



INSTITUTO NACIONAL
DA PROPRIEDADE INDUSTRIAL

(11) Número de Publicação: **PT 1257573 E**

(51) Classificação Internacional:

C07K 14/47 (2006.01) **C12N 15/09** (2006.01)
G01N 33/68 (2006.01)

(12) **FASCÍCULO DE PATENTE DE INVENÇÃO**

(22) Data de pedido: **2001.02.20**

(30) Prioridade(s): **2000.02.23 EP 0010379**

(43) Data de publicação do pedido: **2002.11.20**

(45) Data e BPI da concessão: **2006.10.17**
001/2007

(73) Titular(es):

MERCK PATENT GMBH
FRANKFURTER STRASSE 250 D-64293
DARMSTADT **DE**

(72) Inventor(es):

FRANZ-WERNER KLUXEN **DE**

(74) Mandatário:

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

(54) Epígrafe: **NOVO FACTOR DE TRANSCRIÇÃO CARP-2**

(57) Resumo:

RESUMO

"NOVO FACTOR DE TRANSCRIÇÃO CARP-2"

São descritos polipéptidos e polinucleótidos de CARP-2 e métodos para produzir tais polipéptidos por técnicas recombinantes. São também descritos métodos para utilizar os polipéptidos e polinucleótidos de CARP-2 em ensaios de diagnóstico.

DESCRIÇÃO

"NOVO FACTOR DE TRANSCRIÇÃO CARP-2"

Área da invenção

Esta invenção relaciona-se com polipéptidos recentemente identificados e com polinucleótidos que codificam tais polipéptidos por vezes referidos daqui em diante como "novo Factor de Transcrição (CARP-2)", com a sua utilização em diagnóstico e na identificação de compostos que podem ser agonistas, antagonistas potencialmente úteis em terapêutica e com a produção de tais polipéptidos e polinucleótidos.

Antecedentes da invenção

O processo de descoberta de fármacos está presentemente a sofrer uma revolução fundamental à medida que vai endogeneizando a "genómica funcional", isto é, a biologia à base de genoma ou de genes de elevado débito. Esta abordagem como um meio para identificar genes e produtos génicos como alvos terapêuticos está a ultrapassar rapidamente as abordagens antecedentes com base na "clonagem posicional". Um fenótipo, que é uma função biológica ou doença genética, seria identificado e este seria rastreado ao gene responsável, com base na sua posição no mapa genético.

A genómica funcional assenta fortemente em tecnologias de sequenciação de ADN de elevado débito e em várias ferramentas de bioinformática para identificar sequências de genes com interesse potencial a partir dos muitos bancos

de dados de biologia molecular actualmente disponíveis. Existe uma necessidade contínua para identificar e caracterizar outros genes e os seus polipéptidos/proteínas relacionados, como alvos para a descoberta de fármacos.

Sumário da invenção

A presente invenção relaciona-se com o CARP-2, em particular com polipéptidos de CARP-2 e polinucleótidos de CARP-2, com produtos recombinantes e com métodos para a sua produção. Tais polipéptidos e polinucleótidos têm interesse em relação a métodos de tratamento de certas doenças, incluindo, mas não se limitando a, insuficiência cardíaca aguda e crónica de várias etiologias, enfarte do miocárdio, hipertrofia cardíaca, arritmia, miocardite, hipertensão pulmonar, cardiotoxicidade (por exemplo, induzida por quimioterapia), doença coronária. As insuficiências cardíaca aguda e crónica de várias etiologias, enfarte do miocárdio, hipertrofia cardíaca, arritmia, miocardite, hipertensão pulmonar, cardiotoxicidade (por exemplo, induzida por quimioterapia), doença coronária, são doravante mencionadas como "doenças da invenção". Noutro aspecto, a invenção relaciona-se com métodos para identificar agonistas e antagonistas (por exemplo, inibidores) utilizando os produtos proporcionados pela invenção e para tratar patologias associadas a um desequilíbrio do CARP-2 com os compostos identificados. Ainda noutro aspecto, a invenção relaciona-se com ensaios de diagnóstico para detectar doenças associadas a uma actividade ou níveis inapropriados de CARP-2.

Descrição da invenção

Num primeiro aspecto, a presente invenção relaciona-se com polipéptidos de CARP-2. Tais polipéptidos incluem:

Um polipéptido compreendendo a sequência da SEQ ID NO:1

95%, 96%, 97%, 98% ou 99% de identidade com a sequência polipeptídica da SEQ ID NO:2;

Um polipéptido compreendendo a sequência polipeptídica da SEQ ID NO:2.

Julga-se que os polipéptidos da presente invenção sejam membros da família de polipéptidos dos Factores de Transcrição de Repetições da Anquirina Cardíaca. Eles têm por isso interesse uma vez que o CARP-2 é um membro da família de proteínas que contêm repetições de anquirina que estão envolvidas na interacção proteína-proteína. O homólogo mais próximo, CARP, é um factor de transcrição conhecido, cuja expressão se restringe ao coração, músculo-esquelético e células endoteliais.

As propriedades biológicas do CARP-2 são doravante referidas como "actividade biológica de CARP-2" ou "actividade CARP-2". De um modo preferido, um polipéptido da presente invenção exhibe pelo menos uma actividade biológica de CARP-2.

Os polipéptidos da presente invenção também incluem variantes dos polipéptidos supramencionados, incluindo todas as formas alélicas e variantes de junção. Estes polipéptidos diferem do polipéptido de referência por inserções, deleções e substituições que podem ser conservadoras ou não conservadoras, ou qualquer combinação destas. As variantes particularmente preferidas são aquelas em que são inseridos, substituídos ou eliminados, por exemplo, desde 50 até 30, desde 30 até 20, desde 20 até 10, desde 10 até 5, desde 5 até 3, desde 3 até 2, desde 2 até 1 ou 1 aminoácidos, em qualquer combinação.

Os fragmentos preferidos de polipéptidos da presente invenção incluem um polipéptido isolado compreendendo uma sequência de aminoácidos possuindo pelo menos 30, 50 ou 100 aminoácidos contíguos da sequência de aminoácidos de SEQ ID NO: 2, ou um polipéptido isolado compreendendo uma sequência de aminoácidos possuindo pelo menos 30, 50 ou 100 aminoácidos contíguos truncados ou eliminados da sequência de aminoácidos de SEQ ID NO: 2. Os fragmentos preferidos são fragmentos biologicamente activos que medeiam a actividade biológica do CARP-2, incluindo aqueles com uma actividade semelhante ou uma actividade melhorada, ou com uma actividade diminuída indesejável. Também são preferidos aqueles fragmentos que são antigénicos ou imunogénicos num animal, especialmente num ser humano.

Os fragmentos dos polipéptidos da invenção podem ser utilizados para produzir o polipéptido completo correspondente por síntese de péptidos; deste modo, estas variantes podem ser utilizadas como intermediários para produzir os polipéptidos completos da invenção. Os

polipéptidos da presente invenção podem estar na forma da proteína "madura" ou podem ser uma parte de uma proteína maior tal como um precursor ou uma proteína de fusão. É frequentemente vantajoso incluir uma sequência de aminoácidos adicional que possua sequências secretoras ou líder, pró-sequências, sequências que ajudem na purificação, por exemplo resíduos múltiplos de histidina, ou uma sequência adicional para induzir estabilidade durante a produção recombinante.

Os polipéptidos da presente invenção podem ser preparados de qualquer modo adequado, por exemplo por isolamento a partir de fontes naturais, de células hospedeiras geneticamente modificadas compreendendo sistemas de expressão (*vide infra*) ou por síntese química, utilizando, por exemplo, sintetizadores automáticos de péptidos, ou uma combinação desses métodos. Os meios para preparar tais polipéptidos são bem compreendidos na técnica.

Os fragmentos preferidos de polinucleótidos da presente invenção incluem um polinucleótido isolado compreendendo uma sequência de nucleótidos possuindo pelo menos 15, 30, 50 ou 100 nucleótidos contíguos da sequência de SEQ ID NO: 1, ou um polinucleótido isolado compreendendo uma sequência possuindo pelo menos 30, 50 ou 100 nucleótidos contíguos truncados ou eliminados da sequência de SEQ ID NO: 1.

As variantes preferidas de polinucleótidos da presente invenção incluem variantes de junção, variantes alélicas e polimorfismos, incluindo polinucleótidos com um ou mais polimorfismos de nucleótido único (SNP).

