

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2009.04.24	(73) Titular(es): DYAX CORP.	
(30) Prioridade(s): 2008.04.24 US 47529 P	300 TECHNOLOGY SQUARE, 8TH FLOOR	
(43) Data de publicação do pedido: 2011.02.09	CAMBRIDGE, MA 02139	US
(45) Data e BPI da concessão: 2014.10.22 025/2015	(72) Inventor(es):	
	(74) Mandatário: NUNO MIGUEL OLIVEIRA LOURENÇO	
	RUA CASTILHO, Nº 50 - 9º 1269-163 LISBOA	PT

(54) Epígrafe: **BIBLIOTECAS DE PACOTES GENÉTICOS COMPREENDENDO DESENHOS DE NOVAS CDR1, CDR2, E CDR3 DE HC E NOVAS CDR1, CR2, E CDR3 DE LC**

(57) Resumo:

SÃO PROPORCIONADAS COMPOSIÇÕES E MÉTODOS PARA PREPARAÇÃO E IDENTIFICAÇÃO DE ANTICORPOS TENDO CDR3S QUE VARIAM EM SEQUÊNCIA E EM COMPRIMENTO DE MUITO CURTAS A MUITO LONGAS QUE, EM CERTAS FORMAS DE REALIZAÇÃO, PODEM LIGAR-SE A UMA FRAÇÃO CARBOIDRATO OU AO LOCAL ATIVO DE UMA ENZIMA. SÃO TAMBÉM PROPORCIONADAS BIBLIOTECAS CODIFICANDO ANTICORPOS COM AS CDR3S. AS BIBLIOTECAS PODEM SER PROPORCIONADAS POR MODIFICAÇÃO DE UMA BIBLIOTECA DE ÁCIDOS NUCLEICOS PRÉ-EXISTENTE.

RESUMO

"BIBLIOTECAS DE PACOTES GENÉTICOS COMPREENDENDO DESENHOS DE NOVAS CDR1, CDR2, E CDR3 DE HC E NOVAS CDR1, CR2, E CDR3 DE LC"

São proporcionadas composições e métodos para preparação e identificação de anticorpos tendo CDR3s que variam em sequência e em comprimento de muito curtas a muito longas que, em certas formas de realização, podem ligar-se a uma fração carboidrato ou ao local ativo de uma enzima. São também proporcionadas bibliotecas codificando anticorpos com as CDR3s. As bibliotecas podem ser proporcionadas por modificação de uma biblioteca de ácidos nucleicos pré-existente.

DESCRIÇÃO

"BIBLIOTECAS DE PACOTES GENÉTICOS COMPREENDENDO DESENHOS DE NOVAS CDR1, CDR2, E CDR3 DE HC E NOVAS CDR1, CR2, E CDR3 DE LC"

Referência cruzada a pedidos relacionados

Este pedido reivindica prioridade em relação ao Pedido dos E.U.A. No. em Série 61/047,529, depositado a 24 de abril, 2008. A divulgação do pedido prévio é considerada parte da divulgação deste pedido.

Antecedentes

É agora prática comum na técnica preparar bibliotecas de pacotes genéticos que exibem individualmente, exibem e expressam, ou compreendem um membro de uma família diversa de péptidos, polipéptidos ou proteínas e exibem coletivamente, exibem e expressam, ou compreendem pelo menos uma porção da diversidade de aminoácidos da família. Em muitas bibliotecas comuns, os péptidos, polipéptidos ou proteínas estão relacionados com anticorpos (p.ex., Fv de cadeia única (scFv), Fv, Fab, anticorpos inteiros ou minicorpos (*i.e.*, dímeros que consistem em V_H ligada a V_L)). Frequentemente compreendem uma ou mais das CDRs e regiões de *framework* das cadeias pesadas e leves de anticorpos humanos.

Bibliotecas de péptidos, polipéptidos ou proteínas têm sido produzidas de várias maneiras. Ver, p.ex., Knappik *et al.*, *J. Mol. Biol.*, 296, pp. 57-86 (2000), que é incorporado aqui por referência. Um método é capturar a diversidade de dadores nativos, ingênuos ou imunizados. Outra maneira é gerar bibliotecas tendo diversidade sintética. Um terceiro método é uma combinação dos primeiros dois. Tipicamente, a

diversidade produzida por estes métodos está limitada a diversidade de sequências, *i.e.*, cada membro da biblioteca tem o mesmo comprimento mas difere dos outros membros da família por ter diferentes aminoácidos ou variação a uma dada posição na cadeia de péptidos, polipéptidos ou proteínas. Péptidos, polipéptidos ou proteínas naturalmente diversos, no entanto, não estão limitados a diversidade somente nas suas sequências de aminoácidos. Por exemplo, os anticorpos humanos não estão limitados a diversidade de sequências nos seus aminoácidos, divergem também nos comprimentos das suas cadeias de aminoácidos.

Hoet *et al.*, *Nature Biotechnology*, Vol. 23 (3), 1 de março de 2005, páginas 344-348, descrevem a geração de anticorpos humanos de elevada afinidade por combinação de diversidade de regiões determinadoras da complementaridade derivadas de dadores e sintéticas.

Para anticorpos, a diversidade de cadeias pesadas em comprimento ocorre, por exemplo, durante rearranjos das regiões variáveis. Ver, *p.ex.*, Corbett *et al.*, *J. Mol. Biol.*, 270, pp. 587-97 (1997). A união de genes V a genes J, por exemplo, resulta na inclusão de um segmento D reconhecível Em CDR3 em cerca de metade das sequências de anticorpo de cadeia pesada, criando deste modo regiões codificando comprimentos variáveis de aminoácidos. Os segmentos D são mais comuns em anticorpos tendo CDR3s de HC longas. O seguinte pode também ocorrer durante união de segmentos de gene de anticorpos: (i) a extremidade do gene V pode ter zero a várias bases eliminadas ou mudadas; (ii) a extremidade do segmento D pode ter zero a muitas bases removidas ou mudadas; (iii) um número de bases aproximadamente aleatórias pode ser inserido entre V e D ou

entre D e J; e (iv) a extremidade 5' de J pode ser modificada para remover ou para mudar várias bases. Estes rearranjos resultam em anticorpos que são diversos na sequência de aminoácidos e no comprimento. CDR3s de HC de diferentes comprimentos podem dobrar-se em diferentes formas, dando aos anticorpos novas formas com as quais se ligam a antígenos. As conformações dependem do comprimento e da sequência da CDR3. Deve ser recordado que uma CDR3 de HC de comprimento 8, por exemplo, e de qualquer sequência não consegue mimetizar adequadamente o comportamento de uma CDR3 de comprimento 22, por exemplo.

As bibliotecas que contêm somente diversidade de sequências de aminoácidos têm, deste modo, desvantagem na medida em que não refletem a diversidade natural do péptido, polipéptido ou proteína que se pretende que a biblioteca mimetize. Adicionalmente, a diversidade em comprimento pode ser importante para o funcionamento derradeiro da proteína, péptido ou polipéptido. Por exemplo, no que diz respeito a uma biblioteca compreendendo regiões de anticorpos, muitos dos péptidos, polipéptidos, proteínas exibidos, exibidos e expressos, ou compreendidos pelos pacotes genéticos da biblioteca podem não se dobrar apropriadamente ou a sua ligação a um antígeno pode estar em desvantagem, se a diversidade em sequência e comprimento não estiver representada na biblioteca.

Uma desvantagem adicional de tais bibliotecas de pacotes genéticos que exibem, exibem e expressam, ou compreendem péptidos, polipéptidos e proteínas é que não estão focadas naqueles membros que são baseados em diversidade ocorrendo naturalmente e assim em membros que é o mais provável que sejam funcionais e o menos provável que sejam imunogênicos.

Ao invés, as bibliotecas, tipicamente, tentam incluir tanta diversidade ou variação como possível em cada posição de CDR. Isto torna a construção de bibliotecas morosa e menos eficaz do que o necessário. O grande número de membros que são produzidos por tentativa de capturar a diversidade completa torna também o rastreio mais pesado do que necessita de ser. Isto é particularmente verdade dado que muitos membros da biblioteca não serão funcionais ou serão não especificamente pegajosos.

Adicionalmente ao trabalho de construção de bibliotecas sintéticas está a questão da imunogenicidade. Por exemplo existem bibliotecas nas quais todos os resíduos de CDR são Tyr (Y) ou Ser (S). Embora anticorpos (Abs) selecionados destas bibliotecas mostrem elevadas afinidade e especificidade, a sua composição muito incomum pode torná-los imunogénicos. A presente invenção está dirigida na direção de fabrico de Abs que poderiam bem ter vindo do sistema imune humano e assim é menos prováveis que sejam imunogénicos. As bibliotecas da presente invenção retêm tantos resíduos de fusões V-D-J ou V-J como possível. Para reduzir o risco de imunogenicidade pode ser prudente mudar cada aminoácido não da linha germinativa em *framework* e CDRs de volta para a linha germinativa para determinar se a mudança a partir da linha germinativa é necessária para reter a afinidade de ligação. Assim sendo, uma biblioteca que esteja enviesada em cada posição variada na direção da linha germinativa reduzirá a probabilidade de isolamento de Abs que tenham aminoácidos não da linha germinativa desnecessários.

Abs são grandes proteínas e estão sujeitos a várias formas de degradação. Uma forma de degradação é a desamidação de

resíduos Asn e Gln (especialmente em Asn-Gly ou Gln-Gly) e a isomerização de resíduos Asp. Outra forma de degradação é a oxidação de metioninas, triptofano, e cisteína. Outra forma de degradação é a clivagem de dipéptidos Asp-Pro. Outra forma de degradação é a formação de piroglutamato a partir de Glu ou Gln N-terminal. É vantajoso proporcionar uma biblioteca na qual a ocorrência de sequências problemáticas seja minimizada.

São proporcionadas bibliotecas de vetores ou pacotes que codificam membros de uma família diversa de anticorpos humanos compreendendo CDR3s de cadeia pesada (HC) que tenham entre cerca de 3 aminoácidos de comprimento e cerca de 35 aminoácidos de comprimento. As CDR3s de HC podem também, em certas formas de realização, ser ricas em Tyr (T) e Ser (S) e/ou compreender regiões D diversificadas e/ou compreender regiões JH prolongadas. Por exemplo, as CDR3s de HC podem conter mais do que cerca de 40% (p.ex., entre cerca de 43% e cerca de 80%; p.ex., mais do que cerca de 40% mas menos do que cerca de 100%) de resíduos Y e/ou S, p.ex., como proporcionado nos exemplos aqui. São também proporcionadas bibliotecas focadas compreendendo tais CDR3s de HC. São também proporcionados desenhos para CDR1 de HC, CDR2 de HC, e uma biblioteca de VKIII A27 com diversidade nas CDRs. Uma biblioteca de vetores ou pacotes que codifica membros de uma família diversa de anticorpos humanos compreendendo CDR3s de HC descrita aqui pode ter adicionalmente diversidade numa ou mais (p.ex., numa, em duas ou três) de CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, e CDR3 de LC. Por exemplo, a biblioteca pode ter diversidade numa ou mais (p.ex., numa, em duas ou três) de CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, e CDR3 de LC como descrito aqui.

Uma região D diversificada é uma região D na qual uma ou mais mudanças de aminoácidos foram introduzidas (p.ex., em comparação com a sequência de uma região D ocorrendo naturalmente; por exemplo, um codão de paragem pode ser mudado para um resíduo Tyr).

Uma região JH prolongada é uma região JH na qual um ou mais resíduos de aminoácidos presentes no terminal amino da sequência de *framework* da região JH (p.ex., terminal amino em sequências FR4, p.ex., que começam com WGQ...). Por exemplo, JH1 é uma região JH prolongada. Como outros exemplos, JH2, JH3, JH4, JH5, e JH6 são regiões JH prolongadas.

São também proporcionados métodos de fabrico e rastreio das bibliotecas acima e das CDR3s de HC e anticorpos obtidos em tal rastreio. Composições e estojos para a prática destes métodos são também descritos aqui.

Em alguns aspetos, a divulgação apresenta uma biblioteca focada de vetores ou pacotes genéticos que exibem, exibem e expressam, ou compreendem um membro da uma família diversa de péptidos, polipéptidos e proteínas relacionados com anticorpos humanos (p.ex., uma família diversa de anticorpos) e exibem coletivamente, exibem e expressam, ou compreendem pelo menos uma porção da diversidade da família, em que os vetores ou pacotes genéticos compreendem sequências de ADN variegadas que codificam uma CDR3 de cadeia pesada (HC) selecionada do grupo consistindo em:

- (a) uma CDR3 de HC que tem cerca de 3 ou cerca de 4 ou cerca de 5 aminoácidos de comprimento;

(b) uma CDR3 de HC que tem cerca de 23, cerca de 24, cerca de 25, cerca de 26, cerca de 27, cerca de 28, cerca de 29, cerca de 30, cerca de 31, cerca de 32, cerca de 33, cerca de 34 ou cerca de 35 aminoácidos de comprimento (p.ex., cerca de 23 a cerca de 35 aminoácidos de comprimento); e

c) um CDR3 de HC que tem de cerca de 6 a cerca de 20 aminoácidos de comprimento (p.ex., cerca de 6, cerca de 7, cerca de 8, cerca de 9, cerca de 10, cerca de 11, cerca de 12, cerca de 13, cerca de 14, cerca de 15, cerca de 16, cerca de 17, cerca de 18, cerca de 19, ou cerca de 20 aminoácidos de comprimento);

em que a CDR3 de HC compreende aminoácidos de uma região D (p.ex., uma região D diversificada) (ou seu fragmento (p.ex., 3 ou mais aminoácidos da região D, p.ex., região D diversificada)) ou uma região JH (p.ex., uma região JH prolongada).

Em algumas formas de realização, a CDR3 de HC está enriquecida em Tyr (Y) e Ser (S) (p.ex., mais do que 40% dos resíduos da CDR3 de HC são Y e/ou S).

Em algumas formas de realização, a biblioteca (p.ex., os seus vetores ou pacotes genéticos) compreende uma região D ou um fragmento de uma região D (p.ex., em que a região D é adjacente a uma região JH).

Em algumas formas de realização, a biblioteca compreende uma região JH, p.ex., uma região JH prolongada.

Em algumas formas de realização, a CDR3 de HC compreende aminoácidos de uma região D ou um fragmento de uma região D (p.ex., em que a região D é adjacente a uma região JH).

Em algumas formas de realização, a região D é selecionada do grupo consistindo em D2-2 (RF 2), D2-8 (RF 2), D2-15 (RF 2), D2-21 (RF 2), D3-16 (RF 2), D3-22 (RF 2), D3-3 (RF-2), D3-9 (RF 2), D3-10 (RF 2), D1-26 (RF 3), D4-11 (RF 2), D4-4 (RF 2), D5-5 (RF 3), D5-12 (RF 3), D5-18 (RF 3), D6-6 (RF 1), D6-13 (RF 1), e D6-19 (RF 1).

Em algumas formas de realização, a CDR3 de HC compreende aminoácidos de uma região JH. A região JH pode ser uma região JH prolongada. Em algumas formas de realização, a região JH prolongada é selecionada do grupo consistindo em JH1, JH2, JH3, JH4, JH5, e JH6. Em algumas formas de realização, a região JH pode estar enriquecida em resíduos Y e/ou S, por exemplo, pode conter mais do que cerca de 40% (p.ex., entre cerca de 43% e cerca de 80%; p.ex., mais do que cerca de 40% mas menos do que cerca de 100%) de resíduos Y e/ou S.

Em algumas formas de realização, a região D compreende um ou mais resíduos de cisteína (Cys) e, em algumas formas de realização, o um ou mais resíduos de cisteína não mantidos constantes (p.ex., não são variados).

Em algumas formas de realização, a CDR3 de HC (p.ex., o ADN codificando a CDR3 de HC) compreende um ou mais códons de enchimento entre FR3 e a região D e cada codão de enchimento é individualmente NNK, TMY, TMT, ou TMC (TMY, TMT, ou TMC codifica S ou Y).

Em algumas formas de realização, a CDR3 de HC (p.ex., o ADN codificando a CDR3 de HC) compreende um ou mais codões de enchimento entre a região D e JH e cada codão de enchimento é individualmente NNK, TMY, TMT, ou TMC.

Em algumas formas de realização, a biblioteca (p.ex., os vetores ou pacotes genéticos da biblioteca) compreende adicionalmente uma CDR1 de HC, CDR2 de HC, e/ou uma cadeia leve e compreende também diversidade na CDR1 de HC, CDR2 de HC, ou cadeia leve compreende diversidade na CDR1 de HC e/ou CDR2 de HC, e/ou uma cadeia leve (p.ex., cadeia leve kappa ou lambda) (respectivamente). For exemplo, a diversidade de CDR3 de HC pode ser construída no fundo da diversidade em CDR1 de HC, CDR2 de HC, e/ou cadeias leves. Por exemplo, a diversidade de cadeias leves pode ser codificada na mesma molécula de ADN como a diversidade de HC ou as diversidades de LC e HC podem ser codificadas em moléculas de ADN separadas.

Em alguns aspetos, a divulgação apresenta uma biblioteca compreendendo uma CDR3 de HC que tem 3, 4, ou 5 aminoácidos de comprimento, em que a CDR3 compreende aminoácidos de uma região JH (p.ex., região JH prolongada) ou de uma região D (p.ex., uma região D diversificada) (ou seu fragmento (p.ex., 3 ou mais aminoácidos da região D, p.ex., região D diversificada)) unidos à porção FR4 de uma região JH.

Em algumas formas de realização, a CDR3 de HC é de uma região D unida à porção FR4 de uma região JH e compreende um trímero, um tetrâmero, ou um pentâmero, em que o trímero, tetrâmero, ou pentâmero não compreende um resíduo de cisteína.

Em algumas formas de realização, a CDR3 de HC é de uma região D unida à porção FR4 de uma região JH e compreende um trímero, um tetrâmero, ou um pentâmero, em que o trímero, tetrâmero, ou pentâmero não compreende um codão de paragem.

Em algumas formas de realização, a região D (p.ex., o ADN codificando a região D) compreende um codão TAG e o codão TAG está substituído por um codão selecionado do grupo consistindo em TCG, TTG, TGG, CAG, AAG, TAT, e GAG.

Em algumas formas de realização, a região D (p.ex., o ADN codificando a região D) compreende um codão TAA e o codão TAA está substituído por um codão selecionado do grupo consistindo em TCA, TTA, CAA, AAA, TAT, e GAA.

Em algumas formas de realização, a região D (p.ex., o ADN codificando a região D) compreende um codão TGA e o codão TGA está substituído por um codão selecionado do grupo consistindo em TGG, TCA, TTA, AGA, e GGA.

Em algumas formas de realização, a biblioteca compreende adicionalmente diversidade em CDR1 de HC e/ou CDR2 de HC, e/ou uma cadeia leve (p.ex., cadeia leve kappa ou lambda). For exemplo, a diversidade de CDR3 de HC pode ser construída no fundo da diversidade em CDR1 de HC, CDR2 de HC, e/ou cadeias leves. Por exemplo, a diversidade de cadeias leves pode ser codificada na mesma molécula de ADN como a diversidade de HC ou as diversidades de LC e HC podem ser codificadas em moléculas de ADN separadas.

Em alguns aspetos, a divulgação proporciona um método de diversificação de uma biblioteca, compreendendo o método mutagenização de uma biblioteca descrita aqui.

Em algumas formas de realização, a mutagenização compreende PCR propenso a erros.

Em algumas formas de realização, a mutagenização compreende *wobbling*.

Em algumas formas de realização, a mutagenização compreende *dobbling*.

Em algumas formas de realização, a mutagenização introduz em média cerca de 1 a cerca de 10 mutações (p.ex., cerca de 1, cerca de 2, cerca de 3, cerca de 4, cerca de 5, cerca de 6, cerca de 7, cerca de 8, cerca de 9, cerca de 10 mutações; p.ex., mudanças de bases) por CDR3 de HC.

"*Wobbling*" é um método de fabrico de ADN variegado tal que seja favorecida uma sequência original. Se a sequência original tiver, por exemplo, uma Ala que poderia ser codificada com GCT, a mistura (0,7 de G, 0,1 de A, 0,1 de T, 0,1 de C) pode ser usada para a primeira porção, (0,7 de C, 0,1 de A, 0,1 de T, 0,1 de G) na segunda posição, e (0,7 de T, 0,1 de A, 0,1 de G, 0,1 de C) na terceira. Outras razões de "dopagem" podem ser usadas. Isto permite que Ala apareça cerca de 50% das vezes enquanto V, D, G, T, P, e S ocorrem cerca de 7% das vezes. Outros tipos de AA ocorrem a frequência mais baixa.

Em alguns aspetos, a presente divulgação destina-se, p.ex., a manter um repertório de CDR1-2 de HC (purificadas), e construir diversidade de CDR3 de HC e LC sintética.

Em algumas formas de realização, a divulgação proporciona uma cassette para exibição de uma CDR3 de cadeia pesada (HC) *wobbled*, por exemplo, em que a cassette compreende a cassette mostrada na Tabela 400.

Em alguns aspetos, a presente divulgação apresenta uma biblioteca na qual os níveis de Tyr são controlados na CDR3 de HC. Em algumas formas de realização, as regiões CDR3 de HC contêm cerca de 15% ou mais (p.ex., cerca de 16%, cerca de 18%, cerca de 20%, ou mais) de resíduos de Tyr. Em algumas formas de realização, elevados níveis (p.ex., mais do que cerca de 20%) de Tyr são inseridos na CDR3 de HC de membros da biblioteca, p.ex., em regiões D e J *stumps* (ou sequências sintéticas correspondendo a eles) que contêm Tyr. Em algumas formas de realização, em posições *leadin* ou de enchimento de DJ (ou sequências sintéticas correspondendo a elas), Tyr é permitida, mas a não mais do que 20%. Em algumas formas de realização, as regiões CDR3 de HC contêm menos do que cerca de 15% (p.ex., cerca de 14%, cerca de 12%, cerca de 10%, cerca de 8%, cerca de 6% ou menos) de resíduos de Tyr. Em algumas formas de realização, as posições *leadin* de HC ou de enchimento de DJ (ou sequências sintéticas correspondendo a eles) contêm menos do que cerca de 15% (p.ex., cerca de 14%, cerca de 12%, cerca de 10%, cerca de 8%, cerca de 6% ou menos) de resíduos de Tyr.

Em alguns aspetos, a divulgação apresenta uma biblioteca de pacotes genéticos que codificam uma cadeia pesada de

anticorpo humano na qual uma sequência de aminoácidos parente compreende uma sequência de VH seguida por zero a dez aminoácidos selecionados do grupo consistindo em (Y, S, D, L, R), seguidos por uma região D ou fragmento de uma região D de humano, seguido por zero a dez aminoácidos selecionados do grupo consistindo em (Y, S, R, D, L), seguidos por um segmento JH que compreende pelo menos W103 para a frente em que o ADN variável codificando esta sequência é sintetizado de um modo que a sequência de aminoácidos parente seja a mais provável (p.ex., por *wobbling*).

Em alguns aspetos, a divulgação apresenta uma biblioteca de cadeias leves tendo regiões de *framework* da linha germinativa e em que as CDRs são variadas tal que os resíduos longe do local de combinação ou tendo grupos laterais enterrados sejam mantidos constantes. Em algumas formas de realização é usado um método de síntese de ADN variável tal que a sequência da linha germinativa seja a mais provável (p.ex., por *wobbling*).

Em alguns aspetos, a divulgação apresenta uma biblioteca de diversos membros codificando regiões variáveis de ligação ao antigénio como divulgado aqui.

Em alguns aspetos, a divulgação apresenta uma biblioteca de diversos membros codificando regiões CDR3 de HC como divulgado aqui. Em algumas formas de realização, a biblioteca é uma biblioteca da Tabela 1097.

Em alguns aspetos, a divulgação apresenta uma biblioteca de diversos membros, codificando compreendendo cada membro uma CDR3 de HC, em que

pelo menos 1, 2, 3, 4, 5, 6, 7, ou 8 posições na CDR3 de HC, respetivamente, estão ocupadas por G, S, R, D, L, e Y na biblioteca nas seguintes proporções [1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y] e, opcionalmente, as últimas 4 posições de CDR3 de HC são representadas como se segue:

o aminoácido parente está presente 3, 4, 5, 6, 7, 8, 10 vezes mais provavelmente como outros tipos de aminoácidos, em que os outros tipos de aminoácidos compreendem Y, S, D, R, G.

Em alguns aspetos, a divulgação apresenta uma biblioteca de diversos membros, codificando compreendendo cada membro uma CDR3 de HC, em que pelo menos uma das e preferencialmente todas as primeiras 1, 2, 3, 4, 5, 6, 7, ou 8 posições na CDR3 de HC estão ocupadas por G, S, R, D, L, e Y na biblioteca nas seguintes proporções [1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y] e opcionalmente as últimas 4 posições de CDR3 de HC são representadas como se segue:

o aminoácido parente está presente 3, 4, 5, 6, 7, 8, 10 vezes mais provavelmente como outros tipos de aminoácidos, em que os outros tipos de aminoácidos compreendem Y, S, D, R, G.

Em alguns aspetos, a divulgação apresenta uma biblioteca de diversos membros, codificando cada membro uma CDR3 de HC, em que o comprimento de CDR3 de HC tem 10, 11, ou 12 posições;

cada um das primeiras 6, 7, ou 8 posições na CDR3 de HC, respectivamente, está ocupada por G, S, R, D, L, e Y na biblioteca nas seguintes proporções [1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y]; as últimas 4 posições de CDR3 de HC são representadas como se segue:

o aminoácido parente está presente e 3, 4, 5, 6, 7, 8, 10 vezes mais provavelmente como outros tipos de aminoácidos, em que os outros tipos de aminoácidos compreendem Y, S, D, R, G.

Em algumas formas de realização, cada uma das últimas 4 posições de CDR3 de HC está representada na biblioteca como 7/12 do parente, mais 1/12 cada um de Y, S, D, R, e G.

Em algumas formas de realização, cada uma das últimas 4 posições de CDR3 de HC está representada na biblioteca como A6 = 7/12 de A, mais 1/12 cada um de Y, S, D, R, e G; F7 = 7/12 de F mais 1/12 cada um de Y, S, D, R, e G; D8 = 7/11 de D mais 1/11 de Y, S, R, e G; I9 = 7/12 de I mais 1/12 de Y, S, R, D, G.

Em algumas formas de realização, os membros codificam adicionalmente CDR1 de HC, CDR2 de HC.

Em algumas formas de realização, os membros codificam adicionalmente regiões de *framework* (FR) 1-4.

Em algumas formas de realização, os membros codificam CDR1 de HC, CDR2 de HC e regiões FR 1-4.

Em algumas formas de realização, os membros compreendem uma *framework* de HC 3-23.

Em algumas formas de realização, a biblioteca compreende adicionalmente uma região variável de LC.

Em algumas formas de realização, a biblioteca compreende membros codificando diversas regiões variáveis de LC.

Em algumas formas de realização, os membros compreendendo uma região variável de LC compreendem uma *framework* de LC A27.

Em algumas formas de realização, a biblioteca é uma biblioteca de exibição, p.ex., uma biblioteca de exibição de fagos.

Em algumas formas de realização, a biblioteca tem pelo menos 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} membros diversos.

Em alguns aspetos, a divulgação apresenta um método de seleção de um membro da biblioteca, compreendendo contacto de uma biblioteca descrita aqui com um alvo, permitindo que um membro se ligue ao referido alvo, e recuperação do membro que se liga ao alvo.

Estas formas de realização da presente invenção, outras formas de realização, e as suas funcionalidades e características serão aparentes a partir da descrição, desenhos, e reivindicações que se seguem.

Descrição detalhada

Os anticorpos ("Ab") concentram a sua diversidade naquelas regiões que estão envolvidas na determinação da afinidade e especificidade do Ab para alvos particulares. Estas regiões podem ser diversas em sequência ou em comprimento. Geralmente são diversas de ambos os modos. No entanto, dentro de famílias de anticorpos humanos, as diversidades, tanto em sequência como em comprimento, não são verdadeiramente aleatórias. Ao invés, alguns resíduos de aminoácidos são preferenciais em certas posições das CDRs e alguns comprimentos de CDR são preferenciais. Estas diversidades preferenciais são responsáveis pela diversidade natural da família dos anticorpos.

De acordo com esta invenção, e como mais completamente descrito em baixo, bibliotecas de vetores e pacotes genéticos que codificam membros de uma família diversa de anticorpos humanos compreendendo CDR3s de cadeia pesada (HC) que têm entre cerca de 3 e cerca de 35 aminoácidos de comprimento podem ser preparadas e usadas. As CDR3s de HC podem também, em certas formas de realização, ser ricas em Y e S e/ou compreender regiões D diversificadas. São também proporcionadas bibliotecas focadas compreendendo tais CDR3s de HC.

Quando uma célula imune constrói uma cadeia pesada de anticorpo conecta um segmento V com um segmento D e esse a um segmento J. O segmento D é opcional e cerca de 50% dos Abs humanos têm Ds reconhecíveis. A célula pode realizar modificação considerável nos locais de junção (V a D, D a J, ou V a J) quer removendo quer adicionando bases, mas não exatamente aleatoriamente. O anticorpo inicialmente rearranjado é apresentado à superfície da célula e, se se

ligar a um antígeno (Ag), a célula é estimulada para realizar mutações somáticas para melhorar a afinidade. Existem pontos fulcrais codificados nos genes da linha germinativa de imunoglobulinas tal que seja muito provável que certos locais no gene de Ab sofram um conjunto particular de mutações na procura de um melhor ligante a um Ag persistente. Na natureza, algumas das mutações são em posições de *framework* mas a maioria está nas regiões determinadoras da complementaridade (CDRs). A CDR3 da cadeia pesada (HC) é de particular interesse porque mostra não só um elevado grau de diversidade de sequências mas também diversidade de comprimentos. Foram construídas bibliotecas de anticorpos (Ab) nas quais as CDRs são substituídas por ADN aleatório, e foram obtidos Abs úteis. No entanto, alguns Abs terapêuticos mostram um grau de antigenicidade significativo. É possível que Abs que estejam mais próximos da linha germinativa humana sejam menos antigénicos.

Definições

Por conveniência, antes de descrição adicional da presente invenção, certos termos empregues na especificação, exemplos e reivindicações anexas são definidos aqui.

As formas singulares "um", "uma", e "o/a" incluem referências plurais a não ser que o contexto dite claramente de outro modo.

O termo "afinidade" ou "afinidade de ligação" refere-se à constante de associação aparente ou K_a . A K_a é o recíproco da constante de dissociação (K_d). Uma proteína de ligação pode, por exemplo, ter uma afinidade de ligação de pelo menos 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} e 10^{11} M^{-1} para uma

molécula alvo particular. A ligação por afinidade mais elevada de uma proteína a um primeiro alvo em relação a um segundo alvo pode ser indicada por uma K_A mais elevada (ou um valor numérico mais pequeno K_D) quanto a ligação do primeiro alvo do que a K_A (ou valor numéricos K_D) para ligação do segundo alvo. Em tais casos, a proteína de ligação tem especificidade quanto ao primeiro alvo (p.ex., uma proteína numa primeira conformação ou seu mímico) em relação ao segundo alvo (p.ex., a mesma proteína numa segunda conformação ou seu mímico; ou uma segunda proteína). As diferenças na afinidade de ligação (p.ex., para especificidade ou outras comparações) podem ser pelo menos 1,5, 2, 3, 4, 5, 10, 15, 20, 37,5, 50, 70, 80, 91, 100, 500, 1000, ou 10^5 vezes.

A afinidade de ligação pode ser determinada por uma variedade de métodos incluindo diálise de equilíbrio, ligação de equilíbrio, filtração em gel, ELISA, ressonância de plasmão de superfície, ou espectroscopia (p.ex., usando um ensaio de fluorescência). Condições exemplares para avaliação de afinidade de ligação são em tampão TRIS (TRIS a 50 mM, NaCl a 150 mM, CaCl_2 a 5 mM a pH 7,5). Estas técnicas podem ser usadas para medir a concentração de proteína de ligação ligada e livre como uma função da concentração da proteína de ligação (ou alvo). A concentração de proteína de ligação ligada ([Ligada]) está relacionada com a concentração de proteína de ligação livre ([Livre]) e a concentração de locais de ligação para a proteína de ligação no alvo onde (N) é o número de locais de ligação por molécula alvo pela seguinte equação:

$$[\text{Ligada}] = N \cdot [\text{Livre}] / ((1/K_A) + [\text{Livre}]).$$

Não é sempre necessário fazer uma medição exata da K_A , no entanto, uma vez que por vezes é suficiente obter-se uma medição quantitativa da afinidade, p.ex., determinada usando um método tal como análise ELISA ou FACS, é proporcional a K_A , e pode ser assim usada para comparações, tal como determinação de se uma afinidade mais elevada é, p.ex., 2 vezes mais elevada, para se obter uma medição qualitativa da afinidade, ou para se obter uma inferência da afinidade, p.ex., por atividade num ensaio funcional, p.ex., um ensaio *in vitro* ou *in vivo*.

O termo "anticorpo" refere-se a uma proteína que inclui pelo menos um domínio variável de imunoglobulina ou sequência de domínio variável de imunoglobulina. Por exemplo, um anticorpo pode incluir uma região variável de cadeia pesada (H) (abreviada aqui como VH), e uma região variável de cadeia leves (L) (abreviada aqui como VL). Noutro exemplo, um anticorpo inclui duas regiões variáveis de cadeia pesada (H) e duas regiões variáveis de cadeia leve (L). A cadeia pesada e a cadeia leve podem ser também abreviadas como HC e LC, respetivamente. O termo "anticorpo" engloba fragmentos de ligação do antígeno de anticorpos (p.ex., anticorpos de cadeia única, fragmentos Fab e sFab, $F(ab')_2$, fragmentos Fd, fragmentos Fv, scFv, e fragmentos de domínio de anticorpos (dAb) (de Wildt et al., *Eur J Immunol.* 1996; 26(3):629-39)) bem como anticorpos completos. Um anticorpo pode ter as características estruturais de IgA, IgG, IgE, IgD, IgM (bem como seus subtipos). Os anticorpos podem ser de qualquer fonte, mas primata (humano e primata não humano) e primatizados são preferenciais.

As regiões VH e VL podem ser adicionalmente subdivididas em regiões de hipervariabilidade, denominadas "regiões determinadoras da complementaridade" ("CDR"), intercaladas com regiões que estão mais conservadas, denominadas "regiões de *framework*" ("FR"). A extensão da região de *framework* e das CDRs foi precisamente definida (ver, Kabat, E.A., et al. (1991) *Sequences of Proteins of Immunological Interest*, Quinta Edição, U.S. Department of Health and Human Services, Publicação NIH No. 91-3242, e Chothia, C. et al. (1987) *J. Mol. Biol.* 196:901-917, ver também www.hgmp.mrc.ac.uk). As definições de Kabat são usadas aqui. Cada VH e VL é tipicamente composta por três CDRs e quatro FRs, dispostas a partir do terminal amino para o terminal carboxi na seguinte ordem FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4.

A cadeia de VH e VL do anticorpo pode adicionalmente incluir toda a ou parte de uma região constante de cadeia pesada ou leve, para formar desse modo uma cadeia de imunoglobulina pesada ou leve, respetivamente. Numa forma de realização, o anticorpo é um tetrâmero de duas cadeias de imunoglobulina pesadas e duas cadeias de imunoglobulinas leves, em que as cadeias de imunoglobulinas pesadas e leves estão interconectadas por, p.ex., ligações de dissulfeto. Em IgGs, a região constante de cadeia pesada inclui três domínios de imunoglobulina, CH1, CH2 e CH3. A região constante de cadeia leve inclui um domínio CL. A região variável das cadeias pesadas e leves contém um domínio de ligação que interage com um antigénio. As regiões constantes dos anticorpos medeiam tipicamente a ligação do anticorpo a tecidos hospedeiros ou fatores, incluindo várias células do sistema imune (p.ex., células efectoras) e o primeiro componente (CIq) do sistema do complemento

clássico. As cadeias leves da imunoglobulina podem ser dos tipos kappa ou lambda. Numa forma de realização, a anticorpo está glicosilado. Um anticorpo pode ser funcional quanto a citotoxicidade dependente de anticorpo e/ou citotoxicidade dependente de complemento.

Uma ou mais regiões de um anticorpo podem ser humanas ou efetivamente humanas. Por exemplo, uma ou mais das regiões variáveis podem ser humanas ou efetivamente humanas. Por exemplo, uma ou mais das CDRs podem ser humanas, p.ex., CDR1 de HC, CDR2 de HC, CDR3 de HC, CDR1 de LC, CDR2 de LC, e CDR3 de LC. Cada uma das CDRs de cadeia leve pode ser humana. A CDR3 de HC pode ser humana. Uma ou mais das regiões de *framework* podem ser humanas, p.ex., FR1, FR2, FR3, e FR4 da HC ou LC. Por exemplo, a região de Fc pode ser humana. Numa forma de realização, todas as regiões de *framework* são humanas, p.ex., derivadas de uma célula somática humana, p.ex., uma célula hematopoiética que produz imunoglobulinas ou uma célula não hematopoiética. Numa forma de realização, as sequências humanas são sequências da linha germinativa, p.ex., codificadas por um ácido nucleico da linha germinativa. Numa forma de realização, os resíduos de *framework* (FR) de um Fab selecionado podem ser convertidos no tipo de aminoácido do resíduo correspondente no gene da linha germinativa de primata mais similar, especialmente o gene da linha germinativa humana. Uma ou mais das regiões constantes podem ser humanas ou efetivamente humanas. Por exemplo, pelo menos 70, 75, 80, 85, 90, 92, 95, 98, ou 100% de um domínio variável de imunoglobulina, da região constante, dos domínios constantes (CH1, CH2, CH3, CL), ou do anticorpo inteiro podem ser humanos ou efetivamente humanos.

Todo o ou parte de um anticorpo pode ser codificado por um gene de imunoglobulina ou um seu segmento. Genes de imunoglobulina de humano exemplares incluem os genes da região constante kappa, lambda, alfa (IgA1 e IgA2), gama (IgG1, IgG2, IgG3, IgG4), delta, épsilon e mu, bem como os muitos genes da região variável de imunoglobulina. "Cadeias leves" de imunoglobulina de comprimento total (cerca de 25 kDa ou cerca de 214 aminoácidos) são codificadas por um gene de região variável no terminal de NH₂ (cerca de 110 aminoácidos) e um gene de região constante kappa ou lambda no terminal de COOH. "Cadeias pesadas" de imunoglobulina de comprimento total (cerca de 50 KDa ou cerca de 446 aminoácidos) são similarmente codificadas por um gene de região variável (cerca de 116 aminoácidos) e um dos outros genes de região constante acima mencionados, p.ex., gama (codificando cerca de 330 aminoácidos). O comprimento de HC de humano varia consideravelmente porque CDR3 de HC varia de cerca de 3 resíduos de aminoácidos a mais do que 35 resíduos de aminoácidos.

Aqui, os termos "segmento D" e "região D" são usados indistintamente e são idênticos. É para ser entendido que estes itens têm representações de ADN e aminoácidos e que qual se pretende é claro a partir do contexto.

Uma "biblioteca" ou "biblioteca de exibição" refere-se a uma coleção de sequências de nucleótidos, p.ex., ADN, dentro de clones; ou um coleção geneticamente diversa de polipéptidos exibidos em pacotes de exibição replicáveis capazes de seleção ou rastreamento para proporcionar um polipéptido individual ou uma população mista de polipéptidos.

O termo "pacote" como usado aqui refere-se a um pacote de exibição genético replicável no qual a partícula está a exibir um polipéptido à sua superfície. O pacote pode ser um bacteriófago que exibe um domínio de ligação do antigénio à sua superfície. Este tipo de pacote tem sido chamado um anticorpo de fago (pAb).

Um "alvo pré-determinado" refere-se a uma molécula alvo cuja identidade é conhecida antes do seu uso em qualquer um dos métodos divulgados.

O termo "pacote de exibição replicável" como usado aqui refere-se a uma partícula biológica que tem informação genética proporcionando a partícula com a capacidade de se replicar. A partícula pode exibir à sua superfície pelo menos parte de um polipéptido. O polipéptido pode ser codificado por informação genética nativa da partícula e/ou artificialmente colocada na partícula ou num seu antepassado. O polipéptido exibido pode ser qualquer membro de um par de ligação específico, p.ex., domínios de cadeia pesada ou leve baseados numa molécula de imunoglobulina, uma enzima ou um recetor, etc. A partícula pode ser, por exemplo, um vírus, p.ex., um bacteriófago tal como fd ou M13.

O termo "vetor" refere-se a uma molécula de ADN, capaz de replicação num organismo hospedeiro, no qual um gene é inserido para construir uma molécula de ADN recombinante. Um "vetor de fago" é um vetor derivado por modificação de um genoma de fago, contendo uma origem de replicação para um bacteriófago, mas não uma para um plasmídeo. Um "vetor de fagemídeo" é um vetor derivado por modificação de um

genoma de plasmídeo, contendo uma origem de replicação para um bacteriófago bem como uma origem de replicação de plasmídeo.

Na discussão de oligonucleótidos, a notação "[RC]" indica que o Complemento Reverso do oligonucleótido mostrado é o que está a ser usado.

CDR3s de Cadeia Pesada de Anticorpo de Humano

O Gene da Linha Germinativa (GLCG) de cadeia pesada ("HC") (também conhecido como VP-47) é responsável por cerca de 12% de todos os Abs de humano e é preferencial como a *framework* na forma de realização preferencial da invenção. Deve, no entanto, ser entendido que outras *frameworks* bem conhecidas, tais como 4-34, 3-30, 3-30.3 e 4-30.1, podem ser também usadas sem sair dos princípios das diversidades focadas desta invenção.

Adicionalmente, JH4 (YFDYW₁₀₃GQGTLVTVSS (SEQ ID NO:1)) ocorre mais frequentemente do que JH3 em anticorpos nativos. Consequentemente é preferencial para as bibliotecas focadas desta invenção. No entanto, JH3 (AFDIW₁₀₃GQGTMTTVSS (SEQ ID NO:2)), JH6 (YYYYYGMDVW₁₀₃GQGTTVTVSS (SEQ ID NO:3)), JH1, JH2, ou JH5 poderiam ser usados também. JH2 tem a vantagem de ter RG a 105-106 em vez de QG em todos os outros JHs de humano. JH3 tem a desvantagem de M108. Numa coleção de 1419 Abs que eram positivos quanto a ELISA para pelo menos um alvo vimos 17 JH1s, 31 JH2s, 452 JH3s, 636 JH4s, 32 JH5s, e 251 JH6s. Se presentes, as porções duplamente sublinhadas dos JHs são consideradas como sendo parte da CDR3. Na Tabela 3, as partes de FR4 dos JHs estão sublinhadas.

A frequência à qual cada aminoácido apareceu nas CDR3s de HC destes 1419 Abs foi tabulada e registrada na Tabela 75. Note-se que o aminoácido mais comum é Tyr com Gly, Asp, Ser, e Arg seguindo-se por essa ordem. Rel. para Cima é a abundância relativa de cada tipo em comparação com Cys, a Rel. para Baixo mais comum é a abundância relativa de cada tipo em comparação com Tyr, o mais comum. Consequentemente, os tipos de aminoácidos preferenciais para substituir em CDR3s de HC são Y, G, D, S, e R.

Naturalmente, as CDR3s de HC variam em comprimento. Cerca de metade de HCs de humano consiste nos componentes: V::nz::D::ny::JHn onde V é um V gene, nz é uma série de bases que são essencialmente aleatórias, D é um segmento D, frequentemente com muita modificação em ambas as extremidades, ny é uma série de bases que são essencialmente aleatórias, e JHn é um dos seis segmentos JH, frequentemente com muita modificação na extremidade 5'. Os segmentos D parecem proporcionar segmentos espaçadores que permitem dobragem da IgG. A maior diversidade é nas junções de V com D e de D com JH.

Corbett *et al.* (Corbett SJ, Tomlinson IM, Sonnhhammer EL, Buck D, Winter G. *J Mol Biol.* 1997 V270:587-97) mostraram que o sistema imune humano não insere múltiplos segmentos D e segmentos D recombinantes. Não obstante, os segmentos D foram selecionados como sendo bons componentes de CDR3s de HC e a presente invenção compreende CDR3 de HC que contêm mais do que um segmento D.

Os segmentos D de humano têm alguns enviesamentos muito fortes. O inventário dos 522 aminoácidos em segmentos D de humano é Y 70 (13,4%), L 63 (12,1%), V 52 (10%), G 49

(9,4%), I 41 (7,9%), T 40 (7,7%), S 33 (6,3%), W 27 (5,2%), D 21 (4%), A 19 (3,6%), R 16 (3,1%), TAG 15 (2,9%), N 14 (2,7%), Q 11 (2,1%), C 9 (1,7%), E 9 (1,7%), F 8 (1,5%), M 8 (1,5%), TGA 8 (1,5%), TAA 7 (1,3%), P 1 (0,2%), H 1 (0,2%), e K 0 (0%). Existe um D (2-8 RF 1) que tem uma Cys não emparelhada mas também um codão de paragem TGA, logo é pouco usado. Assim sendo, os segmentos D são principalmente hidrofóbicos. As frequências de aminoácidos em CDR3s de HC de humano são mostradas na Tabela 75. Existem similaridades e diferenças nas frequências. Em CDR3s de HC globalmente, Tyr é o mais comum e somente Gly chega perto (96% tão comum como Tyr). Asp (75% tão comum como Tyr), Ser (53% tão comum como Tyr), Leu, Val, e Ile são relativamente comuns se todos os segmentos D forem contados como iguais. O sistema imune não usa os segmentos D com igual frequência. A Tabela 77 mostra a frequência de utilização de segmentos D. Os segmentos D que são frequentemente usados são muito ricos em Tyr, Gly, Ser, e Asp. Arg não é encontrada nos segmentos D mais frequentemente usados nem é Arg codificada em qualquer uma das porções de CDR de segmentos JH. Arg torna-se proeminente por mutação de V, D, e J ou nas regiões de enchimento entre V e D, D e J, ou V e J. Nesta amostra, 50% de todos os aminoácidos são Tyr, Gly, Asp, Ser, ou Arg. Numa forma de realização da presente invenção, a substituição de sequências de CDR3 de HC parente está limitada ao conjunto de aminoácidos consistindo em Tyr, Gly, Ser, Asp, e Arg. Numa forma de realização da presente invenção, Arg é tornada comum nas regiões de enchimento entre V e D, entre D e J, ou entre V e J.

Nas bibliotecas preferenciais desta invenção são usados ambos os tipos de CDR3s de HC. Em CDR3s de HC que não têm nenhum segmento D identificável, a estrutura é V::nz::JHn

(n = 1,6) onde JH está geralmente modificado na extremidade 5. Em CDR3s de HC que têm um segmento D identificável, a estrutura é V::nz::D::ny::JHn.

São proporcionadas aqui CDR3s de HC que tenham entre cerca de 3 e uma cerca de 35 aminoácidos de comprimento. As CDR3s de HC podem também, em certas formas de realização, ser ricas em Y e S e/ou compreender regiões D diversificadas, onde uma região D está presente. Por exemplo, as CDR3s de HC podem conter entre cerca de 43% e cerca de 80% de resíduos de Y e/ou S, p.ex., cerca de 43%, cerca de 48%, cerca de 69%, cerca de 63%, cerca de 71%, cerca de 62%, cerca de 58%, cerca de 68%, cerca de 80%, cerca de 77%, ou mais do que cerca de 40%, ou cerca de 40% a menos do que cerca de 100%, dos resíduos são Y e/ou S. Por exemplo, nem todos os resíduos na CDR3 são Y e/ou S. As CDR3s de HC podem, em certas formas de realização, compreender uma região JH prolongada. Desenhos de componentes de CDR3 de HC exemplares das bibliotecas preferenciais desta invenção são mostrados e descritos nos Exemplos 1, 2, e 3.

Em algumas formas de realização, a diversidade (p.ex., numa CDR, p.ex., CDR3 de HC, ou região de *framework* (p.ex., região de *framework* próxima de ou adjacente a uma CDR, p.ex., CDR3, p.ex., CDR3 de HC)) é gerada para criar em média cerca de 1, cerca de 2, cerca de 3, cerca de 4, cerca de 5, cerca de 6, cerca de 7, cerca de 8, cerca de 9, cerca de 10, ou cerca de 1 a cerca de 10 mutações (p.ex., mudança de bases), p.ex., por CDR (p.ex., CDR3 de HC) ou região de *framework* (p.ex., região de *framework* próxima de ou adjacente a uma CDR, p.ex., CDR3, p.ex., CDR3 de HC). Em algumas implementações, a mutagenese é dirigida a regiões que se sabe que estão ou que é provável que estejam na

interface de ligação. Adicionalmente, a mutagênese pode ser dirigida a regiões de *framework* próximas das ou adjacentes às CDRs. No caso de anticorpos, a mutagênese pode estar também limitada a uma ou algumas das CDRs, p.ex., para fazer melhorias passo a passo precisas. Do mesmo modo, se os ligandos identificados forem enzimas, a mutagênese pode proporcionar anticorpos que sejam capazes de se ligarem ao local ativo e vizinhança. A CDR ou região de *framework* (p.ex., uma CDR3 de HC descrita aqui) pode ser, em certas formas de realização, sujeita a PCR propensa a erros para gerar a diversidade. Esta abordagem usada uma versão "atabalhada" de PCR, na qual a polimerase tem uma taxa de erros relativamente elevada (até 2%), para amplificar a sequência de tipo selvagem, e é geralmente descrita em Pritchard, *et al.* (2005) *J. Theor. Biol.* 234: 497-509 e Leung *et al.* (1989) *Technique* 1:11-15. Outras técnicas de mutagênese exemplares incluem baralhamento de ADN usando clivagem aleatória (Stemmer (1994) *Nature* 389-391; denominado "baralhamento de ácidos nucleicos), RACHITT™ (Coco *et al.* (2001) *Nature Biotech.* 19:354), mutagênese sítio-dirigida (Zoller *et al.* (1987) *Nucl Acids Res* 10:6487-6504), mutagênese de cassete (Reidhaar-Olson (1991) *Methods Enzymol.* 208:564-586) e incorporação de oligonucleótidos degenerados (Griffiths *et al.* (1994) *EMBO J.* 13:3245).

Em algumas formas de realização da invenção são escolhidos segmentos D nos quais a maioria dos resíduos é Ser ou Tyr. Em algumas formas de realização, quando o ADN codificando a região D é sintetizado, cada resíduo de Ser ou Tyr é codificado por TMT, TMC, ou TMY tal que o aminoácido codificado seja Ser ou Tyr.

Em algumas formas de realização, as sequências de CDR3 de HC descritas aqui podem ser sujeitas a seleção quanto a grelhas de leitura aberta por fusão da sequência codificando a CDR3 de HC de interesse em grelha com um gene de resistência a antibióticos, tal como gene *Kan^R* e seleção quanto a resistência à canamicina. Células nas quais a CDR3 potencial tem um codão de paragem ou uma troca de grelhas não terão a resistência a antibióticos e essa sequência será eliminada.

Métodos de Construção de Bibliotecas compreendendo CDR3s de Cadeia Pesada de Anticorpo de Humano e Bibliotecas compreendendo CDR3s de Cadeia Pesada de Anticorpo de Humano

Uma biblioteca de anticorpos é uma coleção de proteínas que incluem proteínas que têm pelo menos uma sequência de domínio variável de imunoglobulina. Por exemplo podem ser usados domínios variáveis camelizados (p.ex., domínios VH) como um molde para uma biblioteca de proteínas que incluem somente uma sequência de domínio variável de imunoglobulina. Noutro exemplo, as proteínas incluem duas sequências de domínios variáveis, p.ex., um domínio VH e VL, que são capazes de se emparelharem. Uma biblioteca de anticorpos pode ser preparada a partir de uma biblioteca de ácidos nucleicos (uma biblioteca codificando anticorpos) que inclua sequências codificando anticorpos, p.ex., compreendendo as sequências codificando as CDR3s de HC proporcionadas aqui.

