



República Federativa do Brasil  
Ministério da Economia  
Instituto Nacional da Propriedade Industrial

**(11) BR 112017002251-6 B1**



**(22) Data do Depósito: 06/08/2015**

**(45) Data de Concessão: 25/01/2022**

**(54) Título: LIGANTE DE NUCLEOTÍDEO MODIFICADO**

**(51) Int.Cl.: C12N 9/10; C12N 15/00; C12P 21/06; C12P 21/04.**

**(30) Prioridade Unionista: 08/08/2014 GB 1414098.2.**

**(73) Titular(es): ILLUMINA CAMBRIDGE LIMITED.**

**(72) Inventor(es): XIAOLIN WU; XIAOHAI LIU.**

**(86) Pedido PCT: PCT GB2015052282 de 06/08/2015**

**(87) Publicação PCT: WO 2016/020691 de 11/02/2016**

**(85) Data do Início da Fase Nacional: 02/02/2017**

**(57) Resumo:** Algumas concretizações do presente pedido de patente referem-se a novos ligantes de nucleotídeos modificados para aumentar a eficiência da incorporação de nucleotídeos em aplicações de Sequenciamento e Síntese. Os métodos de preparação destes ligantes de nucleotídeos modificados são também fornecidos aqui.

## **"LIGANTE DE NUCLEOTÍDEO MODIFICADO".**

### CAMPO DA INVENÇÃO

[001] Algumas concretizações do presente pedido de patente referem-se a novos ligantes de nucleosídeos ou nucleotídeos para aumentar a incorporação de nucleotídeos no sequenciamento de DNA e outras aplicações de diagnóstico, por exemplo, sequenciamento por síntese.

### ANTECEDENTES DA INVENÇÃO

[002] Os avanços no estudo das moléculas foram conduzidos, em parte, pela melhoria das tecnologias utilizadas para caracterizar as moléculas ou as suas reações biológicas. Em particular, o estudo dos ácidos nucleicos DNA e RNA tem se beneficiado do desenvolvimento de tecnologias utilizadas para a análise de seqüências e o estudo de eventos de hibridização.

[003] Um exemplo das tecnologias que melhoraram o estudo de ácidos nucleicos é o desenvolvimento de matrizes fabricadas de ácidos nucleicos imobilizados. Estas matrizes têm tipicamente uma matriz de alta densidade de polinucleotídeos imobilizados num material de suporte sólido. Ver, por exemplo, Fodor et al., Trends Biotech. 12:19 - 26, 1994, que descreve modos de reunir diferentes ácidos nucleicos utilizando uma superfície de vidro sensibilizada quimicamente protegida por uma máscara, mas exposta em áreas definidas para permitir a ligação de fosforamiditos de nucleotídeos adequadamente modificados. As matrizes fabricadas podem também ser fabricadas pela técnica de "localizar" polinucleotídeos conhecidos num suporte sólido em posições predeterminadas (por exemplo, Stimpson et al., Proc. Nat. Acad. Sci. 92: 6379 - 6383, 1995).

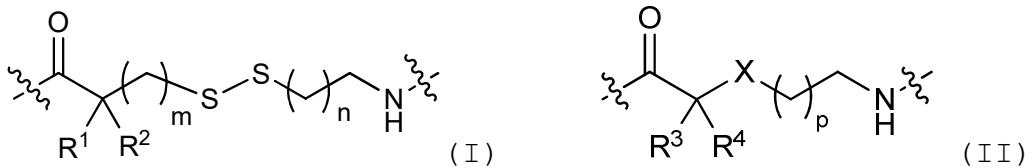
[004] Um modo de determinar a sequência nucleotídica de um ácido nucleico ligado a uma matriz é chamado de

"sequenciamento por síntese" ou "SBS". Esta técnica para determinar a sequência nucleotídica de DNA requer, idealmente, a incorporação controlada (isto é, uma de cada vez) do nucleotídeo complementar correto oposto ao ácido nucleico a ser sequenciado. Isto permite o sequenciamento exato por adição de nucleotídeos em múltiplos ciclos, à medida que cada resíduo de nucleotídeo é sequenciado um de cada vez, evitando, assim, a incorporação de uma série descontrolada de nucleotídeos. Cada nucleotídeo incorporado é lido utilizando um marcador apropriado ligado a ele antes da remoção da porção marcador e da subsequente sequência seguinte de sequenciamento.

[005] Consequentemente, no contexto de reações de sequenciamento de ácido nucleico seria desejável ser capaz de aumentar a taxa de incorporação de nucleotídeos durante o sequenciamento por síntese, de modo que a eficiência do método de sequenciamento possa ser melhorada.

#### SUMÁRIO DA INVENÇÃO

[006] Algumas concretizações aqui reveladas referem-se a um nucleosídeo ou nucleotídeo ligado covalentemente a um fluoróforo através de um ligante, em que o referido ligante compreende uma estrutura de fórmula (I) ou (II), ou uma combinação de ambos.



em que

$\text{R}^1$  é selecionado de entre hidrogênio ou  $\text{C}_{1-6}$  alquil opcionalmente substituído;

$\text{R}^2$  é selecionado de entre hidrogênio ou  $\text{C}_{1-6}$  alquil opcionalmente substituído;

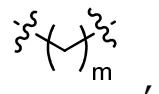
$\text{R}^3$  é selecionado de entre hidrogênio,  $\text{C}_{1-6}$  alquil opcionalmente substituído,  $-\text{NR}^5-\text{C}(=\text{O})\text{R}^6$ , ou  $\text{NR}^7-\text{C}(=\text{O})-\text{OR}^8$ ;

$R^4$  é selecionado de entre hidrogênio ou  $C_{1-6}$  alquil opcionalmente substituído;

cada  $R^5$  e  $R^7$  é selecionado, independentemente, de entre hidrogênio,  $C_{1-6}$  alquil opcionalmente substituído, fenil opcionalmente substituído, ou  $C_{7-12}$  aralquil opcionalmente substituído;

cada  $R^6$  e  $R^8$  é selecionado, independentemente, de entre  $C_{1-6}$  alquil opcionalmente substituído, fenil opcionalmente substituído,  $C_{7-12}$  aralquil opcionalmente substituído,  $C_{3-7}$  cicloalquil opcionalmente substituído, ou heteroaril opcionalmente substituído de 5 a 10 membros;

cada um da unidade de repetição de metileno no



$\text{---}[CH_2]_n\text{---}$  ou  $\text{---}[CH_2]_p\text{---}$  é opcionalmente substituído;

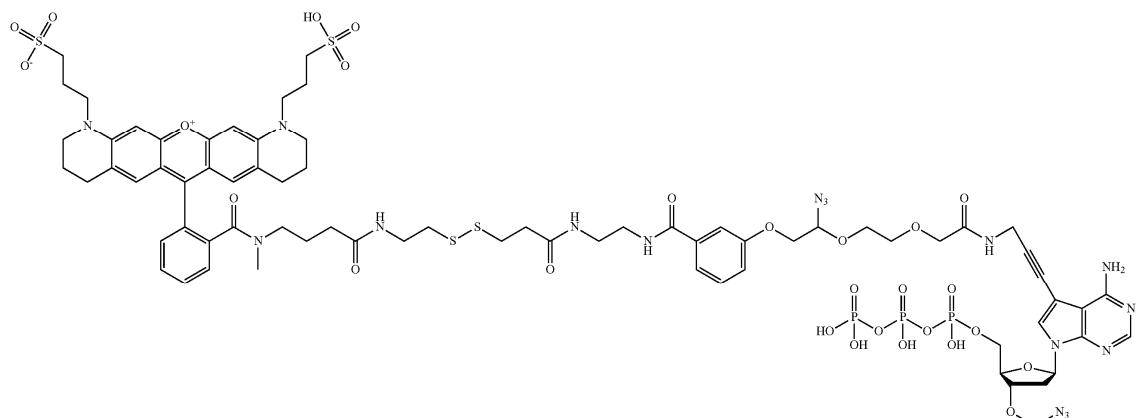
$X$  é selecionado a partir de metileno ( $CH_2$ ), oxigênio (O) ou enxofre (S);

$m$  é um número inteiro de 0 a 20;

$n$  é um número inteiro de 1 a 20; e

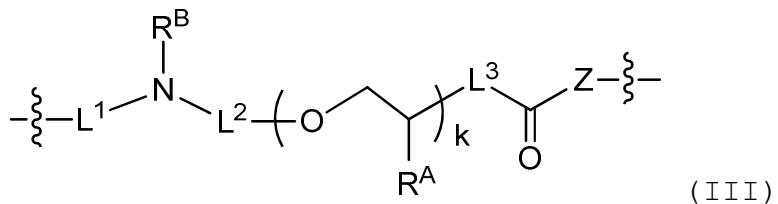
$p$  é um número inteiro de 1 a 20.

[007] Em algumas concretizações, o nucleotídeo ou nucleotídeo marcado com fluoróforo compreende uma estrutura de fórmula (I) não tem a estrutura

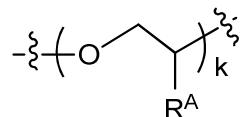


[0008] Algumas concretizações aqui descritas referem-se a um nucleosídeo ou nucleotídeo ligado covalentemente a um

fluoroforo através de um ligante, em que o referido ligante compreende uma estrutura de fórmula (III):



em que L1 está ausente ou compreende um ligante de qualquer um descrito na fórmula (I) ou (II), ou uma porção protetora, ou suas combinações; L2 é selecionado de alquíleno C<sub>1-20</sub> opcionalmente substituído, heteroalquíleno C<sub>1-20</sub> opcionalmente substituído, alquíleno C<sub>1-20</sub> opcionalmente substituído interrompido por um grupo aromático substituído ou heteroalquíleno C<sub>1-20</sub> opcionalmente substituído interrompido por um grupo aromático substituído; L3 é selecionado de alquíleno C<sub>1-20</sub> opcionalmente substituído, ou heteroalquíleno C<sub>1-20</sub> opcionalmente substituído; RA é seleccionado de hidrogênio, ciano, hidroxila, halogênio, C<sub>1-6</sub> alquilo, C<sub>1-6</sub> alcoxi, C<sub>1-6</sub> haloalquilo, C<sub>1-6</sub> haloalcoxi ou azido e em que pelo



menos uma das unidades repetidas de  $\text{RA}$  compreende um grupo azido;  $Z$  é selecionado a partir de oxigênio (O) ou NRB; Cada RB e RC é independentemente selecionado de hidrogênio ou alquilo  $C_{1-6}$  opcionalmente substituído; e  $k$  é um número inteiro de 1 a 50.

[0009] Algumas concretizações aqui descritas referem-se a kits que compreendem um nucleosídeo ou nucleotídeo marcado compreendendo um ligante entre o fluoróforo e o nucleosídeo ou nucleotídeo, em que o ligante compreende uma estrutura de qualquer das fórmulas (I), (II) ou (III), ou combinações Do mesmo.

[0010] Algumas concretizações aqui descritas referem-se a reagentes para modificar um nucleosídeo ou um nucleotídeo que compreende um fluoróforo e um ligante, em que o ligante

compreende uma estrutura de qualquer uma das fórmulas (I), (II) ou (III), ou suas combinações.

[0011] Algumas concretizações aqui reveladas referem-se a métodos para detectar um nucleosídeo que foi incorporado num polinucleotídeo, compreendendo: (a) incorporar um nucleosídeo ou nucleotídeo marcado compreendendo um ligante num polinucleotídeo; e (b) detectar um sinal fluorescente a partir do referido nucleosídeo ou nucleotídeo marcado que foi incorporado na etapa (a), em que o ligante compreende uma estrutura de qualquer um de fórmula (I), (II) ou (III), ou suas combinações. Em algumas concretizações, o método compreende ainda: proporcionar uma cadeia de ácido nucleico padrão e uma cadeia de ácido nucleico parcialmente hibridizada, em que a etapa (a) incorpora na cadeia hibridizada pelo menos um nucleosídeo ou nucleotídeo que é complementar a um nucleosídeo ou nucleotídeo no correspondente. Em que a etapa (b) identifica a base do nucleosídeo ou nucleotídeo incorporado, indicando desse modo a identidade do nucleosídeo ou nucleotídeo complementar da cadeia molde.

[0012] Algumas concretizações aqui reveladas referem-se a métodos de sequenciamento de uma molécula de ácido nucleico molde, compreendendo: a incorporação de um ou mais nucleotídeos marcados numa cadeia de ácido nucleico complementar ao ácido nucleico molde; Determinar a identidade da base presente num ou mais nucleotídeos marcados incorporados de modo a determinar a sequência da molécula de ácido nucleico molde; em que a identidade da base presente num ou mais nucleotídeos marcados é determinada pela detecção de um sinal fluorescente produzido pelos referidos nucleotídeos marcados; E em que pelo menos um nucleotídeo marcado incorporado compreendendo um ligante como descrito acima, em que o ligante compreende uma estrutura de qualquer

um de fórmula (I), (II) ou (III), ou suas combinações. Em algumas concretizações, a identidade da base presente num ou mais nucleotídeos é determinada após cada etapa de incorporação de nucleotídeos.

#### BREVE DESCRIÇÃO DOS DESENHOS

[0013] A FIG. 1A ilustra um grupo estrutural de ligação parcial de um nucleotídeo marcado padrão.

[0014] A FIG. 1B ilustra o nucleotídeo marcado da FIG. 1A com dois ligantes possíveis 125 e 130 para serem inseridos no grupo de ligação padrão da FIG. 1A.

[0015] A FIG. 2 mostra um gráfico da taxa de incorporação de nucleotídeos utilizando o nucleotídeo marcado da FIG. 1A e os nucleotídeos marcados modificados da FIG. 1B.

[0016] As FIGs. 3A-3E ilustram as fórmulas estruturais de ligantes adicionais a serem inseridos no grupo de ligação padrão da FIG. 1A.