Os polinucleótidos da presente invenção também incluem polinucleótidos que codificam variantes de polipéptidos que compreendem a sequência de aminoácidos de SEQ ID NO:2 e no qual foram substituídos, eliminados ou adicionados vários, por exemplo desde 50 até 30, desde 30 até 20, desde 20 até 10, desde 10 até 5, desde 5 até 3, desde 3 até 2, desde 2 até 1 ou 1 resíduos de aminoácidos, em qualquer combinação. Num outro aspecto, a presente invenção proporciona polinucleótidos que são transcritos de ARN das sequências de ADN da presente invenção. Em conformidade é proporcionado um polinucleótido de ARN que:

(a) compreende um transcrito de ARN da sequência de ADN que codifica o polipéptido da SEQ ID NO:2;

(b) é um transcrito de ARN da sequência de ADN que codifica o polipéptido de SEQ ID NO:2;

(c) compreende um transcrito de ARN da sequência de ADN da SEQ ID NO:1; ou

(d) é um transcrito de ARN da sequência de ADN da SEQ ID NO:1; e polinucleótidos de ARN que são complementares daquele.

A sequência de polinucleótidos da SEQ ID NO:1 apresenta homologia com o ARNm da proteína nuclear induzida por citocinas de *H. sapiens* (Chu, W. et al., J. Biol. Chem. 270 (17), 10236-10245 (1995)). A sequência de polinucleótidos da SEQ ID NO:1 é uma sequência de ADNc que codifica o polipéptido da SEQ ID NO:2. A sequência de polinucleótidos que codifica o polipéptido da SEQ ID NO:2 pode ser idêntica

à da sequência que codifica o polipéptido da SEQ ID NO:1 ou ela pode ser uma sequência diferente da SEQ ID NO:1, a qual, em consequência da redundância (degenerescência) do código genético, também codifica o polipéptido da SEQ ID NO:2. O polipéptido da SEQ ID NO:2 está relacionado com outras proteínas da família dos Factores de Transcrição das Repetições de Anquirina Cardíaca, com homologia e/ou semelhança estrutural com o ARNm da proteína nuclear induzida por citocinas de *H. sapiens* (Chu, W. *et al.*, *J. Biol. Chem.* 270 (17), 10236-10245 (1995)).

É expectável que os polipéptidos e polinucleótidos preferidos da presente invenção possuam, *inter alia*, funções/propriedades biológicas semelhantes aos polipéptidos e polinucleótidos seus homólogos. Além disso, os polipéptidos e polinucleótidos preferidos da presente invenção têm pelo menos uma actividade CARP-2.

Os polinucleótidos da presente invenção podem ser obtidos utilizando técnicas correntes de clonagem e selecção a partir de uma biblioteca de ADNc derivada do ARNm de células do coração humano, (ver, por exemplo, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Os polinucleótidos da invenção também podem ser obtidos a partir de fontes naturais tais como bibliotecas de ADN genómico ou podem ser sintetizados utilizando técnicas bem conhecidas e comercialmente acessíveis.

Quando os polinucleótidos da presente invenção são utilizados para a produção recombinante de polipéptidos da presente invenção, o polinucleótido pode incluir a

sequência de codificação para o polipéptido maduro, por si só, ou a sequência de codificação para o polipéptido maduro numa grelha de leitura com outras sequências de codificação, tais como as que codificam uma sequência líder ou secretora, uma sequência pré-, ou pró- ou prepró-proteína, ou outras partes de péptidos de fusão. Por exemplo pode estar codificada uma sequência marcadora que facilita a purificação do polipéptido fundido. Em certas formas de realização deste aspecto da invenção, a sequência marcadora é um péptido de hexa-histidina, como proporcionado no vector pQE (Qiagen, Inc.) e descrito em Gents *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, ou é uma etiqueta HA. O polinucleótido também pode conter sequências 5' e 3' não codificadoras, tais como sequências transcritas, não traduzidas, ligantes e sinais de poliadenilação, sítios de ligação de ribossomas e sequências que estabilizam o ARNm.

Pode utilizar-se polinucleótidos que são idênticos, ou têm identidade suficiente com uma sequência de polinucleótidos da SEQ ID NO:1, como sondas de hibridação para o ADNc e ADN genómico ou como iniciadores de uma reacção de amplificação de ácidos nucleicos (por exemplo, PCR). Pode utilizar-se tais sondas e iniciadores para isolar ADNc completo e clones genómicos que codificam polipéptidos da presente invenção e para isolar ADNc e clones genómicos de outros genes (incluindo genes que codificam parálogos de fontes humanas e ortólogos e parálogos de espécies que não humana) que têm uma semelhança elevada de sequência a com a SEQ ID NO:1, tipicamente, pelo menos 95% de identidade. As sondas e os iniciadores preferidos compreenderão, dum maneira geral, pelo menos 15 nucleótidos, de um modo preferido,

pelo menos 30 nucleótidos e podem ter pelo menos 50, se não pelo menos 100 nucleótidos. As sondas particularmente preferidas terão entre 30 e 50 nucleótidos. Os iniciadores particularmente preferidos terão entre 20 e 25 nucleótidos.

Um polinucleótido que codifica um polipéptido da presente invenção, incluindo homólogos de espécies que não a humana, pode ser obtido por um processo compreendendo os passos de seleccionar uma biblioteca em condições rigorosas de hibridação com uma sonda marcada possuindo a sequência da SEQ ID NO: 1 ou um fragmento desta, de um modo preferido, com pelo menos 15 nucleótidos; e isolar o ADNc completo e os clones genómicos contendo a referida sequência de polinucleótido. Tais técnicas de hibridação são bem conhecidas do especialista. As condições rigorosas de hibridação incluem incubação dum dia para o outro a 42 °C numa solução compreendendo: 50% de formamida, 5×SSC (NaCl 150 mM, citrato trissódico 15 mM), fosfato de sódio 50 mM (pH7,6), 5× solução de Denhardt, sulfato de dextrano a 10% e 20 micrograma/ml de ADN de esperma de salmão fraccionado, desnaturado; seguida de lavagem dos filtros em 0,1× SSC a cerca de 65 °C. Assim, a presente invenção também inclui polinucleótidos isolados, de um modo preferido com uma sequência de nucleótidos com pelo menos 100, obtida seleccionando uma biblioteca em condições rigorosas de hibridação com uma sonda marcada possuindo a sequência da SEQ ID NO:1 ou um fragmento desta, de um modo preferido com pelo menos 15 nucleótidos.

O especialista na técnica compreenderá que, em muitos casos, uma sequência de ADNc isolada será incompleta, uma vez que a região que codifica o polipéptido não se prolonga

em toda a extensão até à extremidade 5'. Isto é uma consequência da transcriptase inversa, uma enzima com uma "processabilidade" inerentemente baixa (uma medida da capacidade da enzima para permanecer ligada à cadeia molde durante a reacção de polimerização), que não conclui uma cópia do ADN a partir da cadeia de ARNm durante a síntese da primeira cadeia de ADNc.

Existem vários métodos disponíveis e bem conhecidos dos especialistas na técnica para se obter ADNc completos, ou para se prolongar ADNc curtos, por exemplo os baseados no método de Amplificação Rápida de extremidades de ADNc (RACE) (ver, por exemplo, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Modificações recentes à técnica, ilustradas por exemplo pela tecnologia Marathon (marca comercial) (Clontech Laboratories Inc.), simplificaram significativamente a pesquisa de ADNc mais compridos. Na tecnologia Marathon (marca comercial), os ADNc foram preparados a partir de ARNm extraído de um tecido seleccionado e ligada uma sequência 'adaptadora' a cada uma das extremidades. Realiza-se em seguida a amplificação de ácidos nucleicos (PCR) para amplificar a extremidade 5' "em falta" do ADNc utilizando uma combinação de iniciadores oligonucleótidos específicos para o gene e específicos para o adaptador. Repete-se em seguida a reacção de PCR utilizando iniciadores 'internos', isto é, iniciadores concebidos para emparelhar no produto amplificado (tipicamente um iniciador específico para o adaptador que emparelha depois de 3' na sequência adaptadora e um iniciador específico para o gene que emparelha depois de 5' na sequência genica conhecida). Os produtos desta reacção podem ser depois analisados por sequenciação de ADN e um

ADNc completo construído quer juntando o produto directamente com o ADNc existente para dar uma sequência completa, ou realizando um PCR completo separado utilizando a informação da nova sequência para a concepção do iniciador 5'.

Os polipéptidos recombinantes da presente invenção podem ser preparados por processos bem conhecidos na técnica a partir de células hospedeiras geneticamente modificadas compreendendo sistemas de expressão. Assim sendo, noutro aspecto, a presente invenção relaciona-se com sistemas de expressão compreendendo um polinucleótido ou polinucleótidos da presente invenção, com células hospedeiras que estão geneticamente modificadas com tais sistemas de expressão e com a produção de polipéptidos da invenção por técnicas recombinantes. Também se pode utilizar sistemas de tradução sem células para produzir tais proteínas utilizando ARN derivados das construções de ADN da presente invenção.

