

(11) Número de Publicação: **PT 1569961 E**

(51) Classificação Internacional:

C07K 14/61 (2007.10) **C12N 15/63** (2007.10)
C12N 5/10 (2007.10) **C07K 16/26** (2007.10)
C12Q 1/68 (2007.10) **A61K 38/16** (2007.10)
G01N 33/50 (2007.10) **A01K 67/27** (2007.10)

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: **2003.12.05**

(30) Prioridade(s): **2002.12.05 GB 0228441**

(43) Data de publicação do pedido: **2005.09.07**

(45) Data e BPI da concessão: **2010.01.27**
025/2010

(73) Titular(es):

ARES TRADING S.A.
ZONE INDUSTRIELLE DE L'OURIETTAZ 1170
AUBONNE **CH**

(72) Inventor(es):

RICHARD JOSEPH FAGAN **GB**
CHRISTOPHER BENJAMIN PHELPS **GB**
TANIA MARIA RODRIGUES **GB**
MELANIE YORKE-SMITH **CN**
MARIASTELLA DE TIANI **CH**

(74) Mandatário:

MARIA SILVINA VIEIRA PEREIRA FERREIRA
RUA CASTILHO, N.º 50, 5º - ANDAR 1269-163 LISBOA **PT**

(54) Epígrafe: **VARIANTE DE SPLICING DA HORMONA DE CRESCIMENTO PITUITÁRIA HUMANA**

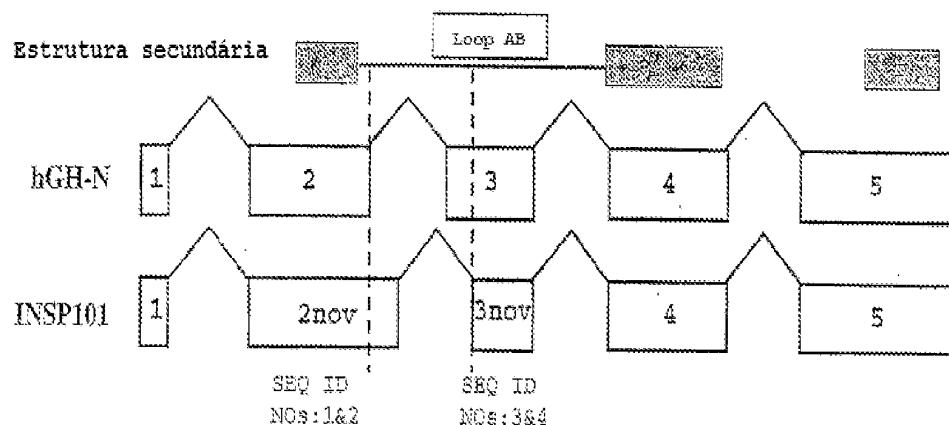
(57) Resumo:

RESUMO

"VARIANTE DE *SPLICING* DA HORMONA DE CRESCIMENTO PITUITÁRIA HUMANA"

A presente invenção é relativa a uma nova proteína, denominada INSP101, identificada no presente documento como uma nova variante de *splicing* (excisão) da hormona de crescimento pituitária humana, e à utilização desta proteína e da sequência de ácidos nucleicos a partir dos genes codificadores no diagnóstico, na prevenção e no tratamento de uma patologia.

Estrutura do Gene



DESCRIÇÃO

"VARIANTE DE *SPLICING* DA HORMONA DE CRESCIMENTO PITUITÁRIA HUMANA"

A presente invenção é relativa a uma nova proteína, denominada INSP101, identificada neste documento como uma nova variante de *splicing* da hormona de crescimento pituitária humana (GH-N; P01241), e à utilização desta proteína e das sequências de ácidos nucleicos a partir do gene codificador no diagnóstico, na prevenção e no tratamento de uma doença. A variante tem um loop A-B alterado e prevê-se, por isso, que possua propriedades alteradas de ligação ao receptor.

ANTECEDENTES

Actualmente, o processo de descoberta de medicamentos está a passar por uma revolução fundamental à medida que a genómica funcional amadurece. O termo "genómica funcional" aplica-se a uma abordagem com recurso a ferramentas de bioinformática para atribuir função às sequências proteicas com interesse. Estas ferramentas são cada vez mais necessárias, à medida que a velocidade de geração de dados de sequências está a ultrapassar a passos largos a capacidade de os laboratórios de investigação atribuírem funções a estas sequências proteicas.

Com o aumento que se verifica da potência e da precisão das ferramentas bioinformáticas, estas estão rapidamente a substituir as técnicas convencionais de caracterização bioquímica. De facto, as ferramentas de bioinformática avançadas, empregues na identificação da presente invenção, são actualmente capazes de fornecer resultados com um

elevado grau de fiabilidade.

Diversas instituições e organizações comerciais vão analisando os dados das sequências à medida que estes vão ficando disponíveis e têm sido feitas descobertas significativas numa base contínua. No entanto, continua a haver a necessidade de identificar e caracterizar mais genes e os polipéptidos por eles codificados, como alvos de investigação e da descoberta de medicamentos.

O *splicing* de pré-ARNm alternativo é um processo celular crucial, pelo qual proteínas com diferentes funções podem ser produzidas a partir do transcrito primário de um único gene, com frequência em padrões específicos de tecidos.

A nível experimental, identifica-se variantes de *splicing* pelo isolamento fortuito e pela subsequente sequenciação de ARNm variantes. No entanto, esta abordagem experimental não foi completada de forma exaustiva para o transcriptoma humano (visto isso requerer o isolamento e a sequenciação sistemáticos de todos os ARNm de todos os tecidos humanos, sob todas as condições ambientais possíveis) e, devido a esta limitação experimental, ainda existe um grande número de variantes de *splicing* por identificar.

Recorreu-se a abordagens de bioinformática patenteada para levar a cabo uma investigação dirigida e objectiva da existência de variantes de *splicing* do gene da hormona de crescimento humana. Com este método, é possível alargar o limitado conjunto de dados de variantes de *splicing* conhecidas a nível experimental, para um conjunto muito maior de variantes de *splicing* previstas.

HORMONAS ENDÓCRINAS

As hormonas regulam uma grande variedade de funções fisiológicas que incluem o metabolismo intermediário, o crescimento e a diferenciação celular. Possuem dois mecanismos fundamentais de ação, em função das respectivas características físico-químicas. As hormonas esteróides lipofílicas e as hormonas tiroideias são hidrofóbicas e actuam em primeiro de forma intracelular, modulando a transcrição genética, enquanto as hormonas peptídicas, como a adrenalina e a melatonina, são hidrofílicas e actuam na membrana celular, desencadeando uma cascata de eventos de transdução de sinal, o que leva a efeitos reguladores a nível intracelular (Lodish et al. (1995) "Molecular Cell Biology", Scientific American Books Inc., Nova Iorque, NY, pág. 856 – 864.

As hormonas são produzidas em células especializadas das glândulas endócrinas e atingem as respectivas células-alvos através da circulação sanguínea. As hormonas esteróides derivam do colesterol por meio de uma série de reacções enzimáticas, que têm lugar no citosol e nas mitocôndrias principalmente de células do córtex adrenal, dos ovários e dos testículos. Em alguns casos, é necessário sujeitar a hormona esteróide a modificação no tecido-alvo, seja para ser activada seja para produzir um derivado mais activo. A maioria das hormonas peptídicas são sintetizadas na forma de proteínas precursoras (pró-hormonas) e são armazenadas na célula endócrina. Antes da sua libertação na circulação, a pró-hormona é clivada na hormona activa. Várias hormonas (principalmente as hormonas esteróides e as hormonas tiroideias) são transportadas na circulação enquanto ligadas a proteínas de ligação específicas. Estas proteínas funcionam como depósitos de hormonas, libertando a hormona

quando necessário e protegendo-a também da sua rápida desactivação.

Devido à natureza central das hormonas na fisiologia geral do *H. sapiens*, a desregulação da função hormonal deu mostras de desempenhar um papel em muitos processos patológicos, incluindo, sem ficarem a ela limitados, a oncologia (Sommer S. and Fuqua S. A. (2001) "Semin Cancer Biol." Out.; 11(5): 339-52, Bartucci M., Morelli C., Mauro L., Ando S., and Surmacz E. (2001) "Cancer Res." Set. 15; 61(18): 6 747-54, Oosthuizen G. M., Joubert G., and du Toit R. S. (2001) "S. Afr. Med. J." Jul.; 91(7): 576-79, Nickerson T., Chang F., Lorimer D., Smeekens S. P., Sawyers C. L., and Pollak M. (2001) "Cancer Res." Agosto 15; 61(16): 6 276-80), "cardiovascular disease" (Liu Y., Ding J., Bush T. L., Longenecker J. C., Nieto F. J., Golden S. H., and Szklo M. (2001) "Am J. Epidemiol." Set. 15; 154(6): 489-94), "metabolic diseases" (Flyvbjerg A. (2001) "Growth Horm. IGF Res." Junho; 11 Supl. A: S115-9, Diamond T., Levy S., Smith A., Day P. and Manoharan A. (2001) "Intern. Med. J." Jul.; 31(5): 272-8, Toprak S., Yonem A., Cakir B., Guler S., Azal O., Ozata M., and Corakci A. (2001) "Horm. Res."; 55(2): 65 - 70, "inflammation" (McEvoy A. N., Bresnihan B., FitzGerald O., and Murphy E. P. (2001) "Arthritis Rheum." Agosto; 44(8): 1761-7, Lipsett P. A. (2001) "Crit. Care Med." Agosto; 29(8): 1642-4) e "CNS related diseases" (Bowen R. L. (2001) "JAMA". Agosto 15; 286(7): 790-1).

Família da hormona de crescimento

A hormona de crescimento (GH) é um membro de uma família de hormonas polipeptídicas, que partilham semelhanças estruturais e actividades biológicas e são produzidas nas glândulas pituitárias de todos os vertebrados e na placenta de alguns mamíferos. Os membros desta família incluem prolactina pituitária, lactogénios placentários (também denominados somatomamotropina coriónica nos humanos [hCS]), proteínas relacionadas com a prolactina em ruminantes e roedores, proliferinas nos ratos e somatolactina nos peixes.

Os genes que codificam a maior parte dos membros da família das GH contêm cinco exões e quatro intrões e parecem ter surgido por duplicação de um único gene ancestral antes do aparecimento dos vertebrados. Foi feita a descrição de variantes de *splicing* e processamento para vários elementos da família.

A família genética relacionada com a GH humana, localizada no cromossoma 17q22-24, consiste num *cluster* (grupo) de genes com grande conservação das suas sequências e num único gene da prolactina no cromossoma 6 (Owerbach D. et al. "Science" 1981). O *cluster* de genes engloba cinco genes estruturais, duas GH e três genes CS, cuja expressão é específica de tecido: hGH-N (N = normal), hGH-V (V = variante), somatomamotropina coriónica humana semelhante à hormona (hCS-L), somatomamotropina coriónica humana A e B (hCS-A e hCS-B) (Misra-Press, A et al. JBC 1994; Boguszewski C. et al. JBC 1998).

A família de proteínas relacionada com a GH tem semelhanças estruturais em comum, dado que a sua estrutura terciária forma quatro hélices α , também conhecida por feixe de quatro hélices antiparalelas. As hélices α estão muito juntas e dispostas numa orientação antiparalela cima-cima-baixo-baixo, com dois *loops* (anéis) compridos ligando os pares paralelos.

A família de genes hGH/hCS é importante na regulação do metabolismo materno e fetal e no crescimento e no desenvolvimento do feto. Durante a gravidez, a expressão da GH pituitária (hGH-N) é suprimida na mãe; e a hGH-V, uma variante da GH expressa pela placenta, torna-se a GH predominante na mãe. A hCS, que é o produto dos genes hCS-A e hCS-B, é secretada para as circulações tanto materna como fetal após as seis semanas de gravidez. A hGH-V e a hCS actuam em parceria na mãe para estimular a produção do factor de crescimento semelhante à insulina (IGF) e para modular o metabolismo intermediário, resultando num aumento da disponibilidade de glucose e de aminoácidos para o feto. No feto, a hCS actua por meio de receptores lactogénicos e possivelmente de um único receptor da CS para modular o desenvolvimento embrionário, regular o metabolismo intermediário e estimular a produção de IGF, insulina, hormonas adrenocorticais e surfactante pulmonar. A hGH-N, que é expressa pela pituitária fetal, tem poucas ou nenhuma actuações fisiológicas no feto até a um estádio avançado da gravidez, devido à falta de receptores da GH funcionais nos tecidos fetais. A hGH-V, que também é uma hormona somatogénica potente, não é libertada no feto. No seu conjunto, os estudos acerca da família de genes hGH/hCS durante a gravidez revelam uma interacção complexa das hormonas entre si e com outros factores de crescimento. São

necessárias investigações adicionais para esclarecer os papéis relativos dos elementos da família na regulação do crescimento e do desenvolvimento fetais e os factores que modulam a expressão dos genes (Handwerger S. & Freemark M. J., "Pediatr. Endocrinol. Metab." 2000, Abril; 13(4): 343-56).

A hormona de crescimento humana, também conhecida por somatotropina, é uma hormona proteica com cerca de 190 aminoácidos, a qual é sintetizada e secretada por células denominadas somatotrofos na pituitária anterior. É um participante crucial no controlo de vários processos fisiológicos complexos, inclusive crescimento e metabolismo. A hormona de crescimento também tem um interesse considerável como medicamento ministrado tanto a humanos como a animais.

A hormona de crescimento produz dois tipos distintos de efeitos. Os efeitos directos resultam da ligação da hormona de crescimento ao seu receptor em células-alvos. As células gordas (adipócitos), por exemplo, têm receptores da hormona de crescimento e a hormona de crescimento estimula-as a decomporem triglicéridos e suprime a sua capacidade de aumentar e acumular lípidos em circulação. Os efeitos indirectos são principalmente mediados pelo factor de crescimento 1 semelhante à insulina (IGF-1). O papel principal da hormona de crescimento na estimulação do crescimento corporal é o de estimular o fígado e outros tecidos a segregar o IGF-1. A maior parte dos efeitos promotores do crescimento da hormona de crescimento deve-se na realidade à actuação do IGF-1 nas suas células-alvos. Por exemplo, o IGF-1 estimula a proliferação dos condrócitos (células das cartilagens), o que leva ao

crescimento ósseo. A hormona de crescimento também produz efeitos importantes no metabolismo proteico, lipídico e dos carboidratos. Em alguns casos, demonstrou-se de forma clara um efeito directo da hormona de crescimento, noutras casos pensa-se que o IGF-1 é o mediador crítico e, em alguns casos, parece que estão em jogo tantos efeitos directos como indirectos.

Além dos seus efeitos complexos no crescimento, os estados tanto de défice como de excesso da hormona de crescimento atestam de forma muito visível o papel desta hormona na fisiologia normal. Estes distúrbios podem reflectir lesões no hipotálamo, na pituitária ou em células-alvos. Um estado de défice poderá resultar não só de um défice de produção da hormona como também da resposta da célula-alvo à hormona.

A nível clínico, um défice da hormona de crescimento ou defeitos do receptor traduzem-se num atraso do crescimento ou no nanismo. A manifestação do défice da hormona de crescimento depende da idade em que o distúrbio é desencadeado e pode ser o resultado de uma patologia hereditária ou adquirida.

O efeito da segregação em excesso da hormona de crescimento também depende muito da idade em que se dá o seu desencadeamento e é visto como dois distúrbios distintos. O gigantismo resulta de uma segregação em excesso da hormona de crescimento que se inicia em crianças jovens ou adolescentes. É um distúrbio muito raro, sendo usualmente resultado de um tumor dos somatotrofos.

A acromegalia resulta de uma segregação em excesso da

hormona de crescimento nos adultos. O desencadeamento deste distúrbio é tipicamente insidioso. A nível clínico, um excessivo crescimento do tecido ósseo e do tecido conjuntivo leva a uma modificação da aparência que poderá ser descrita como "traços fisionómicos grosseiros". Um excesso de hormona de crescimento e IGF-1 também leva a desarranjos metabólicos, inclusive intolerância à glucose.

Há muito que se emprega a hormona de crescimento purificada das pituitárias de cadáveres humanos, para tratar crianças com um atraso grave do crescimento. Mais recentemente, a disponibilidade de hormona de crescimento recombinante levou a diversas outras aplicações a populações humanas e animais. Por exemplo, a hormona de crescimento humana é comumente empregue no tratamento de crianças com uma estatura patologicamente baixa. O papel da hormona de crescimento no envelhecimento normal ainda é pouco compreendido, mas alguns dos sintomas cosméticos do envelhecimento parecem susceptíveis à terapia com a hormona de crescimento. A hormona de crescimento encontra-se actualmente aprovada e é comercializada para aumentar a produção de leite do gado leiteiro; outra aplicação da hormona de crescimento na produção animal é o tratamento de porcos em crescimento com a hormona de crescimento suína. Este tratamento deu mostras de estimular de forma significativa o crescimento muscular e reduzir a deposição de gordura.

Dado que a hormona de crescimento desempenha um papel de tal maneira chave em processos celulares, existe um interesse fulcral no estudo deste meio e do respectivo método de regulação. Seria de grande importância científica a identificação das variantes de *splicing* desta hormona.

Matsuda et al (1988) "Biochimica et Biophysica acta 949", 125 - 131 revela as sequências da variante de *splicing* de 20 kDa da hormona de crescimento humana.

Proctor et al (1998) "Hum. Genet." 103, 255 - 272 é uma análise da estrutura da hormona de crescimento humana e das mutações nela identificadas. São reveladas numerosas substituições de um único par de bases com efeito no ARNm.

A entrada da base de dados OMIM com o N.º de Registo 139250 resume o estado da técnica relativamente à hormona de crescimento humana.

A INVENÇÃO

A invenção tem por base a descoberta de que a proteína INSP101 é uma nova variante de *splicing* da hormona de crescimento pituitária humana (GH-N; P01241).

É feita a descrição dum polipéptido que:

- (i) contém a sequência de aminoácidos indicada na SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 ou SEQ ID NO:10;
- (ii) é um fragmento desta que funciona como hormona de crescimento ou tem um determinante antigénico em comum com um polipéptido de acordo com (i); ou
- (iii) é um equivalente funcional de (i) ou (ii).

Este péptido pode:

- (i) conter a sequência de aminoácidos indicada na SEQ ID NO:8 ou SEQ ID NO:10;

- (ii) ser um fragmento desta que funciona como hormona de crescimento ou tem um determinante antigénico em comum com um polipéptido de acordo com (i); ou
- (iii) ser um equivalente funcional de (i) ou (ii).

É feita a descrição de um polipéptido que consiste na sequência de aminoácidos indicada na SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 ou SEQ ID NO:10.

O polipéptido que tem a sequência indicada na SEQ ID NO:2 é seguidamente designado por "o polipéptido do exão 2nov da INSP101". O polipéptido que tem a sequência definida na SEQ ID NO:4 é seguidamente designado por "o polipéptido do exão 3nov da INSP101". O polipéptido que tem a sequência definida na SEQ ID NO:6 é seguidamente designado por "o polipéptido dos exões 2nov-3nov contíguos da INSP101". O polipéptido que tem a sequência definida na SEQ ID NO:8 é seguidamente designado por "o polipéptido de comprimento total da INSP101". O polipéptido que tem a sequência definida na SEQ ID NO:10 é seguidamente designado por "o polipéptido maduro de comprimento total da INSP101".

A Figura 2 faz a comparação do padrão de *splicing* da P01241, hormona de crescimento pituitária (GH-N) do *H. sapiens*, com o padrão de *splicing* da nova variante de *splicing* INSP101. Com se torna claro com esta figura, a proteína engloba 5 exões separados. O polipéptido do exão 2nov da INSP101 e o polipéptido do exão 3nov da INSP101 são exões alternativos produzidos por *splicing* alternativo; a variante de *splicing* tem um exão2 (2nov) aumentado e um exão3 (3nov) truncado.

A figura também mostra os principais elementos da estrutura secundária da hormona de crescimento pituitária (GH-N). A GH-N é composta por quatro hélices alfa (A, B, C e D), e particularmente importante é o "loop A-B" que liga a hélice A à hélice B. O loop A-B é um componente crítico da superfície de interacção da GH-N, que liga o receptor da Hormona de Crescimento (Wells JA, "PNAS" vol. 93, pág. 1 - 6 1996, "Binding in the growth hormone receptor complex"). É evidente que a nova variante de *splicing* INSP101 tem novos resíduos inseridos no loop A-B (devido à extensão do exão 2). De forma semelhante, o truncamento do exão 3 levará à remoção de alguns resíduos da GH-N no loop A-B. Assim, a INSP101 é diferente da GH-N principalmente em termos da composição do loop A-B e, dado que este loop é um determinante primário na ligação ao receptor da hormona de crescimento, prevê-se que a INSP101 mostre propriedades alteradas de ligação ao receptor (em termos de afinidade de ligação e/ou de selectividade do receptor).

O termo "polipéptidos da INSP101" empregue no pressente documento engloba os polipéptidos que contêm ou são compostos pelo polipéptido do exão 2nov da INSP101, pelo polipéptido do exão 3nov da INSP101, pelo polipéptido dos exões 2nov-3nov contíguos da INSP101, pelo polipéptido de comprimento total da INSP101 e pelo polipéptido maduro de comprimento total da INSP101.

"Funcionamento como hormona de crescimento" refere-se a polipéptidos que contêm características de sequência de aminoácidos ou características estruturais que podem ser identificadas como características conservadas no âmbito das hormonas de crescimento humanas, de tal modo que a

actividade do polipéptido não seja substancialmente afectada de forma negativa em comparação com o funcionamento do polipéptido tipo selvagem de comprimento total. Por exemplo, poderá recorrer-se a uma série de diferentes testes para determinar os efeitos das hormonas de crescimento humanas na ligação (ver, por exemplo, Well J. A., "PNAS" vol. 93, pág. 1-6, 1996), inclusive a utilização de anticorpos monoclonais para precipitar complexos de 1 : 1 de hormona de crescimento e receptor, e a dimerização induzida pela hGH de moléculas hGHbp em solução por extinção de um tag fluorescente colocado perto da terminação C do hGHbp (ver Well J. A., "PNAS" vol. 93, pág. 1-6, 1996). [hGHbp = domínio extracelular do receptor da GH].

Encontram-se disponíveis vários testes que permitem detectar a actividade da hormona de crescimento. Estes incluem os seguintes testes endocrinológicos metabólicos:

i) Teste de diferenciação em adipócito:

Crê-se ser importante a inibição da diferenciação de adipócitos, num modelo *in vitro* para redução da massa adiposa, na redução da resistência à insulina no caso de doenças como a diabetes e a síndrome do ovário policístico (SOP). O objectivo é identificar uma proteína ou proteínas que inibam a diferenciação dos pré-adipócitos em adipócitos. Induz-se a linha celular de pré-adipócitos murinos 3T3-L1 para que estes se diferenciem em adipócitos com insulina + IBMX. A descoberta de que a diferenciação é inibida por TNF α + ciclo-hexamida é utilizada como controlo positivo.

ii) Aumento da glucose tritiada (3T3 L1):

O objectivo é identificar uma proteína ou proteínas que estimule(m) o aumento de glucose como modelo da resistência à insulina na adipose durante a diabetes ou a SOP. Os adipócitos empregues são os pré-adipócitos murinos 3T3-L1 que sofreram diferenciação.

iii) Aumento da glucose tritiada (adipócitos humanos primários):

O objectivo é identificar uma proteína ou proteínas que estimule(m) o aumento de glucose como modelo da resistência à insulina na adipose durante a diabetes ou a SOP. Empregam-se adipócitos humanos primários.

iv) Aumento da glucose tritiada (células musculares esqueléticas humanas primárias):

O objectivo é identificar uma proteína ou proteínas que estimule(m) o aumento de glucose como modelo da resistência à insulina no tecido muscular durante a diabetes ou a SOP. Diferencia-se células musculares esqueléticas humanas primárias em miotubos e aplica-se as mesmas em seguida no teste.