Nos casos onde é usada uma biblioteca de exibição, cada membro da biblioteca codificando anticorpos pode ser associado ao anticorpo que codifica. No caso de exibição de fagos, a proteína de anticorpo é fisicamente associada (diretamente ou indiretamente) a uma proteína de

revestimento do fago. Um membro típico da biblioteca de exibição de anticorpos exibe um polipéptido que inclui um domínio VH e um domínio VL. O membro da biblioteca de exibição pode exibir o anticorpo como um fragmento Fab (p.ex., usando duas cadeias de polipéptidos) ou um Fv de cadeia única (p.ex., usando uma única cadeia de polipéptidos). Podem ser também usados outros formatos.

Como no caso do Fab e outros formatos, o anticorpo exibido pode incluir uma ou mais regiões constantes como parte de uma cadeia leve e/ou pesada. Numa forma de realização, cada cadeia inclui uma região constante, p.ex., como no caso de um Fab. Noutras formas de realização são incluídas regiões constantes adicionais. É também possível adicionar uma ou mais regiões constantes a uma molécula após ser identificada como tendo local de ligação do antígeno útil. Ver, p.ex., US 2003-0224408.

Bibliotecas de anticorpos podem ser construídas por um número de processos (ver, p.ex., de Haard *et al.* (1999) *J. Biol. Chem* 274:18218-30; Hoogenboom *et al.* (1998) *Immunotechnology* 4:1-20, Hoogenboom *et al.* (2000) *Immunol Today* 21:371-8, e Hoet *et al.* (2005) *Nat Biotechnol.* 23(3):344-8).

Em certas formas de realização para construção de bibliotecas, as cadeias pesadas compreendendo as CDR3s descritas aqui e as cadeias leves kappa e lambda são melhor construídas em vetores separados. Em primeiro lugar é desenhado um gene sintético para incluir cada um dos domínios variáveis sintéticos. As cadeias leves podem ser delimitadas por locais de restrição para ApaLI (posicionada na extremidade final da sequência de sinal) e AscI

(posicionada após o codão de paragem). A cadeia pesada pode ser delimitada por SfiI (posicionada dentro da sequência de sinal PelB) e NotI (posicionada no ligante entre CH1 e a proteína âncora). Podem ser também usadas sequências de sinal sem serem PelB, p.ex., uma sequência de sinal M13 pIII.

Os genes iniciais podem ser feitos com sequências "*stuffer*" em lugar das CDRs desejadas. Uma "*stuffer*" é uma sequência que é para ser cortada e substituída por ADN diverso, mas que não permite expressão de um gene de anticorpo funcional. Por exemplo, a *stuffer* pode conter vários codões de paragem e locais de restrição que não ocorrerão no vetor da biblioteca acabada correta. São usadas *stuffers* para evitar ter qualquer uma das sequências de CDR altamente representada.

A cadeia pesada e as cadeias leves kappa ou lambda podem ser construídas num único vetor ou pacotes genéticos (p.ex., para exibição ou exibição e expressão) tendo locais de restrição apropriados que permitem clonagem destas cadeias. Os processos para construir tais vetores são bem conhecidos e amplamente usados na técnica. Preferencialmente, uma biblioteca de cadeias pesadas e cadeias leves kappa e uma biblioteca de cadeias pesadas e cadeias leves lambda seriam preparadas separadamente.

O mais preferencialmente, a exibição é à superfície de um derivado do fago M13. O vetor mais preferencial contém todos os genes de M13, um gene de resistência a antibióticos, e a cassete de exibição. O vetor preferencial é proporcionado com locais de restrição que permitem introdução e excisão de membros da família diversa de

genes, como cassetes. O vetor preferencial é estável contra rearranjo sob as condições de crescimento usadas para amplificar o fago.

A diversidade capturada pela presente invenção pode ser exibida e/ou expressa num vetor de fagemídeo (p.ex., pMID21 (sequência de ADN mostrada na Tabela 35)) que exhibe e/ou expressa o péptido, polipéptido ou proteína. Tais vetores podem ser também usados para armazenar a diversidade para exibição e/ou expressão subsequentes usando outros vetores ou fago.

Ainda noutras formas de realização, um método denominado Otimização Rápida de Cadeias Leves ou "ROLIC", descrito em U.S.S.N. 61/028,265 depositado a 13 de fevereiro, 2008, U.S.S.N. 61/043,938 depositado a 10 de abril, 2008, e U.S.S.N. 12/371,000 depositado a 13 de fevereiro, 2009, uma grande população de LCs é colocada num vetor de fago que fazem com que sejam exibidos no fago. Uma pequena população (p.ex., 3, 10, ou 25) de HCs é clonada em *E. coli* tal que as HCs sejam secretadas no periplasma, p.ex., aquelas HCs tendo as CDR3s descritas aqui. As *E. coli* são depois infetadas com os vetores de fago codificando a grande população de LCs para produzir os emparelhamentos de proteína HC/LC no fago. As partículas de fago transportam somente um gene de LC.

Noutro aspeto, num método denominado a Seleção Económica de Cadeias Pesadas ou "ESCH", também descrito em U.S.S.N. 61/028,265 depositado a 13 de fevereiro, 2008, U.S.S.N. 61/043,938 depositado a 10 de abril, 2008, e U.S.S.N. 12/371,000 depositado a 13 de fevereiro, 2009, uma pequena população de LCs pode ser colocada num vetor que faz com

que elas sejam secretadas. Uma nova biblioteca de HCs no fago é construída, tal como aquelas proporcionadas aqui compreendendo as CDR3s. As LCs e HCs podem ser depois combinadas pelo método de infecção muito mais eficaz. Logo que um pequeno conjunto de HC eficazes seja selecionado, estas podem ser usadas como tal, alimentadas a ROLIC para se obter um emparelhamento de HC/LC ótimo, ou clonadas numa biblioteca de Fab de LCs para seleção clássica.

A diversidade capturada pela presente invenção pode ser exibida e/ou expressa usando um vetor adequado para expressão numa célula eucariótica, p.ex., um vetor de levedura, p.ex., para expressão numa célula de levedura.

Outros tipos de exibição de proteínas incluem exibição baseada em células (ver, p.ex., WO 03/029,456); exibição de ribossomas (ver, p.ex., Mattheakis et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:9022 e Hanes et al. (2000) *Nat Biotechnol.* 18:1287-92); fusões proteína-ácido nucleico (ver, p.ex., Pat. dos E.U.A. No. 6,207,446); e imobilização num marcador não biológico (ver, p.ex., Pat. dos E.U.A. No. 5,874,214).

Os anticorpos isolados das bibliotecas da presente divulgação podem ser analisados para se determinar o tipo da LC e o gene de linha germinativa mais próximo. Numa forma de realização preferencial, resíduos de *framework* não da linha germinativa são mudados de volta para o aminoácido da linha germinativa desde que a afinidade e especificidade de ligação não sejam adversamente afetadas num grau inaceitável. As substituições podem ser feitas como um grupo ou unicamente. Sequências de linha germinativa de humano são divulgadas em Tomlinson, I.A. et al., 1992, *J.*

Mol. Biol. 227:776-798; Cook, G. P. et al., 1995, *Immunol. Today* 16 (5): 237-242; Chothia, D. et al., 1992, *J. Mol. Bio.* 227:799-817. O diretório V BASE proporciona um diretório extensivo de sequências de região variável de imunoglobulina de humano (compilado por Tomlinson, I.A. et al. MRC Centre for Protein Engineering, Cambridge, RU). Os anticorpos são "devolvidos à linha germinativa" por reversão de um ou mais aminoácidos não da linha germinativa em regiões de *framework* nos correspondentes aminoácidos da linha germinativa do anticorpo, desde que as propriedades de ligação sejam substancialmente retidas. Métodos similares podem ser também usados na região constante, p.ex., em domínios de imunoglobulina constantes.

Por exemplo, um anticorpo pode incluir uma, duas, três, ou mais substituições de aminoácidos, p.ex., numa região de *framework*, CDR, ou constante, para a tornar mais similar a uma sequência da linha germinativa de referência. Um método de devolução à linha germinativa exemplar pode incluir identificação de uma ou mais sequências da linha germinativa que sejam similares (p.ex., o mais similar numa base de dados particular) à sequência do anticorpo isolado. Mutações (ao nível dos aminoácidos) são depois feitas no anticorpo isolado, incrementalmente ou em combinação com outras mutações. Por exemplo é feita uma biblioteca de ácidos nucleicos que inclui sequências codificando algumas das ou todas as mutações da linha germinativa possíveis. Os anticorpos mutados são depois avaliados, p.ex., para identificar um anticorpo que tenha um ou mais resíduos da linha germinativa adicionais em relação ao anticorpo isolado e que seja ainda útil (p.ex., tenha uma atividade funcional). Numa forma de realização são introduzidos

tantos resíduos da linha germinativa num anticorpo isolado quanto possível.

Numa forma de realização é usada mutagénese para substituir ou inserir um ou mais resíduos da linha germinativa numa região de *framework* e/ou constante. Por exemplo, um resíduo da região de *framework* e/ou constante da linha germinativa pode ser de uma sequência da linha germinativa que seja similar (p.ex., o mais similar) à região não variável sendo modificada. Após mutagénese, a atividade (p.ex., ligação ou outra atividade funcional) do anticorpo pode ser avaliada para determinar se o resíduo ou resíduos da linha germinativa são tolerados (i.e., não abrogam a atividade). Pode ser realizada mutagénese similar nas regiões de *framework*.

A seleção de uma sequência da linha germinativa pode ser realizada em diferentes modos. Por exemplo, uma sequência da linha germinativa pode ser selecionada se cumprir um critério pré-determinado quanto a seletividade ou similaridade, p.ex., pelo menos uma certa percentagem de identidade, p.ex., pelo menos 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 99,5% de identidade. A seleção pode ser realizada usando pelo menos 2, 3, 5, ou 10 sequências da linha germinativa. No caso de CDR1 e CDR2, a identificação de uma sequência da linha germinativa similar pode incluir seleção de uma tal sequência. No caso de CDR3, a identificação de uma sequência da linha germinativa similar pode incluir seleção de uma tal sequência, mas pode incluir usado de duas sequências da linha germinativa que contribuem separadamente para a porção amino-terminal e a porção carboxi-terminal. Noutras implementações podem ser

usadas mais do que uma ou duas sequências da linha germinativa, p.ex., para formar uma sequência de consenso.

Diversidade de CDR1, CDR2, e cadeias leves

É para ser entendido que as bibliotecas de CDR3 de HC são construídas no fundo da diversidade em CDR1 de HC, CDR2 de HC, e cadeias leves. A diversidade de cadeias leves pode ser codificada na mesma molécula de ADN como a diversidade de HC ou as diversidades de LC e HC podem ser codificadas em moléculas de ADN separadas. Na Tabela 22, a fusão de uma sequência de sinal::VH::CH1::His6::Myc::IIIstump. CDR1 compreende os resíduos 31-35; existe diversidade nos resíduos 31, 33, 35. Numa forma de realização, os resíduos 31, 33, e 35 podem ser qualquer tipo de aminoácido exceto cisteína. CDR2 compreende os resíduos 50 até 65. Existe diversidade nas posições 50, 52, 52a, 56, e 58. Numa forma de realização, os resíduos 50, e 52 podem ser qualquer um dos tipos Ser, Gly, Val, Trp, Arg, Tyr; o resíduo 52a pode ser Pro ou Ser e os resíduos 56 e 58 podem ser qualquer tipo de aminoácido exceto Cys. A diversidade de CDR3 de HC é clonada na diversidade de CDR1 e 2 de HC que é pelo menos $1,0 \times 10^4$, $1,0 \times 10^5$, $1,0 \times 10^6$, $1,0 \times 10^7$, $5,0 \times 10^7$, ou $1,0 \times 10^8$.

Numa forma de realização, os resíduos 31, 33, 35, 50, 52, 56, e 58 podem ser qualquer tipo de aminoácido exceto Cys ou Met e o resíduo 52a pode ser Gly, Ser, Pro, ou Tyr. A diversidade de CDR3 de HC é clonada na diversidade de CDR1 e 2 de HC que é pelo menos $1,0 \times 10^4$, $1,0 \times 10^5$, $1,0 \times 10^6$, $1,0 \times 10^7$, $5,0 \times 10^7$, ou $1,0 \times 10^8$.

Numa forma de realização, a diversidade da HC é clonada num vetor (fago ou fagemídeo) que contém uma diversidade de cadeias leves. Esta diversidade é pelo menos 25, 50, 100,

500, $1,0 \times 10^3$, $1,0 \times 10^4$, $1,0 \times 10^5$, $1,0 \times 10^6$, ou $1,0 \times 10^7$. A diversidade de CDR3 de HC é pelo menos 221, 272, 500, 1000, $1,0 \times 10^4$, $1,0 \times 10^5$, $1,0 \times 10^6$, $1,0 \times 10^7$, $1,0 \times 10^8$, ou $1,0 \times 10^9$.

Numa forma de realização, a diversidade da HC é clonada num vetor de fago que exhibe a HC numa proteína de fago tal como III, VIII, VII, VI, ou IX ou um fragmento de uma destas suficiente para causar exibição e as cadeias leves são combinadas com a HC por infecção de uma coleção de células em que cada célula secreta uma cadeia leve. A diversidade das cadeias leves nas células é pelo menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, ou 100. A diversidade de CDR3 de HC é pelo menos 221, 272, 500, 1000, $1,0 \times 10^4$, $1,0 \times 10^5$, $1,0 \times 10^6$, $1,0 \times 10^7$, $1,0 \times 10^8$, ou $1,0 \times 10^9$.

A Tabela 30 mostra a sequência do vetor de fago DY3FHC87 (SEQ ID NO:894) que transporta um gene *bla*, uma cassete de exibição para cadeias pesadas sob controlo de um promotor P_{lac} . DY3FHC87 contém todos os genes de M13 também. A infecção de células *E. coli* F^+ que abrigam uma diversidade de cadeias leves num vetor tal como pLCSK23 (Sequência na Tabela 40) (SEQ ID NO:896). O vetor pLCSK23 transporta um gene *Kan^R*. Sob o controlo do promotor P_{lac} existe um gene começando na base 2215 tendo uma sequência de sinal (bases 2215-2277), uma VL (nesta sequência, a VL codifica a sequência mostrada em (SEQ ID NO:897) da base 2278 à base 2598, Ckappa da base 2599 à 2922, um ligante que permite um local NotI de 2923 a 2931, e um marcador V5 (bases 2932-2973). Existem um local SfiI a 2259-2271 e um local *KpnI* a 2602-2605 para permitir substituição fácil de Vkappas. (SEQ ID NO:897) é um exemplo das proteínas que são secretadas. É para ser entendido que Ckappa e o marcador V5 são constantes. Todas as proteínas mostradas na Tabela 19

(VK102gl-JK3, VK102var1, VK102var2, VK102var3, VK102var4, VK102var5, VK3L6gl-JK4, VK3L6var1, VK3L6var2, VK3L6var3, VK3L6var4, VK3L6var5, VK3L6var6, VK3L6var7, VK3L6var8, VK3A27gl-JK3, VK3A27var1, VK3A27var2, VK3A27var3, VK3A27var4, VK3A27var5, VK3A27var6, VK3A27var7, VK3L2gl-JK3, e VK1gIL8-JK5) terão estas sequências anexadas à extremidade de carboxi.

Diversidade de Cadeias Leves

A Tabela 800 mostra uma LC (cadeia leve) kappa que se sabe que emparelha bem com 3-23 e com cinco mutações de CDR com uma HC baseada em 3-23, LC K1(012)::JK1 faz um Ab de elevada afinidade para um alvo de proteína. 012 é um VKI frequentemente usado. O gene foi desenhado para ter locais de restrição distintos, úteis na sequência de sinal (ApaLI), FR1 (XhoI, SgfI), FR2 (KpnI), FR3 (XbaI), e Fr4::Ckappa (BsiWI) tal que cada CDR e seja substituída por uma população variada.

Em LCs de humano, CDR3 é a mais importante e CDR1 é a segunda mais importante. CDR2 raramente faz contacto com o Ag. A diversidade é introduzida nas CDRs como mostrado nas Tabela 900 e Tabela 1000 (CDR1), Tabela 1100 e Tabela 1200 (CDR2), Tabelas 1300, 1400, e 1500 (CDR3). Para Seleção Económica de Cadeias Pesadas (ESHC), um pequeno número, por exemplo, 50 LCs com diversidade em CDR3 como na Tabela 1200 é escolhido para expressão em pLCSK24 para secreção no periplasma. Podem ser usadas mais LCs se forem mantidas várias linhas celulares tal que cada linha celular contenha, por exemplo, 50 ou menos LC.

A Tabela 900 mostra diversidade para CDR1 de LC. A biblioteca pode conter o resíduo 012 com a diversidade

adicionada de tipos de AA mostrados como “permitidos”; lendo-se “permitidos” como “tipos permitidos adicionais” nas Tabelas 900, 1000, 1100, 1200, 1300, 1400. 012 tem $R_{24}ASQSISSYLN_{34}$. Outros *loci* de VK1 têm Q em 24. Outros *loci* têm M em 25. S26 e Q27 são invariantes em VKI. Outros *loci* de VKI têm D ou G em 28. I29 e L33 são invariantes em VKI e os grupos laterais estão orientados para a frente. Outros *loci* de VKI permitem a diversidade mostrada na Tabela 900 nas posições 30, 31, 32, e 34. Na Tabela 900, somente sete das onze posições são variadas e a diversidade total é 576.

A Tabela 1000 mostra um nível mais elevado de diversidade para CDR1 de LC. Aqui, 8 de 11 posições foram variadas. Aquelas que são constantes estão longe do local de combinação ou têm grupos laterais enterrados.

A Tabela 1100 mostra um baixo nível de variação para CDR2. CDR2 está longe do local de combinação com o antigénio e a diversidade aqui pode não ser muito útil. De facto, a diversidade de GL é muito limitada. A Tabela 1100 inclui a diversidade de GL. A Tabela 1200 contém um nível mais elevado de diversidade, 1920 sequências permitidas.

A Tabela 1300 mostra um baixo nível de diversidade para CDR3 de LC, 2160 sequências. A Tabela 1400 mostra um nível mais elevado que permite 105.840 sequências.

Para ROLIC são produzidas cerca de 3×10^7 LC tendo a diversidade mostrada nas Tabelas 900, 1100, e 1300.

Diversidade de Cadeias Pesadas

As HC (cadeia pesada) de Ab têm diversidade em CDR1, CDR2, e CDR3. A diversidade em CDR3 é especialmente complexa

porque existe diversidade de sequências e comprimentos. A diversidade de sequências não é aleatória. As células fabricando genes de Ab unem um segmento V a um segmento D a um segmento JH. O segmento D é opcional; cerca de metade dos Abs de humano naturais têm um D reconhecível. Pode existir modificação extensa das fronteiras V-D, D-J, ou V-J com nenhuma a muitas bases adicionadas ou removidas. Um Ab que tenha uma linha germinativa de V::D::JH poderia ser visto como um Ab da linha germinativa.

Segmentos D de humano são mostrados na Tabela 21. Cada segmento D da linha germinativa (GL) pode aparecer num gene de Ab em qualquer uma das três grelhas de leitura aberta. Em algumas grelhas de leitura, alguns dos segmentos D codificam codões de paragem. Estes segmentos D ocorrem raramente com o codão de paragem modificado. A Tabela 600 mostra a frequência de cada segmento D como uma percentagem de todos os segmentos D observados. A maioria dos exemplos aqui que contêm segmentos D usa Ds que são relativamente comuns (> 2% de todos os Ds observados).

Num aspeto, a presente divulgação envolve composição de genes de HC de Ab por fusão de 3-23 (ou outra VH, tal como 4-34) a um de a) um número de aminoácidos escolhidos do conjunto compreendendo (S, Y, D, R, N), b) uma região D, c) uma região JH, e d) a porção FR4 de uma região JH. Estas fusões podem ser uma GL 3-23 ou uma 3-23 que tem diversidade sintética em CDR1 e/ou CDR2. Os comprimentos da CDR3 de HC podem ser qualquer número de cerca de 3 a cerca de 24. Preferencialmente, a biblioteca conteria membros com CDR3 de HC de comprimentos 6, 8, 10, 12, 14, 16, 18, e 20. Alternativamente, os comprimentos poderiam ser 5, 8, 11, 14, 17, e 20 ou qualquer outra combinação.

A Tabela 21 mostra um número de exemplos de desenhos de CDR3s adequadas com comprimentos de 6 a 20. Os codões que especificam as letras maiúsculas na coluna 2 são para ser sintetizados com *wobbling*. A coluna 3 mostra o nível de *doping*. A Tabela 100 mostra razões nas quais os vários comprimentos de CDR3 de HC poderiam ser combinados para formar uma biblioteca que se espera que contenha Abs que se ligam a quase todos os alvos de proteína.

Tabela 100							
Comprimento	6	8	10	12	14	16	20
Diversidade	1,0 × 10 ⁵	2,0 × 10 ⁵	4,0 × 10 ⁵	8,0 × 10 ⁵	8,0 × 10 ⁵	8,0 × 10 ⁵	4,0 × 10 ⁵

Para o comprimento 6, Tabela 21 são dados quatro exemplos. Por exemplo, 6a tem VH(3-23) unida diretamente a JH1 com os primeiros seis AAs *wobbled*, 6b tem Tyr unida a D4-17 na segunda grelha de leitura unida aos AAs de FR4 de JH1, e 6c tem D5-5(3) unida aos resíduos de FR de JH1. Uma vez que estes dão diferentes tipos de diversidade, a inclusão de todos é preferencial, mas uma biblioteca contendo somente um destes deve dar Abs úteis.

Para comprimento 8, a Tabela 21 mostra três exemplos. 8a tem YY fundida com a totalidade de JH1 enquanto 8b tem uma Y fundida com D6-13(1) fundida com a região FR de JH1. Os comprimentos 10, 12, 14, 16, e 20 são também mostrados na Tabela 21. A diversidade de CDR3 de HC poderia ser construída numa linha germinativa 3-23 ou 3-23 contendo diversidade sintética. Alternativamente poderia ser usada uma VH diferente, tal como 4-34.

ROLIC é um método no qual uma pequena população de HCs é expressa em *E. coli* F⁺ como proteínas solúveis. A população é infectada com fagos que transportam fusões LC::III_{stump}. Os fagos produzidos obtêm uma HC do periplasma da célula que as produz. Estes fagos podem ser ligados a alvo imobilizado e os ligantes são separados dos não ligantes. O tamanho da população é importante porque, quando os fagos recuperados são propagados, os fagos recuperados têm de encontrar o mesmo tipo de célula de onde veio para continuar a associação entre LC e HC. Assim sendo é desejável que o número de HC seja pequeno em cada linha de células. Assim sendo pode ser desejável manter um número de linhas de células com até 10, 20, 30, ou 40 HCs diferentes em cada linha de células. Assim sendo podemos ter 1, 2, 4, 6, 8, 10, 24, 48, ou 96 linhas de células e realizamos o mesmo número de produções, seleções, e amplificações de fagos paralelas. Após uma ou duas rondas testamos colônias quanto à produção de fagos que se ligam ao alvo por um ensaio ELISA. Cada colônia ELISA⁺ contém uma LC útil e uma HC útil, mas não estão no mesmo pedaço de ADN. Não obstante sabemos o início e final de cada LC e cada HC e podemos portanto usar PCR na colônia para produzir uma cassete de exibição de Fab ou secreção de Fab que pode ser colocada num fago ou fagemídeo de exibição ou num plasmídeo de produção de Fab.

Na Seleção Eficaz de HCs (ESHC) revertermos os papéis de LC e HC em ROLIC e temos LCs num plasmídeo tal que sejam produzidas como proteínas solúveis no periplasma de *E. coli* F⁺. Produzimos a diversidade de HC num vetor de fago que não tem nenhum gene de LC. Infetamos a *E. coli* F⁺ produzindo LC com fagos transportando HC. Obtemos fagos que transportam um gene de HC e ambas as proteínas HC e LC.

Selecionamos estes fagos para ligação ao alvo. Em muitos Abs, a LC é permissiva e não contribui grandemente para a afinidade de ligação. A escolha da melhor LC pode aumentar grandemente a afinidade, mas é usualmente possível selecionar um Fab com um repertório muito limitado de LCs. Assim sendo colocamos um pequeno conjunto de LCs, preferencialmente linha germinativa nas regiões de *framework* na *E. coli* F⁺ produzindo LC. Se existirem, por exemplo, 25 LC na linha celular de LC, então obtemos uma redução de 25 vezes no número de transformantes de células que necessitam de ser feitos.

As bibliotecas descritas têm uma gama de comprimentos de CDR3 de HC. Para favorecer a dobragem apropriada, as CDR3 de HC têm um segmento D ou nenhum segmento D unido à maioria do, todo o, ou a porção de *framework* de um segmento. As sequências são diversificadas por uso de síntese de ADN *wobble*. Embora isto permita teoricamente qualquer tipo de aminoácido em qualquer posição, na prática, as sequências reais estão fortemente enviesadas na direção das sequências parente e tipos de AA que estão próximos na tabela do código genético.

Por uso de ESHC podemos amostrar novos desenhos de diversidade de CDR3 de HC sintéticos. Nos exemplos dados usamos um agrupamento de, por exemplo, 50 LCs. Uma biblioteca de 5×10^8 HC deve desempenhar-se tão bem como uma biblioteca de estilo antigo de $2,5 \times 10^{10}$ mas requer muito menos esforço.

Quando *wobbling* uma sequência, a escolha dos códons iniciais afeta a mistura real de AAs vista na biblioteca. A Tabela 300 mostra quais as substituições de aminoácidos que

requerem 1, 2, ou 3 mudanças de bases de cada codão parente de início. Por exemplo, se começarmos com gct ou gcc para Ala, todos os três codões de paragem requerem três mudanças de bases e logo são raros. Se se usar misturas 76:8:8:8, Ala aparecerá em 57% dos casos ($0,76 \times 0,76$). V, G, T, P, S aparecerão cada um em cerca de 6% e D cerca de 3%. E, I, L, F, Y, H, N, C, e R estarão para baixo cerca de 10 vezes. M, W, Q, K, Am, Oc, e Op serão ainda mais raros. Se começássemos com gca, então E substituiria D na necessidade de somente uma mudança de bases, mas as paragens opal e ochre requerem somente duas mudanças de bases, o que é indesejável. Os codões preferenciais estão marcados com um asterisco (*). A escolha para serina é complicou o nosso desejo de ter Y como substituição de S com elevada frequência. Isto traz Op e Oc ao grupo que diferem do parente por somente duas bases. Este problema pode ser ultrapassado por clonagem do repertório de CDR3 de HC antes de um gene de resistência a antibióticos tais como *Kan^R* ou *Amp^R* e seleção quanto a resistência, eliminando deste modo os membros que contêm codões de paragem. Adicionalmente, a biblioteca pode ser produzida em *E. coli* supE que inserem Q em vez de paragem.

Tabela 300

Aminoácido	Codão parente	1 mudança de bases	2 mudanças de bases	3 mudanças de bases
A *	gct, gcc	V, D, G, T, P, S	E, I, L, F, Y, H, N, C, R	M, W, Q, K, Am, Oc, Op
A	gca	V, E, G, T, P, S	D, I, L, Oc, Q, K, Op, R	M, W, H, N, C, Am, F, Y
A	gcg	V, E, G, T, P, S	D, M, L, Am, Q, K, R, W	I, F, Y, Oc, Op, H, N, C

Aminoácido	Codão parente	1 mudança de bases	2 mudanças de bases	3 mudanças de bases
C	tgt, tgc	Y, S, F, W, Op, R, G	L, H, N, D, P, T, A, V, I	Am, Oc, Q, K, E, M
D	gat, gac	E, G, A, V, N, H, Y	F, S, C, L, P, Q, K, R, Oc, Am, I, T	M, W, Op
E	gaa	D, G, A, V, K, Q, Oc	Am, L, I, S, P, T, R, Op, Y, H, N	M, F, C, W
E *	gag	D, G, A, V, K, Q, Am	M, L, S, P, T, Y, H, N, Oc, R, W	F, C, I, Op
F	ttt, ttc	L, I, V, S, Y, C	M, Am, Op, Oc, W, P, T, A, H, N, D, R, G	Q, K, E
G *	ggt, ggc	D, A, V, S, R, C	E, W, F, L, I, T, P, Y, H, N	Am, Oc, Op, M, Q, K
G	gga	E, A, V, R, Oc	D, W, L, I, S, P, T, Op, Q, K	Am, Oc, M, F, Y, H, N
G	ggg	E, A, V, R, W	D, Oc, L, M, S, P, T, Am, Op, Q, K	Oc, I, F, Y, H, N
H	cat, cac	Q, Y, N, D, L, P, R	F, S, C, I, T, V, A, D, G, Am, Oc	Op, W, M, E
I *	att, atc	M, L, F, V, T, N, S	Y, C, P, H, R, A, D, G	Am, Op, Oc, W, Q, K, E
I	ata	M, L, V, T, K, R	Op, Oc, S, P, Q, A, E, G, F, N	Am, C, D, H, W, Y
K	aaa	N, Q, Oc, E, P, I, R	H, Y, D, M, L, V, S, T, A, Am, Op, G	C, F, W
K *	aag	N, Q, Am, E, P, M, R	H, Y, D, I, L, V, S, T, A, Oc, G, W	C, F, Op
L	tta	F, S, Oc, Op, I, V	Y, C, W, M, P, T, A, Q, K, E, R, G, Am	D, H, N
L	ttg	F, S, Am, W,	Y, C, Oc, Op, P, T,	D, H, N

Aminoácido	Codão parente	1 mudança de bases	2 mudanças de bases	3 mudanças de bases
		M, V	A, Q, K, E, R, G, I	
L *	ctt, ctc	F, I, V, P, H, R	M, S, Y, C, T, N, A, D, G	Am, Oc, Op, W, E, K, Q
L	cta	I, V, P, Q, R	F, M, S, Oc, Op, T, K, A, E, G, H	Am, W, D, N, C, Y
L	ctg	M, V, P, Q, R	F, I, S, Am, T, K, A, E, G, H, W	Oc, Op, D, N, C, Y
M	atg	L, V, T, K, R, I	F, N, S, P, A, Am, Q, E, W, G	Oc, Op, Y, C, H, D
N	aat, aac	K, Y, H, D, I, T, S	F, C, L, P, R, V, A, G, M, Q, E, Am, Oc	Op, W
P*	cct, ccc	S, T, A, L, H, R	F, Y, C, I, N, V, D, G, Q	Am, Oc, Op, W, M, E, K
P	cca	S, T, A, L, Q, R	Oc, Op, I, K, V, E, G, H	Am, W, M, D, N, C, F, Y
P	ccg	S, T, A, L, Q, R	Am, M, K, V, E, G, H	C, D, F, I, N, W, Y, Oc, Op
Q	caa	Oc, K, E, R, P, L, H	Y, Am, N, D, S, T, A, I, V, G, Op	F, C, W, M
Q*	cag	H, Am, K, E, R, P, L	N, D, Y, M, T, V, A, G, W, Oc, S	C, F, Op, I
R*	cgt, cgc	C, S, G, H, P, L	Op, W, Q, F, Y, I, T, N, V, A, D	Am, Oc, M, E, K
R	cga	G, Op, Q, P, L	Oc, S, C, W, H, I, V, T, A, E, K	Am, M, C, D, N, F, Y
R	cgg	G, W, Q, P, L	Am, Op, S, M, V, T, A, K, E, H, C	F, Y, I, Oc, D, N
R	aga	G, Op, S, K, T, I	C, W, N, M, L, V, P, A, Oc, Q, E	F, Y, H, D, Am
R	agg	G, W, S, K,	C, Op, Am, L, I, V,	F, Y, H, D, Oc

Aminoácido	Codão parente	1 mudança de bases	2 mudanças de bases	3 mudanças de bases
		T, M	A, Q, P, E, N	
S *	tct, tcc	F, Y, C, P, T, A	L, Oc, Op, Am, W, I, V, N, D, R, G, H	E, K, M, Q
S	tca	L, Oc, Op, P, T, A	F, Y, C, W, Q, R, I, K, V, E, G, Am	M, W, D, N, H
S	tcg	L, Am, W, P, T, A	F, Y, C, Op, Oc, Q, R, M, K, V, E, G	I, D, N, H
S	agt, agc	C, R, G, N, T, I	F, Y, L, P, H, V, A, D, K, W, Op	Am, Oc, M, E, Q
T*	act, acc	S, P, A, I, N	F, Y, C, L, H, R, M, K, V, D, G	Am, Oc, Op, W, E, Q
T	aca	S, P, A, I, K, R	L, Oc, Op, Q, M, E, G, V, N	F, Y, C, Am, W, D, H
T	acg	S, P, A, M, K, R	I, N, L, Am, W, Q, V, E, G	C, F, Y, Oc, Op, D, H
V*	gtt, gtc	F, L, I, A, D, G	S, P, T, Y, H, N, E, C, R, M	Am, Oc, Op, W, Q, K
V	gta	L, I, A, E, G	F, M, D, S, P, T, Oc, Op, Q, R, K	Am, W, C, Y, H, N
V	gtg	L, M, A, E, G	F, I, D, S, P, T, Am, Q, R, K, W	Oc, Op, C, Y, H, N
W	tgg	C, R, G, Am, S, L, Op	P, Q, F, M, T, K, V, A, E, Oc, Y	D, N, H, I
Y	tat, tac	C, S, F, N, H, D, Oc, Am	L, W, Q, K, E, P, I, T, V, A, G, Op, R	M
Am é paragem TAG, Op é TGA, Oc é TAA				

Métodos de Uso das Bibliotecas

Seleção por Dissociação. Uma vez que uma taxa de dissociação lenta pode ser preditora de elevada afinidade, particularmente no que diz respeito a interações entre polipéptidos e os seus alvos, os métodos descritos aqui podem ser usados para isolar ligandos com uma taxa de dissociação cinética desejada (*i.e.*, reduzida) para uma interação de ligação a um alvo.

Para seleccionar anticorpos de dissociação lenta de uma biblioteca de exibição, a biblioteca é contactada com um alvo imobilizado. O alvo imobilizado é depois lavado com uma primeira solução que remove anticorpos não especificamente ou fracamente ligados. Depois, os anticorpos ligados são eluídos com uma segunda solução que inclui uma quantidade saturante de alvo livre, *i.e.*, réplicas do alvo que não estão anexadas à partícula. O alvo livre liga-se a anticorpos que se dissociam do alvo. A religação dos anticorpos eluídos é eficazmente prevenida pela quantidade saturante do alvo livre em relação à concentração muito mais baixa do alvo imobilizado.

A segunda solução pode ter condições de solução que sejam substancialmente fisiológicas ou que sejam estridentes (p.ex., baixo pH, elevado pH, ou muito sal). Tipicamente, as condições de solução da segunda solução são idênticas às condições de solução da primeira solução. Frações da segunda solução são recolhidas por ordem temporal para se distinguirem frações iniciais de tardias. As frações tardias incluem anticorpos que se dissociam a uma taxa mais lenta do alvo do que biomoléculas nas frações iniciais. Adicionalmente é também possível recuperar anticorpos que

permanecem ligados ao alvo mesmo após incubação prolongada. Estes podem ser dissociados usando condições caotrópicas ou podem ser amplificados enquanto anexados ao alvo. Por exemplo, os fagos ligados ao alvo podem ser contactados com células bacterianas.

Seleção ou Rastreio quanto a Especificidade. Os métodos de rastreio de biblioteca de exibição descritos aqui podem incluir um processo de seleção ou rastreio que descarta anticorpos que se ligam a uma molécula não alvo. Exemplos de moléculas não alvo incluem, p.ex., uma molécula de carboidrato que difere estruturalmente da molécula alvo, p.ex., uma molécula de carboidrato que tem uma propriedade biológica diferente da molécula alvo. No caso de um carboidrato sulfatado, um não alvo pode ser o mesmo carboidrato sem o sulfato ou com o sulfato numa posição diferente. No caso de um fosfopéptido, o não alvo pode ser o mesmo péptido sem o fosfato ou um fosfopéptido diferente.

Numa implementação, um assim chamado passo de "seleção negativa" é usado para discriminar entre a molécula alvo e não alvo relacionada e uma molécula não alvo distinta, relacionada. A biblioteca de exibição ou um seu agrupamento é contactado com a molécula não alvo. Os membros que não se ligam ao não alvo são recolhidos e usados em seleções subsequentes para ligação à molécula alvo ou mesmo para seleções negativas subsequentes. O passo de seleção negativa pode ser antes da ou após a seleção de membros da biblioteca que se ligam à molécula alvo.

Noutra implementação é usado um passo de rastreio. Após os membros da biblioteca de exibição serem isolados para ligação à molécula alvo, cada membro da biblioteca isolado

é testado quanto à sua capacidade de se ligar a uma molécula não alvo (p.ex., um não alvo listado acima). Por exemplo pode ser usado um rastreio de ELISA de elevada produtividade para se obterem estes dados. O rastreio de ELISA pode ser também usado para se obterem dados quantitativos para ligação de cada membro da biblioteca ao alvo. Os dados de ligação não alvo e alvo são comparados (p.ex., usando um computador e *software*) para se identificarem membros que se ligam especificamente ao alvo.

Em certas formas de realização, os anticorpos compreendendo as CDR3s da invenção podem ser capazes de se ligarem a carboidratos. Métodos para avaliação de anticorpos quanto a ligação de carboidratos incluem ELISA, imunohistoquímica, imunotransferência, e separação de células ativada por fluorescência. Estes métodos podem ser usados para identificar anticorpos que tenham um K_D de melhor do que um limite, p.ex., 100 nM, 50 nM, 10 nM, 5 nM, 1 nM, 500 pM, 100 pM, ou 10 pM.

ELISA. As proteínas codificadas por uma biblioteca de exibição podem ser também rastreadas quanto a uma propriedade de ligação usando um ensaio ELISA. Por exemplo, cada proteína é contactada com uma placa de microtitulação cuja superfície do fundo foi revestida com o alvo, p.ex., uma quantidade limitante do alvo. A placa é lavada com tampão para remover polipéptidos não especificamente ligados. Depois, a quantidade da proteína ligada à placa é determinada por sondagem da placa com um anticorpo que consegue reconhecer o polipéptido, p.ex., um marcador ou porção constante do polipéptido. O anticorpo é ligado a uma enzima tal como fosfatase alcalina, que produz um produto calorimétrico quando são proporcionados substratos

apropriados. A proteína pode ser purificada a partir de células ou avaliada num formato de biblioteca de exibição, p.ex., como uma fusão com um revestimento de bacteriófago filamentosos. Alternativamente, as células (p.ex., vivas ou fixadas) que expressam a molécula alvo, p.ex., um alvo que contém uma fração carboidrato, podem ser colocadas numa placa de microtitulação e usadas para testar a afinidade dos péptidos/anticorpos presentes na biblioteca de exibição ou obtidos por seleção a partir da biblioteca de exibição.

Noutra versão do ensaio ELISA, cada polipéptido de uma biblioteca de cadeias de diversidade é usado para revestir um poço diferente de uma placa de microtitulação. O ELISA procede depois usando uma molécula alvo constante para questionar cada poço.

Ensaio de Ligação a Células. Os anticorpos podem ser avaliados quanto à sua capacidade de interagirem com um ou mais tipos de células, p.ex., uma célula hematopoiética. A separação de células ativada por fluorescência (FACS) é um método exemplar para teste de uma interação entre uma proteína e uma célula. O anticorpo é marcado diretamente ou indiretamente com um fluoróforo, antes de ou após ligação às células, e depois as células são contadas num separador FACS.

Outros tipos de células podem ser preparados para FACS por métodos conhecidos na técnica.

Ensaio de Ligação Homogéneos. A interação de ligação do polipéptido candidato com um alvo pode ser analisada usando um ensaio homogéneo, i.e., após todos os componentes do ensaio serem adicionados, manipulações de fluidos

adicionais não são requeridas. Por exemplo pode ser usada transferência de energia por ressonância de fluorescência (FRET) como um ensaio homogêneo (ver, por exemplo, Lakowicz *et al.*, Pat. dos E.U.A. No. 5,631,169; Stavrianopoulos, *et al.*, Pat. dos E.U.A. No. 4,868,103). Um marcador de fluoróforo na primeira molécula (p.ex., a molécula identificada na fração) é selecionado tal que a sua energia fluorescente emitida possa ser absorvida por um marcador fluorescente na segunda molécula (p.ex., o alvo) se a segunda molécula estiver na proximidade da primeira molécula. O marcador fluorescente na segunda molécula floresce quando absorbe a energia transferida. Uma vez que a eficácia da transferência de energia entre os marcadores está relacionada com a distância separando as moléculas, a relação espacial entre as moléculas pode ser avaliada. Numa situação na qual ocorre ligação entre as moléculas, a emissão fluorescente do marcador de molécula "aceitadora" no ensaio deve ser máxima. Um evento de ligação que é configurado para monitorização por FRET pode ser convenientemente medido através de meios de deteção fluorométrica padrão bem conhecidos na técnica (p.ex., usando um fluorímetro). Por titulação da quantidade da primeira ou segunda molécula de ligação pode ser gerada uma curva de ligação para estimar a constante de ligação de equilíbrio.

Outro exemplo de um ensaio homogêneo é Alpha Screen (Packard Bioscience, Meriden Conn.). Alpha Screen usa duas esférulas marcadas. Uma esférula gera oxigénio em singlete quando excitada por um *laser*. A outra esférula gera um sinal de luz quando o oxigénio em singlete se difunde da primeira esférula e colide com ela. O sinal é somente gerado quando as duas esférulas estão em proximidade. Uma

esférula pode ser anexada ao membro da biblioteca de exibição, a outra ao alvo. Os sinais são medidos para determinar a extensão da ligação.

Os ensaios homogêneos podem ser realizados enquanto o polipéptido candidato está anexado ao veículo da biblioteca de exibição, p.ex., um bacteriófago.

Ressonância de Plasmão de Superfície (SPR). A interação de ligação de uma molécula isolada de uma biblioteca de exibição e um alvo pode ser analisada usando SPR. SPR ou Análise de Interação Biomolecular (BIA) deteta interações bioespecíficas em tempo real, sem marcação de qualquer um dos intervenientes. Mudanças na massa à superfície de ligação (indicativas de um evento de ligação) do *chip* de BIA resultam em alterações do índice de refração da luz próximo da superfície (o fenómeno ótico da ressonância de plasmão de superfície (SPR)). Mudanças na refratividade geram um sinal detetável, que são medidas como uma indicação de reações em tempo real entre moléculas biológicas. Métodos para uso de SPR são descritos, por exemplo, na Pat. dos E.U.A. No. 5,641,640; Raether (1988) *Surface Plasmons* Springer Verlag; Sjolander e Urbaniczky (1991) *Anal. Chem.* 63:2338-2345; Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705 e recursos online proporcionados pela BIAcore International AB (Uppsala, Suécia).

A informação de SPR pode ser usada para proporcionar uma medição precisa e quantitativa da constante de dissociação de equilíbrio (K_D), e parâmetros cinéticos, incluindo k_{on} e k_{off} , para a ligação de uma biomolécula a um alvo. Tais dados podem ser usados para se compararem diferentes

biomoléculas. Por exemplo, as proteínas codificadas por ácido nucleico selecionado de uma biblioteca de cadeias de diversidade podem ser comparadas para se identificarem indivíduos que tenham elevada afinidade para o alvo ou que tenham uma k_{off} lenta. Esta informação pode ser também usada para se desenvolverem relações estrutura-atividade (SAR). Por exemplo, os parâmetros cinéticos e de ligação de equilíbrio de versões maduras de uma proteína parente podem ser comparados com os parâmetros da proteína parente. Podem ser identificados aminoácidos variantes a dadas posições que se correlacionem com parâmetros de ligação particulares, p.ex., elevada afinidade e k_{off} lenta. Esta informação pode ser combinada com modelação estrutural (p.ex., usando modelação de homologia, minimização de energia, ou determinação da estrutura por cristalografia ou NMR). Como resultado, um entendimento da interação física entre a proteína e o seu alvo pode ser formulado e usado para orientar outros processos de desenho.

Redes de Proteínas. As proteínas identificadas a partir da biblioteca de exibição podem ser imobilizadas num suporte sólido, por exemplo, numa esférula ou numa rede. Para uma rede de proteínas, cada um dos polipéptidos é imobilizado num único local num suporte. Tipicamente, o local é um local bidimensional. Métodos de produção de redes de polipéptidos são descritos, p.ex., em De Wildt et al. (2000) *Nat. Biotechnol.* 18:989-994; Lueking et al. (1999) *Anal. Biochem.* 270:103-111; Ge (2000) *Nucleic Acids Res.* 28, e3, I-VII; MacBeath e Schreiber (2000) *Science* 289:1760-1763; WO 01/40803 e WO 99/51773A1. Os polipeptídeos para a rede podem ser localizados a elevada velocidade, p.ex., usando dispositivos robóticos comercialmente disponíveis, p.ex., da Genetic MicroSystems

ou BioRobotics. O substrato da rede pode ser, por exemplo, nitrocelulose, plástico, vidro, p.ex., vidro com superfície modificada. A rede pode também incluir uma matriz porosa, p.ex., acrilamida, agarose, ou outro polímero.

Estojo

São também proporcionados estojos para uso quando se leva a cabo um método de acordo com qualquer aspeto da invenção. Os estojos podem incluir os vetores necessários. Um tal vetor terá tipicamente uma origem de replicação para bacteriófago de cadeia simples e contém o ácido nucleico membro sbp ou tem um local de restrição para a sua inserção na região da extremidade 5' da sequência codificante madura de uma proteína da cápside do fago, e com uma sequência codificante líder secretora a montante do referido local que dirige uma fusão do polipéptido exógeno da proteína da cápside com o espaço periplasmático.

São também proporcionados pacotes codificando as CDR3s de HC como definidos acima e polipéptidos compreendendo as CDR3s de HC e seus fragmentos e derivados, obteníveis por uso de qualquer um dos métodos acima definidos. Os derivados podem compreender polipéptidos fundidos com outra molécula tal como uma enzima ou uma cauda de Fc.

O estojo pode incluir um vetor de fago (p.ex., DY3F87HC) que tenha um local para inserção de CDR3s de HC para expressão do polipéptido codificado em forma livre. O estojo pode também incluir um vetor de plasmídeo para expressão de cadeias leves solúveis, p.ex., pLCSK23. O estojo pode também incluir uma linha de células adequada (p.ex., TG1). A diversidade de cadeias leves codificadas por pLCSK23 pode ser 10, 15, 20, 25, 30, ou 50. As LCs na

diversidade podem ser construídas ou escolhidas de modo a terem certas propriedades desejáveis, tais como sendo linha germinativa nas regiões de *framework* e tendo diversidade em CDR3 e/ou CDR1. As linhas germinativas podem ser daquelas altamente utilizadas, p.ex., VK1_2-02, VK3_1-A27, VK3_5-L6, VK3_3-L2 para kappa e VL2_2a2, VL1_1c, VL1_1g, VL3_3r para lambda.

Por exemplo poder-se-iam clonar genes para VK102g1-JK3, VK102var1, VK102var2, VK102var3, VK102var4, VK102var5, VK3L6g1-JK4, VK3L6var1, VK3L6var2, VK3L6var3, VK3L6var4, VK3L6var5, VK3L6var6, VK3L6var7, VK3L6var8, VK3A27g1-JK3, VK3A27var1, VK3A27var2, VK3A27var3, VK3A27var4, VK3A27var5, VK3A27var6, VK3A27var7, VK3L2g1-JK3, VK1g1L8-JK5, e VK1GL012-JK3 (sequências de aminoácidos mostradas na Tabela 19) em pLCSK23.

Tabela 19: 26 VL a serem usadas em pLCSK23.

VK102g1-JK3	(SEQ ID NO:4)	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK102var1	(SEQ ID NO:5) S28D	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK102var2	(SEQ ID NO:6) S91R	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK102var3	(SEQ ID NO:7) S91E	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK102var4	(SEQ ID NO:8) S31R	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK102var5	(SEQ ID NO:9) S31E, S93R	
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQK GKAPKLLIYA ASSLSGVPS	60	
RFSGSGSGTD FTLTISSLQF EDFATYYCQ SYSTPFTFGP GTKVDIK	107	
VK3L6g1-JK4	(SEQ ID NO:10)	
EIVLTQSPAT LSLSPGERAT LSCRASQSV SYLAWYQQK GPAPRLIYD ASNRATGIPA	60	
RFSGSGSGTD FTLTISLEP EDFAVYCCQ ASNNPLTFG GTKVEIK	107	
VK3L6var1	(SEQ ID NO:11) S31R	
EIVLTQSPAT LSLSPGERAT LSCRASQSV SYLAWYQQK GPAPRLIYD ASNRATGIPA	60	
RFSGSGSGTD FTLTISLEP EDFAVYCCQ ASNNPLTFG GTKVEIK	107	
VK3L6var2	(SEQ ID NO:12) S92R	

```

EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RRNWPLTFGG GTKVEIK 107

VK3L6var3 (SEQ ID NO:13) S92G
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RGNWPLTFGG GTKVEIK 107

VK3L6var4 (SEQ ID NO:14) S92Y
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RYNWPLTFGG GTKVEIK 107

VK3L6var5 (SEQ ID NO:15) S92E
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RENNPLTFGG GTKVEIK 107

VK3L6var6 (SEQ ID NO:16) Y32F
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RSNWPLTFGG GTKVEIK 107

VK3L6var7 (SEQ ID NO:17) Y32D
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RSNWPLTFGG GTKVEIK 107

VK3L6var8 (SEQ ID NO:18) N93G
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTD FTLTISSELP EDFAVYYCQ RSGWPLTFGG GTKVEIK 107

VK3A27g1-JK3 (SEQ ID NO:19)
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSPPFTFG PGTKVDIK 108

VK3A27var1 (SEQ ID NO:20) S31R
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSPPFTFG PGTKVDIK 108

VK3A27var2 (SEQ ID NO:21) S32K
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSPPFTFG PGTKVDIK 108

VK3A27var3 (SEQ ID NO:22) S32D
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSPPFTFG PGTKVDIK 108

VK3A27var4 (SEQ ID NO:23) G93E
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYESSPFTFG PGTKVDIK 108

VK3A27var5 (SEQ ID NO:24) G91R
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYRSSPFTFG PGTKVDIK 108

VK3A27var6 (SEQ ID NO:25) S36D, G93E
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYESSPFTFG PGTKVDIK 108

VK3A27var7 (SEQ ID NO:26) S94R
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQOK PGQAPRLLIY GASSRATGIP 60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYCRSPFTFG PGTKVDIK 108

VK3L2g1-JK3 (SEQ ID NO:27)
EIVMTQSPAT LSVSPGERAT LSCRASQSVS SNLAWYQOKP GQAPRLLIYD ASNRATGIPA 60
RFSGSGSGTE FTLTISSELP EDFAVYYCQ YNNWPFITFG GTKVDIK 107

VK1g1L8-JK5 (SEQ ID NO:28)
SIQLTQSPSF LSASVGRVY ITCRASQGIS SYLAWYQOKP GRAPRLLIYA ASTLQSGVPS 60
RFSGSGSGTE FTLTISSELP EDFAVYYCQ LNSYPITFGG GTKLEIK 107

VK1GL012-JK3 (SEQ ID NO:29)
SIQMTQSPFS LSASVGRVY ITCRASQGIS SSYLNMTQOK PGKAPRLLIY AASSLQSGVPS 60
SAFSGSGSGT DFTLTISSELP QPEDPATYYC QQSYETPTFT GPGTKVDIKR GTVAAPSVFI 120
FFFSQGLKS GTASVCLL NNFYPREAKV QMKVDNALGS GNSQSYTEQ QSKDSTYLS 180
STLTLSKADY ENKRVYACE VIKGLSSPV TKSFWRGECA AAGKPIPNPL LGLDST 236

```

Os conjuntos podem incluir componentes auxiliares requeridos para levar a cabo o método, dependendo a natureza de tais componentes do curso do método particular empregue. Componentes auxiliares úteis podem compreender fagos ajudantes, iniciadores de PCR, tampões, e/ou enzimas de vários tipos. Tampões e enzimas são tipicamente usados para permitir preparação de sequências de nucleótidos codificando fragmentos Fv, scFv ou Fab derivados de genes de imunoglobulina rearranjados ou não arranjados de acordo com as estratégias descritas aqui.