Para produção recombinante, as células hospedeiras podem ser geneticamente manipuladas de modo a incorporar sistemas de expressão, ou partes destes, para polinucleótidos da presente invenção. Os polinucleótidos podem ser introduzidos nas células hospedeiras por métodos descritos em muitos manuais correntes de laboratório, tais como Davis *et al.*, *Basic Methods in Molecular Biology* (1986) e Sambrook *et al.* (*ibid*). Os métodos preferidos para introduzir polinucleótidos em células hospedeiras incluem, por exemplo, transfecção com fosfato de cálcio, transfecção mediada por DEAE-dextrano, transvecção, microinjecção, transfecção catiónica mediada por lípidos, electroporação,

transdução, inserção por raspagem, introdução balística ou infecção.

Os exemplos representativos de hospedeiros apropriados incluem células bacterianas, tais como células de estreptococos, estafilococos, *E coli*, *Streptomyces* e *Bacillus subtilis*; células de fungos, tais como células de levedura e células de *Aspergillus*; células de insecto tais como células de *Drosophila* S2 e *Spodoptera* Sf9; células de animais tais como células CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 e células de melanoma de Bowes; e células vegetais.

Pode utilizar-se uma grande diversidade de sistemas de expressão, por exemplo, sistemas cromossómicos, epissomais e derivados de vírus, por exemplo, vectores derivados de plasmídeos bacterianos, de bacteriófagos, de translocadores, de epissomas de levedura, de elementos de inserção, de elementos cromossómicos de levedura, de vírus tais como baculovírus, papova vírus, tais como SV40, vírus vaccinia, adenovírus, vírus da varíola, vírus da pseudo-raiva e retrovírus, e vectores derivados de combinações destes, tais como os derivados de plasmídeos e elementos genéticos de bacteriófagos, tais como cosmídeos e fagemídeos. Os sistemas de expressão podem conter regiões de controlo que regulam e originam a expressão. Duma maneira geral pode utilizar-se qualquer sistema ou vector que seja capaz de manter, propagar ou expressar um polinucleótido para produzir um polipéptido num hospedeiro. A sequência de polinucleótido apropriada pode ser inserida num sistema de expressão por qualquer um de uma variedade de técnicas de rotina bem conhecidas, tais como, por exemplo, as descritas em Sambrook et al., (*ibid*). Pode

incorporar-se sinais de secreção no polipéptido desejado para permitir a secreção da proteína traduzida para o lúmen do retículo endoplasmático, para o espaço periplasmático ou para o ambiente extracelular. Estes sinais podem ser endógenos ao polipéptido ou eles podem ser sinais heterólogos.

Se um polipéptido da presente invenção for expressado para ser utilizado em ensaios de selecção é geralmente preferido que o polipéptido seja produzido à superfície da célula. Neste caso, as células podem ser colhidas antes de serem utilizadas no ensaio de selecção. Se o polipéptido é segregado para o meio, o meio pode ser recuperado para se recuperar e purificar o polipéptido. Se for produzido intracelularmente, as células têm de ser primeiro submetidas a lise antes de se recuperar o polipéptido.

Os polipéptidos da presente invenção podem ser recuperados e purificados a partir de células de culturas recombinantes por métodos bem conhecidos incluindo precipitação com sulfato de amónio ou etanol, extracção ácida, cromatografia de permuta aniónica e catiónica, cromatografia em fosfocelulose, cromatografia de interacção hidrófoba, cromatografia de afinidade, cromatografia sobre hidroxiapatite e cromatografia sobre lectina. De um modo muito preferido utiliza-se cromatografia de alta eficiência para purificação. Pode utilizar-se técnicas bem conhecidas para reenrolamento de proteínas para regenerar a conformação activa quando o polipéptido é desnaturado durante a síntese intracelular, isolamento e/ou purificação.

Os polinucleótidos da presente invenção podem ser utilizados como reagentes de diagnóstico, através da detecção de mutações no gene associado. A detecção de uma forma mutada do gene caracterizada pelo polinucleótido da SEQ ID NO:1 no ADNc ou na sequência genómica e a qual está associada com uma disfunção proporcionará uma ferramenta de diagnóstico que pode ajudar, ou definir, um diagnóstico de uma doença, ou uma susceptibilidade a uma doença, a qual resulta da subexpressão, sobre-expressão ou da expressão alterada em termos espaciais ou temporais do gene. Os indivíduos portadores de mutações no gene podem ser detectados ao nível do ADN por uma diversidade de técnicas bem conhecidas na arte.

Os ácidos nucleicos para diagnóstico podem ser obtidos a partir de células individuais, tal como a partir de sangue, urina, saliva, biópsia de tecido ou material de autópsia. Pode utilizar-se directamente ADN genómico para detecção ou ele pode ser amplificado enzimaticamente utilizando PCR, de um modo preferido RT-PCR, ou outras técnicas de amplificação antes da análise. Também se pode utilizar de um modo semelhante ARN ou ADNc. As deleções e inserções podem ser detectadas por uma alteração no tamanho do produto amplificado por comparação com o genótipo normal. Pode identificar-se mutações pontuais hibridizando ADN amplificado com sequências de nucleótidos de CARP-2 marcadas. As sequências perfeitamente coincidentes podem ser distinguidas das hélices duplas não coincidentes por digestão com RNase ou por diferenças nas temperaturas de fusão. As diferenças na sequência de ADN também podem ser detectadas através de alterações na mobilidade electroforética de fragmentos de ADN em geles, com ou sem

agentes desnaturantes, ou através de sequenciação directa de ADN (ver, por exemplo, Myers et al., Science (1985) 230:1242). As alterações na sequência em locais específicos também podem ser reveladas por ensaios de protecção com nucleases, tais como protecção com RNase e S1 ou pelo método de clivagem química (ver Cotton et al., Proc Natl Acad Sci USA (1985) 85: 4397-4401).

Pode construir-se uma matriz de sondas de oligonucleótidos compreendendo a sequência polinucleótida de CARP-2 ou fragmentos deste para efectuar uma selecção eficiente, por exemplo, de mutações genéticas. Estas matrizes são, de um modo preferido, matrizes ou grelhas de alta densidade. Os métodos de tecnologia de matriz são bem conhecidos e têm uma aplicabilidade geral e podem ser utilizados para abordar uma diversidade de questões em genética molecular incluindo expressão de genes, ligação genética e variabilidade genética, ver, por exemplo, M.Chee et al., Science, 274, 610-613 (1996) e outras referências ali citadas.

A detecção de níveis anormalmente reduzidos ou aumentados de polipéptido ou expressão de ARNm também pode ser utilizada para diagnosticar ou determinar a susceptibilidade de um indivíduo para uma doença da invenção. A expressão reduzida ou aumentada pode ser medida ao nível do ARN utilizando qualquer um dos métodos bem conhecidos na técnica para a quantificação de polinucleótidos, tal como, por exemplo, amplificação de ácidos nucleicos, por exemplo PCR, RT-PCR, protecção com RNase, transferência de Northern e outros métodos de hibridação. As técnicas de doseamento que podem ser

utilizadas para determinar os níveis de uma proteína, tal como um polipéptido da presente invenção, numa amostra proveniente de um hospedeiro são bem conhecidas dos especialistas na técnica. Estes métodos de doseamento incluem radioimunoensaios, ensaios de ligação competitiva, análise por transferência de Western e ensaios de ELISA.

Assim noutro aspecto, a presente invenção relaciona-se com um estojo de diagnóstico compreendendo:

(a) um polinucleótido da presente invenção, de um modo preferido a sequência de nucleótidos de SEQ ID NO: 1, ou um fragmento ou um ARN transcrito daquela,

(b) uma sequência de nucleótidos complementar à de (a);

(c) um polipéptido da presente invenção, de um modo preferido o polipéptido da SEQ ID NO:2 ou um fragmento deste; ou

(d) um anticorpo para um polipéptido da presente invenção, de um modo preferido do polipéptido da SEQ ID NO:2.

Entender-se-á que em qualquer um desses estojos, (a), (b), (c) ou (d) pode compreender um componente substancial. Um tal estojo terá utilidade no diagnóstico de uma doença ou da susceptibilidade para uma doença, em particular doenças da invenção, entre outras.

As sequências de polinucleótidos da presente invenção são úteis para estudos de localização de cromossomas. A sequência é especificamente dirigida, e pode hibridizar, com uma localização particular de um cromossoma humano

individual. O mapeamento de sequências relevantes para os cromossomas de acordo com a presente invenção é um primeiro passo importante para correlacionar as sequências com as doenças associadas aos genes. Uma vez localizada uma sequência numa localização cromossómica precisa, a posição física da sequência no cromossoma pode ser correlacionada com dados do mapa genético. Estes dados encontram-se, por exemplo, em V. McKusick, Mendelian Inheritance in Man (acessível em linha através da Johns Hopkins University Welch Medical Library). As relações entre genes e doenças que foram mapeadas na mesma região cromossómica são depois identificadas através de análise de ligação (co-herança de genes fisicamente adjacentes). As localizações cromossómicas humanas exactas para uma sequência genómica (fragmento de gene etc.) podem ser determinadas utilizando Mapeamento de Híbridos de Radiação (RH) (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., e Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, Nature Genetics 7, 22-28). Encontra-se disponível um número de painéis RH a partir de Research Genetics (Huntsville, Ala., EUA) por exemplo o painel GeneBridge4 RH (Hum Mol Genet March 1996; 5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme J F, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow P N). Para determinar a localização cromossómica de um gene utilizando este painel realiza-se 93 PCRs utilizando iniciadores concebidos a partir do gene de interesse do ADN de RH. Cada um destes ADN contém fragmentos genómicos humanos aleatórios mantidos num fundo de hamster (linhas celulares híbridas humanas/hamster). Estes PCR originam 93 resultados que

indicam a presença ou ausência do produto de PCR do gene de interesse. Estes resultados são comparados com os resultados gerados utilizando produtos de PCR a partir de sequências genómicas de localização conhecida. Esta comparação é realizada em <http://www.genome.wi.mit.edu/>.