A invenção fornece uma molécula de ácidos nucleicos purificada que codifica um polipéptido da invenção.

A molécula de ácidos nucleicos purificada pode conter a sequência de ácidos nucleicos definida na SEQ ID NO:1 (que codifica o polipéptido do exão 2nov da INSP101), a SEQ ID NO:3 (que codifica o polipéptido do exão 3nov da INSP101), a SEQ ID NO:5 (que codifica o polipéptido dos exões 2nov-3nov contíguos da INSP101), a SEQ ID NO:7 (que codifica o polipéptido de comprimento total da INSP101) ou a SEQ ID NO:9 (que codifica o polipéptido maduro de comprimento

total da INSP101), ou é um equivalente redundante ou um fragmento de qualquer uma destas sequências.

A molécula de ácidos nucleicos purificada pode consistir na sequência de ácidos nucleicos definida na SEQ ID NO:1 (que codifica o polipéptido do exão 2nov da INSP101), na SEQ ID NO:3 (que codifica o polipéptido do exão 3nov da INSP101), na SEQ ID NO:5 (que codifica o polipéptido dos exões 2nov-3nov contíguos da INSP101), na SEQ ID NO:7 (que codifica o polipéptido de comprimento total da INSP101) ou na SEQ ID NO:9 (que codifica o polipéptido maduro de comprimento total da INSP101), ou é um fragmento ou um equivalente redundante de qualquer uma destas sequências.

A sequência codificadora que codifica a hormona de crescimento humana pituitária do tipo selvagem, NM_000515, está especificamente excluída do âmbito da presente invenção.

É feita a descrição de uma molécula de ácidos nucleicos purificada, que sofre hibridação sob condições altamente estritas com uma molécula de ácidos nucleicos da invenção.

A invenção fornece um vector, como um vector de expressão, que contém uma molécula de ácidos nucleicos da invenção.

A invenção fornece uma célula hospedeira transformada com um vector da invenção.

É feita a descrição de um ligando que se liga de forma específica às hormonas de crescimento do primeiro aspecto da invenção. Preferencialmente, o ligando inibe a função de um polipéptido da invenção que é uma hormona de

crescimento. Os ligandos para um polipéptido de acordo com a invenção podem assumir diversas formas, inclusive substratos naturais ou modificados, enzimas, receptores, pequenas moléculas orgânicas, tais como pequenas moléculas orgânicas naturais ou sintéticas de até 2 000 Da, de preferência 800 Da ou menos, peptidomiméticos, moléculas inorgânicas, péptidos, polipéptidos, anticorpos, miméticos estruturais ou funcionais dos supracitados.

É feita a descrição de um composto eficaz na modificação da expressão de um gene natural que codifica um polipéptido da invenção, ou na regulação da actividade de um polipéptido da invenção.

Um composto deste tipo poderá aumentar (agonizar) ou diminuir (antagonizar) o nível de expressão do gene ou a actividade do polipéptido.

De forma importante, a identificação da função do polipéptido INSP101 permite a concepção de métodos de rastreio, capazes de identificar compostos que sejam eficazes no tratamento e/ou no diagnóstico de patologias. É possível identificar os ligandos e os compostos de acordo com a invenção recorrendo a estes métodos. Estes métodos encontram-se englobados como aspectos da presente invenção.

A invenção fornece um polipéptido da invenção ou uma molécula de ácidos nucleicos da invenção, ou um vector da invenção, ou um ligando da invenção, a ser empregue na terapia ou no diagnóstico de uma patologia em que esteja implicada a hormona de crescimento humana. Estas patologias e estes distúrbios poderão incluir problemas reprodutivos, problemas de gravidez, como a doença trofoblástica

gestacional, problemas de desenvolvimento como o sindroma de Silver-Russel, problemas de crescimento, défice da hormona de crescimento, doença de Cushing, problemas endócrinos, distúrbios proliferativos celulares, incluindo neoplasma, carcinoma, tumor da pituitária, tumor dos ovários, melanoma, tumor do pulmão, colo-rectal, da mama, do pâncreas, da cabeça e do pescoço, tumor trofoblástico de localizaçāo placentária, adenocarcinoma, coriocarcinoma, osteossarcoma e outros tumores sólidos; angiogéneses, distúrbios mieloproliferativos; distúrbios autoimunes/inflamatórios; problemas cardiovasculares; distúrbios neurológicos, dor; distúrbios metabólicas inclusive diabetes mellitus, osteoporose e obesidade, caquexia, SIDA, doença renal; lesão pulmonar; envelhecimento; infecções incluindo as virais, bacterianas, fúngicas e parasitárias, e outras condições patológicas. De preferência, a patologia trata-se de uma patologia em que esteja envolvida a função endócrina, em particular hormonas de crescimento (ver, por exemplo, Arato G., Fulop V., Degrell P., Szigetvari J. "Pathol. Oncol. Res." 2000 6(4): 292-4; Hitchins M. P., Stanier P., Preece M. A. and Moore GE., "J. Med. Genet" 2001 Dez. 38(12): 810-9; Rhoton-Vlasak A., Wagner J. M., Rutgers J. L, Baergen R. N., Young R. H., Roche P. C., Plummer T. B. and Gleich G. J., "Hum Pathol" 1998 Março 29(3): 280-8; Llovera M., Pichard C., Bernichttein S., Jeay S., Touraine P., Kelly P. A. and Goffin V., "Oncogene", 2000 Setembro 28 19(41): 4 695-705; Savage M. O., Scommegna S., Carroll P. V., Ho J. T., Monson J. P., Besser G. M. and Grossman AB., "Horm. Res." 2002 58 Supl. 1: 39 - 43; Aimaretti G., Corneli G., Bellone S., Baffoni C., Camanni F. and Ghigo E., "J. Pediatr. Endocrinol. Metab." 2001 14 Supl. 5: 1 233-42; Berger P., Untergasser G., Hermann M., Hittmair A., Madersbacher S.

and Dirnhofer S., "Hum. Pathol." 1999 Outubro 30(10): 1 201-6; Hamilton J., Chitayat D., Blaser S., Cohen L. E., Phillips J. A. 3rd and Daneman D., "Am. J. Med. Genet." 1998 Novembro 2 80(2): 128-32; Gonzalez-Rodriguez E., Jaramillo-Rangel G. and Barrera-Saldana H. A., "Am. J. Med. Genet." 1997 Novembro 12 72(4): 399 - 402; Perez Jurado L. A., Argente J., Barrios V., Pozo J., Munoz M. T., Hernandez M. and Francke U., "J. Pediatr. Endocrinol. Metab." 1997 Março - Abril 10(2): 185-90; Saeger W. and Lubke D., "Endocr. Pathol." 1996 Primavera 7(1): 21 - 35; Conzemius M. G., Graham J. C., Haynes J. S. and Graham C. A., "Am. J. Vet. Res." 2000 Junho 61(6): 646-50; Bartlett D. L., Charland S., Torosian M. H., "Cancer" 1994 Março 1 73(5): 1 499-504). Também se pode utilizar estas moléculas na produção de um medicamento destinado ao tratamento destes distúrbios.

É feita a descrição de um método de diagnóstico de uma patologia num doente, incluindo a avaliação do nível de expressão de um gene natural que codifica um polipéptido da invenção ou a actividade de um polipéptido da invenção no tecido do referido doente, e a comparação do referido nível de expressão ou de actividade com um nível de controlo, sendo que um nível diferente do referido nível de controlo é indicador de doença. Um método destes será de preferência levado a cabo *in vitro*. É possível recorrer a métodos similares na monitorização do tratamento terapêutico da doença no doente, em que uma alteração do nível de expressão ou de actividade de um polipéptido ou de uma molécula de ácidos nucleicos durante o período de tempo em direcção a um nível de controlo é indicador da regressão da doença.

Um método preferido de detecção de polipéptidos da invenção engloba as etapas de: (a) fazer contactar um ligando, como um anticorpo, do sexto aspecto da invenção, com uma amostra biológica sob condições apropriadas para a formação de um complexo de ligando-polipéptido; e (b) detectar o referido complexo.

Existe uma série de métodos diferentes deste diagnóstico, como é do conhecimento do leitor especializado, como se verifica com os métodos de hibridação de ácidos nucleicos com sondas curtas, análise de mutação pontual, amplificação por reacção em cadeia de polimerase (PCR) e métodos empregando anticorpos para detectar níveis proteicos aberrantes. É possível empregar métodos similares a curto ou a longo prazo para permitir o tratamento terapêutico de uma doença a ser monitorizada num doente. É igualmente feita a descrição de kits que sejam úteis nestes métodos de diagnóstico de patologias.

A invenção contempla a utilização de polipéptidos do primeiro aspecto da invenção como hormona de crescimento ou como um modulador da actividade da hormona de crescimento. Os usos apropriados dos polipéptidos da invenção como hormonas de crescimento incluem o uso como regulador do crescimento celular, do metabolismo ou da diferenciação, o uso como parte de um par de receptor/ligando e o uso como um marcador de diagnóstico de uma condição fisiológica ou patológica. Como acima mencionado, a hormona de crescimento estimula a decomposição dos triglicéridos nos adipócitos. A hormona de crescimento humana também estimula o crescimento corporal ao estimular o fígado a segregar o IGF-1. Isto, por sua vez, resulta na proliferação de condrócitos, sendo o resultado o crescimento ósseo. A hormona de crescimento

também produz efeitos no metabolismo. A administração da hormona de crescimento constitui um tratamento conhecido do nanismo. Também se conhece a sua utilização para aumentar a produção de leite no gado leiteiro e para aumentar o crescimento dos porcos. Pode-se explorar todas estas utilidades por meio dos polipéptidos da invenção. Além disso, é possível empregar antagonistas dos polipéptidos da invenção no tratamento da sobreprodução da hormona de crescimento, que poderá resultar em gigantismo.

Como acima debatido, poderá recorrer-se a uma série de testes diferentes para determinar os efeitos das hormonas de crescimento humanas na ligação (ver, por exemplo, Well J. A., "PNAS" vol. 93 pág. 1-6, 1996), inclusive a utilização de anticorpos monoclonais para precipitar complexos de 1 : 1 de hormona de crescimento e receptor, e a dimerização induzida por hGH de moléculas hGHbp em solução, pela extinção de um tag fluorescente colocado perto da terminação C do hGHbp (ver Well J. A. "PNAS" vol. 93 pág. 1-6, 1996). [hGHbp = domínio extracelular do receptor da GH]

A invenção proporciona uma composição farmacêutica contendo um polipéptido da invenção, ou uma molécula de ácidos nucleicos da invenção, ou um vector da invenção ou um ligando da invenção em conjunção com uma substância de suporte aceitável em farmácia.

A presente invenção proporciona um polipéptido da invenção, ou uma molécula de ácidos nucleicos da invenção, ou um vector da invenção ou um ligando da invenção para fins terapêuticos ou de diagnóstico. Também é possível empregar estas moléculas na produção de um medicamento destinado ao

tratamento de uma patologia.

A invenção proporciona um método de tratar uma patologia num doente, o qual engloba a administração ao doente de um polipéptido da invenção, ou de uma molécula de ácidos nucleicos da invenção, ou de um vector da invenção ou de um ligando da invenção.

No caso de patologias em que a expressão de um gene natural, que codifica um polipéptido da invenção, ou em que a actividade de um polipéptido da invenção é mais baixa num doente com essa patologia do que o nível de expressão ou a actividade num paciente saudável, o polipéptido, a molécula de ácidos nucleicos ou o ligando que se ministra ao doente deverá ser um agonista. Em torno, no caso de patologias em que a expressão do gene natural ou a actividade do polipéptido é maior num doente com essa patologia do que o nível de expressão ou a actividade num paciente saudável, o polipéptido, a molécula de ácidos nucleicos ou o ligando ministrado/a ao doente deverá ser um antagonista. Os exemplos destes antagonistas poderão ser moléculas de ácidos nucleicos anti-sentido, ribozimas e ligandos, como os anticorpos.

É feita a descrição de animais não-humanos transgénicos ou *knockout* (= com um gene desactivado) que foram transformados para expressar níveis mais elevados, mais baixos ou inexistentes de um polipéptido da invenção. Estes animais transgénicos são modelos muito úteis no estudo de patologias e também podem ser empregues em regimes de rastreio para identificar compostos que sejam eficazes no tratamento ou no diagnóstico dessas patologias.

Em baixo encontra-se um resumo de técnicas e procedimentos padrão a que se pode recorrer de modo a aplicar a invenção. A presente invenção não ficará limitada à metodologia, aos protocolos, às linhas celulares, aos vectores e aos reagentes particulares descritos. Também a terminologia empregue no presente documento se destina somente a descrever as execuções particulares e não tem por função limitar o âmbito da presente invenção. O âmbito da invenção é apenas limitado pelos termos das revindicações anexas.

Emprega-se as abreviaturas padrão para os nucleótidos e para os aminoácidos nesta especificação.

A prática da presente invenção recorrerá, se nada for expresso em contrário, a técnicas convencionais da biologia molecular, da microbiologia, da tecnologia de ADN recombinante e da imunologia, que se encontram dentro da especialidade dos que trabalham na técnica.

Estas técnicas encontram-se totalmente explicadas na literatura. Os exemplos de textos particularmente apropriados para consulta incluem os seguintes: Sambrook "Molecular Cloning; A Laboratory Manual", 2.^a edição (1989); "DNA Cloning", Volumes I e II (D. N. Glover ed. 1985); "Oligonucleotide Synthesis" (M. J. Gait ed. 1984); "Nucleic Acid Hybridization" (B. D. Hames & S. J. Higgins eds. 1984); "Transcription and Translation" (B. D. Hames & S. J. Higgins eds. 1984); "Animal Cell Culture" (R. I. Freshney ed. 1986); "Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, "A Practical Guide to Molecular Cloning" (1984); "the Methods in Enzymology series" (Academic Press, Inc.) em especial os volumes 154 e 155; "Gene Transfer Vectors for Mammalian Cells" (J. H. Miller

and M. P. Calos eds. 1987, Cold Spring Harbor Laboratory); "Immunochemical Methods in Cell and Molecular Biology" (Mayer and Walker, eds. 1987, Academic Press, Londres); Scopes, (1987) "Protein Purification: Principles and Practice", 2.^a edição (Springer Verlag, Nova Iorque); e "Handbook of Experimental Immunology", volumes I - IV (D. M. Weir and C. C. Blackwell eds. 1986).

Como empregue no presente texto, o termo "polipéptido" engloba todos os péptidos ou proteínas que contenham dois ou mais aminoácidos ligados entre si por ligações peptídicas ou por ligações peptídicas modificadas, ou seja, isósteros peptídicos. O termo é relativo tanto a cadeias curtas (péptidos e oligopéptidos) como a cadeias mais compridas (proteínas).

O polipéptido da presente invenção poderá encontrar-se na forma de uma proteína madura ou poderá ser uma pré-proteína, uma pró-proteína ou uma pré-pró-pró-proteína que poderá ser activada por clivagem da pré-porção, da pró-porção ou da pré-pró-porção para produzir um polipéptido maduro activo. Nestes polipéptidos, a pré-sequência, a pró-sequência ou a pré-pró-sequência poderá ser uma sequência líder ou secretora, ou uma sequência que é empregue na purificação da sequência polipeptídica madura.

O polipéptido descrito poderá formar parte de uma proteína de fusão. Por exemplo, é com frequência vantajoso incluir uma ou mais sequências adicionais de aminoácidos, que poderão conter sequências secretoras ou líderes, pró-sequências, sequências que auxiliam à purificação, ou sequências que conferem uma maior estabilidade proteica, por exemplo, durante a produção recombinante. Em

alternativa ou adicionalmente, o polipéptido maduro poderá ser fundido com outro composto, como um composto para aumentar o tempo de semi-vida do polipéptido (por exemplo, polietilenoglicol).

Os polipéptidos poderão conter aminoácidos diferentes dos 20 aminoácidos codificados pelos genes, modificados por processos naturais como o processamento pós-tradução ou por técnicas de modificação química bem conhecidas da técnica. Entre as modificações conhecidas que são mais comuns estarem presentes nos polipéptidos tem-se glicosilação, ligação lipídica, sulfatação, gama-carboxilação, por exemplo, de resíduos do ácido glutâmico, hidroxilação e ADP-ribosilação. Outras modificações potenciais incluem acetilação, acilação, amidação, ligação covalente da flavina, ligação covalente de um grupo heme, ligação covalente de um nucleótido ou de um derivado nucleotídico, ligação covalente de um derivado lipídico, ligação covalente do fosfatidilinositol, ligação cruzada, ciclização, formação de ligação dissulfureto, demetilação, formação de ligações cruzadas covalentes, formação da cisteína, formação do piroglutamato, formilação, formação da âncora GPI, iodoação, metilação, miristoilação, oxidação, processamento proteolítico, fosforilação, fenilação, racemização, selenoilação, adição mediada por ARN de transferência de aminoácidos a proteínas, tais como arginilação e ubiquitinação.

As modificações podem ocorrer em qualquer ponto de um polipéptido, inclusive na cadeia principal do polipéptido, nas cadeias laterais de aminoácidos e nas terminações amina ou carboxilo. De facto, o bloqueio da terminação amina ou da terminação carboxilo num polipéptido, ou em ambos, por

modificação covalente é comum em polipéptidos que ocorrem naturalmente ou sintéticos.

As modificações que ocorrem num polipéptido serão com frequência em função de como o polipéptido é constituído. No caso dos polipéptidos constituídos de forma recombinante, a natureza e a extensão das modificações serão determinadas em grande parte pela capacidade de modificação pós-tradução da célula hospedeira particular e pelos sinais de modificação que estão presentes na sequência de aminoácidos do polipéptido em questão. Por exemplo, os padrões de glicosilação variam entre diferentes tipos de células hospedeiras.

É possível preparar os polipéptidos da presente invenção de qualquer maneira adequada. Estes polipéptidos incluem polipéptidos que ocorrem naturalmente isolados (por exemplo, purificados a partir de uma cultura celular), polipéptidos produzidos de forma recombinante (incluindo proteínas de fusão), polipéptidos produzidos de forma sintética ou polipéptidos produzidos pela combinação destes métodos.

Os polipéptidos funcionalmente equivalentes podem ser polipéptidos homólogos aos polipéptidos INSP101. Diz-se que dois polipéptidos são "homólogos", tal como o termo é aqui empregue, se a sequência de um dos polipéptidos tiver um grau suficientemente grande de identidade ou de semelhança com a sequência do outro polipéptido. A "identidade" indica que, em qualquer posição particular nas sequências alinhadas, o resíduo de aminoácidos é idêntico entre as sequências. "Semelhança" indica que, em qualquer posição particular nas sequências alinhadas, o resíduo de

aminoácidos é de tipo semelhante entre as sequências. Os graus de identidade e de semelhança podem ser facilmente calculados ("Computational Molecular Biology", Lesk, A. M., ed., Oxford University Press, Nova Iorque, 1988; "Biocomputing. Informatics and Genome Projects", Smith, D. W., ed., Academic Press, Nova Iorque, 1993; "Computer Analysis of Sequence Data", Parte 1, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; "Sequence Analysis in Molecular Biology", von Heinje, G., Academic Press, 1987; e "Sequence Analysis Primer", Gribskov, M. and Devereux, J. eds., M Stockton Press, Nova Iorque, 1991).

Por conseguinte, os polipéptidos homólogos incluem variantes biológicas naturais (por exemplo, variantes alélicas ou variações geográficas dentro da espécie de que derivam os polipéptidos) e mutantes (como os mutantes que contêm substituições, inserções ou deleções de aminoácidos) dos polipéptidos INSP101. Estes mutantes poderão incluir polipéptidos nos quais um ou mais dos resíduos de aminoácidos estão substituídos por um resíduo de aminoácidos conservado ou não conservado (de preferência um resíduo conservado de aminoácidos), e esse resíduo de aminoácidos substituído poderá ou não ser codificado pelo código genético. Tipicamente, estas substituições são entre Ala, Val, Leu e Ile; entre Ser e Thr; entre os resíduos ácidos Asp e Glu; entre Asn e Gln; entre os resíduos básicos Lys e Arg; ou entre os resíduos aromáticos Phe e Tyr. Em particular preferidas são as variantes em que vários aminoácidos, ou seja, entre 5 e 10, 1 e 5, 1 e 3, 1 e 2 ou apenas 1, são substituídos, delectados ou adicionados em qualquer combinação. Prefere-se em especial as substituições, adições e deleções

silenciosas, que não alterem as propriedades e as actividades das proteínas. Relativamente a isto, também se prefere em especial as substituições conservadoras. Estes mutantes também incluem os polipéptidos em que um ou mais dos resíduos de aminoácidos inclua um grupo substituinte.

Tipicamente, uma identidade superior a 30% entre dois polipéptidos é considerada ser indicadora de equivalência funcional. De preferência, os polipéptidos funcionalmente equivalentes descritos têm um grau de identidade de sequência com o polipéptido INSP101, ou com os respectivos fragmentos activos, superior a 90% em todo o comprimento da sequência do INSP101. Os polipéptidos mais preferidos têm graus de identidade superiores a 92%, 95%, 98% ou 99% em todo o comprimento da sequência do INSP101, respectivamente.

Os polipéptidos funcionalmente equivalentes descritos também podem ser polipéptidos que foram identificados recorrendo a uma ou mais técnicas de alinhamento estrutural. Por exemplo, pode empregar-se a tecnologia Inpharmatica Genome ThreaderTM que constitui um aspecto das ferramentas de pesquisa empregues para gerar a base de dados de pesquisa Biopendium (ver o Pedido co-pendente de Patente Internacional N.º PCT/GB01/01105), para identificar os polipéptidos com uma função presentemente desconhecida que, embora tendo uma baixa identidade de sequência em relação o polipéptido INSP101, se prevê tratar de proteínas de hormonas de crescimento, indo o referido método empregar um polipéptido do primeiro aspecto da invenção por partilhar homologia estrutural significativa com as sequências do polipéptido INSP101. "Homologia estrutural significativa" significa que a Inpharmatica Genome

Threader™ prevê que duas proteínas partilham homologia estrutural com uma certeza de pelo menos 10% ou mais.

Os polipéptidos descritos também incluem os fragmentos dos polipéptidos INSP101 e os fragmentos dos equivalentes funcionais dos polipéptidos INSP101, desde que esses fragmentos conservem actividade de hormona de crescimento ou tenham um determinante antigénico em comum com os polipéptidos INSP101.

Como aplicado no presente documento, o termo "fragmento" refere-se a um polipéptido tendo uma sequência de aminoácidos que é igual a parte, mas não a toda, a sequência de aminoácidos dos polipéptidos INSP101 ou de um dos seus equivalentes funcionais. Os fragmentos deverão conter pelo menos n aminoácidos consecutivos da sequência e, em função da sequência particular, n é de preferência 7 ou mais (por exemplo, 8, 10, 12, 14, 16, 18, 20 ou mais). Os fragmentos pequenos poderão formar um determinante antigénico.

Os fragmentos dos polipéptidos INSP101 de comprimento total poderão ser compostos por combinações de 2 ou 3 sequências de exões contíguas nas sequências dos polipéptidos INSP101, respectivamente. Por exemplo, estas combinações incluem os exões 1 e 2, 2 e 3 ou 1, 2 e 3.

Estes fragmentos poderão ser "autónomos", ou seja, não serem parte ou não estarem fundidos outros aminoácidos ou polipéptidos, ou podem estar contidos num polipéptido maior do qual formam uma parte ou uma região. Se contido num polipéptido maior, o fragmento da invenção formará com maior preferência uma única região contínua. Por exemplo,

certas execuções preferidas são relativas a um fragmento tendo uma região de pré-polipéptido e/ou pró-polipéptido fundida na terminação amina do fragmento e/ou uma região adicional fundida na terminação carboxilo do fragmento. No entanto, é possível que vários fragmentos estejam contidos num único polipéptido maior.