Métodos de Introdução de Diversidade

Existem muitos modos de gerar ADN que seja variável. Um modo é usar síntese de nucleótidos mistos (MNS). Uma versão de MNS usa misturas equimolares de nucleótidos como mostrado na Tabela 5. Por exemplo, o uso de codões NNK dá todos os vinte aminoácidos e um codão de paragem TAG. A distribuição é 3(R/S/L): 2(A/G/V/T/P): 1(C/D/E/F/H/I/K/M/N/Q/W/Y) (p.ex., 3 de cada um de Arg, Ser, e Leu, e assim por diante). Uma alternativa, aqui denominada "wobbling", usa nucleótidos mistos mas não em quantidades equimolares. Por exemplo, se um codão parente fosse TTC (codificando Phe) poderíamos usar uma mistura de (0,082 de T, 0,06 de C, 0,06 de A, e 0,06 de G) em lugar de T e uma mistura de (0,082 de C, 0,06 de T, 0,06 de A, e 0,06 de G) em lugar de C. Isto daria TTC ou TTT (codificando Phe) 59% das vezes e Leu 13%, S/V/I/C/Y ~5%, e outros tipos de aminoácidos menos frequentemente.

Van den Brulle *et al.* (*Biotechniques* 45:340-3 (2008)) descrevem um método de síntese de ADN variável no qual enzimas de restrição do tipo II são usadas para transferir

trinucleótidos de um oligonucleótido em gancho de cabelo ancorado (PHONs) para um assim chamado "*splinker*". Ver também patentes EP 1 181 395, EP 1 411 122, EP 1 314 783 e pedidos EP 01127864.5, EP 04001462.3, EP 08006472.8. Por uso de misturas de PHONs ancorados e *splinkers* podem-se construir bibliotecas nas quais tipos de aminoácidos desejados sejam permitidos em razões determinadas pelo desenhador. Assim sendo pode-se decretar que esse tipo de aminoácidos esteja presente, por exemplo 82% das vezes e 18 outros tipos de aminoácidos (todos os tipos de aminoácidos não parente exceto Cys) estejam presentes a 2% cada um. Aqui referir-nos-emos a uma síntese como "*dobbling*" (*wobbling* digital). Em alguns aspetos, o *dobbling* é preferencial ao *wobbling*, mas o *wobbling* proporciona formas de realização úteis, parcialmente porque a estrutura da tabela do código genético faz com que o *wobbling* faça substituições maioritariamente conservativas. O *dobbling* oferece a possibilidade de excluir tipos de aminoácidos indesejados. Em CDRs, as cisteínas não emparelhadas são conhecidas, mesmo em Abs aprovados como terapêuticos, mas, em algumas formas de realização gostar-se-ia de as evitar. Em algumas formas de realização, quando se diversifica uma região D que contenha um par de cisteínas, não se permite que as cisteínas variem porque a ansa de dissulfeto é um elemento estrutural importante e porque não se deseja cisteínas não emparelhadas.

Adicionalmente pode-se sintetizar uma molécula de ADN que codifique uma sequência de aminoácidos parente e sujeitar tal ADN a PCR propenso a erros usando iniciadores que abrangem as regiões de *framework* tal que mutações nas regiões de *framework* sejam evitadas.

Tabela 5: Códigos padrão para nucleótidos mistos

N é equimolar A, C, G, T	
B é equimolar C, G, T	(não A)
D é equimolar A, G, T	(não C)
H é equimolar A, C, T	(não G)
V é equimolar A, C, G	(não T)
K é equimolar G, T	(Ceto)
M é equimolar A, C	(aMino)
R é equimolar A, G	(puRina)
S é equimolar C, G	(Forte)
W é equimolar A, T	(fraco)
Y é equimolar C, T	(pIrimidina)

Tabela 6: Exemplo de nucleótidos mistos para *wobbling*

e = 0,82 de A + 0,06 de C + 0,06 de G + 0,06 de T
q = 0,06 de A + 0,82 de C + 0,06 de G + 0,06 de T
j = 0,06 de A + 0,06 de C + 0,82 de G + 0,06 de T
z = 0,06 de A + 0,06 de C + 0,06 de G + 0,82 de T

Exemplificação

A presente invenção é adicionalmente ilustrada pelos seguintes exemplos que não devem ser interpretados como limitantes de qualquer modo.

Exemplo Profético 1: Bibliotecas Com CDR3s de HC Muito Curtas

CDR3s de HC muito curtas foram descritas na técnica. Kadirvelraj *et al.* (2006) *Proc. Natl. Acad. Sci. USA* 103:8149-54 descreveram uma sequência de CDR3 de HC de quatro aminoácidos num anticorpo que se liga a Ag de Tipo B II de *Streptococcus* (GBS-Ag) mas não a Ag capsular de

Streptococcus pneumoniae. GBS-Ag está sialilado a intervalos regulares. Ag capsular de *S. pneumoniae* (SPC-Ag) é muito similar mas não tem grupos ácido siálico. Uma tal CDR3 de HC curta cria um amplo sulco no qual um carboidrato se poderia ligar, e tais Abs são muito, muito raros em bibliotecas de anticorpos existentes. Assim sendo, as bibliotecas correntes não originam uma grande variedade de potenciais ligantes a carboidratos.

Ab 1B1 é o mAb de murino que se liga a GBS-Ag; Ab 1QFU é o mAb tendo uma estrutura 3D conhecida e a sequência mais próxima; e 1NSN é um anticorpo de estrutura 3D conhecida tendo um CDR3 de HC de comprimento 4. O exame de uma estrutura de HC 3-23 dá uma distância de C α de R₉₄ (que termina FR3) ao C α da W₁₀₄ (que começa FR4) de ~10 Å. A CDR3 de 1B1 (NWDY (SEQ ID NO:29)) mostra que os AAs não necessitam de ter somente grupos laterais pequenos ou serem maioritariamente de glicina. Três aminoácidos (AAs) podem ligar em ponte 10 Å, embora PPP possa não funcionar. De facto obtivemos alguns Fabs com CDR3s tão curtas quanto 3 AAs, mas são muito raros.

Embora tenham sido descritas CDR3s de HC curtas e muito curtas ninguém sugeriu fabrico de uma biblioteca de Ab tendo muito membros (p.ex., mais do que cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, ou cerca de 95% dos membros) com CDR3s de HC curtas (p.ex., CDR3s de HC de 3 a 5 aminoácidos). Uma abordagem à construção de uma biblioteca eficaz é em primeiro lugar desenhar sequências de aminoácidos que poderiam ter origem de acoplamento V-J ou V-D-J. Para comprimento de CDR3 3, 4, ou 4 começamos com as sequências de aminoácidos mostradas na Tabela 7. Por exemplo, a Sequência V-3JH1 mostra a extremidade C-terminal

de 3-23 FR3 (TAVYYCAK (SEQ ID NO:30)) seguida por JH1 que foi encurtada a partir da extremidade N-terminal até três aminoácidos antes da Trp-Gly que começa FR4. V-3JH2 mostra a extremidade de FR3 seguida pela JH2 encurtada. As sequências após V-3JH6 são construídas por união de FR4 com um encurtador tirado de um segmento D de humano seguido pela região FR4 de um segmento JH de humano. 3D3-3.3.2 seria um encurtador do segmento D3-3, terceira grelha de leitura começando no segundo aminoácido. 5D5-12.2.3 é um pentâmero de D5-12 na grelha de leitura 2 começando no aminoácido 3. Alguns dos segmentos D da linha germinativa contêm codões de paragem, todavia aparecem em anticorpos naturais quando os codões de paragem são retirados: Aqui assumimos que a mudança mais provável de codões TAA e TAG é para Tyr (Y) e essas paragens TGA são o mais provavelmente mudadas para Trp (W). A Tabela 20 mostra as sequências de aminoácidos dos segmentos de D de humano; os tipos de codões de paragem são indicados pelo uso de * para TAG, @ para TAA, e \$ para TGA. Na Tabela 11 estão 266 trímeros distintos que podem ser construídos a partir de segmentos de D de humano. As paragens TAA e TAG foram mudadas para Tyr mostradas como "y" (*i.e.*, minúsculas). Estas poderiam ser também mudadas para Ser, Cys, Phe, Gln, Lys, ou Glu por mudanças de bases únicas. TAG poderia ser mudado por mudanças de bases únicas para Trp bem como Tyr, Gln, Lys, Glu, Ser, e Leu. A Tabela 12 mostra os 266 tetrâmeros distintos que podem ser obtidos por encurtamento de segmentos de D de humano. A Tabela 14 mostra os 144 hexâmeros que podem ser obtidos por encurtamento de segmentos D de humano. As bibliotecas a serem construídas têm diversidade substancial em CDR1 de HC e CDR2 de HC. A diversidade de sequências de CDR3 de HC pode ser menos importante do que ter uma sequência curta, mas aceitável. A

diversidade de segmentos JH ou fragmentos (p.ex., 3 ou mais aminoácidos) de segmentos D proporciona sequências que poderiam ser construídas pelo sistema imune humano e logo é menos provável que sejam imunogênicas.

Numa forma de realização, os trímeros, tetrâmeros, e pentâmeros que contenham uma Cys são eliminados.

Numa forma de realização, os trímeros, tetrâmeros, e pentâmeros que contenham uma Cys ou que vieram de um fragmento D contendo uma paragem são eliminados.

As bibliotecas curtas construídas usando os trímeros da Tabela 11, tetrâmeros da Tabela 12, pentâmeros da Tabela 13 têm diversidade substancial: 266, 266, e 215, respectivamente. Isto é para ser comparado com o número de péptidos destes comprimentos: 8000, 160000, e 3200000, respectivamente.

V-3D1-1.1.1-JH1 contém a porção final de FR3 seguida por três aminoácidos de D1-1 (RF1), viz. GTT (SEQ ID NO:257). V-3D1-1.2-JH1 usa os aminoácidos 2-4 de D1-1 (RF1) como a CDR3 parente. V-3D3-3.3.3-JH2 mostra a extremidade de FR3 seguida pelos aminoácidos 3-5 de D3-3 (RF 3). A invenção compreende qualquer sequência de aminoácidos compreendendo FR3::(três, quatro, ou cinco AAs isentos de paragem de um segmento de D de humano)::FR4 de uma JH de humano. Fragmentos de regiões D contendo resíduos de Cys não emparelhados são menos preferenciais do que aqueles que estão isentos de resíduos de Cys não emparelhados. Em V-5JH3 existe uma Tyr mostrada como "y" porque JH3 tem somente 4 codões antes dos codões para Trp-Gly que definem o início de FR4. V-5JH4 tem uma Ser mostrada como "s" pela

mesma razão. Se for usado *wobbling*, o nível preferencial de pureza está entre 0,75 e 0,90. A invenção compreende as sequências V-3JH1 até V-3JH6, V-4JH1 até V-4JH6, e V-5JH1 até V-5JH6, e bibliotecas contendo as mesmas. A invenção compreende também as sequências nas quais a região CDR está substituída por um segmento de 3, 4, ou 4 aminoácidos de uma região D de humano, e bibliotecas contendo as mesmas. A invenção compreende adicionalmente ADN no qual a sequência parente foi mutada na região CDR3, e bibliotecas contendo o mesmo. Uma forma de realização preferencial é uma na qual o número médio de mudanças de bases por CDR3 é um, dois, ou três. Os métodos de mutagênese incluem PCR propenso a erros, *wobbling*, e *dobbling*.

Tabela 7: Sequências de aminoácidos de CDR3s parente de
comprimentos 3, 4, 5

Comprimento 3	...FR3-----	CDR3-	FR4-----	
V-3JH1	TAVYYCAK	FQH	WGQGT LVTVSS	(SEQ ID NO:31)
V-3JH2	TAVYYCAK	FDL	WGRGT LVTVSS	(SEQ ID NO:32)
V-3JH3	TAVYYCAK	FDI	WGQGT MVTVSS	(SEQ ID NO:33)
V-3JH4	TAVYYCAK	FDY	WGQGT LVTVSS	(SEQ ID NO:34)
V-3JH5	TAVYYCAK	FDP	WGQGT LVTVSS	(SEQ ID NO:35)
V-3JH6	TAVYYCAK	MDV	WGQGT TVTVSS	(SEQ ID NO:36)
V-3D1-1.1.1-JH1	TAVYYCAK	GTT	WGQGT LVTVSS	(SEQ ID NO:37)
V-3D1-1.1.2-JH1	TAVYYCAK	TTG	WGQGT LVTVSS	(SEQ ID NO:38)
V-3D3-3.3.3-JH2	TAVYYCAK	IFG	WGRGT LVTVSS	(SEQ ID NO:39)
Comprimento 4				
V-4JH1	TAVYYCAK	YFQH	WGQGT LVTVSS	(SEQ ID NO:40)
V-4JH2	TAVYYCAK	YFDL	WGRGT LVTVSS	(SEQ ID NO:41)
V-4JH3	TAVYYCAK	AFDI	WGQGT MVTVSS	(SEQ ID NO:42)

V-4JH4	TAVYYCAK	YFDY	WGQGT LVTVSS	(SEQ ID NO:43)
V-4JH5	TAVYYCAK	WFDP	WGQGT LVTVSS	(SEQ ID NO:44)
V-4JH6	TAVYYCAK	GMDV	WGQGT TVTVSS	(SEQ ID NO:45)
V-4D3-10.1a-JH2	TAVYYCAK	LLWF	WGRGT LVTVSS	(SEQ ID NO:46)

Comprimento 5

V-5JH1	TAVYYCAK	EYFQH	WGQGT LVTVSS	(SEQ ID NO:47)
V-5JH2	TAVYYCAK	WYFDL	WGRGT LVTVSS	(SEQ ID NO:48)
V-5JH3	TAVYYCAK	yAFDI	WGQGT MVTVSS	(SEQ ID NO:49)
V-5JH4	TAVYYCAK	sYFDY	WGQGT LVTVSS	(SEQ ID NO:50)
V-5JH5	TAVYYCAK	NWFDP	WGQGT LVTVSS	(SEQ ID NO:51)
V-5JH6	TAVYYCAK	YGMDV	WGQGT TVTVSS	(SEQ ID NO:52)
V-5D2-8.2a-JH2	TAVYYCAK	DIVLM	WGRGT LVTVSS	(SEQ ID NO:53)

Tabela 8: ADN codificando V-5D2-8.2a-JH2 para wobbling

```

!                                     CDR3.....
!   A   E   D   T   A   V   Y   Y   C   A   K   D   I   V   L   N
!   |gct|gag|gat|act|GCA|GtT|taT|taC|tgc|gct aag jez eqz jzz qzz ezj
!
!   W   G   Q   G   T   T   V   T   V   S   S   (SEQ ID NO:54)
!   tgg ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:55)
!   BstEII...
```

Alternativamente poder-se-ia sintetizar três fragmentos de ADN que correspondem à região de XbaI a BstEII e tendo o resíduo 94 sendo K ou R seguido por 3, 4, ou 5 códons NNK, seguidos por WG... de FR4. A variação permitida é $20^3 + 20^4 + 20^5 = 3.368.000$. Após amplificação, estas moléculas de ADN seriam misturadas na razão 1:10:100 (tal que sequências mais curtas estejam relativamente sobreamostradas) e clonadas no fagemídeo codificando a biblioteca de kappa com diversidade de CDR1/2 de HC. Uma biblioteca de 1×10^9 daria diversidade significativa e permitiria isolamento de

anticorpos que se ligam a alvos que tenham protuberâncias pequenas a médias. Por exemplo, vários carboidratos, ansas de proteínas que não estejam bem ordenadas (tais como GPCRs), podem beneficiar de um sulco no anticorpo criado por se ter uma CDR3 de HC muito curta. Podemos também construir uma biblioteca de lambda. A razão de sequências de AA é 1:20:400, e pode ser importante para amostrar as sequências mais curtas mais densamente. Ter-se um saco coletor grande, amplo no Ab pode requerer exatamente uma CDR3 de 3 AA, mas, com uma CDR3 de 4 AA poder-se-ia ter mais margem de manobra e, com 5 Aas, ainda mais margem de manobra. Neste Exemplo usamos a versão JH6 de FR4 do motivo WG para a frente.

Podemos selecionar da nossa biblioteca de kappa corrente uma coleção de, por exemplo, 24 cadeias leves kappa que sejam a) linha germinativa nas regiões de *framework*, b) mostrem diversidade adequada em CDRs, e c) sejam dos tipos que produzam bem e emparelhem bem com 3-23. Estas LCs serão feitas em *E. coli* a partir de um vetor que transporta *Kan^R* e nenhum sinal de empacotamento de fago. Construiríamos depois a nossa biblioteca de HC num vetor de fago que não tenha nenhuma LC. HC e LC serão cruzadas por infecção das células produzindo LC com os fagos de HC. Os fagos de HC que são selecionados podem ser combinados com a LC da célula que produz fagos ELISA⁺ ou as HCs podem ser clonadas em pMID21 que tenham a diversidade inteira de LC. Alternativamente, a HC selecionada pode ser movida para pHCSK85 e usada com ROLIC para combinar todas as LCs da nossa coleção. Poderiam ser também usadas LCs de lambda. Assim sendo, uma biblioteca de 1×10^9 HC em fagos pode ser expandida numa biblioteca de Fab de $1,2 \times 10^{11}$ ($1,0 \times 10^9 \times 117$). Se combinássemos 1×10^7 CDR1-2s com 10^6 CDR3s de HC

poderíamos fazer uma biblioteca de 5×10^7 na qual cada CDR3 está acoplada com 50 CDR1-2s. Uma biblioteca de 5×10^7 HCs em fagos poderia dar resultados similares a uma biblioteca de estilo antigo de 6×10^9 .

Tabela 1: Desenhos de CDR3 de HC exemplares muito curtas

```

C3XX
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|acc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
! XbaI...
!
! CDR3.....
! L R A E D T A V Y Y C A KIR qq qq qq W G
! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk tgg ggc-
!
! Q G T T V T V S S (SEQ ID NO:56)
! cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:57)
! BstEII...
!
(C3XX) 5'-T|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk tgg ggc cag ggt act ac-3'
(SEQ ID NO:58)
(ON_5) 5'-AcTggAgAcggTgAcggTAGTAcccTggcccca-3' ! 33 bases (SEQ ID NO:58)
(ON_5 é complemento reverso de 5'-tgg ggc cag ggt act acG GTC ACC gtc tcc
agt-3' (SEQ ID NO:59))
! Usar ON-1 e ON-2 mostrados em baixo
! -----
!
C3X4
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
! XbaI...
!
! CDR3.....
! L R A E D T A V Y Y C A KIR qq qq qq qq W
! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk tgg-
!
! G Q G T T V T V S S (SEQ ID NO:60)
! ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:61)
! BstEII...
!
(C3X4) 5'-GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk tgg-
ggc cag ggt act ac-3' (SEQ ID NO:62)
! Usar ON-1, ON-3, e ON-5
! -----
!
C3X5
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
! XbaI...
!
! CDR3.....
! L R A E D T A V Y Y C A KIR qq qq qq qq W
! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk-
!
! W G Q G T T V T V S S (SEQ ID NO:63)
! tgg ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:64)
! BstEII...
!
(C3X5) 5'-GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk nnk tgg-
ggc cag ggt act ac-3' (SEQ ID NO:65)
! -----
!
aRg codifica K ou R

```

Alternativamente, a diversidade de HC corrente pode ser clonada em DY3F87HC e a diversidade de CDR3 acima é clonada nessa diversidade como fragmentos XbaI - BstEII. Uma biblioteca de, por exemplo, 25 LC é clonada em pLCSK23 e usada para criar uma linha celular em *E. coli* TG1. Estas células são infectadas com o fago DY3F87HC que abriga a nova diversidade de CDR3 (e CDR1-2) de HC. Os fagos obtidos desta infecção são selecionados quanto a ligação a um alvo desejado. Após duas a quatro rondas de seleção, as HCs selecionadas são transferidas para pHCSK22 e usadas para criar uma linha de células que possa ser usada com ROLIC para combinar as HC selecionadas com todas as LCs na biblioteca de LC por ROLIC. Deste modo, uma biblioteca de $1,0 \times 10^9$ pode ser dar Abs que requereriam normalmente construção de uma biblioteca de $1,0 \times 10^{16}$ (assumindo uma diversidade de LC de $1,0 \times 10^7$).

Exemplo Profético 2: Bibliotecas com CDR3s de HC Muito Longas

Sidhu *et al.* (*J Mol Biol.* 2004 338:299-310 e pedido dos EUA 20050119455A1) relatam Abs de elevada afinidade selecionados de uma biblioteca na qual somente Y e S foram permitidos nas CDRs que estavam limitadas no comprimento a 20 aminoácidos. Pode ser possível gerar Abs de elevada afinidade a partir de uma biblioteca que tem CDR3s de HC com uma ou mais das seguintes formas de diversidade: a) vários locais (mas não todos) permitindo Y ou S, b) incluindo 4-6 códons NNK, c) introduzindo segmentos D (com ou sem diversificação no D), e/ou d) usando PCR propenso a erros. Já amostrámos o espaço de Ab no qual a CDR3 de HC está na gama -8 a ~22 com um comprimento médio de 13. Assim sendo, bibliotecas nas quais CDR3 de HC tem ~23 AAs ou ~35 AAs são possíveis e podem ter vantagens com certos

tipos de alvos. Por exemplo, GPCRs são proteínas da membrana integrais com sete segmentos helicoidais atravessando a camada de lípidos da célula que se pensa que têm estados múltiplos. Um anticorpo tendo uma CDR3 de HC muito longa poderia formar uma protuberância que se ajuste ao canal formado pelas sete cadeias. A procura de Abs que se liguem a GPCRs tem sido difícil e a construção intencional de bibliotecas nas quais todos os membros tenham CDR3s de HC muito longas pode melhorar este problema. Os comprimentos podem ser tornados algo variáveis, digamos 23, 24, ou 25, numa biblioteca e 33, 34, ou 35 numa segunda.

Em baixo está um número de desenhos representativos. As CDR3 foram quebradas e a diversidade gerada que permite que as várias partes tenham relações diferentes dependendo do valor de X. Foi usado JH1 de comprimento total e em alguns desenhos a diversidade permitiu diversidade na parte de CDR3 de JH1. Outros JHs poderiam ser usados. Nos desenhos, os segmentos D são ricos em Y ou têm uma ansa de dissulfeto rica em S. As sequências de aminoácidos de segmentos D de humano são mostradas na Tabela 3. Os locais onde a região D tem S ou Y ou permitiu outras combinações foram em particular variados. A Tabela 4 mostra as sequências de aminoácidos de regiões J de humano.

Cada uma das bibliotecas poderia ser construída em pelo menos quatro modos: a) ADN codificando uma sequência de aminoácidos particular é em primeiro lugar sintetizado e sujeito a PCR propenso a erros, 2) a biblioteca pode ser sintetizada por *wobbling* ou com misturas de nucleótidos, 3) a biblioteca pode ser construída usando *dobbling*, e 4) a rota (2) ou (3) poderia ser seguida por PCR propenso a

erros. Como um exemplo da rota (1), no Desenho 12, poderia ser sintetizado ADN codificando SEQ ID NO:908, como mostrado em SEQ ID NO:911. Este ADN poderia ser sujeito a PCR propenso a erros usando os iniciadores mostrados em SEQ ID NO:909 e SEQ ID NO:910. Porque estes iniciadores abrangem as regiões de *framework*, os erros ocorrerão somente na CDR3.

Uma biblioteca de HCs com CDR3 com comprimento 23 de, por exemplo, 2×10^9 membros e uma segunda biblioteca com CDR3s de HC de comprimento ~ 35 tendo também 2×10^9 membros poderiam ser construídas. Alternativamente, o ADN poderia ser misturado para construir uma biblioteca de 4×10^9 .

Tabela 4: Sequências de aminoácidos de JH de humano

	H3	
	CDR3	
	100	110
JH1	---AEYFQHWGQGT ^{LVTVSS}	(SEQ ID NO:66)
JH2	---YWYFDLNGRGTL ^{LVTVSS}	(SEQ ID NO:67)
JH3	-----AFDINGQGT ^{MTVTVSS}	(SEQ ID NO:2)
JH4	-----YFDYWGQGT ^{LVTVSS}	(SEQ ID NO:1)
JH5	-----NWFDPWGQGT ^{LVTVSS}	(SEQ ID NO:68)
JH6	YYYYYGMDVWGQGT ^{TVTVSS}	(SEQ ID NO:3)

Em cada um dos seguintes desenhos, a sequência de aminoácidos começa com YYCA(K/R) que é o final de FR3. FR4 começa com WG e é mostrado negrito.

Desenho 1

SEQ ID NO:898 compreende a extremidade de FR3 unida com dois resíduos (DY) de tipos frequentemente encontrados na sequência de enchimento que o sistema imune coloca entre V e D. Estes são seguidos por D2-2.2, preferencial porque tem uma ansa de dissulfeto e é rico em resíduos de Ser e Tyr.

Isto é seguido por YGYSY, que é rico em resíduos de Tyr e Ser, que é seguido por JH1 de comprimento total.

```

XX::D2-2.2::XX::JH1
      1      1      2      2
FR3 1      5      0      5      0 3FR4
YYCAK DYGYCSSTSCYTYGYSYAEYFQHWGQGLVTVSS (SEQ ID NO:898)
YYCAK XXGYCSXXSCYTXGYSYAEYFQHWGQGLVTVSS (SEQ ID NO:69)
      R   GYCSSTSCYT      AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:70)      (SEQ ID NO:66)
      1 1      1      1
9 9      0 0      0      1
4 5      0 2abcde fghijklmnp3      0

```

```

Diversidade de aminoácidos    = 1,28 E 8
Diversidade de ADN            = 2,15 E 9
Isentos de paragem           = 83%
Isentos de Cys injustificadas = 83%
Isentos de paragem e Cys      = 68%

```

O desenho 1 (C23D222) tem 94 sendo R ou K, depois 2 Xs, D2-2 na segunda grelha de leitura com dois Xs na ansa, seguido por dois Xs, e JH1. A 2ª grelha de leitura de D2-2 tem uma ansa fechada de dissulfeto na qual a diversidade em dois pontos foi introduzida. Esta CDR3 tem comprimento 23. Usando iniciadores que incluem ADN até ...YCA e a partir de WGQG..., PCR propenso a erros poderia ser realizada na CDR3 antes de amplificação em *XbaI* e *BstEII* para clonagem na biblioteca de LC de kappa e CDR1/2 de HC. Assim sendo será permitido que os AAs que são mostrados como fixados variem alguma coisa. Os AAs que são parte da região de sobreposição de PCR serão reforçados pela PCR não propensa a erros final. A PCR propensa a erros não é uma parte necessária do desenho.

```

C23D222JH1
! ADN de scab S R D N S R N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
XbaI...
! L R A E D T A V Y Y C A KIR
!TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|tAT|tAc|tgc|gct aRg -
!
! CDR3-----
! X X D2-2 RF2..... X X JH1..
! qg qg G Y C S qg qg S C Y T qg qg Y S Y A
! nnk nnk ggt tat tgt tcc nnk nnk tct tgc tat act nnk nnk tat tcc tac gct-
!
! CDR3-----
! E Y F Q H
! gaa tat ttc cag cac-
!
! W G Q G T L V T V S S (SEQ ID NO : 71)
! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO : 72)
! BstEII...

(ON_C23D222) 5'-GCAIGtTitaTitaCItgc|gct aRg nnk nnk ggt tat tgt tcc nnk-
nnk tct tgc tat act nnk nnk tat tcc tac gct gaa tat ttc cag cac-
tgg ggc cag ggt act ct-3' : 107 bases (SEQ ID NO : 73)
(ON_1) 5'-GCAIGtTitaTitaCItgc|gct-3' (SEQ ID NO : 74)
(ON_2) 5'-AgAgTAcccTggccccAgAgcTccATAccgTAATAgT-3' : 37 bases (SEQ ID NO :
75)
(ON_2 é complemento reverso de 5'-ac tat tac ggt atg gac gtc tgg
ggc cag ggt act ct-3') (SEQ ID NO : 76)
(ON_3) 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|~
aac|agC|TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|tAT|tAc|tgc|gct-3' (SEQ ID
NO : 77)
(ON_4) 5'-AcTggAgAgcgTgAccAgAgTAcccTggccccA-3' : 33 bases (SEQ ID NO : 78)
(5'-tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (RC) (SEQ ID NO : 79))

```

Desenho 2

```

          1      1      2      2
        1  S      0      5      0      3
YYCAK GSYYYGSGSYYNVDSSYYAEYFQHWGQGLVTVSS (SEQ ID NO:899)
YYCAK XXYYYGXG SXYNXXSYIAEYFQHWGQGLVTVSS (SEQ ID NO:80)
      R   YYYGSGSYYN      AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:81)      (SEQ ID NO:66)

```

```

Diversidade de aminoácidos    = 1,28 E 8
Diversidade de ADN            = 2,15 E 9
Isentos de paragem           = 83%
Isentos de Cys injustificadas = 83%
Isentos de paragem e Cys      = 68%

```

O desenho 2 (C23D310) tem 94 como R ou K, dois Xs, D3-10 com os 5º e 8º resíduos mudados para X, 2 Xs, SY, e JH1. A CDR3 tem 23 AA de comprimento e poderia ser adicionalmente diversificada por uso de PCR propensa a erros.

```

C23D310JH1
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
! XbaI...
! L R A E D T A V Y Y C A K/R
! TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg -
!
! CDR3-----
!
! qq qq Y Y Y G qq G S qq Y N qq qq S Y Y
! nnk nnk tac tac tat ggt nnk ggc tct nnk tac aat nnk nnk tct tat tac
!
! A C Y F Q H
! gct gag tac ttt caa cat
!
! JH1.....
! W G Q G T L V T V S S (SEQ ID NO:82)
! tgg ggc cag ggt act ctG GTC ACC ggc tcc agt-3' (SEQ ID NO:83)
! BarEII...

```

(C23D310) 5'-GCA|GtT|taT|taC|tgc|gct aRg nnk nnk tac tac tat ggt nnk ggc-
tct nnk tac aat nnk nnk tct tat tac gct gag tac ttt caa cat tgg ggc cag-
ggt agt ct-3' (SEQ ID NO:84)
ON_1, ON_2, ON_3, e ON_4 como acima.

Desenho 3

```

      . 1      1      2 2
      1 5      0 5      0 3
YYCAK DYYYYGSGSYYNSDSYSAEYFQHNWGQGLVTVSS (SEQ ID NO:900)
YYCAK XZYZZGZGEXYNZXZYZAXZFQHNWGQGLVTVSS (SEQ ID NO:84)
      R   YYYGSGSYYN   AEYFQHNWGQGLVTVSS (JH1)
      (SEQ ID NO:81)   (SEQ ID NO:66)
Diversidade de aminoácidos      = 1,64 E 8
Diversidade de ADN              = 1,07 E 9
Isentos de paragem              = 88%
Isentos de Cys injustificadas   = 88%
Isentos de paragem e Cys        = 77%

```

O desenho 3 (C23D310B) tem 94 como R ou K, XZ, D3-10 (RF2) com os 2°, 3°, 5°, e 7° como Z(Y|S) e 8° resíduo mudado para X, ZXZYZ, e JH1 (com o E mudado para X). Z é Y ou S. A CDR3 tem 23 AA de comprimento e poderia ser adicionalmente diversificada por uso de PCR propensa a erros.

```

      A  V  Y  Y  C  A  RIK  qq  YIS  Y  YIS  YIS  G  YIS
(C230310b) 5'-GCAIGTITaTItaCItgcIcgt aRq nnk tmc tac tmc tmt ggc tmc ggc-
      YIS  qq  Y  N  YIS  qq  YIS  Y  YIS  A  qq  YIS  F  Q  H  W  G  Q
      tmt nnk tac aat tmt nnk tmc tat tmc gct nnk tmc ttt caa cat tgg ggc cag-

      G  T  L      (SEQ ID NO:85)
      ggt act ct-3' (SEQ ID NO:86)

```

ON_1, ON_2, ON_3, e ON_4 como acima.

Desenho 4

```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK YYSFSYYPYYDSSGYYYGYSDYSYSYAEYFOHNGGTLTVSS (SEQ ID NO:901)
YYCAK YYSXSYYXZYDSEGYYXYYXYYXYYZZAZZFQHNGGTLTVSS (SEQ ID NO:87)
      R      YYDSSGYYY      AEYFOHNGGTLTVSS (JH1)
      (SEQ ID NO:88)      (SEQ ID NO:66)
      1 1      1      1
      9 9      0 0      0      1
      4 5      0 2ab cdefghijklmnopqrstuvwxyab3      0

```

Diversidade de aminoácidos	≈ 1,64 E 8
Diversidade de ADN	≈ 1,07 E 9
Isentos de paragem	= 88%
Isentos de Cys injustificadas	= 88%
Isentos de paragem e Cys	= 77%

O desenho 4 tem CDR3 de comprimento 35. O resíduo 94 pode ser K ou R, depois YYS::X::SY::X::D3-22(2ª RF com uma S como X e 3 Zs)::X::YYS::X::YZZZ::JH1(com 2 Zs). PCR propensa a erros poderia ser usada para adicionar mais diversidade.

```

C35B322JH1
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|ACA|gac|aac|tct|aag|aat|act|ctc|cac|ttg|cag|atg|aac|agC-
XbaI...

L R A E D T A V Y Y C A KIR
TTAAAGg|gct|gag|gat|act|GCA|GTT|TAT|TAC|tgc|gct arg -

CDR3-----
Y Y S qq S Y Y qq Y Y|S Y D S Y|S G Y Y|S Y
tac tat tcc nnk tct tac tat nnk tat tmt tac gat agt tmt ggt tcc tmc tat

qq Y Y S qq Y Y|S Y Y|S Y|S A Y|S Y|S F Q H
nnk tac tat agc nnk tat tmc tac tmc tmt tmc gct tmt tmc ttc caa cac

M Q Q Q T L V T V S S (SEQ ID NO:89)
tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:90)
BacII...

(c35b322b) 5'-GCA|Gt|Tira|Tita|Circ|gct arg tcc tat tcc nnk tct tac tat nnk-
tat tmt tac gat agt tmt ggt tac tmc tat nnk tac tat agc nnk tat tmc tac-
tmc tmt tmc gct tmt tmc ttc caa cac tgg ggc cag ggt act ct-3' (SEQ ID NO:91)
DN_1, DN_2, DN_3, e DN_4 como acima

```

Desenho 5

```

          1      1      2      2
        1  S      0      5      0      3
YYCAK SSGYCSSTSCYTGYYYYAEYFQHWGQGLVTVSS (SEQ ID NO:90)
YYCAK ZZGZCZZXZCZTZXZYXKEYFQHWGQGLVTVSS (SEQ ID NO:92)
      R GYCSSTSCYT AEYFQHWGQGLVTVSS (JH1)
        (SEQ ID NO:70) (SEQ ID NO:66)

Diversidade de aminoácidos      = 1,64 E 8
Diversidade de ADN              = 1,07 E 9
Isentos de paragem              = 88%
Isentos de Cys injustificadas   = 88%
Isentos de paragem e Cys        = 77%

```

O desenho 5 (C23D222b) é como o desenho 1 mas usa muitos codões variáveis Z (Y ou S). Esta CDR3 tem comprimento de 23.

```

C23D222JH1b
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
! XbaI...

L R A E D T A V Y Y C A KIR
:TTA|AGg|gct|gag|gat|act|GCA|Gt|T|aT|T|aC|tgc|gct aRg -

CDR3-----
Y|S Y|S G Y|S C Y|S Y|S Gg Y|S C Y|S T Gg Gg Y Y|S Y Gg
tmc tmt ggt tmt tgc tmc tmt nnk tmt tgt tmc acc nnk nnk tat tmt tac nnk

Y|S Y F Q H
tmt tat ttc cag cac

W G Q G T L V T V S S (SEQ ID NO:93)
tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:94)
BstEII...

(C23D222JH1b) 5'-GCA|Gt|T|aT|T|aC|tgc|gct aRg tmc tmt ggt tmt tgc tmc tmt-
nnk tmt tgt tmc acc nnk nnk tat tmt tac nnk tmt tat ttc cag cac tgg ggc-
cag ggt act ct-3' (SEQ ID NO:95)

```

Desenho 6

```

          1   1   2 2 2   3   3
        1   5   0   5   0 3 5   0   5
YYCAK SYDYYGYCSSTSCYTYYSYVSYSYSSYYSAEYFQHNQGGTLVTVSS (SEQ ID NO:903)
YYCAK SYXZYGCZEXSCETYSZXZYSEYSSZYAEFQHNQGGTLVTVSS (SEQ ID NO:96)
R      GYCSSTSCYT D2-2.2      AEYFQHNQGGTLVTVSS (JH1)
          (SEQ ID NO:70)      (SEQ ID NO:66)

```

```

Diversidade de aminoácidos    = 2,00 E 8
Diversidade de ADN            = 5,37 E 8
Isentos de paragem            = 91%
Isentos de Cys injustificadas = 91%
Isentos de paragem e Cys      = 83%

```

C35D222JH1

ADN de scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
 XbaI...

L R A E D T A V Y Y C A KIR
 ITTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct aRg -

CDR3-----
 YIS Y QQ YIS Y G YIS C YIS YIS QQ S C YIS T Y YIS S
 tmt tac nnk tmc tmc ggc tmt tgc tmt tmc nnk tCt tgt tmc acc tat tmt tcc

YIS QQ YIS Y S QQ Y YIS S YIS Y A E Y P Q H
 tmt nnk tmc tct tct nnk tmc tmc agt tmt tat gct gag tat ttc cag cac

W G Q G T L V T V S S (SEQ ID NO:97)
 tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:98)
 BstEII...

(C35D222JH1)5'-GCA|GtI|taT|taC|tgc|gct aRg tmt tac nnk tmc tmc ggc tat- tgc tmt tmc
 nnk tmt tgt tmc acc tat tmt tcc tat nnk tmc tat tct nnk tct-
 tmc agt tmt tat gct gag tat ttc cag cac tgg ggc cag ggt act ct-3' (SEQ ID NO:99)

Desenho 7

1 1 2 2 2 3 3
 1 5 0 5 0 3 5 0 5
 YYCAK YYSYYGYCSSTSCYTYSSSVSYSYSSYYAEEYFQHNQGGTLVTYSS (SEQ ID NO:904)
 YYCAK ZYZZYGZCZKZCZTYSSZXZYSSZYSSZYAEEYFQHNQGGTLVTYSS (SEQ ID NO:100)
 R GYCSSTSCYT D2-2.2 AEYFQHNQGGTLVTYSS (JH1)
 (SEQ ID NO:70) (SEQ ID NO:66)
 (J=FSY, B=YHND, #=EKQ)

Diversidade de aminoácidos = 9,44 E 8
 Diversidade de ADN = 2,42 E 9
 Isentos de paragem = 93%
 Isentos de Cys injustificadas = 93%
 Isentos de paragem e Cys = 88%

C35D222JH1B

ADN de scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
 XbaI...

L R A E D T A V Y Y C A KIR
 ITTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct aRg -

CDR3-----
 YIS Y YIS YIS Y G YIS C YIS YIS QQ YIS C YIS T Y YIS S
 tmt tac tmc tmc tmc ggc tmt tgc tmt tmc nnk tmt tgt tmc acc tat tmt tcc

YIS QQ YIS Y S YIS Y YIS S YIS Y A EIK YIS FIS Q HIY
 tmt nnk tmc tat tct tmt tac tmc agt tmt tat gct Vag tmt tmc cag Nac

W G Q G T L V T V S S (SEQ ID NO:101)
 tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:102)
 BstEII...

Desenho 8

```

          1      1      2 2 2      3      3
        1 5      0      5      0 3 5      0      5
YYCAK SRSYDYVWGSYRYTSSSYSYSYSSSYAEYFQHWGQGLTVTVSS (SEQ ID NO:905)
YYCAK ZXZYEBZVWGZRETZSZXZYZZZYVZSAZEFQHWGQGLTVTVSS (SEQ ID NO:103)
      R      YYDYVWGSYRYT D3-16.2      AEYFQHWGQGLTVTVSS (JH1)
          (SEQ ID NO:104)          (SEQ ID NO:66)
(J=FSY, B=YHND, Q=EKQ)

```

```

Diversidade de aminoácidos      = 9,44 E 8
Diversidade de ADN              = 1,61 E 9
Isentos de paragem              = 93%
Isentos de Cys injustificadas   = 93%
Isentos de paragem e Cys        = 88%

```

C34D316JH1A

```

!
! ADN de scab S R D N S K N T L Y L O M N S
5'-ctc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
! XbaI...
!
! L R A E D T A V Y Y C A K R
! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|tAt|tAt|tgc|gct aRg -
!
! CDR3-----
! NID
! Y|S QQ Y|S Y Y|S Y|H Y|S V W G Y|S Y|S R Y|S T Y|S
! tmt nnk tmc tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act tmt
!
! S Y|S QQ Y|S Y Y|S Y|S Y|S Y Y|S S Y|S
! agt tmc nnk tmt tac tmc tmt tmc tat tmc agt tmt
!
! Q
! A, E|K Y|S F Q H
! GCT vag tmc ttc cag cat
!
! W G Q G T L Y T V S S (SEQ ID NO:105)
! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:106)
! BstEII...
!
! (C34D316JH1A) 5'-GCA|GtT|tAt|tAt|tgc|gct aRg tmt nnk tmc tac tmt Nat tmt-
! gtt tgg ggt tmt tmc cgt tmt act tmt agt tmc nnk tmt tac tmc tmt tmc tat-
! tmc agt tmt GCT vag tmc ttc cag cat tgg ggc cag ggt act ct -3' (SEQ ID NO:107)

```

Desenho 9

O desenho 9 é como 8 exceto que o segmento D está movido para a direita

```

          1      1      2  2  2      3      3
        1  5      0      5      0  3  5      0      5
YYCAK YGYSSDSYSSSYDYVWGSYRYTYSSSYAEYFOHNGOGTLVTVSS (SEQ ID NO:906)
YYCAK XXXXXXXXXXZZZZZZZZVWGGZZRETYZZSYAAEYFOHNGOGTLVTVSS (SEQ ID NO:108)
      R  D3-16.2  YYDYVWGSYRYT  AEYFOHNGOGTLVTVSS (JH1)
              (SEQ ID NO:104)  (SEQ ID NO:66)
(J=FSY,B=YHND,g=EKQ)

```

```

Diversidade de aminoácidos    = 1,31 E 8
Diversidade de ADN            = 5,37 E 8
Isentos de paragem           = 91%
Isentos de Cys injustificadas = 91%
Isentos de paragem e Cys      = 83%

```

C34D316JH1B

```

!
! ADN de scab S R D N S K N T L Y L Q M N S
5'-ctc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
! XbaI...
!
! L R A E D T A V Y Y C A KIR
! ITTA|AGg|gct|gag|gat|act|GCA|GtT|tAT|tAC|tgc|acc aBq -
!
! CDR3-----
! YIS qq YIS YIS YIS qq YIS Y YIS YIS YIS
! tmt nnk tmc tmt tmc nnk tmt tac tmc tmt tmc
!
! MID
! Y YIS YIH YIS V W G YIS YIS R YIS T
! tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act
!
! Y YIS S YIS Y
! tat tmc agt tmt tac
!
! Q
! A EIK YIS F Q H
! GCT vag tmc ttc cag cat
!
! W G Q G T L V T V S S (SEQ ID NO:109)
! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:110)
! BstEII...

```

(C35D316JH1B)

```

5'-GCA|GtT|tAT|tAC|tgc|gct aBq tmt nnk tmc tmt tmc nnk tmt tac tmc tmt tmc
tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act tat tmc agt tmt tac GCT vag
tmc ttc cag cat tgg ggc cag ggt act ct-3' (SEQ ID NO:111)

```

Desenho 10

```

          1      1      2  2
        1  5      0      5      0  4
YYCAK GSSYYYGSGSYNSDYSSAEYFOHNGOGTLVTVSS (SEQ ID NO:907)
YYCAK XXZZZZGGGZZXNXXZYXAXZFOHNGOGTLVTVSS (SEQ ID NO:112)
      R  YYYGSGSYYN  AEYFOHNGOGTLVTVSS (JH1)
              (SEQ ID NO:81)  (SEQ ID NO:66)

```

O desenho 10 (C24D310B) é como o Desenho 3, mas a CDR3 tem o comprimento 24. O desenho 10 tem 94 como R ou K, XZZ, D3-10 (CRF2) com 2°, 3°, 5°, e 7° como Z(YIS) e o 8° resíduo mudado para X, ZXZYZ, e JH1 (com o E mudado para X). Z é Y ou S. A CDR3 tem 24 AA de comprimento e poderia ser diversificada adicionalmente por uso de PCR propenso a erros.

```
(C24D310b) 5'-GCAIGtTitaTitaCItgcigct aRg nnk tmc tmc tac tmc tmt ggt tmc-
ggc tmt nnk tac aat tmt nnk tmc tat tmc gct nnk tmc ttt caa cat tgg ggc-
cag ggt act ct-3' (SEQ ID NO:113)
```

ON_1, ON_2, ON_3, e ON_4 como acima.

Desenho 11

		1	1	2	2
1	5	0	5	0	5

```
YYCAR SSRSGYCTNGVCYRSGSYWYFDLNGRGTLVTVSS (SEQ ID NO:907)
YYCAR ZZX2GZC32GVC23ZXZZ4212LNGRGTLVTVSS (SEQ ID NO:114)
K GYCTNGVCYT YWYFDLNGRGTLVTVSS D2-8.2 JH2
(SEQ ID NO:115) (SEQ ID NO:67)
(1=FYS(THT), 2=YHND(NAT), 3=ITKR(ANA), 4=LSW(TBG))
```

```
(C24D282) 5'-GCAIGtTitaTitaCItgcigct aRg tmc tmt nnk tmt ggt tmc tgc ana-
nat ggt gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tht nat ctg tgg ggc-
cag ggt act ct-3' (SEQ ID NO:116)
```

```
(C24D282.1) 5'-GCAIGtTitaTitaCItgcigct aRg tmc tmt nnk tmc ggt tmc tgc ana-
nat ggt gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tht nat ctg tgg ggc-
cag ggt act ct-3' (SEQ ID NO:117)
```

```
(C24D282.1) 5'-GCAIGtTitaTitaCItgcigct aRg tmc tmt nnk tmc ggt tmc tgc ana-
nat ggt gtc tgc t-3' (SEQ ID NO:118) (necessita R, M, N, K)
(C24D282.2) 5'-Ag AqT Acc cTs ggc cca cAg ATN ADA AKA cVA AKA AKA MNV gKA TNT AKA gCA
gAc gcc ATN TNT gCA gKA Acc g-3' (SEQ ID NO:119) ! 75 bases
(5'-g ggt tmc tgc ana-
nat ggt gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tht nat ctg tgg ggc-
cag ggt act ct-3' [RC] (SEQ ID NO:120) (necessita N, M, K, S, H)
```

Desenho 12


```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK YSYYSYSGSYYSYDYVWGSYRYTSYDSYYYAEYFQHWGQGTILVTVSS (SEQ ID NO:913)
YYCAK ZZZZZZZZZZDYVWGSYRZTZZZZZZZAEYFQHWGQGTILVTVSS (SEQ ID NO:127)
      R D3-16.2 YYDYVWGSYRYT      AEYFQHWGQGTILVTVSS (JH1)
              (SEQ ID NO:104)      (SEQ ID NO:66)

(C34D316.2A) 5'-GCAIGtTitaTitaCitgcigct aRg tmt tmc tmc tmt tmt tmc tmc tmt-
              tmc tmc tmc gat tat gtc tgg ggt act tat cgt-3' (SEQ ID
NO:128)
(C34D316.2B) 5'-Ag AgT Acc cTg gcc cCA ATg cTg gAA AKA cTc Agc gKA gKA gKA-
              gKA gKA gKA AKA AgT gKA Acg ATA AgT Acc cCA gAc ATA ATc-3' (SEQ ID
NO:129) ! 86 bases
(5'-gat tat gtc tgg ggt act tat cgt tmc act tmt tmc tmc tmc tmc-
              tmc tmc gct gag tmt ttc cag cat tgg ggc cag ggt act ct-3'
(SEQ ID NO:130) [RC])

```

Desenho 15

O desenho 15 permite alguma diversidade na sobreposição, 5 *flip-flops* de dois sentidos. Existem somente 32 sequências de sobreposição e, mesmo se existirem não correspondências, elas não mudarão a diversidade permitida.

```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK SYDYSSYSYYYDYVWGSYRYTSYSGDSYYAEYFQHWGQGTILVTVSS (SEQ ID NO:914)
YYCAK ZZZZZZZZZZDYVWGSYRZTZZZZZZZAEYFQHWGQGTILVTVSS (SEQ ID NO:131)
              YYDYVWGSYRYT      AEYFQHWGQGTILVTVSS
              (SEQ ID NO:104)      (SEQ ID NO:66)

(C35D316.2A) 5'-GCAIGtTitaTitaCitgcigct aRg tmt tmc tmc tmt tmt tmc tmc tmt-
              tmc tmc tmc gat tmt gtc tgg ggt tmc tmc cgt tmc acc t-3' (SEQ ID NO:
132)
(C35D316.2B) 5'-Ag AgT Acc cTg gcc cCA ATg cTg gAA AKA cTc Agc gKA gKA-
              gKA gKA gKA gKA gKA ggt gKA Acg gKA gKA Acc cCA gAc AKA gTc gKA g-3'
(SEQ ID NO:133)
(5'-g tmc gac tmt gtc tgg ggt tmc tmc cgt tmc acc tmt tmc tmc-
              tmc tmc tmc tmc tmc gct gag tmt ttc cag cat tgg ggc cag ggt acc ct-3' (SEQ ID
NO:134) [RC])

```

Desenho 16

O desenho 16 proporciona uma CDR3 de 35. Existem 4 *flip-flops* de dois sentidos na sobreposição, logo 16 sequências.

```

      1      1      2 2 2      3      3
      1      5      0      5      0 3 5      0      5
YYCAK SSSYSYSYSGYCSGGSCYSSYYYSSYYSAEYFQGWGGTILVTIVSS (SEQ ID NO:915)
YYCAK ZZZZZZZZZGZCZGGZCZSEZZZZZZZZAEZFQHWGGTILVTIVSS (SEQ ID NO:135)
      R      GYCSGGSCYS 2-25.2 AEYFQHWGGTILVTIVSSJH1
              (SEQ ID NO:136)      (SEQ ID NO:66)

(C35D225.2A) 5'-GCAIGtTitaTitaCItgcIgst aRg tmt tmt tmt tmt tmt tmt tmt-
              tmc tmc ggc tmc tgc tmc ggt ggc tmc tgc tmc tcc t-3' (SEQ ID NO:137)
(C35D225.2B) 5'-Ag AgT Acc cTg gcc cCA ATg TTg gAA AKA TTc Agc gKA gKA-
              gKA gKA gKA gKA gKA gKA gKA gKA gga gCA gKA gcc Acc gKA Aca gKA gcc gKA g-3'
              (SEQ ID NO:138)! 96 bases

```

Se adicionarmos C34D225.2A e C34D225.2B à mistura obtemos então CDR3s de comprimentos 33, 34, e 35.

```

(C34D225.2A) 5'-GCAIGtTitaTitaCItgcIgst aRg tmt tmt tmt tmt tmt tmt tmt-
              tmc tmc ggc tmc tgc tmc ggt ggc tmc tgc tmc tcc t-3' (SEQ ID NO:139)
(C34D225.2B) 5'-Ag AgT Acc cTg gcc cCA ATg TTg gAA AKA TTc Agc gKA gKA-
              gKA gKA gKA gKA gKA gKA gKA gKA gga gCA gKA gcc Acc gKA Aca gKA gcc gKA g-3'
              (SEQ ID NO:140)! 93 bases

```

Desenho 17

```

      1      1      2 2 2      3      3
      1      5      0      5      0 3 5      0      5
YYCAK YSSYSYDYVWGSYRYTSSSYSYSSYYYAEYFQGWGGTILVTIVSS (SEQ ID NO:916)
YYCAK ZZZZZZZZDVNGZZRZTZZZZZZZZZZAEZFQHWGGTILVTIVSS (SEQ ID NO:141)
      R      YYDYVWGSYRYT D3-16.2 AEYFQHWGGTILVTIVSS (JH1)
              (SEQ ID NO:104)      (SEQ ID NO:66)

(C35D3162A) 5'-GCAIGtTitaTitaCItgcIgst aRg tmt tmt tmt tmt tmt tmt tmc gac-
              tmc gtc tgg ggt tmt tmc ggt tmt acc t-3' (SEQ ID NO:142)
(C35D3162B) 5'-Ag AgT Acc cTg gcc cCA gTg cTg gAA gKA cTc Agc gKA gKA gKA-
              gKA gKA gKA gKA gKA gKA gKA gKA ggt AKA Acg gKA AKA Acc cCA gAc-
              gKA gTc g-3' (SEQ ID NO:143)

```

Desenho 18

```

          1   1   2 2 2   3   3
        1   5   0   5   0 3 5   0   5
YYCAK SSYYYSSSYDYVWGSYRYTSSYYSSYAEYFQGWGQGLVTVSS (SEQ ID NO:917)
YYCAK ZZZZZZZZZZDQEVWGZRRZTEZZZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:144)
R      YYDYVWGSYRYT D3-16.2AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:104)      (SEQ ID NO:66)

(C35D3162C) 5'- GCAIGtTItaTItaCItgcIgcT aRg tmt tmt tmt tmt tmt tmt tmc-
tmc tmc tmc gac tmc gtc tgg ggt tmc tmc cgt tmc acc t-3' (SEQ ID NO:145)
82 bases
(C35D3162B) 5'-Ag AgT Acc cTg gcc cCA gTg cTg gAA gKA cTc Agc gKA gKA-

gKA gKA gKA gKA gKA gKA gKA gKA ggT gKA Acg gKA gKA Acc cCA gAc gKA-
gTc g-3' (SEQ ID NO:146)

```

Desenho 19

```

          1   1   2 2 2   3   3
        1   5   0   5   0 3 5   0   5
YYCAK YSGDSYSSYYYDSSGGYYYSSYYSSSYAEYFQGWGQGLVTVSS (SEQ ID NO:918)
YYCAK ZZZZZZZZZZDSSGGZZZZZZZZZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:147)
R      YYDSSGGYYY AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:88)      (SEQ ID NO:66)

          1 1          1 1
        9 9      0 0          0 1
        4 5      0 2ab cdefghijklmnopqrstuvwxyab3 0
                        ,,

Diversidade de aminoácidos    ~ 6,7 E 7
Diversidade de ADN           ~ 6,7 E 7
Isentos de paragem           = 100
Isentos de Cys injustificadas = 100
Isentos de paragem e Cys     = 100%

```

O desenho 19 tem CDR3 de comprimento 35. O resíduo 94 pode ser K ou R. A ZZZZZZZZZ::D3-22(2ª RF com seis Ys como Z)::ZZZZZZZZZZ::JH1(com 1 Z). PCR propensa a erros poderia ser usada para adicionar mais diversidade.

