



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

| | |
|--------------------|-------------------|
| RESOLUÇÃO | 28/03/2013 |
| PRESIDÊNCIA | Nº 81/2013 |

Assunto: Dispõe sobre os procedimentos para a apresentação da “Listagem de Sequências”, em meio eletrônico, para fins de complementação do relatório descritivo constante dos pedidos de patentes depositados no INPI, bem como sobre as regras para a representação das sequências de nucleotídeos e de aminoácidos na “Listagem de Sequências”, e revoga o item 16.3 do Ato Normativo nº 127, de 5 de março de 1997, e a Resolução nº 210, de 07 de maio de 2009.

O PRESIDENTE DO INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL - INPI e o DIRETOR DE PATENTES, no uso das suas atribuições regimentais,

RESOLVEM:

Art. 1º Esta Resolução dispõe sobre os procedimentos para a apresentação da “Listagem de Sequências”, em meio eletrônico, para fins de complementação do relatório descritivo constante dos pedidos de patentes depositados no INPI a partir da data da entrada em vigor desta Resolução, bem como sobre as regras para a representação das sequências de nucleotídeos e/ou de aminoácidos na “Listagem de Sequências”.

Art. 2º O requerente de pedido de patente que contenha em seu objeto uma ou mais sequências de nucleotídeos e/ou de aminoácidos, que sejam fundamentais para a descrição da invenção, deverá representá-las em uma “Listagem de Sequências”, com vistas à aferição da suficiência descritiva, de que trata o art. 24 da Lei nº 9.279, de 14 de maio de 1996 (Lei da Propriedade Industrial - LPI).

DA “LISTAGEM DE SEQUÊNCIAS” EM ARQUIVO ELETRÔNICO NO FORMATO TEXTO (TXT)

Art. 3º A “Listagem de Sequências” deverá ser apresentada ao INPI, como instrumento complementar ao relatório descritivo, no formato de leitura por computador (arquivo eletrônico), gravado em disco compacto não regravável (CD) ou em disco digital não regravável (DVD), sendo que o arquivo eletrônico da “Listagem de Sequências” deverá ser gerado em formato texto (TXT).

Art. 4º A representação das sequências de nucleotídeos e/ou de aminoácidos na “Listagem de Sequências” deverá seguir o Padrão OMPI ST.25, definido pela Organização Mundial da Propriedade Intelectual - OMPI, de acordo com as regras constantes do Anexo a esta Resolução.

§ 1º Devem ser incluídas na “Listagem de Sequências” todas as sequências lineares de 4 (quatro) ou mais L-aminoácidos contínuos de um peptídeo ou de uma proteína e todas as sequências lineares que tenham 10 (dez) ou mais nucleotídeos contínuos, mesmo as que não tenham sido reivindicadas, como, por exemplo, sondas de PCR, desde que preencham as condições definidas neste parágrafo.

§ 2º As sequências ramificadas, as sequências com menos de 10 (dez) nucleotídeos, as sequências com menos de 4 (quatro) L-aminoácidos e as sequências de aminoácidos que contenham pelo menos um D-aminoácido, bem como as sequências compreendendo nucleotídeos ou aminoácidos diferentes dos que estão listados nas Tabelas 1, 2, 3 e 4, constantes do Anexo desta Resolução, devem ser incluídas no relatório descritivo do pedido de patente, não podendo constar da “Listagem de Sequências”.

DA “LISTAGEM DE SEQUÊNCIAS” EM ARQUIVO ELETRÔNICO NO FORMATO PORTABLE DOCUMENT FORMAT (PDF)

Art. 5º O CD ou o DVD apresentado, contendo o arquivo eletrônico da “Listagem de Sequências” em formato TXT, deverá conter, também, um segundo arquivo eletrônico, em formato *Portable Document Format* (PDF), correspondente à cópia da “Listagem de Sequências”, idêntica e integral àquela apresentada em formato TXT, para fins de disponibilização ao público por parte do INPI.

Parágrafo único. O arquivo eletrônico da “Listagem de Sequências” em formato PDF deverá ser gerado pelo requerente, a partir do arquivo eletrônico da “Listagem de Sequências” em formato TXT, por meio de um programa de computador, denominado SisBioList, disponível no Portal do INPI na Internet, no endereço www.inpi.gov.br.

DO CÓDIGO DE CONTROLE DA “LISTAGEM DE SEQUÊNCIAS”

Art. 6º O CD ou o DVD apresentado, contendo os arquivos eletrônicos da “Listagem de Sequências” em formatos TXT e PDF, deverá conter, ainda, um terceiro arquivo eletrônico correspondente ao Código de Controle Alfanumérico do arquivo eletrônico da “Listagem de Sequências” em formato TXT, destinado a certificar a autenticidade do seu conteúdo.

Parágrafo único. O arquivo eletrônico contendo o Código de Controle Alfanumérico da “Listagem de Sequências” será gerado automaticamente, a partir do arquivo da “Listagem de Sequências” em formato TXT, por meio do SisBioList, quando da geração do arquivo eletrônico da “Listagem de Sequências” em formato PDF.

DA APRESENTAÇÃO DA “LISTAGEM DE SEQUÊNCIAS”

Art. 7º O CD ou o DVD contendo os arquivos eletrônicos da “Listagem de Sequências” nos formatos TXT e PDF e o arquivo eletrônico do Código de Controle Alfanumérico da “Listagem de Sequências”, deverá ser apresentado ao INPI, no ato do depósito do pedido de patente.

§ 1º Quando o CD ou o DVD não for apresentado ao INPI no ato do depósito, poderá ser ele apresentado pelo requerente, independentemente de notificação ou exigência por parte do INPI, até a data do requerimento do exame do pedido de patente, de que trata o art. 33 da LPI, por meio de petição isenta do pagamento de retribuição.

§ 2º A petição apresentada na forma do parágrafo anterior, deverá estar instruída com a declaração expressa do requerente de que *“a informação contida na ‘Listagem de Sequências’ apresentada em formato eletrônico está limitada ao conteúdo da matéria revelada pelas sequências de aminoácidos e/ou de nucleotídeos divulgadas no pedido de patente, conforme depositado”*.

§ 3º Quando a “Listagem de Sequências” no formato de arquivo eletrônico não for apresentada nos prazos previstos no *caput* e no parágrafo primeiro deste artigo, o INPI formulará as exigências necessárias à regularização do pedido de patente, com vistas ao cumprimento do disposto nesta Resolução, que deverão ser atendidas, nos termos e prazos da LPI.

§ 4º Por ocasião do cumprimento da exigência de que trata o parágrafo anterior, o requerente deverá apresentar declaração expressa de que *“a informação contida na ‘Listagem de Sequências’ apresentada em formato eletrônico está limitada ao conteúdo da matéria revelada pelas sequências de aminoácidos e/ou de nucleotídeos divulgadas no pedido de patente, conforme depositado”*.

Art. 8º Se a “Listagem de Sequências” for corrigida subsequentemente a sua apresentação, de ofício ou a requerimento do requerente, este deverá apresentar ao INPI novo CD ou DVD contendo o arquivo eletrônico da “Listagem de Sequências” corrigida, em formatos TXT e PDF, observando, igualmente, as disposições dos arts. 5º e 6º desta Resolução.