É possível utilizar os polipéptidos da pressente invenção ou os seus fragmentos imunogénicos (contendo pelo menos um determinante antigénico) para gerar ligandos, tais como os anticorpos policlonais ou monoclonais, que são imunoespecíficos em relação aos polipéptidos. Pode empregar-se estes anticorpos para isolar ou para identificar os clones que expressam os polipéptidos, ou para purificar os polipéptidos por cromatografia de afinidade. Também é possível utilizar os anticorpos como auxiliares de diagnóstico ou auxiliares terapêuticos, entre outras aplicações, como será aparente ao leitor especialista.

O termo "imunoespecífico" significa que os anticorpos têm uma afinidade substancialmente maior em relação aos polipéptidos da invenção do que a afinidade em relação a outros polipéptidos relacionados no estado da técnica. Como aplicado no presente contexto, o termo "anticorpo" refere-se a moléculas intactas assim como a fragmentos destas, tais como Fab , $\text{F}(\text{ab}')2$ e Fv , que são capazes de se ligar ao determinante antigénico em questão. Estes anticorpos ligam-se assim aos polipéptidos do primeiro aspecto da invenção.

"Afinidade substancialmente maior" significa que existe um aumento mensurável da afinidade em relação a um polipéptido da invenção em comparação com a afinidade em relação a

hormonas de crescimento humanas conhecidas.

Preferencialmente, a afinidade é de pelo menos 1,5, 2, 5, 10, 100, 10^3 , 10^4 , 10^5 , 10^6 (ou mais) vezes maior em relação a um polipéptido da invenção do que em relação a proteínas segregadas conhecidas como as hormonas de crescimento.

Caso se pretenda anticorpos policlonais, poderá imunizar-se um mamífero escolhido, como um rato, um coelho, uma cabra ou um cavalo, com um polipéptido. O polipéptido empregue na imunização do animal pode derivar da tecnologia de ADN recombinante ou pode ser sintetizado por via química. Se pretendido, é possível conjugar o polipéptido com uma proteína de suporte. Os suportes usualmente empregues, aos quais se podem ligar quimicamente os polipéptidos, incluem albumina de soro bovino, tiroglobulina e hemocianina do *Megathura crenulata* (*keyhole limpet haemocyanin* = KLH). Emprega-se depois o polipéptido ligado na imunização do animal. Colhe-se o soro do animal imunizado e trata-se o mesmo de acordo com procedimentos conhecidos, por exemplo, por cromatografia de imunoafinidade.

Um especialista da técnica também consegue facilmente produzir anticorpos monoclonais para os polipéptidos do primeiro aspecto da invenção. É bem conhecida a metodologia geral para marcar anticorpos monoclonais recorrendo à tecnologia dos hibridomas (ver, por exemplo, Kohler, G. and Milstein, C., "Nature" 256: 495-497 (1975); Kozbor *et al.*, "Immunology Today" 4: 72 (1983); Cole *et al.*, 77 - 96 in "Monoclonal Antibodies and Cancer Therapy", Alan R. Liss, Inc. (1985).

É possível rastrear painéis de anticorpos monoclonais,

produzidos contra os polipéptidos da invenção, em busca de diversas propriedades, ou seja, isotipo, epítopo, afinidade, etc. Os anticorpos monoclonais são particularmente úteis na purificação de polipéptidos individuais contra os quais se dirigem. Em alternativa, os genes que codificam os anticorpos monoclonais em questão podem ser isolados de hibridomas, por exemplo, por meio de técnicas de PCR conhecidas da técnica, e ser clonados e expressos em vectores apropriados.

Também poderão ser úteis os anticorpos quiméricos, nos quais regiões variáveis não-humanas estão juntas ou fundidas a regiões constantes humanas (ver, por exemplo, Liu *et al.*, "Proc. Natl. Acad. Sci. USA", 84, 3 439 (1987)).

É possível modificar o anticorpo para o tornar menos imunogénico num indivíduo, por exemplo, por humanização (ver Jones *et al.*, "Nature", 321, 522 (1986); Verhoeven *et al.*, "Science", 239, 1 534 (1988); Kabat *et al.*, "J. Immunol.", 147, 1 709 (1991); Queen *et al.*, "Proc. Natl. Acad. Sci. USA", 86, 10 029 (1989); Gorman *et al.*, "Proc. Natl Acad. Sci. USA", 88, 34 181 (1991); e Hodgson *et al.*, "Bio/Technology", 9, 421 (1991)). O termo "anticorpo humanizado", como empregue no presente documento, refere-se a moléculas de anticorpos em que os aminoácidos CDR e outros aminoácidos seleccionados nos domínios variáveis das cadeias pesadas e/ou leves de um anticorpo dador não-humano foram substituídos em lugar dos aminoácidos equivalentes num anticorpo humano. Assim, o anticorpo humanizado assemelha-se de muito perto a um anticorpo humano, mas tem a capacidade de ligação do anticorpo dador.

Noutra alternativa, o anticorpo pode ser um anticorpo "biespecífico", ou seja, um anticorpo que tem dois domínios de ligação de抗énios diferentes, sendo cada domínio dirigido contra um epítopo diferente.

É possível empregar a tecnologia de exibição de fagos para seleccionar genes que codifiquem anticorpos com actividades de ligação em relação aos polipéptidos da invenção, seja a partir de reportórios de genes V amplificados por PCR de linfócitos de humanos rastreados quanto a terem os anticorpos relevantes, ou de bibliotecas naives (McCafferty, J. et al., (1990), "Nature" 348, 552-554; Marks, J. et al., (1992) "Biotechnology" 10, 779-783). A afinidade destes anticorpos também pode ser melhorada por *shuffling* (recombinação) de cadeias (Clackson, T. et al., (1991) "Nature" 352, 624-628).

Os anticorpos gerados pelas técnicas acima referidas, sejam policlonais ou monoclonais, têm uma utilidade adicional na medida em que podem ser empregues como reagentes em testes de imunidade, testes de radioimunidade (RIA) ou ensaios de imunoabsorção enzimática (ELISA). Nestas aplicações, pode-se titular os anticorpos com um reagente de detecção analítica como um radioisótopo, uma molécula fluorescente ou uma enzima.

As moléculas de ácidos nucleicos preferidas da invenção são as que codificam uma sequência polipeptídica como definida pela SEQ ID NO:8 ou SEQ ID NO:10, e polipéptidos funcionalmente equivalentes. Também se encontram descritas moléculas de ácidos nucleicos que codificam uma sequência polipeptídica como definida nas SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 e polipéptidos funcionalmente equivalentes. É

possível empregar estas moléculas de ácidos nucleicos nos métodos e nas aplicações descritos no presente documento. As moléculas de ácidos nucleicos contêm, de preferência, pelo menos n nucleótidos consecutivos das sequências aqui reveladas, em que, em função da sequência particular, n é 10 ou mais (por exemplo, 12, 14, 15, 18, 20, 25, 30, 35, 40 ou mais).

As moléculas de ácidos nucleicos também englobam sequências que são complementares às moléculas de ácidos nucleicos acima descritas (por exemplo, para fins de anti-sentido ou de teste).

As moléculas de ácidos nucleicos da presente invenção poderão encontrar-se na forma de ARN, como o ARNm, ou na forma de ADN, incluindo, por exemplo, ADNc, ADN sintético ou ADN genómico. Consegue-se obter estas moléculas de ácidos nucleicos por clonagem, por técnicas de síntese química ou pela combinação das mesmas. É possível preparar as moléculas de ácidos nucleicos, por exemplo, por síntese química recorrendo a técnicas como a síntese química de fosforamidite em fase sólida, de bibliotecas de ADNc ou genómicas ou pela separação de um organismo. As moléculas de ARN podem ser em geral geradas pela transcrição *in vitro* ou *in vivo* de sequências de ADN.

As moléculas de ácidos nucleicos podem ser de cadeia dupla ou de cadeia simples. O ADN de cadeia simples pode ser a cadeia codificadora, também conhecida pela cadeia sentido, ou pode ser a cadeia não codificadora, também denominada cadeia anti-sentido.

O termo "molécula de ácidos nucleicos" também engloba

análogos do ADN e do ARN, como os que contêm cadeias principais modificadas, e ácidos nucleicos peptídicos (PNA). O termo "PNA", como aqui aplicado, refere-se a uma molécula anti-sentido ou a um agente antigénico, que contém um oligonucleótido com pelo menos cinco nucleótidos de comprimento ligados a uma cadeia principal peptídica de resíduos de aminoácidos, que termina de preferência em lisina. O terminal lisina confere solubilidade à composição. Os PNA poderão ser peguilados para aumentar o seu tempo de vida na célula, onde de preferência ligam o ADN de cadeia simples complementar e o ARN e param o alongamento da transcrição (Nielsen, P. E. et al. (1993) "Anticancer Drug Des." 8: 53-63).

Uma molécula de ácidos nucleicos, que codifica um polipéptido da presente invenção, poderá ser idêntica à sequência codificadora de uma ou mais moléculas de ácidos nucleicos aqui reveladas.

Estas moléculas também poderão ter uma sequência diferente que, em resultado da degeneração do código genético, codifica um polipéptido da SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 ou SEQ ID NO:10.

Estas moléculas de ácidos nucleicos poderão incluir, sem ficaram a elas limitadas, a sequência codificadora do polipéptido maduro por ele próprio; a sequência codificadora do polipéptido maduro e sequências codificadoras adicionais, como as que codificam uma sequência líder ou secretora, como uma sequência pró-polipeptídica, pré-polipeptídica ou pré-pró-polipeptídica; a sequência codificadora do polipéptido maduro, com ou sem as sequências codificadoras adicionais supracitadas,

juntamente com mais sequências não codificadoras adicionais, incluindo sequências 5' e 3' não codificadoras, tais como as sequências transcritas, não traduzidas que desempenham um papel na transcrição (inclusive sinais de terminação), ligação a ribossomas e estabilidade do ARNm. As moléculas de ácidos nucleicos também podem conter sequências adicionais que codificam aminoácidos adicionais, como aqueles que proporcionam funcionalidades adicionais.

As moléculas de ácidos nucleicos da invenção também podem codificar os fragmentos ou os equivalentes funcionais dos polipéptidos e fragmentos da invenção. Uma molécula de ácidos nucleicos deste tipo poderá ser uma variante que ocorra naturalmente, como é o caso de uma variante alélica que ocorre naturalmente, ou a molécula pode ser uma variante em que se desconhece a sua ocorrência natural. Estas variantes que não ocorrem naturalmente da molécula de ácidos nucleicos poderão ser produzidas por técnicas de mutagénese, inclusive as que se aplicam a moléculas de ácidos nucleicos, a células ou a organismos.

Entre este género de variantes tem-se as variantes que são diferentes das moléculas de ácidos nucleicos supracitadas devido a substituições, deleções ou inserções de nucleótidos. Estas substituições, deleções ou inserções poderão envolver um ou mais nucleótidos. É possível alterar estas variantes em regiões codificadoras ou não codificadoras ou em ambas. As alterações nas regiões codificadoras poderão produzir substituições, deleções ou inserções de aminoácidos conservadoras ou não conservadoras.

Também é possível sujeitar as moléculas de ácidos nucleicos

da invenção a engenharia genética recorrendo a métodos geralmente conhecidos da técnica, por uma série de motivos, inclusive a modificação de clonagem, processamento e/ou expressão do produto genético (o polipéptido). A recombinação/*shuffling* do ADN por fragmentação aleatória e remontagem por PCR de fragmentos genéticos e de oligonucleótidos sintéticos estão incluídas como técnicas que podem ser empregues para manipular as sequências de nucleótidos. É possível recorrer a mutagénese específica de um *locus* para inserir novos locais de restrição, alterar padrões de glicosilação, modificar preferência de códão, produzir variantes de *splicing*, introduzir mutações, etc.

As moléculas de ácidos nucleicos que codificam um polipéptido da invenção podem estar ligadas a uma sequência heteróloga, de modo a que a molécula de ácidos nucleicos combinada codifique uma proteína de fusão. Por exemplo, para rastrear bibliotecas de péptidos em busca de inibidores da actividade do polipéptido, poderá ser útil que a expressão ocorra empregado uma molécula de ácidos nucleicos combinada deste tipo, uma proteína de fusão que possa ser reconhecida por um anticorpo disponível no mercado. Uma proteína de fusão também poderá ser manipulada para conter um local de clivagem entre a sequência do polipéptido e a sequência de uma proteína heteróloga, de modo a que o polipéptido possa ser clivado e purificado longe da proteína heteróloga.

Também é feita a descrição de moléculas anti-sentido, que são parcialmente complementares às moléculas de ácidos nucleicos que codificam polipéptidos da presente invenção e que, por isso, hibridam nas moléculas de ácidos nucleicos codificadoras (hibridação). Estas moléculas anti-sentido,

como os oligonucleótidos, podem ser concebidas para reconhecer, ligarem-se de forma específica e impedir a transcrição de um ácido nucleico-alvo que codifique um polipéptido da invenção, como será do conhecimento do especialista da técnica (ver, por exemplo, Cohen, J. S., "Trends in Pharm. Sci.", 10, 435 (1989), Okano, "J. Neurochem." 56, 560 (1991); O'Connor, "J. Neurochem" 56, 560 (1991); Lee et al., "Nucleic Acids Res" 6, 3 073 (1979); Cooney et al., "Science" 241, 456 (1988); Dervan et al., "Science" 251, 1 360 (1991).

O termo "hibridação" empregue no presente documento é relativo à associação de duas moléculas de ácidos nucleicos entre elas por ligação de hidrogénio. Tipicamente, uma molécula estará fixa a um suporte sólido e a outra encontrar-se-á livre em solução. Depois, as duas moléculas poderão ser postas em contacto uma com a outra sob condições que favoreçam a ligação de hidrogénio. Os factores que afectam esta ligação incluem: o tipo e o volume de solvente; a temperatura de reacção; o tempo de hibridação; agitação; os agentes de bloqueio da ligação não específica da molécula de fase líquida ao suporte sólido (reagente de Denhardt ou BLOTO); a concentração das moléculas; a utilização dos compostos para aumentar a taxa de associação das moléculas (sulfato de dextrano ou polietilenoglicol); e o rigor das condições de lavagem a seguir à hibridação (ver Sambrook et al. [*supra*]).

É possível estudar a inibição da hibridação de uma molécula totalmente complementar com uma molécula-alvo recorrendo a um teste de hibridação, como conhecido do estado da técnica (ver, por exemplo, Sambrook et al. [*supra*]). Uma molécula substancialmente homóloga irá então competir e inibir a

ligação de uma molécula totalmente homóloga à molécula-alvo sob diversas condições de rigor, como se encontra *in Wahl, G. M. and S. L. Berger (1987); "Methods Enzymol." 152: 399-407* e *Kimmel, A. R. (1987); "Methods Enzymol." 152: 507-511*).

"Rigor" refere-se a condições numa reacção de hibridação que favoreçam a associação de moléculas muito semelhantes, através da associação de moléculas diferentes. As condições de hibridação muito rigorosas encontram-se definidas como incubação durante a noite nos 42°C numa solução contendo 50% de formamida, 5 x SSC (NaCl 150 mM, citrato de trissódio 15 mM), fosfato de sódio 50 mM (pH de 7,6), 5 x solução de Denhardt, 10% de sulfato de dextrano e 20 micrograma/mL de ADN de esperma de salmão cortado desnaturado, seguido de lavagem dos filtros em 0,1 x SSC nos cerca de 65°C. As condições de baixo rigor implicam que a reacção de hibridação seja realizada nos 35°C (ver Sambrook *et al.* [*supra*]). Em termos preferenciais, as condições aplicadas à hibridação são as de elevado rigor.

Também é descrito um processo para detectar uma molécula de ácidos nucleicos da invenção, englobando as etapas de: (a) fazer contactar uma sonda nucleica de acordo com a invenção com uma amostra biológica sob condições de hibridação para formar duplexes; e (b) detectar estes duplexes que se formem.

Como adicionalmente debatido em baixo em ligação com os testes que podem ser realizados de acordo com a invenção, é possível empregar uma molécula de ácidos nucleicos como acima descrito como uma amostra de hibridação para ARN, ADNc ou ADN genómico, de modo a isolar os ADNc de

comprimento total e os clones genómicos que codificam os polipéptidos INSP101, e a isolar o ADNc e os clones genómicos de genes homólogos ou ortólogos que têm uma grande semelhança de sequências com o gene que codifica este polipéptido.

Relativamente a isto, as técnicas que se seguem, entre outras conhecidas do estado da técnica, podem ser empregues e são debatidas em baixo para fins ilustrativos. Os métodos de sequenciação e de análise de ADN são bem conhecidos e encontram-se geralmente disponíveis no estado da técnica e podem, de facto, ser empregues para pôr em prática muitas das execuções da invenção discutidas no presente documento. Estes métodos poderão empregar enzimas como o fragmento de Klenow do ADN polimerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polimerase (Perkin Elmer), polimerase T7 termostável (Amersham, Chicago, IL), ou combinações de polimerase e exonucleases correctoras (*proof-reading*) como as que se encontram no ELONGASE Amplification System, comercializado pela Gibco/BRL (Gaithersburg, MD). De preferência, o processo de sequenciação pode ser automatizado utilizando máquinas como a Hamilton Micro Lab 2200 (Hamilton, Reno, NV), a Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) e a ABI Catalyst e os sequenciadores de ADN: 373 e 377 DNA Sequencers (Perkin Elmer).

Um método de isolamento de uma molécula de ácidos nucleicos, que codifica um polipéptido com uma função equivalente à dos polipéptidos INSP101, é o de analisar uma biblioteca genómica ou de ADNc com uma sonda natural ou de concepção artificial, recorrendo a procedimentos padrão reconhecidos no estado da técnica (ver, por exemplo,

"Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, Nova Iorque, 1989, 1992). As sondas contendo pelo menos 15, de preferência pelo menos 30 e, com maior preferência, pelo menos 50, bases contíguas que correspondem ou são complementares às sequências de ácidos nucleicos do gene codificador apropriado (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 ou SEQ ID NO:9) são sondas particularmente úteis. É possível titular estas sondas com um reagente de detecção analítica para facilitar a sua identificação. Os reagentes úteis englobam, sem ficarem a eles limitados, radioisótopos, corantes fluorescentes e enzimas capazes de catalisar a formação de um produto detectável. Utilizando estas sondas, o especialista comum será capaz de isolar cópias complementares dos polinucleótidos de ADN genómico, ADNc ou ARN que codificam proteínas com interesse de fontes humanas, mamíferos ou de outras fontes animais, e de rastrear estas fontes quanto a sequências relacionadas, por exemplo, quanto a elementos adicionais da família, do tipo e/ou do subtípico.

Em muitos casos, as sequências de ADNc isoladas estarão incompletas, na medida em que a região que codifica o polipeptídeo ficará demasiado curta, normalmente na extremidade 5'. Estão disponíveis vários métodos para obter os ADNc de comprimento total, ou para alongar os ADNc curtos. É possível alongar estas sequências recorrendo a uma sequência nucleotídica parcial e empregando vários métodos conhecidos do estado da técnica para detectar sequências a montante, tais como promotores e elementos reguladores. Por exemplo, um método que pode ser empregue tem por base o método de amplificação rápida das

extremidades de ADNc = *Rapid Amplification of cDNA Ends* (RACE; ver, por exemplo, Frohman *et al.*, PNAS USA 85, 8 998-9002, 1988). Modificações recentes a esta técnica, exemplificadas pela tecnologia MarathonTM (Clontech Laboratories Inc.), por exemplo, simplificaram de forma significativa a procura de ADNc mais compridos. Uma técnica ligeiramente diferente, denominada PCR "de locais de restrição", emprega *primers* universais para recuperar a sequência desconhecida de ácidos nucleicos adjacente a um *locus* conhecido (Sarkar, G. (1993) "PCR Methods Applic". 2:318-322). Também é possível empregar PCR inversa para amplificar ou alongar sequências empregando *primers* divergentes com base numa região conhecida (Triglia, T. *et al.* (1988) "Nucleic Acids Res." 16:8186). Outro método que pode ser empregue é PCR de captura que engloba amplificação por PCR de fragmentos de ADN adjacentes a uma sequência conhecida no ADN cromossómico artificial de levedura e humano (Lagerstrom, M. *et al.* (1991) "PCR Methods Applic.", 1, 111-119). Outro método a que se pode recorrer para recuperar sequências desconhecidas é o de Parker, J. D. *et al.* (1991); "Nucleic Acids Res." 19:3055-3060). Adicionalmente, é possível empregar bibliotecas de PromoterFinderTM, PCR e *primers* internos para percorrer ADN genómico (Clontech, Palo Alto, CA). Este processo evita que se tenha de rastrear bibliotecas e é útil na descoberta de junções de intrão/exão.

No rastreio de ADNc de comprimento total, é preferível utilizar bibliotecas que tenham sido seleccionadas em termos de tamanho para englobar ADNc de maiores dimensões. Também se prefere bibliotecas com *primers* aleatórios (*random-primed libraries*) na medida em que conterão mais sequências que tenham as regiões 5' dos genes. A utilização

de uma biblioteca de *primers* aleatórios poderá ser especialmente preferida em situações em que uma biblioteca de oligo d(T) não forneça um ADNc de comprimento total. As bibliotecas genómicas poderão ser úteis no alongamento da sequência para regiões reguladoras não transcritas 5'.

Também é feita a descrição da utilização de moléculas de ácidos nucleicos na localização cromossómica. Nesta técnica, define-se como alvo especificamente uma molécula de ácidos nucleicos e a mesma pode hibridar com uma localização particular num cromossoma humano individual. O mapeamento de sequências relevantes para os cromossomas de acordo com a presente invenção é uma etapa importante na correlação confirmativa destas sequências com a doença associada ao gene. Uma vez feito o mapeamento de uma sequência para uma localização cromossómica precisa, será possível correlacionar a posição física da sequência no cromossoma com os dados do mapa genético. É possível encontrar estes dados, por exemplo, *in* V. McKusick, "Mendelian Inheritance in Man" (disponível on-line através da Johns Hopkins University Welch Medical Library). As relações entre os genes e as doenças que foram mapeadas para a mesma região cromossómica são depois identificadas por análise de ligação (co-hereditariedade de genes fisicamente adjacentes). Isto fornece informação valiosa aos investigadores que procuram genes de doenças utilizando a clonagem posicional ou outras técnicas de descoberta de genes. Uma vez que se tenha localizado de grosso modo a doença ou o sintoma, por meio de ligação genética a uma particular região genómica, quaisquer sequências que mapeiem essa área poderão representar genes associados ou reguladores para uma maior investigação. Também é possível empregar a molécula de ácidos nucleicos na detecção de

diferenças na localização cromossómica devido a translocação, inversão, etc. entre indivíduos normais, portadores ou afectados.

As moléculas de ácidos nucleicos também são valiosas na localização tecidual. Estas técnicas permitem a determinação de padrões de expressão do polipéptido em tecidos, por meio da detecção dos ARNm que os codificam. Estas técnicas englobam técnicas de hibridação *in situ* e técnicas de amplificação de nucleótidos, tais como PCR. Os resultados destes estudos dão indicação das funções normais do polipéptido no organismo. Além disso, estudos comparativos do padrão normal de expressão dos ARNm com o dos ARNm codificados por um gene mutante fornecem informações valiosas acerca do papel desempenhado por polipéptidos mutantes na patologia. Esta expressão desadequada poderá ser de natureza temporal, espacial ou quantitativa.

Também serão possíveis abordagens de silenciamento de genes para regular para baixo a expressão endógena de um gene que codifica um polipéptido da invenção. O ARN de interferência (ARNi) (Elbashir, SM et al., "Nature" 2001, 411, 494-498) é um método que se pode empregar de silenciamento de genes pós-trancripcional específico de sequências. Sintetiza-se oligonucleótidos de ARN de cadeia dupla curtos *in vitro* e introduz-se os mesmos numa célula. A ligação específica de sequência destes oligonucleótidos de ARN de cadeia dupla desencadeia a degradação do ARNm-alvo, reduzindo ou eliminando a expressão de proteínas-alvos.