[0017] A FIG. 4 mostra uma tabela de dados de uma corrida de sequenciamento de dois corantes utilizada para avaliar o efeito do inserto 125 da FIG. 1B e o inserto 315 da FIG. 3B sobre a qualidade da sequenciamento.

[0018] As FIGs. 5A e 5B mostram um gráfico da taxa de erro para a leitura 1 e um gráfico da taxa de erro para a leitura 2 da corrida de sequenciamento da FIG. 4 utilizando ligantes inserto 125 e 315.

[0019] A FIG. 6 mostra uma tabela de dados de uma corrida de sequenciamento utilizada para avaliar o efeito do inserto 125 da FIG. 1B e o inserto 310 da FIG. 3A na qualidade de sequenciamento.

[0020] As FIGs. 7A e 7B mostram um gráfico da taxa de erro para a leitura 1 e um gráfico da taxa de erro para a leitura 2 da sequência de sequenciamento da FIG. 6 utilizando o inserto de ligação 125.

[0021] A FIG. 8A ilustra um exemplo de uma estrutura de

ligação LN3 padrão.

[0022] As FIGs. 8B, 8C e 8D mostram três exemplos de estruturas modificadas do ligante LN3 da FIG. 8A.

[0023] A FIG. 8E ilustra a inserção de uma unidade de proteção no ligante. A FIG. 8D.

[0024] A FIG 9A é um cromatograma que mostra o aparecimento de uma impureza num ffA com ligante SS. A FIG 9B é uma tabela comparando a estabilidade de ffAs com ligante SS e ligante AEDI. A FIG 9C é um cromatograma que mostra uma comparação do ligante SS e do ligante AEDI ffAs em IMX 60° durante 22 horas, mostrando novamente uma impureza com o ligante SS.

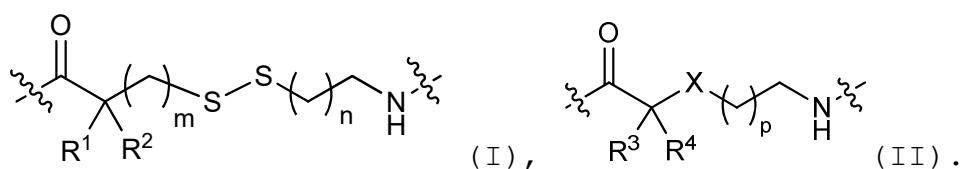
[0025] As FIGS 10A, 10B e 10C mostram o aumento inesperado da velocidade de incorporação de nucleotídeos em solução com alterações de ligante. A FIG. 10A é um gráfico que mostra a taxa de incorporação a 1 uM com a FIG. 10B mostrando os resultados tabulados. A FIG. 10C mostra esquematicamente os ligantes AEDI e SS com NR550S0.

[0026] A FIG. 11A mostra gráficos de dispersão para combinações V10 com diferentes A-550S0 (mesma concentração). A FIG. 11B mostra ligantes Kcat FFA em solução.

[0027] As FIGS 12A e 12B mostram as métricas de sequenciamento em ciclos M111, Human550, 2x151.

#### DESCRIÇÃO DETALHADA DAS CONCRETIZAÇÕES PREFERIDAS

[0027] Algumas concretizações aqui descritas referem-se a um nucleosídeo ou nucleotídeo ligado covalentemente a um fluoróforo através de um ligante, em que o ligante compreende uma estrutura de fórmula (I) ou (II) abaixo ou uma combinação de ambas, em que as definições das variáveis São definidos acima.



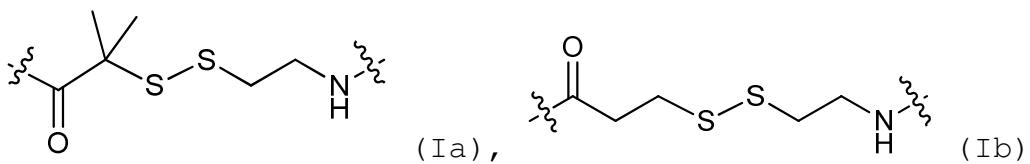
[0028] Em algumas concretizações da estrutura de fórmula (I), R<sup>1</sup> é hidrogênio. Em algumas outras concretizações, R<sup>1</sup> é alquilo C<sub>1-6</sub> opcionalmente substituído. Em algumas destas concretizações, R<sup>1</sup> é metil.

[0029] Em quaisquer concretizações de R<sup>1</sup>, tal como aqui descrito da fórmula (I), R<sup>2</sup> é hidrogênio. Em algumas outras concretizações, R<sup>2</sup> é alquilo C<sub>1-6</sub> opcionalmente substituído. Em algumas destas concretizações, R<sup>2</sup> é metila. Numa concretização, tanto R<sup>1</sup> como R<sup>2</sup> são metil. Em outra concretização, tanto R<sup>1</sup> como R<sup>2</sup> são hidrogênio.

[0030] Em algumas concretizações da estrutura de fórmula (I), m é 0. Em algumas outras concretizações, m é 1.

[0031] Em algumas concretizações da estrutura de fórmula (I), n é 1.

[0032] Em algumas concretizações da estrutura de fórmula (I), a estrutura de fórmula (I) também pode ser representada pela fórmula (Ia) ou (Ib):



[0033] Em algumas concretizações aqui descritas, a fórmula (Ia) é referida como "AEDI" e a fórmula (Ib) é referida como "SS".

[0034] Em algumas concretizações da estrutura de fórmula (II), R<sup>3</sup> é hidrogênio. Em algumas outras concretizações, R<sup>3</sup> é alquil C<sub>1-6</sub> opcionalmente substituído. Em algumas destas concretizações, R<sup>3</sup> é metil. Em algumas concretizações, R<sup>3</sup> é -NR<sub>5</sub>-C(=O)R<sup>6</sup>. Em algumas destas concretizações, R<sup>5</sup> é hidrogênio. Em algumas destas concretizações, R<sup>6</sup> é alquilo C<sub>1-6</sub> opcionalmente substituído, por exemplo, metil. Em algumas concretizações, R<sup>3</sup> é -NR<sub>7</sub>-C(=O)OR<sup>8</sup>. Em algumas destas concretizações, R<sup>7</sup> é hidrogênio. Em algumas destas concretizações, R<sup>8</sup> é alquil C<sub>1-6</sub> opcionalmente substituído,

por exemplo, t-butil.

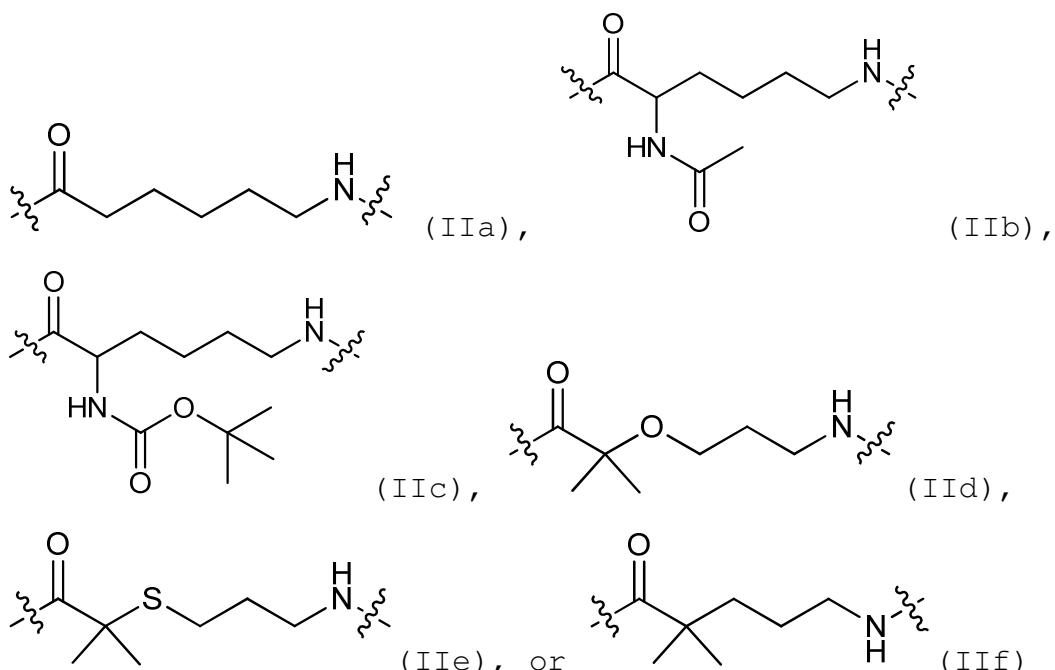
[0035] Em quaisquer concretizações de  $R^3$  como aqui descrito da fórmula (II),  $R^4$  é hidrogênio. Em algumas outras concretizações,  $R^4$  é alquil C1-6 opcionalmente substituído.

[0036] Em algumas destas concretizações,  $R^4$  é metilo. Numa concretização, tanto  $R^3$  como  $R^4$  são metil. Em outra concretização, tanto  $R^3$  como  $R^4$  são hidrogênio. Numa concretização,  $R^3$  é  $-\text{NH}(\text{C=O})\text{CH}_3$  e  $R^4$  é hidrogênio. Em outra concretização,  $R^3$  é  $-\text{NH}(\text{C=O})\text{OtBu}(\text{Boc})$  e  $R^4$  é hidrogênio.

[0037] Em algumas concretizações da estrutura de fórmula (II), X é metileno, que pode ser opcionalmente substituído. Em outra concretização, X é oxigênio (O). Ainda em outra concretização, X é enxofre (S).

[0038] Em algumas concretizações da estrutura de fórmula (II), p é 1. Em algumas outras concretizações, p é 2.

[0039] Em algumas concretizações da estrutura de fórmula (II), a estrutura de fórmula (II) também pode ser representada pela fórmula (IIa), (IIb), (IIc), (IId), (IIe) ou (IIIf):

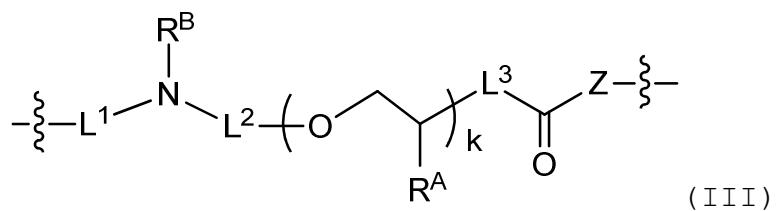


[0040] Em algumas concretizações aqui descritas, a fórmula (IIa) é referida como "ACA", a fórmula (IIb) é

referida como "AcLys", a fórmula (IIc) é referida como "BocLys", a fórmula (IId) é referida como "DMeO", a fórmula (IIe) é referida como "dMeS" e a fórmula (IIIf) é referida como "DMP".

[0041] Em quaisquer concretizações, o nucleosídeo ou nucleotídeo marcado com fluoróforo através de um ligante compreendendo uma estrutura de fórmula (I) ou (II) como aqui descrita, o nucleosídeo ou nucleotídeo pode ser ligado ao lado esquerdo do ligante, quer diretamente, ou através de uma porção de ligação adicional.

[0042] Algumas concretizações aqui descritas referem-se a um nucleosídeo ou nucleotídeo ligado covalentemente a um fluoróforo, através de um ligante, compreendendo o referido ligante uma estrutura de fórmula (III), e em que as definições das variáveis são definidas acima.

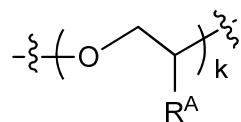


[0043] Em algumas concretizações da estrutura de fórmula (III), L1 está ausente. Em algumas outras concretizações, L1 é um ligante descrito acima compreendendo a estrutura de fórmula (I) ou (II), em particular a fórmula (Ia), (Ib), (II), (IIa), (IIb), (IIc), (IId), (IIe) ou (IIIf). Nalgumas outras concretizações, L1 pode ser uma metade protetora compreendendo molécula que protege contra danos no DNA. Em algumas destas concretizações, a unidade de protecção compreende Trolox, ácido gálico, p-nitro-benzilo (pNB) ou ascorbato, ou combinações destes.

[0044] Em algumas concretizações da estrutura de fórmula (III), L2 é alquíleno C<sub>1-20</sub> opcionalmente substituído. Em algumas concretizações adicionais, L2 é alquíleno C<sub>4-10</sub> opcionalmente substituído. Em algumas destas concretizações,

L2 é heptileno. Em algumas outras concretizações, L2 é heteroalquíleno C<sub>1-20</sub> opcionalmente substituído. Em algumas destas concretizações, o heteroalquíleno C<sub>1-20</sub> opcionalmente substituído compreende um ou mais átomos de azoto. Em algumas destas concretizações, pelo menos um dos átomos de carbono do heteroalquíleno C<sub>1-20</sub> está substituído com oxo (= O). Em algumas concretizações adicionais, L2 é heteroalquíleno C<sub>3-6</sub> opcionalmente substituído. Em algumas concretizações, L2 é interrompido por um grupo aromático substituído, tal como um grupo arilo C<sub>6-10</sub> substituído, ou grupos heteroarilo substituído com 5 a 10 membros compreendendo um a três heteroátomos. Em algumas destas concretizações, L2 é interrompido por um grupo fenilo substituído. Nalgumas destas concretizações, o grupo fenilo é substituído com um ou mais substituintes (até quatro) seleccionados de entre nitro, ciano, halo, hidroxi, C<sub>1-6</sub> alquilo, C<sub>1-6</sub> alcoxi, C<sub>1-6</sub> haloalquilo, C<sub>1-6</sub> haloalcoxi ou Sulfonil. Em algumas outras concretizações, o grupo fenil está substituído com um a quatro substituintes seleccionados de nitro, ciano, halo ou hidróxido de sulfônico (i.e., -S (= O) 2OH).