Parágrafo único. Na hipótese de que trata o *caput* deste artigo, o requerente deverá apresentar ao INPI o CD ou o DVD contendo os arquivos eletrônicos da “Listagem de Sequências” corrigidas, nos formatos TXT e PDF, e o arquivo eletrônico do Código de Controle Alfanumérico referente à “Listagem de Sequências” corrigida, por meio de petição, acompanhada do comprovante do recolhimento da retribuição correspondente ao ato processual, bem como da declaração expressa do requerente de que *“a informação contida na ‘Listagem de Sequências’ corrigida, apresentada em formato eletrônico, não configura acréscimo de matéria àquela constante do correspondente pedido de patente depositado, conforme depositado”*.

DA IDENTIFICAÇÃO DO CD OU DO DVD CONTENDO A “LISTAGEM DE SEQUÊNCIAS”

Art. 9º O CD ou o DVD apresentado, contendo os arquivos eletrônicos das “Listagens de Sequências”, nos formatos TXT e PDF, e o arquivo eletrônico do Código de Controle Alfanumérico da “Listagem de Sequências”, deverá estar identificado com uma etiqueta, a qual deverá conter o Código de Controle Alfanumérico da “Listagem de Sequências” e o número da Guia de Recolhimento Único - GRU relativa ao ato processual correspondente, se for o caso.

Parágrafo único. No caso do CD ou do DVD apresentado referir-se a um pedido de patente já depositado no INPI e que já tenha numeração própria, a etiqueta deverá conter, também, a numeração do pedido de patente.

DA ENTREGA DO CD OU DO DVD CONTENDO A “LISTAGEM DE SEQUÊNCIAS”

Art. 10 O CD ou o DVD contendo os arquivos eletrônicos das “Listagens de Sequências”, nos formatos TXT e PDF, e o arquivo eletrônico do Código de Controle Alfanumérico da “Listagem de Sequências”, deverá ser apresentado com uma duplicata, acomodados em porta CD ou DVD individuais de plástico transparente modelo *slim* (cerca de 5 mm de espessura), acompanhados de declaração expressa do requerente de que *“os arquivos eletrônicos contidos nos dois CDs ou DVDs são idênticos”*.

DOS FORMULÁRIOS

Art. 11 Quando da apresentação do CD ou do DVD contendo os arquivos eletrônicos das “Listagens de Sequências”, nos termos e prazos previstos nesta Resolução, o requerente de pedido de patente deverá informar ao INPI, no campo específico do formulário, que está apresentando a “Listagem de Sequências”, informando, ainda, o Código de Controle Alfanumérico da Listagem de Sequências, na forma indicada no próprio formulário.

DO PEDIDO INTERNACIONAL DE PATENTE

Art. 12 As disposições desta Resolução aplicam-se ao pedido de patente oriundo de pedido internacional de patente depositado nos termos do Tratado de Cooperação em Matéria de Patentes - PCT, quando da sua entrada na fase nacional, apresentado ao INPI em conformidade com a legislação vigente.

DA “LISTAGEM DE SEQUÊNCIAS” ADICIONAL EM FORMATO IMPRESSO

Art. 13 A “Listagem de Sequências” poderá ser adicionalmente apresentada em formato impresso, como parte integrante do pedido de patente.

§ 1º A “Listagem de Sequências” que for adicionalmente apresentada no formato impresso quando do depósito do pedido de patente, deverá ser incluída após o relatório descritivo, sendo iniciada em uma página separada, sob o título “Listagem de Sequências”, e entregue em 3 (três) vias, para uso do INPI, sendo facultada a apresentação de mais uma via, para restituição ao requerente.

§ 2º As páginas da “Listagem de Sequências” de que trata o *caput* deverão ser numeradas de forma sequencial e independente, com algarismos arábicos, no centro da parte superior, entre 1 e 2 cm do limite da página.

§ 3º A “Listagem de Sequências” referida no *caput* deverá apresentar conteúdo idêntico àquela apresentada no formato de arquivo eletrônico, em TXT e PDF, exceto quanto à numeração das suas respectivas páginas, e estar acompanhada da declaração expressa do requerente de que “*a ‘Listagem de Sequências’ apresentada em formato impresso é idêntica àquela contida no formato de arquivo eletrônico, exceto quanto à numeração das suas respectivas páginas*”.

DA CARTA PATENTE

Art. 14 Constará da Carta-Patente, além das informações e dos documentos de que trata o art. 39 da LPI, um CD ou DVD contendo os arquivos da “Listagem de

Sequências”, em formatos TXT e PDF, e o arquivo eletrônico com o Código de Controle Alfanumérico, bem como a “Listagem de Sequências” em formato impresso, se houver.

DAS DISPOSIÇÕES FINAIS

Art. 15 O requerente de pedido de patente em andamento no INPI que, na data da entrada em vigor desta Resolução, tenha apresentado a “Listagem de Sequências” em formato impresso, poderá apresentar a “Listagem de Sequências” em formato de arquivo eletrônico, nas condições estabelecidas por esta Resolução, voluntariamente ou a requerimento do INPI, por meio de petição isenta do pagamento de retribuição.

Art. 16 O requerente de pedido de patente em andamento no INPI que, na data da entrada em vigor desta Resolução, não tenha apresentado a “Listagem de Sequências” em formato impresso, deverá apresentar a “Listagem de Sequências” em formato de arquivo eletrônico, nas condições estabelecidas por esta Resolução, voluntariamente ou a requerimento do INPI, por meio de petição, acompanhada do comprovante do recolhimento da retribuição correspondente ao ato processual.

Art. 17 Fica revogada a Resolução/INPI nº 70, de 18 de março de 2013.

Art. 18 Esta Resolução entra em vigor na data de sua publicação no Diário Oficial da União.

Jorge de Paula Costa Ávila
Presidente

Julio César Castelo Branco Reis Moreira
Diretor de Patentes

ANEXO

REGRAS PARA APRESENTAÇÃO E REPRESENTAÇÃO DE SEQUÊNCIAS DE AMINOÁCIDOS E DE NUCLEOTIDEOS NA “LISTAGEM DE SEQUÊNCIAS” NO FORMATO OMPI ST.25¹

1. Das definições:

- 1.1. **Identificador de sequências** é um número inteiro único que corresponde a SEQ ID NO: assinalada para cada sequência da listagem de sequências, sendo que a primeira sequência definida na “Listagem de Sequências”, SEQ ID NO: 1, deve ser a sequência mais importante da invenção.
- 1.2. **Identificador numérico** é um número de três dígitos que representa um elemento específico de dados, alocado entre os símbolos <>.
- 1.3. **Vocabulário linguisticamente neutro**² corresponde a um vocabulário padrão que se utiliza na listagem de sequências para representar os termos científicos no formato prescrito por provedores de dados de sequências (incluindo o nome científico, os qualificadores e seus valores em relação ao vocabulário, os símbolos das Tabelas 1, 2, 3 e 4 e as chaves de caracterização que figuram nas Tabelas 5 e 6).
- 1.4. **Texto livre**³ é a descrição textual das características de uma sequência em virtude do identificador numérico <223> (Outra informação), na qual se emprega um vocabulário distinto do vocabulário linguisticamente neutro definido no item 1.3.