As sequências de polinucleótidos da presente invenção são também ferramentas úteis para estudos de expressão em tecidos. Estes estudos permitem a determinação de padrões de expressão de polinucleótidos da presente invenção os quais podem dar um indício para os padrões de expressão dos polipéptidos codificados em tecidos, através da detecção dos ARN que os codificam. As técnicas utilizadas são bem conhecidas na técnica e incluem técnicas de hibridação *in situ* com clones dispostos numa grelha, tal como a hibridação de ADNc em micromatriz (Schena *et al*, Science, 270, 467-470, 1995 e Shalon *et al*, Genome Res, 6, 639-645, 1996) e técnicas de amplificação de nucleótidos tal como PCR. Um método preferido utiliza a tecnologia TAQMAN (Marca comercial) disponível de Perkin Elmer. Os resultados deste estudo podem fornecer uma indicação da função normal do polipéptido no organismo. Além disso, os estudos comparativos do padrão de expressão normal de ARNm com os de ARNm codificados por uma forma alternativa do mesmo gene (por exemplo, um possuindo uma alteração no polipéptido que codifica uma mutação potencial ou reguladora) podem fornecer conhecimentos valiosos sobre o papel dos polipéptidos da presente invenção, ou o da expressão inadequada daqueles na doença. Esta expressão inadequada pode ser de natureza temporal, espacial ou simplesmente quantitativa.

Os polipéptidos da presente invenção são expressados no coração e músculo-esquelético.

Um outro aspecto da presente invenção relaciona-se com anticorpos. Os polipéptidos da invenção ou os seus fragmentos, ou as células que os expressam, podem ser utilizados como imunogénios para produzir anticorpos que são imuno-específicos para polipéptidos da presente invenção. O termo "imuno-específico" significa que os anticorpos têm uma afinidade consideravelmente maior para os polipéptidos da invenção do que a sua afinidade para outros polipéptidos afins do estado anterior da técnica.

Os anticorpos gerados contra polipéptidos da presente invenção podem ser obtidos administrando os polipéptidos ou os fragmentos portadores do determinante antigénico, ou a células a um animal, de um modo preferido um animal não humano, utilizando protocolos de rotina. Para a preparação de anticorpos monoclonais pode utilizar-se qualquer técnica que proporcione anticorpos produzidos por culturas contínuas de linhas celulares. Os exemplos incluem a técnica de hibridoma (Kohler, G. e Milstein, C., *Nature* (1975) 256:495-497), a técnica trioma, a técnica de hibridoma de células B humanas (Kozbor *et al.*, *Immunology Today* (1983) 4:72) e a técnica de hibridoma de EBV (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

As técnicas para a produção de anticorpos de cadeia única, tais como os descritos na Patente U.S. nº 4 946 778, também podem ser adaptadas para produzir anticorpos de cadeia única para polipéptidos desta invenção. Também se pode

utilizar ratinhos transgênicos, ou outros organismos, incluindo outros mamíferos, para expressar anticorpos humanizados.

Pode utilizar-se os anticorpos acima descritos para isolar ou identificar clones que expressam o polipéptido ou para purificar os polipéptidos por cromatografia de afinidade. Os anticorpos contra polipéptidos da presente invenção também podem ser utilizados para tratar doenças da invenção, entre outras.

Os polipéptidos e polinucleótidos da presente invenção também podem ser utilizados como vacinas. Assim sendo, num outro aspecto, a presente invenção relaciona-se com um método para induzir uma resposta imunológica num mamífero o qual compreende inocular um mamífero com um polipéptido da presente invenção, adequado para produzir um anticorpo e/ou resposta imunológica de células T, incluindo, por exemplo, células T produtoras de citocinas ou células T citotóxicas, para proteger o referido animal da doença, independentemente do facto dessa doença se encontrar já estabelecida no indivíduo ou não. Também se pode induzir uma resposta imunológica num mamífero por um método que compreende administrar um polipéptido da presente invenção via um vector que orienta a expressão do polinucleótido e que codifica o polipéptido *in vivo* para induzir uma tal resposta imunológica que produza anticorpos para proteger o referido animal das doenças da invenção. Uma forma de administrar o vector consiste em acelerá-lo para as células desejadas como um revestimento sobre partículas ou de outro modo. Este vector de ácido nucleico pode compreender ADN, ARN, um ácido nucleico modificado ou um híbrido ADN/ARN.

Para utilizar uma vacina, um vector de polipéptido ou um ácido nucleico será normalmente proporcionado como uma formulação (composição) para vacina. A formulação pode compreender ainda um veículo adequado. Uma vez que um polipéptido pode ser quebrado no estômago, ele é de um modo preferido administrado por via parentérica (por exemplo, injeção subcutânea, intramuscular, intravenosa ou intradérmica). As formulações adequadas para administração parentérica incluem soluções aquosas e não aquosas estéreis para injeção que possam conter antioxidantes, tampões, bacteriostáticos e solutos que tornam a formulação isotónica com o sangue do receptor; e suspensões aquosas e não aquosas estéreis que podem incluir agentes de suspensão ou agentes espessantes. As formulações podem ser apresentadas em recipientes de dose única ou doses múltiplas, por exemplo, ampolas e frascinhos selados e podem ser conservadas num estado liofilizado o qual requer apenas a adição do veículo líquido estéril imediatamente antes da utilização. A formulação de vacina também pode incluir sistemas adjuvantes para intensificar a imunogenicidade da formulação, tais como sistemas de óleo em água e outros sistemas conhecidos na técnica. A dosagem dependerá da actividade específica da vacina e pode ser facilmente determinada por experimentação em rotina.

Os polipéptidos da presente invenção têm uma ou mais funções biológicas que são relevantes em um ou mais estados patológicos, em particular nas doenças da invenção mencionadas acima. Por conseguinte é útil identificar compostos que estimulem ou inibam a função ou nível do polipéptido. Assim sendo, num outro aspecto, a presente invenção proporciona um método de selecção de compostos

para identificar aqueles que estimulam ou inibem a função ou nível do polipéptido. Estes métodos identificam agonistas ou antagonistas que podem ser utilizados para efeitos terapêuticos e profilácticos para tais doenças da invenção como mencionadas acima. Os compostos podem ser identificados a partir de uma diversidade de fontes, por exemplo, células, preparados sem células, bibliotecas químicas, colecções de compostos químicos e misturas de produtos naturais. Estes agonistas ou antagonistas assim identificados podem ser substratos, ligandos, receptores, enzimas, etc. naturais ou modificados consoante o caso, do polipéptido; um seu mimético estrutural ou funcional (ver Coligan *et al.*, *Current Protocols in Immunology* 1(2):Capítulo 5 (1991)) ou uma molécula pequena.

O método de selecção pode medir simplesmente a ligação de um composto candidato ao polipéptido, ou às células ou membranas portadoras do polipéptido, ou uma proteína de fusão deste, por meio de uma etiqueta associada directa ou indirectamente ao composto candidato. Alternativamente, o método de selecção pode envolver a medição ou detecção (qualitativa ou quantitativa) da ligação competitiva de um composto candidato ao polipéptido relativamente a um competidor marcado (por exemplo agonista ou antagonista). Além disso, estes métodos de selecção podem avaliar se o composto candidato origina um sinal gerado pela activação ou inibição do polipéptido, utilizando sistemas de detecção apropriados para as células portadoras do polipéptido. Os inibidores de activação são geralmente avaliados na presença de um agonista conhecido e observa-se o efeito do agonista na activação na presença do composto candidato. Além disso, os métodos de selecção podem compreender

simplesmente os passos de misturar um composto candidato com uma solução contendo um polipéptido da presente invenção, para formar uma mistura, medir a actividade CARP-2 na mistura e comparar a actividade CARP-2 da mistura com uma mistura de controlo sem composto candidato.