[0045] Em algumas concretizações da estrutura de fórmula



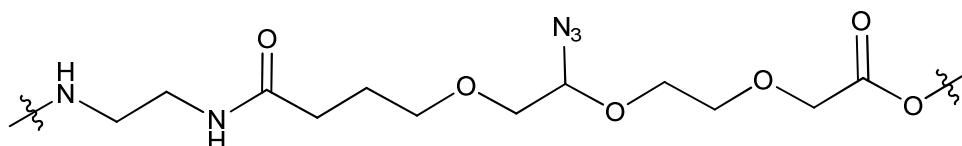
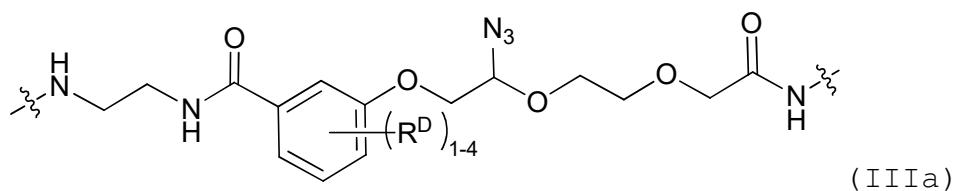
(III), RA em R<sup>A</sup> é seleccionado de hidrogênio ou azido. Em algumas destas concretizações, k é 2 com um RA sendo azido e o outro sendo hidrogênio.

[0046] Em algumas concretizações da estrutura de fórmula (III), L<sup>3</sup> é alquíleno C<sub>1-20</sub> opcionalmente substituído. Em algumas concretizações adicionais, L<sup>3</sup> é alquíleno C<sub>1-6</sub> opcionalmente substituído. Em algumas destas concretizações, L<sup>3</sup> é etileno. Em algumas outras concretizações, L<sup>3</sup> é heteroalquíleno C<sub>1-20</sub> opcionalmente substituído. Em algumas destas concretizações, o heteroalquíleno C<sub>1-20</sub> opcionalmente substituído compreende um ou mais átomos de oxigénio.

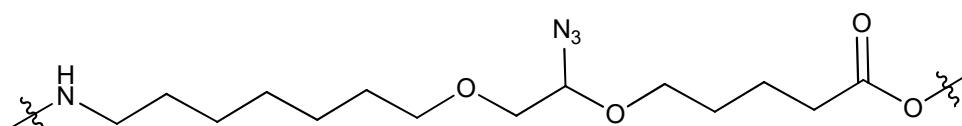
Nalgumas destas concretizações, L<sup>1</sup> é óxido de alquíleno C<sub>1-6</sub> opcionalmente substituído, por exemplo, óxido de C<sub>1-3</sub> alquíleno.

[0047] Em algumas concretizações da estrutura de fórmula (III), RB é hidrogênio. Em algumas concretizações, RC é hidrogênio. Em algumas concretizações adicionais, tanto RB como RC são hidrogênio.

[0048] Em algumas concretizações da estrutura de fórmula (III), a estrutura de fórmula (III) também pode ser representada pela fórmula (IIIa), (IIIb) ou (IIIc):



(IIIb),



(IIIc);

Em que RD é selecionado de entre nitro, ciano, halo, hidroxi, C<sub>1-6</sub> alquilo, C<sub>1-6</sub> alcoxi, C<sub>1-6</sub> haloalquilo, C<sub>1-6</sub> haloalcoxi ou hidróxido sulfonila. Em algumas concretizações adicionais, RD é selecionado a partir de nitro, ciano, halo, ou hidróxido de sulfonila.

[0049] Em quaisquer concretizações do nucleosídeo ou nucleotídeo marcado com fluoróforo através de um ligante compreendendo uma estrutura de fórmula (III) como aqui descrita, o fluoróforo pode ser ligado ao lado esquerdo do ligante, quer directamente, quer através de porção de ligação adicional.

Em quaisquer concretizações aqui descritas com respeito a um ligante que compreende uma estrutura de fórmula (I), (II) ou (III), quando o termo "opcionalmente substituído" é usado para definir uma variável, tal variável pode ser não substituída.

#### Definições

[0050] Salvo definição em contrário, todos os termos técnicos e científicos aqui utilizados têm o mesmo significado que é comumente entendido por um especialista na técnica. O uso do termo "incluso", bem como outras formas, tais como "incluir", "inclui" e "incluso", não é limitativo. O uso do termo "tendo", bem como outras formas, como "ter", "tem" e "teve", não é limitativo. Tal como utilizado nesta especificação, quer numa frase de transição quer no corpo da reivindicação, os termos "compreende (m)" e "compreendendo" devem ser interpretados como tendo um significado aberto. Isto é, os termos acima devem ser interpretados como sinónimo com as frases "ter pelo menos" ou "incluir pelo menos". Por exemplo, quando usado no contexto de um processo, o termo "compreendendo" significa que o processo inclui pelo menos os passos descritos, mas pode incluir passos adicionais. Quando utilizado no contexto de um composto, composição ou dispositivo, o termo "compreendendo" significa que o composto, composição ou dispositivo inclui pelo menos as características ou componentes citados, mas também pode incluir características ou componentes adicionais.

[0051] Os cabeçalhos de seção utilizados aqui são apenas para fins organizacionais e não devem ser interpretados como limitando a matéria descrita. Tal como aqui utilizado, abreviaturas orgânicas comuns são definidas como se segue:

Ac	acetila
Ac2O	Anidrido acético

Aq.	Aquoso
Bn	Benzil
Bz	Benzoil
BOC ou Boc	terc-butoxicarbonilo
Bu	n-butil
cat.	Catalisador
° C	Temperatura em graus Centígrados
CHAPS 3 -	[ (3-Colamidopropil)dimetilamônio] -1-
propanossulfonato	
DATP	Desoxiadenosina trifosfato
DCTP	Deoxicitidina trifosfato
DGTP	Deoxiguanosina trifosfato
DTTP	Trifosfato de desoxitimidina
DdNTP (s)	Dideoxinucleotídeo (s)
DCM	Cloreto de metíleno
DMA	Dimetilacetamida
DMAP	4-Dimetilaminopiridina
DMF	N, N'-dimetilformamida
DMSO	Dimetilsulfóxido
DSC	carbonato de N, N'-dissuccinimidilo
EDTA	Etilenodiamina tetra-acético
Et	Et
EtOAc	Acetato de etila
FfN	Nucleotídeo totalmente funcional
FfA	Nucleotídeo de adenosina completamente funcionalizado
G	Gram (s)
GPC	Cromatografia de troca em gel
H ou h	Hora (s)
Base de Hunig	N,N-Diisopropiletilamina
IPr	Isopropil
KPi	tampão fosfato de potássio 10 mM a pH 7,0
IPA	Álcool isopropílico

LCMS	Cromatografia líquida-espectrometria de massa
LDA	Diisopropilamida de lítio
M ou min	Minuto (s)
MeCN	Acetonitril
ML	MillilitRO (s)
PEG	Polietilenoglicol
PG	Grupo de proteção
Ph	FenIl
PNB	p-nitro-benzila
Ppt	Precipitado
Rt	Temperatura ambiente
SBS	Sequenciamento por Síntese
-S (O) 2OH	Hidróxido de sulfonila
TEA	Trietilamina
TEAB	Brometo de tetraetilamônio
TFA	Ácido trifluoroacético
Tert 1	Tertário
THF	Tetrahidrofurano
TLC	Cromatografia em camada fina
TSTU	tetrafluoroborato de O-(N-succinimidil)-N,N,N',N'-tetrametilurônio
µL	Microlitro (s).

[0052] Tal como aqui utilizado, o termo "matriz" refere-se a uma população de diferente sondas de moleculares que estão ligadas a um ou mais substratos, de tal modo que as diferentes sondas moleculares podem ser diferenciadas uma da outra, de acordo com a localização relativa. Uma matriz pode incluir diferentes sondas de moleculares que estão localizadas numa posição endereçável diferente num substrato. Alternativamente ou adicionalmente, uma matriz pode incluir substratos separados cada um com uma molécula de sonda diferente, em que as diferentes sondas moleculares podem ser

identificadas, de acordo com as localizações dos substratos numa superfície à qual os substratos estão ligados ou de acordo com as localizações dos substratos em um líquido. As matrizes exemplificativas nas quais os substratos separados estão localizados numa superfície incluem, sem limitação, as que incluem as pérolas nos poços, como descrito, por exemplo, na Patente U.S. N.º 6 355 431 B1, US 2002/0102578 e na Publicação PCT N.º WO 00/63437. Exemplos de formatos que podem ser utilizados na invenção para distinguir as pérolas numa disposição de líquido, por exemplo, utilizando um dispositivo microfluídico, tal como um separador de células activado por fluorescência (FACS), são descritos, por exemplo, na Pat. U.S. No. 6,524,793. Outros exemplos de arranjos que podem ser utilizados na invenção incluem, sem limitação, os descritos nas patentes norte-americanas 5,429,807; 5,436,327; 5,561,071; 5,583,211; 5,658,734; 5,837,858; 5,874,219; 5,919,523; 6,136,269; 6,287,768; 6,287,776; 6,288,220; 6,297,006; 6,291,193; 6,346,413; 6,416,949; 6,482,591; 6,514,751 e 6,610,482; and WO 93/17126; WO 95/11995; WO 95/35505; EP 742 287; e EP 799 897.

[0053] Tal como aqui utilizado, o termo "ligado covalentemente" ou "ligado covalentemente" refere-se à formação de uma ligação química que é caracterizada pela partilha de pares de electrões entre átomos. Por exemplo, um revestimento de polímero ligado covalentemente refere-se a um revestimento de polímero que forma ligações químicas com uma superfície funcionalizada de um substrato, em comparação com a ligação à superfície através de outros meios, por exemplo, adesão ou interacção electrostática. Será apreciado que os polímeros que estão ligados covalentemente a uma superfície também podem ser ligados através de meios para além da ligação covalente.

[0054] Tal como aqui utilizado, "Ca para Cb" ou "Ca-b" em que "a" e "b" são números inteiros referem-se ao número de átomos de carbono no grupo especificado. Ou seja, o grupo pode conter de "a" a "b", inclusive, átomos de carbono. Assim, por exemplo, um grupo "alquilo C1 a C4" ou "alquilo C1-4" refere-se a todos os grupos alquilo com 1 a 4 átomos de carbono, isto é,  $\text{CH}_3-$ ,  $\text{CH}_3\text{CH}_2-$ ,  $\text{CH}_3\text{CH}_2\text{CH}_2-$ ,  $(\text{CH}_3)_2\text{CH}-$ ,  $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2-$ ,  $\text{CH}_3\text{CH}_2\text{CH}(\text{CH}_3)-$  e  $(\text{CH}_3)_3\text{C}-$ .

[0055] O termo "halogênio" ou "halo", tal como aqui utilizado, significa qualquer um dos átomos radio-estáveis da coluna 7 da Tabela Periódica dos Elementos, por exemplo, flúor, cloro, bromo ou iodo, com flúor sendo preferido o cloro.

[0056] Tal como aqui utilizado, "alquil" refere-se a uma cadeia de hidrocarboneto linear ou ramificada que está totalmente saturada (isto é, não contém ligações duplas ou triplas). O grupo alquilo pode ter 1 a 20 átomos de carbono (sempre que aparece aqui, uma gama numérica tal como "1 a 20" refere-se a cada número inteiro na gama dada, p.e., "1 a 20 átomos de carbono" significa que o grupo alquilo pode consistir em 1 átomo de carbono, 2 átomos de carbono, 3 átomos de carbono, etc., até 20 átomos de carbono inclusive, embora a presente definição também cubra a ocorrência do termo "alquilo" onde não é designada uma gama numérica). O grupo alquilo pode também ser um alquilo de tamanho médio possuindo 1 a 9 átomos de carbono. O grupo alquilo pode também ser um alquilo inferior possuindo 1 a 4 átomos de carbono. O grupo alquilo pode ser designado como "alquil C1-4" ou designações semelhantes. Apenas a título de exemplo, "alquilo C1-4" indica que há um a quatro átomos de carbono na cadeia alquilo, isto é, a cadeia alquilo é selecionada do grupo que consiste em metilo, etilo, propilo, isopropilo, n-Butilo, iso-butilo, sec-butilo e t-butilo. Os grupos alquilo típicos incluem, mas

não estão de modo algum limitados a, metilo, etilo, propilo, isopropilo, butilo, isobutilo, butilo terciário, pentilo, hexilo e semelhantes. O grupo alquilo pode estar substituído ou não substituído.

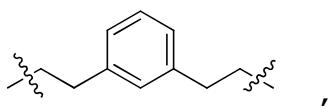
[0057] Tal como aqui utilizado, "alcoxi" refere-se à fórmula -OR em que R é um alquilo como definido acima, tal como "alcoxi C1-9", incluindo mas não limitado a metoxi, etoxi, n-propoxi, 1- Metiletoxi (isopropoxi), n-butoxi, iso-butoxi, sec-butoxi e terc-butoxi e semelhantes.

[0058] Tal como aqui utilizado, "heteroalquil" refere-se a uma cadeia hidrocarboneto linear ou ramificada contendo um ou mais heteroátomos, isto é, um elemento diferente de carbono, incluindo, mas não limitado a, azoto, oxigénio e enxofre, no esqueleto da cadeia. O grupo heteroalquilo pode ter 1 a 20 átomos de carbono, embora a presente definição também cubra a ocorrência do termo "heteroalquilo" em que não é designada uma gama numérica. O grupo heteroalquilo pode também ser um heteroalquilo de tamanho médio possuindo 1 a 9 átomos de carbono. O grupo heteroalquilo pode também ser um heteroalquilo inferior possuindo 1 a 4 átomos de carbono. O grupo heteroalquilo pode ser designado como "heteroalquilo C1-4" ou designações semelhantes. O grupo heteroalquilo pode conter um ou mais heteroátomos. Apenas a título de exemplo, "heteroalquilo C1-4" indica que há um a quatro átomos de carbono na cadeia heteroalquilo e adicionalmente um ou mais heteroátomos no esqueleto da cadeia.