¹ A “Listagem de Sequências” em formato Padrão OMPI ST.25 poderá ser criada por quaisquer meios ou por um software específico, tal como o PatentIn¹, ou por outros programas de informática personalizados, desde que o arquivo eletrônico gerado possa ser lido pelo sistema operacional de um computador pessoal em uso no INPI. O software PatentIn foi concebido pelo Escritório de Patentes e Marcas dos Estados Unidos (USPTO) para acelerar o processo de preparar listagem de seqüências num formato eletrônico padronizado segundo as normas recomendadas pela OMPI para representação de seqüências de nucleotídeos e de aminoácidos em listagem de seqüências. O software PatentIn se encontra disponível para download gratuito a partir do sitio do USPTO na internet (www.uspto.gov).

² Este vocabulário deverá estar presente na “Listagem de Sequências” seguindo os padrões internacionais, não devendo ser traduzido para a língua vernácula, nos seguintes identificadores numéricos <212>, <213> e <221>.

³ Texto livre deverá ser introduzido na “Listagem de Sequências” em língua vernácula

2. Da representação das sequências biológicas no formato OMPI ST.25:

- 2.1. Cada sequência deverá ser assinalada com um identificador de sequência distinto. Os identificadores de sequências deverão ser iniciados com o número 1 e irão aumentando sequencialmente por números inteiros tais como "SEQ ID NO:1", "SEQ ID NO:2", "SEQ ID NO:3", etc.
- 2.2. No relatório descritivo, nas reivindicações e nos desenhos do pedido, as sequências representadas na listagem de sequências deverão ser referidas mediante o identificador de sequência precedido de "SEQ ID NO:".
- 2.3. As sequências de nucleotídeos e de aminoácidos deverão estar representadas por pelo menos uma das três possibilidades seguintes:
 - (i) uma sequência de nucleotídeos pura;
 - (ii) uma sequência de aminoácidos pura;
 - (iii) uma sequência de nucleotídeos e a correspondente sequência de aminoácidos.
- 2.4. Nas sequências representadas no formato especificado na opção (iii), a sequência de aminoácidos deverá ser adicionalmente revelada na listagem de sequências como uma sequência de aminoácidos pura e com um identificador de sequência diferente, composto por um número inteiro.

3. Do formato e dos símbolos que devem ser utilizados em sequências de nucleotídeos:

- 3.1. Toda sequência de nucleotídeos deverá ser representada unicamente por fita simples, no sentido 5' para 3' e da esquerda para a direita.
- 3.2. Toda sequência de nucleotídeos deverá ser representada por um máximo de 60 bases por linha, tendo um espaço entre cada grupo de 10 bases.
- 3.3. As bases das regiões codificadoras de uma sequência de nucleotídeos deverão figurar como tripletes (códons).
- 3.4. As bases de uma sequência de nucleotídeos deverão ser representadas usando o código de uma letra para os caracteres de nucleotídeos de sequência. Somente deverão ser usadas letras minúsculas, em conformidade com a listagem fornecida na Tabela 1.
- 3.5. As bases modificadas deverão ser representadas mediante as bases correspondentes não modificadas ou mediante o caractere "n" na própria sequência, caso a base modificada é uma das que figurem na Tabela 2.

4. Do formato e dos símbolos que devem ser utilizados em sequências de aminoácidos:

- 4.1. Toda sequência de proteína ou de peptídeo deverá ser representada com um máximo de 16 aminoácidos por linha, deixando um espaço entre cada aminoácido.
- 4.2. Os aminoácidos correspondentes aos códons das regiões codificadoras de uma sequência de nucleotídeos, deverão figurar imediatamente abaixo dos códons correspondentes. Quando um códon estiver interrompido por um ítron, o símbolo do aminoácido figurará debaixo da porção do códon que contenha dois nucleotídeos.
- 4.3. A numeração dos aminoácidos deverá ser iniciada no primeiro aminoácido da sequência com o número 1.
- 4.4. Alternativamente, os aminoácidos que precedem a proteína madura, por exemplo, as pré-sequências, as pró-sequências e as pré-pró-sequências, assim como as sequências sinal, quando existentes, poderão ter números negativos, contados em forma regressiva, a partir do aminoácido adjacente ao número 1.
- 4.5. Não se empregará o zero (0) quando a numeração dos aminoácidos empregar números negativos para distinguir a proteína madura.
- 4.6. Toda sequência de aminoácidos composta por um ou mais segmentos não contínuos de uma sequência maior ou de segmentos de sequências diferentes, deverá ser numerada como uma sequência distinta e com um identificador de sequência diferente.
- 4.7. Os aminoácidos de uma sequência de proteína ou de peptídeo deverão ser representados no sentido do grupamento amino para o grupamento carboxila e da esquerda para a direita.
- 4.8. Os aminoácidos deverão ser representados utilizando o código de três letras, sendo a primeira letra uma letra maiúscula, em conformidade com a listagem dada na Tabela 3.

5. Dos elementos de dados obrigatórios:

- 5.1. A listagem de sequências deverá incluir, em adição a, e imediatamente antes da sequência de nucleotídeos e/ou aminoácidos, os seguintes elementos de informação (elementos de dados obrigatórios):

| | |
|-------|---------------------------------|
| <110> | Nome do requerente |
| <120> | Título da invenção ⁴ |
| <160> | Número total de SEQ ID NOs |
| <210> | SEQ ID NO: # |
| <211> | Comprimento |
| <212> | Tipo |
| <213> | Organismo |
| <400> | Sequência |

Quando o nome do requerente (identificador numérico <110>) estiver escrito em caracteres outros que não os pertencentes ao alfabeto latino, também deverá aparecer em caracteres do alfabeto latino, seja como uma simples transliteração do nome ou através da sua tradução para o inglês.

- 5.2. Se for empregado na sequência o caractere "n" ou Xaa", ou uma base modificada, ou um L-aminoácido modificado ou pouco comum, os seguintes elementos de dados serão obrigatórios:

| | |
|--------|-------------------------------|
| <220>> | Característica ⁵ |
| <221> | Nome/chave ⁶ |
| <222> | Localização ⁷ |
| <223> | Outra informação ⁸ |

- 5.3. Se o organismo (identificador numérico <213>) é uma "Sequência artificial" ou "Desconhecida", os seguintes elementos de dados são obrigatórios:

| | |
|-------|-------------------------------|
| <220> | Característica ⁶ |
| <223> | Outra informação ⁹ |

⁴ Em língua vernácula

⁵ Descrição dos pontos de importância biológica para a sequência na SEQ ID NO: # (pode repetir-se em função do número de características indicadas).

⁶ Só se empregarão as chaves definidas nas Tabelas 5 e 6

⁷ Especifique localização dentro da seqüência; quando necessário; especifique o número do primeiro e do último aminoácido ou base

⁸ Em língua vernácula, vide item 1.4. Neste indicador, deve ser dada qualquer outra informação pertinente utilizando um vocabulário linguisticamente neutro, ou em forma de texto livre; todo texto livre deverá estar presente no relatório descritivo no idioma do mesmo; quando estiver presente na sequência uma base modificada ou um L-aminoácido modificado ou pouco comum que figure nas Tabelas 2 e 4, o símbolo a ser usado associado com a dita base ou aminoácido será o que está representado nas Tabelas 2 e 4.