Os polipéptidos da presente invenção podem ser utilizados em métodos de selecção de baixa capacidade convencionais e também em formatos de selecção de alto débito (HTS). Estes formatos HTS incluem não só a utilização bem estabelecida de placas microtítulo de 96, e mais recentemente 384, poços mas também métodos emergentes tal como o método de nanopocões descrito por Schullek *et al*, *Anal Biochem.*, 246, 20-29, (1997).

Também se pode utilizar proteínas de fusão, tais como as feitas a partir da fracção Fc e do polipéptido CARP-2, como se descreveu acima, para ensaios de selecção de alto débito para identificar antagonistas para o polipéptido da presente invenção (ver D. Bennett *et al.*, *J Mol Recognition*, 8:52-58 (1995); e K. Johanson *et al.*, *J Biol Chem*, 270(16):9459-9471 (1995)).

Técnicas de selecção

Os polinucleótidos, polipéptidos e anticorpos para o polipéptido da presente invenção também podem ser utilizados para configurar métodos de selecção para detectar o efeito de compostos adicionados na produção de ARNm e polipéptidos nas células. Por exemplo, pode conceber-se um ensaio ELISA para medir níveis segregados ou associados às células de polipéptidos utilizando anticorpos

monoclonais e policlonais por métodos correntes conhecidos na técnica. Este pode ser utilizado para descobrir agentes que possam inibir ou intensificar a produção de polipéptidos (também designados por antagonistas ou agonistas, respectivamente) a partir de células ou tecidos adequadamente manipulados.

Pode utilizar-se um polipéptido da presente invenção para identificar receptores ligados à membrana ou solúveis, através de técnicas correntes de ligação ao receptor conhecidas na técnica. Estes incluem, entre outros, ensaios de ligação ao ligando e ensaios de reticulação nos quais o polipéptido é marcado com um isótopo radioactivo (por exemplo, ^{125}I), quimicamente modificado (por exemplo, biotinilado) ou fundido com uma sequência peptídica adequada para detecção ou purificação, e incubado com uma fonte do receptor putativo (células, membranas celulares, sobrenadantes celulares, extractos de tecidos, fluidos orgânicos). Outros métodos incluem técnicas biofísicas tais como a ressonância plásmica de superfície e espectroscopia. Estes métodos de selecção também podem ser utilizados para identificar agonistas e antagonistas do polipéptido que competem com a ligação do polipéptido para os seus receptores. Os métodos correntes para realizar tais ensaios são bem compreendidos na técnica.

Os exemplos de antagonistas de polipéptidos da presente invenção incluem anticorpos ou, nalguns casos, oligonucleótidos ou proteínas que estão intimamente relacionados com os ligandos, substratos, receptores, enzimas, etc., consoante o caso, do polipéptido, por exemplo, um fragmento dos ligandos, substratos, receptores,

enzimas, etc.; ou uma molécula pequena que se liga ao polipéptido da presente invenção mas que não viabiliza uma resposta, pelo que a actividade do polipéptido é impedida.

Os métodos de selecção também podem envolver a utilização de tecnologia transgénica e o gene do CARP-2. A técnica para preparar animais transgénicos está bem estabelecida. Por exemplo, o gene do CARP-2 pode ser introduzido através de microinjecção no pronúcleo macho de ovócitos fertilizados, transferência retroviral para embriões pré- ou pós-implante ou injecção de células pluripotentes embrionárias geneticamente modificadas, tal como por electroporação, em blastocistos hospedeiros. Os animais transgénicos particularmente úteis são os chamados animais "knock-in" nos quais um gene do animal está substituído pelo equivalente humano no do genoma do animal. Os animais transgénicos "knock-in" são úteis no processo de descoberta de fármacos, para validação do alvo, nos casos em que o composto é específico para o alvo humano. Outros animais transgénicos úteis são os chamados animais "knock-out" nos quais a expressão do ortólogo animal de um polipéptido da presente invenção e codificado por uma sequência de ADN endógena numa célula está parcial ou completamente anulada. A anulação do gene pode ter como objectivo células ou tecidos específicos, pode ocorrer apenas em determinadas células ou tecidos como uma consequência das limitações da tecnologia, ou pode ocorrer em todas, ou praticamente todas, as células no animal. A tecnologia transgénica animal também oferece um sistema de expressão-clonagem para todo o animal no qual os genes introduzidos são expressados para dar grandes quantidades de polipéptidos da presente invenção

Os estojos de selecção para serem utilizados nos métodos descritos acima constituem um outro aspecto da presente invenção. Estes estojos de selecção compreendem:

- (a) um polipéptido da presente invenção;
- (b) uma célula recombinante que expressa um polipéptido da presente invenção;
- (c) uma membrana celular que expressa um polipéptido da presente invenção; ou
- (d) um anticorpo para um polipéptido da presente invenção;

Polipéptido esse que é de um modo preferido o da SEQ ID NO:2.

Entender-se-á que em qualquer estajo deste tipo, (a), (b), (c) ou (d) pode compreender um componente substancial.

Glossário

As definições seguintes são proporcionadas para facilitar a compreensão de determinados termos utilizados frequentemente acima.

“Anticorpos” como aqui utilizado inclui anticorpos policlonais e monoclonais, quiméricos, de cadeia simples e anticorpos humanizados, assim como fragmentos Fab,

incluindo os produtos de um Fab ou outra biblioteca de expressão de imunoglobulina.

“Isolado” significa alterado “pela mão do homem” a partir do seu estado natural, isto é, se ele ocorrer na natureza, ele foi alterado ou retirado do seu ambiente original, ou ambos. Por exemplo, um polinucleótido ou um polipéptido naturalmente presente num organismo vivo não está “isolado,” mas o mesmo polinucleótido ou polipéptido separado dos materiais coexistentes no seu estado natural está “isolado”, nas condições em que o termo é aqui utilizado. Além disso, um polinucleótido ou polipéptido que é introduzido num organismo por transformação, manipulação genética ou por qualquer outro método recombinante está “isolado” mesmo se ele ainda estiver presente no referido organismo, organismo esse que pode ser vivo ou não vivo.

“Polinucleótido” refere-se, duma maneira geral, a qualquer polirribonucleótido (ARN) ou polidesoxirribonucleótido (ADN), o qual pode ser ARN ou ADN não modificado ou modificado. Os “polinucleótidos” incluem, sem restrições, ADN de cadeia simples e dupla, ADN que é uma mistura de regiões de cadeia simples e cadeia dupla, ARN de cadeia simples e dupla e ARN que é uma mistura de regiões de cadeia simples e cadeia dupla, moléculas híbridas compreendendo ADN e ARN que podem ser de cadeia simples ou, mais tipicamente, de cadeia dupla ou uma mistura de regiões de cadeia simples e dupla. Além disso, “polinucleótido” refere-se a regiões de cadeia tripla compreendendo ARN ou ADN ou ambos, ARN e ADN. O termo “polinucleótido” também inclui ADN ou ARN contendo uma ou mais bases modificadas e ADN ou ARN com cadeias principais modificadas por razões de

estabilidade ou outras. Bases “modificadas” incluem, por exemplo, bases tritiladas e bases invulgares tais como inosina. Pode fazer-se uma diversidade de modificações ao ADN ou ARN; assim, “polinucleótido” abrange formas química, enzimática ou metabolicamente modificadas de polinucleótidos como encontrados tipicamente na natureza, assim como as formas químicas de ADN e ARN características de vírus e células. “Polinucleótido” também abrange polinucleótidos relativamente curtos, frequentemente mencionados como oligonucleótidos.

“Polipéptido” refere-se a qualquer polipéptido compreendendo dois ou mais aminoácidos ligados um ao outro por ligações peptídicas ou ligações peptídicas modificadas, isto é, isósteres de péptidos. “Polipéptido” refere-se tanto a cadeias curtas, geralmente referidas como péptidos, oligopéptidos ou oligómeros, como a cadeias mais compridas, geralmente referidas como proteínas. Os polipéptidos podem conter aminoácidos que não os 20 aminoácidos codificados por genes. Os “polipéptidos” incluem sequências de aminoácidos modificadas quer por processos naturais, tal como processamento pós-tradução, quer por técnicas de modificação química que são bem conhecidas na técnica. Tais modificações estão bem descritas em textos básicos e em monografias mais pormenorizadas, assim como numa pesquisa bibliográfica extensa. As modificações podem ocorrer em qualquer parte do polipéptido, incluindo na cadeia principal do péptido, nas cadeias laterais dos aminoácidos e nas extremidades amino ou carboxilo.