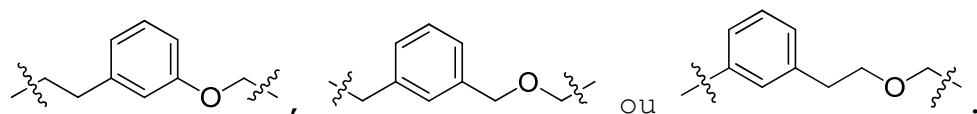
[0059] Tal como aqui utilizado, "alquileno" significa um grupo químico de radicais ramificados ou de cadeia linear completamente saturados contendo apenas carbono e hidrogénio que está ligado ao resto da molécula através de dois pontos de ligação (isto é, um alcanodiil). O grupo alquileno pode ter 1 a 20 átomos de carbono, embora a presente definição também cubra a ocorrência do termo alquileno em que não é

designada uma gama numérica. O grupo alquíleno também pode ser um alquíleno de tamanho médio com 1 a 9 átomos de carbono. O grupo alquíleno também pode ser um alquíleno inferior possuindo 1 a 4 átomos de carbono. O grupo alquíleno pode ser designado como "alquíleno C1-4" ou designações semelhantes. A título de exemplo apenas, "alquíleno C1-4" indica que há um a quatro átomos de carbono na cadeia alquíleno, isto é, a cadeia alquíleno é seleccionada do grupo que consiste em metíleno, etíleno, etan-1,1-diílo, Propileno, propan-1, 1-diílo, propano-2,2-diílo, 1-metil-etíleno, butíleno, butan-1,1-diílo, butan-2,2-diílo, 2-metil-propan- 1-diílo, 1-metil-propileno, 2-metil-propileno, 1,1-dimetil-etíleno, 1,2-dimetil-etíleno e 1-etil-etíleno. Tal como aqui utilizado, quando um alquíleno é interrompido por um grupo aromático, refere-se à inserção de um grupo aromático entre uma ligação carbono-carbono da cadeia alquíleno através de dois pontos de ligação ou a ligação de um grupo aromático a um terminal do Alquíleno através de um ponto de ligação. Por exemplo, quando um n-butíleno é interrompido por um grupo fenilo,

estruturas exemplificativas incluem



$\text{NHCH}_2\text{CH}_2-$ ,  $\text{--CH}_2\text{NHCH}_2-$ ,  $-\text{CH}_2\text{NHCH}_2\text{CH}_2-$  e semelhantes. Tal como aqui utilizado, quando um heteroalquíleno é interrompido por um grupo aromático, refere-se à inserção de um grupo aromático entre uma ligação carbono-carbono ou ligação carbono-heteroátomo da cadeia heteroalquíleno através de dois pontos de ligação ou a ligação de um grupo aromático A um terminal da cadeia heteroalquíleno através de um ponto de ligação. Por exemplo, quando um óxido de n-propileno é interrompido por um grupo fenilo, estruturas exemplificativas incluem



[0061] Tal como aqui utilizado, "alquenil" refere-se a um grupo alquílico que contém na cadeia hidrocarboneto linear ou ramificada uma ou mais ligações duplas. Um grupo alquenil pode estar não substituído ou substituído. Tal como aqui utilizado, "alcinílico" refere-se a um grupo alquílico que contém na cadeia hidrocarboneto linear ou ramificada uma ou mais ligações triplas. Um grupo alcinílico pode ser não substituído ou substituído.

[0062] Tal como aqui utilizado, "cicloalquílico" refere-se a um sistema de anel de hidrocarboneto mono- ou multicíclico completamente saturado (sem dupla ou tripla ligação). Quando compostos por dois ou mais anéis, os anéis podem ser unidos em conjunto por fusão. Os grupos cicloalquílico podem conter 3 a 10 átomos no anel (s). Em algumas formas de realização, os grupos cicloalquílico podem conter 3 a 8 átomos no anel (s). Um grupo cicloalquílico pode estar não substituído ou substituído. Os grupos cicloalquílico típicos incluem, mas não estão de modo algum limitados a, ciclopropílico, ciclobutílico, ciclopentílico, ciclo-hexílico, ciclo-heptílico e ciclooctílico.

[0063] O termo "aromático" refere-se a um sistema de anel ou anel que tem um sistema de electrões pi conjugado e

inclui grupos aromáticos carbocíclicos aromáticos (por exemplo, fenilo) e heterocíclicos (por exemplo, piridina). O termo inclui grupos policíclicos monocíclicos ou de anel fundido (isto é, anéis que partilham pares de átomos adjacentes) desde que o sistema de anel inteiro seja aromático.

[0064] Tal como aqui utilizado, "aril" refere-se a um anel aromático ou sistema de anel (isto é, dois ou mais anéis fundidos que compartilham dois átomos de carbono adjacentes) contendo apenas carbono na estrutura do anel. Quando o arilo é um sistema de anel, cada anel no sistema é aromático. O grupo arilo pode ter 6 a 18 átomos de carbono, embora a presente definição também cubra a ocorrência do termo "arilo" onde não é designada uma gama numérica. Em algumas concretizações, o grupo arilo tem 6 a 10 átomos de carbono. O grupo arilo pode ser designado como "arilo C6-10", "arilo C6 ou C10" ou designações semelhantes. Exemplos de grupos arilo incluem, mas não estão limitados a, fenil, naftilo, azulenilo e antraquenil.

[0065] Um "aralquil" ou "arilalquilo" é um grupo arilo ligado, como um substituinte, através de um grupo alquíleno, tal como "aralquilo C7-14" e semelhantes, incluindo mas não limitado a benzilo, 2-feniletilo, 3-fenilpropilo e naftilalquilo. Em alguns casos, o grupo alquíleno é um grupo alquíleno inferior (isto é, um grupo alquíleno C1-4).

[0066] Tal como aqui utilizado, "heteroarilo" refere-se a um anel aromático ou sistema de anel (isto é, dois ou mais anéis fundidos que partilham dois átomos adjacentes) que contêm um ou mais heteroátomos, isto é, um elemento diferente de carbono, incluindo mas não limitado a, azoto, oxigénio e enxofre, no esqueleto do anel. Quando o heteroarilo é um sistema de anel, cada anel no sistema é aromático. O grupo heteroarilo pode ter 5-18 membros do anel (isto é, o número

de átomos que compõem o esqueleto do anel, incluindo átomos de carbono e heteroátomos), embora a presente definição também cubra a ocorrência do termo "heteroarilo" onde não é designada uma gama numérica. Em algumas concretizações, o grupo heteroarilo tem 5 a 10 membros de anel ou 5 a 7 membros de anel. O grupo heteroarilo pode ser designado como "heteroarilo com 5-7 membros", "heteroarilo com 5-10 membros" ou designações semelhantes. Exemplos de anéis heteroarilo incluem, mas não estão limitados a, furilo, tienilo, ftalazinilo, pirrolilo, oxazolilo, tiazolilo, imidazolilo, pirazolilo, isoxazolilo, isotiazolilo, triazolilo, tiadiazolilo, piridinilo, piridazinilo, pirimidinilo, pirazinilo, triazinilo, quinolinilo, Benzimidazolilo, benzoxazolilo, benzotiazolilo, indolilo, isoindolilo e benzotienilo.

[0067] Um "heteroaralquilo" ou "heteroarilalquilo" é um grupo heteroarilo ligado, como um substituinte, através de um grupo alquíleno. Exemplos incluem, mas não estão limitados a 2-tienilmetilo, 3-tienilmetilo, furilmetilo, tieniletilo, pirrolilalquilo, piridilalquilo, isoxazolilalquilo e imidazolilalquilo. Em alguns casos, o grupo alquíleno é um grupo alquíleno inferior (isto é, um grupo alquíleno C1-4). Tal como aqui utilizado, "cicloalquilo" significa um anel ou sistema de anel carbociclico totalmente saturado. Exemplos incluem ciclopropilo, ciclobutilo, ciclopentilo e ciclohexilo.

[0068] Um grupo "O-carboxi" refere-se a um grupo "-OC(=O)R", em que R é selecionado de hidrogénio, alquiloC1-6, alquenil C2-6, alquinil C2-6, carbociclicoC3-7, Arilo C6-10, heteroarilo com 5-10 membros e heterociclico com 5-10 membros, como aqui definido.

[0069] Um grupo "C-carboxi" refere-se a um grupo "C(=O)OR" em que R é seleccionado de hidrogénio, , C<sub>1-6</sub> alquil,

$C_{2-6}$  alquenil,  $C_{2-6}$  alquinil,  $C_{3-7}$  carbociclico,  $C_{6-10}$  aril, heteroarilo de 5-10 membros e heterociclico de 5-10 membros, como aqui definido. Um exemplo não limitativo inclui carboxilo (i.e.,  $-C(=O)OH$ ).

[0070] Um grupo "ciano" refere-se a um grupo "CN".

[0071] Um grupo "azido" refere-se a um grupo "-N<sub>3</sub>".

[0072] Um grupo "O-carbamilo" refere-se a um grupo "OC(=O)NRARB", em que RA e RB são cada um independentemente seleccionados a partir de hidrogénio,  $C_{1-6}$  alquil,  $C_{2-6}$  alquenil,  $C_{2-6}$  alquinil,  $C_{3-7}$  carbociclico,  $C_{6-10}$  aril, heteroarilo de 5-10 membros e heterociclico de 5-10 membros, como definido aqui.

[0073] Um grupo "C-amido" refere-se a um grupo "C(=O)NRARB" em que RA e RB são cada um, independentemente, seleccionados a partir de hidrogénio, alquil  $C_{1-6}$ , alquenil  $C_{2-6}$ , alquinil  $C_{2-6}$ , carbociclico  $C_{3-7}$ , Aril  $C_{6-10}$ , heteroarilo com 5-10 membros e heterociclico com 5-10 membros, como aqui definido.

[0074] Um grupo "N amido" refere-se a um grupo "-N(RA)C(=O)RB" em que RA e RB são cada um independentemente seleccionados de hidrogénio, alquil  $C_{1-6}$ , alquenil  $C_{2-6}$ , carbociclico  $C_{3-7}$ , arilo  $C_{6-10}$ , heteroarilo com 5-10 membros e heterociclico com 5-10 membros, tal como aqui definido. Um grupo "amino" refere-se a um grupo "-NRARB" em que RA e RB são cada um independentemente seleccionados a partir de hidrogénio, alquilo  $C_{1-6}$ , alcenilo  $C_{2-6}$ , alcinilo  $C_{2-6}$ , carbociclico  $C_{3-7}$ , 10 arilo, heteroarilo de 5-10 membros e heterociclico de 5-10 membros, como aqui definido. Um exemplo não limitativo inclui amino livre (i.e.,  $-NH_2$ ).

[0075] Tal como aqui utilizado, o termo "Trolox" refere-se a ácido 6-hidroxi-2,5,7,8-tetrametilcroman-2-carboxílico.

[0076] Tal como aqui utilizado, o termo "ascorbato" refere-se ao sal de ácido ascórbico.

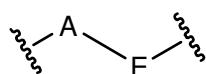
[0077] Tal como aqui utilizado, o termo "ácido gálico" refere-se a ácido 3,4,5-tri-hidroxibenzóico.

[0078] Tal como aqui utilizado, um grupo substituído é derivado do grupo pai não substituído em que houve uma troca de um ou mais átomos de hidrogénio para outro átomo ou grupo. Salvo indicação em contrário, quando se considera que um grupo é "substituído", significa que o grupo está substituído com um ou mais substituintes seleccionados independentemente entre alquilo C1-C6, alcenilo C1-C6, alcinilo C1-C6, heteroalquilo C1-C6, Carbociclilo C3-C7 (opcionalmente substituído com halo, alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), C3-C7- carbociclil-alquilo C1-C6 (opcionalmente substituído com halo, Alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), heterociclilo com 5-10 membros (opcionalmente substituído com halo, alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6), arilo (opcionalmente substituído com halo, alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), Alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), arilalquilo C1-C6 (opcionalmente substituído com halo, C1-C6alquilo, C1-C6alcoxi, C1-C6haloalquilo, C1-C6haloalcoxi), 5-10 membros Heteroarilo (opcionalmente substituído com halo, alquilo C1-C6, alcoxilo C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), heteroarilalquilo (C1-C6) 5-10 membros (opcionalmente substituído com halo, alquilo C1-C6, Alcoxi C1-C6, haloalquilo C1-C6 e haloalcoxi C1-C6), halo, ciano, hidroxilo, alcoxi C1-C6, alcoxi C1-C6 alquilo (isto é, éter), ariloxi, sulfidrilo (mercapto), Haloalquilo (C1-C6) (por exemplo, -CF3), haloalcoxi (C1-C6) (eg, -OCF3), alquiltio C1-C6, ariltio, amino, aminoalquilo C1-C6, nitro, O N-carbamilo, O-tiocarbamilo, N-tiocarbamilo, C-amido, N-amido, S-sulfonamido, N-sulfonamido, C-carboxilo, O-carboxilo, acilo, cianato, isocianato, tiocianato,

isotiocianato, sulfinilo, Sulfônico e oxo (= O). Sempre que um grupo é descrito como "opcionalmente substituído" esse grupo pode ser substituído com os substituintes acima.

[0079] Deve ser entendido que certas convenções de nomenclatura de radicais podem incluir um mono-radical ou um di-radical, dependendo do contexto. Por exemplo, quando um substituinte requer dois pontos de ligação ao resto da molécula, entende-se que o substituinte é um di-radical. Por exemplo, um substituinte identificado como alquilo que requer dois pontos de ligação inclui di-radicais tais como -CH<sub>2</sub>-, -CH<sub>2</sub>CH<sub>2</sub>-, -CH<sub>2</sub>CH (CH<sub>3</sub>) CH<sub>2</sub>- e semelhantes. De forma semelhante, um grupo identificado como amino que requer dois pontos de ligação inclui radicais di tais como -NH-, -N (CH<sub>3</sub>)- e semelhantes. Outras convenções de denominação radical indicam claramente que o radical é um di-radical tal como "alquileno" ou "alcenileno".