- 5.4. Quando uma listagem de seqüências é apresentada em conjunto com o pedido de patente no ato de seu depósito ou em qualquer momento **antes da designação de um número de depósito ao mesmo**, o seguinte elemento de dados deverá estar incluído obrigatoriamente na listagem de seqüências:

| | |
|--------------------|---------------------------------------------------------|
| <130> | Número de referência pessoal (indicado pelo requerente) |
|--------------------|---------------------------------------------------------|

- 5.5. Quando uma listagem de sequências é apresentada em resposta a uma exigência emitida por este INPI ou a qualquer momento **após a designação de um número de depósito**, os seguintes elementos de dados deverão estar obrigatoriamente incluídos na “Listagem de Seqüências”:

| | |
|--------------------|----------------------------------------|
| <140> | Número do pedido de patente em trâmite |
| <141> | Data de depósito do pedido de patente |

- 5.6. Além dos elementos de dados identificados acima, quando uma listagem de sequências é apresentada em relação a um pedido na qual se reivindica a prioridade de um pedido anterior, os seguintes elementos de dados deverão constar na “Listagem de Seqüências”:

| | |
|--------------------|--------------------------------------------------------------|
| <150> | Pedido de patente anterior (documento de prioridade) |
| <151> | Data de depósito do pedido de patente anterior (dia/mês/ano) |

6. Da apresentação das características:

- 6.1. Quando características da sequência são apresentadas (ou seja, identificador numérico **<220>⁶**), as mesmas deverão ser descritas mediante as “chaves de caracterização” definidas nas Tabelas 5 e 6.

7. Texto Livre:

- 7.1. A utilização do texto livre deverá estar limitada a uns poucos termos curtos que sejam indispensáveis para o entendimento da sequência.
- 7.2. Cada elemento de dados não excederá a quatro linhas com um máximo de 65 caracteres por linha.

7.3. Qualquer informação adicional deverá ser incluída na parte principal do relatório descritivo.

Identificadores Numéricos Obrigatórios

| identificador numérico | Descrição do identificador numérico | Comentário |
|------------------------|-------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <110> | Nome do requerente | Quando o nome do requerente estiver escrito em caracteres diferentes dos que compõem o alfabeto latino, também deverá ser indicado em caracteres do alfabeto latino, seja como simples transliteração ou mediante a sua tradução para o inglês; havendo mais de um requerente, listar um nome por linha. |
| <120> | Título da invenção ⁹ | |
| <130> | Número de referência do pedido | Obrigatório somente nas condições especificadas pelo item 5.4. |
| <140> | Pedido de patente em trâmite | Obrigatório somente nas condições especificadas pelo item 5.5. |
| <141> | Data de depósito do pedido de patente em trâmite | Obrigatório somente nas condições especificadas pelo item 5.5. |
| <150> | Pedido de patente anterior (prioridade) | Obrigatório somente na condição especificada pelo item 5.6. |
| <151> | Data de depósito do pedido de patente anterior (prioridade) | Obrigatório somente na condição especificada pelo item 5.6. |
| <160> | Número de SEQ ID NOs | Inclui o número total de SEQ ID NOs compreendidas na listagem de sequências |
| <210> | Informação sobre a SEQ ID NO: # | A resposta deverá estar composta por um número inteiro que represente a SEQ ID NO mostrada |
| <211> | Comprimento | Comprimento da sequência expressa em número de pares de bases ou de resíduos de aminoácidos |

⁹ em língua vernácula.

| | | |
|-------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <212> | Tipo | Tipo de molécula DNA/RNA/PROTEÍNA que é mostrada na SEQ ID NO: #, ou seja, DNA, RNA ou PRT (proteína); se a sequência de nucleotídeos contiver fragmentos de DNA e de RNA, o tipo será "DNA"; além disso, a molécula combinada de DNA/RNA também deverá ser objeto de descrição na seção de características <220> a <223> |
| <213> | Organismo | Gênero e espécie (ou seja, o nome científico) ou "Sequência Artificial" (<i>Artificial Sequence</i>) ou "Desconhecido" (<i>Unknown</i>); adicionalmente, a sequência artificial ou o organismo desconhecido deverá ser também objeto de descrição na seção de características <220> a <223> |
| <220> | Característica | Obrigatório somente nas condições especificadas pelos itens 5.2 e 5.3. Caso contrário, deixe em branco. |
| <221> | Nome/chave | Obrigatório somente na condição especificada pelo item 5.2. |
| <222> | Localização | Obrigatório somente na condição especificada pelo item 5.2. |
| <223> | Outras informações | Obrigatório somente nas condições especificadas pelos itens 5.2 e 5.3. |
| <400> | Sequência | O elemento SEQ ID NO: deve ir depois do identificador numérico e deve figurar na linha anterior a sequência de fato (ver exemplo) |

Tabela 1: Listagem de nucleotídeos

| Símbolo | Significado | Origem da designação |
|---------|------------------|-----------------------------------------------------------------------------------|
| a | a | adenina |
| g | g | guanina |
| c | c | citosina |
| t | t | timina |
| u | u | uracila |
| r | g ou a | purína |
| y | t/u ou c | pirimidina (<i>pyrimidine</i>) |
| m | a ou c | amino |
| k | g ou t/u | ceto (<i>keto</i>) |
| s | g ou c | interações fortes (<i>strong interactions</i>) 3 (três) pontes de hidrogênio |
| w | a ou t/u | interações fracas (<i>weak interactions</i>) 2 (duas) pontes de hidrogênio |
| b | g ou c ou t/u | que não seja a |
| d | a ou g ou t/u | que não seja c |
| h | a ou c ou t/u | que não seja g |
| v | a ou g ou c | que não seja t e nem u |

| | | |
|---|----------------------------------------------|----------------|
| n | a ou g ou c ou t/u, desconhecido ou outro | qualquer (any) |
|---|----------------------------------------------|----------------|

Tabela 2: Listagem de nucleotídeos modificados

| Símbolo | Significado |
|----------|---------------------------------------------------------------------|
| ac4c | 4-acetilcitidina |
| chm5u | 5-(carboxihidroximetil)uridina |
| cm | 2'-O-metilcitidina |
| cmnm5s2u | 5-carboximetilaminometil-2-tiouridina |
| cmnm5u | 5-carboximetilaminometiluridina |
| d | dihidrouridina |
| fm | 2'-O-metilpseudouridina |
| gal q | beta, D-galactosilqueosine |
| gm | 2'-O-metilguanosina |
| i | Inosina |
| i6a | N6-isopenteniladenosina |
| m1a | 1-metiladenosina |
| m1f | 1-metilpseudouridina |
| m1g | 1-metilguanosina |
| m1i | 1-metilinosina |
| m22g | 2,2-dimetilguanosina |
| m2a | 2-metiladenosina |
| m2g | 2-metilguanosina |
| m3c | 3-metilcitidina |
| m5c | 5-metilcitidina |
| m6a | N6-metiladenosina |
| m7g | 7-metilguanosina |
| mam5u | 5-metilaminometiluridina |
| mam5s2u | 5-metoxiaminometil-2-tiouridina |
| man q | beta, D-manosilqueosina |
| mcm5s2u | 5-metoxicarbonilmetyl-2-tiouridina |
| mcm5u | 5-metoxicarbonilmetyluridina |
| mo5u | 5-metoxiuridina |
| ms2i6a | 2-metiltio-N6-isopenteniladenosina |
| ms2t6a | N-((9-beta-D-ribofuranosil-2-metiltiopurina-6-il)carbamoil)treonina |
| mt6a | N-((9-beta-D-ribofuranosilpurina-6-il)N-metilcarbamoil)treonina |
| mv | 5-metoxicarbonilmetoxiuridina |
| o5u | uridina-5-ácido oxiacético |
| osyw | wybutoxosina |
| p | pseudouridina |
| q | queosina |
| s2c | 2-tiocitidina |
| s2t | 5-metil-2-tiouridina |
| s2u | 2-tiouridina |
| s4u | 4-tiouridina |
| t | 5-metiluridina |
| t6a | N-((9-beta-D-ribofuranosilpurina-6-il)-carbamoil)treonina |
| tm | 2'-O-metil-5-metiluridina |
| um | 2'-O-metiluridina |