A eficácia das abordagens de silenciamento de genes anteriormente mencionada pode ser avaliada pela medição da

expressão polipeptídica [por exemplo, por meio de *Western blotting* (imunotransferência enzimática)], e ao nível do ARN por meio de metodologias com base em TaqMan.

Os vectores da presente invenção contêm moléculas de ácidos nucleicos da invenção e poderão ser vectores de clonagem ou de expressão. As células hospedeiras da invenção, as quais podem ser transformadas, transfectadas e transduzidas com os vectores da invenção, poderão ser procarióticas ou eucarióticas.

É possível preparar os polipéptidos da invenção na forma recombinante, por meio da expressão das respectivas moléculas de ácidos nucleicos codificadoras em vectores contidos numa célula hospedeira. Estes métodos de expressão são bem conhecidos dos especialistas da técnica e muitos foram descritos em pormenor por Sambrook *et al* (*supra*) e Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, Londres, Boston, Nova Iorque, Sidnei, Tóquio, Toronto).

Dum modo geral, é possível empregar qualquer sistema ou vector que se adegue a manter, propagar ou expressar moléculas de ácidos nucleicos para que produzam um polipéptido no hospedeiro requerido. Pode inserir-se a sequência nucleotídica apropriada num sistema de expressão por uma série de técnicas de rotina e bem conhecidas, tais como, por exemplo, as descritas *in Sambrook et al.*, (*supra*). Em geral, o gene codificador pode ser posto sob o controlo de um elemento de controlo como um promotor, um local de ligação aos ribossomas (para expressão bacteriana) e, opcionalmente, um operador, pelo que a sequência de ADN

que codifica o polipéptido pretendido é transcrita no ARN na célula hospedeira transformada.

Os exemplos de sistemas de expressão adequados incluem, por exemplo, sistemas cromossómicos, episomais ou derivados de vírus, como, por exemplo, vectores derivados de: plasmídeos bacterianos, bacteriófagos, transposões, episomas de levedura, elementos de inserção, elementos cromossómicos de levedura, vírus como os baculovírus, papovavírus como o SV40, vírus vaccinia, adenovírus, vírus da variola aviária, vírus da pseudo-raiva e retrovírus, ou combinações destes, como os que derivam de elementos genéticos de plasmídeos e de bacteriófagos, inclusive cosmídeos e fagemídeos. Também se pode empregar cromossomas artificiais humanos (HAC) para produzir fragmentos maiores de ADN, que possam estar contidos e ser expressos num plasmídeo.

Os sistemas de expressão particularmente apropriados incluem microorganismos como as bactérias transformadas com vectores de expressão de ADN de bacteriófagos, plasmídeos ou cosmídeos recombinantes; levedura transformada com vectores de expressão de levedura; sistemas celulares de insectos infectados com vectores de expressão virais (por exemplo, o baculovírus); sistemas celulares vegetais transformados com vectores de expressão virais (por exemplo, vírus do mosaico da couve-flor, CaMV; vírus do mosaico do tabaco, TMV) ou com vectores de expressão bacterianos (por exemplo, plasmídeos pBR322 ou Ti); ou sistemas celulares animais. Também se pode empregar sistemas de tradução sem células.

É possível introduzir moléculas de ácidos nucleicos, que codificam um polipéptido da presente invenção, em células

hospedeiras por métodos descritos em muitos manuais laboratoriais padrão, tais como Davis *et al.*, "Basic Methods in Molecular Biology" (1986) e Sambrook *et al.*, (*supra*). Os métodos particularmente apropriados englobam transfecção por fosfato de cálcio, transfecção mediada por DEAE-dextrano, transvecção, microinjecção, transfecção mediada por lípidos catiónica, electroporação, transdução, *scrape loading*, infecção ou introdução balística (ver Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). Nas células eucarióticas, os sistemas de expressão podem ser transitórios (por exemplo, episomais) ou permanentes (integração cromossómica) de acordo com as necessidades do sistema.

A molécula de ácidos nucleicos codificadora poderá ou não conter uma sequência que codifica uma sequência de controlo, como um péptido de sinal ou uma sequência líder, como pretendido, por exemplo, para segregar o polipéptido traduzido no lúmen do retículo endoplasmático, no espaço periplásмico ou no ambiente extracelular. Estes sinais podem ser endógenos ao polipéptido ou podem ser sinais heterólogos. O hospedeiro bacteriano pode remover as sequências líderes no processamento pós-tradução.

Além das sequências de controlo, poderá ser desejável adicionar sequências reguladoras, que permitam a regulação da expressão do polipéptido em relação ao crescimento da célula hospedeira. Os exemplos de sequências reguladoras são as que causam o aumento ou a redução da expressão de um gene em resposta a um estímulo químico ou físico, inclusive à presença de um composto regulador, ou a diversas condições de temperatura ou metabólicas. As sequências

reguladoras são as regiões não traduzidas do vector, tais como intensificadores, promotores e regiões não traduzidas 5' e 3'. Estas interagem com as proteínas da célula hospedeira para realizar a transcrição e a tradução. Estas sequências reguladoras podem variar em força e especificidade. Em função do sistema vector e do hospedeiro empregue, poderá empregar-se uma série de elementos de transcrição e de tradução apropriados, inclusive promotores constitutivos e indutíveis. Por exemplo, aquando da clonagem em sistemas bacterianos, é possível empregar promotores indutíveis como o promotor híbrido lacZ do fagemido Bluescript (Stratagene, LaJolla, CA) ou o plasmídeo pSport1TM (Gibco BRL) e outros do género. O promotor da poliedrina do baculovírus pode ser empregue em células de insectos. Os promotores ou intensificadores derivados dos genomas de células vegetais (por exemplo, choque térmico, RUBISCO e genes de proteínas de reserva) ou de vírus das plantas (por exemplo, promotores virais ou sequências líderes) podem ser clonados no vector. Nos sistemas celulares dos mamíferos, é dada preferência aos promotores de genes de mamíferos ou de vírus de mamíferos. Se for necessário produzir uma linha celular que contenha múltiplas cópias da sequência, será possível empregar vectores com base no SV40 ou EBV com um marcador seleccionável apropriado.

Constrói-se um vector de expressão de modo a que a particular sequência codificadora de ácidos nucleicos esteja localizada no vector com as sequências reguladoras apropriadas, em que o posicionamento e a orientação da sequência codificadora em relação às sequências reguladoras sejam tais que a sequência codificadora seja transcrita sob o "controlo" das sequências reguladoras, ou seja, a ARN

polimerase que se liga à molécula de ADN nas sequências de controlo transcreve a sequência codificadora. Em alguns casos poderá ser necessário modificar a sequência, de modo a que ela possa ficar ligada às sequências de controlo com a orientação apropriada; ou seja, para manter a fase de leitura.

As sequências de controlo e outras sequências reguladoras podem estar ligadas à sequência codificadora de ácidos nucleicos antes da inserção num vector. Em alternativa, a sequência codificadora pode ser directamente clonada num vector de expressão que já contenha as sequências de controlo e um local de restrição apropriado.

Para a produção de alto rendimento a longo prazo de um polipéptido recombinante, é dada preferência à expressão estável. Por exemplo, é possível transformar linhas celulares que expressem de forma estável o polipéptido em questão, empregando vectores de expressão que possam ter origens virais de elementos de replicação e/ou de expressão endógena e um gene marcador seleccionável no mesmo vector ou num vector em separado. A seguir à introdução do vector, poderá deixar-se crescer as células durante 1-2 dias num meio enriquecido antes de serem mudadas para meios selectivos. A função do marcador seleccionável é a de conferir resistência à selecção, e a sua presença permite o crescimento e a recuperação de células que expressam com êxito as sequências introduzidas. Os clones resistentes de células transformadas de forma estável podem proliferar, recorrendo a técnicas de cultura de tecidos adequadas ao tipo de célula.

As linhas celulares de mamíferos, disponíveis como

hospedeiras para expressão, são conhecidas do estado da técnica e incluem muitas linhas celulares imortalizadas disponibilizadas pela American Type Culture Collection (ATCC), inclusive, sem ficarem a elas limitadas, células de ovário de hamster chinês (CHO), HeLa, rim de bebé hamster (BHK), rim de macaco (COS), C127, 3T3, BHK, HEK 293, melanoma de Bowes e carcinoma hepatocelular humano (por exemplo, Hep G2) e uma série de outras linhas celulares.

No sistema baculovírus, os materiais para os sistemas de expressão de células de insecto/baculovírus encontram-se disponíveis no mercado na forma de kit da, entre outros, Invitrogen, San Diego CA (o kit "MaxBac"). Estas técnicas são geralmente conhecidas dos especialistas da técnica e encontram-se descritas na totalidade *in Summers and Smith, "Texas Agricultural Experiment Station Bulletin No. 1 555"* (1987). As células hospedeiras particularmente apropriadas para serem empregues neste sistema incluem células de insectos, tais como as células *Drosophila S2* e *Spodoptera Sf9*.

Conhece-se do estado da técnica muitos sistemas de expressão genética de culturas de células vegetais e de plantas inteiras. Os exemplos de sistemas de expressão genética de células vegetais apropriados incluem os descritos nos US 5,693,506; US 5,659,122; e US 5,608,143. Foi feita a descrição de exemplos adicionais da expressão genética na cultura de células vegetais por Zenk, "Phytochemistry" 30, 3 861 - 3 863 (1991).

Em particular, é possível utilizar todas as plantas das quais se possa isolar protoplastos e pô-los em cultura para produzir plantas inteiras regeneradas, para que se recupere

plantas inteiras que contenham o gene transferido. Praticamente todas as plantas podem ser regeneradas a partir de tecidos ou de células de cultura, inclusive, sem ficarem a elas limitadas, todas as principais espécies de cana-de-açúcar, beterraba sacarina, algodão, árvores de fruta e outras árvores, legumes e vegetais.

Os exemplos de células hospedeiras bacterianas particularmente preferidas englobam células de estreptococos, estafilococos, *E. coli*, *Streptomyces* e *Bacillus subtilis*.

Os exemplos de células hospedeiras particularmente adequadas à expressão fúngica englobam células de leveduras (por exemplo, *S. cerevisiae*) e células de *Aspergillus*.

Conhece-se uma série de sistemas de selecção no estado da técnica, que podem ser empregues para recuperar linhas celulares transformadas. Os exemplos englobam os genes da timidina quinase do vírus da herpes vulgar (Wigler, M. et al. (1997) "Cell" 11:223-32) e da adenina fosforribosiltransferase (Lowy, I. et al. (1980) "Cell" 22:817-23), que podem ser empregues em células tk⁻ ou aprt⁺, respectivamente.

É igualmente possível empregar a resistência a antimetabólitos, a antibióticos ou a herbicidas como base de selecção; por exemplo, a di-hidrofolato reductase (DHFR) que confere resistência ao metotrexato (Wigler, M. et al. (1980) "Proc. Natl. Acad. Sci." 77:3567-70); npt, que confere resistência à neomicina de aminoglicósidos e G-418 (Colbere-Garapin, F. et al (1981) "J. Mol. Biol." 150:1-14) e als ou pat, que conferem resistência ao cloro-sulfurão e

à fosfinotricina acetiltransferase, respectivamente. Foi feita a descrição de genes seleccionáveis adicionais, cujos exemplos serão claros ao especialista.

Embora a presença ou a ausência de expressão de genes marcadores sugira que o gene em questão também se encontra presente, poderá ser necessário confirmar a sua presença e a sua expressão. Por exemplo, caso se insira a sequência relevante numa sequência do gene marcador, será possível identificar as células transformadas contendo as sequências apropriadas, devido à ausência da função do gene marcador. Em alternativa, é possível colocar um gene marcador em série com uma sequência que codifica um polipéptido da invenção sob o controlo de um único promotor. A expressão do gene marcador em resposta a indução ou selecção também indica usualmente a expressão do gene em série.

Em alternativa, é possível identificar as células hospedeiras, que contêm uma sequência de ácidos nucleicos que codifica um polipéptido da invenção e que expressa o referido polipéptido, por meio de uma série de procedimentos conhecidos do especialista da técnica. Estes procedimentos englobam, sem a eles ficarem limitados, hibridações de ADN-ADN ou de ADN-ARN e biotestes proteicos, por exemplo, separação celular activada por fluorescência (FACS-fluorescence activated cell sorting) ou técnicas de ensaios imunológicos (como o ensaio de imunoabsorção enzimática [ELISA - *enzyme-linked immunosorbent assay*] e o ensaio radioimunológico [RIA-radioimmunoassay], que englobam tecnologias com base em membranas, soluções ou chips, com vista à detecção e/ou à quantificação de ácido nucleico ou proteína (ver Hampton, R. et al. (1990) "Serological Methods, a Laboratory Manual", APS Press, St

Paul, MN) e Maddox, D. E. et al. (1983) "J. Exp. Med", 158, 1211-1216).

Os especialistas da técnica conhecem uma grande variedade de títulos e de técnicas de conjugação e podem empregá-los em diversos ensaios de ácidos nucleicos e de aminoácidos. Os meios destinados à produção de hibridação titulada ou sondas de PCR para detectar sequências relacionados com moléculas de ácidos nucleicos que codificam polipéptidos da presente invenção incluem oligotitulação, *nick-translation*, *end-labelling* ou amplificação por PCR recorrendo a um polinucleótido titulado. Em alternativa, é possível clonar as sequências que codificam o polipéptido da invenção num vector destinado à produção de uma sonda de ARNm. Estes vectores são conhecidos do estado da técnica, encontram-se disponíveis no mercado e podem ser empregues na síntese de sondas de ARN *in vitro*, por meio da adição de uma ARN polimerase apropriada, como T7, T3 ou SP6 e nucleótidos titulados. É possível seguir estes procedimentos com recurso a uma série de kits disponíveis no mercado (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); e U. S. Biochemical Corp., Cleveland, OH)).

Os títulos ou as moléculas repórter adequados/as, que podem ser empregues para facilitar a detecção, englobam radionuclídos, enzimas e agentes fluorescentes, quimioluminescentes e cromogénicos, assim como substratos, cofactores, inibidores, partículas magnéticas e outros do género.

Também se pode empregar moléculas de ácidos nucleicos para criar animais transgénicos, em particular roedores. Isto pode ser realizado localmente por modificação das células

somáticas, ou por terapia na linha germinal para incorporar modificações herdáveis. Estes animais transgénicos poderão ser particularmente úteis na produção de modelos animais para moléculas farmacológicas eficazes como moduladores dos polipéptidos da presente invenção.

É possível recuperar e purificar o polipéptido a partir de culturas celulares recombinantes, por métodos bem conhecidos, inclusive precipitação com sulfato de amónio ou etanol, extracção ácida, cromatografia de permuta aniónica ou catiónica, cromatografia em fosfocelulose, cromatografia de interacção hidrófoba, cromatografia de afinidade, cromatografia com hidroxilapatita e cromatografia com lectinas. A cromatografia líquida de alta resolução (HPLC) é particularmente útil para fins de purificação. É possível utilizar técnicas bem conhecidas para o *refolding* de proteínas, para regenerar uma conformação activa quando o polipéptido é desnaturado durante o isolamento e ou a purificação.

Também se pode empregar construções de vectores especializadas para facilitar a purificação das proteínas, como pretendido, ao juntar as sequências que codificam os polipéptidos a uma sequência nucleotídica que codifica um domínio polipeptídico que irá facilitar a purificação de proteínas solúveis. Os exemplos destes domínios facilitadores de purificação incluem péptidos quelantes de metais, tais como módulos de histidina-triptofano que permitem a purificação em metais imobilizados, domínios da proteína A que permitem a purificação em imunoglobulina imobilizada, e o domínio utilizado no sistema de purificação de afinidade/extensão FLADS (Immunex Corp., Seattle, WA). É possível realizar a inclusão de sequências

de ligação cliváveis, como as específicas para o Factor XA ou a enteroquinase (Invitrogen, San Diego, CA), entre o domínio de purificação e o polipéptido da invenção para facilitar a purificação. Um vector de expressão deste tipo permite a expressão de uma proteína de fusão contendo o polipéptido da invenção fundido em vários resíduos histidina que antecedem um local de clivagem tiorredoxina ou enteroquinase. Os resíduos histidina facilitam a purificação por IMAC (cromatografia de afinidade por iões metálicos imobilizados, como descrito *in* Porath, J. et al. (1992), "Prot. Exp. Purif." 3:263-281), enquanto o local de clivagem tiorredoxina ou enteroquinase proporciona um meio de purificação do polipéptido da proteína de fusão. Um debate acerca de vectores que contêm proteínas de fusão encontra-se *in* Kroll, D. J. et al. (1993; "DNA Cell Biol." 12:441-453).

Caso o polipéptido deva ser expresso para ser empregue em testes de rastreio, prefere-se dum modo geral que seja produzido na superfície da célula hospedeira na qual é expresso. Se for este o caso, as células hospedeiras podem ser colhidas antes da sua utilização no teste de rastreio, por exemplo, recorrendo a técnicas como separação celular activada por fluorescência (FACS) ou técnicas de imunoafinidade. Se o polipéptido for secretado para o meio, pode recuperar-se o meio com vista a recuperar e a purificar o polipéptido expresso. Se o polipéptido for produzido a nível intracelular, será necessário que as células sofram lise antes de se recuperar o polipéptido.

Pode empregar-se o polipéptido para rastrear bibliotecas de compostos, escolhendo dentre uma série de técnicas de rastreio de medicamentos. Estes compostos podem activar

(agonizar) ou inibir (antagonizar) o nível de expressão do gene ou a actividade do polipéptido. Os compostos preferidos são eficazes em modificar a expressão de um gene natural que codifique um polipéptido, ou em regular a actividade de um polipéptido.

Pode isolar-se os compostos agonistas ou antagonistas a partir, por exemplo, de células, preparações sem células, bibliotecas químicas ou misturas de produtos naturais. Estes agonistas ou antagonistas podem ser substratos naturais ou modificados, ligandos, enzimas, receptores ou miméticos estruturais ou funcionais. Para uma análise apropriada destas técnicas de rastreio, ver Coligan *et al.*, "Current Protocols in Immunology" 1 (2): Capítulo 5 (1991).

Os compostos mais prováveis de serem bons antagonistas são as moléculas que se ligam ao polipéptido, sem induzir os efeitos biológicos do polipéptido quando se ligam a ele. Os antagonistas potenciais incluem moléculas orgânicas pequenas, péptidos, polipéptidos e anticorpos que se ligam ao polipéptido e, com isso, inibem ou extinguem a actividade do mesmo. Deste modo, é possível inibir a ligação do polipéptido a moléculas de ligação celular normal, impedindo-se a actividade biológica normal do polipéptido.

O polipéptido empregue nesta técnica de rastreio pode encontrar-se livre em solução, fixo a um suporte sólido, situado numa superfície celular ou localizado no espaço intracelular. Em geral, estes procedimentos de rastreio poderão implicar a utilização de células ou de membranas celulares apropriadas que expressem o polipéptido, as quais entram em contacto com um composto de teste para se

observar a ligação, a estimulação ou a inibição de uma resposta funcional. A resposta funcional das células que entram em contacto com o composto de teste é depois comparada com as células de controlo que não entraram em contacto com o composto de teste. Um teste destes consegue avaliar se o composto de teste resulta num sinal gerado pela activação do polipéptido, empregando um sistema de detecção apropriado. Os inibidores da activação são geralmente testados na presença de um agonista conhecido e observa-se o efeito de activação pelo agonista na presença do composto de teste.

Um método preferido de identificação de um composto agonista ou antagonista de um polipéptido contempla:

- (a) fazer contactar uma célula, em cuja superfície é expresso o polipéptido, estando o polipéptido associado a um segundo componente capaz de fornecer um sinal detectável em resposta à ligação de um composto ao polipéptido, com um composto a ser rastreado sob condições que permitam a ligação ao polipéptido; e
- (b) determinar se o composto se liga e activa ou inibe o polipéptido, através da medição do nível de um sinal gerado a partir da interacção do composto com o polipéptido.

Outro método preferido de identificação de um agonista ou antagonista de um polipéptido contempla:

- (a) fazer contactar uma célula, em cuja superfície é expresso o polipéptido, estando o polipéptido associado a um segundo componente capaz de fornecer um sinal detectável em resposta à ligação de um composto ao polipéptido, com um composto a ser rastreado sob condições que permitam a

ligação ao polipéptido; e

(b) determinar se o composto se liga e activa ou inibe o polipéptido, através da comparação do nível de um sinal gerado a partir da interacção do composto com o polipéptido com o nível de um sinal na ausência do composto.

Os métodos gerais acima descritos poderão ainda englobar a identificação do agonista ou do antagonista na presença de ligando titulado ou não titulado para o polipéptido.

Noutra execução, o método de identificação do agonista ou do antagonista de um polipéptido contempla:

Determinar a inibição da ligação de um ligando a células que têm na sua superfície um polipéptido, ou a membranas celulares contendo esse polipéptido, na presença de um composto candidato sob condições que permitam a ligação ao polipéptido, e determinar a quantidade de ligando ligado ao polipéptido. Um composto que seja capaz de provocar a redução da ligação de um ligando é considerado um agonista ou um antagonista. De preferência o ligando está titulado.

Mais em particular, um método de rastreio de um composto agonista ou antagonista de um polipéptido engloba as etapas de:

(a) incubar um ligando titulado com uma célula inteira que expressa um polipéptido na superfície celular, ou com uma membrana celular contendo um polipéptido,

(b) medir a quantidade de ligando titulado ligado à célula inteira ou à membrana celular;

(c) adicionar um composto candidato a uma mistura do ligando titulado e da célula inteira ou da membrana celular

da etapa (a), e permitir que a mistura atinja o equilíbrio; (d) medir a quantidade de ligando titulado ligado à célula inteira ou à membrana celular após a etapa (c); e (e) comparar a diferença no ligando titulado ligado na etapa (b) e na etapa (d), de modo a que o composto que provoca a redução de ligação na etapa (d) seja considerado um agonista ou antagonista.

Os polipéptidos INSP101 da presente invenção podem modular uma série de processos fisiológicos e patológicos, inclusive processos como a proliferação e a migração celulares dentro do sistema imunitário. Assim, é possível analisar a actividade biológica dos polipéptidos INSP101 em sistemas que permitam o estudo destas actividades moduladoras, realizando uma série de ensaios apropriados.

Por exemplo, com vista a observar a inibição do crescimento celular, é possível empregar um meio sólido ou líquido. Num meio sólido, as células sujeitas a inibição do crescimento podem ser facilmente seleccionadas do grupo celular em questão, ao comparar os tamanhos das colónias formadas. Num meio líquido, é possível rastrear a inibição do crescimento ao medir a turvação do meio de cultura ou ao incorporar timidina titulada no ADN. Tipicamente, é possível recorrer à incorporação de um nucleósido de forma análoga em ADN acabado de sintetizar, de modo a medir a proliferação (ou seja, o crescimento celular activo) numa população de células. Por exemplo, pode empregar-se bromodeoxiuridina (BrdU) como reagente de titulação de ADN, e pode empregar-se anticorpos monoclonais murinos anti-BrdU como reagente de detecção. Este anticorpo liga-se apenas a células que contêm ADN que tenha bromodeoxiuridina nele incorporada. É possível aplicar uma série de métodos de detecção em

conjunção com este teste, inclusive imunofluorescência, imuno-histoquímica, ELISA e métodos colorimétricos. Os kits que contêm a bromodeoxiuridina (BrdU) e anticorpo monoclonal murino anti-BrdU encontram-se disponíveis no mercado através da Boehringer Mannheim (Indianapolis, IN).