[0080] Sempre que um substituinte é representado como um di-radical (isto é, tem dois pontos de ligação ao resto da molécula), deve entender-se que o substituinte pode ser ligado em qualquer configuração direccional, a não ser que seja indicado o contrário. Assim, por exemplo, um

substituinte representado como -AE- ou  Inclui o substituinte orientado de tal modo que o A está ligado no ponto de ligação mais à esquerda da molécula bem como o caso em que A está ligado no ponto de ligação mais à direita da molécula.

[0081] Quando os compostos aqui revelados têm pelo menos um estereocentro, podem existir como enantiómeros e diastereómeros individuais ou como misturas de tais isômeros, incluindo racematos. A separação dos isômeros individuais ou a síntese seletiva dos isômeros individuais é conseguida pela aplicação de vários métodos que são bem conhecidos dos profissionais da técnica. Salvo indicação em contrário, todos

estes isômeros e suas misturas estão incluídos no âmbito dos compostos aqui revelados.

[0082] Tal como aqui utilizado, um "nucleotídeo" inclui uma base heterocíclica contendo azoto, um açúcar e um ou mais grupos fosfato. São unidades monoméricas de uma sequência de ácido nucleico. No RNA, o açúcar é uma ribose e no DNA uma desoxirribose, isto é, um açúcar que não possui um grupo hidroxilo que está presente na ribose. A base heterocíclica contendo azoto pode ser purina ou pirimidina base. As bases de purina incluem adenina (A) e guanina (G), e seus derivados modificados ou análogos. As bases de pirimidina incluem citosina (C), timina (T) e uracilo (U), e seus derivados modificados ou análogos. O átomo C-1 de desoxirribose está ligado a N-1 de uma pirimidina ou N-9 de uma purina.

[0083] Tal como aqui utilizado, um "nucleosídeo" é estruturalmente semelhante a um nucleotídeo, mas está faltando as porções fosfato. Um exemplo de um análogo nucleosídico seria aquele em que o marcador está ligado à base e não há nenhum grupo fosfato ligado à molécula de açúcar. O termo "nucleosídeo" é aqui utilizado no seu sentido corrente, tal como é entendido pelos especialistas na matéria. Exemplos incluem, mas não estão limitados a, um ribonucleosídeo compreendendo uma porção ribose e um desoxirribonucleosídeo compreendendo uma porção desoxirribose. Uma porção de pentose modificada é uma porção de pentose na qual um átomo de oxigénio foi substituído por um carbono e / ou um carbono foi substituído por um enxofre ou um átomo de oxigénio. Um "nucleosídeo" é um monómero que pode ter uma base substituída e / ou uma unidade de açúcar. Adicionalmente, um nucleosídeo pode ser incorporado em polímeros e oligómeros de DNA e / ou RNA maiores.

[0084] Tal como aqui utilizado, o termo "polinucleotídeo" refere-se a ácidos nucleicos em geral,

incluindo DNA (por exemplo DNAc de DNA genómico), RNA (por exemplo RNAm), oligonucleotídeos sintéticos e análogos de ácido nucleico sintéticos. Os polinucleotídeos podem incluir bases naturais ou não naturais, ou suas combinações e ligações estruturais naturais ou não naturais, e. Fosforotioatos, PNA ou 2'-O-metil-RNA, ou suas combinações.

[0085] Tal como aqui utilizado, o termo "faseamento" refere-se a fenômenos em SBS que são causados por remoção incompleta dos terminadores 3' e fluoróforos e falha na conclusão da incorporação de uma porção de cadeias de DNA dentro de clusters por polimerases a um dado Ciclo de sequenciamento. A pré-fase é causada pela incorporação de nucleotídeos sem extremidades 3' eficazes e o evento de incorporação vai um ciclo à frente. A fase e o pré-fase fazem com que as intensidades extraídas para um ciclo específico consistam no sinal do ciclo actual, bem como no ruído dos ciclos anteriores e seguintes. À medida que o número de ciclos aumenta, a fração de sequências por aglomerado afetado pelo phasing aumenta, dificultando a identificação da base correta. A pré-fase pode ser causada pela presença de uma quantidade traço de nucleotídeos 3'-OH desprotegidos ou não bloqueados durante a sequenciamento por síntese (SBS). Os nucleotídeos 3'-OH não protegidos podem ser gerados durante os processos de fabrico ou possivelmente durante os processos de armazenagem e manuseamento de reagentes. Consequentemente, as modificações de análogos de nucleotídeos ou os grupos de ligação que resultam num tempo de ciclo de SBS mais rápido, valores de fase mais baixos e pré-fase e comprimento de leitura de sequenciamento mais longo proporcionam maiores vantagens em aplicações de SBS.

[0086] Tal como aqui utilizado, o termo "grupo protetor" inclui, mas não se limita a, moléculas que podem proteger contra danos no DNA (por exemplo, danos por foto ou outros

danos químicos). Alguns exemplos específicos incluem antioxidantes, tais como vitamina C, derivados de vitamina E, ácido fenólico, polifenóis, e seus derivados e análogos. Deve ser entendido que em certos contextos onde o termo "grupo protector" é definido, refere-se à porção resultante da reacção entre um ou mais grupos funcionais da unidade protetora com o grupo funcional correspondente do ligante tal como aqui descrito. Por exemplo, quando a porção protetora é "ácido gálico", pode referir-se às amidas e ésteres do ácido gálico em vez do ácido gálico propriamente dito com o grupo carboxilo livre.

#### Marcadores detectáveis

[0087] Algumas concretizações aqui descritas referem-se à utilização de marcadores detectáveis convencionais. A detecção pode ser realizada por qualquer método adequado, incluindo espectroscopia de fluorescência ou por outros meios ópticos. O marcador preferido é um fluoróforo, que, após absorção de energia, emite radiação a um comprimento de onda definido. São conhecidos muitos marcadores fluorescentes adequados. Por exemplo, Welch et al. (Chem. Eur. J. 5 (3): 951-960, 1999) descreve frações fluorescentes funcionalizadas em dansilo que podem ser utilizadas na presente invenção. Zhu et al. (Citometria 28: 206-211, 1997) descreve o uso dos marcadores fluorescentes Cy3 e Cy5, que também podem ser utilizados na presente invenção. As etiquetas adequadas para utilização são também descritas em Prober et al. (Science 238: 336-341, 1987); Connell et al. (BioTechniques 5 (4): 342-384, 1987), Ansorge et al. (Nucl. Acids Res. 15 (11): 4593-4602, 1987) e Smith et al. (Nature 321: 674, 1986). Outros marcadores fluorescentes comercialmente disponíveis incluem, mas não estão limitados a, fluoresceína, rodamina (incluindo TMR, texas red e Rox), alexa, bodipy, acridina, cumarina, pireno, benzanceno e

as cianinas. Podem também ser utilizados vários marcadores na presente aplicação, por exemplo, cassetes bi-fluoróforo FRET (Tet. 46: 8867-8871, 2000). Podem também utilizar-se sistemas dendriméricos multi-fluor (J. Am. Chem. Soc. 123: 8101-8108, 2001). Embora os marcadores fluorescentes sejam preferidos, outras formas de marcadores detectáveis serão evidentes como úteis para os especialistas na técnica. Por exemplo, as micropartículas, incluindo os pontos quânticos (Empodocles et al., Nature 399: 126-130, 1999), nanopartículas de ouro (Reichert et al., Anal. Chem. 72: 6025-6029, 2000) e microesferas (Lacoste et al., Proc. Nat. Acad. Sci USA 97 (17): 9461- 9466, 2000) podem ser todos utilizados. Os rótulos de múltiplos componentes também podem ser utilizados na presente aplicação. Um marcador de múltiplos componentes é aquele que é dependente da interacção com um composto adicional para detecção. O marcador multi-componente mais utilizado na biologia é o sistema biotina-estreptavidina. A biotina é utilizada como o marcador ligado à base nucleotídica. A estreptavidina é então adicionada separadamente para permitir a detecção. Outros sistemas multi-componente estão disponíveis. Por exemplo, dinitrofenol tem um anticorpo fluorescente comercialmente disponível que pode ser utilizado para detecção. A menos que indicado de outro modo, a referência a nucleotídeos também se destina a ser aplicável a nucleosídeos. O presente pedido de patente será também descrito com referência ao DNA, embora a descrição seja também aplicável a RNA, PNA e outros ácidos nucleicos, a menos que indicado de outra forma.

#### Métodos de sequenciamento

[0088] Os nucleosídeos ou nucleotídeos aqui descritos podem ser utilizados em conjunto com uma variedade de técnicas de sequenciamento. Em algumas concretizações, o processo para determinar a sequência nucleotídica de um ácido nucleico alvo pode ser um processo automatizado.

[0089] Os análogos de nucleotídeos aqui apresentados podem ser utilizados num procedimento de sequenciamento, tal como uma técnica de sequenciamento por síntese (SBS). Resumidamente, a SBS pode ser iniciada por contacto dos ácidos nucleicos alvo com um ou mais nucleotídeos marcados, DNA polimerase, etc. As características em que um iniciador é estendido utilizando o ácido nucleico alvo como molde incorporarão um nucleotídeo marcado que pode ser detectado. Opcionalmente, os nucleotídeos marcados podem incluir adicionalmente uma propriedade de terminação reversível que termina a extensão do iniciador mais uma vez que um nucleotídeo tenha sido adicionado a um iniciador. Por exemplo, um análogo de nucleotídeos com uma porção terminadora reversível pode ser adicionado a um iniciador de tal modo que a extensão subsequente não pode ocorrer até que um agente de desbloqueio seja libertado para remover a porção. Assim, para concretizações que utilizam terminação reversível, um reagente de desbloqueamento pode ser fornecido à célula de fluxo (antes ou depois da detecção ocorrer). As lavagens podem ser realizadas entre os vários passos de fornecimento. O ciclo pode então ser repetido n vezes para prolongar o iniciador por n nucleotídeos, detectando assim uma sequência de comprimento n. Exemplos de procedimentos SBS, sistemas fluídicos e plataformas de detecção que podem ser prontamente adaptados para utilização com uma matriz produzida pelos métodos da presente divulgação são descritos, por exemplo, em Bentley et al., Nature 456: 53-59 (2008), WO 04/018497; WO 91/06678; WO 07/123744; US Pat. Nos. 7,057,026;

7,329,492; 7,211,414; 7,315,019 or 7,405,281, and US Pat. App. Pub. No. 2008/0108082 A1, cada um dos quais é aqui incorporado por referência.

[0090] Outros procedimentos de sequenciamento que utilizam reações cíclicas, tais como pirosequenciamento podem ser utilizados. O pirosequenamento detecta a liberação de pirofosfato inorgânico (PPi) à medida que são incorporados nucleotídeos particulares numa cadeia de ácido nucleico nascente (Ronaghi, et al., *Analytical Biochemistry* 242 (1), 84-9 (1996), Ronaghi, *Genome Res.*), 3-11 (2001), Ronaghi et al., *Science* 281 (5375), 363 (1998), US 6,210,891, 6,258,568 e 6,274,320, cada um dos quais aqui incorporado como referência). Na pirosequenciamento, o PPi libertado pode ser detectado por conversão em trifosfato de adenosina (ATP) por ATP sulfúrilase, e o ATP resultante pode ser detectado através de fótons produzidos por luciferase. Assim, a reação de sequenciamento pode ser monitorizada através de um sistema de detecção de luminescência. As fontes de radiação de excitação utilizadas para sistemas de detecção baseados em fluorescência não são necessárias para os procedimentos de pirosequenciamento. Sistemas de fluidos úteis, detectores e procedimentos que podem ser utilizados para a aplicação de pirosequenciamento a matrizes da presente divulgação estão descritos, por exemplo, no pedido de patente Ser. PCT / US11 / 57111, pedido de patente. Bar. 2005/0191698 A1, Pat. US 7,595,883 e Pat. 7,244,559, cada um dos quais é aqui incorporado por referência.

[0091] As reações de sequenciamento por ligação são também úteis incluindo, por exemplo, as descritas em Shendure et al. *Science* 309: 1728 - 1732 (2005); A Pat. 5 599 675; E Pat. 5 750 341, cada um dos quais é aqui incorporado por referência. Algumas concretizações podem incluir procedimentos de sequenciamento por hibridação como descrito,

por exemplo, em Bains et al., *Journal of Theoretical Biology* 135 (3), 303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251 (4995), 767 - 773 (1995); E WO 1989/10977, cada um dos quais é aqui incorporado por referência. Nos procedimentos de sequenciamento por ligação e sequenciamento por hibridação, os ácidos nucleicos que estão presentes em poços contendo gel (ou outras características côncavas) são sujeitos a ciclos repetidos de entrega e detecção de oligonucleotídeos. Os sistemas fluidos para os métodos de SBS, tal como aqui apresentados, ou nas referências aqui citadas, podem ser prontamente adaptados para distribuição de reagentes para procedimentos de sequenciamento por ligação ou sequenciamento por hibridização. Tipicamente, os oligonucleotídeos são marcados fluorescentemente e podem ser detectados utilizando detectores de fluorescência semelhantes aos descritos em relação aos procedimentos de SBS aqui descritos ou nas referências aqui citadas.

[0092] Algumas concretizações podem utilizar métodos que envolvem a monitoração em tempo real da atividade da DNA polimerase. Por exemplo, as incorporações de nucleotídeos podem ser detectadas através de interacções de transferência de energia de ressonância de fluorescência (FRET) entre uma polimerase portadora de fluoróforo e nucleotídeos marcados com  $\gamma$ -fosfato, ou com guias de onda zeromódicas. Técnicas e reagentes para a sequenciamento baseada em FRET são descritos, por exemplo, em Levene et al. *Science* 299, 682-686 (2003); Lundquist et al. *Optar. Lett.* 33, 1026-1028 (2008); Korlach et al. *Proc. Natl. Acad. Sei. USA* 105, 1176-1181 (2008), cujas descrições são aqui incorporadas por referência.