| | |
|----|----------------------------------------------|
| yw | wybutosina |
| x | 3-(3-amino-3-carboxi-propil)uridina, (acp3)u |

Tabela 3: Listagem de aminoácidos

| Símbolo | Significado |
|---------|-----------------------|
| Ala | Alanina |
| Cys | Cisteína |
| Asp | Ácido Aspártico |
| Glu | Ácido Glutâmico |
| Phe | Fenilalanina |
| Gly | Glicina |
| His | Histidina |
| Ile | Isoleucina |
| Lys | Lisina |
| Leu | Leucina |
| Met | Metionina |
| Asn | Asparagina |
| Pro | Prolina |
| Gln | Glutamina |
| Arg | Arginina |
| Ser | Serina |
| Thr | Treonina |
| Val | Valina |
| Trp | Triptofano |
| Tyr | Tirosina |
| Asx | Asp ou Asn |
| Glx | Glu ou Gln |
| Xaa | desconhecido ou outro |

Tabela 4: Listagem de aminoácidos modificados ou pouco usuais

| Símbolo | Significado |
|---------|-------------------------------------------|
| Aad | Ácido 2-aminoadípico |
| bAad | Ácido 3-aminoadípico |
| bAla | beta-Alanina, ácido beta-aminopropiônico |
| Abu | Ácido 2-aminobutírico |
| 4Abu | Ácido 4-aminobutírico, ácido piperidínico |
| Acp | Ácido 6-aminocaprólico |
| Ahe | Ácido 2-aminoheptanóico |
| Aib | Ácido 2-aminoisobutírico |
| bAib | Ácido 3-aminoisobutírico |
| Apm | Ácido 2-aminopimélico |
| Dbu | Ácido 2,4 diaminobutírico |

| | |
|-------|-----------------------------|
| Des | Desmosina |
| Dpm | Ácido 2,2'-diaminopimélico |
| Dpr | Ácido 2,3-diaminopropiônico |
| EtGly | N-etilglicina |
| EtAsn | N-etylaspargina |
| Hyl | Hidroxilisina |
| aHyl | alo-Hidroxilisina |
| 3Hyp | 3-Hidroxiprolina |
| 4Hyp | 4-Hidroxiprolina |
| Ide | Isodesmosina |
| alle | alo-Isoleucina |
| MeGly | N-metilglicina, sarcosina |
| Melle | N-metilisoleucina |
| MeLys | 6-N-metillisina |
| MeVal | N-metilvalina |
| Nva | Norvalina |
| Nle | Norleucina |
| Orn | Ornitina |

Tabela 5: Listagem das Chaves de Caracterização de Sequências de Nucleotídeos

| Chave ¹⁰ | Descrição |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| allele (alelo) | Existência de indivíduos ou estirpes relacionadas que contém formas estáveis e diferentes do mesmo gene e que diferem da sequência apresentada nesta localização (e talvez em outras) |
| attenuator (atenuador) | 1) região do DNA onde ocorre controle da terminação da transcrição que controla a expressão de certos operadores bacterianos; 2) segmento de sequência localizado entre o promotor e o primeiro gene estrutural que causa terminação parcial da transcrição |
| C_region (região-C) | Região constante das cadeias leve e pesada das imunoglobulinas e das cadeias alfa, beta e gama do receptor de linfócitos T; inclui um ou mais exons, dependendo da cadeia em particular |

¹⁰ Entre parênteses é apresentada a tradução mais usualmente empregada na língua vernácula do correspondente termo científico (vide item 1.3).

| | |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CAAT_signal (sinal CAAT) | Região CAAT box; parte de uma sequência conservada situada à cerca de 75 pares de bases a montante do local de iniciação das unidades de transcrição eucarióticas e que pode estar envolvida na ligação da RNA polimerase sequência consenso= GG (C ou T) CAATCT |
| CDS (sequência codificadora) | Sequência codificadora (<i>coding sequence</i>); sequência de nucleotídeos que se corresponde com a sequência de aminoácidos de uma proteína (a localização inclui o códon de terminação); contém a tradução conceptual dos aminoácidos |
| conflict (conflito) | Determinações independentes da "mesma" sequência diferem neste local ou nesta região |
| D-loop (alça de deslocamento) | Alça de deslocamento (<i>Displacement loop</i>); região do DNA mitocondrial na qual uma sequência curta de RNA fita simples é pareada com uma das fitas do DNA, deslocando nesta região a outra fita de DNA pareada; também usada para descrever o deslocamento de uma região de fita simples em um DNA duplex por um invasor fita simples, na reação catalisada pela proteína RecA |
| D-segment (segmento de diversidade) | Segmento de diversidade (<i>Diversity segment</i>) da cadeia pesada das imunoglobulinas e da cadeia pesada do receptor de linfócitos T |
| enhancer (acentuador) | <i>Enhancer</i> ou acentuador é uma sequência que aumenta a utilização de (certos) promotores eucarióticos situados na mesma fita de DNA (efeito em <i>cis</i>) e cuja ação pode efetuar-se com independência da orientação e da localização (5' ou 3') em relação ao promotor |
| exon (éxon) | Região do genoma que codifica para a porção do RNA mensageiro processado (<i>spliced mRNA</i>); pode conter a região 5'UTR, todas as sequências codificadoras (CDS) e a região 3'UTR |
| GC_signal (sinal GC) | Região GC box; região conservada rica em GC e localizada antes do ponto de iniciação das unidades de transcrição eucarióticas e que pode adotar a forma de múltiplas cópias e produzir-se em ambos os sentidos (5' ou 3') sequência consenso= GGGCGG |
| gene (gene) | Região de interesse biológico identificada como sendo um gene e para a qual foi designado um nome; ácido nucléico codificador |
| iDNA (DNA de intervenção) | DNA de intervenção (<i>intervening DNA</i>); DNA que é eliminado em diferentes tipos de recombinação |