Entender-se-á que o mesmo tipo de modificação pode estar presente no mesmo ou em vários graus em vários sítios num

dados polipéptido. De igual modo, um dado polipéptido pode conter muitos tipos de modificação. Os polipéptidos podem ser ramificados em resultado de ubiquitinação, e eles podem ser cíclicos, com ou sem ramificação. Os polipéptidos cíclicos, ramificados e cíclicos ramificados podem resultar de processos de pós-tradução naturais ou podem ser preparados por métodos sintéticos. As modificações incluem acetilação, acilação, ADP-ribosilação, amidação, biotinilação, ligação covalente de flavina, ligação covalente de uma unidade heme, ligação covalente de um nucleótido ou derivado de nucleótido, ligação covalente de um lípido ou derivado de lípido, ligação covalente de fosfatidilinositol, reticulação, ciclização, formação de ligação dissulfureto, desmetilação, formação de reticulações covalentes, formação de cistina, formação de piroglutamato, formilação, gama-carboxilação, glicosilação, formação de âncoras de GPI, hidroxilação, iodação, metilação, miristoilação, oxidação, processamento proteolítico, fosforilação, fenilação, racemização, selenoilação, sulfatação, adição de aminoácidos a proteínas mediada por ARN de transferência e ubiquitinação (ver, por exemplo, *Proteins -Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., *Post-translational Protein Modifications: Perspectives and Prospects*, 1-12, in *Post-translational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter et al., "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol*, 182, 626-646, 1990, e Rattan et al., "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci*, 663, 48-62, 1992).

“Fragmento” de uma sequência de polipéptido refere-se a uma sequência de polipéptido que é mais curta do que a sequência de referência mas que retém essencialmente a mesma função ou actividade biológica que o polipéptido de referência. “Fragmento” de uma sequência de polinucleótido refere-se a uma sequência de polinucleótido que é mais curta do que a sequência de referência da SEQ ID NO:1.

“Variante” refere-se a um polinucleótido ou polipéptido que difere de um polinucleótido ou polipéptido de referência, mas retém as propriedades essenciais deste. Uma variante típica de um polinucleótido difere na sequência de nucleótidos do polinucleótido de referência. As alterações na sequência de nucleótidos da variante podem ou podem não alterar a sequência de aminoácidos de um polipéptido codificado pelo polinucleótido de referência.

As alterações de nucleótidos podem resultar em substituições, adições, deleções, fusões e truncagem de aminoácidos no polipéptido codificado pela sequência de referência, como se discute a seguir. Uma variante típica de um polipéptido difere na sequência de aminoácidos do polipéptido de referência. Duma maneira geral, as alterações estão limitadas de modo a que as sequências do polipéptido de referência e da variante sejam muito semelhantes na globalidade e, em muitas regiões, idênticas. Uma variante e o polipéptido de referência podem diferir na sequência de aminoácidos em uma ou mais substituições, inserções, deleções em qualquer combinação. Um resíduo de aminoácido substituído ou inserido pode ser (ou não) um codificado pelo código genético. Substituições conservadoras típicas incluem Gli, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Tre; Lis, Arg; e Fen e Tir. Uma

variante de um polinucleótido ou polipéptido pode ocorrer naturalmente tal como um alelo, ou ela pode ser uma variante que não seja conhecida como ocorrendo naturalmente. As variantes de polinucleótidos e polipéptidos que não ocorrem naturalmente podem ser preparadas por técnicas de mutagénese ou por síntese directa. Também estão incluídos como variantes os polipéptidos possuindo uma ou mais modificações pós-tradução, por exemplo glicosilação, fosforilação, metilação, ADP ribosilação e semelhantes. As formas de realização incluem metilação do aminoácido N-terminal, fosforilações de serinas e treoninas e modificação de glicinas C-terminais.

“Alelo” refere-se a uma de duas ou mais formas alternativas de um gene que ocorre num dado local no genoma.

“Polimorfismo” refere-se a uma variação na sequência de nucleótidos (e na sequência do polipéptido codificado, se relevante) numa dada posição no genoma dentro de uma população.

“Polimorfismo de Nucleótido Único” (SNP) refere-se à ocorrência de variabilidade de nucleótidos numa única posição nucleotídica do genoma, dentro de uma população. Um SNP pode ocorrer dentro de um gene ou em regiões intergénicas do genoma. Os SNP podem ser detectados utilizando Amplificação Específica de Alelo (ASA). No processo são necessários pelo menos 3 iniciadores. Um iniciador comum é utilizado em complemento inverso do polimorfismo a ser detectado. Este iniciador comum pode ter entre 50 e 1500 pb da base polimórfica. Os outros dois (ou

mais) iniciadores são idênticos um ao outro à excepção da base 3' final variar para condizer com um dos dois (ou mais) alelos que fazem parte do polimorfismo. Em seguida são realizadas duas (ou mais) reacções de PCR no ADN amostra, utilizando cada uma o iniciador comum e um dos Iniciadores Específicos para o Alelo.

A "Variante de Junção" como aqui utilizada refere-se a moléculas de ADNc produzidas a partir de moléculas de ARN inicialmente transcritas a partir da mesma sequência genómica de ADN mas a qual sofreu junção alternativa de ARN. A junção alternativa de ARN ocorre quando um transcrito primário de ARN sofre junção, geralmente para a remoção de intrões, a qual resulta na produção de mais do que uma molécula de ARNm cada uma das quais pode codificar sequências de aminoácidos diferentes. O termo variante de junção também se refere às proteínas codificadas pelas moléculas de ADNc acima.

"Identidade" reflecte uma relação entre duas ou mais sequências de polipéptidos ou duas ou mais sequências de polinucleótidos, determinada comparando as sequências. Duma maneira geral, identidade refere-se a uma correspondência exacta nucleótido a nucleótido ou aminoácido a aminoácido das duas sequências de polinucleótidos ou polipéptidos, respectivamente, ao longo do comprimento das sequências a serem comparadas.

"% de identidade" - Para as sequências em que não existe uma correspondência exacta pode determinar-se uma "% de identidade". Duma maneira geral, as duas sequências a serem comparadas são alinhadas para dar uma correlação máxima

entre as sequências. Isto pode incluir a inserção de “vazios” quer em uma quer em ambas as sequências, para aumentar o grau de alinhamento. Pode determinar-se uma % de identidade ao longo de todo o comprimento de cada uma das sequências a serem comparadas (o chamado alinhamento global), o que é particularmente adequado para sequências com o mesmo comprimento ou com comprimentos muito semelhantes, ou ao longo de comprimentos definidos mais curtos (os chamados alinhamentos locais), o que é mais adequado para sequências de comprimento desigual.

“Similaridade” é uma outra medida mais sofisticada da relação entre duas sequências de polipéptidos. Duma maneira geral, “similaridade” significa uma comparação entre os aminoácidos de duas cadeias de polipéptido, numa base resíduo a resíduo, tendo em consideração não apenas a correspondência exacta entre pares de resíduos, um de cada uma das sequências a serem comparadas (como para a identidade) mas também, quando não existe uma correspondência exacta, se, numa base evolutiva, um resíduo é um substituto provável de outro. Esta possibilidade tem uma “classificação” associada a partir da qual se pode determinar a “% de semelhança” das duas sequências.

Os métodos para comparar a identidade e similaridade de duas ou mais sequências são bem conhecidos na técnica. Assim, por exemplo, pode utilizar-se os programas disponíveis no Wisconsin Sequence Analysis Package, versão 9.1 (Devereux J et al, *Nucleic Acids Res*, 12, 387-395, 1984, acessível de Genetics Computer Group, Madison, Wis., EUA), por exemplo os programas BESTFIT e GAP, para determinar a % de identidade entre dois polinucleótidos e a

% de identidade e a % de similaridade entre duas sequências de polipéptidos. O BESTFIT utiliza o algoritmo de "homologia local" de Smith e Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) e determina a região única com melhor similaridade entre duas sequências. O BESTFIT é mais adequado para comparar duas sequências de polinucleótidos ou polipéptidos que têm um comprimento diferente, assumindo o programa que a sequência mais curta representa uma fracção da mais comprida. Em comparação, a GAP alinha duas sequências, encontrando uma "similaridade máxima", segundo o algoritmo de Neddleman e Wunsch (J Mol Biol, 48, 443-453, 1970). O GAP é mais adequado para comparar sequências que têm aproximadamente o mesmo comprimento e em que é expectável um alinhamento ao longo de todo o comprimento. De um modo preferido, os parâmetros "Peso do Vazio" e "Peso do Comprimento" utilizados em cada um dos programas são 50 e 3, para sequências de polinucleótidos e 12 e 4 para sequências de polipéptidos, respectivamente. De um modo preferido, as % de identidade e similaridade são determinadas quando as duas sequências a serem comparadas estão optimamente alinhadas.

Também são conhecidos na técnica outros programas para determinar a identidade e/ou similaridade entre sequências, por exemplo a família BLAST de programas (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, disponível no National Center for Biotechnology Information (NCBI), Bethesda, Md., EUA e acessível através da página principal do NCBI em www.ncbi.nlm.nih.gov) e FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R e

Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448,1988, disponível como parte do Wisconsin Sequence Analysis Package).

De um modo preferido utiliza-se a matriz de substituição de aminoácidos BLOSUM62 (Henikoff S e Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) nas comparações das sequências de polipéptidos incluindo nos casos em que as sequências de nucleótidos são primeiramente traduzidas em sequências de aminoácidos antes da comparação.