[0093] Algumas concretizações SBS incluem a detecção de um protão libertado após a incorporação de um nucleotídeo num

produto de extensão. Por exemplo, a sequenciamento com base na detecção de prótons liberados pode utilizar um detector elétrico e técnicas associadas que estão comercialmente disponíveis a partir de Ion Torrent (Guilford, CT, uma subsidiária Life Technologies) ou métodos e sistemas de sequenciamento descritos na Pat. Aplicativo. Bar. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; Ou 2010/0282617 A1, cada um dos quais é aqui incorporado por referência.

#### Ligantes modificados exemplificativos

[0094] Concretizações adicionais são reveladas com mais detalhes nos seguintes exemplos, os quais não são de modo algum destinados a limitar o escopo das reivindicações.

[0095] A FIG. 1A ilustra uma fórmula estrutural parcial de um nucleotídeo 100 marcado. O nucleotídeo 100 identificado inclui um nucleotídeo de adenosina (ffA) 110 completamente funcionalizado, uma porção de ligação padrão 115 e um corante fluorescente 120. A porção de ligação padrão 115 pode ser uma porção de ligação tipicamente utilizada na síntese de nucleotídeos marcados para sequenciamento por síntese (SBS). Num exemplo, o corante fluorescente 120 é NR550S0. Neste exemplo, o nucleotídeo 100 marcado pode ser descrito como "ffA-NR550S0". Em outro exemplo, o corante fluorescente 120 é S07181 e o nucleotídeo 100 marcado pode ser descrito como "ffA-S07181".

[0096] A FIG. 1B ilustra o nucleotídeo 100 marcado da FIG. 1A com duas modificações estruturais possíveis para a porção de ligação padrão 115. Num exemplo, a porção ligadora padrão 115 inclui um inserto AEDI 125 entre a porção carbonilo (isto é, -C (= O) - e a porção amino (isto é, -NH-) Da porção amido. Neste exemplo, o nucleotídeo marcado modificado pode ser descrito como "ffA-AEDI-NR550S0". Noutro exemplo, a porção de ligador padrão 115 inclui um inserto SS 130 e o

nucleotídeo marcado modificado pode ser designado "ffa-SS-NR550S0".

[0097] A FIG. 2 mostra um gráfico da taxa de incorporação de nucleotídeos usando o nucleotídeo 100 da FIG. 1A e os nucleotídeos 100 marcados modificados da FIG. 1B. O ensaio foi realizado a 55 ° C, em etanolamina 40 mM (pH 9,8), MgCl<sub>2</sub> 9 mM, NaCl 40 mM, EDTA 1 mM, CHAPS 0,2% com iniciador 20nM: DNA molde e 30 µg / ml de Polimerase 812 (MiSeq Kit V2) Nucleotídeo 1 mM. A enzima é ligada a DNA e depois rapidamente misturada com nucleotídeo numa máquina de fluxo de arrefecimento durante um curto período de tempo (até 10s) antes da extinção com EDTA a 500 mM. Vários pontos de tempo são tomados para cada nucleotídeo. As amostras geradas são então analisadas num gel desnaturante e a percentagem de DNA que está a ser convertido em DNA + 1 é determinada e traçada em função do tempo para determinar a constante de velocidade de primeira ordem para cada nucleotídeo. Os dados estão resumidos na Tabela 1 abaixo. Os dados mostram que a taxa de incorporação para um nucleotídeo marcado compreendendo o inserto SS 130 (ffa-SS-NR550S0) era cerca de 2 vezes mais rápida em comparação com a taxa de incorporação de um nucleotídeo marcado compreendendo um ligante padrão 115 (ffa-NR550S0). A taxa de incorporação para um nucleotídeo marcado compreendendo um inserto AEDI 125 (ffa-AEDI-NR550S0) foi cerca de 4 vezes mais rápida em comparação com a taxa de incorporação de ffa-NR550S0. Os dados também mostram que a taxa de incorporação de um nucleotídeo marcado compreendendo o ligante padrão 115 e o corante fluorescente S07181 (ffa-S07181) era cerca de 4 vezes mais rápida comparada com a taxa de incorporação de ffa-NR550S0.

**Tabela 1.**

ffa	K (µM/min)
ffa-NR550S0	22 (1X)

ffA-SS-NR550S0	54 (2X)
ffA-AEDI- NR550S0	97 (4X)
ffA-S07181	99 (4X)

[0098] As FIGs. 3A-3F ilustram as fórmulas estruturais de inserções adicionais 310, 315, 320, 325, 330 e 335 para a porção de ligação padrão 115 da FIG. 1A. Numa inserção ACA 310, as substituições dimetilo são removidas e a ligação enxofre-enxofre (SS) é substituída por ligação carbono-carbono em comparação com o inserto 125. A ligação enxofre-enxofre não é necessária para SBS (por exemplo 2-corante ou 4 (SBS)).

[0099] Numa inserção AcLys 315, utiliza-se lisina acetil protegida para substituir o inserção 125.

[00100] Numa inserção 320 de BocLys, utiliza-se lisina protegida com terc-butoxicarbonilo para substituir o inserção 125.

[00101] Numa inserção 325 de dMeO, a ligação enxofre-enxofre (S-S) é substituída por ligação oxigénio-carbono (O-CH<sub>2</sub>) em comparação com o inserção 125.

[00102] Numa inserção dMeS 330, a ligação enxofre-enxofre (S-S) é substituída por ligação enxofre-carbono (S-CH<sub>2</sub>) em comparação com o inserção 125.

[00103] Numa inserção DMP 335, a ligação enxofre-enxofre (S-S) é substituída por ligação enxofre-carbono (CH<sub>2</sub>) em comparação com o inserção 125.

[00104] Em vários exemplos aqui descritos, verificou-se que as inserções incluindo um padrão de substituição dimetilo (por exemplo, inserção AEDI 125, inserção dMeO 325 e inserção dMeS 330) tinham uma taxa aumentada de incorporação de nucleotídeos durante SBS.

[00105] Em vários exemplos aqui descritos, o comprimento das cadeias de carbono nas inserções pode também ser variado.

[00106] A FIG. 4 mostra uma tabela de dados de uma corrida de sequenciamento de dois corantes utilizada para avaliar o efeito do inserção AEDI 125 da FIG. 1B e inserção AcLys 315 da FIG. 3B sobre a qualidade da sequenciamento. O sequenciamento foi executado em uma plataforma híbrida Miseq com modelo 550bp humano e 2 vezes 150 ciclos. O novo conjunto de corantes, V10 / ciano-peg4 A-AEDI550S0, V10 / ciano-peg4 A-AcLys550S0 e V10 / ciano A-AcLys 550S0 foram comparados com o conjunto de corante comercial standard V4 e um conjunto de corante melhorado da plataforma Nova V5 .75. Para cada uma das amostras V10 / ciano-peg4 A-AEDI550S0, V10 / ciano-peg4 A-AcLys550S0 e V10 / ciano A-AcLys 550S0, o valor de faseamento (Ph R1) foi inferior aos valores de faseamento das amostras sem o AEDI ou Inserções AcLys. Por conseguinte, os nucleotídeos marcados compreendendo inserções adicionais 125 e 315 demonstraram melhorias na qualidade de sequenciamento.

[00107] A FIG. 5A e 5B mostram um gráfico das taxas de erro para a leitura 1 e um gráfico das taxas de erro para a leitura 2 da sequência de sequenciamento da FIG. 4 respectivamente. Para a leitura 1, as taxas de erro de V10 / ciano-peg4 A-AEDI550S0 e V10 / ciano-peg4 A-AcLys550S0 eram inferiores ao mesmo conjunto de corante V10 / Cian-peg4 sem o inserção. Para a leitura 2, é ainda mais pronunciada quando se utiliza a inserção AcLys 315, onde a taxa de erro final foi reduzida em 30% em comparação com o conjunto de corantes sem inserção. Deste modo, as inserções 125 e 315 provaram melhorar significativamente a qualidade da sequenciamento.

[00108] A FIG. 6 mostra uma tabela de dados de uma corrida de sequenciamento utilizada para avaliar o efeito do inserção AEDI 125 e da inserção ACA 310 sobre a qualidade de sequenciamento. A sequenciamento foi executada em uma plataforma híbrida Miseq com modelo 550bp humano e 2 vezes 150 ciclos. O novo conjunto de corantes, V10 / ciano-peg4 A-

AEDI550S0, V10 / ciano-peg4 A-ACALys550S0, foi comparado com o conjunto de corante comercial standard V4 e um conjunto de corante melhorado da plataforma Nova V5.75. Mais uma vez, cada uma das amostras de V10 / ciano-peg4 A-AEDI550S0 e V10 / ciano-peg4 A-ACA550S0 tem um menor valor de fase (Ph R1) em comparação com amostras sem os inserções AEDI ou ACA e assim demonstrou uma melhoria na qualidade do sequenciamento.

[00109] A FIG. 7A e 7B mostram um gráfico das taxas de erro para a leitura 1 e um gráfico das taxas de erro para a leitura 2 da corrida de sequenciamento da FIG. 6 respectivamente. As taxas de erro para a leitura 1 para o conjunto que compreende a nova inserção AEDI (V10 / ciano-peg4 A-AEDI550S0), era inferior ao mesmo conjunto de corante sem o inserção, V10 / Cyan-peg4. A inserção ACA (V10 / Cyan-peg4 ACA550S0) deu um gráfico da taxa de erro semelhante em ambos os lidos para o padrão V10 / Cyan-peg4. A AEDI mostrou novamente para melhorar a qualidade do seqüenciamento. Esses dados mostraram também a estrutura da própria inserção está influenciando na melhora da qualidade de sequenciamento.

[00110] A FIG. 8A mostra a fórmula estrutural de um ligador LN3 padrão 800. O ligante LN3 800 inclui uma primeira porção funcional amido 810 substituída, uma segunda porção funcional PEG 815 substituída por azido e uma terceira porção funcional éster 820 que pode ser desejada numa estrutura ligante para Ligando uma molécula de corante 825 a um nucleótido 830. A primeira porção funcional 810 pode, por exemplo, ser utilizada para ligar a molécula de corante 825 ao ligador 800 de LN3. A segunda porção funcional 815 pode, por exemplo, ser um grupo funcional clivável que pode ser utilizado para Clivam a molécula de corante 825 a partir do ligador LN3 800. A terceira porção funcional 820 pode, por exemplo, ser utilizada para ligar o nucleótido 830 ao ligador 800 de LN3. 8B ilustra alguma modificação ao ligador LN3

padrão onde a porção fenoxi 850 é substituída com um a quatro substituintes seleccionados de -NO<sub>2</sub>, -CN, halo ou -SO<sub>3</sub>H. Além disso, a porção éster 820 é substituída por uma porção amido 855.

[00111] A Figura 9A é um gráfico que mostra o aparecimento de uma impureza num ffa com ligante SS. Com ffa-SS-NR550S0 após purificação por HPLC: a impureza apareceu durante a noite numa condição ligeiramente básica (pH 8-9). R.t. Durante a noite em TEAB / CH<sub>3</sub>CN 0,1 M. A Figura 9B compara a estabilidade de ffAs com ligante SS e ligante AEDI, com o ligante AEDI mostrando estabilidade significativamente melhorada em comparação com ligantes SS. Utilizou-se ffa-LN3-NR550S0 como referência interna. O produto secundário dissulfureto: (NR550S0-S-) 2. A Figura 9C mostra uma comparação do ligante SS e do ligante AEDI ffAs em IMX 60° durante 22 horas, mostrando novamente uma impureza com o ligante SS. O controlo interno é ffa-LN3-NR550S0. Di-P: difosfato.

[00112] As Figuras 10A, 10B e 10C mostram o aumento inesperado da velocidade de incorporação de nucleotídeos em solução com alterações do ligante. A Fig. 10A mostra a velocidade de incorporação a 1 uM. Os resultados mostram que o corante e o ligante tinham um efeito significativo na cinética de incorporação com a tabela 10B mostrando claramente o benefício do ligante AEDI na cinética de incorporação. A Figura 10C ilustra esquematicamente os ligantes AEDI e SS com NR550S0.

[00113] A Figura 11A mostra gráficos de dispersão para combinações V10 com diferentes A-550S0 (mesma concentração). As parcelas de dispersão são para a telha 1 ciclo 2. A Figura 11B mostra Kcat FFA Ligantes em solução. Pode-se ver que na superfície em termos de taxa de incorporação o ligante AEDI é mais rápido do que nenhum ligante: a nuvem A (no gráfico

de dispersão) ligeiramente movendo-se em direcção ao centro. Ligador AcLys é mais lento que AEDI e nenhum linker: 'A'-nuvem movendo-se em direção ao eixo x. BocLys linker parece semelhante a nenhum linker e não muito longe do linker AEDI. Em solução, pode ver-se que o ligante ACA era mais lento, enquanto AEDI e ACLys têm Kcat semelhante, seguido por BocLys.

[00114] As Figuras 12A e 12B mostram métricas de sequenciamento em ciclos M111, Human550, 2x151. Use ffNs em combinação com diferentes A-550S0 (mesma concentração). Pode ver-se que tanto o AEDI como o ligante BocLys proporcionaram resultados de sequenciamento semelhantes. Embora a solução Kcat de AEDI e AcLys sejam semelhantes, o resultado de sequenciamento de AEDI é ligeiramente melhor.

[00115] O ligante LN3 800 inclui uma porção fenoxi opcional 835 que pode ser removida do ligante LN3 800 como mostrado na FIG. 8C. O ligante LN3 800 inclui também uma porção amida opcional 840 e uma porção éter opcional 845, podendo ambos ser removidos do ligante LN3 800 como mostrado na FIG. 8D. O objectivo de remover certos grupos funcionais como a porção amido 810, a fracção fenoxi 835 é testar se eles têm interacções negativas com a enzima durante a incorporação de nucleotídeos, o que pode reduzir a eficiência de incorporação.