| | |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| intron (ítron) | Segmento de DNA que é transcrito, porém logo removido da nova molécula de RNA pelo processo de <i>splicing</i> do RNA, ocasionando junção dos exons que flanqueiam os ítrons |
| J_segment (segmento de ligação) | Segmento de ligação (<i>Joining segment</i>) das cadeias leve e pesada das imunoglobulinas e das cadeias alfa, beta e gama do receptor de célula T |
| LTR (sequências repetitivas longas) | LTGs (<i>Long Terminal Repeat</i>) são sequências repetitivas longas encontradas em cada extremidade (5' e 3") de uma sequência tal como a que é tipicamente encontrada nos retrovírus |
| mat_peptide (sequência codificadora de um peptídeo) | Sequência codificadora de um peptídeo ou de uma proteína madura; sequência codificadora do peptídeo ou da proteína em sua condição madura ou final, seguida de modificação pós-tradução; a localização não inclui o códon de terminação (diferentemente da CDS correspondente) |
| misc_binding ¹¹ | Região em um ácido nucléico que se liga covalentemente ou não com outra molécula e que não pode ser descrito por qualquer outra chave de ligação (primer_bind ou protein_bind) |
| misc_difference ¹² | A sequência caracterizada é diferente nesta posição, daquela apresentada na entrada e não pode ser descrita por nenhuma outra chave de diferença (conflict, unsure, old_sequence, mutation, variation, allele ou modified_base) |
| misc_feature ¹⁰ | Região de interesse biológico que não pode ser descrita por nenhuma outra chave de característica; uma característica nova ou pouco comum |
| misc_recomb ¹⁰ | Sítio de qualquer recombinação generalizada, sitio-específica ou replicativa, por onde se produz a excisão e ligação de DNA duplex e que não pode ser descrita por nenhuma outra chave de recombinação (iDNA ou virion) e nem por qualificadores da chave de origem (/insertion_seq, /transposon, /proviral) |
| misc_RNA ¹⁰ | Qualquer porção transcrita ou RNA que não pode ser definida por nenhuma outra chave de RNA (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA ou snRNA) |

¹¹

Sem termo correspondente na língua vernácula

¹²

Sem termo correspondente na língua vernácula

| | |
|-------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| misc_signal ¹⁰ | Qualquer região que contenha um sinal que controla ou modifica uma função ou expressão de um gene, que não pode ser descrito por nenhuma outra chave de sinal (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator ou rep_origin) |
| misc_structure ¹⁰ | Qualquer conformação ou estrutura secundária ou terciária que não pode ser descrita por nenhuma outra chave de estrutura (stem_loop ou D-loop) |
| modified_base (nucleotídeo modificado) | O nucleotídeo indicado é um nucleotídeo modificado e deve ser substituído pela molécula indicada (que figura no valor qualificador de mod_base) |
| mRNA (RNA mensageiro) | RNA mensageiro; inclui a região 5' não traduzida (5'UTR), a sequência codificadora (CDS, exon) e a região 3' não traduzida (3'UTR) |
| mutation (mutação) | Uma estirpe relacionada apresenta uma alteração brusca e não transmissível na sequência, nesta localização |
| região N (N_region) | Região de inserção de nucleotídeos adicionais entre os segmentos reordenados das imunoglobulinas |
| old_sequence (prévia sequência) | A sequência apresentada é uma versão revisada de uma prévia sequência nesta localização |
| polyA_signal (sinal de poliadenilação) | Região indispensável de reconhecimento para clivagem por uma endonuclease seguida por poliadenilação de uma porção transcrita de RNA sequência consenso= AATAAA |
| polyA_site (sítio de poliadenilação) | Região de um transcrito de RNA no qual se adicionam resíduos de adenina por poliadenilação pós-transcricional |
| precursor_RNA (RNA precursor) | Precursor de RNA, qualquer RNA imaturo; pode incluir a região cortada em 5' (5'clip), a região 5' não traduzida (5'UTR), as sequências codificadoras (CDS, exon), as sequências intervenientes (intron), a região 3' não traduzida (3'UTR) e a região cortada em 3' (3'clip) |
| prim_transcript (transcrito primário) | Transcrito primário (inicial, não processado); inclui a região cortada em 5' (5'clip), a região 5' não traduzida (5'UTR), as sequências codificadoras (CDS, exon), as sequências intervenientes (intron), a região 3' não traduzida (3'UTR) e a região cortada em 3' (3'clip) |

| | |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| primer_bind (região de ligação de um iniciador) | Região de ligação não covalente de um iniciador (<i>primer</i>) na iniciação da replicação, da transcrição ou da transcrição reversa; inclui as regiões para iniciadores sintéticos, por exemplo, os que são usados na reação em cadeia da polimerase (PCR) |
| promoter (promotor) | Região de uma molécula de DNA na qual se liga a RNA polimerase para iniciar a transcrição |
| protein_bind (ligação de proteína) | Região de ligação não covalente de proteínas em um ácido nucléico |
| RBS (sítio de ligação de ribossomo) | Região de ligação do ribossomo (<i>ribosome binding site</i>) |
| repeat_region (região repetitiva) | Região do genoma que contém unidades de repetição |
| repeat_unit (unidade de repetição) | Elemento (unidade de repetição) que se repete na <i>repeat_region</i> |
| rep_origin (origem de replicação) | Origem de replicação; região onde se inicia a duplicação de um ácido nucléico para obter duas cópias idênticas |
| rRNA (rRNA) | RNA ribossomal maduro; RNA que compõe a partícula ribonucleoprotéica (ribossomo) que sintetiza proteínas a partir de aminoácidos |
| S_region (região S) | Região de mudança (<i>switch region</i>) das cadeias pesadas das imunoglobulinas; envolvida no rearranjo do DNA que codifica para a cadeia pesada levando à expressão de uma classe diferente de imunoglobulina por um mesmo linfócito B |
| satellite (satélite) | Múltiplas repetições em <i>tander</i> (idênticas ou parecidas) de uma unidade de repetição básica curta; muitas delas têm uma composição de bases ou uma outra propriedade diferente do genoma em geral, o que permite separá-las do resto do DNA genômico (banda principal) |
| scRNA (RNA citoplasmático pequeno) | RNA citoplasmático de tamanho pequeno (<i>small cytoplasmic RNA</i>); uma das diversas pequenas moléculas de RNA presentes no citoplasma e (algumas vezes) no núcleo de uma célula eucariótica |

| | |
|-------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sig_peptide (peptídeo sinal) | Sequência codificadora para um peptídeo sinal; sequência codificadora do domínio amino-terminal de uma proteína secretada; este domínio está envolvido na integração do polipeptídeo nascente na membrana; sequência leader |
| snRNA (RNA nuclear pequeno) | RNA nuclear de tamanho pequeno (<i>small nuclear RNA</i>); qualquer uma das muitas espécies de RNA pequeno que estão confinadas no núcleo; vários dos snRNA estão envolvidos em <i>splicing</i> ou em outras reações de processamento de RNA |
| source (fonte) | Identifica a fonte biológica do intervalo de sequência especificamente indicado; esta chave é obrigatória; cada entrada deve estar composta por no mínimo, de uma chave única de fonte englobando a sequência inteira; é permitido o uso de mais de uma chave de fonte por sequência |
| stem_loop (alça em forma de grampo) | Alça em forma de grampo (<i>hairpin</i>); região de dupla hélice formada pelo pareamento de bases entre sequências complementares adjacentes (invertidas) que pertencem a uma mesma fita de RNA ou de DNA (pareamento intramolecular) |
| STS (região marcadora de DNA) | Regiões marcadoras na sequência (<i>Sequence Tagged Site</i>); trata-se de sequências curtas de DNA que ocorrem uma única vez no genoma humano e cuja posição exata e ordem de bases, uma vez conhecidas, identificam um local no genoma, sendo detectadas por PCR; o mapa de uma região do genoma pode efetuar-se determinando a ordem de uma série de STS |
| TATA_signal (sinal TATA) | TATA-box; Goldberg-Hogness box; é um heptâmero conservado rico em A•T, situado a cerca de 25 pares de bases antes do sítio de iniciação de cada unidade transcrita pela RNA polimerase II das células eucarióticas; pode estar envolvido no posicionamento da enzima para a iniciação correta da transcrição sequência consensual= TATA(A ou T)A(A ou T) |
| terminator (terminador) | <i>Terminator</i> ou terminador; sequência de DNA localizada no final do transcrito ou adjacente a um promotor e que faz com que a RNA polimerase termine a transcrição; também pode ser o sítio de ligação da proteína represora |
| transit_peptide (peptídeo de trânsito) | Sequência codificadora para um peptídeo de trânsito; sequência codificadora do domínio amino-terminal de uma proteína de organela codificada no núcleo; este elemento está envolvido na importação pós-tradução da proteína para dentro da organela |