C35D322AJH1
 ! ADN de scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 ! XbaI...
 !
 ! L R A E D T A V Y Y C A KIR
 ! TTA|ACg|gct|gag|gaT|ACT|GCA|GtT|taT|taC|tgc|gct aRg -
 !
 ! CDR3-----
 !
 ! YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS D S S G YIS YIS YIS
 ! tmc tmt tmc tmc tmt tmc tmt tmc tmc tmc tmc gac agc tcc ggc tmc tmc tmt
 !
 ! YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS A E YIS F Q H
 ! tmc tmt tmc tmc tmt tmc tmt tmc tmc tmc tmc gct gac tmc ttc caa cac
 !
 ! W G Q G T L V T V S S (SEQ ID NO:148)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:149)
 ! BstEII...

(C35D322AJH1_T) 5'-GCAIGtT|taT|taC|tgc|gct aRg tmc tmt tmc tmc tmt-
 tmc tmt tmc tmc tmc gac agc tcc ggc tmc tmc t-3' (SEQ ID NO:150)
 (C35D322AJH1_B) 5'-cAg AgT Acc cTg gcc cCA gTg TTg gAA gKA TTc Agc gKA-
 gKA gKA gKA AKA gKA AKA gKA gKA AKA gKA gKA ggc gGA gCT gTc-
 gKA gKA g-3' (SEQ ID NO:151)

ON_1, ON_2, ON_3, e ON_4 como acima.

Desenho 20

		1	1	2	2	2	3	3	
	1	5	0	5	0	3	5	0	5
YYCAK	YSSYS	YYYDSSG	YYYSSYS	SSYS	YYYAEYFQGW	CGGTLVT	VSS	(SEQ ID NO:919)	
YYCAK	ZZZZZ(Z)	ZZZEDSSG	ZZZZZZZZZZ(Z)	ZZZAEZFQHW	CGGTLVT	VSS	(SEQ ID NO:152)		
R		YYYDSSG	YYY		AEYFQHW	CGGTLVT	VSS (JH1)		
		(SEQ ID NO:88)		(SEQ ID NO:66)					
	1	1			1	1			
9	9	0	0		0	1			
4	5	0	3abcde fghijklmnop q rstuvwx yz				0		

Diversidade de aminoácidos = 6,7 E 7
 Diversidade de ADN = 6,7 E 7
 Isentos de paragem = 100
 Isentos de Cys injustificadas = 100
 Isentos de paragem e Cys = 100%

O desenho 20 tem CDR3s de comprimento 33, 34, ou 35. O resíduo 94 pode ser K ou R. A ZZZZZZ(Z)ZZ::D3-22(2ª RF com seis Ys como Z)::ZZZZZZZ(Z)ZZZ::JH1 (com 1 Z). A PCR combinando (C35D322AJH1_T), (C34D322AJH1_T),

(C35D322AJH1_B), e (C34D322AJH1_B) permite diversidade de comprimentos bem como de sequências.

```
(C35D322AJH1_T)5'-GCAIGtTItaTItaCItgcIgcT aRg tmc tmt tmc tmc-  
tmt tmc tmt tmc tmc tmc tmc gac agc Icc ggc tmc tmc t-3' (SEQ ID NO:153)  
(C34D322AJH1_T)5'-GCAIGtTItaTItaCItgcIgcT aRg tmc tmc tmc tmt-  
tmc tmt tmc tmc tmc tmc gac agc Icc ggc tmc tmc t-3' (SEQ ID NO:154)  
(C35D322AJH1_B) 5'-cAg AgT Acc cTg gcc cCA gTg TTg qAA qKA Ttc Agc gKA-  
gKA gKA qKA AKA qKA AKA gKA gKA AKA gKA AKA gKA gKA gcc gga gCT gTc-  
gKA gKA g-3' (SEQ ID NO:920)  
(C34D322AJH1_B) 5'-cAg AgT Acc cTg gcc cCA gTg TTg qAA qKA Ttc Agc gKA-  
gKA gKA qKA AKA qKA AKA gKA gKA AKA AKA gKA gKA gcc gga gCT gTc-  
gKA gKA g-3' (SEQ ID NO:155)
```

Seleção contra codões de paragem:

Porque algumas destas bibliotecas têm codões NNK elas terão alguns codões de paragem TAG. Poderíamos remover os clones com TAG por clonagem do ADN amplificado num local *XbaI-BstEII* entre a sequência de sinal para um gene *bla* e a proteína real *bla* e expressar em células *Sup*⁰. As colónias *Bla*^R não contêm paragens TAG. Alternativamente poderíamos clonar os fragmentos de *XbaI-BstEII* depois de um gene de resistência à canamicina e selecionar quanto a *Kan*^R. Moveríamos depois a cassette *XbaI-BstEII* para a biblioteca de fagos.

Igualmente, porque o *wobbling* permite alguns codões de paragem, podemos melhorar a biblioteca por remoção dos clones com paragens por clonagem do ADN amplificado num local *XbaI-BstEII* entre a sequência de sinal para um gene *bla* e a proteína real *bla* e expressar em células *Sup*⁰. As colónias *Bla*^R não contêm paragens. Alternativamente podemos clonar os fragmentos de *XbaI-BstEII* depois de um gene de resistência à canamicina e selecionar quanto a *Kan*^R. Podemos depois mover a cassette *XbaI-BstEII* para a biblioteca de fagos.

Tabela 20: Regiões D de humano

!* para TAG; @ para TAA; \$ para TGA

D - Alinhamento de sequência de aminoácidos (RF: grelha de leitura)

		RF 1 (SEQ ID NO:156)	RF 2 (SEQ ID NO:157)	RF 3 (SEQ ID NO:158)	Usada nos desenhos
D1	1-1	GTTGT	VQLER	YNWNO	
		(SEQ ID NO:159)	(SEQ ID NO:160)	(SEQ ID NO:161)	
	1-7	GTTGT	V*LEL	YNWNY	
		(SEQ ID NO:159)	(SEQ ID NO:162)	(SEQ ID NO:163)	
	1-20	GTTGT	V*LER	YNWNO	
		(SEQ ID NO:164)	(SEQ ID NO:165)	(SEQ ID NO:166)	
	1-26	GIVCAT	V*WELL	YSGSY	
		(SEQ ID NO:167)	(SEQ ID NO:170)	(SEQ ID NO:168)	
D2	2-2	RIL**YQLLY	GYSSTSCYT	GIVVVPAAI	1, 5, 6, 7,
		(SEQ ID NO:169)	(SEQ ID NO:115)	(SEQ ID NO:170)	
	2-8	RILSWCHLY	GCTNSVCYT	GIVLMVYAI	20, 21, 22,
		(SEQ ID NO:171)	(SEQ ID NO:136)	(SEQ ID NO:172)	
	2-15	RIL*WW*LLL	GYSGGSCYS	GIVVVVAAT	25,
		(SEQ ID NO:173)	(SEQ ID NO:174)	(SEQ ID NO:175)	
	2-21	SILWELLF	AYCGGDCYS	HIVVVTAI	
		(SEQ ID NO:176)	(SEQ ID NO:177)	(SEQ ID NO:178)	
D3	3-3	VLRFLWLLY	YDFWEGYIT	ITIFGVVII	
		(SEQ ID NO:179)	(SEQ ID NO:180)	(SEQ ID NO:181)	
	3-9	VLRYPWLLQ	YYDILTGYM	ITIF*LVII	
		(SEQ ID NO:182)	(SEQ ID NO:81)	(SEQ ID NO:183)	
	3-10	VLRWFCGLLQ	YYGSGSYM	ITMVRGVII	
		(SEQ ID NO:184)	(SEQ ID NO:104)	(SEQ ID NO:185)	
	3-16	VLSRLGELSLY	YDYVWGSYXT	IMITFGGVIVI	8, 9, 14, 15, 17, 18
		(SEQ ID NO:186)	(SEQ ID NO:187)	(SEQ ID NO:188)	
	3-22	VLL***MLLL	YYDSGGYY	ITMIVVVIT	4, 19, 20
		(SEQ ID NO:189)	(SEQ ID NO:82)	(SEQ ID NO:190)	
D4	4-4	SLQEL	DYSNY	TTVT	
		(SEQ ID NO:191)	(SEQ ID NO:192)	(SEQ ID NO:193)	
	4-11	SLQEL	DYSNY	TTVT	
		(SEQ ID NO:194)	(SEQ ID NO:195)	(SEQ ID NO:196)	
	4-17	SLRRL	DYGDY	TTVT	
		(SEQ ID NO:197)	(SEQ ID NO:198)	(SEQ ID NO:199)	
	4-23	SLRWEL	DYGENS	TTVVT	

Tabela 3: Segmentos de JH de humano

JH - Alinhamento de sequências de aminoácidos

JH - Alinhamento de seqüências de aminoácidos

	H3		
	CDR3		
	100	110	
	I	FR4-----	Usada nos desenhos
JH1	---AEYFONWGQGLVTVSS	1-8,	(SEQ ID NO:66)
JH2	---YHYFDLWGRGTLVTVSS		(SEQ ID NO:67)
JH3	---AEDINGOGTMTVTVSS		(SEQ ID NO:2)
JH4	---YFDYWGQGLVTVSS		(SEQ ID NO:1)
JH5	---NWFDPWGQGLVTVSS		(SEQ ID NO:68)
JH6	YYYYYGNVWGQGLTVTVSS		(SEQ ID NO:3)
	123456789		

Tabela 10: ADN codificando V-5D2-8.2a-JH2 para wobbling

	CDR3.....	
! A E D T A V Y Y C A K D I V L M		
! gct gag gaT aCT GCA GtT taT taC tgc gct aag jez ezq jzz qzz ezj		
! W G Q G T T V T V S S	(SEQ ID NO:224)	
! tgg ggc cag ggt act acG GTC ACC gtc tcc agt-3'	(SEQ ID NO:225)	
! BstEII...		

Tabela 11: Trímeros que podem ser extraídos de segmentos D
de humano

Nas Tabelas 11-14, o uso de uma letra minúscula numa sequência de aminoácidos indica que um codão de paragem foi mudado para o resíduo listado como a letra minúscula. Por exemplo, na sequência de aminoácidos "yLE", um resíduo de Tyr foi introduzido em lugar de um codão de paragem.

GTT	D1-1.1.1	1
VQL	D1-1.2.1	2
YNW	D1-1.3.1	3
TTG	D1-1.1.2	4
QLE	D1-1.2.2	5
NWN	D1-1.3.2	6
TGT	D1-1.1.3	7
LER	D1-1.2.3	8
WND	D1-1.3.3	9

GIT	D1-7.1.1	10	
VyL	D1-7.2.1	11	*
ITG	D1-7.1.2	12	
yLE	D1-7.2.2	13	*
LEL	D1-7.2.3	14	
WNY	D1-7.3.3	15	
GIV	D1-26.1.1	16	
VyW	D1-26.2.1	17	*
YSG	D1-26.3.1	18	
IVG	D1-26.1.2	19	
yWE	D1-26.2.2	20	*
SGS	D1-26.3.2	21	
VGA	D1-26.1.3	22	
WEL	D1-26.2.3	23	
GSY	D1-26.3.3	24	
GAT	D1-26.1.4	25	
ELL	D1-26.2.4	26	
SY Y	D1-26.3.4	27	
RIL	D2-2.1.1	28	
GYC	D2-2.2.1	29	#
DIV	D2-2.3.1	30	
ILy	D2-2.1.2	31	*
YCS	D2-2.2.2	32	#
IVV	D2-2.3.2	33	
Lyy	D2-2.1.3	34	*
CSS	D2-2.2.3	35	#
VVV	D2-2.3.3	36	
yyY	D2-2.1.4	37	*
SST	D2-2.2.4	38	

VVP	D2-2.3.4	39	
yYQ	D2-2.1.5	40	*
STS	D2-2.2.5	41	
VPA	D2-2.3.5	42	
YQL	D2-2.1.6	43	
TSC	D2-2.2.6	44	#
PAA	D2-2.3.6	45	
QLL	D2-2.1.7	46	
SCY	D2-2.2.7	47	#
AAI	D2-2.3.7	48	
LLY	D2-2.1.8	49	
CYT	D2-2.2.8	50	#
ILY	D2-8.1.2	51	
YCT	D2-8.2.2	52	#
IVL	D2-8.3.2	53	
LYy	D2-8.1.3	54	*
CTN	D2-8.2.3	55	#
VLM	D2-8.3.3	56	
YyW	D2-8.1.4	57	*
TNG	D2-8.2.4	58	
LMV	D2-8.3.4	59	
yWC	D2-8.1.5	60	* #
NGV	D2-8.2.5	61	
MVY	D2-8.3.5	62	
WCM	D2-8.1.6	63	#
GVC	D2-8.2.6	64	#
VYA	D2-8.3.6	65	
CML	D2-8.1.7	66	#
VCY	D2-8.2.7	67	#

YAI	D2-8.3.7	68	
MLY	D2-8.1.8	69	
LyW	D2-15.1.3	70	*
CSG	D2-15.2.3	71	#
yWW	D2-15.1.4	72	*
SGG	D2-15.2.4	73	
WWy	D2-15.1.5	74	*
GGs	D2-15.2.5	75	
VVA	D2-15.3.5	76	
WyL	D2-15.1.6	77	*
GSC	D2-15.2.6	78	#
VAA	D2-15.3.6	79	
yLL	D2-15.1.7	80	*
AAT	D2-15.3.7	81	
LLL	D2-15.1.8	82	
CYS	D2-15.2.8	83	#
SIL	D2-21.1.1	84	
AYC	D2-21.2.1	85	#
HIV	D2-21.3.1	86	
ILW	D2-21.1.2	87	
YCG	D2-21.2.2	88	#
LWW	D2-21.1.3	89	
CGG	D2-21.2.3	90	#
WWw	D2-21.1.4	91	*
GGD	D2-21.2.4	92	
VVT	D2-21.3.4	93	
WwL	D2-21.1.5	94	*
GDC	D2-21.2.5	95	#
VTA	D2-21.3.5	96	

WLL	D2-21.1.6	97	*
DCY	D2-21.2.6	98	#
TAI	D2-21.3.6	99	
LLF	D2-21.1.7	100	
VLR	D3-3.1.1	101	
YYD	D3-3.2.1	102	
ITI	D3-3.3.1	103	
LRF	D3-3.1.2	104	
YDF	D3-3.2.2	105	
TIF	D3-3.3.2	106	
RFL	D3-3.1.3	107	
DFW	D3-3.2.3	108	
IFG	D3-3.3.3	109	
FLE	D3-3.1.4	110	
FWS	D3-3.2.4	111	
FGV	D3-3.3.4	112	
LEW	D3-3.1.5	113	
WSG	D3-3.2.5	114	
GVV	D3-3.3.5	115	
EWL	D3-3.1.6	116	
SGY	D3-3.2.6	117	
VVI	D3-3.3.6	118	
WLL	D3-3.1.7	119	
GYI	D3-3.2.7	120	
VII	D3-3.3.7	121	
YYT	D3-3.2.8	122	
LRY	D3-9.1.2	123	
YDI	D3-9.2.2	124	
RYF	D3-9.1.3	125	

DIL	D3-9.2.3	126	
IFy	D3-9.3.3	127	*
YFD	D3-9.1.4	128	
ILT	D3-9.2.4	129	
FyL	D3-9.3.4	130	*
FDW	D3-9.1.5	131	
LTG	D3-9.2.5	132	
yLV	D3-9.3.5	133	*
DWL	D3-9.1.6	134	
TGY	D3-9.2.6	135	
LVI	D3-9.3.6	136	
LLy	D3-9.1.8	137	*
YYN	D3-9.2.8	138	
VLL	D3-10.1.1	139	
YYY	D3-10.2.1	140	
ITM	D3-10.3.1	141	
LLW	D3-10.1.2	142	
YYG	D3-10.2.2	143	
TMV	D3-10.3.2	144	
LWF	D3-10.1.3	145	
YGS	D3-10.2.3	146	
MVR	D3-10.3.3	147	
WFG	D3-10.1.4	148	
GSG	D3-10.2.4	149	
VRG	D3-10.3.4	150	
FGE	D3-10.1.5	151	
RGV	D3-10.3.5	152	
GEL	D3-10.1.6	153	
GVI	D3-10.3.6	154	

VLw	D3-16.1.1	155	*
IMI	D3-16.3.1	156	
LwL	D3-16.1.2	157	*
YDY	D3-16.2.2	158	
MIT	D3-16.3.2	159	
wLR	D3-16.1.3	160	*
DYV	D3-16.2.3	161	
ITF	D3-16.3.3	162	
LRL	D3-16.1.4	163	
YVW	D3-16.2.4	164	
TFG	D3-16.3.4	165	
RLG	D3-16.1.5	166	
VWG	D3-16.2.5	167	
FGG	D3-16.3.5	168	
LGE	D3-16.1.6	169	
WGS	D3-16.2.6	170	
GGV	D3-16.3.6	171	
ELS	D3-16.1.8	172	
SYR	D3-16.2.8	173	
VIV	D3-16.3.8	174	
LSL	D3-16.1.9	175	
YRY	D3-16.2.9	176	
IVI	D3-16.3.9	177	
SLY	D3-16.1.10	178	
RYT	D3-16.2.10	179	
LLw	D3-22.1.2	180	*
TMI	D3-22.3.2	181	
Lwy	D3-22.1.3	182	*
YDS	D3-22.2.3	183	

MIV	D3-22.3.3	184	
wyy	D3-22.1.4	185	*
DSS	D3-22.2.4	186	
yyW	D3-22.1.5	187	*
SSG	D3-22.2.5	188	
yWL	D3-22.1.6	189	*
VIT	D3-22.3.7	190	
wLQ	D4-4.1.1	191	*
DYS	D4-4.2.1	192	
TTV	D4-4.3.1	193	
LQy	D4-4.1.2	194	*
YSN	D4-4.2.2	195	
TVT	D4-4.3.2	196	
QyL	D4-4.1.3	197	*
SNY	D4-4.2.3	198	
DYG	D4-17.2.1	199	
LRw	D4-17.1.2	200	*
YGD	D4-17.2.2	201	
RwL	D4-17.1.3	202	*
GDY	D4-17.2.3	203	
LRW	D4-23.1.2	204	
YGG	D4-23.2.2	205	
TVV	D4-23.3.2	206	
RWy	D4-23.1.3	207	*
GGN	D4-23.2.3	208	
GNS	D4-23.2.4	209	
VDT	D5-5.1.1	210	
WIQ	D5-5.2.1	211	
GYS	D5-5.3.1	212	

DTA	D5-5.1.2	213	
IQL	D5-5.2.2	214	
YSY	D5-5.3.2	215	
TAM	D5-5.1.3	216	
QLW	D5-5.2.3	217	
SYG	D5-5.3.3	218	
AMV	D5-5.1.4	219	
LWL	D5-5.2.4	220	
YGY	D5-5.3.4	221	
VDI	D5-12.1.1	222	
WIy	D5-12.2.1	223	*
IyW	D5-12.2.2	224	*
IVA	D5-12.1.3	225	
VAT	D5-12.1.4	226	
WLR	D5-12.2.4	227	
GYD	D5-12.3.4	228	
ATI	D5-12.1.5	229	
VEM	D5-24.1.1	230	
yRW	D5-24.2.1	231	*
RDG	D5-24.3.1	232	
EMA	D5-24.1.2	233	
RWL	D5-24.2.2	234	
DGY	D5-24.3.2	235	
MAT	D5-24.1.3	236	
WLQ	D5-24.2.3	237	
GYN	D5-24.3.3	238	
LQL	D5-24.2.4	239	
YNY	D5-24.3.4	240	
EYS	D6-6.1.1	241	

SIA	D6-6.2.1	242	
VyQ	D6-6.3.1	243	*
YSS	D6-6.1.2	244	
IAA	D6-6.2.2	245	
yQL	D6-6.3.2	246	*
SSS	D6-6.1.3	247	
AAR	D6-6.2.3	248	
QLV	D6-6.3.3	249	
GIA	D6-13.2.1	250	
yQQ	D6-13.3.2	251	*
AAA	D6-13.2.3	252	
QQL	D6-13.3.3	253	
SSW	D6-13.1.4	254	
AAG	D6-13.2.4	255	
SWY	D6-13.1.5	256	
IAV	D6-19.2.2	257	
yQW	D6-19.3.2	258	*
AVA	D6-19.2.3	259	
QWL	D6-19.3.3	260	
SGW	D6-19.1.4	261	
VAG	D6-19.2.4	262	
WLV	D6-19.3.4	263	
GWY	D6-19.1.5	264	
yLG	D7-27.2.1	265	*
NWG	D7-27.3.1	266	

Tabela 12: Tetrâmeros distintos que podem ser extraídos de
segmentos D de humano

GTTG	D1-1.1.1	(SEQ ID NO:257)	1
VQLE	D1-1.2.1	(SEQ ID NO:258)	2
YNWN	D1-1.3.1	(SEQ ID NO:259)	3
TTGT	D1-1.1.2	(SEQ ID NO:263)	4
QLER	D1-1.2.2	(SEQ ID NO:264)	5
NWND	D1-1.3.2	(SEQ ID NO:265)	6
GITG	D1-7.1.1	(SEQ ID NO:266)	7
VyLE	D1-7.2.1	(SEQ ID NO:267)	8
ITGT	D1-7.1.2	(SEQ ID NO:271)	9
yLEL	D1-7.2.2	(SEQ ID NO:272)	10
NWNY	D1-7.3.2	(SEQ ID NO:273)	11
yLER	D1-20.2.2	(SEQ ID NO:275)	12
GIVG	D1-26.1.1	(SEQ ID NO:276)	13
VyWE	D1-26.2.1	(SEQ ID NO:277)	14
YSGS	D1-26.3.1	(SEQ ID NO:278)	15
IVGA	D1-26.1.2	(SEQ ID NO:285)	16
yWEL	D1-26.2.2	(SEQ ID NO:286)	17
SGSY	D1-26.3.2	(SEQ ID NO:287)	18
VGAT	D1-26.1.3	(SEQ ID NO:291)	19
WELL	D1-26.2.3	(SEQ ID NO:292)	20
GSYY	D1-26.3.3	(SEQ ID NO:293)	21
RILy	D2-2.1.1	(SEQ ID NO:294)	22
GYCS	D2-2.2.1	(SEQ ID NO:295)	23
DIVV	D2-2.3.1	(SEQ ID NO:296)	24
ILyy	D2-2.1.2	(SEQ ID NO:303)	25
YCSSL	D2-2.2.2	(SEQ ID NO:304)	26

IVVV	D2-2.3.2	(SEQ ID NO:305)	27
LyyY	D2-2.1.3	(SEQ ID NO:312)	28
CSST	D2-2.2.3	(SEQ ID NO:313)	29
VVVP	D2-2.3.3	(SEQ ID NO:314)	30
yyYQ	D2-2.1.4	(SEQ ID NO:321)	31
SSTS	D2-2.2.4	(SEQ ID NO:322)	32
VVPA	D2-2.3.4	(SEQ ID NO:323)	33
yYQL	D2-2.1.5	(SEQ ID NO:330)	34
STSC	D2-2.2.5	(SEQ ID NO:331)	35
VPAA	D2-2.3.5	(SEQ ID NO:332)	36
YQLL	D2-2.1.6	(SEQ ID NO:338)	37
TSCY	D2-2.2.6	(SEQ ID NO:339)	38
PAAI	D2-2.3.6	(SEQ ID NO:340)	39
QLLY	D2-2.1.7	(SEQ ID NO:343)	40
SCYT	D2-2.2.7	(SEQ ID NO:344)	41
RILY	D2-8.1.1	(SEQ ID NO:345)	42
GYCT	D2-8.2.1	(SEQ ID NO:346)	43
DIVL	D2-8.3.1	(SEQ ID NO:347)	44
ILYy	D2-8.1.2	(SEQ ID NO:354)	45
YCTN	D2-8.2.2	(SEQ ID NO:355)	46
IVLM	D2-8.3.2	(SEQ ID NO:356)	47
LYyW	D2-8.1.3	(SEQ ID NO:363)	48
CTNG	D2-8.2.3	(SEQ ID NO:364)	49
VLMV	D2-8.3.3	(SEQ ID NO:365)	50
YyWC	D2-8.1.4	(SEQ ID NO:372)	51
TNGV	D2-8.2.4	(SEQ ID NO:373)	52
LMVY	D2-8.3.4	(SEQ ID NO:374)	53
yWCM	D2-8.1.5	(SEQ ID NO:381)	54
NGVC	D2-8.2.5	(SEQ ID NO:382)	55

MVYA	D2-8.3.5	(SEQ ID NO:383)	56
WCML	D2-8.1.6	(SEQ ID NO:389)	57
GVCY	D2-8.2.6	(SEQ ID NO:390)	58
VYAI	D2-8.3.6	(SEQ ID NO:391)	59
CMLY	D2-8.1.7	(SEQ ID NO:394)	60
VCYT	D2-8.2.7	(SEQ ID NO:395)	61
ILyW	D2-15.1.2	(SEQ ID NO:401)	62
YCSG	D2-15.2.2	(SEQ ID NO:402)	63
LyWW	D2-15.1.3	(SEQ ID NO:409)	64
CSGG	D2-15.2.3	(SEQ ID NO:410)	65
VVW	D2-15.3.3	(SEQ ID NO:411)	66
yWWy	D2-15.1.4	(SEQ ID NO:418)	67
SGGS	D2-15.2.4	(SEQ ID NO:419)	68
VVVA	D2-15.3.4	(SEQ ID NO:420)	69
WWyL	D2-15.1.5	(SEQ ID NO:427)	70
GGSC	D2-15.2.5	(SEQ ID NO:428)	71
WAA	D2-15.3.5	(SEQ ID NO:429)	72
WyLL	D2-15.1.6	(SEQ ID NO:435)	73
GSCY	D2-15.2.6	(SEQ ID NO:436)	74
VAAT	D2-15.3.6	(SEQ ID NO:437)	75
yLLL	D2-15.1.7	(SEQ ID NO:440)	76
SCYS	D2-15.2.7	(SEQ ID NO:441)	77
SILW	D2-21.1.1	(SEQ ID NO:442)	78
AYCG	D2-21.2.1	(SEQ ID NO:443)	79
HIW	D2-21.3.1	(SEQ ID NO:444)	80
ILWW	D2-21.1.2	(SEQ ID NO:451)	81
YCGG	D2-21.2.2	(SEQ ID NO:452)	82
LWWw	D2-21.1.3	(SEQ ID NO:459)	83
CGGD	D2-21.2.3	(SEQ ID NO:460)	84

VVVT	D2-21.3.3	(SEQ ID NO:461)	85
WWwL	D2-21.1.4	(SEQ ID NO:468)	86
GGDC	D2-21.2.4	(SEQ ID NO:469)	87
VVTA	D2-21.3.4	(SEQ ID NO:470)	88
WwLL	D2-21.1.5	(SEQ ID NO:476)	89
GDCY	D2-21.2.5	(SEQ ID NO:477)	90
VTAI	D2-21.3.5	(SEQ ID NO:478)	91
wLLF	D2-21.1.6	(SEQ ID NO:481)	92
DCYS	D2-21.2.6	(SEQ ID NO:482)	93
VLRF	D3-3.1.1	(SEQ ID NO:483)	94
YYDF	D3-3.2.1	(SEQ ID NO:484)	95
ITIF	D3-3.3.1	(SEQ ID NO:485)	96
LRFL	D3-3.1.2	(SEQ ID NO:492)	97
YDFW	D3-3.2.2	(SEQ ID NO:493)	98
TIFG	D3-3.3.2	(SEQ ID NO:494)	99
RFLE	D3-3.1.3	(SEQ ID NO:501)	100
DFWS	D3-3.2.3	(SEQ ID NO:502)	101
IFGV	D3-3.3.3	(SEQ ID NO:503)	102
FLEW	D3-3.1.4	(SEQ ID NO:510)	103
FWSG	D3-3.2.4	(SEQ ID NO:511)	104
FGVV	D3-3.3.4	(SEQ ID NO:512)	105
LEWL	D3-3.1.5	(SEQ ID NO:519)	106
WSGY	D3-3.2.5	(SEQ ID NO:520)	107
GWI	D3-3.3.5	(SEQ ID NO:521)	108
EWLL	D3-3.1.6	(SEQ ID NO:527)	109
SGYY	D3-3.2.6	(SEQ ID NO:528)	110
VVII	D3-3.3.6	(SEQ ID NO:529)	111
WLLY	D3-3.1.7	(SEQ ID NO:532)	112
GYYT	D3-3.2.7	(SEQ ID NO:533)	113

VLRY	D3-9.1.1	(SEQ ID NO:534)	114
YYDI	D3-9.2.1	(SEQ ID NO:535)	115
LRyF	D3-9.1.2	(SEQ ID NO:542)	116
YDIL	D3-9.2.2	(SEQ ID NO:543)	117
TIFy	D3-9.3.2	(SEQ ID NO:544)	118
RYFD	D3-9.1.3	(SEQ ID NO:551)	119
DILT	D3-9.2.3	(SEQ ID NO:552)	120
IFyL	D3-9.3.3	(SEQ ID NO:553)	121
YFDW	D3-9.1.4	(SEQ ID NO:560)	122
ILTG	D3-9.2.4	(SEQ ID NO:561)	123
FyLV	D3-9.3.4	(SEQ ID NO:562)	124
FDWL	D3-9.1.5	(SEQ ID NO:569)	125
LTGY	D3-9.2.5	(SEQ ID NO:570)	126
yLVI	D3-9.3.5	(SEQ ID NO:571)	127
DWLL	D3-9.1.6	(SEQ ID NO:577)	128
TGYY	D3-9.2.6	(SEQ ID NO:578)	129
LVII	D3-9.3.6	(SEQ ID NO:579)	130
WLLy	D3-9.1.7	(SEQ ID NO:582)	131
GYYN	D3-9.2.7	(SEQ ID NO:583)	132
VLLW	D3-10.1.1	(SEQ ID NO:584)	133
YYYG	D3-10.2.1	(SEQ ID NO:585)	134
ITMV	D3-10.3.1	(SEQ ID NO:586)	135
LLWF	D3-10.1.2	(SEQ ID NO:593)	136
YYGS	D3-10.2.2	(SEQ ID NO:594)	137
TMVR	D3-10.3.2	(SEQ ID NO:595)	138
LWFG	D3-10.1.3	(SEQ ID NO:602)	139
YGSG	D3-10.2.3	(SEQ ID NO:603)	140
MVRG	D3-10.3.3	(SEQ ID NO:604)	141
WFGE	D3-10.1.4	(SEQ ID NO:611)	142

GSGS	D3-10.2.4	(SEQ ID NO:612)	143
VRGV	D3-10.3.4	(SEQ ID NO:613)	144
FGEL	D3-10.1.5	(SEQ ID NO:620)	145
RGVI	D3-10.3.5	(SEQ ID NO:621)	146
GELL	D3-10.1.6	(SEQ ID NO:626)	147
GVII	D3-10.3.6	(SEQ ID NO:627)	148
ELLy	D3-10.1.7	(SEQ ID NO:630)	149
SYYN	D3-10.2.7	(SEQ ID NO:631)	156
VLwL	D3-16.1.1	(SEQ ID NO:632)	151
YYDY	D3-16.2.1	(SEQ ID NO:633)	152
IMIT	D3-16.3.1	(SEQ ID NO:634)	153
LwLR	D3-16.1.2	(SEQ ID NO:641)	154
YDYV	D3-16.2.2	(SEQ ID NO:642)	155
MITF	D3-16.3.2	(SEQ ID NO:643)	156
wLRL	D3-16.1.3	(SEQ ID NO:650)	157
DYVW	D3-16.2.3	(SEQ ID NO:651)	158
ITFG	D3-16.3.3	(SEQ ID NO:652)	159
LRLG	D3-16.1.4	(SEQ ID NO:659)	160
YVWG	D3-16.2.4	(SEQ ID NO:660)	161
TFGG	D3-16.3.4	(SEQ ID NO:661)	162
RLGE	D3-16.1.5	(SEQ ID NO:668)	163
VWGS	D3-16.2.5	(SEQ ID NO:669)	164
FGGV	D3-16.3.5	(SEQ ID NO:670)	165
LGEL	D3-16.1.6	(SEQ ID NO:677)	166
WGSY	D3-16.2.6	(SEQ ID NO:678)	167
GGVI	D3-16.3.6	(SEQ ID NO:679)	168
GELS	D3-16.1.7	(SEQ ID NO:686)	169
GSYR	D3-16.2.7	(SEQ ID NO:687)	170
GVIV	D3-16.3.7	(SEQ ID NO:688)	171

ELSL	D3-16.1.8	(SEQ ID NO:694)	172
SYRY	D3-16.2.8	(SEQ ID NO:695)	173
VIVI	D3-16.3.8	(SEQ ID NO:696)	174
LSLY	03-16.1.9	(SEQ ID NO:699)	175
YRYT	D3-16.2.9	(SEQ ID NO:700)	176
VLLw	D3-22.1.1	(SEQ ID NO:701)	177
YYYD	D3-22.2.1	(SEQ ID NO:702)	178
ITMI	D3-22.3.1	(SEQ ID NO:703)	179
LLwy	D3-22.1.2	(SEQ ID NO:710)	180
YYDS	D3-22.2.2	(SEQ ID NO:711)	181
TMIV	D3-22.3.2	(SEQ ID NO:712)	182
Lwyy	D3-22.1.3	(SEQ ID NO:719)	183
YDSS	D3-22.2.3	(SEQ ID NO:720)	184
MIVV	D3-22.3.3	(SEQ ID NO:721)	185
wyyW	D3-22.1.4	(SEQ ID NO:728)	186
DSSG	D3-22.2.4	(SEQ ID NO:729)	187
yyWL	D3-22.1.5	(SEQ ID NO:736)	188
SSGY	D3-22.2.5	(SEQ ID NO:737)	189
VVVI	D3-22.3.5	(SEQ ID NO:738)	190
yWLL	D3-22.1.6	(SEQ ID NO:744)	191
WIT	D3-22.3.6	(SEQ ID NO:745)	192
WLLL	D3-22.1.7	(SEQ ID NO:748)	193
GYYY	D3-22.2.7	(SEQ ID NO:749)	194
wLQy	D4-4.1.1	(SEQ ID NO:750)	195
DYSN	D4-4.2.1	(SEQ ID NO:751)	196
TTVT	D4-4.3.1	(SEQ ID NO:752)	197
LQyL	D4-4.1.2	(SEQ ID NO:755)	198
YSNY	D4-4.2.2	(SEQ ID NO:756)	199
wLRw	D4-17.1.1	(SEQ ID NO:757)	200

DYGD	D4-17.2.1	(SEQ ID NO:758)	201
LRwL	D4-17.1.2	(SEQ ID NO:761)	202
YGDY	D4-17.2.2	(SEQ ID NO:762)	203
wLRW	D4-23.1.1	(SEQ ID NO:763)	204
DYGG	D4-23.2.1	(SEQ ID NO:764)	205
TTVV	D4-23.3.1	(SEQ ID NO:765)	206
LRWy	D4-23.1.2	(SEQ ID NO:771)	207
YGGN	D4-23.2.2	(SEQ ID NO:772)	208
TVVT	D4-23.3.2	(SEQ ID NO:773)	209
RWyL	D4-23.1.3	(SEQ ID NO:776)	210
GGNS	D4-23.2.3	(SEQ ID NO:777)	211
VDTA	D5-5.1.1	(SEQ ID NO:778)	212
WIQL	D5-5.2.1	(SEQ ID NO:779)	213
GYSY	D5-5.3.1	(SEQ ID NO:780)	214
DTAM	D5-5.1.2	(SEQ ID NO:787)	215
IQLW	D5-5.2.2	(SEQ ID NO:788)	216
YSYG	D5-5.3.2	(SEQ ID NO:789)	217
TAMV	D5-5.1.3	(SEQ ID NO:793)	218
QLWL	D5-5.2.3	(SEQ ID NO:794)	219
SYGY	D5-5.3.3	(SEQ ID NO:795)	220
VDIV	D5-12.1.1	(SEQ ID NO:796)	221
WIyW	D5-12.2.1	(SEQ ID NO:797)	222
GYSG	D5-12.3.1	(SEQ ID NO:798)	223
DIVA	D5-12.1.2	(SEQ ID NO:805)	224
IyWL	D5-12.2.2	(SEQ ID NO:806)	225
YSGY	D5-12.3.2	(SEQ ID NO:807)	226
IVAT	D5-12.1.3	(SEQ ID NO:814)	227
yWLR	D5-12.2.3	(SEQ ID NO:815)	228
SGYD	D5-12.3.3	(SEQ ID NO:816)	229

VATI	D5-12.1.4	(SEQ ID NO:820)	230
WLRL	D5-12.2.4	(SEQ ID NO:821)	231
GYDY	D5-12.3.4	(SEQ ID NO:822)	232
VEMA	D5-24.1.1	(SEQ ID NO:823)	233
yRWL	D5-24.2.1	(SEQ ID NO:824)	234
RDGY	D5-24.3.1	(SEQ ID NO:825)	235
EMAT	D5-24.1.2	(SEQ ID NO:832)	236
RWLQ	D5-24.2.2	(SEQ ID NO:833)	237
DGYN	D5-24.3.2	(SEQ ID NO:834)	238
MATI	D5-24.1.3	(SEQ ID NO:838)	239
WLQL	D5-24.2.3	(SEQ ID NO:839)	240
GYNY	D5-24.3.3	(SEQ ID NO:840)	241
EYSS	D6-6.1.1	(SEQ ID NO:841)	242
SIAA	D6-6.2.1	(SEQ ID NO:842)	243
VyQL	D6-6.3.1	(SEQ ID NO:843)	244
YSSS	D6-6.1.2	(SEQ ID NO:848)	245
IAAR	D6-6.2.2	(SEQ ID NO:849)	246
yQLV	D6-6.3.2	(SEQ ID NO:850)	247
SSSS	D6-6.1.3	(SEQ ID NO:852)	248
GYSS	D6-13.1.1	(SEQ ID NO:853)	249
GIAA	D6-13.2.1	(SEQ ID NO:854)	250
VyQQ	D6-13.3.1	(SEQ ID NO:855)	251
IAAA	D6-13.2.2	(SEQ ID NO:862)	252
yQQQL	D6-13.3.2	(SEQ ID NO:863)	253
SSSW	D6-13.1.3	(SEQ ID NO:868)	254
AAAG	D6-13.2.3	(SEQ ID NO:869)	255
QQLV	D6-13.3.3	(SEQ ID NO:870)	256
SSWY	D6-13.1.4	(SEQ ID NO:872)	257
GIAV	D6-19.2.1	(SEQ ID NO:873)	258

VyQW	D6-19.3.1	(SEQ ID NO:874)	259
YSSG	D6-19.1.2	(SEQ ID NO:881)	260
IAVA	D6-19.2.2	(SEQ ID NO:882)	261
yQWL	D6-19.3.2	(SEQ ID NO:883)	262
SSGW	D6-19.1.3	(SEQ ID NO:888)	263
AVAG	D6-19.2.3	(SEQ ID NO:889)	264
QWLV	D6-19.3.3	(SEQ ID NO:890)	265
SGWY	D6-19.1.4	(SEQ ID NO:892)	266

Tabela 13: Pentâmeros que podem ser extraídos de segmentos
de D de humano

GTTGT	D1-1.1.1	(SEQ ID NO:260)	1
VQLER	D1-1.2.1	(SEQ ID NO:261)	2
YNWND	D1-1.3.1	(SEQ ID NO:262)	3
GITGT	D1-7.1.1	(SEQ ID NO:268)	4
VyLEL	D1-7.2.1	(SEQ ID NO:269)	5
YNWNY	D1-7.3.1	(SEQ ID NO:270)	6
VyLER	D1-20.2.1	(SEQ ID NO:274)	7
GIVGA	D1-26.1.1	(SEQ ID NO:279)	8
VyWEL	D1-26.2.1	(SEQ ID NO:280)	9
YSGSY	D1-26.3.1	(SEQ ID NO:281)	10
IVGAT	D1-26.1.2	(SEQ ID NO:288)	11
yWELL	D1-26.2.2	(SEQ ID NO:289)	12
SGSYY	D1-26.3.2	(SEQ ID NO:290)	13
RILyy	D2-2.1.1	(SEQ ID NO:297)	14
GYCSS	D2-2.2.1	(SEQ ID NO:298)	15
DIWV	D2-2.3.1	(SEQ ID NO:299)	16
ILyyY	D2-2.1.2	(SEQ ID NO:306)	17
YCSST	D2-2.2.2	(SEQ ID NO:307)	18

IVVVP	D2-2.3.2	(SEQ ID NO:308)	19
LyYQ	D2-2.1.3	(SEQ ID NO:315)	20
CSSTS	D2-2.2.3	(SEQ ID NO:316)	21
VVVP	D2-2.3.3	(SEQ ID NO:317)	22
yyYQL	D2-2.1.4	(SEQ ID NO:324)	23
SSTSC	D2-2.2.4	(SEQ ID NO:325)	24
VVPAA	D2-2.3.4	(SEQ ID NO:326)	25
yYQLL	D2-2.1.5	(SEQ ID NO:333)	26
STSCY	D2-2.2.5	(SEQ ID NO:334)	27
VPAAI	D2-2.3.5	(SEQ ID NO:335)	28
YQLLY	D2-2.1.6	(SEQ ID NO:341)	29
TSCYT	D2-2.2.6	(SEQ ID NO:342)	30
RILYy	D2-8.1.1	(SEQ ID NO:348)	31
GYCTN	D2-8.2.1	(SEQ ID NO:349)	32
DIVLM	D2-8.3.1	(SEQ ID NO:350)	33
ILYyW	D2-8.1.2	(SEQ ID NO:357)	34
YCTNG	D2-8.2.2	(SEQ ID NO:358)	35
IVLMV	D2-8.3.2	(SEQ ID NO:359)	36
LYyWC	D2-8.1.3	(SEQ ID NO:366)	37
CTNGV	D2-8.2.3	(SEQ ID NO:367)	38
VLMVY	D2-8.3.3	(SEQ ID NO:368)	39
YyWCM	D2-8.1.4	(SEQ ID NO:375)	40
TNGVC	D2-8.2.4	(SEQ ID NO:376)	41
LMVYA	D2-8.3.4	(SEQ ID NO:377)	42
yWCML	D2-8.1.5	(SEQ ID NO:384)	43
NGVCY	D2-8.2.5	(SEQ ID NO:385)	44
MVYAI	D2-8.3.5	(SEQ ID NO:386)	45
WCMLY	D2-8.1.6	(SEQ ID NO:392)	46
GVCYT	D2-8.2.6	(SEQ ID NO:393)	47

RILyW	D2-15.1.1	(SEQ ID NO:396)	48
GYCSG	D2-15.2.1	(SEQ ID NO:397)	49
ILyWW	D2-15.1.2	(SEQ ID NO:403)	50
YCSGG	D2-15.2.2	(SEQ ID NO:404)	51
IVVVV	D2-15.3.2	(SEQ ID NO:405)	52
LyWWy	D2-15.1.3	(SEQ ID NO:412)	53
CSGGS	D2-15.2.3	(SEQ ID NO:413)	54
VVVVA	D2-15.3.3	(SEQ ID NO:414)	55
yWWyL	D2-15.1.4	(SEQ ID NO:421)	56
SGGSC	D2-15.2.4	(SEQ ID NO:422)	57
VVVAA	D2-15.3.4	(SEQ ID NO:423)	58
WWyLL	D2-15.1.5	(SEQ ID NO:430)	59
GGSCY	D2-15.2.5	(SEQ ID NO:431)	60
VVAAT	D2-15.3.5	(SEQ ID NO:432)	61
WyLLL	D2-15.1.6	(SEQ ID NO:438)	62
GSCYS	D2-15.2.6	(SEQ ID NO:439)	63
SILWW	D2-21.1.1	(SEQ ID NO:445)	64
AYCGG	D2-21.2.1	(SEQ ID NO:446)	65
HIVVV	D2-21.3.1	(SEQ ID NO:447)	66
ILWWw	D2-21.1.2	(SEQ ID NO:453)	67
YCGGD	D2-21.2.2	(SEQ ID NO:454)	68
IVVVT	D2-21.3.2	(SEQ ID NO:455)	69
LWWwL	D2-21.1.3	(SEQ ID NO:462)	70
CGGDC	D2-21.2.3	(SEQ ID NO:463)	71
VVVTA	D2-21.3.3	(SEQ ID NO:464)	72
WWwLL	D2-21.1.4	(SEQ ID NO:471)	73
GGDCY	D2-21.2.4	(SEQ ID NO:472)	74
VVTAI	D2-21.3.4	(SEQ ID NO:473)	75
WwLLF	D2-21.1.5	(SEQ ID NO:479)	76

GDCYS	D2-21.2.5	(SEQ ID NO:480)	77
VLRFLL	D3-3.1.1	(SEQ ID NO:486)	78
YYDFW	D3-3.2.1	(SEQ ID NO:487)	79
ITIFG	D3-3.3.1	(SEQ ID NO:488)	80
LRFLE	D3-3.1.2	(SEQ ID NO:495)	81
YDFWS	D3-3.2.2	(SEQ ID NO:496)	82
TIFGV	D3-3.3.2	(SEQ ID NO:497)	83
RFLEW	D3-3.1.3	(SEQ ID NO:504)	84
DFWSG	D3-3.2.3	(SEQ ID NO:505)	85
IFGVV	D3-3.3.3	(SEQ ID NO:506)	86
FLEWL	D3-3.1.4	(SEQ ID NO:513)	87
FWSGY	D3-3.2.4	(SEQ ID NO:514)	88
FGVVI	D3-3.3.4	(SEQ ID NO:515)	89
LEWLL	D3-3.1.5	(SEQ ID NO:522)	90
WSGY	D3-3.2.5	(SEQ ID NO:523)	91
GVVII	D3-3.3.5	(SEQ ID NO:524)	92
EWLLY	D3-3.1.6	(SEQ ID NO:530)	93
SGYYT	D3-3.2.6	(SEQ ID NO:531)	94
VLRYF	D3-9.1.1	(SEQ ID NO:536)	95
YYDIL	D3-9.2.1	(SEQ ID NO:537)	96
ITIFy	D3-9.3.1	(SEQ ID NO:538)	97
LRYFD	D3-9.1.2	(SEQ ID NO:545)	98
YDILT	D3-9.2.2	(SEQ ID NO:546)	99
TIFyL	D3-9.3.2	(SEQ ID NO:547)	100
RYFDW	D3-9.1.3	(SEQ ID NO:554)	101
DILTG	D3-9.2.3	(SEQ ID NO:555)	102
IFyLV	D3-9.3.3	(SEQ ID NO:556)	103
YFDWL	D3-9.1.4	(SEQ ID NO:563)	104
ILTGY	D3-9.2.4	(SEQ ID NO:564)	105