[00116] A FIG. 8E ilustra a inserção ou adição de uma parte protetora 860 no ligante. A porção protetora 860 é inserida entre a porção funcional 810 (ou pode estar ligada à porção fenoxi 835 ou 850 na Figura 8A-8C) e a molécula 825 de corante. A parte protetora 860 pode, por exemplo, ser uma molécula que proteja contra danos no ADN. Danos no DNA incluindo danos por foto ou outros danos químicos são um dos efeitos cumulativos (isto é, ciclo por ciclo) de SBS. Substancialmente reduzindo ou eliminando danos no DNA pode fornecer SBS mais eficiente e leituras de seqüenciamento mais

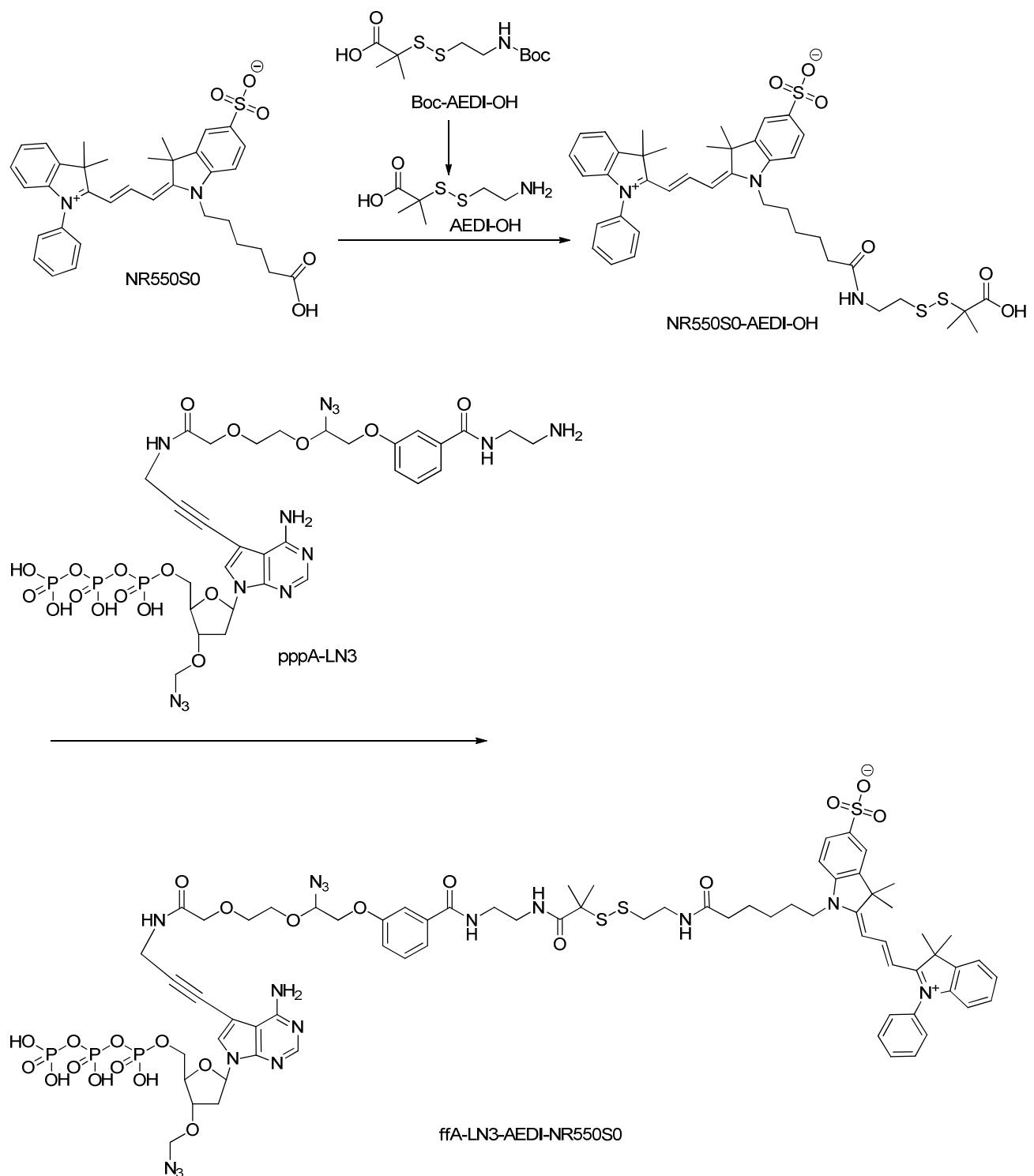
longas. Em algumas concretizações, a parte protetora 860 pode ser seleccionada a partir de um inibidor de estado de tripleto tal como Trolox, ácido gálico, 2-mercaptoetanol (BME), etc. Em algumas outras concretizações, a unidade protetora 860 pode ser seleccionada a partir de um reagente de extinção ou protector tal como 4 -nitrobenzílico ou um sal de ácido ascórbico, tal como ascorbato de sódio. Em algumas outras formas de realização, um agente de protecção pode ser fisicamente misturado no tampão em vez de formar ligação covalente com o nucleósido ou nucleótido marcado. No entanto, esta abordagem pode requerer uma concentração mais elevada do agente de protecção e pode ser menos eficiente. Como alternativa, a porção protetora ligada covalentemente aos nucleósidos ou nucleotídeos pode proporcionar uma melhor protecção contra danos ao ADN. Em algumas formas de realização adicionais da FIG. 8C, 8D e 8E, a porção éster 820 também pode ser substituída com a porção amido 855 e a porção fenoxy 835 ser adicionalmente substituída.

[00117] Em qualquer dos exemplos demonstrados na FIG. 8A-8E, inserção AEDI 125 e inserção SS 130 da FIG. 1B e inserções 300 das FIGs. 3A-3F pode, por exemplo, ser inserido entre a primeira porção funcional 810 e a molécula de corante 825 do ligante 800.

#### EXEMPLOS

[00118] Concretizações adicionais são reveladas com mais detalhes nos seguintes exemplos, os quais não são de modo algum destinados a limitar o escopo das reivindicações.

Procedimento Geral de Reação para Preparação de ffa-LN3-AEDI-NR550S0:



[00119] Num balão de fundo redondo de 50 ml, dissolveu-se Boc-AEDI-OH (1 g, 3,4 mmol) em DCM (15 mL) e adicionou-se TFA (1,3 mL, 17 mL) à ta. A mistura reaccional foi agitada durante 2 horas. A TLC (DCM: MeOH = 9: 1) indicou o consumo completo de Boc-AEDI-OH. A mistura reaccional foi evaporada

até à secura. Foi então adicionado TEAB (2 M, ~ 15 ml) ao resíduo e o pH foi monitorizado até ficar neutro. A mistura foi então dissolvida em H<sub>2</sub>O / CH<sub>3</sub>CN (1: 1, ~ 15 ml) e evaporada até à secura. Repetiu-se o procedimento durante 3 vezes para remover o excesso de sal de TEAB. Tratou-se o resíduo sólido branco com CH<sub>3</sub>CN (20 ml) e agitou-se durante 0,5 hora. Filtrou-se a solução e lavou-se o sólido com CH<sub>3</sub>CN, obteve-se sal de AEDI-OH-TFA puro (530 mg, 80%). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O, δ(ppm)): 3.32 (t, J=6.5 Hz, 2H, NH<sub>2</sub>-CH<sub>2</sub>); 2.97 (t, J=6.5 Hz, 2H, S-CH<sub>2</sub>); 1.53 (s, 6H, 2x CH<sub>3</sub>). <sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O, δ(ppm)): 178.21 (s, CO); 127.91, 117.71 (2s, TFA); 51.87 (s, C-(CH<sub>3</sub>)<sub>2</sub>); 37.76 (t, CH<sub>2</sub>-NH<sub>2</sub>); 34.23 (t, S-CH<sub>2</sub>); 23.84 (q, 2x CH<sub>3</sub>). <sup>19</sup>F NMR (400 MHz, D<sub>2</sub>O, δ(ppm)): -75.64.

[00120] Num balão de fundo redondo de 50 ml, o corante NR550S0 (114 mg, 176 umol) foi dissolvido em DMF (anidro, 20 mL) e evaporado para secar. Repetiu-se o procedimento durante 3 vezes. DMA anidra (10 mL) e base de Hunig (92 μl, 528 μmol, 3 equivalentes) foi então pipetada para o balão de fundo redondo. Adicionou-se TSTU (69 mg, 228 umol, 1,3 equivalentes) numa porção. A mistura reaccional foi mantida à ta. Após 30 min, a análise de TLC (CH<sub>3</sub>CN: H<sub>2</sub>O = 85: 15) indicou que a reacção estava completa. Adicionou-se AEDI-OH (68 mg, 352 μmol, 2 equivalentes) em TEAB 0,1 M à mistura reaccional e agitou-se à temperatura ambiente durante 3 h. A TLC (CH<sub>3</sub>CN: H<sub>2</sub>O = 8: 2) mostrou o consumo completo do éster activado e apareceu uma mancha vermelha abaixo do éster activado. Entretanto, a HPLC analítica indicou também o consumo completo do éster activado e a formação do produto. A reacção foi desactivada com tampão TEAB (0,1 M, 10 ml) e o solvente volátil foi removido por evaporação sob pressão reduzida (HV) e purificado em coluna Axia para obter NR550S0-AEDI-OH. Rendimento: 60%.

[00121] Num balão de fundo redondo de 25 ml, dissolveu-se NR550S0-AEDI-OH (10 umol) em DMF (anidro, 5 mL) e evaporou-se para secar. Repetiu-se o procedimento durante 3 vezes. Adicionou-se então DMA anidro (5 mL) e DMAP (1,8 mL, 15  $\mu$ mol, 1,5 equivalentes) ao balão de fundo redondo. Adicionou-se DSC (5,2 mg, 20  $\mu$ mol, 2 equivalentes) em uma porção. A mistura reacional foi mantida à temperatura ambiente. Após 30 min, a análise de TLC (CH<sub>3</sub>CN: H<sub>2</sub>O = 8:2) indicou que a reacção estava completa. A base de Hunig (3,5  $\mu$ l, 20 umol) foi adicionada à mistura reacional. Depois adicionou-se à mistura reacional uma solução de pppA-LN3 (20  $\mu$ mol em 0,5 ml de H<sub>2</sub>O, 2 equivalentes) e Et<sub>3</sub>N (5  $\mu$ l) e agitou-se à temperatura ambiente durante a noite. A TLC (CH<sub>3</sub>CN: H<sub>2</sub>O = 8: 2) mostrou o consumo completo do éster activado e apareceu uma mancha vermelha na linha de base. Entretanto, a HPLC analítica indicou também o consumo completo do éster activado e a formação do produto final. A reacção foi desactivada com tampão TEAB (0,1 M, 10 ml) e carregada numa coluna DEAE Sephadex (25 g de coluna Biotage). A coluna foi eluída com gradiente como mostrado na Tabela 2 abaixo.

A: tampão TEAB 0,1 M (CH<sub>3</sub>CN a 10%)

B: tampão TEAB 1 M (CH<sub>3</sub>CN a 10%).

Gradiente:

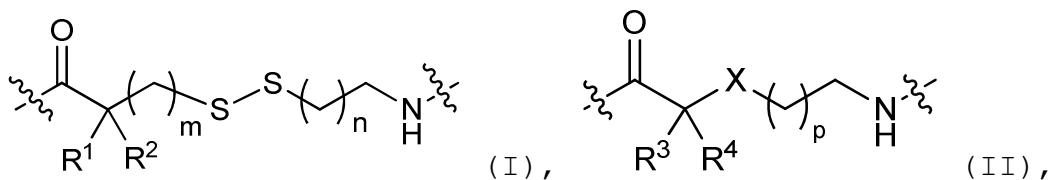
Tabela 2

ETAPA	Mix de solventes (B%)	Comprimento (ml)
1	0	100
2	0 - 45	50
3	45	100
4	45 - 100	50
5	100	100

[00122] O produto desejado foi eluído para fora de 45% para 100% de tampão TEAB 1 M. A fracção que contém o produto

foi combinada, evaporada e purificada por HPLC (coluna YLC, 8 mL / min). Rendimento: 53%.