| | |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| tRNA (RNA transportador) | RNA de transferência maduro, RNA de tamanho pequeno (75-85 bases) que media a tradução de uma sequência de ácido nucléico em uma sequência de aminoácidos |
| unsure (incerto) | O autor não está seguro sobre a exatidão da sequência nesta região |
| V_region (região V) | Região variável das cadeias leve e pesada das imunoglobulinas e das cadeias alfa, beta e gama do receptor de linfócitos T; codifica para a região variável na extremidade amino-terminal; pode estar composta por: V_segment, D_segment, N_region e J_segment |
| V_segment (segmento V) | Segmento variável das cadeias leve e pesada das imunoglobulinas e das cadeias alfa, beta e gama do receptor de linfócitos T; codifica para a maior parte da região variável (V_region) e para os últimos aminoácidos do peptídeo líder (<i>leader peptide</i>) |
| variation (variante) | Existência de uma estirpe relacionada que contém mutações estáveis do mesmo gene (por exemplo, RFLP, polimorfismos, etc) e que diferem da sequência apresentada nesta localização (e talvez em outras) |
| 3'clip ¹³ | Região na extremidade 3' de um RNA precursor que é cortado durante processamento |
| 3'UTR ¹⁴ | Região na extremidade 3' (posterior ao códon de terminação) de um RNA maduro que não se traduz em proteína |
| 5'clip ¹⁴ | Região na extremidade 5' de um RNA precursor que é cortado no processamento |
| 5'UTR ¹⁴ | Região na extremidade 5' (anterior ao códon de terminação) de um RNA maduro que não se traduz em proteína |
| -10_signal (sinal -10) | Sequência -10 (<i>pribnow box</i>); sequência conservada centrada aproximadamente 10 pares de bases antes do sítio de início da transcrição de um gene bacteriano e que pode participar na ligação da RNA polimerase sequência consenso= TATAaT |
| -35_signal (sinal -35) | Sequência -35; sequência centrada aproximadamente 35 pares de bases antes do sítio de início da transcrição de um gene bacteriano sequência consenso= TTGACa ou TGTTGACA |

¹³

Sem termo correspondente na língua vernácula

Tabela 6: Listagem de Chaves de Caracterização de Sequências de Aminoácidos

| Chave ¹⁴ | Descrição |
|----------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| CONFLICT (CONFLITO) | Diferentes documentos reportam diferentes sequências |
| VARIANT (VARIANTE) | Os autores assinalam que existem variações da seqüência |
| VARSPlicing (VARIANTE DE EDIÇÃO) | Descrição das variações da sequência produzidas por um <i>splicing</i> alternativo |
| MUTAGEN (SÍTIO ALTERADO POR MUTAÇÃO) | Sítio que foi experimentalmente alterado |
| MOD_RES (RESÍDUO PÓS-MODIFICADO) | Modificação pós-tradução de um resíduo |
| ACETYLATION (ACETILAÇÃO) | Acetilação na extremidade amino-terminal ou outra |
| AMIDATION (AMIDAÇÃO) | Amidação geralmente na extremidade carboxi-terminal de um peptídeo maduro e ativo |
| BLOCKED (SÍTIO BLOQUEADO) | Grupo de bloqueio indeterminado na extremidade amino-terminal ou carboxi-terminal |
| FORMYLATION (FORMILAÇÃO) | Formilação da metionina da extremidade amino-terminal |
| GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION (HIDROXILAÇÃO ÁCIDO GAMA-CARBOXIGLUTÂMICO) | da asparagina, do ácido aspártico, da prolina ou da lisina |
| METHYLATION (METILAÇÃO) | Metilação geralmente da lisina ou da arginina |
| PHOSPHORYLATION (FOSFORILAÇÃO) | Fosforilação da serina, da treonina, da tirosina, do ácido aspártico ou da histidina |
| PYRROLIDONE CARBOXYLIC ACID (ÁCIDO CARBOXI PIRROLIDÔNICO) | Glutamato amino-terminal que formou uma lactama cíclica interna |

¹⁴ Entre parênteses é apresentada a tradução mais usualmente empregada na língua vernácula do correspondente termo científico (vide item 1.3).

| | |
|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SULFATION (SULFATAÇÃO) | Sulfatação geralmente da tirosina |
| LIPID (LIPÍDIO) | Ligaçāo covalente de um fragmento lipídico |
| MYRISTATE (MIRISTATO) | Grupo miristato unido por uma ligação amida a um resíduo de glicina da extremidade amino-terminal da forma madura de uma proteína ou de um resíduo interno de lisina |
| PALMITATE (PALMITATO) | Grupo palmitato unido por uma ligação tioéter a um resíduo de cisteína ou por uma ligação éster a um resíduo de serina ou de treonina |
| FARNESYL (FARNESIL) | Grupo farnesil ligado por uma ligação tioéter a um resíduo de cisteína |
| GERANYL-GERANYL (GERANIL-GERANIL) | Grupo geranil-geranil ligado por uma ligação tioéter a um resíduo cisteína |
| GPI-ANCHOR (GRUPO GLICOSIL- FOSFATIDILINOSITOL ANCORADO) | Grupo glicosil-fosfatidilinositol (GPI) unido a um grupo alfa-carboxila do resíduo carboxi-terminal da forma madura de uma proteína |
| N-ACYL DIGLYCERIDE (N-ACIL DICLICERÍDEO) | Cisteína amino-terminal da forma madura de uma lipoproteína de procarioto unida por uma ligação amida a um ácido graxo e um grupo gliceril, na qual dois ácidos graxos estão unidos por ligação éster |
| DISULFID (PONTE DISSULFETO) | Ponte dissulfeto; os extremos "DE" ("FROM") e "PARA" ("TO") representam os dois resíduos que estão ligados por uma ponte dissulfeto intra-cadeia peptídica; se os extremos "DE" ("FROM") e "PARA" ("TO") são idênticos, a ponte dissulfeto é uma ligação inter-cadeia peptídica e o campo descritivo indica a natureza das ligações cruzadas (cross-link) |
| THIOEST (LIGAÇÃO TIOÉSTER) | Ligaçāo tioéster; os extremos "DE" ("FROM") e "PARA" ("TO") representam os dois resíduos que estão unidos pela ligação tioéster |
| THIOETH (LIGAÇÃO TIOÉTER) | Ligaçāo tioéster; os extremos "DE" ("FROM") e "PARA" ("TO") representam os dois resíduos que estão unidos pela ligação tioéster |
| CARBOHYD (SÍTIO DE GLICOSILAÇĀO) | Sítio de glicosilação; a natureza do carboidrato (se conhecido) está indicada no campo descritivo |
| METAL (SÍTIO DE LIGAÇÃO DE METAL) | Sítio de ligação para um íon de metal; no campo descritivo é indicada a natureza do metal |
| BINDING (SÍTIO DE LIGAÇÃO) | Sítio de ligação para qualquer grupo químico (coenzima, grupo prostético, etc.); no campo descritivo é indicada a natureza química do grupo |