FyLVI	D3-9.3.4	(SEQ ID NO:565)	106
FDWLL	D3-9.1.5	(SEQ ID NO:572)	107
LTGY Y	D3-9.2.5	(SEQ ID NO:573)	108
yLVII	D3-9.3.5	(SEQ ID NO:574)	109
DWLLy	D3-9.1.6	(SEQ ID NO:580)	110
TGYYN	D3-9.2.6	(SEQ ID NO:581)	111
VLLWF	D3-10.1.1	(SEQ ID NO:587)	112
YYYGS	D3-10.2.1	(SEQ ID NO:588)	113
ITMVR	D3-10.3.1	(SEQ ID NO:589)	114
LLWFG	D3-10.1.2	(SEQ ID NO:596)	115
YYGSG	D3-10.2.2	(SEQ ID NO:597)	116
TMVRG	D3-10.3.2	(SEQ ID NO:598)	117
LWFGE	D3-10.1.3	(SEQ ID NO:605)	118
YGS GS	D3-10.2.3	(SEQ ID NO:606)	119
MVRGV	D3-10.3.3	(SEQ ID NO:607)	120
WFGEL	D3-10.1.4	(SEQ ID NO:614)	121
GSGSY	D3-10.2.4	(SEQ ID NO:615)	122
VRGVI	D3-10.3.4	(SEQ ID NO:616)	123
FGELL	D3-10.1.5	(SEQ ID NO:622)	124
RGVII	D3-10.3.5	(SEQ ID NO:623)	125
GELLy	D3-10.1.6	(SEQ ID NO:628)	126
GSYYN	D3-10.2.6	(SEQ ID NO:629)	127
VLwLR	D3-16.1.1	(SEQ ID NO:635)	128
YYDYV	D3-16.2.1	(SEQ ID NO:636)	129
IMITF	D3-16.3.1	(SEQ ID NO:637)	130
LwLRL	D3-16.1.2	(SEQ ID NO:644)	131
YDYVW	D3-16.2.2	(SEQ ID NO:645)	132
MITFG	D3-16.3.2	(SEQ ID NO:646)	133
wLRLG	D3-16.1.3	(SEQ ID NO:653)	134

DYVWG	D3-16.2.3	(SEQ ID NO:654)	135
ITFGG	D3-16.3.3	(SEQ ID NO:655)	136
LRLGE	D3-16.1.4	(SEQ ID NO:662)	137
YVWGS	D3-16.2.4	(SEQ ID NO:663)	138
TFGGV	D3-16.3.4	(SEQ ID NO:664)	139
RLGEL	D3-16.1.5	(SEQ ID NO:671)	140
VWGSY	D3-16.2.5	(SEQ ID NO:672)	141
FGGVI	D3-16.3.5	(SEQ ID NO:673)	142
LGELS	D3-16.1.6	(SEQ ID NO:680)	143
WGSYR	D3-16.2.6	(SEQ ID NO:681)	144
GGVIV	D3-16.3.6	(SEQ ID NO:682)	145
GELSL	D3-16.1.7	(SEQ ID NO:689)	146
GSYRY	D3-16.2.7	(SEQ ID NO:690)	147
GVIVI	D3-16.3.7	(SEQ ID NO:691)	148
ELSLY	D3-16.1.8	(SEQ ID NO:697)	149
SYRYT	D3-16.2.8	(SEQ ID NO:698)	150
VLLwy	D3-22.1.1	(SEQ ID NO:704)	151
YYYDS	D3-22.2.1	(SEQ ID NO:705)	152
ITMIV	D3-22.3.1	(SEQ ID NO:706)	153
LLwyy	D3-22.1.2	(SEQ ID NO:713)	154
YYDSS	D3-22.2.2	(SEQ ID NO:714)	155
TMIVV	D3-22.3.2	(SEQ ID NO:715)	156
LwyyW	D3-22.1.3	(SEQ ID NO:722)	157
YDSSG	D3-22.2.3	(SEQ ID NO:723)	158
MIVVV	D3-22.3.3	(SEQ ID NO:724)	159
wyyWL	D3-22.1.4	(SEQ ID NO:730)	160
DSSGY	D3-22.2.4	(SEQ ID NO:731)	161
IVVVI	D3-22.3.4	(SEQ ID NO:732)	162
yyWLL	D3-22.1.5	(SEQ ID NO:739)	163

SSGY	D3-22.2.5	(SEQ ID NO:740)	164
VVVIT	D3-22.3.5	(SEQ ID NO:741)	165
yWLLL	D3-22.1.6	(SEQ ID NO:746)	166
SGYYY	D3-22.2.6	(SEQ ID NO:747)	167
wLQyL	D4-4.1.1	(SEQ ID NO:753)	168
DYSNY	D4-4.2.1	(SEQ ID NO:754)	169
wLRwL	D4-17.1.1	(SEQ ID NO:759)	170
DYGDY	D4-17.2.1	(SEQ ID NO:760)	171
wLRWy	D4-23.1.1	(SEQ ID NO:766)	172
DYGGN	D4-23.2.1	(SEQ ID NO:767)	173
TTVVT	D4-23.3.1	(SEQ ID NO:768)	174
LRWyL	D4-23.1.2	(SEQ ID NO:774)	175
YGGNS	D4-23.2.2	(SEQ ID NO:775)	176
VDTAM	D5-5.1.1	(SEQ ID NO:781)	177
WIQLW	D5-5.2.1	(SEQ ID NO:782)	178
GYSYG	D5-5.3.1	(SEQ ID NO:783)	179
DTAMV	D5-5.1.2	(SEQ ID NO:790)	180
IQLWL	D5-5.2.2	(SEQ ID NO:791)	181
YSYGY	D5-5.3.2	(SEQ ID NO:792)	182
VDIVA	D5-12.1.1	(SEQ ID NO:799)	183
WIyWL	D5-12.2.1	(SEQ ID NO:800)	184
GYSGY	D5-12.3.1	(SEQ ID NO:801)	185
DIVAT	D5-12.1.2	(SEQ ID NO:808)	186
IyWLR	D5-12.2.2	(SEQ ID NO:809)	187
YSGYD	D5-12.3.2	(SEQ ID NO:810)	188
IVATI	D5-12.1.3	(SEQ ID NO:817)	189
yWLRL	D5-12.2.3	(SEQ ID NO:818)	190
SGYDY	D5-12.3.3	(SEQ ID NO:819)	191
VEMAT	D5-24.1.1	(SEQ ID NO:826)	192

yRWLQ	D5-24.2.1	(SEQ ID NO:827)	193
RDGYN	D5-24.3.1	(SEQ ID NO:828)	194
EMATI	D5-24.1.2	(SEQ ID NO:835)	195
RWLQL	D5-24.2.2	(SEQ ID NO:836)	196
DGYNY	D5-24.3.2	(SEQ ID NO:837)	197
EYSSS	D6-6.1.1	(SEQ ID NO:844)	198
SIAAR	D6-6.2.1	(SEQ ID NO:845)	199
VyQLV	D6-6.3.1	(SEQ ID NO:846)	200
YSSSS	D6-6.1.2	(SEQ ID NO:851)	201
GYSSS	D6-13.1.1	(SEQ ID NO:856)	202
GIAAA	D6-13.2.1	(SEQ ID NO:857)	203
VyQQL	D6-13.3.1	(SEQ ID NO:858)	204
YSSSW	D6-13.1.2	(SEQ ID NO:864)	205
IAAAG	D6-13.2.2	(SEQ ID NO:865)	206
yQQLV	D6-13.3.2	(SEQ ID NO:866)	207
SSSWY	D6-13.1.3	(SEQ ID NO:871)	208
GYSSG	D6-19.1.1	(SEQ ID NO:875)	209
GIAVA	D6-19.2.1	(SEQ ID NO:876)	210
VyQWL	D6-19.3.1	(SEQ ID NO:877)	211
YSSGW	D6-19.1.2	(SEQ ID NO:884)	212
IAVAG	D6-19.2.2	(SEQ ID NO:885)	213
yQWL	D6-19.3.2	(SEQ ID NO:886)	214
SSGWY	D6-19.1.3	(SEQ ID NO:891)	215

Tabela 14: Todos os hexâmeros que podem ser extraídos de segmentos D de humano

GIVGAT	D1-26.1.1	(SEQ ID NO:282)	1
VyWELL	D1-26.2.1	(SEQ ID NO:283)	2
YSGSY	D1-26.3.1	(SEQ ID NO:284)	3

RILyyY	D2-2.1.1	(SEQ ID NO:300)	4
GYCSST	D2-2.2.1	(SEQ ID NO:301)	5
DIVVVP	D2-2.3.1	(SEQ ID NO:302)	6
ILyyYQ	D2-2.1.2	(SEQ ID NO:309)	7
YCSSTS	D2-2.2.2	(SEQ ID NO:310)	8
IVVVPA	D2-2.3.2	(SEQ ID NO:311)	9
LyyYQL	D2-2.1.3	(SEQ ID NO:318)	10
CSSTSC	D2-2.2.3	(SEQ ID NO:319)	11
VVVPAA	D2-2.3.3	(SEQ ID NO:320)	12
yyYQLL	D2-2.1.4	(SEQ ID NO:327)	13
SSTSCY	D2-2.2.4	(SEQ ID NO:328)	14
VVPAAI	D2-2.3.4	(SEQ ID NO:329)	15
yYQLLY	D2-2.1.5	(SEQ ID NO:336)	16
STSCYT	D2-2.2.5	(SEQ ID NO:337)	17
RILYyW	D2-8.1.1	(SEQ ID NO:351)	18
GYCTNG	D2-8.2.1	(SEQ ID NO:352)	19
DIVLMV	D2-8.3.1	(SEQ ID NO:353)	20
ILYyWC	D2-8.1.2	(SEQ ID NO:360)	21
YCTNGV	D2-8.2.2	(SEQ ID NO:361)	22
IVLMVY	D2-8.3.2	(SEQ ID NO:362)	23
LYyWCM	D2-8.1.3	(SEQ ID NO:369)	24
CTNGVC	D2-8.2.3	(SEQ ID NO:370)	25
VLMVYA	D2-8.3.3	(SEQ ID NO:371)	26
YyWCML	D2-8.1.4	(SEQ ID NO:378)	27
TNGVCY	D2-8.2.4	(SEQ ID NO:379)	28
LMVYAI	D2-8.3.4	(SEQ ID NO:380)	29
yWCMLY	D2-8.1.5	(SEQ ID NO:387)	30
NGVCYT	D2-8.2.5	(SEQ ID NO:388)	31
RILyWW	D2-15.1.1	(SEQ ID NO:398)	32

GYCSGG	D2-15.2.1	(SEQ ID NO:399)	33
DIVVVV	D2-15.3.1	(SEQ ID NO:400)	34
ILyWWy	D2-15.1.2	(SEQ ID NO:406)	35
YCSGGS	D2-15.2.2	(SEQ ID NO:407)	36
IVVVVA	D2-15.3.2	(SEQ ID NO:408)	37
LyWWyL	D2-15.1.3	(SEQ ID NO:415)	38
CSGGSC	D2-15.2.3	(SEQ ID NO:416)	39
VVVVAA	D2-15.3.3	(SEQ ID NO:417)	40
yWWyLL	D2-15.1.4	(SEQ ID NO:424)	41
SGGSCY	D2-15.2.4	(SEQ ID NO:425)	42
VVVAAT	D2-15.3.4	(SEQ ID NO:426)	43
WWyLLL	D2-15.1.5	(SEQ ID NO:433)	44
GGSCYS	D2-15.2.5	(SEQ ID NO:434)	45
SILWWw	D2-21.1.1	(SEQ ID NO:448)	46
AYCGGD	D2-21.2.1	(SEQ ID NO:449)	47
HIVVVT	D2-21.3.1	(SEQ ID NO:450)	48
ILWWwL	D2-21.1.2	(SEQ ID NO:456)	49
YCGGDC	D2-21.2.2	(SEQ ID NO:457)	50
IVVVTA	D2-21.3.2	(SEQ ID NO:458)	51
LWWwLL	D2-21.1.3	(SEQ ID NO:465)	52
CGGDCY	D2-21.2.3	(SEQ ID NO:466)	53
VVVTAI	D2-21.3.3	(SEQ ID NO:467)	54
WWwLLF	D2-21.1.4	(SEQ ID NO:474)	55
GGDCYS	D2-21.2.4	(SEQ ID NO:475)	56
VLRFLE	D3-3.1.1	(SEQ ID NO:489)	57
YYDFWS	D3-3.2.1	(SEQ ID NO:490)	58
ITIFGV	D3-3.3.1	(SEQ ID NO:491)	59
LRFLEW	D3-3.1.2	(SEQ ID NO:498)	60
YDFWSG	D3-3.2.2	(SEQ ID NO:499)	61

TIFGVV	D3-3.3.2	(SEQ ID NO:500)	62
RFLEWL	D3-3.1.3	(SEQ ID NO:507)	63
DFWSGY	D3-3.2.3	(SEQ ID NO:508)	64
IFGVVI	D3-3.3.3	(SEQ ID NO:509)	65
FLEWLL	D3-3.1.4	(SEQ ID NO:516)	66
FWSGYY	D3-3.2.4	(SEQ ID NO:517)	67
FGVVII	D3-3.3.4	(SEQ ID NO:518)	68
LEWLLY	D3-3.1.5	(SEQ ID NO:525)	69
WSGYYT	D3-3.2.5	(SEQ ID NO:526)	70
VLRYFD	D3-9.1.1	(SEQ ID NO:539)	71
YYDILT	D3-9.2.1	(SEQ ID NO:540)	72
ITIFyL	D3-9.3.1	(SEQ ID NO:541)	73
LRYFDW	D3-9.1.2	(SEQ ID NO:548)	74
YDILTG	D3-9.2.2	(SEQ ID NO:549)	75
TIFyLV	D3-9.3.2	(SEQ ID NO:550)	76
RYFDWL	D3-9.1.3	(SEQ ID NO:557)	77
DILTGy	D3-9.2.3	(SEQ ID NO:558)	78
IFyLVI	D3-9.3.3	(SEQ ID NO:559)	79
YFDWLL	D3-9.1.4	(SEQ ID NO:566)	80
ILTGYY	D3-9.2.4	(SEQ ID NO:567)	81
FyLVII	D3-9.3.4	(SEQ ID NO:568)	82
FDWLLY	D3-9.1.5	(SEQ ID NO:575)	83
LTGYYN	D3-9.2.5	(SEQ ID NO:576)	84
VLLWFG	D3-10.1.1	(SEQ ID NO:590)	85
YYYGSG	D3-10.2.1	(SEQ ID NO:591)	86
ITMVRG	D3-10.3.1	(SEQ ID NO:592)	87
LLWFGE	D3-10.1.2	(SEQ ID NO:599)	88
YYGSGS	D3-10.2.2	(SEQ ID NO:600)	89
TMVRGV	D3-10.3.2	(SEQ ID NO:601)	90

LWFGEL	D3-10.1.3	(SEQ ID NO:608)	91
YGSGSY	D3-10.2.3	(SEQ ID NO:609)	92
MVRGVI	D3-10.3.3	(SEQ ID NO:610)	93
WFGELL	D3-10.1.4	(SEQ ID NO:617)	94
GSYSYY	D3-10.2.4	(SEQ ID NO:618)	95
VRGVII	D3-10.3.4	(SEQ ID NO:619)	96
FGELLy	D3-10.1.5	(SEQ ID NO:624)	97
SGSYYN	D3-10.2.5	(SEQ ID NO:625)	98
VLwLRL	D3-16.1.1	(SEQ ID NO:638)	99
YYDYVW	D3-16.2.1	(SEQ ID NO:639)	100
IMITFG	D3-16.3.1	(SEQ ID NO:640)	101
LwLRLG	D3-16.1.2	(SEQ ID NO:647)	102
YDYVWG	D3-16.2.2	(SEQ ID NO:648)	103
MITFGG	D3-16.3.2	(SEQ ID NO:649)	104
wLRLGE	D3-16.1.3	(SEQ ID NO:656)	105
DYVWGS	D3-16.2.3	(SEQ ID NO:657)	106
ITFGGV	D3-16.3.3	(SEQ ID NO:658)	107
LRLGEL	D3-16.1.4	(SEQ ID NO:665)	108
YVWGSY	D3-16.2.4	(SEQ ID NO:666)	109
TFGGVI	D3-16.3.4	(SEQ ID NO:667)	110
RLGELS	D3-16.1.5	(SEQ ID NO:674)	111
VWGSYR	D3-16.2.5	(SEQ ID NO:675)	112
FGGVIV	D3-16.3.5	(SEQ ID NO:676)	113
LGELSL	D3-16.1.6	(SEQ ID NO:683)	114
WGSYRY	D3-16.2.6	(SEQ ID NO:684)	115
GGVIVI	D3-16.3.6	(SEQ ID NO:685)	116
GELSLY	D3-16.1.7	(SEQ ID NO:692)	117
GSYRYT	D3-16.2.7	(SEQ ID NO:693)	118
VLLWyy	D3-22.1.1	(SEQ ID NO:707)	119

YYYDSS	D3-22.2.1	(SEQ ID NO:708)	120
ITMIVV	D3-22.3.1	(SEQ ID NO:709)	121
LLwyyW	D3-22.1.2	(SEQ ID NO:716)	122
YYDSSG	D3-22.2.2	(SEQ ID NO:717)	123
TMIVVV	D3-22.3.2	(SEQ ID NO:718)	124
LwyyWL	D3-22.1.3	(SEQ ID NO:725)	125
YDSSGY	D3-22.2.3	(SEQ ID NO:726)	126
MIVVVI	D3-22.3.3	(SEQ ID NO:727)	127
wyyWLL	D3-22.1.4	(SEQ ID NO:733)	128
DSSGYY	D3-22.2.4	(SEQ ID NO:734)	129
IVVVIT	D3-22.3.4	(SEQ ID NO:735)	130
yyWLLL	D3-22.1.5	(SEQ ID NO:742)	131
SSGYYY	D3-22.2.5	(SEQ ID NO:743)	132
wLRWyL	D4-23.1.1	(SEQ ID NO:769)	133
DYGGNS	D4-23.2.1	(SEQ ID NO:770)	134
VDTAMV	D5-5.1.1	(SEQ ID NO:784)	135
WIQLWL	D5-5.2.1	(SEQ ID NO:785)	136
GYSYGY	D5-5.3.1	(SEQ ID NO:786)	137
VDIVAT	D5-12.1.1	(SEQ ID NO:802)	138
WIyWLR	D5-12.2.1	(SEQ ID NO:803)	139
GYSGYD	D5-12.3.1	(SEQ ID NO:804)	140
DIVATI	D5-12.1.2	(SEQ ID NO:811)	141
IyWLRL	D5-12.2.2	(SEQ ID NO:812)	142
YSGYDY	D5-12.3.2	(SEQ ID NO:813)	143
VEMATl	D5-24.1.1	(SEQ ID NO:829)	144
YRWLQL	D5-24.2.1	(SEQ ID NO:830)	145
RDGYNY	D5-24.3.1	(SEQ ID NO:831)	146
EYSSSS	D6-6.1.1	(SEQ ID NO:847)	147
GYSSSW	D6-13.1.1	(SEQ ID NO:859)	148

GIAAAG	D6-13.2.1	(SEQ ID NO:860)	149
VyQQLV	D6-13.3.1	(SEQ ID NO:861)	150
YSSSWY	D6-13.1.2	(SEQ ID NO:867)	151
GYSSGW	D6-19.1.1	(SEQ ID NO:878)	152
GIAVAG	D6-19.2.1	(SEQ ID NO:879)	153
VyQWLV	D6-19.3.1	(SEQ ID NO:880)	154
YSSGWY	D6-19.1.2	(SEQ ID NO:887)	155

Exemplo 3: CDR3 de comprimento 6-20.

A inserção de segmentos D em CDR3s de HC pode levar a maior estabilidade e menor imunogenicidade. As bibliotecas são desenhadas ao nível dos aminoácidos por união de uma VH com um enchimento opcional de algum comprimento que é unido com um segmento D, um segundo enchimento opcional e uma JH. Para bibliotecas de comprimento seis ou oito, uma JH de comprimento total pode seguir-se a VH e um curto enchimento. A Tabela 77 mostra a frequência de segmentos D numa amostragem de 1419 Abs selecionados de FAB-310 ou FAB-410 para ligação a um alvo ou outro. Na amostra, 1099 Abs não tinham nenhum segmento D detetável (*i.e.*, correspondência de menos do que 70%). Onde são usados segmentos S, os segmentos D D1-1.3, D1-26.3, D2-2.2, D2-8.2, D2-15.2, D2-21.2, D3-16.2, **D3-22.2**, **D3-3.2**, D3-9.1, D3-9.2, D3-10.2, D3-16.2, D4-4.2, D4-4.3, D4-11.2, D4-4.2, D4-17.2, D4-23.2, **D5-5.3**, **D5-12.3**, D5-18.3, D6-6.1, D6-6.2, **D6-13.1**, D6-13.2, **D6-19.1**, D6-19.2, e D7-27.1 são preferenciais.

Logo que a sequência de aminoácidos parente tenha sido desenhada pode ser diversificada de vários modos: PCR propensa a erros, *wobbling*, e *dobbling*. A Tabela 14 mostra um número de hexâmeros que podem ser derivados de regiões D

de humano. Numa forma de realização, os hexâmeros que contêm resíduos de cisteína são excluídos. Numa forma de realização, os fragmentos de regiões D que contêm paragens são excluídos. Numa forma de realização, qualquer codão TAG encontrado na região D está substituído por um codão escolhido do conjunto compreendendo TCG, TTG, TGG, CAG, AAG, TAT, e GAG. Numa forma de realização, qualquer codão TAA encontrado na região D está substituído por um codão escolhido do conjunto compreendendo TCA, TTA, CAA, AAA, TAT, e GAA. Numa forma de realização, qualquer TGA da região D está substituído por um codão escolhido do conjunto compreendendo TGG, TCA, TTA, AGA, e GGA.

A Tabela 21 mostra sequências de aminoácidos parente exemplares para CDR3s de 6 a 20 aminoácidos. Estas sequências parente podem ser combinadas com diversidade em CDR1 e CDR2 de HC para formar uma biblioteca. É provável que a utilidade melhore se as regiões CDR3 forem diversificadas por, por exemplo, *wobbling*, *dobbling*, ou PCR propensa a erros das CDR3. Na Tabela 21, a sequência 6a compreende a extremidade de VH de 3-23 fundida com JH1 inteira. A sequência 6b contém a extremidade de 3-23 unida a um Y unido a D4-17 (RF 2) unido à região FR4 de JH1. A sequência 6c contém a extremidade de 3-23 seguida por D5-5 (RF 3) seguido pela parte de FR4 de JH1. A sequência 6d contém a extremidade de 3-23 unida a SY unido à JH4 inteira. A Tabela 21 mostra o nível de *doping* que seria apropriado para o *wobbling* da CDR3; outros níveis poderiam também ser usados. Outras regiões D ou fragmentos de regiões D poderiam ser usados. Outras sequências de JH poderiam ser usadas.

Tabela 21: Sequências de aminoácidos parente para CDR3s de
HC de 6-20 AAs.

Comprimento	Sequência parente	nível de <i>doping</i>	Comentário	SEQ ID NO:
6a	yycakAEYFQHwgqgtltvss	70:10:10:10	JH1 (inteira)	226
6b	yycakYDYGDYwgqgtltvss	70:10:10:10	Y::D4-17(2)::FR4 de JH1	227
6c	yycakGYSYGYwgqgtltvss	70:10:10:10	D5-5(3)::FR4 de JH1	228
6d	yycakSYFFDYwgqgtltvss	70:10:10:10	SY::JH4 (inteira)	229
8a	yycakYYAEYFQHwgqgtltvss	73:9:9:9	YY:JH1 (inteira)	230
8b	yycakYGYSSSWYwgqgtltvss	73:9:9:9	Y::D6-13(1)::FR4 de JH1	231
8c	yycakYGDYYFDYwgqgtltvss	73:9:9:9	D4-17(2) [2-5]::JH4 (inteira)	232
10a	yycakYYYDSSGYYYwgqgtltvss	73:9:9:9	D3-22(2)::Fr4 de JH1	233
10b	yycakGYcSSTScYTwgqgtltvss	73:9:9:9	D2-2(2)::Fr4 de JH1	234
10c	yycakYYSSAEYFQHwgqgtltvss	73:9:9:9	YYSS::JH1 (inteira)	235
10d	yycakGYSYGYFFDYwgqgtltvss	73:9:9:9	DS-5(3)::JH4 (inteira)	236
12a	yycakYYYDSSGYYYQHwgqgtltvss	85:5:5:5	D3-22(2)::QH::Fr4 de JH1	237
12b	yycakGYcSSTScYQHWgqgtltvss	85:5:5:5	D2-2(2)::QH::Fr4 de JH1	238
12c	yycakYDGSYSAEYFQHwgqgtltvss	85:5:5:5	YDGSYS::JH1 (inteira)	239
12d	yycakYYDYVWGSYRYTwgqgtltvss	85:5:5:5	D3-16(2)::Fr de JH1	240
12e	yycakGYSYGYWYFDLwgrgtltvss	85:5:5:5	D5-5(3)::JH2 (inteira)	241
14a	yycakYYYDSSGYYYFQHwgqgtltvss	73:9:9:9	D3-22(2)::YFQH::Fr de JH1	242
14b	yycakGYcSSTScYTYFQHwgqgtltvss	73:9:9:9	D2-2(2)::YFQH::Fr de JH1	243
14c	yycakSYGYcSSTScYQHWgqgtltvss	73:9:9:9	SY::D2-2(2)::QH::Fr de JH1	244
14d	yycakSYRYSGYSAEYFQHwgqgtltvss	73:9:9:9	SYRYSGYS::JH1 (inteira)	245
14e	yycakAYcGGDcYSNWFDPwgqgtltvss	73:9:9:9	D2-21(2)::JH5 (inteira)	246
15a	vycakSDGYYYDSSGYYYDYwgqgtltvss	73:9:9:9	SD::D3-22.2::JH4 (101ff)	930

Comprimento	Sequência parente	nível de <i>doping</i>	Comentário	SEQ ID NO:
15b	yycakGSGYCSGGSCYSFDYwgqgtltvtvss	73:9:9:9	GS::D2-15.2::JH4 (100ff)	931
15c	yycakGGRGYSSGWYRAFDIwgqgtmvtvss	73:9:9:9	GGR::D6- 19.1::R::JH3 (toda)	932
16a	yycakYYYDSSGYYYAEYFQHwgqgtltvtvss	73:9:9:9	D3-22(2)::JH1 (inteira)	247
16b	yycakGYcSSTScYTAEYFQHwgqgtltvtvss	73:9:9:9	D2-2(2)::JH1 (inteira)	248
16c	yycakSYDSYRSYGS AEYFQHwgqgtltvtvss	73:9:9:9	SYDSYRSYGS::JH1 (inteira)	249
16d	yycakSYSYGYcSSTScYTQHwgqgtltvtvss	73:9:9:9	SYSY::D2-2(2)::QH::Fr JH1	250
17a	yycakSRPGYSSSWYYYYGMDVwgqgtltvss	73:9:9:9	SRP::6-13.1::JH6 (-1Y)	933
18a	yycakGYCSGGSCYSYYYYGMDVwgqgtltvss	73:9:9:9	2-15.2::1H6 (-1Y)	221
18b	yycakDGYCSGGSCYSYYYYGMDVwgqgtltvss	73:9:9:9	D::2-15.2::JH6 (-2Ys)	222
19a	yycakDGYYDSSGYYYRGYFDYwgqgtltvss	73:9:9:9	D::D3- 22.2::RGY::JH4 (toda)	223
20a	yycakYSSYYYYDSSGYYYAEYFQHwgqgtltvss	73:9:9:9	YSSY::D3-22(2)::JH (inteira)	1 251
20b	yycakSYYSGYcSSTScYTAEYFQHwgqgtltvss	73:9:9:9	SYYS::D2-2(2)::JH 1 (inteira)	252
20c	yycakSGYcSSTScYTYSAEYFQHwgqgtltvss	73:9:9:9	S::D2- 2(2)::YYS::JH1 (inteira)	253
20d	yycakYYYYDYVWGSYRYTSNWFDpwgqgtltvss	73:9:9:9	Y::D3- 16(2)::S::JH5 (inteira)	254
20e		73:9:9:9	Y::D3-	255

Comprimento	Sequência parente	nível de <i>doping</i>	Comentário	SEQ ID NO:
	yycakYYYYDYVWGSYRYTSSYFDYwgqgtl vtvss		16(2)::SS::JH4 (inteira)	

Tabela 22: Cassete de exibição de HC

A sequência de aminoácidos mostrada na Tabela 22 é SEQ ID NO:892.

A sequência de ADN mostrada na Tabela 22 é SEQ ID NO:893.

```

1      Sinal para VH-Chi-IIIstump
2      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
3      M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L
4      946  atg aaa cac cta itg cct acg gcc gcc gct gga ttg tta tta ctc
5
6      16 17 18 19 20 21 22
7      A  A  Q  P  A  M  A
8      991 gcc GCC cag ccg GCC atg gcc
9      Sfil.....
10     NgmI... (1/2)
11     NcoI....
12
13 VH
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
```

A sequência de aminoácidos mostrada na Tabela 22 é SEQ ID NO:892.

A sequência de ADN mostrada na Tabela 22 é SEQ ID NO:893.

```

      N   S   L   R   A   E   D   T   A   V   Y   Y   C   A   K
1261  taac|agC|TTA|AGg|gct|qag|gac|act|GCA|Gtc|tac|tat|tgc|ggt|aaa|
      |ATIII| :                               | PstI |(2/2)
      .....CDR3.....|-----FR4-----
      93  96  97  98  98a 98b 98c  99  100 101 102 103 104 105 106
      G   Y   E   G   T   G   Y   A   F   D   I   W   G   Q   G
1306  |gac|tat|gaag|ggt|act|agt|cat|gct|ttc|gaC|ATA|TCg|ggg|caa|ggt|
      | NdeI |
      -----FR4----->|
      107 108 109 110 111 112 113
      T   M   V   T   V   S   S
1351  |act|atG|GTC|ACC|gct|tet|agt
      | NotSI| : c tcg ag = XhoI.
CH1
      A   S   I   R   G   P   S   V   F   P   L   A   P   S   S
1372  ggc tcc acc aag ggc cca tcg gtc ttc ccG CTA GCa ccc tcc tcc
      NheI...
      151 152 153 154 155 156 157 158 159 160 161 162 163 164 165
      R   S   T   S   G   G   T   A   A   L   G   C   L   V   K
1417  aag agc acc tat ggg ggc aca gcg gcc ctg ggc tgc ctg gtc aag
      166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
      D   Y   F   P   E   P   V   T   V   S   M   H   S   G   A
1462  gac tac ttc ccc gaa ccg gtg acg gtg tcc tgg sac tca ggc gcc
      181 182 183 184 185 186 187 188 189 190 191 192 193 194 195
      L   T   S   G   V   H   T   F   P   A   V   E   Q   S   S
1507  ctg acc agc ggc gtc caa acc ttc ccg gct gtc cta caa tcc tca
      196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
      G   L   Y   S   L   S   S   V   V   T   V   P   S   S   S
1552  gga ctc tac tcc ctg acc agc agc gta gtg acc ctg ccc tCC Agc agc
      BatKI.....
      211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
      L   G   T   Q   T   Y   I   C   H   V   H   H   K   P   S
1597  tTQ Ggc acc caa acc tac atc tgc aac gtg aat cac aag ccc agc
      NotXI.....
      226 227 228 229 230 231 232 233 234 235 236 237 238
      R   T   K   V   D   R   K   V   E   P   R   S   C
1642  aac acc aag gtg gac aag AAA GTT CAG CCC AAA TCT TGT
      139 140 141 marcador His ..... marcador cMyo .....
      A   A   A   H   H   H   H   H   H   G   A   A   E   Q   K   L   I
1681  GCG GCC GCa cat cat cat cac cat cac ggg gcc gca gaS caS aaS ctS atS
      NotI.....
      EagI....
      S   E   S   D   L   N   G   A   A   E   A   S   S   A   S   N   A   S
1732  tca gaa gag gat ctg aat ggg GCC gca gaG GCT agt tct gct agt AAC CCS Tc

```

A sequência de aminoácidos mostrada na Tabela 22 é SEQ ID NO:892.

A sequência de ADN mostrada na Tabela 22 é SEQ ID NO:893.

```

!                               BglI.....(3/4)                               MluI....
!
! Domínio 3 (III stump) -----
!   S   G   D   F   D   Y   E   K   M   A   N   A   N   K   G   A
1786 tcc ggt gat ttc gat tat gaa aag atg gca aac gct aat aag ggg gct
!
!   M   T   E   N   A   D   E   N   A   L   Q   S   D   A   K   G
1834 atg acc gaa aat gcc gat gaa aac gag cta cag tcc gac gct aaa ggc
!
!   K   L   D   S   V   A   T   D   Y   G   A   A   I   D   G   F
1882 aaa cct gat tct gtc gct act gat tcc ggt gct gct atc gat ggt ttc
!
!   I   G   D   V   S   G   L   A   N   G   N   G   A   T   G   D
1930 att ggt gac gtt tcc ggc ctt gct aat ggt aat ggt gct act ggt gat
!
!   F   A   G   S   N   S   Q   M   A   Q   V   G   G   G   D   N
1978 ttt gct ggc tct aat tcc caa atg gct caa gtc ggt gac ggt gat aat
!
!   S   P   L   M   N   N   F   R   Q   Y   L   P   S   L   P   Q
2026 tca cct tta atg aat aat ttc cgt caa tct tta cct tcc ctc cct aaa
!
!   S   V   E   C   S   F   F   V   F   G   A   G   K   F   Y   E
2074 teg gtt gaa tgt cgc cct ttt gtc ttt ggc gct ggt aaa cca tat gaa
!
!   F   S   I   D   C   D   K   I   N   L   F   R
2122 ttt tct att gat tgt gac aaa ata aac tta ttc cgt
!
!                               Final do Domínio 3
!
!   G   V   F   A   F   L   L   Y   V   A   T   F   M   Y   V   P140
2158 ggt gtc ttt ggc ttc ctt tta tat gtt gcc acc ttt atg tat gta ttt
!
!   início do segmento transmembranar
!
!   S   T   F   A   N   I   L
2206 tca acc ttt gct aac ata cag
!
!   N   N   K   E   S   (SEQ ID NO:892)
2227 cct aat aag gag tct TAA   tga aac ccc tga tga GAATTC (SEQ ID NO:893)
!   âncora intracelular,      MluI...,      EcoRI.

```

Tabela 25: A sequência de ADN de DY3F85LC contendo uma cadeia leve kappa 012 da linha germinativa de amostra. As sequências de anticorpos mostradas são da forma do anticorpo real, mas não foram identificadas como ligação a um antígeno particular.

Em cada linha, tudo após um ponto de exclamação (!) é comentário.

O ADN de DY3F85LC é SEQ ID NO:27

Em cada linha, tudo após um ponto de exclamação (!) é comentário.

O ADN de DY3F85LC é SEQ ID NO:27

```

!-----
1      AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT
61     ATAGCTAAAC AGGTTATTGA CCAITTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT
121    CGTTTCGCAGA ATTGGGAATC AACTGTTATA TGGAAATGAAA CTTCCAGACA CCGTACTTTA
181    GTTGCAATAT TAAAACATGT TGAGCTACAG CATTATATTC AGCAATTAAG CTCTAAGCCA
241    TCCGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG
301    TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAACGCG ATATTTGAAG
361    TCTTTCGGGC TTCCTCTTAA TCTTTTGTAT GCAATCCGCT TTGCTTCTGA CTATAATAGT
421    CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAAC TTTTAAAGCA

481    TTIGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT
541    AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT
601    GGTTTTTATC GTCGTCGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT
661    AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCCTAA ATCTCAACTG
721    ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT
781    TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA
841    CAATGATTTA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT
901    CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG
961    AATATCCGGT TCTTGTCAGG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC
1021   TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTCCCTT ATGATTGACC
1081   GTCTGCGCCT CGTTCGGCT AAGTAACAIG GAGCAGGTCG CGGATTTTGA CACAATTTAT
1141   CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT
1201   CAAAGATGAG TGTTTTAGTG TATCTTTTG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA
1261   GTGGCATTAC GTATTTTACC CGTTTAATGG AAACCTCCTC ATGAAAAAGT CTTTAGTCTC
1321   CAAAGCCTCT GTAGCGGTG CTACCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA
1381   CGATCCCGCA AAAGCGGCTT TTAACCTCCT GCAAGCCTCA GCGACCGAAT ATATCGGTGA
1441   TGCGTGGGGG ATGGTTGTGG TCATTGTGCG CGCAACTATC GGTATCAAGG TGTTTAAGAA
1501   ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAG GGCTCCTTTT GGAGCCTTTT
1561   TTTTGGAGAT TTCAACGTG AAAAAATTAT TATTGCAAT TCCTTTAGTT GTTCCTTTCT
1621   ATTCCTACCT CGCTGAAACT GTTGAAAGTT GTTAGCAAA ATCCCATACA GAAAATTCAT
1681   TTAATAACGT CIGGAAAGAC GACAAAACIT TAGATCGTIA CGCTAACTAT GAGGGCTGTC
1741   TGTGGAATGC TACAGGCGTT GTAGTTTGTA CTGGTGACGA AACTCAGTGT TACGGTACAT
1801   GGGTTCCTAT TGGGCTTGGT ATCCCTGAAA ATGAGGGTGG TGGCTCTGAG GGTGGCGGTT
1861   CTGAGGGTGG CGGTTCTGAG GGTGGCGGTA CTAAACCTCC TGAGTACGGT GATACACCTA
1921   TTCCGGGCTA TACTTATAIC AACCTCTCTG ACGGCACCTA TCCGCTTGGT ACTGAGCAAA
1981   ACCCGCTTAA TCCTAATCCT TCTCTTGAGG AGTCTCAGCC TCTTAATACT TTCATGTTTC
2041   AGAATAATAG GTTCCGAAAT AGGCAGGGGG CATTAACTGT TTATACGGGC ACTGTTACTC
2101   AAGGCACTGA CCCCCTTAAA ACTTATTACC AGTACACTCC TGTATCATCA AAAGCCATGT
2161   ATGACGCTTA CTGGAACGGT AAATTCAGAG ACTGCGCTTT CCATTCGCGC TTTAATGAGG
2221   ATTTATTGTT TTGTGAATAT CAAGGCCAAT CGTCTGACCT GCTCAACCT CCGTCAATG
2281   CTGGCGGGGG CTCTGGTGGT GGTCTGGTGG GCGGCTCTGA GGGTGGTGGC TCTGAGGGTG
2341   GCGGTTCTGA GGGTGGCGGG TCTGAGGGAG GCGGTTCCGG TGGTGGCTCT GGTTCGGGTG
2401   ATTTTGATTA TGAAGAAGAT GCAAAACGCTA ATAAGGGGGG TATGACCGAA AATGCGGATG
2461   AAAACGCGCT ACAGTCTGAC GCTAAAGGCA AACTTGATTG TGTCGCTACT GATTACGGTG
2521   CTGCTATCGA TGGTTTCATT GGTGACGTTT CCGGCCCTGC TAATGGTAAT GGTGCTACTG
2581   GTGATTTTGC TGGCTCTAAT TCCCAAATGG CTCAAGTCCG TGACGGTGAT AATTCACCTT
2641   TAATGAATAA TTTCCGICAA TATTTACCTT CCCTCCCTCA ATCGGTTGAA TGTGCGCCCTT
2701   TTGTCATTGG CGCTGGTAAA CCATATGAAT TTTCTATIGA TTGTGACAAA ATAAACTTAT
2761   TCCGTGGTGT CTTTGCCTTT CTTTATATG TTGCCACCTT TATGTATGTA TTTTCTACGT
2821   TTGCTAACAT ACTGCGTAAT AAGGAGTCTT AATCATGCCA GTTCTTTTGG GTATTCCGTT
2881   ATTATTGCGT TTCCTCGGTT TCCTTCTGGT AACTTTGTTC GGCTATCTGC TTACTTTTCT
2941   TAAAAAGGGC TTCGGTAAGA TAGCTATTGC TATTTCAITG TTTCTTGCTC TTATTATTGG
3001   GCTTAACTCA ATTCTTGTGG GTTATCTCTC TGATATTAGC GCTCAATTAC CCTCTGACTT
3061   TGTTCAAGGT GTTCAGTTAA TTCTCCCGTC TAATGCGCTT CCCTGTTTTT ATGTTATTCT
3121   CTCTGTAAAG GCTGCTATTT TCATTTTGA CGTTAAACAA AAAATCGTTT CTTATTTTGA
3181   TTGGGATAAA TAATAAGGCT GTTTATTTTG TAACCTGGCA ATTAGGCTCT GGAAAGACGC
3241   TCGTTAGCGT TGGTAAGATT CAGGATAAAA TTGTAGCTGG GTGCAAAATA GCAACTAATC
3301   TTGATTTAAG GCTTCAAAAC CTCCGCAAG TCGGGAGGTT CGCTAAAACG CCTCGCGTTC
3361   TTGAATAACC GGATAAGCCT TCTATATCTG ATTTGCTTGC TATTGGGCGC GGTAAATGATT
3421   CCTACGATGA AAATAAAAC GGCTTGCTTG TTCTCGATGA GTGCGGTACT TGGTTTAATA
3481   CCCGTTCTTG GAATGATAAG GAAAGACAGC CGATTATIGA TTGGTTTCTA CATGCTCGTA
3541   AATTAGATG GGATAITATT TTTCTTGTTC AGGACTTATC TATTGTGAT AAACAGGCGC
3601   GTTCTGCATT AGCTGAACAT GTTGTTTATT GTCGTCGTCT GGACAGAATT ACTTTACCTT
3661   TTGTCGGTAC TTTATATTCT CTTATTACTG GCTCGAAAAT GCCTCTGCCT AAATTACATG
3721   TTGGCGTTGT TAAATATGGC GATTCTCAAT TAAGCCCTAC TGTGAGCGT TGGCTTTATA
3781   CTGGTAAGAA TTTGATAAAC GCATATGATA CTAAACAGGC TTTTCTAGT AATTATGATT
3841   CCGGIGTTTA TTCTTATTTA ACGCCTTATT TATCACACGG TCGGTATTTT AAACCATTAA

```

Em cada linha, tudo após um ponto de exclamação (!) é comentário.

O ADN de DY3F85LC é SEQ ID NO:27

```

3901      ATTTAGGTCA GAAGATGAAA TTAATAAAAA TATATTGAA AAAGTTTTCT CGCGTTCTTT
3961      GTCTTGCGAT TGGATTGCA TCAGCATTTA CATATAGTTA TATAACCCAA CCTAAGCCGG
4021      AGGTTAAAAA GGTTAGTCTCT CAGACCTATG ATTTTGATAA ATTCCTATT GACTCTTCTC
4081      AGCGTCTTAA TCTAAGCTAT CGCTATGTT TCAAGGATTC TAAGGGAAAA TTAATTAATA
4141      GCGACGATTT ACAGAAGCAA GGTATTTCAC TCACATATAT TGATTTATGT ACTGTTTCCA
4201      TTAAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATTA ATTTTGTTTT CTGTATGTTT
4261      GTTTCATCAT CTCTTTTTC TCAGGTAATT GAAATGAATA ATTCGCCTCT CGCGGATTTT
4321      GTAACTTGGT ATTCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCCGA TGTAAAAGGT
4381      ACTGTTACTG TATATTTCAT TGACGTTAAA CTGAAAATC TACGCAATTT CTTTATTCT
4441      GTTTTACGTG CAAATAATTT TGATATGGTA GGTCTAACC CTTCCATAAT TCAGAAGTAT
4501      AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT
4561      GATAATCCG CTCTTCTGG TGGTTCTTT GTTCCGCAA ATGATAATGT TACTCAAAC
4621      TTTAAATTA ATAACGTTG GGCAGGAT TTAATACGAG TTGTCGAATT GTTTGTAAG
4681      TCTAATACTT CTAAATCCTC AAATGTAAT TCTATTGACG GCTCTAATCT ATTAGTTGTT
4741      AGTGCTCCTA AAGATATTTT AGATAACCTT CCTCAATTCC TTTCAACTGT TGATTTGCCA
4801      ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTT AGCAAGGTGA TGTCTTAGAT
4861      TTTTCATTG CTGCTGGCTC TCAGCGTGGC ACTGTTGCGG CGGGTGTAA TACTGACCGC
4921      CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTCCGTA TTTTAAATGG CGATGTTTTA
4981      GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT
5041      ATCTTACGC TTTCAGGTCA GAAGGTTCT ATCTCTGTTG GCCAGAAATGT CCTTTTATT
5101      ACTGGTCGTG TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTTAGACG GATTGAGCGT
5161      CAAAATGTAG GTATTTCCAT GAGCGTTTTT CCGTTGCAA TGGCTGGCGG TAATATTGTT
5221      CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTTCTTCTA CTCAGGCAAG TGATGTTATT
5281      ACTAATCAA GAAGTATTGC TACAACGGTT AATTGCGGTG ATGGACAGAC TCTTTTACTC
5341      GGTGGCCTCA CTGATTATAA AAACACTTCT CAGGATCTG GCGTACCGTT CCGTCTAAA
5401      ATCCCTTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CTAACGAGGA AAGCACGTTA
5461      TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GCGCGATTAA GCGCGCGGGG
5521      TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCGCTAGCGC CCGCTCTTTT
5581      CGCTTCTTTC CCTTCTTTC TCGCCACGTT CGCCGGCTTT CCGCGTCAAG CTCTAAATCG
5641      GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAAGTTGA
5701      TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC
5761      GTTGAGATCC ACGTTCTTTA ATAGTGGACT CTGTTTCAA ACTGGAACAA CACTCAACCC
5821      TATCTCGGGC TATCTTTTTC ATTTATAAGG GATTTTGCCG ATTTGGAAC CACCATCAAA
5881      CAGGATTTTC GCCTGCTGGG GCAAACGAGC GTGGACCGCT TCGTCAACT CTCTCAGGGC
5941      CAGGCGGTGA AGGGCAATCA GCTGTTGCC GTCTACTGG TGAAGAAGAA AACCAACCTG
6001      GATCCAAGCT TGCAGGTGGC ACTTTTCGGG GAAATGIGCG CGGAACCCCT ATTTGTTTAT
6061      TTTTCTAAAT ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCTGA TAAATGCTTC
6121      AATAATATTG AAAAAGGAAG AGTAAGGTA TTCAACATT CCGTGTGCGC CTTATTCCCT
6181      TTTTTCGGGC ATTTTGCTT CCTGTTTTTG CTCACCCAGA AACGCTGGTG AAAGTAAAG
6241      ATGTGAAGA TCAGTTGGGC GCACTAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA
6301      AGATCCTTGA GATTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAGTTC
6361      TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA
6421      TACACTATTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG
6481      ATGGCATGAC AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG
6541      CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA
6601      TGGGGGATCA TGTAACTCGC CTTGATCGTT GGAACCGGA GCTGAATGAA GCCATACCAA
6661      ACGACGAGCG TGACACCACG ATGCTGTAG CAATGGCAAC AACGTTGCGC AAATATTAA
6721      CTGGCGAAGT ACTTACTCTA GCTTCCCGGC AACAAITTAAT AGACTGGATG GAGGCGGATA
6781      AAGTTGCAGG ACCACTTCTG CGCTCGGCCC TTCCGGCTGG CTGTTTATT GCTGATAAAT
6841      CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAAG
6901      CCTCCCGTAT CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA
6961      GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAAGTGTCA GACCAAGTTT
7021      ACTCATATAT ACTTTAGATT GATTTAAAC TTCAATTTTA ATTTAAAGG ATCTAGGTGA
7081      AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCACTGTA
7141      CGTAAGACCC CCAAGCTTGT CGACTGAATG GCGAATGGCG CTTTGCCTGG TTCCGGCAC
7201      CAGAAGCGGT GCCGGAAGC TGGCTGGAGT GCGATCTTCC TGACGCTCGA GCGCAACGCA
!
XhoI...

7261      ATTAATGTGA GTTAGCTCAC TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT
7321      CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTCACA CAGGAACAG CTATGACCAT
7381      GATTACGCCA AGCTTTGGAG CCTTTTTTTT GGAGATTTT AAC

```

Tabela 30: Sequência de ADN de DY3FHC87 (SEQ ID NO:894)