Poderá verificar-se que os polipéptidos INSP101 modulem uma série de processos fisiológicos e patológicos em função da dose nos testes acima descritos. Por conseguinte, os "equivalentes funcionais" dos polipéptidos INSP101 incluem os polipéptidos que demonstrem quaisquer das mesmas actividades moduladoras nos testes acima descritos em função da dose. Embora o grau de actividade em função da dose precise de ser idêntico ao dos polipéptidos INSP101, preferencialmente os "equivalentes funcionais" exibirão uma dependência da dose substancialmente semelhante num certo teste de actividade em comparação com os polipéptidos INSP101.

Em certas execuções acima descritas, poderá recorrer-se a testes de ligação simples, nos quais a aderência de um composto de teste à superfície que tem o polipéptido é detectada por meio de um título directa ou indirectamente associado ao composto de teste, ou a um teste que envolva a competição com um concorrente titulado. Noutra execução, poderá recorrer-se a testes de rastreio de medicamentos competitivos, nos quais anticorpos neutralizantes, capazes de se ligaram ao polipéptido, competem de forma específica com um composto de teste pela ligação. Deste modo, é possível utilizar os anticorpos para detectar a presença de um composto de teste que tenha uma afinidade de ligação específica relativamente ao polipéptido.

Também se pode conceber testes para detectar o efeito de compostos de teste adicionados na produção de ARNm que codifica o polipéptido nas células. Por exemplo, é possível conceber um ELISA que meça níveis de segregação ou associados à célula do polipéptido, aplicando anticorpos monoclonais ou policlonais por métodos padrão conhecidos no estado da técnica, e isto pode ser usado para procurar compostos que possam inibir ou aumentar a produção do polipéptido a partir de células ou tecidos manipuladas/os de forma adequada. Pode-se em seguida medir a formação de complexos de ligação entre o polipéptido e o composto que está a ser testado.

Outra técnica de rastreio de medicamentos que pode ser empregue proporciona um elevado rendimento de rastreio de compostos tendo uma afinidade de ligação apropriada em relação ao polipéptido em questão (ver o pedido de patente internacional WO84/03564). Neste método, sintetiza-se grandes quantidades de diferentes pequenos compostos de teste num substrato sólido, que podem depois ser feitos reagir com o polipéptido da invenção e lavados. Uma forma de imobilizar o polipéptido é empregar anticorpos não-neutralizantes. É possível depois detectar o polipéptido de ligação recorrendo a métodos bem conhecidos do estado da técnica. Também se pode aplicar o polipéptido purificado directamente sobre placas a serem utilizadas nas técnicas de rastreio de medicamentos supracitadas.

Pode empregar-se o polipéptido para identificar receptores ligados a membrana ou solúveis, através de técnicas padrão de ligação a receptores, conhecidas do estado da técnica, tais como testes de ligação a ligandos e testes de ligação cruzada, nos quais o polipéptido é titulado com um isótopo

radioactivo, é quimicamente modificado ou fundido numa sequência peptídica que facilita a sua detecção ou purificação, e é incubado com uma fonte do receptor putativo (por exemplo, uma composição de células, membranas celulares, sobrenadantes celulares, extractos tecidulares ou fluidos corporais). É possível medir a eficácia de ligação recorrendo a técnicas biofísicas, como a ressonância plasmónica de superfície e a espectroscopia. Pode aplicar-se testes de ligação com vista à purificação e à clonagem do receptor, mas os mesmos também poderão identificar agonistas e antagonistas do polipéptido que compitam com a ligação do polipéptido ao seu receptor. Os métodos padrão de realização de testes de rastreio são bem compreendidos no estado da técnica.

Os polipéptidos INSP101 poderão modular uma série de processos fisiológicos e patológicos, inclusive processos como a segregação hormonal, o crescimento celular e as metástases celulares, incluindo a metastização de células cancerígenas (Torosian, M. H. & Donoway, R. B., 1991, "Cancer", 67(9): 2280-2283). Por conseguinte, é possível analisar a actividade biológica dos polipéptidos INSP101 em sistemas que permitam o estudo destas actividades moduladoras realizando uma série de testes apropriados.

Por exemplo, no intuito de observar o efeito dos polipéptidos INSP101 da presente invenção na metastização celular, é possível aplicar um ou mais dos métodos descritos *in Ohtaki et al.*, "Nature". 2001, 31 de Maio; 411 (6837): 613-7, ou nas publicações aí referidas.

Por exemplo, no intuito de observar o efeito dos polipéptidos INSP101 da presente invenção na segregação de

hormonas, é possível empregar um ou mais dos métodos descritos *in Hinuma et al.*, "Nature". 1998, 21 de Maio; 393 (6 682):272-6 ou *Hinuma et al.*, "Nat Cell Biol." 2000, Outubro; 2(10):703-8, ou nas publicações aí referidas.

Também se pode utilizar os polipéptidos INSP101 na identificação e na caracterização de receptores, em particular de receptores de hormonas de crescimento, que interajam com os polipéptidos INSP101. Os métodos adequados de identificação e de caracterização englobam, sem que fiquem a eles limitados, os descritos *in Hinuma et al.*, "Nat Cell Biol." 2000, Outubro; 2(10): 703-8, e no pedido de patente internacional publicado como WO01/17958 ou nas publicações aí referidas.

Poderá verificar-se que os polipéptidos INSP101 modulem uma série de processos fisiológicos e patológicos em função da dose nos testes acima descritos. Assim, os "equivalentes funcionais" dos polipéptidos INSP101 incluem os polipéptidos que apresentem quaisquer das mesmas actividades moduladoras nos testes acima descritos em função da dose. Embora seja necessário que o grau de actividade em função da dose seja idêntico ao dos polipéptidos INSP101, preferencialmente os "equivalentes funcionais" apresentarão uma dependência da dose substancialmente similar num determinado teste de actividade em comparação com os polipéptidos INSP101.

Também é feita a descrição de um kit de rastreio útil em métodos de identificação de agonistas, antagonistas, ligandos, receptores, substratos, enzimas, que se encontram acima descritos.

Também é feita a descrição de agonistas, antagonistas, ligandos, receptores, substratos e enzimas e de outros compostos que modulam a actividade ou a antigenicidade do polipéptido da invenção descobertos pelos métodos acima descritos.

A invenção fornece igualmente composições farmacêuticas que contêm um polipéptido, um ácido nucleico ou um ligando da invenção, em combinação com uma substância de suporte farmacêutica adequada. Estas composições podem adequar-se como reagentes terapêuticos ou de diagnóstico, como vacinas ou como outras composições imunogénicas, como definido pormenorizadamente em baixo.

De acordo com a terminologia empregue no presente documento, uma composição contendo um polipéptido, um ácido nucleico, um ligando ou um composto [X] está "substancialmente isenta de" impurezas [Y no presente documento], se pelo menos 85% em peso do total de X + Y na composição for X. De preferência, X engloba pelo menos cerca de 90% em peso do total de X + Y na composição, com maior preferência pelo menos cerca de 95%, 98% ou mesmo 99% em peso.

Preferencialmente, as composições farmacêuticas deverão conter uma quantidade de eficácia terapêutica de um polipéptido, de uma molécula de ácidos nucleicos, de um ligando ou de um composto da invenção. O termo "quantidade de eficácia terapêutica" refere-se, no sentido do presente documento, a uma quantidade de uma agente terapêutico necessária para tratar, melhorar ou prevenir uma patologia ou condição alvo, ou para surtir um efeito terapêutico ou preventivo detectável. Para todos os compostos, é possível

estimar inicialmente a dose de eficácia terapêutica seja em testes de culturas celulares, por exemplo, de células neoplásicas, ou em modelos animais, usualmente ratos, coelhos, cães ou porcos. Também se pode empregar o modelo animal para determinar a gama apropriada de concentração e a via de administração. Em seguida, pode aplicar-se esta informação para determinar as doses úteis e as vias de administração nos humanos.

A quantidade eficaz precisa para o indivíduo humano dependerá da gravidade do estado patológico, do estar geral de saúde do indivíduo, da idade, do peso, do sexo, da dieta, do tempo e da frequência de administração, da(s) combinação(ões) de medicamentos, de sensibilidades de reacção e da tolerância/resposta à terapia. É possível determinar esta quantidade por experimentação de rotina e fica no âmbito de decisão do clínico. Em geral, uma dose eficaz irá de 0,01 mg/kg a 50 mg/kg, de preferência de 0,05 mg/kg a 10 mg/kg. As composições podem ser ministradas individualmente a um doente ou podem ser ministradas em combinação com outros agentes, medicamentos ou hormonas.

Uma composição farmacêutica também pode conter uma substância de suporte aceitável a nível farmacêutico, com vista à administração do agente terapêutico. Estas substâncias de suporte incluem anticorpos e outros polipéptidos, genes e outros agentes terapêuticos, tais como lipossomas, desde que a substância de suporte não induza ela própria a produção de anticorpos nocivos ao indivíduo que recebe a composição, e possa ser ministrada sem indevida toxicidade. As substâncias de suporte adequadas poderão ser macromoléculas grandes de metabolismo lento, como é o caso das proteínas, dos polissacáridos, dos

ácidos polilácticos, dos ácidos poliglicólicos, dos aminoácidos poliméricos, dos copolímeros de aminoácidos e de partículas virais inactivas.

Os sais aceitáveis a nível farmacêutico podem ser aí utilizados, por exemplo, sais de ácidos minerais como os hidrocloretos, hidrobrometos, fosfatos, sulfatos e outros do género; e os sais de ácidos orgânicos como os acetatos, propionatos, malonatos, benzoatos e outros do género. Uma discussão minuciosa acerca das substâncias de suporte aceitáveis a nível farmacêutico encontra-se *in "Remington's Pharmaceutical Sciences"* (Mack Pub. Co, N. J. 1991).

As substâncias de suporte aceitáveis a nível farmacêutico em composições terapêuticas poderão ainda conter líquidos como água, meio salino, glicerol e etanol. Além disso, é possível a presença nestas composições de substâncias auxiliares, tais como agentes molhantes e emulsionantes, substâncias-tampão de pH e outros do género. Estas substâncias de suporte permitem a formulação das composições farmacêuticas na forma de comprimidos, pílulas, drageias, cápsulas, líquidos, géis, xaropes, suspensões espessas, suspensões e outros do género, destinadas a serem ingeridas pelo doente.

Uma vez formuladas, as composições da invenção podem ser ministradas directamente no indivíduo. Os indivíduos a serem tratados podem ser animais; em particular, é possível tratar indivíduos humanos.

É possível ministrar as composições farmacêuticas empregues na presente invenção por diversas vias, inclusive, mas sem ficarem a elas limitadas, as aplicações oral, intravenosa,

intramuscular, intra-arterial, intramedular, intratecal, intraventricular, transdérmica ou transcutânea (por exemplo, ver o WO98/20734), as vias subcutânea, intraperitoneal, intranasal, enteral, tópica, sublingual, intravaginal ou rectal. Também é possível empregar pistolas de genes ou *hyposprays* para ministrar as composições farmacêuticas da invenção. Tipicamente, as composições terapêuticas podem ser preparadas na forma de injectáveis, seja na forma de suspensões ou soluções líquidas; também é possível preparar formas sólidas adequadas a solução ou a suspensão em veículos líquidos antes da injecção.

A entrega directa das composições será em geral conseguida por injecção, de forma subcutânea, intraperitoneal, intravenosa ou intramuscular, ou entregue no espaço intersticial de um tecido. Também se pode ministrar as composições numa lesão. O tratamento por dosagem poderá ser um programa de uma única dose ou um programa de doses múltiplas.

Se a actividade do polipéptido for excessiva num particular estado patológico, existem várias abordagens. Uma abordagem tem a ver com o ministrar a um indivíduo um composto inibidor (antagonista) como acima descrito, juntamente com uma substância de suporte aceitável a nível farmacêutico, numa quantidade eficaz à inibição da função do polipéptido, tal como bloqueando a ligação de ligandos, substratos, enzimas, receptores, ou inibindo um segundo sinal, aliviando desse modo a condição anormal. De preferência, estes antagonistas são anticorpos. Com total preferência, estes anticorpos são quiméricos e/ou humanizados para minimizar a respectiva imunogenicidade, como previamente descrito.

Noutra abordagem, é possível ministrar as formas solúveis do polipéptido que mantém afinidade de ligação em relação ao ligando, ao substrato, à enzima, ao receptor em questão. Tipicamente, o polipéptido pode ser ministrado na forma de fragmentos que conservam as porções relevantes.

Numa abordagem alternativa, a expressão do gene que codifica o polipéptido pode ser inibida utilizando técnicas de bloqueio da expressão, tais como a utilização de moléculas de ácidos nucleicos anti-sentido (como acima descrito), sejam de geração interna ou de administração em separado. Consegue-se as modificações da expressão genética pela concepção de sequências complementares ou de moléculas anti-sentido (ADN, ARN ou PNA) para o controlo, 5' ou regiões reguladoras (sequência de sinal, promotores, intensificadores e intrões) do gene que codifica o polipéptido. De forma semelhante, consegue-se a inibição com a metodologia de emparelhamento de bases de "tripla hélice". O emparelhamento de tripla hélice é útil por levar à inibição da capacidade de a hélice dupla abrir o suficiente para que se dê a ligação de polimerases, factores de transcrição ou moléculas reguladoras. Avanços terapêuticos recentes utilizando o ADN triplo foram descritos na literatura (Gee, J. E., et al. (1994) in: Huber, B. E. and B. I. Carr, "Molecular and Immunologic Approaches", Futura Publishing Co., Mt. Kisco, Nova Iorque). A sequência complementar ou a molécula anti-sentido também pode ser concebida para bloquear a tradução do ARNm, ao impedir que o transcrito se ligue aos ribossomos. Estes oligonucleótidos podem ser ministrados ou podem ser gerados *in situ* a partir de expressão *in vivo*.

Além disso, a expressão do polipéptido da invenção pode ser impedida utilizando ribozimas específicas para a sua sequência de ARNm codificadora. As ribozimas são ARN de acção catalítica que podem ser naturais ou sintéticas (ver, por exemplo, Usman, N, et al., "Curr. Opin. Struct. Biol" (1996) 6(4), 573-33). Pode conceber-se ribozimas sintéticas para clivarem especificamente os ARNm em posições seleccionadas, impedindo assim a tradução dos ARNm para polipéptido funcional. As ribozimas podem ser sintetizadas com uma cadeia principal de ribose-fosfato natural e bases naturais, como normalmente se encontra em moléculas de ARN. Em alternativa, as ribozimas podem ser sintetizadas com cadeias principais não naturais, por exemplo, 2'-O-metil-ARN, para proporcionar protecção contra degradação da ribonuclease, e poderão conter bases modificadas.

É possível modificar as moléculas de ARN para aumentar a estabilidade intracelular e o tempo de semi-vida. As modificações possíveis incluem, não ficando a elas limitadas, a adição de sequências de flanco na extremidade 5' e/ou na extremidade 3' da molécula, ou a utilização de tioato de fósforo ou de 2' O-metilo em vez de ligações de fosfodiesterase na cadeia principal da molécula. Este conceito é inerente à produção de PNA e pode ser aumentado em todas estas moléculas pela inclusão de bases não tradicionais como a inosina, a queosina e a butosina, assim como as formas acetil-modificadas, metil-modificadas, tio-modificadas e formas modificadas de forma semelhante das adenina, citidina, guanina, timina e uridina, que não são tão facilmente reconhecidas pelas endonucleases endógenas.

Também existem várias abordagens com vista ao tratamento de condições anormais, relacionadas com uma subexpressão do

polipéptido da invenção e da sua actividade. Uma abordagem contempla a administração a um indivíduo de uma quantidade de eficácia terapêutica de um composto que activa o polipéptido, ou seja, um agonista como acima descrito, para atenuar a condição anormal. Em alternativa, é possível ministrar uma quantidade terapêutica do polipéptido em combinação com uma substância de suporte farmacêutica apropriada, para recuperar o equilíbrio fisiológico relevante do polipéptido.

É possível recorrer à terapia genética para conseguir a produção endógena do polipéptido pelas células relevantes no indivíduo. Recorre-se à terapia genética para tratar de forma permanente a produção desadequada do polipéptido, ao substituir um gene defeituoso por um gene terapêutico corrigido.

Pode realizar-se a terapia genética da presente invenção *in vivo* ou *ex vivo*. A terapia genética *ex vivo* requer o isolamento e a purificação das células do doente, a introdução de um gene terapêutico e a introdução das células geneticamente modificadas de novo no doente. Em contraste, a terapia genética *in vivo* não requer o isolamento e a purificação das células de um doente.

O gene terapêutico é tipicamente "embalado" para ser ministrado no doente. Os veículos de entrega do gene poderão ser não virais, como os lipossomas, ou vírus com deficiência de replicação, tais como os adenovírus, como descritos por Berkner, K. L., *in* "Curr. Top. Microbiol. Immunol.", 158, 39-66 (1992) ou vectores de vírus adeno-associados (AAV) como descritos por Muzyczka, N., *in* "Curr. Top. Microbiol. Immunol.", 158, 97-129 (1992) e na patente

US n.º 5,252,479. Por exemplo, pode manipular-se por engenharia genética uma molécula de ácidos nucleicos que codifica um polipéptido da invenção, com vista à expressão num vector retroviral com deficiência de replicação. Este constructo de expressão poderá ser em seguida isolado e introduzido numa célula de embalagem, transduzida com um vector plasmídeo retroviral contendo ARN que codifica o polipéptido, de modo a que a célula de embalagem passe então a produzir partículas virais infecciosas contendo o gene em questão. Pode ministrarse estas células produtoras a um indivíduo para manipular por engenharia genética células *in vivo* e para expressão do polipéptido *in vivo* (ver o Capítulo 20, "Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (e as referências aí citadas) *in* "Human Molecular Genetics" (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Outra abordagem é a da administração do "ADN nu", no qual o gene terapêutico é directamente injectado na corrente sanguínea ou no tecido muscular.

Em situações em que os polipéptidos ou as moléculas de ácidos nucleicos da invenção sejam agentes patológicos, a invenção contempla que os mesmos possam ser empregues em vacinas para aumento dos anticorpos contra esse agente patológico.

As vacinas de acordo com a invenção podem ser profiláticas (ou seja, para prevenir a infecção) ou terapêuticas (ou seja, para tratar a doença após a infecção). Estas vacinas contêm antigénio(s) imunizante(s), imunogénio(s), polipéptido(s), proteína(s) ou ácido nucleico, usualmente em combinação com substâncias de suporte aceitáveis a nível

farmacêutico como acima descritas, que englobam todas as substâncias de suporte que não induzam elas próprias a produção de anticorpos nocivos ao indivíduo que recebe a composição. Adicionalmente, estas substâncias de suporte poderão funcionar como agentes imunoestimulantes ("adjuvantes"). Além disso, o antigénio ou o imunogénio pode ser conjugado com um toxóide bacteriano, como o toxóide da difteria, do tétano, da cólera, do *H. pylori* e de outros agentes patogénicos.

Uma vez que os polipéptidos podem ser decompostos no estômago, as vacinas que contêm polipéptidos são de preferência ministradas por via parenteral (por exemplo, de forma subcutânea, intramuscular, intravenosa ou por injecção intradérmica). As formulações adequadas à administração parenteral englobam soluções injectáveis estéreis aquosas e não aquosas, que poderão conter antioxidantes, tampões, bacteriostatos e solutos que tornam a formulação isotónica com o sangue do destinatário, e suspensões estéreis aquosas e não aquosas, que poderão incluir agentes de suspensão ou espessantes.

As formulações de vacinas da invenção poderão ser apresentadas em recipientes de dose unitária ou de dose múltipla. Por exemplo, ampolas e frascos selados, e podem ser armazenadas sob condições de liofilização, requerendo apenas a adição da substância de suporte líquida estéril logo antes da utilização. A dosagem dependerá da actividade específica da vacina e pode ser facilmente determinada por experimentação de rotina.

A entrega genética de anticorpos que se ligam a polipéptidos de acordo com a invenção também pode ser

conseguida como descrito, por exemplo, no pedido de patente internacional WO98/55607.

A tecnologia designada por injecção compressiva (ver, por exemplo, www.powderject.com) também poderá ser útil na formulação de composições de vacinas.

Uma série de métodos apropriados de vacinação e de sistemas de entrega de vacinas encontra-se descrita no pedido de patente internacional WO00/29428.

Também é descrita a utilização de moléculas de ácidos nucleicos na qualidade de reagentes de diagnóstico. A detecção de uma forma mutada do gene caracterizado pelas moléculas de ácidos nucleicos, que se encontra associada a uma disfunção, irá constituir uma ferramenta de diagnóstico que poderá contribuir ou definir o diagnóstico de uma doença ou da susceptibilidade a uma doença que resulte da subexpressão, da sobreexpressão ou da expressão temporal ou espacial alterada do gene. Os indivíduos portadores de mutações no gene poderão ser detectados ao nível do ADN por uma série de técnicas.

As moléculas de ácidos nucleicos para fins de diagnóstico podem ser obtidas das células de um indivíduo, como do sangue, da urina, da saliva, de biopsia de tecido ou de material de autópsia. Pode-se empregar o ADN genómico directamente na detecção ou amplificá-lo de forma enzimática recorrendo a PCR, reacção em cadeia de ligase (LCR), amplificação por deslocamento de cadeia (SDA), ou a outras técnicas de amplificação (ver Saiki *et al.*, "Nature", 324, 163-166 (1986); Bej, *et al.*, "Crit. Rev. Biochem. Molec. Biol.", 26, 301-334 (1991); Birkenmeyer *et*

al., "J. Virol. Meth.", 35, 117-126 (1991); Van Brunt, J., "Bio/Technology", 8, 291-294 (1990)) antes da análise.

Numa execução, isto proporciona um método de diagnóstico de uma doença num doente, englobando a avaliação do nível de expressão dum gene natural que codifica um polipéptido, e a comparação desse nível de expressão com um nível de controlo, sendo que um nível diferente do referido nível de controlo é indicador de uma doença. O método pode ter as seguintes etapas:

- a) fazer contactar uma amostra de tecido do doente com uma sonda de ácidos nucleicos, sob condições estritas que permitam a formação de um complexo híbrido entre uma molécula de ácidos nucleicos e a sonda;
- b) fazer contactar uma amostra de controlo com a referida sonda sob as mesmas condições aplicadas na etapa a);
- c) detectar a presença de complexos híbridos nas referidas amostras;

em que a detecção de níveis do complexo híbrido na amostra do paciente que sejam diferentes dos níveis do complexo híbrido nas amostra de controlo é indicadora de doença.

Também está descrito um método de diagnóstico tendo as seguintes etapas:

- a) obter uma amostra de tecido de um doente que está a ser testado quanto a uma doença;
- b) isolar uma molécula de ácidos nucleicos da referida amostra de tecido; e

- c) fazer um diagnóstico do doente quanto a uma doença, por meio da detecção da presença de uma mutação na molécula de ácidos nucleicos associada à doença.

Para auxiliar à detecção de moléculas de ácidos nucleicos nos métodos acima descritos, poderá incluir-se uma etapa de amplificação, por exemplo, recorrendo a PCR.

É possível detectar as deleções e as inserções pela alteração do tamanho do produto amplificado em comparação com o genótipo normal. Consegue-se identificar mutações pontuais por meio de hibridação de ADN amplificado com ARN titulado da invenção ou, em alternativa, sequências de ADN anti-sentido tituladas da invenção. As sequências com uma correspondência perfeita podem ser distinguidas dos duplexes não correspondentes por digestão com RNase ou por avaliação das diferenças em temperaturas de fusão. A presença ou a ausência da mutação no doente pode ser detectada ao fazer contactar ADN com uma sonda de ácidos nucleicos que hibride com o ADN sob condições estritas, de modo a formar uma molécula de cadeia dupla híbrida, tendo essa molécula de cadeia dupla híbrida uma porção não-hibridada da cadeia da sonda de ácidos nucleicos em qualquer porção correspondendo a uma mutação associada a doença; e ao detectar a presença ou a ausência de uma porção não-hibridada da cadeia da sonda como indicação da presença ou da ausência de uma mutação associada a uma patologia na porção correspondente da cadeia de ADN.