| | |
|------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| SIGNAL (SINAL) | Extensão de uma sequência-sinal (pré-peptídeo) |
| TRANSIT (TRÂNSITO) | Extensão de um peptídeo de trânsito (mitocondrial, cloroplástico ou destinado para microssoma) |
| PROPEP (PROPEP) | Extensão de um pró-peptídeo |
| CHAIN (CADEIA) | Extensão da cadeia polipeptídica na proteína madura |
| PEPTIDE (PEPTÍDEO) | Extensão de um peptídeo ativo liberado |
| DOMAIN (DOMÍNIO) | Extensão de um domínio de interesse na sequência; no campo descritivo é indicada a natureza deste domínio |
| CA_BIND (SÍTIO DE LIGAÇÃO DE CÁLCIO) | Extensão de uma região de ligação de cálcio |
| TRANSMEM (TRANSMEMBRANA) | Extensão de uma região transmembrana |
| ZN_FING (MOTIVO DEDO DE ZINCO) | Extensão de uma região contendo o motivo dedo de zinco (<i>zinc finger</i>) |
| SIMILAR (SIMILAR) | Extensão da similaridade de uma região com uma outra sequência protéica; no campo descritivo são indicadas informações detalhadas sobre esta sequência |
| REPEAT (SEQUÊNCIA INTERNA REPETITIVA) | Extensão de uma sequência interna repetitiva |
| HELIX (HÉLICE) | Estrutura secundária: Hélices, por exemplo, a alfa-hélice, a hélice 310 ou a hélice Pi |
| STRAND (FITA) | Estrutura secundária: folha beta (folha-b), por exemplo, folha beta-pregueada unida por pontes de hidrogênio, o resíduo isolado em uma ponte beta |
| TURN (VOLTA) | Estrutura secundária: voltas (<i>turns</i>), por exemplo, voltas mantidas por pontes de hidrogênio (voltas de 3, 4 ou 5 resíduos de aminoácidos) |
| ACT_SITE (SÍTIO ATIVO) | Aminoácidos envolvidos na atividade de uma enzima |
| SITE (SÍTIO) | Qualquer outro sítio de interesse na sequência |

| | |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| INIT_MET (INICIA COM METIONINA) | A sequência começa com uma metionina de iniciação |
| NON_TER (NÃO TERMINAL) | O resíduo em uma extremidade da sequência não é o resíduo terminal; se aplicado à posição 1, significa que a primeira posição não é a posição amino-terminal da molécula completa; se aplicado para a última posição, significa que esta posição não é a posição carboxi-terminal da molécula completa; não há nenhum campo descritivo para esta chave |
| NON_CONS (NÃO CONSECUTIVOS) | Resíduos não consecutivos; indica que dois resíduos de uma sequência não são consecutivos e que existem vários resíduos não sequenciados entre eles |
| UNSURE (INCERTO) | Zonas de incertezas na sequência; usado para descrever as regiões da sequência para as quais os autores não estão certos de sua definição |

8. Dos elementos de dados não obrigatórios:

8.1. Todos os elementos de dados citados a seguir são facultativos de comporem a "Listagem de Sequências":

| | |
|-------|-------------------------------------------------------------------------------------------------------------|
| <170> | Programa de computador usado para gerar a listagem de seqüências |
| <300> | Informações sobre publicação; havendo várias publicações, repita a seção para cada publicação relevante |
| <301> | Autores, especifique um nome por linha, preferencialmente no formato: sobrenome, outros nomes e/ou iniciais |
| <302> | Título da publicação |
| <303> | Nome do periódico no qual se publicaram os dados |
| <304> | Volume do periódico no qual se publicaram os dados |
| <305> | Número do periódico no qual se publicaram os dados |
| <306> | Número das páginas do periódico no qual se publicaram os dados |
| <307> | Data do periódico no qual se publicaram os dados; usar formato Dia/Mês/Ano |
| <308> | Número de acesso assinalado pela base de dados, incluindo o nome da base de dados |
| <309> | Data de entrada na base de dados (dia/mês/ano) |
| <310> | Número do documento de patente, unicamente para as patentes citadas |
| <311> | Data de submissão do documento de patente, unicamente para as patentes citadas (dia/mês/ano) |
| <312> | Data de publicação do documento de patente; unicamente para as patentes citadas (dia/mês/ano) |
| <313> | Resíduos relevantes na SEQ ID NO: #: DE (from) _ PARA (to) |

LISTAGEM DE SEQUÊNCIAS (EXEMPLO)

Exemplo de Listagem de Sequências

<110> . . .

<120> Example of a Sequence Listing

<130> 01-00001

<140> PCT/EP98/00001
<141> 31-12-1998

<150> US 08/999,999
<151> 15-10-1997

<160> 4

<170> PatentIn ver. 3.5

<210> 1
<211> 389
<212> DNA
<213> Paramecium sp.

<220>
<221> CDS
<222> (279) . . . (389)

<300>
<301> Doe, Richard
<302> Isolamento e Caracterização de um Gene Codificador de uma
Protease de Paramecium sp.
<303> Journal of Genes
<304> 1
<305> 4
<306> 1-7
<307> 31-06-1988
<308> 123456
<309> 31-06-1988

<400> 1
agctgttagtc attcctgtgt cctcttctct ctgggcttct caccctgcta atcagatctc 60
agggagagtg tcttgaccct cctctgcctt tgcagcttca caggcaggca ggcaggcagc 120
tcatgtggca attgctggca gtgccacagg ctttcagcc aggcttaggg tgggttcogc 180
cgccccggcg ggccccctct cgegctctc tggcgcctct ctctcgctc cctctegtc 240

ggacctgatt aggtgagcag gaggaggggg cagttgc atg
Met 1 gtt tca atg ttc agc 296
5

ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgt ttg ttc
Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
caa Gln 344

tgt ccc aaa gtc ctc ccc tgt cac tca tca ctg cag ccg aat ctt
Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu
25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110
389

<210> 2
<211> 37
<212> PRT
<213> Paramecium sp.

<400> 2
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu
1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Leu Gln Pro Asn Leu

Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Leu Gln Pro Asn Leu

<210> 3
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Peptídeo desenhado baseado em tamanho e polaridade para atuar como
um ligante entre as cadeias alfa e beta da proteína XYZ

<400> 3
Met Val Asn Leu Glu Pro Met His Thr Glu Ile
1 5 10

<210> 4
<400> 4
000