De um modo preferido utiliza-se o programa BESTFIT para determinar a % de identidade de uma sequência de polinucleótido ou polipéptido investigada em relação a uma sequência de polinucleótido ou polipéptido de referência, estando a sequência investigada e de referência optimamente alinhadas e os parâmetros do programa definidos para os valores de base, como se descreveu acima.

O "Índice de Identidade" é uma medida do parentesco da sequência o qual pode ser utilizado para comparar uma sequência candidata (polinucleótido ou polipéptido) e uma sequência de referência. Assim, por exemplo, uma sequência de polinucleótido candidata possuindo, por exemplo, um Índice de Identidade de 0,95 relativamente a uma sequência de polinucleótido de referência é idêntica à sequência de referência excepto que a sequência de polinucleótido candidata pode incluir em média até cinco diferenças por cada 100 nucleótidos da sequência de referência. Tais diferenças são seleccionadas do grupo consistindo de pelo menos uma deleção, substituição, incluindo transição e transversão, ou inserção de nucleótido. Estas diferenças

podem ocorrer nas posições terminais 5' ou 3' da sequência de polinucleótido de referência ou em qualquer parte entre estas posições terminais, dispersadas quer individualmente entre os nucleótidos na sequência de referência ou em um ou mais grupos contíguos dentro da sequência de referência. Por outras palavras, para se obter uma sequência de polinucleótido com um Índice de Identidade de 0,95 relativamente a uma sequência de polinucleótido de referência, pode eliminar-se, substituir ou inserir, ou qualquer combinação destes, uma média de até 5 por cada 100 dos nucleótidos na sequência de referência, como se descreveu acima. O mesmo se aplica *mutatis mutandis* a outros valores do Índice de Identidade, por exemplo 0,96, 0,97, 0,98 e 0,99.

De modo semelhante, para um polipéptido, uma sequência de polipéptido candidata possuindo, por exemplo, um Índice de Identidade de 0,95 relativamente a uma sequência de polipéptido de referência é idêntica à sequência de referência excepto que a sequência de polipéptido pode incluir uma média de até cinco diferenças por cada 100 aminoácidos da sequência de referência. Tais diferenças são seleccionadas do grupo consistindo de pelo menos uma deleção, substituição, incluindo substituição conservadora e não conservadora, ou inserção de aminoácido. Estas diferenças podem ocorrer nas posições terminais amino ou carboxilo da sequência de polipéptido de referência ou em qualquer parte entre estas posições terminais, dispersadas quer individualmente entre os aminoácidos na sequência de referência ou em um ou mais grupos contíguos dentro da sequência de referência. Por outras palavras, para se obter uma sequência de polipéptido com um Índice de Identidade de

0,95 relativamente a uma sequência de polipéptido de referência, pode eliminar-se, substituir ou inserir, ou qualquer combinação destes, uma média de até 5 por cada 100 dos aminoácidos na sequência de referência, como se descreveu acima. O mesmo se aplica *mutatis mutandis* para outros valores do Índice de Identidade, por exemplo 0,96, 0,97, 0,98 e 0,99.

A relação entre o número de diferenças de nucleótidos ou aminoácidos e o Índice de Identidade pode ser expressa na equação seguinte:

$$n_a \leq x_a - (x_a \bullet I),$$

na qual:

n_a é o número de diferenças de nucleótidos ou aminoácidos,

x_a é o número total de nucleótidos ou aminoácidos na SEQ ID NO:1 ou SEQ ID NO:2, respectivamente,

I é o Índice de Identidade,

\bullet é o símbolo para o operador multiplicação, e

na qual qualquer produto não inteiro de x_a e I é arredondado para baixo para o inteiro mais próximo antes de o subtrair de x_a .

“Homólogo” é um termo genérico utilizado na técnica para indicar uma sequência de polinucleótido ou polipéptido possuindo um grau elevado de parentesco de sequência com a

sequência de referência. Um tal parentesco pode ser quantificado determinando o grau de identidade e/ou similaridade entre as duas sequências como se definiu acima. Dentro deste termo genérico estão os termos "ortólogo" e "parólogo". "Ortólogo" refere-se a um polinucleótido ou polipéptido que é o equivalente funcional do polinucleótido ou polipéptido noutras espécies. "Parólogo" refere-se a um polinucleótido ou polipéptido que dentro da mesma espécie é funcionalmente semelhante.

"Proteína de fusão" refere-se a uma proteína codificada por dois genes, ou fragmentos destes, não relacionados, fundidos. Exemplos foram descritos nas Patentes U.S. nº 5 541 087, 5 726 044. No caso de Fc-Carp-2 é vantajosa a utilização de uma região Fc de imunoglobulina como uma parte da proteína de fusão para efectuar a expressão funcional da Fc-Carp-2 ou de fragmentos da Carp-2, para melhorar as propriedades farmacocinéticas de uma tal proteína de fusão quando utilizada em terapêutica e para produzir uma Carp-2 dimérica. A construção de ADN Fc-Carp-2 compreende na direcção 5' para 3', uma cassette de secreção, isto é uma sequência de sinalização que inicia a exportação a partir de uma célula de mamífero, ADN que codifica um fragmento da região Fc da imunoglobulina, como um parceiro de fusão, e um ADN que codifica o Carp-2 ou fragmentos deste. Nalgumas utilizações seria desejável alterar as propriedades funcionais intrínsecas (ligação de complemento, ligação do receptor Fc) mutando os lados Fc funcionais enquanto se deixa o resto da proteína de fusão inalterado ou eliminando completamente a parte Fc após expressão.

Todas as publicações e referências, incluindo, entre outros, as patentes e os pedidos de patentes, citadas nesta especificação são aqui incorporadas por referência na sua totalidade como se cada publicação ou referência individual fosse específica e individualmente indicada para ser aqui incorporada por referência como se estivesse totalmente descrita. Qualquer pedido de patente em relação ao qual seja reivindicada prioridade por este pedido é também aqui incorporado por referência na sua totalidade do modo descrito acima para as publicações e referências.

Legendas das figuras:

Quadro: 1 Padrão de expressão do CARP-2

A análise da expressão do ARNm do polipéptido CARP-2 em vários tecidos e órgãos humanos foi realizada como se descreve no Exemplo 2

Figura: 1 Análise por transferência de Northern

Para se analisar a expressão do CARP-2 humano em tecidos humanos realizou-se uma transferência de Northern (clontech). Detectou-se uma banda de cerca de 2 kb no coração e no músculo-esquelético.

Outros Exemplos

Exemplo 1

Transferência de Northern

Transferência de Northern de vários tecidos (Clontech Laboratories Inc, Palo Alto, Calif., EUA), com um fragmento génico de 1000 pb do CARP-2 marcado com ^{32}P radioactivo hibridizado com ARN proveniente de tecido humano. O tamanho do ARNm foi estimado como sendo de cerca de 2 kb e foi detectado no coração e músculo-esquelético (Figura 1).

Exemplo 2

Análise da expressão génica

Aplicou-se vários ARNm de origem humana (MTE, Clontech Laboratories Inc, Palo Alto, Calif., EUA) numa membrana de nylon. A membrana foi hibridizada com um fragmento do CARP-2 com 1000 pb marcado com ^{32}P radioactivo. Examinou-se sessenta e umas amostras individuais (dadas no Quadro 1) de origem humana. Percutiu-se a radioactividade e quantificou-se as manchas (todos os valores foram normalizados ao valor mais elevado fixado = 100%). Além da expressão no coração e músculo-esquelético, detectou-se CARP-2 na glândula tiroideia. No coração, constatou-se que a expressão de CARP-2 é mais acentuada no ventrículo do que na aurícula, tendo-se também encontrado uma expressão elevada no septo interventricular e no vértice do coração.