Estes diagnósticos são em particular úteis nos testes pré-natais e até mesmo neonatais.

As mutações pontuais e outras diferenças de sequência entre

o gene de referência e os genes "mutantes" podem ser identificadas por outras técnicas bem conhecidas, tais como a sequenciação directa do ADN ou o polimorfismo conformacional de cadeia simples, (ver Orita *et al.*, "Genomics", 5, 874-879 (1989)). Por exemplo, é possível empregar um primer de sequenciação com produto de PCR de cadeia dupla ou uma molécula modelo de cadeia simples gerada por uma PCR modificada. Realiza-se a determinação da sequência por procedimentos convencionais com nucleótidos radiotitulados ou por procedimentos de sequenciação automática com tags fluorescentes. É igualmente possível empregar segmentos de ADN clonados como sondas para detectar segmentos de ADN específicos. A sensibilidade deste método é grandemente aumentada se combinada com PCR. Além disso, as mutações pontuais e outras variações de sequências, tais como os polimorfismos, podem ser detectadas como acima descrito através, por exemplo, da utilização de oligonucleótidos específicos de alelos para amplificação por PCR de sequências que são distintas em nucleótidos únicos.

Também é possível detectar as diferenças da sequência de ADN através de alterações na mobilidade electroforética de fragmentos de ADN em géis, com ou sem desnaturantes, ou por sequenciação directa de ADN (por exemplo, Myers *et al.*, "Science" (1985) 230:1242). As modificações da sequência em locais específicos também podem ser reveladas por ensaios de protecção contra as nucleases, tais como a protecção contra RNase e S1, ou o método de clivagem química (ver Cotton *et al.*, "Proc. Natl. Acad. Sci. USA" (1985) 85:4397-4401).

A par da electroforese em gel convencional e da

sequenciação de ADN, também é possível detectar mutações, tais como microdelecções, aneuploidias, translocações, inversões, por meio de análise *in situ* (ver, por exemplo, Keller *et al.*, "DNA Probes", 2.ª ed., Stockton Press, Nova Iorque, N. Y., EUA (1993)), ou seja, é possível analisar sequências de ADN ou de ARN em células quanto a mutações, sem ser necessário o seu isolamento e/ou a sua imobilização numa membrana. A hibridação *in situ* fluorescente (FISH) é no presente o método aplicado mais comum e apareceram diversas revisões do FISH (ver, por exemplo, Trachuck *et al.*, "Science", 250, 559-562 (1990), e Trask *et al.*, "Trends, Genet.", 7, 149-154 (1991)).

Noutra execução, é possível construir uma matriz de sondas oligonucleotídicas englobando uma molécula de ácidos nucleicos, para realizar um rastreio eficiente de variantes genéticas, mutações e polimorfismos. Os métodos de tecnologia de matrizes são bem conhecidos, têm uma aplicabilidade geral e podem ser empregues para abordar uma variedade de questões na genética molecular, inclusive expressão de genes, ligação de genes e variabilidade genética (ver, por exemplo: M. Chee *et al.*, "Science" (1996), vol. 274, pág. 610-613).

Numa execução, a matriz é preparada e empregue de acordo com os métodos descritos na aplicação PCT WO95/11995 (Chee *et al.*); Lockhart, D. J. *et al.* (1996) "Nat. Biotech." 14:1675-1680); e Schena, M. *et al.* (1996) "Proc. Natl. Acad. Sci" 93: 10614-10619). Os pares de oligonucleótidos podem ir de dois a mais de um milhão. Os oligómeros são sintetizados em áreas designadas num substrato, por meio de um processo químico dirigido por luz. O substrato pode ser papel, nylon ou outro tipo de membrana, filtro, chip,

lamela de vidro ou outro suporte sólido apropriado. Noutro aspecto, é possível sintetizar um oligonucleótido na superfície do substrato, recorrendo a um procedimento de ligação química e a um equipamento de aplicação de jacto de tinta, como descrito no pedido PCT WO95/251116 (Baldeschweiler *et al*). Noutro aspecto, uma matriz "com grelha" análoga a um dot (ou slot) blot pode ser empregue para dispor e ligar fragmentos de ADNc ou oligonucleótidos à superfície de um substrato, recorrendo a um sistema de vácuo, procedimentos de ligação térmicos, por UV, mecânicos ou químicos. Uma matriz, como as acima descritas, pode ser produzida manualmente ou empregados dispositivos disponíveis (equipamento de slot-blot ou de dot-blot), materiais (qualquer suporte sólido apropriado) e máquinas (inclusive instrumentos robóticos) e poderá conter 8, 24, 96, 384, 1536 ou 6144 oligonucleótidos, ou qualquer outro número entre 2 e mais de um milhão, que se preste à utilização eficiente da instrumentação disponível no mercado.

Além dos métodos acima debatidos, é possível diagnosticar doenças por métodos que englobem determinar, a partir de uma amostra derivada de um sujeito, um nível anormalmente reduzido ou aumentado do polipéptido ou do ARNm. É possível medir a expressão aumentada ou reduzida ao nível do ARN, recorrendo a qualquer um dos métodos bem conhecidos do estado da técnica para a quantificação de polinucleótidos, tais como, por exemplo, amplificação de ácidos nucleicos, por exemplo, PCR, RT-PCR, protecção contra RNase, Northern blotting e outros métodos de hibridação.

As técnicas de teste, que podem ser empregues na determinação dos níveis de um polipéptido numa amostra

derivada de um hospedeiro, são bem conhecidas dos especialistas da técnica e encontram-se debatidas com certo grau de detalhe acima (incluindo testes de radioimunidade, testes de ligação competitiva, análise por Western Blot e ensaios ELISA). Este aspecto proporciona um método de diagnóstico que engloba as etapas: (a) fazer contactar um ligando como acima descrito com uma amostra biológica, sob condições adequadas à formação de um complexo de ligando-polipéptido; e (b) detectar o referido complexo.

Os protocolos como ELISA, RIA e FACS para medir níveis de polipéptidos podem ainda fornecer uma base de diagnóstico de níveis alterados ou anormais da expressão de polipéptidos. Os valores normais ou padrão da expressão dos polipéptidos são estabelecidos pela combinação de fluidos corporais ou de extractos celulares tirados de sujeitos mamíferos normais, de preferência humanos, com anticorpo para o polipéptido, sob condições adequadas à formação de complexos. É possível quantificar a formação de complexos padrão por diversos métodos, como é o caso da fotometria.

Pode-se empregar os anticorpos que se ligam de forma específica a um polipéptido, com vista ao diagnóstico de condições ou doenças caracterizadas pela expressão do polipéptido, ou em ensaios para monitorizar os doentes que estão a ser tratados com os polipéptidos, moléculas de ácidos nucleicos, ligandos e outros compostos da invenção. Os anticorpos úteis para fins de diagnóstico podem ser preparados da mesma forma que aqueles acima descritos para fins terapêuticos. Os ensaios de diagnóstico para o polipéptido incluem métodos que empregam o anticorpo e um título para detectar o polipéptido em fluidos corporais humanos ou em extractos de células ou tecidos. É possível

empregar os anticorpos com ou sem modificação, e é possível titulá-los ao juntá-los, por covalência ou não, a uma molécula repórter. É possível empregar uma grande variedade de moléculas repórter conhecidas do estado da técnica, muitas das quais se encontram acima descritas.

As quantidades do polipéptido expressas nas amostras do sujeito, do controlo e da doença de tecidos sujeitos a biopsia são comparadas com os valores padrão. O desvio entre os valores padrão e os valores do sujeito estabelece os parâmetros de diagnóstico da doença. Os ensaios de diagnóstico podem ser empregues para fazer a distinção entre ausência, presença e excesso de expressão do polipéptido e para monitorizar a regulação dos níveis de polipéptido durante a intervenção terapêutica. Também é possível empregar estes ensaios para avaliar a eficácia de um regime de tratamento terapêutico particular em estudos animais, em ensaios clínicos ou na monitorização do tratamento de um doente individual.

O kit de diagnóstico da presente invenção pode conter:

- (a) uma molécula de ácidos nucleicos
- (b) um polipéptido; ou
- (c) um ligando.

Num aspecto, um kit de diagnóstico pode conter um primeiro recipiente contendo uma sonda de ácidos nucleicos que hibride, sob condições estritas, com uma molécula de ácidos nucleicos; um segundo recipiente contendo *primers* úteis na amplificação da molécula de ácidos nucleicos; e instruções de utilização da sonda e dos *primers* para facilitar o diagnóstico da patologia. O kit pode ainda conter um

terceiro recipiente tendo um agente para digerir o ARN não hibridado.

Num aspecto alternativo da invenção, um kit de diagnóstico pode conter uma matriz de moléculas de ácidos nucleicos.

Com vista a detectar um polipéptido, um kit de diagnóstico poderá conter um ou mais anticorpos que se liga(m) ao polipéptido; e um reagente útil na detecção de uma reacção de ligação entre o antícorpo e o polipéptido.

Estes kits destinam-se a ser empregues no diagnóstico de uma doença ou de um distúrbio, ou de uma susceptibilidade a uma doença ou a um distúrbio, em que estejam envolvidas proteínas endócrinas. Estas doenças e estes distúrbios poderão incluir problemas reprodutivos, problemas de gravidez, como a doença trofoblástica gestacional, problemas de desenvolvimento como o síndrome de Silver-Russel, problemas de crescimento, défice da hormona de crescimento, doença de Cushing, problemas endócrinos, problemas de proliferação celular, inclusive neoplasma, carcinoma, tumor da pituitária, tumor dos ovários, melanoma, tumor do pulmão, colo-rectal, da mama, do pâncreas, da cabeça e do pescoço, tumor trofoblástico de localização placentária, adenocarcinoma, coriocarcinoma, osteossarcoma e outros tumores sólidos; angiogéneses, distúrbios mieloproliferativos; distúrbios autoimunes/inflamatórios; problemas cardiovasculares; distúrbios neurológicos, dor; distúrbios metabólicos inclusive diabetes mellitus, osteoporose e obesidade, caquexia, SIDA, doença renal; lesão pulmonar; envelhecimento; infecções incluindo as virais, bacterianas, fúngicas e parasitárias, e outras condições patológicas. De

preferência, a patologia trata-se de uma patologia em que esteja envolvida a função endócrina, em particular hormonas de crescimento.

Em seguida, descreve-se em maior detalhe diversos aspectos e execuções da presente invenção a título de exemplo, com particular referência ao polipéptido INSP101.

Breve descrição das Figuras

A **Figura 1** mostra um alinhamento da INSP101 de comprimento total (SEQ ID NO:8) versus a P01241, hormona de crescimento pituitária (GH-N) do *H. sapiens*. O loop A-B está marcado com asteriscos.

A **Figura 2** compara o padrão de *splicing* da P01241, hormona de crescimento pituitária (GH-N) do *H. sapiens*, com o padrão de *splicing* da nova variante de *splicing* INSP101.

Figura 3: sequência de nucleótidos prevista da INSP101 com tradução. A sequência sublinhada indica a sequência de sinal prevista.

Figura 4: mapa do plasmídeo pCRII-TOPO-E00974

Figura 5: alinhamento da INSP101 com o plasmídeo #13686

Figura 6: sequência de nucleótidos e tradução do produto da INSP101 clonado

Figura 7: mapa do pENTR-INSPI01-6HIS

Figura 8: mapa do pEAK12d-INSPI01-6HIS

Exemplos

Exemplo 1

Identificou-se que o INSP101 tinha cinco exões. A Figura 2 faz a comparação entre o padrão de *splicing* da P01241, a hormona de crescimento pituitária (GH-N) do *H. sapiens* e o padrão de *splicing* da nova variante de *splicing* INSP101. O polipéptido do exão 2nov da INSP101 e o polipéptido do exão 3nov da INSP101 são exões alternativos, produzidos por *splicing* alternativo; a variante de *splicing* tem um exão2 (2nov) aumentado e um exão3 (3nov) truncado em comparação com a proteína do tipo selvagem.

O diagrama também mostra os elementos da principal estrutura secundária da hormona de crescimento pituitária (GH-N). A GH-N é composta por quatro hélices alfa (A, B, C e D), e particularmente importante é o "loop A-B" que liga a hélice A à hélice B. O loop A-B é um componente crítico da superfície de interacção da GH-N, que liga o receptor da Hormona de Crescimento (Wells JA, "PNAS" vol. 93, pág. 1-6 1996, "Binding in the growth hormone receptor complex"). É evidente que a nova variante de *splicing* INSP101 tem novos resíduos inseridos no loop A-B (devido à extensão do exão2). De forma semelhante, o truncamento do exão3 levará à remoção de alguns resíduos da GH-N no loop A-B. Assim, a INSP101 difere da GH-N principalmente em termos de composição do loop A-B e, dado que este loop é um determinante primário na ligação ao receptor da hormona de crescimento, prevê-se que a INSP101 mostre propriedades alteradas de ligação ao receptor (em termos de afinidade de ligação e/ou de selectividade do receptor).

Estas previsões experimentais serão subsequentemente

confirmadas por um teste experimental dirigido. Por exemplo, poderá recorrer-se a uma série de diferentes testes para determinar os efeitos das Hormonas de Crescimento humanas na ligação (ver, por exemplo, Well J. A., "PNAS" vol. 93, pág. 1-6 1996), inclusive a utilização de anticorpos monoclonais para precipitar complexos de 1 : 1 de hormona de crescimento e receptor, e a dimerização induzida por hGH de moléculas hGHbp em solução, por extinção de um tag fluorescente colocado perto da terminação C do hGHbp (ver Well J. A., "PNAS" vol. 93, pág. 1-6, 1996). [hGHbp = domínio extracelular do receptor da GH].

Exemplo 2: Clonagem da INSP101

1.1 Preparação do ADNc pituitário humano

Preparou-se ADNc da primeira cadeia a partir do ARN total pituitário humano (Clontech), empregando a Superscript II RNase H⁻ Reverse Transcriptase (Invitrogen) de acordo com o protocolo do fabricante. Combinou-se 1 µL do primer Oligo (dT)₁₅ (500 µg/mL, Promega), 2 µg de ARN total de pulmão humano, 1 µL de mistura de dNTP 10 mM (cada dATP, dGTP, dCTP e dTTP 10 mM com um pH neutro) e água destilada estéril para um volume final de 12 µL, num tubo eppendorf de 1,5 mL, aqueceu-se para os 65°C durante 5 min e, depois, arrefeceu-se em gelo. Colheu-se os ingredientes por breve centrifugação e adicionou-se 4 µL de 5 x tampão First-Strand, 2 µL de DTT 0,1 M e 1 µL de RNaseOUT Recombinant Ribonuclease Inhibitor (40 unidades/µL, Invitrogen). Misturou-se com cuidado os ingredientes do tubo e incubou-se nos 42°C durante 2 min, adicionou-se em seguida 1 µL (200 unidades) da enzima SuperScript II e misturou-se com cuidado com pipeta. Incubou-se a mistura nos 42°C durante

50 min e desactivou-se em seguida por aquecimento nos 70°C durante 15 min. Para remover o ARN complementar ao ADNc, adicionou-se 1 μ L (2 unidades) de *E. coli* RNase H (Invitrogen) e incubou-se a mistura de reacção nos 37°C durante 20 min. A mistura de reacção final de 21 μ L foi diluída pela adição de 179 μ L de água estéril, para originar um volume total de 200 μ L.

1.2 *Primers* de clonagem específicos de genes para PCR

Concebeu-se um par de *primers* de PCR tendo um comprimento entre 18 e 25 bases, para amplificar a sequência de codificação completa do ADNc virtual, utilizando o Primer Designer Software (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, EUA). Optimizou-se os *primers* de PCR para terem uma T_m perto de $55 \pm 10^\circ\text{C}$ e um teor de GC de 40-60%. Seleccionou-se os *primers* que tinham uma grande selectividade em relação à sequência-alvo (INSP101) com pouco ou nenhum *priming* inespecífico.

1.3 **Amplificação por PCR da INSP101 a partir do ADNc pituitário humano**

Concebeu-se *primers* de clonagem específicos de genes (o INSP101-CP1 e o INSP101-CP2, (Tabela 1)) para amplificar um fragmento de ADNc cobrindo toda a sequência codificadora da previsão da INSP101. Utilizou-se os *primers* de clonagem específicos de genes INSP101-CP1 e INSP101-CP2 com o ADNc pituitário humano como modelo. Realizou-se a PCR num volume final de 50 μ L, contendo 1 x tampão AmpliTaqTM, dNTP 200 μ M, INSP101-CP1, 50 pmoles do INSP101-CP2, 2,5 unidades do AmpliTaqTM (Perkin Elmer) e 100 ng de ADNc pulmonar, utilizando um motor de pesquisa de ADN MJ Research DNA Engine, programado como se segue: 94°C, 2 min; 40 ciclos de

94°C, 1 min, 50°C, 1 min, e 72°C, 1 min; seguido de 1 ciclo nos 72°C durante 7 min e um ciclo de espera nos 4°C. Purificou-se directamente os produtos de PCR com o sistema de purificação Wizard PCR Preps DNA Purification System (Promega).

1.4 Subclonagem de produtos de PCR

Subclonou-se produtos de PCR no vector de clonagem modificado com a topoisomerase I (pCRII TOPO), utilizando o kit de clonagem TA da Invitrogen Corporation, sob condições especificadas pelo fabricante. Resumidamente, incubou-se 4 µL de produto de PCR purificado em gel da amplificação do pool P da biblioteca humana, durante 15 min à temperatura ambiente, com 1 µL do vector TOPO e 1 µL de solução salina. Transformou-se em seguida a mistura de reacção na estirpe TOP10 da *E. coli* (Invitrogen) como se segue: um aliquota de 50 µL de células One Shot TOP10 foi imerso em gelo e adicionou-se 2 µL de reacção TOPO. Incubou-se a mistura durante 15 min em gelo e, depois, sujeitou-se a mesma a choque térmico por incubação nos 42°C durante exactamente 30 s. Voltou-se a pôr as amostras em gelo e adicionou-se 250 µL de meio SOC morno (temperatura ambiente). Incubou-se as amostras sob agitação (220 rpm) durante 1 h nos 37°C. A mistura de transformação foi depois colocada em placas L-broth (LB) contendo ampicilina (100 µg/mL) e foi incubada durante a noite nos 37°C.

1.5 PCR de colónias

Inoculou-se as colónias em 50 µL de água estéril, empregando um palito estéril. Depois, sujeitou-se um aliquota de 10 µL do meio de inoculação a PCR, num volume de reacção total de 20 µL, contendo 1 x tampão AmpliTaqTM, dNTP 200 µM, 20 pmoles do primer T7, 20 pmoles do primer

SP6, 1 unidade de AmpliTaqTM (Perkin Elmer), por intermédio de um motor de pesquisa de ADN MJ Research DNA Engine. As condições de ciclo foram as seguintes: 94°C, 2 min; 30 ciclos de 94°C, 30 s, 47°C, 30 s e 72°C durante 1 min. Em seguida, manteve-se as amostras nos 4°C (ciclo de espera) antes de outra análise.

Analisou-se os produtos de reacção PCR em géis de agarose a 1% em 1 x tampão TAE. Fez-se a cultura de uma colónia que originou o tamanho de produto de PCR previsto (+ 187 bp devido ao local de clonagem múltipla ou MCS), durante a noite nos 37°C, em 5 mL de L-Broth (LB) contendo ampicilina (100 µg/mL), sob agitação nas 220 rpm.

1.6 Preparação e sequenciação do ADN plasmídeo

Preparou-se ADN plasmídeo miniprep a partir dos 5 mL de cultura, empregando um sistema robótico Qiaprep Turbo 9600 (Qiagen) ou o kit Wizard Plus SV Minipreps (Promega cat. n.º 1460) de acordo com as instruções do fabricante. Eluiu-se o ADN plasmídeo em 100 µL de água estéril. Mediu-se a concentração de ADN utilizando o fotómetro Eppendorf BO. O ADN plasmídeo (200-500 ng) foi sujeito a sequenciação de ADN com o *primer* T7 e o *primer* SP6, utilizando o sistema BigDyeTerminator (Applied Biosystems cat. n.º 4390246) de acordo com as instruções do fabricante. As sequências dos *primers* encontram-se na Tabela 1. Purificou-se as reacções de sequenciação com colunas Dye-Ex (Qiagen) ou placas de limpeza Montage SEQ 96 (Millipore cat. n.º LSKS09624), analisou-se em seguida num sequenciador Applied Biosystems 3700. A análise das sequências revelou que a sequência clonada no plasmídeo pCRII-TOPO era idêntica a GenBank E00974 (plasmídeo 13686) (Figura 4).

1.7 Amplificação por PCR das extremidades 5' e 3' da INSP101 a partir do plasmídeo 13686

Os primers de PCR foram concebidos para amplificar individualmente as extremidades 5' e 3' da INSP101 (Tabela 1). O primer *forward* da extremidade 5' da INSP101 (INSP101-B1P-5'-F) também contém a sequência parcial do local *Gateway* attB1 (5' GCAGGCTTC) e uma sequência de Kozac (5' GCCACC). O primer *reverse* para a extremidade 5' da INSP101 (INSP101-5'-R) tem uma sobreposição de 25 bases com o primer *forward* da extremidade 3' da INSP101. O primer *forward* para a extremidade 3' da INSP101 (INSP101-3'-F) tem uma sobreposição de 26 bp com o primer *reverse* para a extremidade 5'. O primer *reverse* para a extremidade 3' da INSP101 (INSP101-3'-R) contém uma sequência 5 HIS (Figura 5).

Para amplificar a extremidade 5' da INSP101, realizou-se a reacção PCR num volume final de 50 μ L e este continha 0,5 μ L do ADN miniprep do plasmídeo 13686, 2 μ L de dNTP 5 mM (Amersham Pharmacia Biotech), 6 μ L do INSP101-B1p-5'-F (10 μ M), 6 μ L do INSP101-5'R, 5 μ L de 10 x tampão Pfu e 0,5 μ L de polimerase Pfu (3 U/ μ L) (# M774B, Promega). As condições de PCR foram 94°C durante 2 min; 30 ciclos de 94°C durante 30 s, 55°C durante 30 s e 72°C durante 1 min, e um ciclo de espera de 4°C. Carregou-se os produtos de reacção num gel de agarose a 1,5% (1 x TAE) e purificou-se os produtos de PCR com o tamanho correcto (211 bp) em gel, com um Qiaquick Gel Extraction Kit (Qiagen cat. n.º 28704) e eluiu-se os mesmos em 50 μ L de tampão de eluição (Qiagen).

Amplificou-se a extremidade 3' da INSP101 sob as mesmas condições de reacção anteriores, mas com os primers INSP101-3'-F e INSP101-3'-R, 5 μ L de 10 x tampão AmpliTaq e

0,5 µL de AmpliTaq DNA Polymerase (Applied Biosystems, # N808-0155, 5 U/uL). Purificou-se os produtos de PCR de 441 bp em gel como anteriormente.

2. Montagem das extremidades 5' e 3' da INSP101 para gerar o INSP101 ORF

Montou-se as extremidades 5' e 3' da INSP101 numa reacção PCR contendo 5 µL da extremidade 5' purificada em gel, 2 µL da extremidade 3' purificada em gel, 2 µL de dNTP 5 mM, 6 µL do INSP101-B1P-5'-F (10 µM), 6 µL do INSP101-3'-R (10 µM), 5 µL de 10 x tampão Pfu, 0,5 µL de polimerase Pfu (3 U/µL; Promega cat. n.º M774B). As condições de reacção foram: 94°C, 4 min; 10 ciclos de 94°C durante 30 s, 48°C durante 30 s e 70°C durante 2 min; 25 ciclos de 94°C durante 30 s, 52°C durante 30 s e 70°C durante 2 min; uma etapa de prolongamento adicional de 70°C durante 10 min; e um ciclo de espera nos 4°C. Analisou-se os produtos de reacção num gel de agarose a 1,5% (1 x TAE). Purificou-se em gel os produtos de PCR com o tamanho previsto (627 bp) utilizando um Qiaquick Gel Extraction Kit (Qiagen cat. n.º 28704) e eluiu-se em 50 µL de tampão de eluição (Qiagen). O produto resultante (INSP101 ORF) contém o ORF da INSP101, tendo adjacente à extremidade 5' um local attB1 e a sequência de Kozak, e à extremidade 3' uma sequência codificadora 5HIS tag (Figura 6).