1 aatgctacta ctattatag aattgatgcc aacttttcaa ctgcgcgcgc aaatgaaaat
 61 atagctaac aggttactga caatttgcga aatgtatcta atggtcaaaa taattccact
 121 cgttcgcaga attggaatc aactgttata tggaaatgaaa ctccagaca cgtacttta
 181 gctgcctatt caaaacatgt tgcgtacag cattatattc agcaattaaq ctctagcca
 241 tccgcacaaa tgcctctta tcaaaaggag caattaaagg tactctctaa tccctgacctg
 301 ttggagtttg ctkecgtctt ggttcgcttt gaagctcgaa ttaaaacgag atatttgaag
 361 tctttcgggc tccctcttaa cctttttagt gcaatcgcct tgcctctga ctataacagt
 421 cagggcraag acctgatttt tgaattatgg tcaattctgt tctcgaact gtttaagca
 481 tttgaggggg atcaaatgaa tatttatgac gattccgcag tactggagcg tactccagctc
 541 aaacatttta ctattacccc ctctggcaaa cttctttttg caaaagctc tgcctacttt
 601 ggttttttat gtcgtcttgg aaacggagggt tatgatagtg ttgcctttac tatgcctcgt
 661 aattcctttt ggcgttatgt atctgcctta gttgaatgtg gtattcctaa atctcaactg
 721 atgaactctt ctacccgtta caatgttgtt ccgttagttc gttttatcaa cgttagatttt
 781 tcttcccaac gtcctgactg gtaaatgag caagtcttta aaatcgcta aggttaattca
 841 caatgattaa agtgaatc aaacatctc aggcacatt tactactcgt tctggtgttt
 901 ctctgcaggg caagccttat tcactgaatg agcagctttg ttacgttgat ttgggtaatg
 961 aataccggt tctgtcaag attactcttg atgaaggtca gcaagcctat ggcctgtgtc
 1021 tgcacacagt tcatctgtcc tcttcaaaag ttggtcagtt aggttccct atgattgacc
 1081 gtcctgcctt cgttccgctt aagttaacatg gaggcagctc cggatttcta caaatcttat
 1141 cagggcagtg tacaatctc cgttctactt tgtttcgcgc ttggtataat cgttgggggt
 1201 caaagatggg tgttttagc tttctttttg cctctttcgt tctaggttgg tgccttcgtt
 1261 gtcgcattac gtaatttacc cgttttaagg aaacttcttc atgaaaaagt ctttaagctc
 1321 caaagcctct gtagccttg ctacccctct tccgatgctg tctttcgtg ctgaggggtg
 1381 agatccgcga aaagcggcct ttaacccctt gcaagcctc ggcagcgaat atctcgttta
 1441 tgcgtggggc atggtttgtt tcaattgttg ccgaactata ggtatcaagg tgttaagaa
 1501 attccctctg aaagcaagct gataaaacga tacaattaaa ggcctccttt ggagcctttt
 1561 tttttggaga ttctcaacgt gaasaaatta ttatttcgaa tctcctttagt tgttctcttc
 1621 tattctcact ccgttgaac tgttgaaggt tgttttagaa aatccctaac agaaattca
 1681 tttactaacg tctggaagaa cgaacaaact ctgatttgtt acgctaacta tggggcgtgt
 1741 ctgtggaatg ctacagagct tctagtttgt actgttgagc aaactcagtg tccagtcac
 1801 tgggttctta ttgggccttg tatccctgaa aatgaggggt gtcgctctga ggggtggcgt
 1861 tctgaggggt ggggttctga ggggtggcgt actaaacctc ctgagtcagg tcatcacct
 1921 attcgggctt atacttatat caacctctc gacggcactt atccgcctgg tactgagcaa
 1981 aaacccgcta atcctaactc ctctcttgag gagtctcagc ctcttaatac tttcatgttt
 2041 cagaaataa aggttcgaaa taggcagggt gcatlaactg ttatataggg cactgttact
 2101 caaggcactg accccgttaa aaactattac cagtacactc ctgtatcacc aaagccatg
 2161 tatgcgctt actggaacgg taattccaga gactgcgctt tccattcttg ctttaatatg
 2221 gatttatatt ttgtgaata tcaagggcaa tctctgacc tgcctcaacc tccgttcaat
 2281 gctgagggcg gctctgtgtg tggttcttgt ggcggctatg aggttggttg ctctgaggggt
 2341 ggcgttcttg aggttggttg ctctgagggg ggcgttccg gttgtggtc tggttccgggt
 2401 gattttgatt atgaaaagat ggcacacgct aatcaagggg ctatgacaga aaatgcgat
 2461 gaaaacgcgc taccgtctga cgttaaggtt aaacttgatt ctgtcgtac tcatcaggt
 2521 gctgttatcg atggttctat tggtagcgtt tccggccttg cttaatggta tgggtgtact
 2581 ggtgattttg ctggctctaa tcccaaatg gctcaagctg gtcagcgtga taattcaact
 2641 ttaatgaata actccgtca atatttacct tccctccttc aatcggctga atgtgcctc
 2701 tttgtctctg ggcctgttaa acratatgaa tttcttatg atgtgacaa aataaactta
 2761 ttcgtgtgtg tcttgcgtt tottttatat gttgcacact ttatgtatgt atttctacg
 2821 tctgtctaca taatgcgtta taaggagctt taatcatgac agttcttttg ggtattccgt
 2881 tattatttgc tttctcggc tccctctcgg taacttctt cggctatctg ctacttttc
 2941 taaaaaaggg attcggtaag atagctattg ctatttcaat gttctcttgt ctattattg

```

3001 ggcttaactc aattcttgtg gggtatctct ctgatattag cgctcaatta ccctctgact
3061 ttgttcaggg tggttcagtta attctcccgt ctaatgcgct tccctgtttt tatgttatcc
3121 tctctgtaaa ggctgctatt ttcatTTTTg acgttaaaac aaaaatcggt tcttatttgg
3181 attgggataa ataatatggc tgtttatttt gtaactggca aattagggtc tggaaaagacg
3241 ctcggttagc ttggtaagat tcaggataaa attgtagctg ggtgcaaaat agcaactaat
3301 cttgatttaa ggcttcaaaa cctcccgcga gtcgggaggt tcgctaaaaac gcctcgcgtt
3361 cttagaatac cggataagcc ttctatatct gatttgcttg ctattgggcg cggtaatgat
3421 tcctacgatg aaaataaaaa cggcttgctt gttctcgatg agtgcgggtac ttggtttaat
3481 acccgttctt ggaatgataa ggaaagacag ccgattattg attggtttct acatgctcgt
3541 aaattaggat gggatattat ttttcttgtt caggacttat ctattgttga taaacaggcg
3601 cgttctgcac tagctgaaca tggtgtttat tgcgtcgtc tggacagaat tactttacct
3661 tttgtcggtc ctttatattc tcttattact ggctcgaaaa tgcctctgcc taaattacat
3721 gttggcggtt ttaaatatgg cgattctcaa ttaagcccta ctgttgagcg ttggctttat
3781 actggtaaga atttgtataa cgcataatgat actaaacagg ctttttctag taattatgat
3841 tccggtgttt attcttattt aacgccttat ttatcacacg gtcggtattt caaaccata
3901 aatttaggtc agaagatgaa attaactaaa atatatTTga aaaagtTTtc tcgcgttctt
3961 tgctctgcga ttggatttgc atcagcattt acatatagtt atataaccca acctaaagccg
4021 gaggttaaaa aggtagtctc tcagacctat gattttgata aattcactat tgactcttct
4081 cagcgtctta atctaagcta tcgctatggt ttcaaggatt ctaagggaat attaattaat
4141 agcgacgatt tacagaagca aggttattca ctacatata ttgatttatg tactgtttcc
4201 attaaaaaag gtaattcaaa tgaattgttt aaatgtaatt aattttgttt tcttgatgtt
4261 tgtttcatca tcttcttttg ctcaggtaat tgaatgaat aattcgccct tgcgcgattt
4321 tgtaacttgg tattcaaaagc aatcaggcga atccgttatt gtttctcccg atgtaaaagg
4381 tactgttact gtatatcatc ctgacgttaa acctgaaaaat ctacgcaatt tctttatttc
4441 tgttttacgt gcaataaatt ttgatattgt aggttctaac ccttccataa ttcagaagta
4501 taatccaaac aatcaggatt atattgatga attgccatca tctgataatc aggaatatga
4561 tgataattcc gtccttctg gtgttttctt tgttccgcaa aatgataatg ttactcaaac
4621 ttttaaaatt aataacgttc gggcaaaagga ttaaatcaga gttgtcgaat tgtttgtaa
4681 gtctaatact tctaaatcct caaatgtatt atctattgac ggctctaate tattagtgtt
4741 tagtgctcct aaagatatTT tagataacct tctcaattc ctttcaactg ttgatttgcc
4801 aactgaccag atattgattg agggtttgat atttgaggtt cagcaagggt atgctttaga
4861 tttttcatTT gctgctggct ctgacgttgg cactgttgca ggcggtgtta atactgaccg
4921 cctcacctct gttttatctt ctgctgggtg ttctgtcggg atttttaatg gcgatgtttt
4981 agggctatca gttcgcgcac taaagactaa tagccattca aaaaatttgt ctgtgccacg
5041 tattcttacy ctttcaggtc agaagggttc tatctctggt ggccagaatg tcccttttat
5101 tactggtcgt gtgactggtg aatctgccaa tgtaaataat ccatttcaga cgattgagcg
5161 tcaaaaatga ggtatttcca tgagcgtttt tctgttgca atggctggcg gtaatatgtt
5221 tctgatatatt accagcaagg ccgatagttt gagttcttct actcaggcaa gtgatgttat
5281 tactaatcaa agaagtattg ctacaacggt taatttgctg gatggacaga ctcttttact
5341 cgggtggcctc actgattata aaaacacttc tcaggattct ggcgtaacct tctgtctaa
5401 aatcccttta atcggcctcc tgtttagctc ccgctctgat tctaacgagg aaagcacgtt
5461 atacgtgctc gtcaaagcaa ccatagtacy cgcctgttag cggcgcatTA agcggggcgg
5521 gtgtggtggt tacgcgcagc gtgaccgcta cacttgccag cgccctagcg cccgctcctt
5581 tcgctttctt ccttctcttt ctgcgcacgt tcgcccgttt tccccgtcaa gctctaaatc
5641 gggggctccc tttagggttc cgatttagtg cttlacggca cctcgacccc aaaaaacttg
5701 atttggttga tggttcacgt agtgggcat cgccctgata gacggttttt cgcccttga
5761 cgttggagtc cacgttcttt aatagtggac tctgtttcca aactggaaca acactcaacc
5821 ctatctcggg ctattctttt gatttataag ggattttgcc gatttcggaa ccaccatcaa
5881 acaggatttt cgcttctggt ggcaaaccag cgtggaccgc ttgttgcaac tctctcaggg
5941 ccaggcgggt aagggcaatc agctgttgcc cgtctcactg gtgaaaagaa aaaccaccct
6001 ggatccaagc ttgcagggtg cacttttcgg ggaatgtgc gcggaacccc tatttgttta
6061 tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt
6121 caataatatt gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc ccttattccc
6181 ttttttgcgg cattttgctt tccgtttttt gctcaccag aaacgctggt gaaagtaaaa
6241 gatgctgaag atcagttggg cgcactagtg ggttacatcg aactggatct caacagcggg
6301 aagatccttg agagttttgc ccccgaagaa cgttttccaa tgatgagcac ttttaaagtt
6361 ctgctatgtg gcgcggtatt atccccgatt gacgcccggc aagagcaact cggctcgcgcg

```

```

6421 atacactatt ctcagaatga cttggttgag tactcaccag tcacagaaaa gcatcttacg
6481 gatggcatga cagtaagaga attatgcagt gctgccataa ccatgagtga taacactgcg
6541 gccaaacttac ttctgacaac gatcgaggga ccgaaggagc taaccgcttt tttgcacaac
6601 atgggggatc atgtaactcg ccttgatcgt tgggaaccgg agctgaatga agccatacca
6661 aacgacgagc gtgacaccac gatgcctgta gcaatggcaa caacgttgcg caaactatta
6721 actggcgaac tacttactct agcttcccgg caacaattaa tagactggat ggaggcggat
6781 aaagtgtcag gaccacttct gcgctcggcc cttccggctg gctggtttat tgctgataaa
6841 tctggagcgg gtgagcgtgg gtctcgcggt atcattgcag cactggggcc agatggtaag
6901 cccctcccga tcgtagttag ctacacgacg gggagtcagg caactatgga tgaacgaaat
6961 agacagatcg ctgagatagg tgcctcactg attaagcatt ggtaactgtc agaccaagtt
7021 tactcatata tacttttagat tgatttaaaa cttcattttt aattttaaag gatctagggtg
7081 aagatccctt ttgataatct catgacccaa atcccttaac gtgagttttc gttccactgt
7141 acgtaagacc cccaagcttg tcgactgaat ggcgaatggc gctttgcctg gtttccggca
7201 ccagaagcgg tgcgggaaag ctggctggag tgcgatcttc ctgacgctcg agcgcaacgc
7261 aattaatgtg agttagctca ctcataggc accccaggct ttacacttta tgcttccggc
7321 tcgtatgttg tgtggaattg tgagcggata acaatttcac acaggaaaca gctatgacca
7381 tgattacgcc aagctttgga gcctttttt tggagatttt caacatgaaa taacctattgc
7441 ctacggcagc cgctggattg ttattactcg cGGCCcagcc GGCCatggcc gaagtccaat
7501 tgttagagtc tgggtggcgt cttgttcagc ctggtggttc tttacgtctt tcttgcgctg
7561 cttccggatt cactttctct tcgtacgcta tgtcttgggt tcgccaagct cctggtaaaag
7621 gtttggagtg ggtttctgct atctctggtt ctggtggcag tacttactat gctgactccg
7681 ttaaaaggct cttcactatc tctagagaca actctaagaa tactctctac ttgcagatga
7741 acagcttaag ggctgaggac actgcagtct actattgcgc taaagcctat cgtcctctct
7801 atcatgacat atgggggtcaa ggtactatgg tcaccgtctc tagtgccctc accaaggggc
7861 catcgggtct cccgctagca cctcctcca agagcacctc tgggggcaca gcggccctgg
7921 gctgcctggt caaggactac ttcccgaac cggtgacggt gtcgtggaac tcaggcggcc
7981 tgaccagcgg cgtccacacc ttcccggctg tccacagtc ctcaggactc tactccctca
8041 gcagcgtagt gaccgtgccc tccagcagct tgggcaccca gacctacatc tgcaacgtga
8101 atcacaagcc cagcaacacc aagggtggaca agaaagtga gcccaaactc tgtgcgcccg
8161 cacatcatca tcaccatcac ggggcgcgag aacaaaaact catctcagaa gaggatctga
8221 atggggccgc agaggttagc tctgctagtg gcgacttcga ctacgagaaa atggctaag
8281 ccaacaaagg cggcatgact gagaacgctg acgagaatgc tttgcaaagc gatgccaagg
8341 gtaagttaga cagcgtcgcg accgactatg gcgcgcgat cgacggcttt atcggcgatg
8401 tcagtggttt ggccaacggc aacggagcca ccggagactt cgcaggttcg aattctcaga
8461 tggcccaggt tggagatggg gacaacagtc cgcttatgaa caactttaga cagtaccttc
8521 cgtctcttcc gcagagtgtc gagtgcgctc cattedgttt cgggtccggc aagccttacg
8581 agttcagcat cgactgcgat aagatcaatc ttttccggcg cgttttcgct tcttgcgat
8641 acgtcgctac tttcatgtac gttttcagca ctttcgcaa tattttacgc aacaaagaaa
8701 gctagtgtac tcctaggaag cccgccta at gagcgggctt tttttttctg gtatgcatcc
8761 tgaggccgat actgtcgtcg tcccctcaaa ctggcagatg cacggttacg atgcgcccac
8821 ctacaccaac gtgacctatc ccattacggg caatccgcgg tttgttccca cggagaatcc
8881 gacgggttgt tactcgctca catttaatgt tgatgaaagc tggctacagg aaggccagac
8941 gcgaattatt tttgatggcg ttccctattg ttaaaaaatg agctgattta acaaaaaatt
9001 aatgcgaatt ttaacaaaat attaacgttt acaatttaaa tatttgctta tacaatcttc
9061 ctgttttttg ggcttttctg attatcaacc ggggtacata tgattgacat gctagtttta
9121 cgattaccgt tcctcgatcc tcttgtttgc tccagactct caggcaatga cctgatagcc
9181 tttgtagatc tctcaaaaat agctaccctc tccggcatla atttatcagc tagaacgggt
9241 gaatatcata ttgatggtga tttgactgtc tccggccttt ctcacccttt tgaatcttta
9301 cctacacatt actcaggcat tgcatttaaa atatatgagg gttctaaaaa tttttatcct
9361 tgcgttgaaa taaaggcttc tcccgaaaa gtattacagg gtcataatgt ttttggtaca
9421 accgatttag ctttatgtct tgaggcttta ttgcttaatt ttgctaattc tttgccttgc
9481 ctgtatgatt tattggatgt t

```

Tabela 35: Sequência de ADN de pMID21: 5957 bp (SEQ ID NO:895)

```

1 gacgaaaggg cctcgtgata cgcctatttt tatagtttaa tgtcatgata ctactgggtt

```

```

61  cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttatttt
121 tctaaataca ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat
181 aatattgaaa aaggaagagt atgagtattc aacatttccg tgcgcacctt attccctttt
241 ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg
301 ctgaagatca gttgggtgcc cgagtgggtt acatcgaact ggatctcaac agcggtaaga
361 tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagtctctc
421 tatgtggcgc ggtatttatcc cgtattgacg ccgggcaaga gcaactcggg cgccgcatac
481 actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg
541 gcatgacagt aagagaatta tgcagtgcgt ccataaccat gagtataaac actgcggcca
601 acttacttct gacaacgatac ggaggaccga aggagctaac cgcttttttg cacaacatgg
661 gggatcatgt aactcgcctt gatcgttggg aaccggagct gaataagacc ataccaaaacg
721 acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg
781 gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag
841 ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg
901 gagccggtga gcgtgggtct cgcggtatca ttgcagcaact ggggccagat ggtaagccct
961 cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac
1021 agatcgctga gatagggtgc tcaactgata agcatlggta actgtcagac caagttaact
1181 catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga
1141 tcctttttga taatctcatg accaaaatcc cttaacgtga gttttctgtt cactgagcgt
1201 cagaccccggt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct
1261 gctgcttgca aacaaaaaaa ccaccgctac cagcgggtggg ttgtttgccc gatcaagagc
1321 taccaactct tttccgaag gtaactggct tcagcagagc gcagatacca aatactgttc
1381 ttctagtgtg gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc
1441 tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg
1501 ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acggggggtt
1561 cgtgcataca gccacgcttg gagcgaacga cctacaccga actgagatac ctacagcgtg
1621 agctatgaga aagcgccacg cttcccgaag ggagaaaggg ggacaggtat ccggtaaagc
1681 gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatctctt
1741 atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag
1801 gggggcggag cctatggaaa aacgccagca acgcgccctt ttacggttc ctggcctttt
1861 gctggccttt tgcacacgt ttcttctctg cgttatcccc tgattctgtg gataaccgta
1921 ttaccgcctt tgagttagct gataccgctc gccgcagcgg aacgaccgag cgcagcaggt
1981 cagtgcagca ggaagcgga gagcgcccaa tacgcaaac gcctctcccc gcgcgttggc
2041 cgattcatta atgcagctgg cagcagaggt tccccgactg gaaagcgggg agtgagcgca
2101 acgcaattaa tgtgagttag ctcaactcatt aggcacccca ggctttacac ttatgcttc
2161 cggtcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg
2221 accatgatta cgccaagctt tggagccttt tttttggaga ttttcaacgt gaaaaaatta
2281 ttattcgcaa ttcccttagt tgttccttcc tattctcaca gtgcacaggt ccaactgcag
2341 gagctcgaga tcaaacgtgg aactgtggct gcaccatctg tcttcatctt cccgccatct
2401 gatgagcagt tgaatcttg aactgcctct gttgtgtgcc tgctgaataa cttctatccc
2461 agagaggcca aagtacagt gaaggtggat aacgccttcc aatcgggtaa ctcccaggag
2521 agtgtcacag agcaggacag caaggacagc acctacagcc tcagcagcac cctgacgctg
2581 agcaaaagca actacgagaa acacaaaagtc tacgcctgcg aagtcaccca tcagggcctg
2641 agttcacagg tgacaaaagag cttcaacagg ggagagtgtt aataaggcgc gcctaaccat
2701 ctatttcaag gaacagtctt aatgaaaaag cttttattca tgatcccgtt agttgtaccg
2761 ttctgtggcc agccggcctc tgctgaagtt caattgttag agtctgggtg cggctctgtt
2821 cagcctggtg gttctttacg tctttcttgc gctgcttccg gagcttcaga tctgtttgcc
2881 tttttgtggg gtggtgcaga tcgcgttacg gagatcgacc gactgcttga gcaaaagcca
2941 cgcttaactg ctgatcaggg atgggatgtt attcgccaaa ccagtcgtca ggatcttaac
3001 ctgaggtttt ttttacctac tctgcaagca gcgacatctg gtttgacaca gagcgatccg
3061 cgtcgtcagt tggtagaaac attaacacgt tgggatggca tcaatttgct taatgatgat
3121 ggtaaaacct ggcagcagcc aggtctctgc atcctgaacg tttggctgac cagtatgttg
3181 aagcgtaccg tagtggctgc cgtacctatg ccatttgata agtggtagac cgccagtggc
3241 tacgaaacaa ccaggagcgg cccaactggt tcgctgaata taagtgttgg agcaaaaatt
3301 ttgtatgagg cgggtgcagg agacaaatca ccaatccac aggcgggtga tctgtttgct
3361 gggaaaccac agcaggaggt tgtgttggct gcgctggaag atacctggga gactctttcc
3421 aaacgctatg gcaataatgt gagtaactgg aaaacaccgg caatggcctt aacgttccgg

```

```

3481 gcaaataatt tctttggtgt accgcaggcc gcagcgggag aaacgcgtca tcaggcggag
3541 tatcaaaacc gtggaacaga aaacgatatg attgttttct caccaacgac aagcgatcgt
3601 cctgtgcttg cctgggatgt ggtcgcaccc ggtcagagtg ggtttattgc tcccgatgga
3661 acagttgata agcactatga agatcagctg aaaatgtacg aaaatttttg ccgtaagtcg
3721 ctctggttaa cgaagcagga tgtggaggcg cataaggagt tctagagaca actctaagaa
3781 tactctctac ttgcagatga acagcttaag tctgagcatt cggtcggggc aacattctcc
3841 aaactgacca gacgacacaa acggtttacg ctaaatcccg cgcatgggat ggtaaaagg
3901 tggcgctctt gctggcctgg actcatcaga tgaaggccaa aaattggcag gagtggacac
3961 agcaggcagc gaaacaagca ctgaccatca actggtaacta tgctgatgta aacggcaata
4021 ttggttatgt tcatactggt gcttatccag atcgtcaatc aggcctatgat ccgcgattac
4081 ccgttcctgg tacgggaaaa tgggactgga aagggtattt gccttttgaa atgaacccta
4141 aggtgtataa cccccagcag ctagccatat tctctcggtc accgtctcaa gcgcctccac
4201 caagggccca tcggtcttcc cgctagcacc ctctccaag agcacctctg ggggcacagc
4261 ggccctgggc tgctgtgtca aggactactt ccccgaaaccg gtgacgggtg ctggaactc
4321 aggcgcctgt accagcggcg tccacacctt cccggtgtc ctacagtcta gcggactcta
4381 ctccctcagc agcgtagtga ccgtgccttc ttctagcttg ggcacccaga cctacatctg
4441 caacgtgaat cacaagccca gcaacaccaa ggtggacaag aaagttagac ccaaatcttg
4501 tgcggccgca catcatcatc accatcacgg ggcgcagaa caaaaactca tctcagaaga
4561 ggatctgaat ggggccgcag aggtagtctt tgctagttaac gcgtcttcgg gtgattttga
4621 ttatgaaaag atggcaaacg ctaataaggg ggctatgacc gaaaatgccg atgaaaacgc
4681 gctacagctc gacgctaagg gcaaaactga ttctgtcgtc actgattacg gtgctgctat
4741 cgatgttttc attggtgacg ttcccgccct tgctaattgtt aatggtgcta ctggtgattt
4801 tgctggctct aattcccaaa tggctcaagt cggtagcggg gataattcac ctttaatgaa
4861 taatttccgt caatatttac ctccctccc tcaatcggtt gaatgtcgcc cttttgtctt
4921 tggcgctggt aaaccatagc aattttctat tgattgtgac aaaataaact tattcgttg
4981 tgtctttgct tttcttttat atgttgccac ctttatgtat gtattttcta cgttgctaa
5041 catactgcgt aataaggagt cttaatgaaa cgcgtgatga gaattcactg gccgtcgctt
5101 tacaacgtcg tgactgggaa aaccctggcg ttaccaact taatcgctt gcagcacatc
5161 cccctttcgc cagctggcgt aatagcgaag aggccgcac cgatcgccct tcccaacagt
5221 tgccgcagct gaatggcgaa tggcgccctg tgccgtattt tctccttacg catctgtgctg
5281 gtatttcaca ccgcatacgt caaagcaacc atagtacgcg cctgtagcg gcgcattag
5341 cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccttagcgc
5401 cgctcccttc gctttcttcc ctcccttctc cgcacgcttc gccggtcttc ccgctcaagc
5461 tctaaatcgg gggctccctt taggggtccg atttagtctt ttacggcacc tcgaccccaa
5521 aaaacttgat ttgggtgatg gttcacgtag tgggccatcg cctgataga cggttttctg
5581 ccccttgacg ttggagtcga cgttctttaa tagtggaactc ttgttccaaa ctggaacaac
5641 actcaactct atctcgggct attcttttga ttataaaggg attttgccga tttcggtcta
5701 ttggttaaaa aatgagctga tttacaacaa atttaacgcg aattttaaca aaatattaac
5761 gtttacaatt ttatggtgca gtctcagtag aatctgctct gatgccgat agttaagcca
5821 gcccgcacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc tcccggcatc
5881 cgcttacaga caagctgtga ccgtctccgg gagctgcagtg tgtcagagggt ttccaccgtc
5941 atcaccgaaa cgcgcga

```

Tabela 36: pM21J contendo IIIss::A27::Ckappa

Número de bases 5225 (SEQ ID NO:921)

```

GACGAAAGGG CCCTGTGATA CGCCTATTTT TATAGGTAA TGTATGATA ATATGGTTT 60
CTTAGACGTC AGGTGGGACT TTTCGGGGAA AISTGSCGG AACCCCTATT TGTITATTTT 120
TCTAAATACA TTCAAAATAG TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT 180
ATATATGAAG AAGGAAGAGT ATGAGTATTC AACATITCCG TGTTCGCTTT ATTCCTTTT 240
TTGCGGCAAT TTGCTTTCCT GTTTTTCCTC ACCCAGAAAC GCTGCTGAAA GTAAAGATG 300
CTGAAGATCA GTTGGGTCGC CGAGTGGGTT ACATCGAAGT GGTCTCAAC AGCGGTAAGA 360
TCTTGAAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTCTTGC 420
TATCTGCGGC GGTATTTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC 480
ACTATTTCTC GAATGACTTG GTTGAGTACT CACCACTCAC AGAAAAGCAT CTTACGGATG 540
GCATGACAGT AAGAGAAATA TGCAGTGTCT CCATAACCAT GAGTGATAAC ACTGCGGCCA 600

```

ACTTACTTCT	GACAAACGATC	GGAGGACCGA	AGGAGCTAAC	CGCTTTTTTG	CACAACATGG	660
GGGATCAIGT	AACTCGCCTT	GATCGTTGGG	AACCGGAGCT	GAATGAAGCC	ATACCAAACG	720
ACGAGCGTGA	CACCACGATG	CCTGTAGCAA	TGGCAACAAC	GTTCGCGAAA	CTATTAACATG	780
GCGAACTACT	TACTCTAGCT	TCCCAGCAAC	AATTAATAGA	CTGGATGGAG	GCGGATAAAG	840
TTGCAGGACC	ACTTCTGCGC	TCCGCCCTTC	CGGCTGGCTG	GTTTATTGCT	GATAAATCTG	900
GAGCCGGTGA	GCGTGGGTCT	CGCGGTATCA	TTGCAGCACT	GGGGCCAGAT	GGTAAGCCCT	960
CCCGTATCGT	AGTTATCTAC	ACGACGGGGA	GTGAGGCAAC	TATGGATGAA	CGAAATAGAC	1020
AGATCGCTGA	GATAGGTGCC	TCACGTATTA	AGCATTGGTA	ACTGTCAGAC	CAAGTTTACT	1080
CATATATACT	TTAGATTGAT	TTAAAACTTC	ATTTTAAATT	TAAAAGGATC	TAGGTGAAGA	1140
TCCTTTTTGA	TAACTCTCATG	ACCAAATCC	CTTAACGTGA	GTTTTCGTTC	CACTGAGCGT	1200
CAGACCCCGT	AGAAAAGATC	AAAGGATCTT	CTTGAGATCC	TTTTTTTCTG	CGCGTAATCT	1260
GCTGCTTGCA	AACAAAAAAA	CCACCGCTAC	CAGCGGTGGT	TTGTTTGCCG	GATCAAGAGC	1320
TACCAACTCT	TTTTCCGAAG	GTAACITGGCT	TCAGCAGAGC	GCAGATACCA	AATACTGTTC	1380
TTCTAGTGTA	GCCGTAGTTA	GGCCACCACT	TCAAGAACTC	TGTAGCACCG	CCTACATACC	1440
TCGCTCTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG	CGATAAGTCG	TGTCTTACCG	1500
GGTTGGACTC	AAGACGATAG	TTACCGGATA	AGGCGCAGCG	GTCGGGCTGA	ACGGGGGGTT	1560
CGTGCAATACA	GCCGAGCTTG	GAGCGAACGA	CCTACACCGA	ACTGAGATAC	CTACAGCGTG	1620
AGCTATGAGA	AAGCGCCACG	CTTCCGAAG	GGAGAAAGGC	GGACAGGTAT	CCGGTAAGCG	1680
GCAGGGTCCG	AACAGGAGAG	CGCACGAGGG	AGCTTCCAGG	GGGAAACGCC	TGGTATCTTT	1740
ATAGTCTGTG	CGGGTTTTCG	CACCTCTGAC	TTGAGCGTCG	ATTTTGTGTA	TGCTCGTCAG	1800
GGGGGCGGAG	CCTATGGAAA	AACGCCAGCA	ACGCGGCCTT	TTTACGGTTC	CTGGCCTTTT	1860
GGTGGGCTTT	TGCTCACATG	TTCTTTCCCTG	CGTTATCCCC	TGATTCTGTG	GATAACCGTA	1920
TTACCGCCTT	TGAGTGAGCT	GATACCGCTC	GCCGCAGCCG	AACGACCGAG	CGCAGCGAGT	1980
CAGTGAGCGA	GGAAGCGGAA	GAGCGCCCAA	TACGCAAAAC	GCCTCTCCCC	GCAGCGTTGGC	2040
CGATTCAATTA	ATGCAGCTGG	CACGACAGGT	TTCCCGACTG	GAAAGCGGGC	AGTGAGCGCA	2100
ACGCAATTAA	TGTGAGTTAG	CTCACTCATT	AGGCACCCCA	GGCTTTACAC	TTTATGCTTC	2160
CGGCTCGTAT	GTTGTGTGGA	ATTGTGAGCG	GATAACAATT	TCACACAGGA	AACAGCTATG	2220
ACCATGATTA	CGCCAAGCTT	TGGAGCCTTT	TTTTTGAGAG	TTTTCAACAT	GAAGAACTG	2280
CTGTCTGCTA	TCCCAGTAGT	TGTCCCTTTC	TATTCTCATA	GTGAAATCGT	TCTGACCCAG	2340
TCCCCGGGGA	CCCTGTCTCT	GTCTCCGGGT	GAACGTGCTA	CGCTGAGCTG	TGCTGCTTCT	2400
CAATCCGTTA	GCTCCTCTTA	TTTAGCTTGG	TATCAGCAAA	AGCCGGGTCA	AGCTCCGCGG	2460
CTGTTGATCT	ATGGTGCCTC	TAGTCGTGCT	ACTGGCATCC	CTGATCGTTT	CTCTGGCTCT	2520
GGCTCCGGAA	CCGATTTTAC	TCTGACCAAT	TCTCGTCTCG	AGCCGGAAGA	TTTCGCTGTC	2580
TACTATTGTC	AACAGTATGG	TCTAGTCCG	CTGACTTTTCG	GTGGCGGTAC	CAAAGTCGAA	2640
ATCAAGCGTG	GAAGTGTGGC	TGCACCATCT	GTCTTCATCT	TCCCGCCATC	TGATGAGCAG	2700
TTGAAATCTG	GAAGTGCCTC	TGTTGTGTGC	CTGCTGAATA	ACTTCTATCC	CAGAGAGGCC	2760
AAAGTACAGT	GGAAGGTGGA	TAACGCCCTC	CAATCGGGTA	ACTCCAGGA	GAGTGTACAA	2820
GAGCAGGACA	GCAAGGACAG	CACCTACAGC	CTCAGCAGCA	CCCTGACTCT	GTCCAAAGCA	2880
GACTACGAGA	AACACAAAGT	CTACGCCTGC	GAAGTCAACC	ATCAGGGCCT	GAGTTCACCG	2940
GTGACAAAGA	GCTTCAACAG	GGGAGAGTGT	TAATAAGGCG	CGCCAAATTA	ACCATCTATT	3000
TCAAGGAACA	GCTTTAATGA	AGAAGCTCCT	CTTTGCTATC	CCGCTCGTCG	TTCTTTTGTG	3060
GGCCACGCCG	GCCATGGCCG	AAGTTCAATT	GTTAGAGTCT	GGTGGCGGTC	TTGTTACAGC	3120
TGGTGGTTCT	TTACGTCTTT	CTTGCCTGTC	TTCCGGATTG	ACTTTCTCTC	GTTACAAGAT	3180
GAAGTGGGTT	CGCCAAGCTC	CTGGTAAAGG	TTTGGAGTGG	GTTTCTGTGA	TCTATCCTTC	3240
TGGTGGCGGT	ACTGGTTATG	CTGACTCCGT	TAAAGGTGCG	TTCACTATCT	CTAGAGACAA	3300
CTCTAAGAAAT	ACTCTCTACT	TGCAGATGAA	CAGCTTAAGG	GCTGAGGACA	CTGCAGICTA	3360
CTATTGTGCG	AGAGTCAATT	ACTATGATAG	TAGTGGITAC	GGTCTTATAG	CTCCTGGACT	3420
TGACTACTGG	GGCCAGGGAA	CCCTGGTCAAC	CGTCTCAAGC	GCCTCCACCA	AGGGTCCGTC	3480
GGTCTTCCCG	CTAGCACCTT	CCTCCAAGAG	CACCTCTGGG	GGCACAGCGG	CCCTGGGCTG	3540
CCTGGTCAAG	GACTACTTCC	CCGAACCGGT	GACGGTGTCTG	TGGAACTCAG	GCGCCCTGAC	3600
CAGCGGCGTC	CACACCTTCC	CGGCTGTCTT	ACAGTCTAGC	GGACTCTACT	CCCTCAGCAG	3660
CGTAGTGACC	GTGCCCTCTT	CTAGCTTGGG	CACCCAGACC	TACATCTGCA	ACGTGAATCA	3720
CAAGCCGAGC	AACACCAAGG	TGGACAAGAA	AGTTGAGCCC	AAATCTTGTG	CGGCCGCACA	3780
TCATCATCAC	CATCACGGGG	CCGCAGAACA	AAAACCTCATC	TCAGAAGAGG	ATCTGAATGG	3840
GGCCGCAGAG	GCTAGTTCTG	CTAGTAACGC	GTCTTCCGGT	GATTTTGATT	ATGAAAAGAT	3900
GGCAACCGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	GAAAACGCGC	TACAGTCTGA	3960
CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	GCTGCTATCG	ATGGTTTCAT	4020

```

TGGTGACGTT TCCGGCCCTTG CTAATGGTAA TGGTGCTACT GGTGATTTTG CTGGCTCTAA 4080
TTCCCAAAATG GCTCAAAGTCG GTGACGGTGA TAATTACCTT TTAATGAATA ATTTCCGTC 4140
ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT TTTGTCTTTG GCGCTGGTAA 4200
ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACCTT TTCCGTGGTG TCTTTGCGTT 4260
TCTTTTATAT GTTGCCACCT TTAATGTATG ATTTTCTACG TTTGCTAACA TACTGCGTAA 4320
TAAGGAGTCT TAATGAAACG CGTGATGAGA ATTCAGTGGC CGTCGTTTTA CAACGTCGTG 4380
ACTGGGAAAA CCCTGGCGTT ACCCAACTTA ATCGCCTTGC AGCACATCCC CCTTTCGCCA 4440
GCTGGCGTAA TAGCGAAGAG GCCCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA 4500
ATGGCGAATG GCGCCTGATG CGGTATTTTC TCCTTACGCA TCTGTGCGGT ATTTACACAC 4560
GCATACGTCA AAGCAACCAT AGTACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT 4620
GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC TTAGCGCCCG CTCCTTTCCG 4680
TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCTTTCCC CGTCAAGCTC TAAATCGGGG 4740
GCTCCCTTTA GGGTTCGGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTT 4800
GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT 4860
GGAGTCCACG TTTCTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACTCTAT 4920
CTCGGGCTAT TCTTTTGATT TATAAGGGAT TITGCCGATT TCGGTCTATT GGTAAAAA 4980
TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTACAAA ATATTAACGT TTACAATTT 5040
ATGGTGCAGT CTCAGTACAA TCTGCTCTGA TGCCGCATAG TTAAGCCAGC CCGCACCC 5100
GCCAACACCC GCTGACGCGC CCTGACGGGC TTGTCTGCTC CCGGCATCCG CTTACAGACA 5160
AGCTGTGACC GTCTCCGGA GCTGCATGTG TCAGAGGTTT TCACCGTCAT CACCGAAACG 5220
CGCGA

```

Tabela 40: pLCSk23 (SEQ ID NO:896)

```

1 GACGAAAGGG CCGTCTCTGC CAGTGTTCAC ACCAATTAC CAATTCGTAT TAGAAAACT
61 CATCGAGCAT CAAATGAAAC TGCAATTTAT TCATATCAGG ATATCAATA CCATATTTT
121 GAAAAAGCCG TTTCTGTAAT GAACGAGAAA ACTCACCGAG GCASTTCCAT AGGATGGCAA
181 GATCTCTGTA TCGGTCTGCG ATTCCGACTC GTCCAACTC AATACAACCT ATTAATTTCC
241 CCTCTCTCAA AATAAGTTTA TCAGTCTAGA AATCACCATG AGTGACGACT GAATCCGGTG
301 AGAATGGCAA AAGCTTATGC ATTTCTTTCC AGACTTGTTC AACAGGCCAG CCATTACGCT
361 CGTCATCAAA ATCACTGGCA TCAACCAAC CGTTATTCAT TCGTGATTC GCGTGAAGCA
421 GACGAAATAC GCGATCGCTG TTAAGAGGAC AATTACAAAC AGGAATTEAA TGCAACCGCG
481 GCAGGACAC TGCCAGCGCA TCAACAATAT TTTCACTTGA ATCAGGATAT TCTTCTAATA
541 CCTGGATGCG TGTCTTCCCG GGGATCCGAG TCGTGAGTAA CCATGCATCA TCAGGAGTAC
601 GGATAAAAAT CTTGATGGTC GGAAGAGGCA TAAATTCCTT CAGCCAGTTT AGTCTGACCA
661 TCTCATCTGT AACATCATTG GCAADGGTAC CTTTGCCATG TTTCAAAAC AACCTGCGG
721 CATCGGCTT CCCATACAAT CQATAGATTG TCGCACCTGA TTGCCCCGCA TTATCGCGAG
781 CCCATTTATA CCCATATAAA TCAACATCCA TGTTCGAATT TAATCGCGCG CTCGAGCAAG
841 ACGTTTCCCG TTGAATATGG CTCATAACAC CCCTTGTATT ACTGTTTATG TAAGCAGACA
901 GTTTTATTGT TCATGATGAT ATATTTTAT CTTGTGCAAT GTAACATCAG AGATTTIGAG
961 ACACAACCGG GCCTTCCGCC CCGCCCGCTG CAGGCTCGG GCTATTCCTG TCAGACCAAG
1021 TTAACATATA TATACTTTAG ATTGATTTAA AACTTCATTY TTAATTTAAA AGGATCTAGG
1081 TGAAGATCCY TTTTGATAAT CTCATGACCA AAATCCCTTA ACSTGAGTTT TCGTTCCACT
1141 GAGCGTCAGA CCGCTAGAA AAGATCAAGS GATTTCTTTC AGATCCTTTT TTTCTGCGCG
1201 TAATCTGCTG CTTGCAAAAC AAAAAGCCAC CGCTACCAGC GGTGCTTTGT TTGCGGATC
1261 AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAATA
1321 CTGTTCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA GAATCTCTTA GCACCGCTTA
1381 CATACCTCCG TGTGCTAATC CTGTTACGAG TGGCTGCTGC CAGTGGCAT AASTCGTCT
1441 TACCGGCTT GGACTCAAGA CGATAGTTAC CCGAIAAGGC GCAGCGGTCG GCGTGAACCG
1501 GCGGTTCTGT CATACAGCCC AGCTTGGAGC GACGACCTTA CACCGAAGTG ACATAGCTAC
1561 AGCTGTAGCT ATGAGAAAGC GCCACCTTC CCGAAGGAG AGAGGCGGAC AGGTATCCCG
1621 TAAGCGGCGAG GTTCGSAACA GAGAGCGGCA CCGAGGAGCT TCCAGGCGGA AACGCTGGT
1681 ATCTTATAG TCTGTGCGG TTTGCCACC TGTGACTTGA GCGTCCATTT TTGTGATGCT
1741 CGTCAAGGGG GCGGAGCTTA TGGAAAAACG CCAGCAACGC GCGCTTTTTA CGCTTCTGTG
1801 GCTTTTCTGT GCTTTTGTCT CACATGTTCT TTTCTGCTT ATCCCTGAT TCTGTGATA
1861 ACCGTATTAC CGCTTTGAG TGAGCTGATA CCGCTCCTCG CAGCGGAACG ACCGAGCGCA

```

```

1921 GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC GCCCAATACG CAAACCGCCT CTCCTCCGCGC
1981 GTTGGCCGAT TCATTAATGC AGCTGGCAGC ACAGGTTTCC CGACTGGAAA GCGGGCAGTG
2041 AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT TTACACTTTA
2101 TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTTAC ACAGGAAACA
2161 GCTATGACCA TGATTACGCC AAGCTTTGGA GCCTTTTTTT TGGAGATTTT CAACATGAAG
2221 AAGCTCCTCT TTGCTATCCC GCTCGTCGTT CCTTTTGTGG CCCAGCCGGC CATGGCCGAC
2281 ATCCAGATGA CCCAGTCTCC ATCCTCCCTG TCTGCATCTG TAGGAGACAG AGTCACCATC
2341 ACTTGCCGGG CAAGTCAGAG CATTAGCAGC TATTTAAATT GGTATCAGCA GAAACCAGGG
2401 AAAGCCCCTA AGCTCCTGAT CTATGCTGCA TCCAGTTTGC AAAGTGGGGT CCCATCAAGG
2461 TTCAGTGCCA GTGGATCTGG GACAGATTTC ACTCTACCCA TCAGCAGTCT GCAACCTGAA
2521 GATTTTGCAA CTACTACTG TCAACAGAGT TACAGTACCC CTTTCACTTT CGGCCCTGGG
2581 ACCAAAGTGG ATATCAAACG TGGTACcGTG GCTGCACCAT CTGTCTTCAT CTTCCTGCCA
2641 TCTGATGAGC AGTTGAAATC TGGAACCTGCC TCTGTTGTGT GCCTGCTGAA TAACTTCTAT
2701 CCCAGAGAGG CCAAAGTACA GTGGAAGGTG GATAACGCCC TCCAATCGGG TAACTCCCAG
2761 GAGAGTGTC AAGAGCAGGA CAGCAAGGAC AGCACCTACA GCCTCAGCAG CACCCTGACG
2821 CTGAGCAAAG CAGACTACGA GAAACACAAA GTCTACGCCT GCGAAGTCAC CCATCAGGGC
2881 CTGAGTTTAC CGGTGACAAA GAGCTTCAAC AGGGGAGAGT GTGCGGCCGC TGGTAAGCCT
2941 ATCCCTAACC CTCTCCTCGG TCTCGATTCT ACGTGATAAC TTCACCGGTC AACCGCTGAT
3001 GAGAATTAC TGGCCGTCGT ITTACAACGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA
3061 CTTAATCGCC TTGCAGCACA TCCCCTTTT GCCAGCTGGC GTAATAGCGA AGAGGCCCGC
3121 ACCGATCGCC CTTCCTAACA GTTGCAGCAG CTGAATGGCG AATGGCGCCT GATGCGGTAT
3181 TTTCTCCTTA CGCATCTGTG CGGTATTTCA CACCGCATAC GTCAAAGCAA CCATAGTCTC
3241 AGTACAATCT GCTCTGATGC CGCATAGTTA AGCCAGCCCC GACACCCGCC AACACCCGCT
3301 GACGCGCCCT GACAGGCTTG TCTGCTCCCG GCATCCGCTT ACAGACAAGC TGTGACCGTC
3361 TCCGGGAGCT GCATGTGTCA GAGGTTTTCA CCGTCATCAC CGAAACGCGC GA

```

Exemplo 4: *Dobbling* de CDRs

Os seguintes exemplos exemplificam o uso de *dobbling* na construção de bibliotecas sintéticas. A cadeia pesada (HC) 3-23 parente é diversificada em CDR1, 2, e 3. Esta diversidade é combinada com uma cadeia leve (LC) A27 sinteticamente diversificada. A diversidade será como se segue:

Exemplo 4.1 CDR1 de HC

A seguinte diversidade de *dobbling* permite 5.832 variantes. Ver Tabela 50. Na posição 31, Ser é o tipo de aminoácido da linha germinativa (GL). Consequentemente fazemos Ser três vezes mais provável do que os outros tipos. Uma vez que são permitidos 18 tipos, Ser seria permitida 15% das vezes e todos os outros são permitidos a 5%. Assim sendo, se não existir nenhuma seleção para o tipo de AA em 31, é mais provável que isolemos um Ab com Ser. Similarmente, em 33, o tipo de AA de GL é Ala e fazemos Ala 3 vezes tão provável (15%) como todos os outros (5%). Em 35, Ser é o tipo de AA de GL e tornamo-la três vezes tão provável como os outros.

Em todas as outras três posições excluimos Cys e Met. Excluimos Cys porque não queremos dissulfetos injustificados ou cisteínas não emparelhadas expostas que poderiam afetar adversamente a solubilidade e reatividade do Ab. Excluimos Met porque grupos laterais de metioninas expostos estão sujeitos a oxidação que pode alterar as propriedades de ligação e tempo de prateleira. Poderíamos fazer o tipo de aminoácido da linha germinativa 2, 3, 4, 5, 6, 8, ou 10 vezes mais provável do que os outros tipos de AA.

Tabela 50: Diversidade para CDR1 em 3-23

Posição	AA Parente	Permitidos
31	S (três vezes mais provável do que os outros)	ADEFGHKLNPQIRSTVWY
33	A (3 X mais provável)	ADEFGHKLNPQIRSTVWY
35	S (3 X mais provável)	ADEFGHKLNPQIRSTVWY

Ao longo desta divulgação, os aminoácidos mostrados como "Permitidos" são os aminoácidos que podem ser usados a uma dada posição. Por exemplo, na Tabela 50, na posição 31, os aminoácidos permitidos "ADEFGHKLNPQIRSTVWY" são mostrados. Isto indica que os aminoácidos A, D, E, F, G, H, K, L, N, P, Q, R, S, T, V, W, e Y são todos permitidos na posição 31.

Exemplo 4.2: CDR2 de HC

Em CDR2 permitimos diversidade nas posições 50, 52, 52a, 56, e 58. Em 50, 52, 56, e 58 permitimos todos os tipos de aminoácidos exceto Cys e Met e fazemos os tipos de AA de GL mais prováveis por três vezes. Poderíamos fazer o tipo de

AA de GL 2, 3, 4, 5, 6, 8, ou 10 vezes mais provável do que os outros tipos de AA.

Tabela 51: CDR2 de HC: Diversidade = 419.904

Posição	AA Parente	Permitidos
50	A (3 X mais provável)	ADEFGHKLNPQRSTVWY
52	S (3 X mais provável)	ADEFGHKLNPQRSTVWY
52a	G (3 X mais provável)	GPSY
56	S (3 X mais provável)	ADEFGHKLNPQRSTVWY
58	Y (3 X mais provável)	ADEFGHKLNPQRSTVWY

Diversidade combinada de CDR1 e CDR2 = 2,45 E 9

Exemplo 4.3 CDR3 de HC, comprimentos 3, 4, 5

CDR3 muito curtas podem ser feitas por *dobbling*. A Tabela 7 mostra várias sequências parente para CDR3 de comprimento 3. Em 94, as VH3s têm Arg e permitimos esta mudança, mas Lys é tornada 3 X mais provável. Em 95, F é encontrado nesta posição em JH1. Permitimos também Ser, Tyr, Asp, e Arg permitam carga pequena, carga grande, positiva, e negativa. Em 96, JH1 tem Q. Uma vez que Q é muito similar a Glu permitimos Glu como uma alternativa ácida mais Arg, Ser, Tyr, e Leu. Em97, His é o AA da linha germinativa de JH1. Permitimos carga negativa (D), carga positiva (R), polar pequeno (S), grande hidrofóbico (Y), e alifático (L). A sequência parente constitui até 4,5% da biblioteca, mas isto é combinado com uma grande diversidade em CDR1 e CDR2. O *dobbling* permite 360 sequências no total. As sequências menos prováveis ocorrem em 1 em 1792. A sequência mais provável (parente) ocorre cerca de 1 em 22.

Tabela 60: Uma CDR3 de HC *dobbled* de comprimento 3 (V-3JH1 da Tabela 7)

Posição	Aminoácido parente (fonte)	Permitidos
94	K (VH 3-23)	KR (3:1)
95	F (JH1)	FSYDR (3:1:1:1:1)
96	Q (JH1)	QERSYL (3:1:1:1:1:1)
97	H (JH1)	HDRSYL (3:1:1:1:1:1)
103	W (JH1)	W

A Tabela 61 mostra uma CDR3 de HC *dobbled* de comprimento 3. Aqui, K94 é fixada como o é W103. Tornamos o aminoácido do segmento D “parente” cinco vezes mais provável do que os outros tipos de AA permitidos.

Tabela 61: Uma CDR3 de HC *dobbled* de comprimento 3 de um fragmento D (V-3D1-1.1.2-JH1 da Tabela 7).

Posição	Parente	Permitidos
94	K (V 3-23)	K
95	T (D1-1.1.2)	TYRDL (5:1:1:1:1)
96	T (D1-1.1.2)	TYRDL (5:1:1:1:1)
97	G (D1-1.1.2)	GSYRDL (5:1:1:1:1:1)
103	W (JH1)	W

Neste exemplo (Tabela 62, usando V-4JH2 da Tabela 8), 94 está fixado como Lys. Em 95, JH2 tem Tyr e permitimos Ser, Asp, Arg, e Leu tal que o tamanho, carga, e hidrofobicidade se possam alterar para se adequarem ao antigénio. JH2 tem Phe em 96 e permitimos Ser, Tyr, Asp, Arg, e Leu. Em 97, JH2 tem Asp e permitimos Arg, Ser, Tyr, e Leu. Em 98, JH2 tem Leu e permitimos Ser, Tyr, Asp, e Arg. Este padrão permite 750 sequências distintas, das quais a parente é a

mais provável (1 em 18). As sequências menos prováveis ocorrem em 1 em 4608 ou 256 vezes menos provavelmente do que a mais provável.

Tabela 62: CDR3 de HC comprimento 4 de JH2 (V-4JH2 na Tabela 7)

Posição	AA Parente (fonte)	Permitidos
94	K (VH 3-23)	K
95	Y (JH2)	Y SDRL (4:1:1:1:1)
96	F (JH2)	F SYDRL (4:1:1:1:1:1)
97	D (JH2)	D RSYL (4:1:1:1:1)
98	L (JH2)	L SYDR (4:1:1:1:1)
103	W (JH2)	W

Na Tabela 63 existe um *dobbling* de V-4D3-10.1a-JH2 da Tabela 8. Em 94 permitimos Lys e Arg com Lys (a parente) quatro vezes mais provável como Arg. Em 95, D3-10.1a (*i.e.*, D3-10 na primeira grelha de leita e começando um AA 1) tem Leu; permitimos SYDR também com Leu 4 X tão provável como cada um dos outros tipos de AA. Em 96, D3-10.1a tem Leu novamente, e permitimos o mesmo menu. Em 97, D3-10.1a tem Trp e permitimos Ser, Tyr, Asp, e Arg com Trp 4 X tão provável. Em 98, D3-10.1a tem Phe e permitimos Ser, Tyr, Asp, e Arg também.

Tabela 63: CDR3 de HC de comprimento quatro de V-4D3-10.1a na Tabela 8

Posição	AA Parente (fonte)	Permitidos
94	K (VH 3-23)	K R (4:1)
95	L (D3-10.1a)	L SYDR (4:1:1:1:1)
96	L (D3-10.1a)	L SYDR (4:1:1:1:1)

Posição	AA Parente (fonte)	Permitidos
97	W (D3-10.1a)	W SYDR (4:1:1:1:1)
98	F (D3-10.1a)	F SYDR (4:1:1:1:1)
103	W	W

Exemplo 4.4: CDR3 de HC de comprimento 10 a 20

CDR3 de HC

Duas subbibliotecas, ambas com CDR3 de comprimento 16:

Tabela 52: Biblioteca 1: Diversidade = 5×10^{11} , a sequência “parente” ocorre em 1 em $1,5 \times 10^6$.

Posição	AA “Parente” (fonte)	Permitidos
94	K (3 X mais provável) (3-23)	K R (3:1)
95	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
96	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
97	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
98	D (3 X mais provável) (D2-21(2))	D YSRL (3:1:1:1:1)
99	S (3 X mais provável) (D2-21(2))	S YRDL (3:1:1:1:1)
100	S (3 X mais provável) (D2-21(2))	S YRDL (3:1:1:1:1)
101	G (3 X mais provável) (D2-21(2))	G ASYRDL (3:1:1:1:1:1:1)
102	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
102a	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
102b	Y (3 X mais provável) (D2-21(2))	Y SRDL (3:1:1:1:1)
102c	A (3 X mais provável) (JH1)	A SYRD (3:1:1:1:1)
102d	E (3 X mais provável) (JH1)	E RSYL (3:1:1:1:1)
102e	Y (3 X mais provável) (JH1)	Y SRDL (3:1:1:1:1)
102f	F (3 X mais provável) (JH1)	F YSRD (3:1:1:1:1)
102g	Q (3 X mais provável) (JH1)	Q ERSY (3:1:1:1:1)

Posição	AA "Parente" (fonte)	Permitidos
102h	H (3 X mais provável) (JH1)	H ERSYL (3:1:1:1:1:1)
103	W (JH1, fixado)	W

Tabela 53: Biblioteca 2: CDR3 comprimento 16; Diversidade é $3,0 \times 10^5$ e a sequência parente ocorre uma vez em $3,7 \times 10^5$.