LISTAGEM DE SEQUÊNCIAS

<110> Merck Patent GmbH

<120> Proteína de Repetição de Anquirina Cardíaca 2

<130> CARP-2FWKWS

<140>

<141>

<160> 2
 <170> PatentIn Ver. 2.1
 <210> 1
 <211> 1002
 <212> ADN
 <213> Homo sapiens
 <220>
 <221> CDS
 <222> (1)..(1002)
 <400> 1

```

atg gac ggc acc atg gag gac tcc gag gcc gtg cag agg gcc ace gcc 48
Met Asp Gly Thr Met Glu Asp Ser Glu Ala Val Gln Arg Ala Thr Ala
  1             5             10             15

ctc atc gag cag cgg ctg gca cag gag gag gag aat gag aaa ctc cga 96
Leu Ile Glu Gln Arg Leu Ala Gln Glu Glu Glu Asn Glu Lys Leu Arg
             20             25             30

gga gac ace cgc cag aag ctg ccc atg gac ttg ctg gtg ctg gag gat 144
Gly Asp Thr Arg Gln Lys Leu Pro Met Asp Leu Leu Val Leu Glu Asp
             35             40             45

gag aag cac cac ggg gct cag agt gca gcc ctg cag aag gtg aag gcc 192
Glu Lys His His Gly Ala Gln Ser Ala Ala Leu Gln Lys Val Lys Gly
             50             55             60

caa gag cgc gtg cgc aag acg tcc ctg gac ctg cgg cgg gag atc atc 240
Gln Glu Arg Val Arg Lys Thr Ser Leu Asp Leu Arg Arg Glu Ile Ile
             65             70             75             80

gat gtg gcc ggg atc cag aac ctc atc gag ctg cgg aag aaa cgc aag 288
Asp Val Gly Gly Ile Gln Asn Leu Ile Glu Leu Arg Lys Lys Arg Lys
             85             90             95

cag aag aag cgg gac gct ctg gcc gcc tcg cat gag ccg ccc cca gag 336
Gln Lys Lys Arg Asp Ala Leu Ala Ala Ser His Glu Pro Pro Pro Glu
             100             105             110

ccc gag gag atc act gcc cct gtg gat gag gag acc ttc ctg aaa gct 384
Pro Glu Glu Ile Thr Gly Pro Val Asp Glu Glu Thr Phe Leu Lys Ala
             115             120             125

gcc gtg gag ggg aaa atg aag gtc att gag aag ttc ctg gct gac ggg 432
Ala Val Glu Gly Lys Met Lys Val Ile Glu Lys Phe Leu Ala Asp Gly

```

130	135	140	
ggg tca gcc gac acg tgc gac cag ttc cgt cgg acc gca ctg cac cga 480			
Gly Ser Ala Asp Thr Cys Asp Gln Phe Arg Arg Thr Ala Leu His Arg			
145	150	155	160
gct tcc ctg gaa ggc cac atg gaa atc ctg gag aag ctt cta gat aat 528			
Ala Ser Leu Glu Gly His Met Glu Ile Leu Glu Lys Leu Leu Asp Asn			
	165	170	175
ggg gcc act gtg gac ttc cag gat cgg ctg gac tgc acc gcc atg cat 576			
Gly Ala Thr Val Asp Phe Gln Asp Arg Leu Asp Cys Thr Ala Met His			
	180	185	190
tgg gcc tgc cgc ggg ggc cac tta gag gtg gtg aaa ctt ctg caa agc 624			
Trp Ala Cys Arg Gly Gly His Leu Glu Val Val Lys Leu Leu Gln Ser			
	195	200	205
cat gga gca gac acc aat gtg agg gat aag ctg ctg agc acc cgg ctg 672			
His Gly Ala Asp Thr Asn Val Arg Asp Lys Leu Leu Ser Thr Pro Leu			
	210	215	220
cac gtg gca gtc cgg acc ggg cag gtg gag att gtg gag cac ttc cta 720			
His Val Ala Val Arg Thr Gly Gln Val Glu Ile Val Glu His Phe Leu			
	225	230	235
tcc ctg ggc ctg gaa atc aat gcc aga gac agg gaa ggg gat act gcc 768			
Ser Leu Gly Leu Glu Ile Asn Ala Arg Asp Arg Glu Gly Asp Thr Ala			
	245	250	255
ctg cat gac gct gtg agg ctc aac cgc tac aaa atc atc aaa ctg ctg 816			
Leu His Asp Ala Val Arg Leu Asn Arg Tyr Lys Ile Ile Lys Leu Leu			
	260	265	270
ctc ctg cat ggg gct gac atg atg acc aag aac ctg gca gga aag acc 864			
Leu Leu His Gly Ala Asp Met Met Thr Lys Asn Leu Ala Gly Lys Thr			
	275	280	285
cgg acg gac ctg gtg cag ctc tgg cag gct gat acc cgg cac gcc ctg 912			
Pro Thr Asp Leu Val Gln Leu Trp Gln Ala Asp Thr Arg His Ala Leu			
	290	295	300
gag cat cct gag cgg ggg gct gag cat aac ggg ctg gag ggg cct aat 960			
Glu His Pro Glu Pro Gly Ala Glu His Asn Gly Leu Glu Gly Pro Asn			
	305	310	315
gat agt ggg cga gag acc cct cag cct gtg cca gcc cag tga 1002			
Asp Ser Gly Arg Glu Thr Pro Gln Pro Val Pro Ala Gln			
	325	330	

<210> 2

<211> 333

<212> PRT

<213> Homo Sapiens

<400> 2

```

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

```

Lisboa, 8 de Janeiro de 2007

REIVINDICAÇÕES

1. Polipéptido compreendendo a sequência de SEQ ID NO:1.
2. Polipéptido compreendendo a sequência de SEQ ID NO:2, codificado pelo polinucleótido da reivindicação 1.
3. Vector de expressão compreendendo um polinucleótido da reivindicação 1 capaz de produzir um polipéptido da reivindicação 2 quando o referido vector de expressão está presente numa célula hospedeira compatível.
4. Célula hospedeira recombinante compreendendo o vector de expressão da reivindicação 3 ou uma membrana desta que expressa o polipéptido da reivindicação 2.
5. Processo para produzir um polipéptido da reivindicação 2 compreendendo o passo de cultivar uma célula hospedeira como definida na reivindicação 4 em condições suficientes para a produção do referido polipéptido e recuperar o polipéptido do meio de cultura.
6. Proteína de fusão consistindo da região Fc da imunoglobulina e um polipéptido da reivindicação 2.
7. Anticorpo imuno específico para o polipéptido da reivindicação 2.
8. Método de selecção para identificar compostos que estimulam ou inibem a função ou nível do polipéptido da reivindicação 2 compreendendo um método seleccionado do grupo consistindo de:

(a) medir ou, detectar, quantitativa ou qualitativamente, a ligação de um composto candidato ao polipéptido (ou às células ou membranas que expressam o polipéptido) ou a uma proteína de fusão deste por meio de uma etiqueta directa ou indirectamente associada ao composto candidato;

(b) medir a competição de ligação de um composto candidato ao polipéptido (ou às células ou membranas que expressam o polipéptido) ou a uma proteína de fusão deste na presença de um competidor marcado;

(c) testar se o composto candidato origina um sinal gerado pela activação ou inibição do polipéptido, utilizando sistemas de detecção apropriados para as células ou membranas celulares que expressam o polipéptido;

(d) misturar um composto candidato com uma solução contendo um polipéptido da reivindicação 2, para formar uma mistura, medir a actividade do polipéptido na mistura e comparar a actividade da mistura com uma mistura de controlo sem composto candidato; ou

(e) detectar o efeito de um composto candidato na produção de ARNm que codifica o referido polipéptido ou o referido polipéptido em células, utilizando por exemplo, um ensaio ELISA.

Lisboa, 8 de Janeiro de 2007

Quadro 1:

Órgãos humanos	* %	Órgãos humanos	* %
1. cérebro completo	12,4	31. estômago	11,9
2. córtex cerebral	9,7	32. duodeno	10,4
3. lóbulo frontal	9,4	33. jejuno	7,3
4. lóbulo parietal	10,4	34. íleo	3,1
5. lóbulo occipital	10,0	35. íleo-cego	4,6
6. lóbulo temporal	10,6	36. apêndice	3,9
7. circunvolução paracentral	14,5	37. cólon, ascendente	8,9
8. ponto de Valore	14,0	38. cólon, transverso	5,3
9. cerebelo esquerdo	9,7	39. cólon, descendente	5,8
10. cerebelo direito	7,5	40. recto	6,6
11. corpo caloso	6,9	41. rim	18,8
12. amígdala	10,2	42. músculo-esquelético	100,0
13. núcleo caudado	11,6	43. baço	2,8
14. hipocampo	9,6	44. timo	1,2
15. bolbo raquidiano	11,0	45. leucócitos PB	n.d.
16. putamen	13,5	46. gânglio linfático	5,8
17. substância negra	13,7	47. medula óssea	17,1
18. nucleus accumbens	11,2	48. traqueia	12,8
19. tálamo	14,3	49. pulmão	10,1
20. glândula pituitária	8,5	50. placenta	9,2
21. espinal-medula	9,0	51. bexiga	4,5
22. coração	60,4	52. útero	n.d.
23. aorta	16,0	53. próstata	16,1
24. aurícula, esquerda	29,3	54. testículos	9,2
25. aurícula, direita	27,3	55. ovário	4,2
26. ventrículo, esquerdo	60,7	56. fígado	12,4
27. ventrículo, direito	57,6	57. pâncreas	12,1
28. septo interventricular	79,7	58. glândula supra-renal	12,3
29. vértice do coração	75,7	59. glândula tiroideia	37,9
30. esófago	24,6	60. glândula salivar	18,0
		61. glândula mamária	16,0
*) Os sinais de hibridação foram quantificados e representados como percentagem do valor mais elevado detectado no músculo-esquelético			