3. Subclonagem do INSP101 ORF no pDONR221

Subclonou-se o INSP101 ORF no pDONR221 utilizando o sistema de clonagem Gateway™ (Invitrogen). Adicionou-se um local de recombinação attB1 parcial à extremidade 5' do INSP101 ORF e uma sequência 6HIS tag, adicionou-se um códão stop e um local de recombinação attB2 à extremidade 3' do INSP101 ORF, numa reacção PCR de 50 µL contendo 2 µL de produto de

PCR INSP101 ORF purificado em gel, 2 μ L de dNTP 5 mM (Amersham Pharmacia Biotech), 3 μ L de GCP-Forward (10 μ M), 3 μ L de GCP-Reverse (10 μ M), 5 μ L de 10 x tampão AmpliTaq e 0,5 μ L de AmpliTaq DNA Polymerase (Applied Biosystems, # N808-0155, 5 U/ μ L) num volume final de 50 μ L. As condições da PCR eram 94°C durante 2 min; 30 ciclos de 94°C durante 30 s; 55°C durante 30 s e 72°C durante 1 min; uma etapa de prolongamento adicional de 72°C durante 30 min e um ciclo de espera de 4°C. Analisou-se os produtos de reacção num gel de agarose a 1,5% (1 x TAE), purificou-se os produtos de PCR com o tamanho correcto (685 bp) em gel com um Qiaquick Gel Extraction Kit (Qiagen cat. n.º 28 704) e eluiu-se em 50 μ L de tampão de eluição (Qiagen). O produto de PCR purificado (INSP101 ORF modificado por Gateway) foi depois transferido para pDONR221, utilizando a BP clonase como se segue: incubou-se 1,5 μ L de INSP101 ORF modificado por Gateway com 0,5 μ L de pDONR221 (0,1 μ g/ μ L), 1 μ L de tampão BP e 1 μ L de mistura de enzima BP clonase (Invitrogen) num volume final de 5 μ L, à temperatura ambiente durante 2 h. Interrompeu-se a reacção pela adição de 0,5 μ L de proteinase K (2 μ g/ μ L) e incubou-se nos 37°C durante mais 10 min. Empregou-se um aliquota desta reacção (3 μ L) para transformar 50 μ L de célula *E. coli* DH5alpha (Invitrogen) por choque térmico. Incubou-se a mistura de transformação num tubo de PCR de 0,2 mL em gelo durante 30 min, e transferiu-se depois para um banho de água nos 42°C durante 45 s. Voltou-se a pôr os tubos em gelo durante 2 min e transferiu-se a mistura de transformação para um tubo de polipropileno de 12 mL. Diluiu-se em seguida as amostras pela adição de 900 μ L de meio SOC e incubou-se durante 1 h nos 37°C sob agitação. Os transformados (200 μ L) foram postos em placas de LB contendo 40 μ g/ μ L de canamicina e foram incubados durante 72 h à temperatura ambiente. O ADN

miniprep plasmídeo foi isolado de 8 colónias resultantes, utilizando um sistema robótico Qiaprep Turbo 9600 (Qiagen), e foi sujeito a sequenciação de ADN com os *primers* de sequenciação M13F e M13R com o sistema BigDyeTerminator (Applied Biosystems cat. n.º 4390246) de acordo com as instruções do fabricante. Purificou-se as reacções de sequenciação com colunas Dye-Ex (Qiagen) ou placas de limpeza Montage SEQ 96 (Millipore cat. n.º LSKS09624), depois analisou-se num sequenciador Applied Biosystems 3700.

3. Mutagénese do pDONR221-INSPI01

A verificação da sequenciação mostrou que todos os clones pDONR-221-INSPI01 tinham um códão stop entre ORF e o 6HIS tag. Recorreu-se a mutagénese dirigida ao local para delectar o códão stop. Empregou-se um Stratagene Quik Change XL Site-Directed Mutagenesis kit (cat. # 200517) e os *primers* INSP101-mut-F e INSP101-mut-R de acordo com as instruções do fabricante. Isolou-se ADN miniprep plasmídeo de 8 das colónias resultantes com um sistema robótico Qiaprep Turbo 9600 (Qiagen), e sujeitou-se o mesmo a sequenciação de ADN com os *primers* de sequenciação M13F e M13R com o sistema BigDye Terminator (Applied Biosystems cat. n.º 4390246), de acordo com as instruções do fabricante. Purificou-se as reacções de sequenciação com colunas Dye-Ex (Qiagen) ou placas de limpeza Montage SEQ 96 (Millipore cat. n.º LSKS09624), depois analisou-se num sequenciador Applied Biosystems 3700.

4. Subclonagem do INSP101 ORF no vector de expressão pEAK12d

O eluato de plasmídeo (1,5 µL) de um clone pDONR221 contendo a sequência correcta do INSP101 ORF (pENTR-

INSP101-6HIS, ID do plasmídeo 14577) (Figura 7) foi depois utilizado numa reacção de recombinação contendo 1,5 μ L de pEAK12d (0,1 μ g/ μ L), 2 μ L de tampão LR e 1,5 μ L de LR clonase (Invitrogen), num volume final de 10 μ L. Incubou-se a mistura à temperatura ambiente durante 1 h, interrompeu-se a mesma pela adição de 1 μ L de proteinase K (2 μ g/ μ L) e incubou-se nos 37°C durante mais 10 min. Empregou-se um aliquota desta reacção (1 μ L) para transformar 20 μ L de células *E. coli* DH10B (Invitrogen) (diluídas 1/5 em água estéril) por electroporação com um Biorad Gene Pulser, de acordo com as recomendações do fabricante. As células sujeitas a electroporação foram transferidas para tubos de polipropileno de 12 mL, foram diluídas pela adição de 1 mL de meio SOC e incubadas durante 1 h nos 37°C. Os transformados foram colocados em placas de LB-ampicilina e foram incubados durante a noite nos 37°C. Preparou-se ADN miniprep a partir de 3 colónias com um sistema robótico Qiaprep Turbo 9600 (Qiagen) e eluíu-se com 50 μ L de tampão de eluição. Sujeitou-se em seguida 0,5 μ L de cada miniprep a PCR num volume de reacção total de 50 μ L, contendo 2 μ L de dNTP 5 mM, 3 μ L de pEAK12-F 10 μ M, 3 μ L de pEAK12-R 10 μ M, 5 μ L de 10 x tampão AmpliTaqTM e 0,5 μ L de AmpliTaqTM (Applied Biosystems cat. n.º N808-0155). As condições de ciclo foram as seguintes: 94°C durante 2 min; 30 ciclos de 94°C durante 30 s, 55°C durante 30 s, e 72°C durante 1 min; 1 ciclo, 72°C durante 3 min. Manteve-se depois as amostras nos 4°C (ciclo de espera) antes de outra análise.

Analisou-se os produtos de reacção PCR em géis de agarose a 1% em 1 x tampão TAE. Uma colónia que deu origem ao tamanho de produto de PCR previsto (1 010 bp) foi empregue na preparação de ADN maxiprep purificado em gradiente de CsCl do plasmídeo pEAK12d-INSP101-6HIS (ID do plasmídeo 14578)

(Figura 6), a partir de 500 mL de cultura (Sambrook J. et al., in "Molecular Cloning, a Laboratory Manual", 2.ª edição, 1989, Cold Spring Harbor Laboratory Press). O ADN plasmídeo resultante foi ressuspenso com uma concentração de 1 μ g/ μ L em tampão TE e a sequência foi verificada com os primers pEAK12d F e pEAK12d R como acima descrito.

Tabela 1

Primers para a clonagem e para a sequenciação da INSP101

Sequência sublinhada = sequência de Kozak

A negrito = códão stop

Sequência a itálico = His Tag

 = sobreposição com parte de ADNc adjacente

Exemplo 3: Expressão e purificação da INSP101

Nesta altura é possível realizar mais experiências para determinar a distribuição tecidual e os níveis de expressão dos polipeptídos INSP101 *in vivo*, com base na sequência nucleotídica e de aminoácidos revelada no presente documento.

É possível investigar a presença dos transcritos para a INSP101 por PCR de ADNc de diferentes tecidos humanos. Os transcritos da INSP101 podem estar presentes com níveis muito baixos nas amostras testadas. Por isso, é necessário um extremo cuidado na concepção de experiências para estabelecer a presença de um transcrito em diversos tecidos humanos, visto que uma pequena quantidade de contaminação genómica na preparação de ARN originará um falso resultado positivo. Assim, todo o ARN deverá ser tratado com DNase antes de se empregar a transcrição reversa. Além disso, pode estabelecer-se para cada tecido uma reacção de controlo em que não tenha tido lugar a transcrição reversa (um controlo de TR negativo = *-ve RT control*).

Por exemplo, 1 µg do ARN total de cada tecido pode ser empregue para gerar ADNc, usando a transcriptase reversa Multiscript (ABI) e primers hexaméricos aleatórios. Para cada tecido, estabelece-se uma reacção de controlo em que se adiciona todos os constituintes, à excepção da transcriptase reversa (*-ve RT control*). Estabelece-se as reacções PCR para cada tecido nas amostras de ARN transcrito reverso e os controlos de transcriptase reversa negativos. Pode-se facilmente conceber primers específicos da INSP101 com base na informação de sequência apresentada no presente documento. A presença de um produto com o peso molecular correcto na amostra que sofreu transcrição

reversa, em conjunto com a ausência de um produto no controlo de TR negativo, poderá ser tida como prova da presença de um transcrito nesse tecido. Pode empregar-se qualquer biblioteca adequada de ADNc para rastrear os transcritos da INSP101, não apenas as produzidas como acima descrito.

O padrão de distribuição tecidual dos polipéptidos INSP101 fornecerá mais informação útil acerca da função desses polipéptidos.

Além disso, é possível realizar outras experiências utilizando os vectores de expressão pEAK12d-INSP101-6HIS. A transfecção de linhas celulares de mamíferos com estes vectores poderá permitir o elevado nível de expressão das proteínas da INSP101 e, assim, permitir a continuação da investigação das características funcionais dos polipéptidos INSP101. O material e os métodos que se seguem constituem exemplo dos que se adequam a estas experiências:

Cultura celular

Mantém-se células renais embrionárias humanas (*Human Embryonic Kidney cells*) 293 que expressam o antígeno nuclear do vírus Epstein-Barr (HEK293-EBNA, Invitrogen) em suspensão em meio sem soro Ex-cell VPRO (cultura-mãe, meio de manutenção, JRH). Dezasseis a 20 horas antes da transfecção (dia -1), faz-se a cultura das células em 2 x balões T225 (50 mL por balão em DMEM/F12 (1 : 1), contendo 2% de meio de cultura FBS (JRH) com uma densidade de 2 x 10⁵ células/mL). No dia seguinte (dia de transfecção 0), a transfecção tem lugar com o reagente JetPEITM (2 µL/µg de ADN plasmídeo, transfecção PolyPlus). Para cada balão, co-transfeta-se ADN plasmídeo com ADN de GFP (gene repórter

fluorescente). Adiciona-se em seguida a mistura de transfecção aos balões de 2 x T225 e incuba-se nos 37°C (CO₂ a 5%) durante 6 dias. É possível efectuar a confirmação de transfecção positiva através de análise fluorescente qualitativa no dia 1 e no dia 6 (Axiovert 10 Zeiss).

No dia 6 (dia da colheita), junta-se os sobrenadantes dos dois balões e centrifuga-se os mesmos (por exemplo, 4°C, 400 g) e coloca-se numa cuba tendo um único identificador. Um alíquota (500 µL) é mantido para controlo da qualidade da proteína com o 6HIS-tag (CQ de bioprocessamento interno).

É possível produzir misturas à grande escala seguindo o protocolo denominado "transfecção PEI de células em suspensão", com a referência BP/PEI/HH/02/04, com polietilenoimina (*PolyEthyleneImine*) da Polysciences como agente de transfecção.

Processo de purificação

A amostra de meio de cultura, contendo a proteína recombinante com um 6His tag de terminação C, é diluída com tampão A frio (NaH₂PO₄ 50 mM; NaCl 600 mM; 8,7% (w/v) de glicerol, pH de 7,5). Depois, filtra-se a amostra através de um filtro estéril (Millipore) e mantém-se nos 4°C num frasco de secção quadrangular estéril (Nalgene).

Realiza-se a purificação nos 4°C numa estação de trabalho VISION (Applied Biosystems) ligada a um carregador de amostras automático (Labomatic). O procedimento de purificação é composto por duas etapas sequenciais, cromatografia de afinidade com metal numa coluna Poros 20

MC (Applied Biosystems) carregada com iões de Ni (4,6 x 50 mm, 0,83 mL), a que se segue filtração em gel numa coluna (1,0 x 10 cm) de meio Sephadex G-25 (Amersham Pharmacia).

No caso da primeira etapa cromatográfica, regenera-se a coluna de afinidade com metal com 30 volumes da coluna de solução de EDTA (EDTA 100 mM; NaCl 1 M; pH de 0,8), recarrega-se a mesma com iões de Ni através de lavagem com 15 volumes de coluna de uma solução de NiSO_4 100 mM, lavada com 10 volumes de coluna do tampão A, a que se seguiu 7 volumes de coluna do tampão B (NaH_2PO_4 50 mM; NaCl 600 mM; 8,7% (w/v) de glicerol, 400 mM; imidazol, pH de 7,5) e, por fim, equilibrou-se com 15 volumes de coluna do tampão A contendo imidazol 15 mM. Transfere-se a amostra, com o carregador de amostras Labomatic, para um loop de amostras de 200 mL e, em seguida, carrega-se a mesma na coluna de afinidade com metal Ni com um débito de 10 mL/min. Lava-se a coluna com 12 volumes de coluna do tampão A, a que se segue 28 volumes de coluna do tampão A contendo imidazol 20 mM. Durante a lavagem com imidazol 20 mM, as proteínas contaminantes fracamente ligadas são eluídas para fora da coluna. A proteína recombinante com His tag é por fim eluída com 10 volumes de coluna do tampão B, com um débito de 2 mL/min, e colhe-se a proteína eluída.

No caso da segunda etapa de cromatografia, regenera-se a coluna de filtração em gel Sephadex G-25 com 2 mL do tampão D (NaCl 1,137 M; KCl 2,7 mM; KH_2PO_4 1,5 mM; Na_2HPO_4 8 mM; pH de 7,2), e equilibra-se depois a mesma com 4 volumes de coluna do tampão C (NaCl 137 mM; KCl 2,7 mM; KH_2PO_4 1,5 mM; Na_2HPO_4 8 mM; 20% (w/v) de glicerol; pH de 7,4). A fracção de pico, eluída para fora da coluna de Ni, é automaticamente carregada na coluna Sephadex G-25 através

do carregador de amostras integrado no VISION, e a proteína é eluída com o tampão C com um débito de 2 mL/min. Filtrou-se a fracção através de um filtro de centrifugação estéril (Millipore), congelou-se e armazenou-se a mesma nos -80°C. Analisa-se um alíquota da amostra em SDS-PAGE (4-12% de gel NuPAGE; Novex) Western blot com anticorpos anti-His. Pode-se corar o gel NuPAGE numa solução corante Coomassie blue R250 a 0,1% (30% de metanol, 10% de ácido acético), à temperatura ambiente durante 1 h e, em seguida, descolorá-lo em 20% de metanol, 7,5% de ácido acético, até o fundo estar limpo e as bandas de proteínas estarem claramente visíveis.

A seguir à electroforese, as proteínas são electrotransferidas do gel para uma membrana de nitrocelulose. A membrana é bloqueada com 5% de leite em pó no tampão E (NaCl 137 mM; KCl 2,7 mM; KH₂PO₄ 1,5 mM; Na₂HPO₄ 8 mM; 0,1% de Tween 20, pH de 7,4), durante 1 h à temperatura ambiente, e é depois incubada com uma mistura de 2 anticorpos anti-His policlonais de coelho (G-18 e H-15, 0,2 µg/mL cada; Santa Cruz) em 2,5% de leite em pó no tampão E, durante a noite nos 4°C. Passada mais 1 hora de incubação à temperatura ambiente, lava-se a membrana com o tampão E (3 x 10 min) e incuba-se depois com um antícorpo anti-coelho HRP-conjugado secundário (DAKO, HRP 0399) diluído de 1/3 000 em tampão E contendo 2,5% de leite em pó, durante 2 horas à temperatura ambiente. Após a lavagem com o tampão E (3 x 10 minutos), a membrana é revelada com o kit ECL (Amersham Pharmacia) durante 1 min. Em seguida, a membrana é exposta a uma película Hyperfilm (Amersham Pharmacia), a película é revelada e a imagem de Western blot é analisada visualmente.

No caso das amostras que mostraram bandas de proteínas detectáveis por coloração Coomassie, é possível determinar a concentração proteica utilizando o kit de teste proteico BCA (Pierce) com albumina de soro bovino como padrão.

Além disso, a sobreexpressão ou a redução da expressão dos polipéptidos nas linhas celulares pode ser utilizada para determinar o efeito na activação transcripcional do genoma da célula hospedeira. Os parceiros de dimerização, co-activadores e co-repressores do polipéptido INSP101 podem ser identificados por imunoprecipitação combinada com Western blotting e imunoprecipitação combinada com espectroscopia de massa.

Lista de sequências específicas da INSP101 (nota: no caso de aminoácidos codificados por junções de exão-exão, o aminoácido será atribuído ao exão mais 5')

SEQ ID NO: 1 (sequência nucleotídica do exão 2nov da INSP101)

1 GGCTCCCGGA CGTCCCTGCT CCTGGCTTT GGCCTGCTCT GCCTGCCCTG
51 GCTTCAAGAG GGCAGTGCCT TCCCAACCAT TCCCTTATCC AGGCTTTTG
101 ACAACGCTAT GCTCCGCGCC CATCGTCTGC ACCAGCTGGC CTTTGACACC
151 TACCAGGAGT TTGTAAGCTC TTGGGGAATG G

SEQ ID NO: 2 (sequência proteica do exão 2nov da INSP101)

1 GSRTSLLLAF GLLCLPWLQE GSAFPTIPLS RLFDNAMLRA HRLHQLAFDT
51 YQEFVSSWGM E

SEQ ID NO: 3 (sequência nucleotídica do exão 3nov da INSP101)

1 AGTCTATTCC GACACCCTCC AACAGGGAGG AACACACAACA GAAATCC

SEQ ID NO: 4 (sequência proteica do exão 3nov da INSP101)

1 SIPTPSNREE TQQKS

SEQ ID NO: 5 (sequência nucleotídica contígua dos exões 2nov e 3nov da INSP101)

1 GGCTCCCGGA CGTCCCTGCT CCTGGCTTT GGCCTGCTCT GCCTGCCCTG
51 GCTTCAAGAG GGCAGTGCCT TCCCAACCAT TCCCTTATCC AGGCTTTTG
101 ACAACGCTAT GCTCCGCGCC CATCGTCTGC ACCAGCTGGC CTTTGACACC
151 TACCAGGAGT TTGTAAGCTC TTGGGGAATG GAGTCTATTC CGACACCCCTC
201 CAACAGGGAG GAAACACAAC AGAAATCC

SEQ ID NO: 6 (sequência proteica contígua dos exões 2nov e

3nov da INSP101)

1 GSRTSLLLAF GLLCLPWLQE GSAFPTIPLS RLFDNAMLRA HRLHQLAFDT
51 YQEFVSSWGM ESIPTPSNRE ETQQKS

SEQ ID NO: 7 (sequência nucleotídica de comprimento total da INSP101)

1 ATGGCTACAG GCTCCCGGAC GTCCCTGCTC CTGGCTTTG GCCTGCTCTG
51 CCTGCCCTGG CTTCAAGAGG GCAGTGCCTT CCCAACATT CCCTTATCCA
101 GGCTTTTGA CAACGCTATG CTCCGCGCCC ATCGTCTGCA CCAGCTGGCC
151 TTTGACACCT ACCAGGAGTT TGTAAGCTCT TGGGAAATGG AGTCTATTCC
201 GACACCCTCC AACAGGGAGG AAACACAACA GAAATCCAAC CTAGAGCTGC
251 TCCGCATCTC CCTGCTGCTC ATCCAGTCGT GGCTGGAGCC CGTGCAGTTC
301 CTCAGGAGTG TCTTCGCCAA CAGCCTGGTG TACGGCGCCT CTGACAGCAA
351 CGTCTATGAC CTCCTAAAGG ACCTAGAGGA AGGCATCCAA ACGCTGATGG
401 GGAGGCTGGA AGATGGCAGC CCCCGGACTG GGCAGATCTT CAAGCAGACC
451 TACAGCAAGT TCGACACAAA CTCACACAAAC GATGACGCAC TACTCAAGAA
501 CTACGGGCTG CTCTACTGCT TCAGGAAGGA CATGGACAAG GTCGAGACAT
551 TCCTGCGCAT CGTGCAGTGC CGCTCTGTGG AGGGCAGCTG TGGCTTCTAG

SEQ ID NO: 8 (sequência proteica de comprimento total da INSP101)

1 MATGSRTSLL LAFGLLCLPW LQEGLSAFPTI PLSRLFDNAM LRAHRLHQLA
51 FDTYQEFVSS WGMESIPTPS NREETQQKSN LELLRISLLL IQSWLEPVQF
101 LRSVFANSLV YGASDSNVYD LLKDLEEGIQ TLMGRLEDGS PRTGQIFKQT
151 YSKFDTNSHN DDALLKNYGL LYCFRKDMDK VETFLRIVQC RSVEGSCGF

SEQ ID NO: 9 (sequência nucleotídica de comprimento total da INSP101 - sem região peptídica de sinal)

1 TTCCCAACCA TTCCCTTATC CAGGCTTTTG GACAACGCTA TGCTCCGCGC
51 CCATCGTCTG CACCAGCTGG CCTTTGACAC CTACCAGGAG TTTGTAAGCT
101 CTTGGGAAAT GGAGTCTATT CCGACACCCCT CCAACAGGGA GGAAACACAA
151 CAGAAATCCA ACCTAGAGCT GCTCCGCATC TCCCTGCTGC TCATCCAGTC
201 GTGGCTGGAG CCCGTGCAGT TCCTCAGGAG TGTCTTCGCC AACAGCCTGG
251 TGTACGGCGC CTCTGACAGC AACGTCTATG ACCTCCTAAA GGACCTAGAG
301 GAAGGCATCC AAACGCTGAT GGGGAGGCTG GAAGATGCCA GCCCCCCGGAC
351 TGGGCAGATC TTCAAGCAGA CCTACAGCAA GTTCGACACA AACTCACACA
401 ACGATGACGC ACTACTCAAG AACTACGGGC TGCTCTACTG CTTCAGGAAG
451 GACATGGACA AGGTCGAGAC ATTCCCTGCGC ATCGTGCAGT GCCGCTCTGT
501 GGAGGGCAGC TGTGGCTTCT AG

SEQ ID NO: 10 (sequência proteica de comprimento total da INSP101 - sem região peptídica de sinal)

1 FPTIPLSRLF DNAMLRAHRL HQLAFDTYQE FVSSWGMESI PTPSNREETO
51 QKSNLELLRI SLLLIQSWLE PVQFLRSVFA NSLVYGASDS NVYDLLKDLE
101 EGIQTLMGRL EDGSPRTGQI FKQTYSKFDT NSHNDALLK NYGLLYCFRK
151 DMDKVETFLR IVQCRSVEGS CGF

Lisboa, 29 de Janeiro de 2010

REIVINDICAÇÕES

1. Um polipéptido isolado que funciona como hormona de crescimento e que:
 - (i) contém ou consiste na sequência de aminoácidos indicada na SEQ ID NO:8 ou na SEQ ID NO:10; ou
 - (ii) tem uma identidade de sequência superior a 95% com o polipéptido de (i).
2. Um polipéptido de acordo com a reivindicação 1, que contém ou consiste numa sequência de aminoácidos como indicada na SEQ ID NO:8 ou na SEQ ID NO:10.
3. Uma proteína de fusão incluindo um polipéptido de acordo com uma das reivindicações anteriores.
4. Uma molécula de ácidos nucleicos purificada, que codifica um polipéptido de acordo com uma das reivindicações anteriores.
5. Uma molécula de ácidos nucleicos purificada de acordo com a reivindicação 4, que contém a sequência de ácidos nucleicos indicado na SEQ ID NO:7 ou na SEQ ID NO:9.
6. Uma molécula de ácidos nucleicos purificada de acordo com a reivindicação 4, que consiste na sequência de ácidos nucleicos indicada na SEQ ID NO:7 ou na SEQ ID NO:9.
7. Um vector que contém uma molécula de ácidos nucleicos indicada numa das reivindicações 4 a 6.