Posição	AA "Parente" (fonte)	Permitidos
94	K (3 X mais provável) (3-23)	K R (3:1)
95	G (3 X mais provável) (D2-2(2))	G SYDRL (3:1:1:1:1:1)
96	Y (3 X mais provável) (D2-2(2))	Y SDRL (3:1:1:1:1:1)
97	C (fixado) (D2-2(2))	C
98	S (3 X mais provável) (D2-2(2))	S YRDL (3:1:1:1:1:1)
99	S (3 X mais provável) (D2-2(2))	S YRDL (3:1:1:1:1:1)
100	T (3 X mais provável) (D2-2(2))	T YRDL (3:1:1:1:1:1)
101	S (3 X mais provável) (D2-2(2))	S YRDL (3:1:1:1:1:1)
102	C (fixado) (D2-2(2))	C
102a	Y (3 X mais provável) (D2-2(2))	Y SDRL (3:1:1:1:1:1)
102b	T (3 X mais provável) (D2-2(2))	T YRDL (3:1:1:1:1:1)
102c	A (3 X mais provável) (JH1)	A SYDRL (3:1:1:1:1:1)
102d	E (3 X mais provável) (JH1)	E RSYL (3:1:1:1:1:1)
102e	Y (3 X mais provável) (JH1)	Y SDRL (3:1:1:1:1:1)
102f	F (3 X mais provável) (JH1)	F YSRDL (3:1:1:1:1:1)
102g	Q (3 X mais provável) (JH1)	Q ERSYL (3:1:1:1:1:1)
102h	H (3 X mais provável) (JH1)	H DRSYL (3:1:1:1:1:1)
103	W ((JH1))	W

A Tabela 65 mostra uma variação do *dobbling* de SEQ ID NO:898. A diversidade total permitida é $2,1 \times 10^3$. Uma síntese que produz $1,0 \times 10^8$, $3,0 \times 10^8$, $5,0 \times 10^8$, $1,0 \times 10^9$, ou $5,0 \times 10^9$ amostrará a diversidade adequadamente. O desenho de

SEQ ID NO:898 foi discutido acima. No *dobbling* de SEQ ID NO:898 é permitir o tipo de AA parente a três vezes acima de outros tipos de AA na maioria das posições. Em posições onde o parente é Tyr usamos então Tyr e Ser em quantidades iguais com Leu em metade dessa frequência. Os resíduos de Cys estão fixados. Permite-se que cada tipo de AA parente vá para um de Arg, Asp, Ser, Tyr, ou Leu (Leu pode ser omitida se a parente for hidrofóbica, tal como Phe). A sequência parente ocorrerá uma vez em $1,0 \times 10^8$ membros. As sequências menos prováveis ocorrerão uma vez em $9,5 \times 10^{16}$. Não é importante que a biblioteca contenha realmente a sequência parente, somente que contenha muitas sequências que se assemelhem à parente. Assim sendo, uma biblioteca que contenha $1,0 \times 10^7$, $5,0 \times 10^7$, $1,0 \times 10^8$, $3,0 \times 10^8$, $1,0 \times 10^9$, ou $5,0 \times 10^9$, quando combinada com diversidade em CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, e CDR3 de LC, proporcionará uma biblioteca que conterá muitos Abs variáveis.

Tabela 65: *Dobbling* do Desenho 1 com SEQ ID NO:898 como parente

Posição	Parente (fonte)	Permitidos
94	K (VH 3-23)	K
95	D (Nenhuma fonte)	DSYL (3:1:1:1)
96	Y (Nenhuma fonte)	YSL (2:2:1)
97	G (D2-2.2)	GSYDRL (3:1:1:1:1:1)
98	Y (D2-2.2)	YSL (2:2:1)
99	C (D2-2.2)	C
100	S (D2-2.2)	SYDRL (3:1:1:1:1)
101	S (D2-2.2)	SYDRL (3:1:1:1:1)
102	T (D2-2.2)	TYDRL (3:1:1:1:1)

Posição	Parente (fonte)	Permitidos
102a	S (D2-2.2)	S YDRL (3:1:1:1:1)
102b	C (D2-2.2)	C
102c	Y (D2-2.2)	Y SL (2:2:1)
102d	T (D2-2.2)	T YDRL (3:1:1:1:1)
102e	Y (Nenhuma fonte)	Y DSL (3:1:1:1)
102f	G (Nenhuma fonte)	G SYRD (3:1:1:1:1)
102g	Y (Nenhuma fonte)	Y SL (2:2:1)
102h	S (Nenhuma fonte)	S YDRL (3:1:1:1)
102i	Y (Nenhuma fonte)	Y SL (2:2:1)
102j	A (JH1)	A SYDR (3:1:1:1:1)
102k	E (JH1)	E RSYL (3:1:1:1:1)
102l	Y (JH1)	Y SL (2:2:1)
102m	F (JH1)	F SYDR (3:1:1:1:1)
102n	Q (JH1)	Q YSDRL (3:1:1:1:1:1)
102p	H (JH1)	H SYDRL (3:1:1:1:1:1)
103	W (JH1, FR4)	W

Exemplo 4.5 *Dobbling* de yycakGSGYCSGGSCYSFDYwgqgtltvtvss (SEQ ID NO:931)

A Tabela 80 mostra o *dobbling* de SEQ ID NO:931, um exemplo de uma CDR3 de HC de comprimento 15. A posição 94 é parte de FR3 é mantida constante. As posições 95 e 96 têm tipos de aminoácidos “parente” escolhidos do conjunto altamente usado de (YGDRS) e são G95 e S96. As próximas dez posições são retiradas de D2-15.2 (um segmento D moderadamente altamente usado contendo uma ansa fechada de dissulfeto). As últimas três posições são as posições 100, 101, e 102 de JH4 como mostrado na Tabela 3. Em cada posição tornamos o tipo de aminoácido parente três vezes mais provável do que

os outros tipos permitidos. Os resíduos de Cys estão fixados. Em 102e, Phe é três vezes mais provável do que o são YGSRD (*i.e.*, Phe é três vezes mais provável como o são qualquer um dos aminoácidos Y, G, S, R, ou D). A diversidade permitida é $1,46 \times 10^9$. A sequência parente é esperada em 1 em $6,9 \times 10^4$. Cada uma das sequências unicamente substituídas é cerca de $1/3$ tão provável; as duplamente substituídas são $1/9$ tão prováveis e assim por diante. As sequências que são compostas inteiramente por outros tipos de AA ocorrem em somente 1 em $1,1 \times 10^{11}$.

Cada uma das outras sequências na Tabela 21 pode ser *dobbled* do mesmo modo.

Tabela 80: *Dobbling* de yycakGSGYCSGGSCYSFDYwgqgltlvtvss (SEQ ID NO:931)

Posição	Parente (fonte)	Permitidos
94	K (VH 3-23)	K
95	G (Nenhuma fonte)	G YSRD (3:1:1:1:1)
96	S (Nenhuma fonte)	S GYRD (3:1:1:1:1)
97	G (D2-15.2)	G YSRD (3:1:1:1:1)
98	Y (D2-15.2)	Y GSRD (3:1:1:1:1)
99	C (D2-15.2)	C
100	S (D2-15.2)	S GYRD (3:1:1:1:1)
101	G (D2-15.2)	G YSRD (3:1:1:1:1)
102	G (D2-15.2)	G YSRD (3:1:1:1:1)
102a	S (D2-15.2)	S GYRD (3:1:1:1:1)
102b	C (D2-15.2)	C
102c	Y (D2-15.2)	Y GSRD (3:1:1:1:1)
102d	S (D2-15.2)	S GYRD (3:1:1:1:1)

Posição	Parente (fonte)	Permitidos
102e	F (JH4)	FYGSRD (3:1:1:1:1:1)
102f	D (JH4)	DGSRY (3:1:1:1:1)
102g	Y (JH4)	YGSRD (3:1:1:1:1)
103	W (JH4, FR4)	W

Exemplo 5: Diversidade de cadeias leves sintética

Para fazer anticorpos inteiros necessitamos de combinar uma biblioteca de cadeias pesadas com uma biblioteca de cadeias leves (LC). Em Abs naturais é frequentemente observado que a HC faz a maior parte da ligação e muitas bibliotecas têm dado pouca atenção à LC ou têm obtido a diversidade de LC a partir de dadores humanos. Para se ter diversidade suficiente para dar bons ligantes a quase todos os alvos desenhamos um programa de diversificação que excede o que o sistema imune humano usualmente proporciona. Não obstante, o programa é desenhado para originar LC completamente funcional que tem o mesmo tipo de mudanças como visto em Abs. naturais, somente mais algumas. Vkappa III A27 foi escolhida como a LC.

De uma biblioteca que compreende LCS kappa e lambda doados, uma coleção de 1266 Abs foi digitada. Entre VKIIIs, A27 é a mais frequentemente vista (Tabela 66) e emparelha bem como HC 3-23.

As CDRs de A27 contêm 12, 7, e 9 aminoácidos. A colocação de diversidade em todas estas posições pode não trabalhar bem: a) podem existir muitos membros instáveis ou não funcionais, e b) a diversidade em algumas posições pode não ajudar a melhorar a ligação. Reduzimos o número de posições variáveis de 28 para 16.

Estudamos a estrutura em 3D de 1 QLR que tem uma LC de A27. A estrutura de 1 GLR está publicamente disponível na Base de Dados de Proteínas RCDL. A partir desta, os resíduos marcados na Tabela 68 parecem úteis de variar. O T56 está a cerca de 10 Å de uma His em CDR3 de HC. A variação em 56 pode ser útil. G24 está somente a cerca de 7 Å a partir de um átomo em CDR3 de HC. A linha germinativa é R24; assim sendo, a variação em 24 pode ser útil.

A Tabela 69 mostra uma cassete de exibição que desenhamos para uso em pMID21. Assim sendo, as enzimas de restrição escolhidas não têm outros locais em pMID21. *SpeI* está na sequência de sinal *iii* e permite que a LC inteira seja inserida ou removida. *XmaI*, *PpuMI*, *EcoO109I*, e *BlpI* precedem CDR1. *SacII* está em FR2, separando CDR1 de CDR2. Alternativamente, um local *AvrII* poderia ser inserido na mesma posição. Os locais *BspEI* e *XhoI* estão em FR3 e um local *KpnI* está em FR4.

Reunimos 155 sequências A27 e analisamos o que acontece nas CDRs. A Tabela 70 mostra a análise. Na Tabela 70 mostramos o que é encontrado nos Abs da nossa biblioteca e o que colocaríamos em cada posição.

```
!Tabela 68: onde variar A27
!
!           22      3      3      5      5      89      9
!           45      0a     4      0      5      90      5
!1QLR      GASQSVS_NYLA  DASSRAT  QQYGSSPLT
!A27       RASQSVSSSYLA  GASSRAT  QQYGSSPLT
!           ** **** *  *  *  *  *****
GASQSVS é (SEQID NO:922)  DASSRAT é (SEQID NO:923)
QQYGSSPLY é (SEQID NO:924)
RASQSVSSSYLA é (SEQ ID NO:925)  GASSRAT é (SEQ ID NO:926)
```

A Tabela 68 mostra onde as CDRs de A27 seriam variegadas.

Tabela 67: Comparar seqs de AA de VKIII

VK3 é (SEQ ID NO:927)

```

FR1.....CDR1.....FR2.....CDR2...FR3.....CDR3.....FR4.....
      1      2      3      4      5      6      7      8      9      0
VK3  123456789012345678901234567890123456789012345678901234567890123456789
A27  DIVLTQSPATLSLSPGERATLSCRASGVSSSYLAHWYQKPGQAPRLLIYGASRATGVFARFSGSGSGTDFTLTIS
A11  E-----G-----G-----L-----D-----I-----D-----R-----YGSS-
L2   E--M-----V-----N-----T-----I-----E-----QS-----YNNW-
L16  E--M-----V-----N-----T-----I-----E-----QS-----YNNW-
L6   E-----G-----N-----D--N-----I-----P-----RSNNH-
L20  E-----G-----N-----D--N-----I-----P-----RSNNH-
L25  E--M-----S-----T-----I-----Q-----D--NL-

```

VK3 difere de A27 por E1D, G9A, I58V, D60A, R77S.

A27 e L6 diferem por G9A(FR1), Δ31a (em CDR1), G50D(CDR2), S53N(CDR2), G92S(CDR3), S93N(CDR3), S94W(CDR3) VK3 de US 7,264,963.

CDR1

R24, A25, e S26 estão demasiado longe do local de combinação para ajudar e foram mantidos constantes. O grupo lateral de V29 está enterrado; esta posição foi mantida constante como Val. Nas outras posições permitimos Y ou S e um *flip-flop* de carga (RE ou RD, dependendo de onde a amostra tinha mais de E ou D na posição em questão) mais outros tipos que eram frequentemente vistos. Usámos uma folha de cálculo do Excel para determinar se este padrão de variação daria a sequência parente a 0,8% se os "outros" AAs fossem substituídos a 5%, a 0,1% se os "outros" AAs fossem substituídos a 6,5%, e a 0,02% se o "outro" estivesse aa 9%. Na amostra de 155, 17 têm um AA eliminado (incluindo 1 QLR); assim sendo trataremos de ter S30a eliminado em ~8% dos membros.

CDR2

A partir da inspeção de 1 QLR vemos que CDR2 está algo remota em relação ao local de combinação. Têm existido mesmo sugestões de que mantivéssemos os resíduos nesta CDR constante. O estudo da estrutura em 3D sugere que a variação a G50, S53, e T56 poderia ser útil. S53 é o mais variável na amostra de 155, mas isto não prova que estas mudanças são úteis. Em 1 QLR, G50 foi mutado para R50. O grupo lateral de T56 está apontado na direção de CDR3 de HC e está a cerca de 11 Å a partir de um átomo em CDR3 de HC.

CDR3

Q89 e Q90 estão enterrados e a natureza não os varia frequentemente; estes resíduos não são variados. Y91 é empacotado contra CDR3 de HC e mudanças aqui alterariam o local de combinação e ocorrem. Em G92, $\phi = -80$ e $\psi = -15$ logo colocar um não Glu é factível; a natureza fá-lo em 47/155 casos. S93 está muito frequentemente variado ou eliminado. S94 está altamente exposto e está altamente variado. P95 está exposto e variado, L96 empacota contra CDR3 de HC: as mudanças aqui afetarão o local de ligação e ocorrem na natureza. T97 está enterrado e tem sido mantido constante/o aminoácido não é variado.

A sequência parente aparece a 0,000246 ou 1 em 4,06 E 3. A diversidade permitida é cerca de 2,1 E 12. Com duas eliminações a 8%, 84,6% dos membros terão comprimento total, 7,5% terão CDR1 curta e CDR3 de comprimento total, 7,4% terão CDR1 de comprimento total e CDR3 curta, e 0,6% terão ambas as eliminações.

Outras linhas germinativas não estavam na amostra.

Tabela 66: Distribuição de VLs em 1266 LCs selecionadas.

Kappas				Lambdas			
O12	VKI	313		1a	VL1	9	
O18	VKI	1		1e	VL1	7	
A20	VKI	26		1c	VL1	55	
A30	VKI	26		1g	VL1	46	
L14	VKI	2		1b	VL1	1	118
L1	VKI	5		2c	VL2	18	
L15	VKI	1		2e	VL2	23	
L5	VKI	83		2a2	VL2	79	
L8	VKI	10		2d	VL2	1	121
L12	VKI	77	544	3r	VL3	56	
O11	VKII	4		3j	VL3	4	
A17	VKII	17		3l	VL3	31	
A19	VKII	31	52	3h	VL3	22	113
A27	VKIII	155		4a	VL4	1	1
L2	VKIII	31		5c	VL5	1	1
L6	VKIII	88		6a	VL6	8	8
L25	VKIII	16	290	10a	VL10	6	6
B3	VKIV	12	12	Número de lambdas			368
Número de kappas			898	Abs totais na amostra			1266

!Tabela 69: Um gene de exibição para A27 em pM21J. !

IIIsinal::A27::Ckappa

A sequência de aminoácidos da Tabela 69 é (SEQ ID NO:928).

A sequência de ADN DA Tabela 69 é (SEQ ID NO:929).

A sequência de aminoácidos da Tabela 69 é (SEQ ID NO:928).

A sequência de ADN DA Tabela 69 é (SEQ ID NO:929).

```

!      sequência de sinal-----
!      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
!      M  K  K  L  L  S  A  I  P  L  V  V  P  F  Y
!  1  |atg|aaG|aaA|ctg|ctg|tct|gct|atc|ccA|CTA|GTt|gtc|cct|ttc|tat|
!      SpeI....
!
!      Sinal ----- FRI-----
!      16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
!      S  H  S  E1 I  V3 L  T5 Q  S7 P  G9 T  L  S12
!  46 |tct|cat|agt|gaa|atc|gtt|ctg|acc|cag|tcC|CCG|GGG|aCC|Ctg|tct|
!      XmaI....
!
!      FpuMI....
!      EcoO109I. (1/2)

```

A sequência de aminoácidos da Tabela 69 é (SEQ ID NO:928).

A sequência de ADN DA Tabela 69 é (SEQ ID NO:929).

```

!      FR1----- CDR1-----
!      31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
!      L13 S P G E R A T L S C23 R24 A S Q
!      91 |ctg|tct|ccg|ggg|gaa|cgt|gct|acG|CTg|AGC|tgt|cgt|gct|tct|caa|
!          B1pI.....
!
!      CDR1----- FR2-----
!      46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
!      S28 V S S S30a Y L A34 W Y Q Q K P G
!      136 |tcc|gtt|agC|TCC|TCT|tat|tta|gct|tgg|tat|cag|caa|aag|ccg|ggg|
!          BseRI...
!
!      FR2----- CDR2-----
!      61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
!      Q A P R45 L L I Y G50 A S S R A T56
!      181 |caa|gct|CCG|CGG|ctg|ttg|atc|tat|ggt|gcc|tct|agt|cgt|gct|act|
!          SacII..
!
!      FR3-----
!      76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
!      G I P D60 R F S G S65 G S G T D F
!      226 |ggc|atc|cct|gat|cgt|ttc|tct|ggc|tct|ggc|TCC|GGA|acc|gat|ttc|
!          BspEI..
!
!      FR3-----
!      91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
!      T L T I S R L E P E D F A V Y
!      271 |act|ctg|acc|att|tct|CGT|CTC|GAG|ccg|gaa|gat|ttc|gct|gtc|tac|
!          BsmBI..
!          XhoI...
!
!      FR3----- CDR3----- FR4-----
!      106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
!      Y C Q89 Q Y G S S P95 L T F G G G
!      316 |tat|tgt|caa|cag|tat|ggt|tct|agt|ccg|ctg|act|ttc|ggt|ggc|GGT|
!          KpnI...
!
!      FR4-----
!      121 122 123 124 125 126
!      T K V E I K
!      361 |ACC|aaa|gtc|gaa|atc|aag
!          KpnI..
!
!      Ckappa-----
!      R G T V A A P S V F I F P P S
!      379 |cgt|gga|act|gtg|gCT|GCA|Cca|tct|GTC|TTC|atc|ttc|ccg|cca|tct|
!          BsgI.... BbsI...
!
!      D E Q L K S G T A S V V C L L
!      424 |gat|gag|cag|ttg|aaa|tct|gga|act|gcc|tct|ggt|gtg|tgc|ctg|ctg|
!
!      N N F Y P R E A K V Q W K V D
!      469 |aat|aac|ttc|tat|ccc|aga|gag|gcc|aaa|gta|cag|tgg|aag|gtg|gat|
!
!      N A L Q S G N S Q E S V T E Q
!      514 |aac|gcc|ctc|caa|tcg|ggt|aac|tcc|cag|gag|agt|gtc|aca|gag|cag|

```

A sequência de aminoácidos da Tabela 69 é (SEQ ID NO:928).

A sequência de ADN DA Tabela 69 é (SEQ ID NO:929).

```

!      D S K D S T Y S L S S T L T L
559    gac agc aag gac agc acc tac agc ctc agc agc acc ctg act ctg
!
!      S K A D Y E K H K V Y A C E V
604    tcc aaa gca gac tac gag aaa cac aaa GTC TAC gcc tgc gaa gtc
!
!      T H Q G L S S P V T K S F N R
649    acc cat cAG GGC CTg agt tCA CCG GTG aca aag agc ttc aac agg
!      AlwNI..... SgrAI.....
!      Eco0109I.(2/2) AgeI....
!
!      G E C . .
694    gga gag tgt taa taa
!
!      GG CGCGCCaatt
709
!      AscI.....
!      BssHII.
!

```

Tabela 70: Inventário de mutações em CDRs de ABs A27

CDR1				
R24	1,	3G, 1T, 151-		Fixado
A25	2,	3T, 152-		Fixado
S26	3,	1R, 154-		Fixado
Q27	4,	3E , 1H, 1L, 1P, 4R , 145-	9%	ERYSL
S28	5,	1A, 2F, 2G, 1I, 2L, 5N , 1P, 1R, 10T , 1V, 1Y , 128-	9%	NTYERL
V29	6,	1F, 19I, 6L, 129-		Fixado
S30	7,	2A, 2D , 8G, 2H, 1I, 11N , 9R , 6T , 4V , 2Y , 108-	9%	DNRTY
S30a (8% eliminar 30a)	8,	1A, 2F, 6G , 1H, 6N , 1P, 10R , 6T , 3Y , 119 -	9%	GNRTYD
S31	9,	1A, 5D , 3F , 4G , 1H, 2I, 4K, 1L, 31N , 19R , 7T , 7Y , 70-	9%	DFGNRTY
Y32	10,	5F, 1K, 14L, 4N, 4Q, 2R, 8S, 3V, 1W, 113 -	9%	FDLNQRSY
L33	11,	16A , 1F, 4I, 1N, 1S, 8V, 1Y, 123 -		Fixado
A34	12,	2G, 2L, 1N, 1S, 4V, 128-	9%	SY

CDR1

- 13, 2A, 1G,
 - 14, 1S,
 - 15, 1S,
 - 16, 1Y,
 - 17, 1L,
 - 18, 1A,

Nota: um anticorpo tinha uma inserção de seis AAs em CDR1! Dois outros Abs tinham uma inserção única. Dezassete Abs têm uma eliminação de AA em CDR1.

CDR2

G50 1, 10A, 11D, 1H, 2R, 2S, 1V, 7Y, **121-**, 9% DRSYL
 A51 2, 7G, 2I, 6S, 7T, 2V, 131-, Fixado
 S52 3, 6A, 3F, 1G, 1T, 144-, Fixado
 S53 4, 1A, 1G, 1H, 5I, 2K, 16N, 7R, 16T, 9% NTSYER
106-,
 R54 5, 1A, 1I, 1N, 1S, 3T, 1Y, 147-, Fixado
 A55 6, 2P, 7R, 4S, 2V, 140-, Fixado
 T56 7, 10A¹, 1G, 1H, 2P, 4S, 137-, 9% ERSY
 - 8, 1A, 6T,

Note-se, existem dezassete anticorpos com uma inserção de um AA.

CDR3 (mostrando "_" significa que o Ab tem uma eliminação em CDR3)

Q89 1, 5H, 1L, 2M, 147-, Fixado
 Q90 2, 1E, 1F, 13H, 2K, 2L, 4R, 1S, 1Y, 130-, Fixado
 Y91 3, 2A, 8F, 2G, 2H, 1L, 1P, 13R, 4S, **122-**, 9% FERS
 G92 4, 10A, 3D, 2H, 1I, 1L, 2N, 6R, 12S, 2V, 3Y, **108-**, 9% ADRSTY
 5_,
 S93 5, 1A, **2D**, **2F**, 6G, 2H, 3I, 2K, 2M, **14N**, 1P, 1Q, **8R**, 9% DFNRTY
17T, **2Y**, **86-**, 6_,

CDR3 (mostrando “_” significa que o Ab tem uma eliminação em CDR3)

(8% têm 93 eliminado)

S94 6, 3A, 6F, 1I, 3L, 3P, 2R, 2T, 11W, **117-**, 7_, 9% WERYS

P95 7, 2A, 1E, 1G, 1K, 3L, 1M, 7R, 3S, 3T, 1V, **108-**, 9% ERYs
24_,

L96 8, 2A, **2E**, 3F, 3G, 1H, 1X, **3K**, 7L, 2M, **24P²**, 6Q, 9% ERPYS
28R, 3S, 3T, 7V, 2Y, **58-**,

T97 9, 2A, 1IF, 2G, 3I, 1K, 2L, 3M, 1N, 1R, 6S, 3V, 2Y, Fixado
128-,

- 10, 1A, 1S, 34-,

- 11, 1S, 7-,

- 12, 1A,

Existe um Ab com uma inserção de 3 AAs.

Cinco têm eliminações de 4 AAs, 1 tem uma eliminação de 3 AA, 1 tem uma eliminação de 2 AA, e 17 tem uma eliminação de AA.

Tabela 71: Diversidade permitida em CDR1, 2, e 3 de
A27::JK4.

Posição	parente	permitidos	
CDR1			
42 (24)	R	fixado	
43 (25)	A	fixado	
44 (26)	S	fixado	
45 (27)	Q	ERYSL	55% Q 9% outro
46 (28)	S	NTYERL	46% S 9% outro
47 (29)	V	fixado	
48 (30)	S	DNRTY	55% S 9% outro
49 (30a)	S	GNRTYD	46% S 9% outro
8% têm 30a eliminado			
50 (31)	S	DFGNRTY	44% S 8% outro

Posição	parente	permitidos	
CDR1			
51 (32)	Y	FDLNQRSY	44% Y 7% outro
52 (33)	L	fixado	
53 (34)	A	SY	70% A 15% outro
CDR2			
69 (50)	G	DRSYL	55% G 9% outro
70 (51)	A	Fixado	
71 (52)	S	Fixado	
72 (53)	S	NTSYER	52% S 8% outro
73 (54)	R	Fixado	
74 (55)	A	Fixado	
75 (56)	T	ERSY	64% T 9% outro
CDR3			
108 (89)	Q	fixado	
109 (90)	Q	fixado	
110 (91)	Y	FERS	64% Y 9% outro
111 (92)	G	ADRSTY	52% G 8% outro
112 (93)	S	DFNRTY	52% S 8% outro
113 (94)	S	WERYS	55% S 9% outro
114 (95)	P	ERYS	64% P 9% outro
8% têm P95 eliminado			
115 (96)	L	ERPYS	55% L 9% outro
116 (97)	T	fixado	
¹ Sete destes vêm das inserções. ² Alguns destes aparecem devido às inserções.			

A sequência parente aparece a $5,32 \times 10^{-5}$ ou 1 em $1,88 \times 10^4$.

As sequências com uma única substituição têm uma probabilidade entre $1,1 \text{ E } -5$ e $7,5 \text{ E } -6$.

As sequências que não têm nenhum dos AAs parente ocorrem em 1 em $6,7 \text{ E } 16$.

A diversidade permitida é cerca de $2,35 \text{ E } 12$.

Tabela 75: Frequências de aminoácidos CDR3s de HC.

AA	Número	%	Rel para cima	Rel para baixo
Y	3428	15,64	50,41	1,00
G	3244	14,80	47,71	0,95
D	2622	11,96	38,56	0,76
S	1777	8,11	26,13	0,52
R	1337	6,10	19,66	0,39
F	1328	6,06	19,53	0,39
A	1213	5,53	17,84	0,35
V	1141	5,20	16,78	0,33
L	816	3,72	12,00	0,24
I	745	3,40	10,96	0,22
P	726	3,31	10,68	0,21
T	586	2,67	8,62	0,17
W	566	2,58	8,32	0,17
M	560	2,55	8,24	0,16
N	462	2,11	6,79	0,13
E	363	1,66	5,34	0,11
K	355	1,62	5,22	0,10
H	327	1,49	4,81	0,10
Q	259	1,18	3,81	0,08
C	68	0,31	1,00	0,02

AA	Número	%	Rel para cima	Rel para baixo
Total	21923			

Tabela 76: Distribuição de comprimentos de CDR3 de HC

Comprimento	Número de Anticorpos	Soma	Mediana
1	0		
2	0		
3	2	2	
4	21	23	
5	16	39	
6	100	139	
7	36	175	
8	78	253	
9	155	408	
10	153	561	
11	134	695	11,12
12	123	818	
13	133	951	
14	92	1043	
15	87	1130	
16	71	1201	
17	59	1260	
18	41	1301	
19	40	1341	
20	22	1363	
21	21	1384	
22	15	1399	
23	7	1406	

Comprimento	Número de Anticorpos	Soma	Mediana
24	7	1413	
25	1	1414	
26	1	1415	
27	3	1418	
28	0	1418	
29	0	1418	
30	1	1419	
	1419	709,5	

Tabela 77: Utilização de segmentos D (limite de exclusão a correspondência de 0,70)

3-22.2	38	YYDSSGYYY
4-17.2	27	DYGDY
3-3.2	25	YYDFWSGYT
6-19.1	25	GYSSGWY
7-27.1	19	LTG
5-5.3	18	GYSYGY
6-13.1	18	GYSSSWY
5-12.3	13	GYSGYDY
6-13.2	10	GIAAAG
1-26.3	9	YSGSY
2-15.2	9	GYCSGGSCYS
4-4.3	9	TTVT
3-10.2	8	YYYGSGSYN
1-1.3	7	YNWND
4-4.2	7	DYSNY
2-2.2	6	GYCSSTSCYT

3-16.2	6	YYDYVWGSYRYT
6-6.1	6	EYSSSS
6-19.2	6	GIAVAG
3-9.1	5	VLRYFDWLL@
4-23.2	5	DYGGNS
6-6.2	5	SIAAR
1-7.3	4	YNWNY
2-2.3	4	DIVVVPAAI
4-23.3	4	TTVVT
1-7.1	3	GITGT
1-26.1	3	GIVGAT
7-27.3	3	NWG
3-10.1	2	VLLWFGELL@
3-10.2	2	ITMVRGVII
5-5.1	2	VDTAMV
5-5.2	2	WIQLWL
5-12.1	2	VDIVATI
5-24.3	2	RDGYNY
1-1.1	1	GTTGT
2-21.3	1	HIVVVTAI
3-3.3	1	ITIFGVVII
5-24.2	1	*RWLQL
6-6.3	1	V*QLV
6-19.3	1	V*QWLV

Tabela 78: Utilização de segmentos D (limite de exclusão de 0,667)

Nome	Número	Sequência	%
Nenhum	935		0,517

Nome	Número	Sequência	%
7-27.1	158	LTG	0,087
7-27.3	98	NWG	0,054
5-5.3	72	GYSYGY	0,040
1-26.3	67	YSGSYY	0,037
3-22.2	46	YYYDSSGYYY	0,025
4-17.2	38	DYGDY	0,021
3-3.2	37	YYDFWSGYT	0,020
7-27.2	37	@ LG	0,020
6-19.1	33	GYSSGWY	0,018
6-13.2	31	GIAAAG	0,017
6-13.1	22	GYSSSWY	0,012
6-6.1	18	EYSSSS	0,010
6-19.2	18	GIAVAG	0,010
4-23.2	17	DYGGNS	0,009
5-12.3	17	GYSGYDY	0,009
5-24.3	14	RDGYNY	0,008
2-15.2	13	GYCSGGSCYS	0,007
1-26.1	11	GIVGAT	0,006
4-4.3	11	TTVT	0,006
1-1.3	9	YNWND	0,005
2-2.2	9	GYCSSTSCYT	0,005
3-16.2	9	YYDYVWGSYRYT	0,005
2-2.3	8	DIVVVPAAI	0,004
3-10.2	8	YYYGSGSYYN	0,004
4-4.2	8	DYSNY	0,004
1-7.3	7	YNWNY	0,004
3-3.3	6	ITIFGVVII	0,003
6-6.2	6	SIAAR	0,003

Nome	Número	Sequência	%
3-9.1	5	VLRYFDWLL@	0,003
3-10.2	5	ITMVRGVII	0,003
6-19.3	5	V*QWLV	0,003
1-7.1	4	GITGT	0,002
4-23.3	4	TTVVT	0,002
1-1.1	3	GTTGT	0,002
5-5.1	3	VDTAMV	0,002
5-24.2	3	*RWLQL	0,002
3-10.1	2	VLLWFGELL@	0,001
5-5.2	2	WIQLWL	0,001
5-12.1	2	VDIVATI	0,001
1-26.2	1	V*WELL	0,001
2-21.2	1	AYCGGDCYS	0,001
2-21.3	1	HIWVTAI	0,001
3-3.1	1	VLRFLEWLLY	0,001
3-16.2	1	IMITFGGVIVI	0,001
6-6.3	1	V*QLV	0,001
6-13.3	1	V*QQLV	0,001

Tabela 78: Utilização de segmentos JH

JH1	17
JH2	31
JH3	452
JH4	636
JH5	32
JH6	251

Exemplo 6: ADN *wobbled* para CDR3 de HC 16d

A Tabela 400 mostra um segmento de ADN de um local *XbaI* em FR3 para um local *BstEII* em Fr4. A CDR3 de HC consiste em SYSY::D2-2(2)::QH seguido pela região FR4 de JH1. A QH é encontrada na linha germinativa de JH1. Na união V-D-J, as células imunes modificam frequentemente as extremidades de V, D, e J. Assim sendo, a construção corresponde ao que é muito possível na construção e maturação de genes de imunoglobulina reais. Por *wobbling* da síntese obtemos uma grande coleção de genes que se assemelham ao que resultaria da união de 3-23 com uma região D ou com uma JH1 pouco modificada seguida por algumas mutações. Na biblioteca 16d existem duas cisteínas que formam presumivelmente um dissulfeto, estas não foram *wobbled*.

A Tabela 500 mostra a distribuição esperada de tipos de aminoácidos em cada posição na biblioteca 16d. O *doping* do *wobbling* foi definido a 73:9:9:9. A sequência mais provável é aquela mostrada na Tabela 21 e deve estar presente a uma frequência de $4,8 \times 10^{-5}$. Somente 55% das sequências estão isentas de paragem e 74% estão isentas de ochre ou opel. Se a biblioteca for expressa em células supE, este é um número importante. Seria valioso remover as sequências com codões de paragem como discutido noutro lugar aqui. Pode-se ver que é previsto que aquelas posições que começam como S tenham S 54% das vezes e Y 5,4% enquanto aquelas que começam como Y tenham Y 44% das vezes e S 7,2%. Em cada posição existem 7-9 tipos de AA que aparecem a >1%. Existem 14 posições variegadas. As sequências que serão o mais eficazmente amostradas são cerca de $8^{14} = 4,3 \times 10^{12}$.

Tabela 400: Cassete para exibição de CDR3 de HC 16d wobbled

FR3														
68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
T	I	S	R	D	N	S	K	N	T	L	Y	L	Q	M
1216	act atc TCT AGA gac aac tct aag aat act ctc tac ttg cag atg													
	XbaI													
FR3														
82a	82b	82c	83	84	85	86	87	88	89	90	91	92	93	94
N	S	L	R	A	E	D	T	A	V	Y	Y	C	A	K
1261	aac agC TTA AGg gct ggg gac act gca gtc tac tat tgc gct aaa													
	AflII													

$e = 0.73 A + 0.09 C + 0.09 G + 0.09 T$
 $q = 0.09 A + 0.73 C + 0.09 G + 0.09 T$
 $j = 0.09 A + 0.09 C + 0.73 G + 0.09 T$
 $z = 0.09 A + 0.09 C + 0.09 G + 0.73 T$

Os valores 0.73 e 0.09 são escolhidos tal que $0.73 + 3 \cdot 0.09 = 1.0$
Poderiam ser usadas outras razões.

95	96	97	98	99	100	101	102	102	102	102	102	102	102	102
S	Y	S	Y	G	Y	c	S	S	T	S	c	Y	T	Q
zqz	zez	zqz	zez	jjz	zez	TGT	zqz	zqz	eqz	zqz	TGT	zez	eqz	qez

FR4												
103	104	105	106	107	108	109	110	111	112	113		
W	G	Q	G	T	L	V	T	V	S	S		
TGg ggt caa ggt act ttG GTC ACC gtc tct agt												
BstEII												

Tabela 500: Distribuição esperada de tipos de AA em CDR3 de HC 16d wobbled

"," = TGA ou TAA; "b" = TAG

S Y S Y G Y c S S T S c Y T Q H
 zqz zez zqz zez jjz zez tqt zqz zqz eqz zqz tqt zez eqz qej qez

Pureza da base nominal = 0.7300 outras = 0.0900

s(zqz)	y(zez)	s(zqz)	y(zez)	q(jjz)	y(zez)	C(TGT)	s(zqz)	s(zqz)	t(eqz)
1 s 5.4-01	y 4.4-01	s 5.4-01	y 4.4-01	q 5.3-01	y 4.4-01	c 1.000	s 5.4-01	s 5.4-01	t 5.3-01
2 p 6.6-02	s 7.2-02	p 6.6-02	s 7.2-02	r 7.8-02	s 7.2-02		p 6.6-02	p 6.6-02	s 1.2-01
3 a 6.6-02	f 5.4-02	a 6.6-02	f 5.4-02	a 5.6-02	f 5.4-02		a 6.6-02	a 6.6-02	a 6.6-02
4 t 6.6-02	h 5.4-02	t 6.6-02	h 5.4-02	v 6.6-02	h 5.4-02		t 6.6-02	t 6.6-02	p 6.6-02
5 f 5.4-02	n 5.4-02	f 5.4-02	n 5.4-02	a 6.2-02	n 5.4-02		f 5.4-02	f 5.4-02	i 6.0-02
6 c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02		c 5.4-02	c 5.4-02	n 5.4-02
7 y 5.4-02	d 5.4-02	y 5.4-02	d 5.4-02	d 5.4-02	d 5.4-02		y 5.4-02	y 5.4-02	r 2.0-02
8 l 2.0-02	5.4-02	l 2.0-02	5.4-02	e 1.2-02	5.4-02		l 2.0-02	l 2.0-02	k 1.2-02
9 1.2-02	b 4.8-02	1.2-02	b 4.8-02	l 9.6-03	b 4.8-02		1.2-02	1.2-02	l 9.6-01
10 r 9.6-03	l 2.0-02	r 9.6-03	l 2.0-02	t 8.1-03	l 2.0-02		r 9.6-03	r 9.6-03	q 8.1-03
11 q 8.1-03	k 1.2-02	q 8.1-03	k 1.2-02	p 8.1-03	k 1.2-02		q 8.1-03	q 8.1-03	v 8.1-03
12 v 8.1-03	q 1.2-02	v 8.1-03	q 1.2-02	i 7.4-03	q 1.2-02		v 8.1-03	v 8.1-03	f 6.6-03
13 i 7.4-03	e 1.2-02	i 7.4-03	e 1.2-02	6.6-03	e 1.2-02		i 7.4-03	i 7.4-03	e 6.6-03
14 h 6.6-03	r 9.6-03	h 6.6-03	r 9.6-03	f 6.6-03	r 9.6-03		h 6.6-03	h 6.6-03	h 6.6-03
15 n 6.6-03	t 8.1-03	n 6.6-03	t 8.1-03	h 6.6-03	t 8.1-03		n 6.6-03	n 6.6-03	d 6.6-03
16 d 6.6-03	v 8.1-03	d 6.6-03	v 8.1-03	y 6.6-03	v 8.1-03		d 6.6-03	d 6.6-03	y 6.6-03
17 w 5.9-03	a 8.1-03	w 5.9-03	a 8.1-03	n 6.6-03	a 8.1-03		w 5.9-03	w 5.9-03	m 5.9-03
18 b 5.9-03	g 8.1-03	b 5.9-03	g 8.1-03	w 5.9-03	g 8.1-03		b 5.9-03	b 5.9-03	q 1.5-03
19 q 1.5-03	p 8.1-03	q 1.5-03	p 8.1-03	q 1.5-03	p 8.1-03		q 1.5-03	q 1.5-03	e 1.5-03
20 k 1.5-03	i 7.4-03	k 1.5-03	i 7.4-03	k 1.5-03	i 7.4-03		k 1.5-03	k 1.5-03	1.5-03
21 e 1.5-03	w 5.9-03	e 1.5-03	w 5.9-03	m 7.3-04	w 5.9-03		e 1.5-03	e 1.5-03	w 7.3-04
22 m 7.3-04	m 7.3-04	m 7.3-04	m 7.3-04	b 7.3-04	m 7.3-04		m 7.3-04	m 7.3-04	b 7.3-04

	s(rqz)	C(TGT)	y(tez)	t(eqz)	q(qej)	h(qez)
1	s 5.4-01	c 1.000	y 4.4-01	t 5.3-01	q 4.4-01	h 4.4-01
2	p 6.6-02		s 7.2-02	s 1.2-01	h 9.6-02	q 9.6-02
3	a 6.6-02		f 5.4-02	a 6.6-02	l 7.2-02	l 6.7-02
4	t 6.6-02		h 5.4-02	p 6.6-02	r 7.2-02	r 6.7-02
5	f 5.4-02		n 5.4-02	i 6.0-02	p 6.6-02	p 6.6-02
6	c 5.4-02		c 5.4-02	n 5.4-02	e 5.4-02	n 5.4-02
7	y 5.4-02		d 5.4-02	r 2.0-02	x 5.4-02	d 5.4-02
8	l 2.0-02		.	5.4-02	k 1.2-02	b 4.8-02
9	.		b 4.8-02	l 9.6-03	d 1.2-02	s 1.5-02
10	r 9.6-03		l 2.0-02	g 8.1-03	y 1.2-02	k 1.2-02
11	g 8.1-03		k 1.2-02	v 8.1-03	n 1.2-02	e 1.2-02
12	v 8.1-03		q 1.2-02	f 6.6-03	s 9.6-03	q 8.1-03
13	i 7.4-03		e 1.2-02	c 6.6-03	t 8.1-03	t 8.1-03
14	h 6.6-03		r 9.6-03	h 6.6-03	v 8.1-03	v 8.1-03
15	n 6.6-03		t 8.1-03	d 6.6-03	a 8.1-03	a 8.1-03
16	d 6.6-03		v 8.1-03	y 6.6-03	q 8.1-03	i 7.4-03
17	w 5.9-03		a 8.1-03	m 5.9-03	.	6.6-03
18	b 5.9-03		q 8.1-03	q 1.5-03	w 5.9-03	c 6.6-03
19	q 1.5-03		p 8.1-03	e 1.5-03	m 5.9-03	f 6.6-03
20	x 1.5-03		i 7.4-03	.	1.5-03	i 2.2-03
21	e 1.5-03		w 5.9-03	w 7.3-04	f 1.5-03	w 7.3-04
22	m 7.3-04		m 7.3-04	b 7.3-04	c 1.5-03	m 7.3-04

A sequência mais frequente tem a frequência = 4.8E-05
 Fração isenta de paragem = 5.5E-01
 Fração isenta de (TAA&TGA) = 7.4E-01

		F%	F%	F%
D1	1-1	0,42	0,14	2,90
	1-7	0,42	0,28	1,24
	1-20	0,00	0,00	0,00
	1-26	0,00	0,57	1,80
D2	2-2	0,55	4,30	1,21
	2-8	0,00	0,57	0,41
	2-15	0,28	4,03	0,94
	2-21	0,00	2,22	0,94
D3	3-3	0,94	4,44	3,70
	3-9	0,67	1,82	0,00
	3-10	0,67	5,78	1,55
	3-16	1,08	2,49	0,67
	3-22	0,14	7,87	0,81
D4	4-4	0,28	0,59	0,28
	4-11	0,00	0,00	0,00
	4-17	0,00	4,03	2,76
	4-23	0,14	1,41	0,54
D5	5-5	1,34	0,40	4,30
	5-12	1,08	0,00	1,95
	5-18	0,00	0,00	0,00
	5-24	0,67	1,55	1,82
D6	6-6	1,21	1,55	0,13
	6-13	4,84	2,62	0,27
	6-19	6,66	1,95	0,54
D7	7-27	0,27	0,13	0,27
Fração total %		21,65	49,34	29,01

Tabela 808: LC K1(O12)::JK1

```

.. Seq. líder ->|----- FR1 ----->
      G  V  H  S  A  Q  D  I  Q  M  T  Q  S  P  S  S  L
1  |ggT|GTA|CAG|aGT|GCT|Cag|gat|att|cag|atg|act|caa|tct|ccC|TCG|AGT|ctg|
   |BsrGI...  ApaLI...                               XhoI....

----- FR1 ----->|----- CDR1 ->
      12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
   |S  A  S  V  G  D  R  V  T  I  T  C  R  A  S
46 |tct|gct|tct|gtc|gGC|GAT|CGC|gtt|act|act|act|tgt|cgt|gct|tcc|
   |SgfI.....

----- CDR1 ----->|----- FR2 ----->
      27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
   |Q  S  I  S  S  Y  L  N  W  Y  Q  Q  K  P  G
91 |cag|tcc|att|tct|agc|tat|ctg|aat|tGG|TAC|Cag|caa|aag|ccg|ggt|
   |KpnI....

----- FR2 ----->|----- CDR2 ----->
      42 43 44 45 46 47 48 49 50 51 52 53 54 55 56

```

```

      K  A  P  K  L  L  I  Y  A  A  S  S  L  Q  S
|aag|gct|ccg|aaa|ctg|tta|atc|tat|gcc|gct|tct|agt|ctg|cag|tct|

----- FR3 ----->
57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
G  V  P  S  R  F  S  G  S  G  S  G  T  D  F
|ggg|ggt|ccg|TCT|AGA|ttc|tct|ggc|tct|ggg|tct|ggg|act|gat|ttt|
XbaI...

----- FR3 ----->
72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
T  L  T  I  S  S  L  Q  P  E  D  F  A  T  Y
|act|ctg|act|att|tcc|tct|ctg|caa|ccg|gag|gac|ttt|gct|acc|tat|

- FR3->|---- CDR3 ----->|---- FR4 ----->
87 88 89 90 91 92 93 94 95 96 97 98 99 100 101
Y  C  Q  Q  S  Y  S  T  P  W  T  F  G  Q  G
|tac|tgc|caa|cag|tct|tat|agt|act|ccg|tgg|act|ttc|ggg|caa|ggc|

---- FR4 ----->|---- Ckappa----->
102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
T  K  V  E  I  K  R  T  V  A  A  P  S  V  F
|act|aaa|ggt|gag|att|aag|CGT|ACG|gtg|gct|gct|ccg|tct|gtc|ttc|
BsiWI..

```

Tabela 900: Diversidade de CDR1												
												Diversidade
Posição	24	25	26	27	28	29	30	31	32	33	34	
O12	R	A	S	Q	S	I	S	S	Y	L	N	
diversidade	2	2	1	1	3	1	2	2	4	1	3	576
permitidos	Q	M			D		R	N	D		A	
					G				W		G	
									A			

Tabela 1000: Grande diversidade de CDR1												
												Diversidade
Posição	24	25	26	27	28	29	30	31	32	33	34	
O12	R	A	S	Q	S	I	S	S	Y	L	N	
diversidade	3	2	4	1	5	1	4	5	5	1	6	72000

Tabela 1000: Grande diversidade de CDR1

											Diversidade
permitidos	Q	M	E		D		R	N	D		A
	E		R		G		E	E	W		G
			Y		R		Y	R	A		D
					Y			Y	R		R

Tabela 1100: Diversidade de CDR2

POSIÇÃO	50	51	52	53	54	55	56	Diversidade
O12	A	A	S	S	L	Q	S	
diversidade	2	1	1	3	1	2	2	24
permitidos	D			N		E	T	
				T				

Tabela 1200: Grande diversidade de CDR2

POSIÇÃO	50	51	52	53	54	55	56	Diversidade
O12	A	A	S	S	L	Q	S	
diversidade	4	1	4	6	1	4	5	1920
permitidos	D	D E N	E	N		E	T	
	R		R	T		R	Y	
	Y		Y	E		Y	R	
				R			E	
				Y				

Tabela 1300: Diversidade de CDR3

[illegible]

Tabela 1300: Diversidade de CDR3											
diversidade	2	2	6	3	3	5	2	1	1	2160	
permitidos	L	K	Y	D	N	T	S				
			H	N	Y	L					
			F			Y					
			A			F					
			D								

Tabela 1400: Grande diversidade de CDR3											
Posição	93	94	95	96	97	98	99	100	101	div. tot.	
012	Q	Q	S	Y	S	S	P	W	T		
diversidade	6	1	7	7	6	5	2	6	1	105840	
permitidos	L		Y	D	N	T	S	F			
	E		H	N	Y	L		Y			
	R		F	R	D	Y		H			
	Y		A	A	R	F		L			
	A		D	L	A	E		I			
			R	S		R					

Exemplo 7: Exemplos adicionais de CDR3s de HC sintéticas

Uma coleção de 22.063 Fabs com CDR3 distintas que tinham sido selecionados da biblioteca de FAB-310 ou FAB-410 e que eram positivos quanto a ELISA quanto a pelo menos um antígeno foi examinada. A utilização de cadeias de JH é mostrada na Tabela 1001; a parte de FR4 de cada JH é mostrada negrito. A Tabela 1010 mostra a utilização de aminoácidos nas CDR3 de HC. A Tabela 1020 mostra a distribuição de comprimentos de CDR3. O comprimento médio é 11,5.

A Tabela 1030 mostra a utilização de segmentos D nas CDR3s. Um segmento D foi identificado é 70% dos aminoácidos correspondidos; existiram 5.654 casos (25,6%). Os Ds mais usados foram 3-3.2 (743, sequência: YYDFWSGYT), 3-22.2 (617, sequência: YYDSSGY), 6-19.1 (441, sequência: GYSSGWY), 6-13.1 (399, sequência: GYSSSWY), e 4-17.2 (392, sequência: DYGDY). Dos Ds contendo resíduos de Cys emparelhados, 2-15.2 (sequência: GYCSGGSCYS) foi o mais usado; existiram 139 exemplos que são 0,6% da coleção.

Quando V ou V::D é unido com J existe frequentemente modificação da extremidade 3' de V ou V::D e a extremidade 5' e J. A inspeção de muitas sequências CDR3-FR4 mostra que existe frequentemente uma porção de JH constituindo parte de CDR3. Existem frequentemente mutações nos resíduos de CDR3 correspondendo aos resíduos de JH 1-9. Aqui, a porção de CDR3 que se pensa que deriva de JH é chamada a "*J stump*". A JH usada numa cadeia pesada é determinada por comparação de cada um dos resíduos das seis cadeias de JH a partir da posição 6 a 20 com fusão dos últimos quatro aminoácidos de CDR2 com FR4. A JH que tenha menos não correspondências é selecionada. A sequência de CDR3 é examinada quanto a uma *J stump* trabalhando para trás na JH selecionada a partir da posição 9 na direção da primeira posição da JH selecionada comparando com CDR2 até a procura ser terminada por a) a extremidade de JH, b) a extremidade de CDR3, ou c) duas não correspondências consecutivas. Se uma das cadeias terminar e a última posição comparada for uma correspondência, então é incluída na *J stump*. Se não, não é. A Tabela 1070 mostra vários exemplos. A CDR é escrita acima, a JH está em baixo, e a *J stump* está sublinhada. Em 1070 A começamos em 9, V correspondências V,

e continuamos para a posição 6 com correspondências. A procura para em 4 devido à não correspondência dupla. GMDV vai para a pilha de *J stump* e GL segue para a pilha "Leadin". EM 1070 B, a procura termina com a extremidade de JH6. Os resíduos sublinhados vão para a pilha de *J stump* e EPIWG segue para a pilha *Leadin*. EM 1070 E, a procura termina devido à extremidade de JH4, mas o resíduo final testado (D na CDR vs. J em JH4) é uma não correspondência e logo a *J stump* é FDS e DSGVVAAAD segue para a pilha *Leadin*.