8. Uma célula hospedeira isolada, transformada com um vector de acordo com a reivindicação 7.
9. Um polipéptido de acordo com uma das reivindicações 1 a 3, uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6, um vector de acordo com a reivindicação 7, ou uma célula hospedeira isolada de acordo com a reivindicação 8, destinado/a a ser empregue na terapia ou no diagnóstico de uma patologia.
10. Um polipéptido de acordo com uma das reivindicações 1 a 3, a ser empregue como hormona de crescimento ou como modulador da actividade da hormona de crescimento.
11. Uma composição farmacêutica contendo um polipéptido de acordo com uma das reivindicações 1 a 3, uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6, um vector de acordo com a reivindicação 7, ou uma célula hospedeira isolada de acordo com a reivindicação 8.
12. Uma composição de uma vacina contendo um polipéptido de acordo com uma das reivindicações 1 a 3, ou uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6.
13. Um polipéptido de acordo com uma das reivindicações 1 a 3, uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6, um vector de acordo com a reivindicação 7, uma célula hospedeira isolada de acordo com a reivindicação 8, ou uma composição

farmacêutica de acordo com a reivindicação 11, a ser empregue na produção de um medicamento destinado ao tratamento de problemas de reprodução, problemas de gravidez, problemas de desenvolvimento, problemas de crescimento, défice da hormona de crescimento, doença de Cushing, problemas endócrinos, problemas proliferativos celulares, inclusive neoplasma, carcinoma, tumor da pituitária, tumor nos ovários, melanoma, tumor do pulmão, colo-rectal, da mama, do pâncreas, da cabeça e do pescoço, tumor trofoblástico de localização placentária, adenocarcinoma, coriocarcinoma, osteossarcoma e outros tumores sólidos; angiogénes, distúrbios mieloproliferativos; distúrbios autoimunes/inflamatórios; distúrbios cardiovasculares; distúrbios neurológicos, dor; distúrbios metabólicos; lesão pulmonar; envelhecimento; infecções.

14. A utilização do polipéptido, da molécula de ácidos nucleicos, do vector, da célula hospedeira isolada ou da composição farmacêutica de acordo com a reivindicação 13, em que o problema de gravidez é a doença trofoblástica gestacional.
15. A utilização do polipéptido, da molécula de ácidos nucleicos, do vector, da célula hospedeira isolada ou da composição farmacêutica de acordo com a reivindicação 13, em que o problema de desenvolvimento é o sindroma de Silver-Russel.
16. A utilização do polipéptido, da molécula de ácidos nucleicos, do vector, da célula hospedeira isolada ou da composição farmacêutica de acordo com a

reivindicação 13, em que o problema metabólico é seleccionado dentre diabetes mellitus, osteoporose, obesidade, caquexia, SIDA e doença renal.

17. A utilização do polipéptido, da molécula de ácidos nucleicos, do vector, da célula hospedeira isolada ou da composição farmacêutica de acordo com a reivindicação 13, em que a infecção é seleccionada dentre infecção viral, infecção bacteriana, infecção fúngica e infecção parasitária.
18. Um polipéptido de acordo com uma das reivindicações 1 a 3, uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6, um vector de acordo com a reivindicação 7, uma célula hospedeira isolada de acordo com a reivindicação 8 ou uma composição farmacêutica de acordo com a reivindicação 11, a ser empregue na produção de um medicamento destinado ao tratamento de uma doença em que tomam parte proteínas de hormona do crescimento.
19. Um método de identificar um composto que seja eficaz no tratamento e/ou no diagnóstico de uma patologia, contemplando fazer contactar um polipéptido de acordo com uma das reivindicações 1 a 3 ou uma molécula de ácidos nucleicos de acordo com uma das reivindicações 4 a 6, com um ou mais compostos suspeitos de terem afinidade de ligação em relação ao referido polipéptido ou à referida molécula de ácidos nucleicos, e seleccionar um composto que se liga de forma específica à referida molécula de ácidos nucleicos ou ao referido polipéptido.

Lisboa, 29 de Janeiro de 2010

Figura 1: INSP101 de comprimento total (SEQ ID NO:6) versus P01241, hormona de crescimento pituitária (GH-N) do *H. sapiens*:

Query: 1 MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFVSS 60
 MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEF +
 Sbjct: 1 MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEA 60

Query: 61 W-----GMESIPTPSNREETQQKSNLELLRISLLLQSWLEPVQFLR 102
 + ESIPTPSNREETQQKSNLELLRISLLLQSWLEPVQFLR
 Sbjct: 61 YIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLQSWLEPVQFLR 120

Query: 103 SVFANSLVYGASDSNVYDLLKLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDD 162
 SVFANSLVYGASDSNVYDLLKLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDD
 Sbjct: 121 SVFANSLVYGASDSNVYDLLKLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDD 180

Query: 163 ALLKNYGLLYCFRKDMKDVKETFLRIVQCRSVEGSCGF 199
 ALLKNYGLLYCFRKDMKDVKETFLRIVQCRSVEGSCGF
 Sbjct: 181 ALLKNYGLLYCFRKDMKDVKETFLRIVQCRSVEGSCGF 217

Figura 2: Estrutura do gene

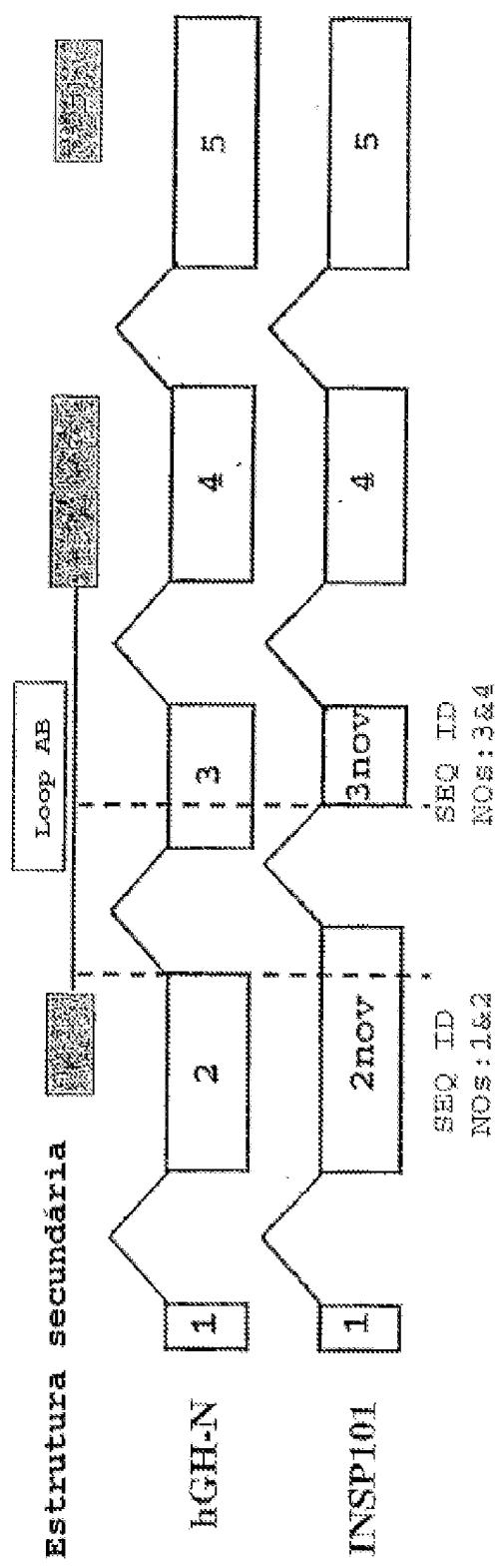


Figura 3: Sequência nucleotídica prevista da INSP101 com tradução. A sequência sublinhada indica a sequência de sinal prevista

```

1 atggctacag gctccggac gtccctgctc ctggctttg gcctgtctg
m a t g s r t s l l l a f g l l

51 cctgcctgg cttaagagg gcagtgcctt cccaaaccatt cccttatcca
c l p w l q e g s a f p t i p l s

101 ggcttttga caacgctatg ctccgcgccc atcgcttgca ccagctggcc
r l f d n a m l r a h r l h q l a

151 tttgacacct accaggagtt tgtaagctct tggggatgg agtctattcc
f d t y q e f v s s w g m e s i

201 gacaccctcc aacagggagg aaacacaaca gaaatccaac ctagagctgc
p t p s n r e e t q q k s n l e l

251 tccgcatctc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc
l r i s l l l i q s w l e p v q f

301 ctcaggagtg tcttcgccaa cagcctggtg tacggcgct ctgacaccaa
l r s v f a n s l v y g a s d s

351 cgtctatgac ctcctaaagg acctagagga aggcatccaa acgctgatgg
n v y d l l k d l e e g i q t l m

401 ggaggctgga agatggcagc ccccgactg ggcagatctt caagcagacc
g r l e d g s p r t q q i f k q t

451 tacagcaagt tcgacacaaa ctcacacaac gatgacgcac tactcaagaa
y s k f d t n s h n d d a l l k

501 ctacgggctg ctctactgct tcaggaagga catggacaag gtcgagacat
n y g l l y c f r k d m d k v e t

551 tcctgcgcac cgtgcagtc cgctctgtgg agggcagctg tggcttc
f l r i v q c r s v e g s c g f

```

Figura 4: Mapa do plasmídeo pCRII-TOPO-E00974

Molécula: produto2, 4 563 bp ADN circular
 Nome ficheiro: 13686[1].cm5

Descrição: Ligação de NoName invertido no PCRII-TOPO-open

Descrição	Nome	Fim	Início	Descrição
GENE	LacZa'	336	1	
REGIÃO	SP6	256	239	Promotor Sp6
GENE	Cds	338	943	cds inserido (E00974)
REGIÃO	Insert	337	949	Produto PCR inserido
GENE	'LacZa	201	950	
REGIÃO	T7	038	019	Promotor T7
GENE	f1	617	203	f1 ori
GENE	KanR	745	951	Gene resistência canamicina
GENE	AmpR	623	763	Gene resistência ampicilina
GENE	pUC	441	768	pUC ori

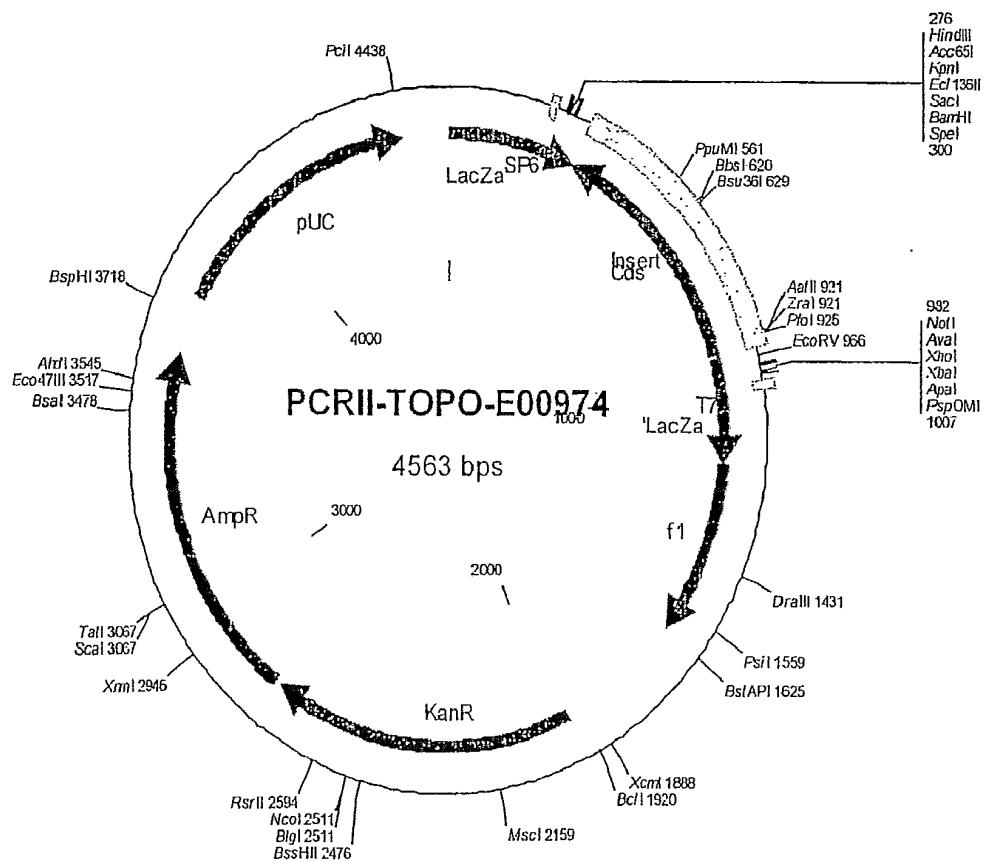


Figura 5: Alinhamento da INSP101 com o plasmídeo #13686

Em cima = INSP101

Em baixo = 13686

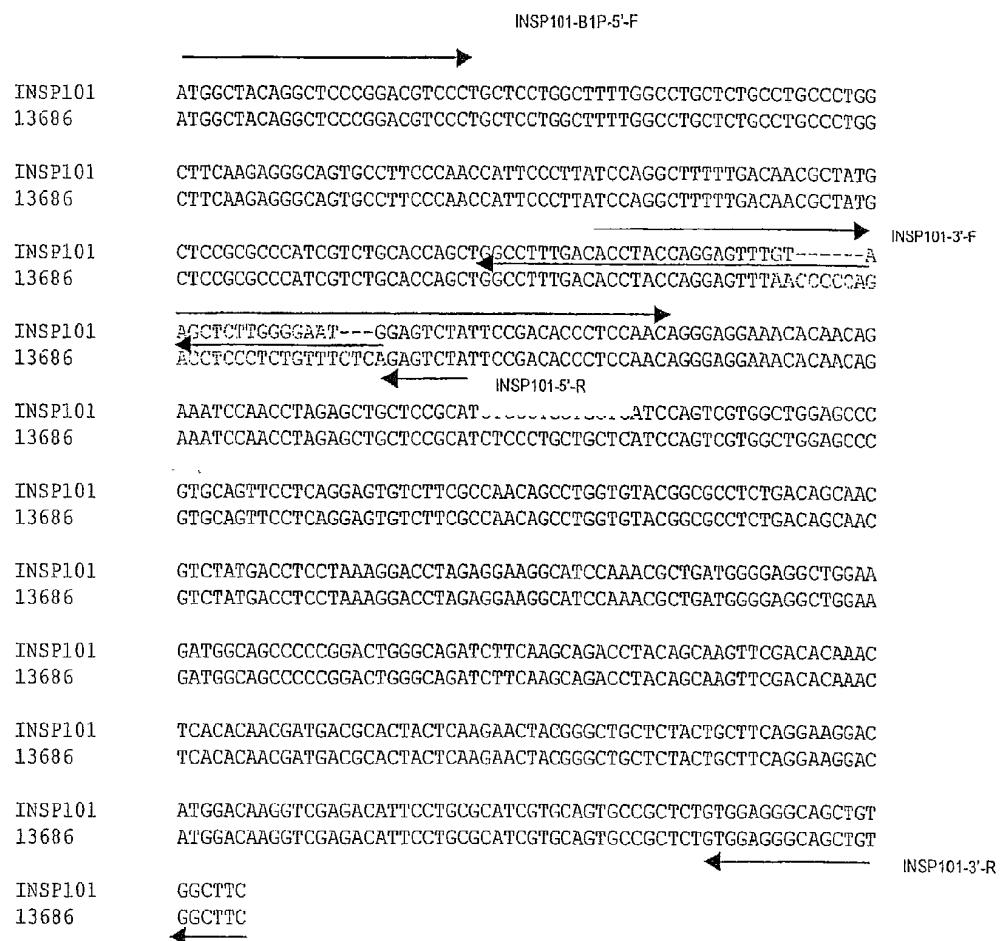


Figura 6: Sequência nucleotídica e tradução do produto da INSP101 clonado

```

1  acaagtttgt acaaaaaagc aggcttgc accatggcta caggctcccg
               m a t g s

51  gacgtccctg ctcctggctt ttggcctgct ctgcctgccc tggcttcaag
      r t s l l a f g l l c l p w l q

101 agggcagtgc cttcccaacc attcccttat ccaggcttt tgacaacgct
      e g s a f p t i p l s r l f d n a

151 atgctccgcg cccatcgtct gcaccagctg gccttgaca cctaccagga
      m l r a h r l h q l a f d t y q

201 gtttgtaagc tcttgggaa tggagtctat tccgacaccc tccaacaggg
      e f v s s w g m e s i p t p s n r

251 aggaaacaca acagaaatcc aacctagagc tgctccgcatttccctgt
      e e t q q k s n l e l l r i s l l

301 ctcatccagt cgtggctgga gcccgtgcag ttccctcagga gtgtcttcgc
      l i q s w l e p v q f l r s v f

351 caacagcctg gtgtacggcg cctctgacag caacgtctat gacctctaa
      a n s l v y g a s d s n v y d l l

401 aggacctaga ggaaggcattc caaacgtga tggggaggct ggaagatggc
      k d l e e q i q t l m g r l e d g

451 agccccccgga ctggcagat cttcaagcag acctacagca agttcgacac
      s p x t g q i f k q t y s k f d

501 aaactcacac aacgatgacg cactactaa gaactacggg ctgctctact
      t n s h n d d a l l k n y g l l y

551 gtttcaggaa ggacatggac aaggtcgaga cattcctgcg catcgtgcag
      c f r k d m d k v e t f l r i v q

601 tgccgcgttg tggagggcag ctgtggcttc caccatcacc atcaccattg
      c r s v e g s c g f h h h h h h h

651 aaacccagct ttcttgtaca aagtgg

```

Figura 7: Mapa do pENTR-INSPI01-6HIS

Molécula: pENTR-INSPI01-6HIS, 3 171 bp ADN circular
 Nome ficheiro: pENTR-INSPI01-6HIS-V1.cm5, data 21 Nov 2003

Descrição: Ligação da Blb2-orf.seq* ao pDONR221*

Descrição	Nome	Fim	Início	Tipo
seq. termin. transcrição	rrnB T2	268	295	REGIÃO
seq. termin. transcrição	rrnB T1	427	470	REGIÃO
primer forward	M13F	552	537	REGIÃO
	attL1	651	570	REGIÃO
INSPI01-6HIS	1 291	677	677	GENE
attL2	1 394	1 306	1 306	REGIÃO
primer reverse	M13 R	1 436	1 452	REGIÃO
Kan r	2 374	1 565	1 565	GENE
pUC ori	3 168	2 495	2 495	GENE

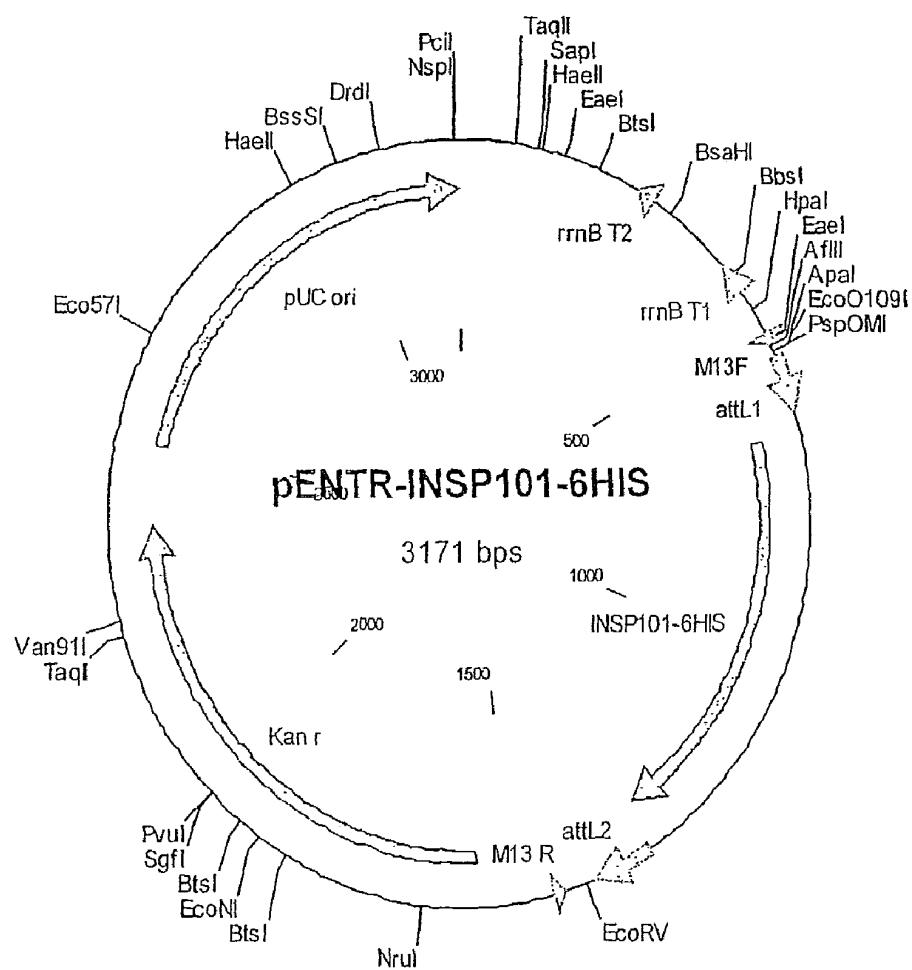


Figura 8: Mapa do pEAK12d-INSPI01-6HIS

Molécula: pEAK12d-INSPI01-6HIS-V1, 7 564 bp ADN circular
Nome ficheiro: pEAK12d-INSPI01-6HIS-V1.cm5, data 17 Jul 2003

Descrição: pEAK12 DES com dois locais de recombinação attR1 e attR2 entre os quais se insere o ADNC

Descrição	Nome	Fim	Início	Tipo
	pmb-ori	595	2	REGIÃO
	Amp	519	596	GENE
	EF-1alpha	795	1 690	REGIÃO
primer forward	peak12-F	722	2 703	REGIÃO
	attB1	874	2 855	REGIÃO
	INSP101-6HIS	502	2 888	GENE
	attB2	531	3 510	REGIÃO
	'A	966	3 538	REGIÃO
poly A/splicing	peak12-R	633	3 652	REGIÃO
primer reverse	PUR	967	4 585	GENE
PUROMICINA	tk	586	4 809	REGIÃO
promotor tk	Ori P	810	5 304	REGIÃO
	EBNA-1	304	7 356	GENE
	sv40	556	7 357	REGIÃO