A Tabela 1015 mostra a distribuição de aminoácidos de CDR3 que não têm nenhuns segmentos D dos quais a *J stump* tenha sido removida. Note-se que a frequência de Tyr é muito mais baixa do que quando as CDR3s inteiras foram compiladas. Isto indica que Tyr vem para as CDR3 em larga medida através de incorporação de segmentos D e *J stumps*. Estas Tyr não são aleatoriamente inseridas, mas ocorrem numa sequência que foi selecionada através da evolução dos mamíferos. É uma característica da presente invenção que elevados níveis (mais do que 20%) de Tyr devam ser inseridos em bibliotecas através da incorporação de Ds e *J stumps* que contenham Tyr. Em posições *leadin* e de enchimento de DJ, Tyr é permitida, mas a não mais do que 20%.

Tabela 1078: Exemplos de atribuição de *J stumps*

```

A)
  6      GLGMDV
JH6      YYYYYGMDVWGQGTTVTVSS
        123456789

B)
  13     EPIWGYYYYGMDV
JH6      YYYYYGMDVWGQGTTVTVSS

C)
  9      DFFTSYFDY
JH4      -----YFDYWGQGITLVTVSS

D)
  12     DRGVSLLGAFDI
JH3      -----AFDIWGQGITMVTVSS

E)
  12     DSGVAAADFDS
JH4      -----YFDYWGQGITLVTVSS
        6789

```

A Tabela 1082 mostra a distribuição do uso de aminoácidos nas *J stumps* de cada JH. Uma vez que as JHs mais comuns são JH3, JH4, e JH6, estas são as JHs preferenciais para se construírem bibliotecas. A Tabela 1082 mostra que a maioria dos exemplos de JH2 retém a sequência tetrapeptídica AFDI em CDR3. Com JH4, uma maioria retém DY e uma grande fração retém a sequência FDY em CDR3. Com JH6, uma grande maioria retém a sequência DV, uma maioria retém a sequência MDV, e uma fração substancial retém a sequência GMDV. Uma fração não negligenciável retém a sequência YGMDV, YYGMDV, ou YYYGMDV.

Bibliotecas tais como 5.001 (Tabela 1097) estão incluídas nas bibliotecas da presente invenção. A biblioteca 5.001 contém CDR1-2 de LC e HC como descrito noutro lugar no presente pedido. A biblioteca contém uma VH de HC (tal como 3-23) seguida 6, 7, ou 8 aminoácidos permitindo [GSRDLY] em proporção mostrada na Tabela 1097. Na *J stump*, o aminoácido parente está presente 3, 4, 5, 6, 7, 8, 10 vezes tão

provavelmente como "outros" tipos de aminoácidos. Os "outros" tipos de aminoácidos compreendem Y, S, D, R, G. Assim sendo, em A6 permitimos 7/12 de A, mais 1/12 cada um de Y, S, D, R, e G. Em F7 permitimos 7/12 de F mais 1/12 cada um de Y, S, D, R, e G. Em D8 permitimos 7/11 de D mais 1/1 de Y, S, R, e G. Em 19 permitimos 7/12 de I mais 1/12 de Y, S, R, D, G. A sequência de aminoácidos parente poderia ser 5, 6, 7, 8, 10 vezes mais provável do que os outros tipos de aminoácidos.

A biblioteca 5.002 na Tabela 1097 está incluída nas bibliotecas da presente invenção. Esta biblioteca compreende CDR3 de comprimentos 13, 14, e 15 e nenhum segmento D. Existem 6, 7, ou 8 resíduos *leadin* permitindo G, S, R, D, L, ou Y nas razões 1:0,57:0,46:0,42:0,36:0,35 ou aproximação razoável a elas. A CDR3 é completada com uma porção de JH6: YYYGMDV. O ADN que codifica a sequência parente YYYGMDV é sintetizado com o aminoácido parente a 5, 6, 7, 8, ou 10 vezes mais provavelmente do que os outros.

A biblioteca 5.003 na Tabela 1097 está incluída na biblioteca da presente invenção. FR3 é seguida por 4, 5, ou 6 resíduos *leadin* permitindo G, S, R, D, L, Y na razão 1,0:0,57:0,46:0,42:0,36:0,35. De seguida vem o segmento D 3-3.2; o ADN que codifica esta região favorece o aminoácido parente por 5 vezes e permite como outros aminoácidos Y, G, D, R, S. Não existe nenhum enchimento de DJ e os quatro aminoácidos finais vêm da *J stump* de JH3. O ADN codificando a *J stump* é sintetizado com o aminoácido parente 5 vezes mais provavelmente do que os outros: YSGRD.

A biblioteca 5.004 na Tabela 1097 é uma parte da presente invenção. Existem 2, 3, ou 4 resíduos *leadin* permitindo

GSRDLY nas razões mostradas. O ADN codificando a sequência GYSSGWY é sintetizado tal que o aminoácido parente seja 6 X tão provável como os outros, dois resíduos de enchimento de DJ são permitidos com GSRDLY permitido nas razões 1,0:0,57:0,46:0,42:0,36:0,35. O ADN para codificar AFDI é sintetizado com o aminoácido parente 6 x tão provavelmente como os outros.

A biblioteca 5.005 é parte da presente invenção. A biblioteca 5.005 compreende membros com comprimentos de CDR3 11-14. Após FR3 existem 0, 1, ou 2 resíduos *leadin* permitindo GSRDLY nas razões mostradas seguidos por ADN que codifica a sequência parente GYSSGWY com variabilidade que permite YGSRD tal que o aminoácido parente seja 6 X mais provável como os outros tipos permitidos. Após a região D existem 0 ou 1 resíduo de enchimento de DJ permitindo GSRDLY nas razões mostradas. Finalmente existe JH3 com variabilidade na *J stump* (sequência: YFDY) que permite YGSRD com o aminoácido parente 6 X tão provável como os outros tipos permitidos.

A biblioteca 5.006 na Tabela 1097 é parte da presente invenção. A CDR3 pode ter comprimento 19-25. Existem zero a três resíduos *leadin* permitindo GSRDLY nas razões mostradas. Após o *leadin* está a região D 2-2.2. O ADN codificando 2-2.2 é sintetizado tal que o aminoácido parente seja 6 X tão provável como os outros (*viz.* YGSRD) exceto que os dois resíduos de Cys estão fixados. Após 2-2.2 estão zero a três resíduos de enchimento de DJ permitindo GSRDLY nas razões mostradas. O ADN que codifica os primeiros nove resíduos de JH6 permite o aminoácido parente mais YSGDR com o tipo parente sendo 6X mais provável do que os outros.

Tabela 1001: Utilização de JHs

JH	Número	%	1111111112 12395678901239567890
JH1	1356	6,15	---AEYFQHWGQGT L VTVSS
JH2	1720	7,80	---YWYFDLWGRGT L VTVSS
JH3	5601	25,39	-----AFDIWGQGT M VTVSS
JH4	7658	34,71	-----YFDYWGQGT L VTVSS
JH5	1062	4,81	----NWFDPWGQGT L VTVSS
JH6	4666	21,15	YYYYYGMDVWGQGT T VTVSS
Total	22063		

Tabela 1010: Utilização de Aminoácidos em CDR3 de HC

AA	Número	%	Rel para cima	Rel para baixo
Y	42863	15,47	35,87	1,00
G	37512	13,54	31,39	0,88
D	34051	12,29	28,49	0,79
S	23068	8,33	19,30	0,54
F	17813	6,43	14,91	0,42
A	15150	5,47	12,68	0,35
R	14090	5,09	11,79	0,33
V	13834	4,99	11,58	0,32
L	12351	4,46	10,34	0,29
I	10014	3,61	8,38	0,23
P	9514	3,43	7,96	0,22
W	9340	3,37	7,82	0,22
T	7544	2,72	6,31	0,18
M	6093	2,20	5,10	0,14
E	6042	2,18	5,06	0,14

AA	Número	%	Rel para cima	Rel para baixo
N	5901	2,13	4,94	0,14
H	4403	1,59	3,68	0,10
K	3147	1,14	2,63	0,07
Q	3097	1,12	2,59	0,07
C	1195	0,43	1,00	0,03
	277022			

Tabela 1015: Frequência de aminoácidos em *Leadin* de CDR3s não tendo regiões D

AA	Número	percentagem	rel para cima	rel para baixo
G	23134	18,24	46,45	1,000
S	13555	10,69	27,22	0,586
R	10562	8,33	21,21	0,457
D	9704	7,65	19,49	0,419
L	8255	6,51	16,58	0,357
Y	8099	6,39	16,26	0,350
A	7188	5,67	14,43	0,311
V	6599	5,20	13,25	0,285
P	5768	4,55	11,58	0,249
W	4804	3,79	9,65	0,208
T	4769	3,76	9,58	0,206
E	4497	3,55	9,03	0,194
N	3733	2,94	7,50	0,161
F	3616	2,85	7,26	0,156
I	3464	2,73	6,96	0,150
H	2787	2,20	5,60	0,120
K	2460	1,94	4,94	0,106
Q	2124	1,67	4,27	0,092

Tabela 1015: Frequência de aminoácidos em <i>Leadin</i> de CDR3s não tendo regiões D				
AA	Número	percentagem	rel para cima	rel para baixo
M	1225	0,97	2,46	0,053
C	498	0,39	1,00	0,022
	126841			

Tabela 1020: Comprimentos de CDR3s de HC		
Comprimento	Número	%
1	0	0,00
2	6	0,03
3	36	0,16
4	153	0,69
5	121	0,55
6	669	3,03
7	756	3,43
8	1066	4,83
9	2227	10,09
10	2701	12,24
11	2240	10,15
12	2071	9,39
13	2006	9,09
14	1594	7,22
15	1396	6,33
16	1254	5,68
17	1102	4,99
18	783	3,55
19	588	2,67
20	474	2,15

Tabela 1020: Comprimentos de CDR3s de HC		
Comprimento	Número	%
21	285	1,29
22	237	1,07
23	133	0,60
24	81	0,37
25	32	0,15
26	25	0,11
27	11	0,05
28	6	0,03
29	2	0,01
30	3	0,01
31	2	0,01
32	1	0,00
33	1	0,00
34	0	0,00
35	0	0,00
36	1	0,00
	22063	

Tabela 1030: Utilização de segmentos D.		
Id	Número	Sequência
1-1.1	29	GTTGT
1-1.2	6	VQLER
1-1.3	151	YNWND
1-7.1	34	GITGT
1-7.2	0	V*LEL
1-7.3	65	YNWNY
1-20.1	0	GITGT

Tabela 1030: Utilização de segmentos D.

Id	Número	Sequência
1-20.2	0	V*LER
1-20.3	0	YNWND
1-26.1	48	GIVGAT
1-26.2	3	V*WELL
1-26.3	220	YSGSYY
2-2.1	0	RIL**YQLLY
2-2.2	102	GYCSSTSCYT
2-2.3	37	DIVVVPAAI
2-8.1	0	RILY@WCMLY
2-8.2	23	GYCTNGVCYT
2-8.3	1	DIVLMVYAI
2-15.1	0	RIL*WW*LLL
2-15.2	139	GYCSGGSCYS
2-15.3	12	DIVVVVAAT
2-21.1	0	SILWWSLLF
2-21.2	24	AYCGGDCYS
2-21.3	6	HIVVVTAI
3-3.1	28	VLRFLEWLLY
3-3.2	743	YYDFWSGYYT
3-3.3	15	ITIFGVVII
3-9.1	41	VLRYFDWLL@
3-9.2	8	YYDILTGYYN
3-9.3	0	ITIF*LVII
3-10.1	26	VLLWFGELL@
3-10.2	136	YYYGSGSYYN
3-10.2	32	ITMVRGVII
3-16.1	0	VLSLRLGELSLY

Tabela 1030: Utilização de segmentos D.		
Id	Número	Sequência
3-16.2	109	YYDYVWGSYRYT
3-16.2	8	IMITFGGVIVI
3-22.1	0	VLL***WLLL
3-22.2	617	YYDSSGYYY
3-22.3	2	ITMIVVVIT
4-4.1	0	SLQ@L
4-4.2	75	DYSNY
4-4.3	165	TTVT
4-11.1	0	SLQ@L
4-11.2	0	DYSNY
4-11.3	0	TTVT
4-17.1	0	SLR@L
4-17.2	392	DYGDY
4-17.3	0	TTVT
4-23.1	0	SLRW@L
4-23.2	60	DYGGNS
4-23.3	16	TTVVT
5-5.1	25	VDTAMV
5-5.2	29	WIQLWL
5-5.3	292	GYSYGY
5-12.1	13	VDIVATI
5-12.2	0	WI*WLRL
5-12.3	200	GYSGYDY
5-18.1	0	VDTAMV
5-18.2	0	WIQLWL
5-18.3	0	GYSYGY
5-24.1	9	VEMATl

Tabela 1030: Utilização de segmentos D.		
Id	Número	Sequência
5-24.2	21	*RWLQL
5-24.3	44	RDGYNY
6-6.1	87	EYSSSS
6-6.2	122	SIAAR
6-6.3	1	V*QLV
6-13.1	399	GYSSSWY
6-13.2	170	GIAAAG
6-13.3	0	V*QQLV
6-19.1	441	GYSSGWY
6-19.2	104	GIAVAG
6-19.3	3	V*QWLV
7-27.1	257	LTG
7-27.2	0	@LG
7-27.3	64	NWG
nenhuma	16409	

Tabela 1040: JH vs. Comprimento						
Comprimento	JH1	JH2	JH3	JH4	JH5	JH6
1	0	0	0	0	0	0
2	1	4	0	1	0	0
3	20	2	3	9	0	2
4	75	3	10	45	8	12
5	47	6	10	38	8	12
6	273	14	43	280	26	33
7	88	27	194	337	30	80
8	134	43	243	503	41	102
9	121	70	855	886	61	234

Tabela 1040: JH vs. Comprimento						
Comprimento	JH1	JH2	JH3	JH4	JH5	JH6
10	116	693	623	979	68	222
11	105	81	675	1003	84	292
12	107	84	552	905	121	302
13	87	274	538	672	113	322
14	48	81	480	532	105	348
15	50	83	372	421	80	390
16	28	54	316	322	87	447
17	27	49	239	334	69	384
18	11	64	174	140	49	345
19	8	28	104	99	41	308
20	4	23	59	56	20	312
21	0	13	40	30	24	178
22	3	14	31	30	13	146
23	1	3	22	12	7	88
24	0	5	9	12	4	51
25	1	0	1	3	1	26
26	0	0	5	5	0	15
27	0	1	2	1	1	6
28	1	0	0	2	0	3
29	0	0	0	0	0	2
30	0	0	0	0	0	3
31	0	0	1	0	0	1
32	0	1	0	0	0	0
33	0	0	0	1	0	0
34	0	0	0	0	0	0
35	0	0	0	0	0	0
36	0	0	0	0	1	0

Tabela 1050: Utilização de aminoácidos em *Leadin* com nenhum segmento D

AA	Número	%	Rel para cima	Rel para baixo
G	23134	18,24	46,45	1,00
S	13555	10,69	27,22	0,59
R	10562	8,33	21,21	0,46
D	9704	7,65	19,49	0,42
L	8255	6,51	16,58	0,36
Y	8099	6,39	16,26	0,35
A	7188	5,67	14,43	0,31
V	6599	5,20	13,25	0,29
P	5768	4,55	11,58	0,25
W	4804	3,79	9,65	0,21
T	4769	3,76	9,58	0,21
E	4497	3,55	9,03	0,19
N	3733	2,94	7,50	0,16
F	3616	2,85	7,26	0,16
I	3464	2,73	6,96	0,15
H	2787	2,20	5,60	0,12
K	2960	1,94	4,94	0,11
Q	2124	1,67	4,27	0,09
M	1225	0,97	2,46	0,05
C	498	0,39	1,00	0,02
	126841			

Tabela 1080: Dipéptidos em CDR3s de HC, parte 1

YY	13565	FG	1073	PL	591	TV	397
FD	11637	RS	1072	TT	589	TP	390

Tabela 1080: Dipéptidos em CDR3s de HC, parte 1							
DY	8337	SW	1014	ID	588	NA	389
SG	5979	DW	1003	DD	583	NS	388
GY	5805	LR	990	AS	570	ER	387
YG	5461	DG	989	KG	566	HG	386
DI	5448	PG	976	VD	556	VW	381
AF	4975	LL	974	VP	551	QL	378
DV	4968	AY	962	LT	540	RI	374
GG	4575	DR	923	LF	539	WN	365
SS	4491	VR	882	VL	539	YT	365
MD	4436	YM	877	FY	534	CS	360
GS	4047	AR	872	PD	533	DH	359
GM	3501	VV	869	RV	531	EA	359
YF	3438	YR	865	RF	525	WD	353
YD	3430	VA	857	AL	521	ES	350
RG	3118	RA	844	PS	510	FR	349
SY	2770	SP	820	EY	508	YC	343
GA	2611	GN	812	LW	508	PT	337
YS	2576	HY	809	PA	505	TL	326
DA	2285	SD	805	LP	500	KR	325
DS	2087	GI	804	VS	497	VF	324
WY	2079	NW	785	IR	493	MG	314
GD	2017	LS	760	YV	493	PN	313
GR	1985	LY	757	VY	478	RE	312
GL	1800	TY	749	IG	476	IV	311
DL	1777	PR	742	VT	475	KS	310
DF	1763	GE	737	TR	472	SC	310
GW	1725	SA	736	DN	471	FL	309
WS	1675	SF	728	SI	469	FF	306

Tabela 1080: Dipéptidos em CDR3s de HC, parte 1							
AA	1671	PF	725	AD	462	CY	303
LD	1651	ND	693	LA	459	SH	302
EG	1610	ST	684	PP	451	LK	300
AG	1606	GH	683	RT	451	IT	298
RY	1558	YP	676	DT	448	LE	298
DP	1547	WL	675	RW	447	FS	296
GV	1500	SN	667	GQ	446	ED	294
RR	1498	TS	652	QG	446	RK	294
LG	1387	RD	648	TD	446	HF	292
GF	1386	YA	648	TA	437	VI	290
VG	1366	SL	644	TF	426	RH	287
GP	1339	RP	643	GK	422	MV	285
WF	1282	YL	638	YW	421	KY	284
FW	1277	IA	634	HD	420	AI	282
NY	1271	RL	627	IL	417	HS	281
PY	1209	EL	622	LV	406	YH	281
GT	1194	YN	607	IS	402	LN	278
WG	1177	AV	605	NG	398	PV	276
SR	1162	AP	600	RN	398	QY	276
TG	1142	AT	592	SV	397	WA	271

Tabela 1080: Dipéptidos em CDR3s de HC, parte 2						
267	KD	176	II	102	NQ	53
264	SK	176	HI	101	CF	51
257	YK	176	KP	101	MP	50
255	EF	174	MY	100	CP	49
253	FN	174	RM	99	RC	47
250	HN	171	AQ	98	HE	46

Tabela 1080: Dipéptidos em CDR3s de HC, parte 2						
250	FH	165	EQ	96	VC	46
248	YQ	165	QT	96	QI	45
242	KN	164	LM	95	MN	44
242	MA	163	HV	94	MF	43
240	NN	160	IK	93	HQ	41
239	KA	159	PM	93	CD	38
239	SQ	157	QN	93	CL	38
238	PE	156	CG	91	NC	38
229	WV	154	QF	91	HM	37
228	EI	153	FI	90	FM	36
227	TH	153	HW	90	ME	36
225	FV	152	WH	90	MK	35
221	AK	151	QV	89	QM	35
220	TK	151	WI	89	NM	34
220	WT	151	KH	88	KM	32
219	PK	150	MI	88	TC	31
217	KK	148	MS	87	CR	29
214	IW	145	TQ	86	CV	25
212	VH	145	NV	85	HC	25
210	VE	141	EM	84	WM	25
210	EE	138	HK	84	AC	24
208	DE	136	IN	83	FC	24
208	KL	136	NH	82	CA	23
201	PQ	136	NI	82	CH	21
201	QP	135	HT	81	CN	21
201	SM	134	WK	79	MW	21
201	QD	133	KF	77	PC	19
199	QS	131	VM	73	LC	17

Tabela 1080: Dipéptidos em CDR3s de HC, parte 2						
198	VQ	130	MT	71	IC	16
198	QQ	129	IH	69	MM	16
196	WW	129	EH	68	MH	15
196	NT	128	IE	67	WC	15
193	DC	118	QK	65	EC	12
192	KT	118	WQ	65	CK	10
191	QA	118	GC	64	CW	10
191	NK	113	KE	61	MQ	10
186	KW	112	KI	61	CI	9
185	EK	109	CT	58	CC	8
182	FT	108	FK	58	CM	8
180	KV	108	IM	57	CQ	6
180	MR	105	KQ	57	QC	6
178	TE	104	ML	55	CE	5
177	HH	103	QE	55	KC	5
177	IQ	103	NE	53	MC	3

Tabela 1060a Comprimentos de CDR3s de HC vs. que segmentos D ocorrem (se algum) para comprimentos 3-17

Nome	Sequência	Comprimento														
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1-1.1	GTTGT	0	0	0	0	0	0	2	2	3	6	2	3	3	4	0
1-1.2	VQLER	0	0	0	0	0	0	0	0	1	0	1	0	1	2	1
1-1.3	YNWND	0	0	0	0	0	2	6	14	16	19	16	14	17	16	9
1-7.1	GITGT	0	0	0	0	0	0	0	1	2	7	6	4	4	0	4
1-7.3	YNWNY	0	0	0	0	1	0	2	5	7	8	8	6	5	9	4
1-26.1	GIVGAT	0	0	0	1	0	0	2	4	10	4	6	9	3	2	2
1-26.2	V*WELL	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1

Tabela 1060a Comprimentos de CDR3s de HC vs. que segmentos D ocorrem
(se algum) para comprimentos 3-17

Nome	Sequência	Comprimento														
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1-26.3	YSGSY	0	0	0	0	1	0	2	10	14	24	24	27	21	26	13
2-2.2	GYCSSTSCYT	0	0	0	0	0	0	0	0	0	2	2	9	15	15	11
2-2.3	DIVVVPAAI	0	0	0	0	0	0	0	0	0	0	1	3	2	5	5
2-8.2	GYCTNGVCYT	0	0	0	0	0	0	0	0	0	1	0	2	4	3	4
2-8.3	DIVLMVYAI	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2-15.2	GYCSGGSCYS	0	0	0	0	0	0	0	0	1	3	5	12	10	25	22
2-15.3	DIVVVAAT	0	0	0	0	0	0	0	0	0	0	1	1	2	0	1
2-21.2	AYCGGDCYS	0	0	0	0	0	0	0	0	0	1	1	3	5	2	5
2-21.3	HIVVVTAI	0	0	0	0	0	0	0	0	0	1	2	0	2	0	1
3-3.1	VLRFLEWLLY	0	0	0	0	0	0	0	0	0	0	1	2	2	2	3
3-3.2	YYDFWSGYT	0	0	0	0	0	0	0	1	5	8	22	38	44	72	69
3-3.3	ITIFGVVII	0	0	0	0	0	0	0	0	0	0	0	1	0	2	5
3-9.1	VLRYFDWLL@	0	0	0	0	0	0	0	0	1	0	0	4	5	5	5
3-9.2	YYDILTGYN	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
3-10.1	VLLWFGELL@	0	0	0	0	0	0	0	0	2	1	2	4	2	3	5
3-10.2	YYYGSGSYN	0	0	0	0	0	0	0	2	4	7	10	13	15	18	14
3-10.2	ITMVRGVII	0	0	0	0	0	0	0	0	0	3	1	2	7	5	2
3-16.2	YYDYVWGSYRYT	0	0	0	0	0	0	0	0	0	0	0	1	7	7	7
3-16.2	IMITFGGVIVI	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2
3-22.2	YYDSSGYYY	0	0	0	0	0	0	0	0	6	30	45	56	59	108	101
3-22.3	ITMIVVVIT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4-4.2	DYSNY	0	0	0	0	0	3	3	4	6	14	4	7	4	10	6
4-4.3	TTVT	0	0	0	0	2	4	11	19	23	19	25	19	10	11	9
4-17.2	DYGDY	0	0	2	6	12	8	38	40	48	47	50	40	29	21	10
4-23.2	DYGGNS	0	0	0	0	0	2	7	4	5	17	4	8	5	1	1

Tabela 1060a Comprimentos de CDR3s de HC vs. que segmentos D ocorrem
(se algum) para comprimentos 3-17

		Comprimento														
Nome	Sequência	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
4-23.3	TTVVT	0	0	0	0	0	0	2	0	1	1	2	1	2	0	0
5-5.1	VDTAMV	0	0	0	0	0	0	0	1	4	8	1	3	2	0	1
5-5.2	WIQLWL	0	0	0	0	0	0	0	3	2	0	3	1	4	3	3
5-5.3	GYSYGY	0	0	0	0	1	6	9	20	43	29	27	22	32	26	27
5-12.1	VDIVATI	0	0	0	0	0	0	0	2	0	1	2	4	0	2	1
5-12.3	GYSGYDY	0	0	0	0	4	10	13	15	19	15	22	27	16	15	9
5-24.1	VEMATI	0	0	0	0	0	0	1	1	0	0	6	0	0	0	0
5-24.2	*RWLQL	0	0	0	0	1	0	3	1	3	2	2	1	2	22	2
5-24.3	RDGYNY	0	0	0	0	0	0	0	1	8	12	6	7	3	2	1
6-6.1	EYSSSS	0	0	0	0	0		0	9	7	16	19	13	2	4	2
6-6.2	SIAAR	0	0	0	1	1	0	17	8	7	13	17		16	16	7
6-6.3	V*QLV	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6-13.1	GYSSSWY	0	0	0	0	0	1	6	11	35	40	56	44	50	42	35
6-13.2	GIAAAG	0	0	0	0	1	2	18	14	15	20	20	15	16	14	11
6-19.1	GYSSGWY	0	0	0	0	1	1	4	27	57	58	48	52	45	35	30
6-19.2	GIAVAG	0	0	0	0	1	1	0	7	8	20	8	13	16	8	10
6-19.3	V*QWLV	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
7-27.1	LTG	0	0	1	0	2	8	12	7	14	11	17	17	24	24	31
7-27.3	NWG	0	0	0	1	2	11	6	5	10	6	7	5	7	1	0
nenhum		36	153	118	660	726	1007	2063	2463	1851	1596	1502	1075	874	681	609

Tabela 1060b Comprimentos de CDR3s de HC vs. que segmentos D
ocorrem (se algum) para comprimentos 18-32

		Comprimento															
Nome	Sequência	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	

Tabela 1060b Comprimentos de CDR3s de HC vs. que segmentos D ocorrem (se algum) para comprimentos 18-32																
		Comprimento														
Nome	Sequência	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1-1.1	GTTGT	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0
1-1.2	VQLER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-1.3	YNWND	6		3	4	1	1	1	0	0	1	0	0	0	0	0
1-7.1	GITGT	2	2	1	0	0	0	1	0	0	0	0	0	0	0	0
1-7.3	YNWNY	5	2	1	0	0	0	0	0	0	0	0	1	0	0	0
1-26.1	GIVGAT	0	0	3	1	1	0	0	0	0	0	0	0	0	0	0
1-26.2	V*WELL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1-26.3	YSGSYY	14	11	7	8	6	3	3	0	3	1	0	0	0	0	0
2-2.2	GYCSSTSCYT	11	7	2	10	11	4	2	0	0	0	1	0	0	0	0
2-2.3	DIVVVPAAI	2	6	3	4	3	2	1	0	0	0	0	0	0	0	0
2-8.2	GYCTNGVCYT	3	0	1	1	3	1	0	0	0	0	0	0	0	0	0
2-8.3	DIVLMVYAI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2-15.2	GYCSGGSCYS	20	10	7	4	8	9	3	0	0	0	0	0	0	0	0
2-15.3	DIVVVVAAT	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
2-21.2	AYCGGDCYS	1	3	1	1	1	0	0	0	0	0	0	0	0	0	0
2-21.3	HIVVVTAI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3-3.1	VLRFLWLLY	6	2	4	3	2	0	0	0	1	0	0	0	0	0	0
3-3.2	YYDFWSGYT	82	97	104	67	61	32	23	7	3	4	0	0	2	1	0
3-3.3	ITIFGVVII	3	2	0	0	2	0	0	0	0	0	0	0	0	0	0
3-9.1	VLR YFDWLL@	7	2	2	6	1	2	1	0	0	0	0	0	0	0	0
3-9.2	YYDILTGYYN	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
3-10.1	VLLWFGELL@	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0
3-10.2	YYYGSGSYYN	15	10	8	7	6	3	2	1	1	0	0	0	0	0	0
3-10.2	ITMVRGVII	3	3	6	0	0	0	0	0	0	0	0	0	0	0	0
3-16.2	YYDYVWGSYRYT	11	11	14	10	18	13	5	2	2	0	0	0	1	0	0
3-16.2	IMITFGGVIVI	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0
3-22.2	YYDSSGYYY	77	54	28	22	18	8	1	2	1	1	0	0	0	0	0

Tabela 1060b Comprimentos de CDR3s de HC vs. que segmentos D ocorrem (se algum) para comprimentos 18-32																
		Comprimento														
Nome	Sequência	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
3-22.3	ITMIVVVIT	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4-4.2	DYSNY	7	2	2	1	1	0	1	0	0	0	0	0	0	0	0
4-4.3	TTVT	4	2	2	1	3	0	0	0	0	0	1	0	0	0	0
4-17.2	DYGDY	17	7	8	3	2	1	1	0	1	0	1	0	0	0	0
4-23.2	DYGGNS	3	1	0	2	0	0	0	0	0	0	0	0	0	0	0
4-23.3	TTVVT	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0
5-5.1	VDTAMV	0	1	1	2	1	0	0	0	0	0	0	0	0	0	0
5-5.2	WIQLWL	3	3	1	0	1	0	0	0	0	1	0	0	0	1	0
5-5.3	GYSYGY	13	18	7	2	2	6	0	0	2	0	0	0	0	0	0
5-12.1	VDIVATI	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5-12.3	GYSGYDY	11	10	6	6	0	2	0	0	0	0	0	0	0	0	0
5-24.1	VEMATl	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5-24.2	*RWLQL	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5-24.3	RDGYNY	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
6-6.1	EYSSSS	9	3	1	1	0	0	0	1	0	0	0	0	0	0	0
6-6.2	SIAAR	2	3	2	6	0	0	0	0	0	0	0	0	0	0	0
6-6.3	V*QLV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-13.1	GYSSSWY	24	22	17	4	5	1	3	2	0	0	1	0	0	0	0
6-13.2	GIAAAG	10	4	2	5	2	1	0	0	0	0	0	0	0	0	0
6-19.1	GYSSGWY	23	25	13	7	7	3	0	2	1	0	1	0	0	0	0
6-19.2	GIAVAG	6	2	2	0	0	0	2	0	0	0	0	0	0	0	0
6-19.3	V*QWLV	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7-27.1	LTG	29	23	12	9	7	4	3	1	0	1	0	0	0	0	0
7-27.3	NWG	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0
nenhum		339	217	198	85	60	36	27	13	9	2	1	1	0	0	1

Tabela 1082: Inventário de *J* stumps

Tabela 1082: Inventário de <i>J stumps</i>																					
JH1	---AEYFQHWGQGTLVTVSS 6,15%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
4	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	605
5	1	0	1	64	0	6	0	1	0	0	0	0	0	1	2	1	1	0	0	0	568
6	0	0	0	0	0	1	1	0	0	3	0	2	1	0	1	1	0	0	0	211	425
7	1	0	1	0	363	3	1	0	0	8	0	0	1	0	1	6	0	3	0	6	252
8	8	0	59	23	4	17	11	3	5	19	0	8	3	221	11	6	8	5	0	4	231
9	2	1	13	2	13	3	447	19	2	6	3	20	1	2	0	20	3	5	4	9	71
JH2	---YWYFDLWGRGTLVTVSS 7,80%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	159	1519
5	0	0	2	0	2	11	0	0	0	1	0	1	0	0	7	1	1	1	929	21	701
6	3	2	9	0	40	5	7	4	1	9	0	7	1	0	1	11	2	0	1	1083	492
7	1	6	1	0	1209	2	6	23	0	89	30	1	2	1	1	12	0	11	1	42	240
8	31	2	1241	90	4	38	30	3	1	0	1	29	3	4	2	2	3	19	0	15	160
9	3	1	9	3	34	2	26	17	5	1064	36	17	30	38	33	20	8	83	1	177	71
JH3	-----AFDIWGQGTMTVTVSS 25,4%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
6	4374	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1074
7	1	4	0	1	4612	2	0	8	0	56	3	2	1	1	0	29	2	15	12	74	625
8	23	0	4765	51	0	28	14	0	3	0	0	15	2	1	1	2	2	6	0	4	531
9	7	5	5	0	73	2	1	4439	4	64	64	43	2	1	11	54	49	113	2	18	491
JH4	-----YFDYWGQGTMLVTVSS 34,7%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1975	5683
7	9	1	10	7	3950	31	6	26	0	109	4	5	24	4	5	35	7	28	16	59	3322
8	26	0	5991	32	5	91	19	2	1	6	1	21	7	0	8	13	3	6	2	14	1410

Tabela 1082: Inventário de <i>J stumps</i>																					
9	5	18	17	0	119	2	0	0	14	0	2	64	15	16	10	216	11	3	6	6317	823
JH5	----NWFDPWGQGTTLVTVSS 4,8%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
5	0	0	0	0	0	0	0	0	0	0	0	274	0	0	0	0	0	0	0	0	764
6	2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	644	0	390
7	1	1	0	0	768	2	1	1	0	13	0	0	0	0	0	1	0	2	0	4	244
8	2	0	810	3	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	1	218
9	3	0	3	0	3	0	3	0	0	4	0	4	814	1	0	14	2	0	0	0	187
JH6	YYYYYGMDVWGQGTTVTVSS 21,1%																				
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	-
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	690	3967
2	4	0	15	0	16	31	12	2	1	4	0	8	7	4	14	18	7	4	1	1694	2815
3	3	0	19	3	14	23	16	1	5	9	0	16	9	5	12	20	4	1	4	2440	2053
4	4	0	14	2	31	16	25	3	2	10	0	13	6	5	15	35	5	5	4	2815	1647
5	2	1	9	1	23	19	21	1	1	3	1	15	3	1	7	26	3	0	1	3350	1169
6	69	5	14	1	4	3057	8	1	0	0	0	8	7	5	4	15	3	7	0	657	792
7	3	0	1	0	4	5	0	18	2	108	3866	0	2	2	3	1	7	18	3	1	613
8	7	0	4064	5	1	17	4	1	0	2	0	11	2	0	0	3	2	3	2	6	527
9	9	0	1	0	6	1	1	19	0	7	2	1	1	0	1	4	0	4092	0	1	511

Tabela 1097: Bibliotecas de CDR3 de HC

Biblioteca	Comprimento de CDR3	Lead in	Sequência da região D	Sequência de enchimento de DJ	Sequência de <i>J stump</i>	Sequência de FR4
5.001	10,11,12	6, 7, ou 8 X X = (1,0 de G, 0,57	nenhuma	nenhuma	AFDI (diversidade no texto)	WGQGTMTVTVSS (JH4)

Tabela 1097: Bibliotecas de CDR3 de HC						
Biblioteca	Comprimento de CDR3	Lead in	Sequência da região D	Sequência de enchimento de DJ	Sequência de <i>J stump</i>	Sequência de FR4
		de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)				
5.002	13, 14, 15	6,7,8 X X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	nenhuma	nenhuma	YYYGMDV (diversidade no texto)	WGQGTTVTVSS (jh6)
5.003	18, 19, 20	4,5,6 X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	YYDFWSGYT (3-3.2)	nenhuma	YFDY	WGQGTTLVTVSS (JH3)
5.004	15, 16, 17	2,3,4 X = (1,0 de G,	GYSSGWY (6- 19.1)	2X,X = (1,0 de G, 0,57 de S,	AFDI	WGQGTMTVTVSS (JH4)

Tabela 1097: Bibliotecas de CDR3 de HC						
Biblioteca	Comprimento de CDR3	Leadin	Sequência da região D	Sequência de enchimento de DJ	Sequência de <i>J stump</i>	Sequência de FR4
		0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)		0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)		
5.005	11-14	0, 1, 2 X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	GYSSGWY (6-19.1)	0, 1 X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	YFDY	WGQGTTLVTVSS (JH3)
5.006	19-25	0,1,2,3 X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	GYCSGGSCYS (2-2.2) (resíduos de Cys constantes)	0, 1, 2, 3 X = (1,0 de G, 0,57 de S, 0,46 de R, 0,42 de D, 0,36 de L, 0,35 de Y)	YYYYYGMDV (AA parente 8X outros)	WGQGTTLVTVSS (jh6)

REIVINDICAÇÕES

1. Uma biblioteca de vetores ou pacotes genéticos que exibem, exibem e expressam, ou compreendem um membro de uma família diversa de péptidos, polipéptidos e proteínas relacionados com anticorpos humanos e exibem coletivamente, exibem e expressam, ou compreendem pelo menos uma porção da diversidade da família dos anticorpos, em que os vetores ou pacotes genéticos compreendem sequências de ADN variegadas que codificam:

(i) uma cadeia pesada, compreendendo:

(a) uma CDR1 tendo a seguinte sequência de aminoácidos: $X_{31}YX_{33}MX_{35}$, em que X_{31} , X_{33} e X_{35} são qualquer aminoácido exceto Cys ou Met;

(b) uma CDR2 tendo a seguinte sequência de aminoácidos: $X_{50}IX_{52}X_{52a}SGGX_{56}TX_{58}YADSVKKG$, em que X_{50} , X_{52} , X_{56} e X_{58} são qualquer aminoácido exceto Cys ou Met e X_{52a} é selecionado do grupo consistindo em Gly, Ser, Pro ou Tyr; e

(c) uma CDR3;

e

(ii) uma cadeia leve.

2. A biblioteca da Reivindicação 1, em que 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, ou 35 das posições na CDR3 de HC da família estão variadas entre resíduos de Tyr, Gly, Asp, Ser, e Arg, e em que cerca de 50% de todas as posições da CDR3 de HC são Tyr, Gly, Asp, Ser, ou Arg.

3. A biblioteca da Reivindicação 1, em que 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, ou 35 das posições na CDR3 de HC da família estão variadas entre resíduos de Tyr, Gly, Asp, Ser, e Arg, e em que um resíduo de Arg está presente numa região de enchimento entre V e D, entre D e J, ou entre V e J.
4. A biblioteca da reivindicação 1, em que a CDR3 de cadeia pesada (HC) consiste em 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, ou 35 resíduos, cada um dos quais varia independentemente entre resíduos de Tyr, Gly, Asp, Ser, e Arg, em que uma região de *framework* (FR) 3 está localizada no terminal amino da CDR3, e uma porção de FR4 de uma região JH está localizada no terminal carboxi da CDR3; e em que a FR3 é uma FR3 de VH 3-23.
5. A biblioteca da Reivindicação 1, em que mais do que cerca de 20% da sequência de CDR3 de HC consiste em resíduos de Tyr, e em que regiões D e *J stumps* da CDR3 de HC consistem em mais do que cerca de 20% de resíduos de Tyr.
6. A biblioteca da Reivindicação 1, em que menos do que cerca de 20% da sequência de CDR3 de HC consiste em resíduos de Tyr, e em que posições *leadin* ou de enchimento de DJ da CDR3 de HC consistem em menos do que cerca de 20% de resíduos de Tyr.

7. A biblioteca da Reivindicação 1, em que a CDR3 de HC tem três aminoácidos de comprimento e: o primeiro resíduo de CDR3 de HC varia entre F, S, Y, D, e R numa razão 3:1:1:1:1; o segundo resíduo de CDR3 de HC varia entre Q, E, R, S, Y, e L numa razão 3:1:1:1:1:1; o terceiro resíduo de CDR3 de HC varia entre H, D, R, S, Y, e L numa razão 3:1:1:1:1:1, em que uma FR3 está localizada no terminal amino da CDR3, e o último resíduo da FR3 varia entre K e R numa razão 3:1.
8. A biblioteca da Reivindicação 1, em que a CDR3 de HC tem quatro aminoácidos De comprimento e: o primeiro resíduo de CDR3 de HC varia entre L, S, Y, D, e R numa razão 4:1:1:1:1; o segundo resíduo de CDR3 de HC varia entre L, S, Y, D, e R numa razão 4:1:1:1:1:1; o terceiro resíduo de CDR3 de HC varia entre F, S, Y, D, e R Numa razão 4:1:1:1:1:1; e o quarto resíduo de CDR3 de HC varia entre F, S, Y, D, e R numa razão 4:1:1:1:1:1, e em que uma FR3 está localizada no terminal amino da CDR3, e o último resíduo da FR3 varia entre K e R numa razão 4:1.
9. A biblioteca da Reivindicação 1, em que a CDR3 de HC tem dezasseis aminoácidos de comprimento e: o primeiro resíduo de CDR3 de HC varia entre Y, S, R, D, e L numa razão 3:1:1:1:1; o segundo resíduo de CDR3 de HC varia entre Y, S, R, D, e L numa razão 3:1:1:1:1; o terceiro resíduo de CDR3 de HC varia entre Y, S, R, D, e L numa razão 3:1:1:1:1; o quarto resíduo de CDR3 de HC varia entre D, Y, S, R, e L numa razão 3:1:1:1:1; o quinto resíduo de CDR3 de HC varia entre S, Y, R, D, e L numa razão 3:1:1:1:1; o sexto resíduo de CDR3 de HC varia entre S, Y, R, D, e L numa razão 3:1:1:1:1; o sétimo

resíduo de CDR3 de HC varia entre G, A, S, Y, R, D, e L numa razão 3:1:1:1:1; o oitavo entre Y, S, R, D, e L numa razão 3:1:1:1:1; o nono resíduo de CDR3 de HC varia entre Y, S, R, D, e L numa razão 3:1:1:1:1; o décimo resíduo de CDR3 de HC varia entre Y, S, R, D, e L numa razão 3:1:1:1:1; o décimo primeiro resíduo de CDR3 de HC varia entre A, S, Y, R, e D numa razão 3:1:1:1:1; o décimo segundo resíduo de CDR3 de HC varia entre E, R, S, Y, e L numa razão 3:1:1:1:1; o décimo terceiro resíduo de CDR3 de HC varia entre Y, S, R, D, and L numa razão 3:1:1:1:1; o décimo quarto resíduo de CDR3 de HC varia entre F, Y, S, R, e D numa razão 3:1:1:1:1; o décimo quinto resíduo de CDR3 de HC varia entre Q, E, R, S, e Y numa razão 3:1:1:1:1; e o décimo sexto resíduo de CDR3 de HC varia entre H, E, R, S, Y, e L numa razão 3:1:1:1:1, e em que uma FR3 está localizada no terminal amino da CDR3, e o último resíduo da FR3 varia entre K e R numa razão 3:1.

10. A biblioteca da Reivindicação 1, em que a CDR3 de HC tem dezasseis aminoácidos de comprimento e: o primeiro resíduo de CDR3 de HC varia entre G, S, Y, D, e L numa razão 3:1:1:1:1; o segundo resíduo de CDR3 de HC varia entre Y, S, D, R, e L numa razão 3:1:1:1:1; o terceiro resíduo de CDR3 de HC é C; o quarto resíduo de CDR3 de HC varia entre S, Y, R, D, e L numa razão 3:1:1:1:1; o quinto resíduo de CDR3 de HC varia entre S, Y, R, D, e L numa razão 3:1:1:1:1; o sexto resíduo de CDR3 de HC varia entre T, Y, R, D, e L numa razão 3:1:1:1:1; o sétimo resíduo de CDR3 de HC varia entre S, Y, R, D, e L numa razão 3:1:1:1:1; o oitavo resíduo de CDR3 de HC varia é C; o nono resíduo de CDR3 de HC varia entre Y, S, D, R, e L numa razão 3:1:1:1:1; o décimo resíduo de

CDR3 de HC varia entre T, Y, R, D, e L numa razão 3:1:1:1:1; o décimo primeiro resíduo de CDR3 de HC varia entre A, S, Y, R, e D numa razão 3:1:1:1:1; o décimo segundo resíduo de CDR3 de HC varia entre E, R, S, Y, e L numa razão 3:1:1:1:1; o décimo terceiro resíduo de CDR3 de HC varia entre Y, S, D, R, and L numa razão 3:1:1:1:1; o décimo quarto resíduo de CDR3 de HC varia entre F, Y, S, R, e D numa razão 3:1:1:1:1; o décimo quinto resíduo de CDR3 de HC varia entre Q, E, R, S, Y, e L numa razão 3:1:1:1:1; e o décimo sexto resíduo de CDR3 de HC varia entre H, D, S, Y, e L numa razão 3:1:1:1:1, e em que uma FR3 está localizada no terminal amino da CDR3, e o último resíduo da FR3 varia entre K e R numa razão 3:1.

11. A biblioteca da reivindicação 1, em que a cadeia leve é uma cadeia leve (LC) A27, em que CDR1 da LC é diversificada, em que a posição 27 é cerca de 55% de Q, e cerca de 9% cada um de E, R, Y, S, e L; a posição 28 é cerca de 46% de S e cerca de 9% cada um de N, T, Y, E, R, e L; a posição 30 é cerca de 55% de S e cerca de 9% cada um de D, N, R, T, e Y; a posição 30a é cerca de 46% de S e cerca de 9% cada um de G, N, R, T, Y, e D, e cerca de 8% com nenhum aminoácido; a posição 31 é cerca de 44% de S e cerca de 8% cada um de D, F, G, N, R, T, e Y; a posição 32 é cerca de 44% de Y e cerca de 7% cada um de F, D, L, N, Q, R, S, e Y; e a posição 34 é cerca de 70% de A e cerca de 15% cada um de S e Y, e em que CDR2 da LC é diversificada, em que a posição 50 é cerca de 55% de G, e cerca de 9% cada um de D, R, S, Y, e L; a posição 53 é cerca de 52% de S e cerca de 8%

cada um de N, T, S, Y, E, e R; e a posição 56 é cerca de 64% de T e cerca de 9% cada um de E, R, S, e Y.

- 12.** A biblioteca da Reivindicação 1, em que a cadeia leve é uma cadeia leve (LC) A27,

em que CDR1 da LC é diversificada, em que a posição 27 é cerca de 55% de Q, e cerca de 9% cada um de E, R, Y, S, e L; a posição 28 é cerca de 46% de S e cerca de 9% cada um de N, T, Y, E, R, e L; a posição 30 é cerca de 55% de S e cerca de 9% cada um de D, N, R, T, e Y; a posição 30a é cerca de 46% de S e cerca de 9% cada um de G, N, R, T, Y, e D, e cerca de 8% com nenhum aminoácido; a posição 31 é cerca de 44% de S e cerca de 8% cada um de D, F, G, N, R, T, e Y; a posição 32 é cerca de 44% de Y e cerca de 7% cada um de F, D, L, N, Q, R, S, e Y; e a posição 34 é cerca de 70% de A e cerca de 15% cada um de S e Y, e

em que CDR3 da LC é diversificada, em que a posição 91 é cerca de 64% de Y, e cerca de 9% cada um de F, E, R, e S; a posição 92 é cerca de 52% de G e cerca de 8% cada um de A, D, R, S, T, e Y; a posição 93 é cerca de 52% de S e cerca de 8% cada um de D, F, N, R, T, e Y; a posição 94 é cerca de 55% de S e cerca de 9% cada um de W, E, R, Y, e S; a posição 95 é cerca de 64% de P e cerca de 9% cada um de E, R, Y, e S, e cerca de 8% com nenhum aminoácido; e a posição 96 é cerca de 55% de L e cerca de 9% cada um de E, R, P, Y, e S.

- 13.** A biblioteca da Reivindicação 1, em que a cadeia leve é uma cadeia leve (LC) A27

em que CDR2 da LC é diversificada, em que a posição 50 é cerca de 55% de G, e cerca de 9% cada um de D, R, S, Y, e L; a posição 53 é cerca de 52% de S e cerca de 8%

cada um de N, T, S, Y, E, e R; e a posição 56 é cerca de 64% de T e cerca de 9% cada um de E, R, S, e Y, e em que CDR3 da LC é diversificada, em que a posição 91 é cerca de 64% de Y, e cerca de 9% cada um de F, E, R, e S; a posição 92 é cerca de 52% de G e cerca de 8% cada um de A, D, R, S, T, e Y; a posição 93 é cerca de 52% de S e cerca de 8% cada um de D, F, N, R, T, e Y; a posição 94 é cerca de 55% de S e cerca de 9% cada um de W, E, R, Y, e S; a posição 95 é cerca de 64% de P e cerca de 9% cada um de E, R, Y, e S, e cerca de 8% com nenhum aminoácido; e a posição 96 é cerca de 55% de L e cerca de 9% cada um de E, R, P, Y, e S.

14. A biblioteca de qualquer uma das Reivindicações 1-10, em que a cadeia leve tem diversidade numa ou em mais de CDR1 de LC, CDR2 de LC, e CDR3 de LC.
15. A biblioteca da Reivindicação 14, em que a cadeia leve compreende uma CDR1 de LC, uma CDR2 de LC, ou uma CDR3 de LC de uma LC A27, em que CDR1 da LC é diversificada, em que a posição 27 é cerca de 55% de Q, e cerca de 9% cada um de E, R, Y, S, e L; a posição 28 é cerca de 46% de S e cerca de 9% cada um de N, T, Y, E, R, e L; a posição 30 é cerca de 55% de S e cerca de 9% cada um de D, N, R, T, e Y; a posição 30a é cerca de 46% de S e cerca de 9% cada um de G, N, R, T, Y, e D, e cerca de 8% com nenhum aminoácido; a posição 31 é cerca de 44% de S e cerca de 8% cada um de D, F, G, N, R, T, e Y; a posição 32 é cerca de 44% de Y e cerca de 7% cada um de F, D, L, N, Q, R, S, e Y; e a posição 34 é cerca de 70% de A e cerca de 15% cada um de S e Y;

em que CDR2 da LC é diversificada, em que a posição 50 é cerca de 55% de G, e cerca de 9% cada um de D, R, S, Y, e L; a posição 53 é cerca de 52% de S e cerca de 8% cada um de N, T, S, Y, E, e R; e a posição 56 é cerca de 64% de T e cerca de 9% cada um de E, R, S, e Y, ou em que CDR3 da LC é diversificada, em que a posição 91 é cerca de 64% de Y, e cerca de 9% cada um de F, E, R, e S; a posição 92 é cerca de 52% de G e cerca de 8% cada um de A, D, R, S, T, e Y; a posição 93 é cerca de 52% de S e cerca de 8% cada um de D, F, N, R, T, e Y; a posição 94 é cerca de 55% de S e cerca de 9% cada um de W, E, R, Y, e S; a posição 95 é cerca de 64% de P e cerca de 9% cada um de E, R, Y, e S, e cerca de 8% com nenhum aminoácido; e a posição 96 é cerca de 55% de L e cerca de 9% cada um de E, R, P, Y, e S, ou suas combinações.

- 16.** Um método de preparação da biblioteca de qualquer uma das Reivindicações 1 a 15, compreendendo o método *wobbling* digital.

Lisboa, 21 de Janeiro de 2015