[00123] Em resumo, a presente invenção pode referir-se a um nucleósido ou nucleótido ligado covalentemente a um fluoróforo através de um ligante, em que o referido ligante compreende uma estrutura de fórmula (I) ou (II), ou uma combinação de ambos.



em que

Cada  $R^1$  e  $R^2$  é independentemente selecionado de hidrogênio ou alquil C1-6 opcionalmente substituído;

R3 é selecionado de hidrogênio, alquil C1-6  
opcionalmente substituído, -NR5-C(=O)R6, ou -NR7-C(=O)-OR8;

R4 é selecionado de hidrogênio ou alquil C1-6  
opcionalmente substituído;

Cada  $R^5$  e  $R^7$  é independentemente selecionado de hidrogênio, alquil C1-6 opcionalmente substituído, fenil opcionalmente substituído ou aralquil C7-12 opcionalmente substituído;

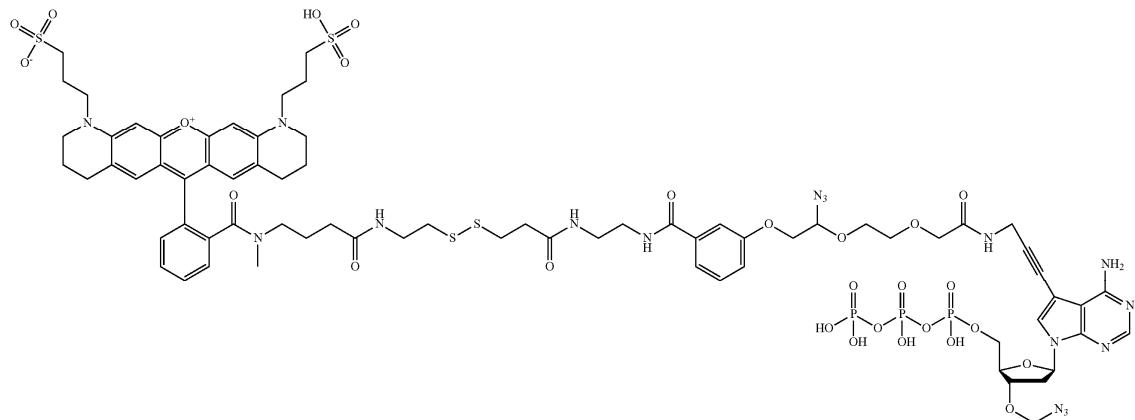
Cada R<sup>6</sup> e R<sup>8</sup> é independentemente selecionado de alquil C1-6 opcionalmente substituído, fenil opcionalmente substituído, aralquil C7-12 opcionalmente substituído, cicloalquil C3-7 opcionalmente substituído ou heteroarilo opcionalmente substituído com 5 a 10 membros;

cada uma das unidades repetitivas de metileno no  $\text{CH}_2(\text{CH}_2)_m$ ,  
 $\text{CH}_2(\text{CH}_2)_n$  ou  $\text{CH}_2(\text{CH}_2)_p$  é opcionalmente substituído;

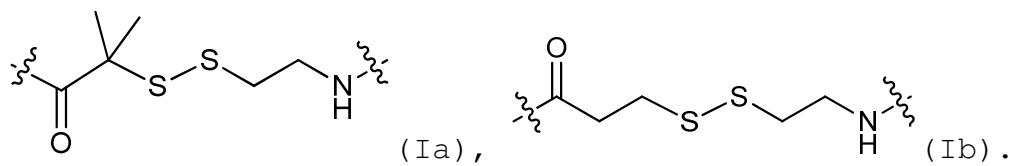
X é selecionado a partir de metileno ( $\text{CH}_2$ ), oxigênio (O) ou enxofre (S);

m é um número inteiro de 0 a 20;

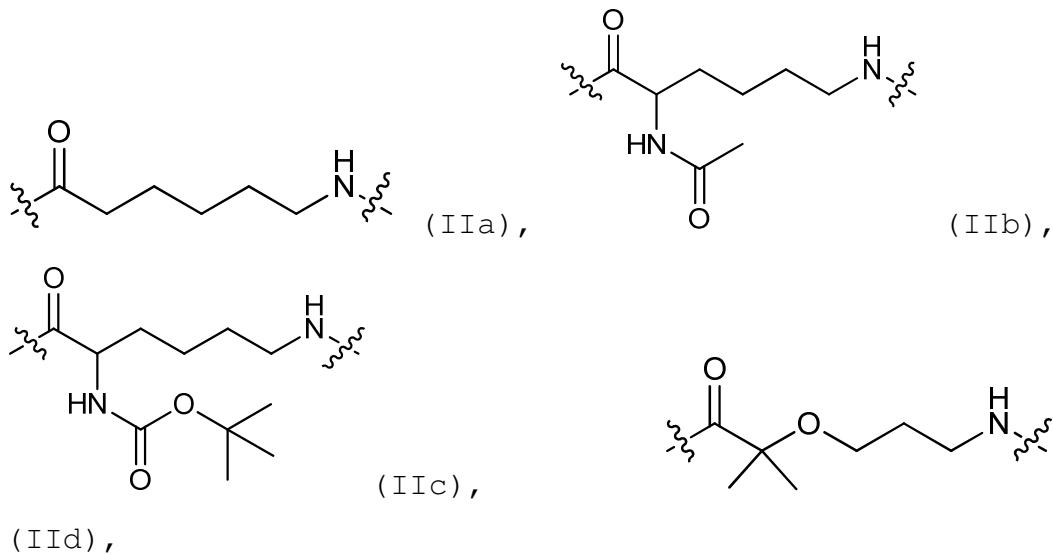
*n* é um número inteiro de 1 a 20; e  
*p* é um número inteiro de 1 a 20, desde que o nucleosídeo  
ou nucleotídeo marcado com fluoróforo não tenha a estrutura.

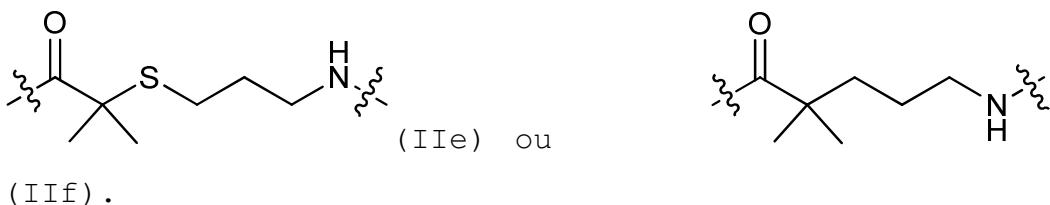


[00124] Em alguns casos, o nucleosídeo ou nucleotídeo referido acima, da estrutura de fórmula (I) também é representada pela fórmula (Ia) ou (Ib):

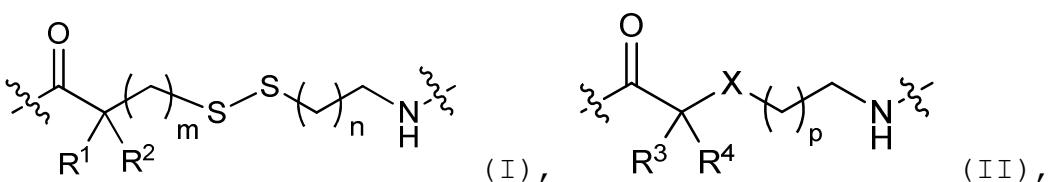


[00125] Além disso, a estrutura de fórmula (II) também pode ser representada pela fórmula (IIa), (IIb), (IIc), (IId), (IIe) ou (IIIf):





[00126] Mais particularmente, a presente invenção pode referir-se a um nucleosídeo ou nucleotídeo ligado covalentemente a um fluoróforo, através de um ligante, em que o referido ligante compreende uma estrutura de fórmula (I) ou (II), ou uma combinação de ambos:



em que:

$R^1$  é selecionado de entre hidrogênio ou  $C_{1-6}$  alquil opcionalmente substituído;

$R^2$  é selecionado de entre hidrogênio ou  $C_{1-6}$  alquil opcionalmente substituído;

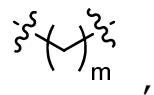
$R^3$  é selecionado de entre hidrogênio,  $C_{1-6}$  alquil opcionalmente substituído,  $-NR^5-C(=O)R^6$ , ou  $NR^7-C(=O)-OR^8$ ;

$R^4$  é selecionado de entre hidrogênio ou  $C_{1-6}$  alquil opcionalmente substituído;

cada  $R^5$  e  $R^7$  é selecionado, independentemente, de entre hidrogênio,  $C_{1-6}$  alquil opcionalmente substituído, fenil opcionalmente substituído, ou  $C_{7-12}$  aralquil opcionalmente substituído;

cada  $R^6$  e  $R^8$  é selecionado, independentemente, de entre  $C_{1-6}$  alquil opcionalmente substituído, fenil opcionalmente substituído,  $C_{7-12}$  aralquil opcionalmente substituído,  $C_{3-7}$  cicloalquil opcionalmente substituído, ou heteroaril opcionalmente substituído de 5 a 10 membros;

cada um da unidade de repetição de metileno no



$\text{---}(\text{---})_n\text{---}$  ou  $\text{---}(\text{---})_p\text{---}$  é opcionalmente substituído;

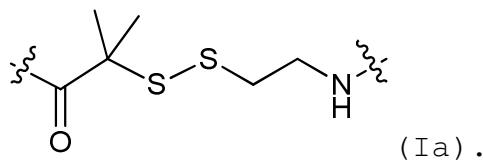
X é selecionado a partir de metileno ( $\text{CH}_2$ ), oxigênio (O) ou enxofre (S);

m é um número inteiro de 0 a 20;

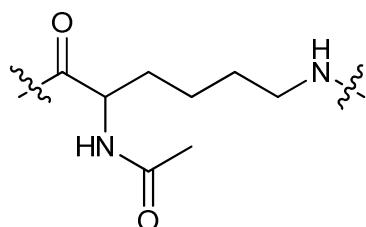
n é um número inteiro de 1 a 20; e

p é um número inteiro de 1 a 20.

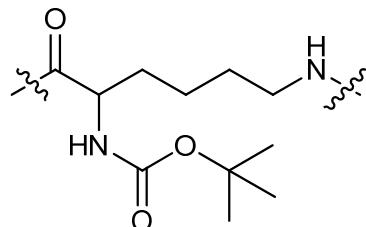
[00127] É preferido que a estrutura de fórmula (I) também seja representada pela fórmula (Ia):



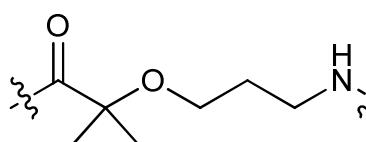
[00128] Além disso, a estrutura de fórmula (II) também é representada pela fórmula (IIb), (IIc), (IId), (IIe) or (IIIf):



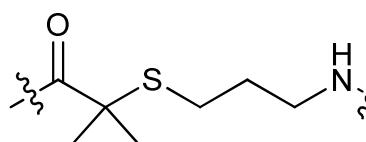
(IIb),



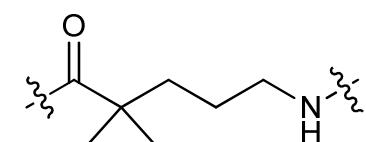
(IIc),



(IId),



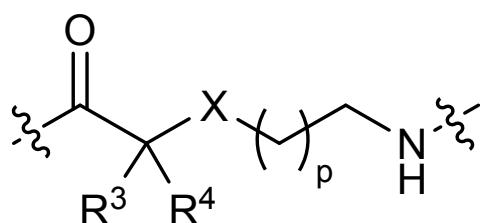
(IIe) or



(IIIf).

### REIVINDICAÇÕES

1. Nucleosídeo ou nucleotídeo ligado covalentemente a um fluoróforo através de um ligante **caracterizado pelo** fato de que o referido ligante compreende uma estrutura de fórmula (II) :



(II)

em que:

R<sup>3</sup> é selecionado de entre C<sub>1-6</sub> alquil opcionalmente substituído, -NR<sup>5</sup>-C(=O)R<sup>6</sup>, ou -NR<sup>7</sup>-C(=O)-OR<sup>8</sup>;

R<sup>4</sup> é selecionado de entre hidrogênio ou C<sub>1-6</sub> alquil opcionalmente substituído;

cada R<sup>5</sup> e R<sup>7</sup> é selecionado, independentemente, de entre hidrogênio ou C<sub>1-6</sub> alquil opcionalmente substituído,

cada R<sup>6</sup> and R<sup>8</sup> é independentemente C<sub>1-6</sub> alquil opcionalmente substituído;

cada um da unidade de repetição de metileno no,

$\left\{ \text{---} \right\}_p$  é opcionalmente substituído;

X é selecionado a partir de metileno (CH<sub>2</sub>), oxigênio (O) ou enxofre (S); e

p é um número inteiro de 1 a 20.

2. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 1, **caracterizado pelo** fato de que  $R^3$  é  $C_{1-6}$  alquil opcionalmente substituído ou metil.

3. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 1, **caracterizado pelo** fato de que  $R^3$  é  $-NH-C(=O)R^6$ .

4. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 3, **caracterizado pelo** fato de que  $R^6$  é metil.

5. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 1, **caracterizado pelo** fato de que  $R^3$  é  $-NH-C(=O)OR^8$ .

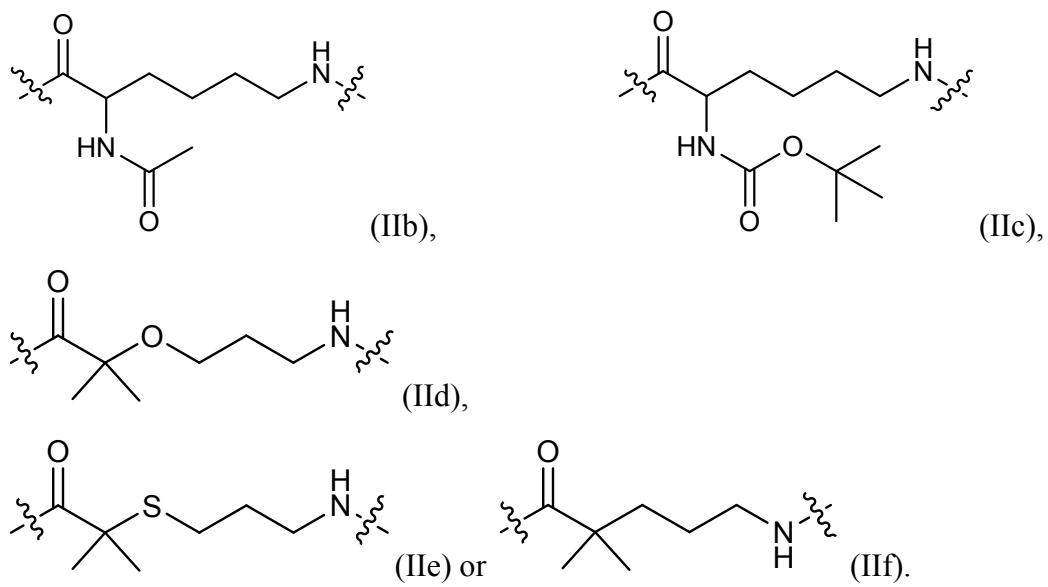
6. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 5, **caracterizado pelo** fato de que  $R^8$  é *t*-butil.

7. Nucleosídeo ou nucleotídeo, de acordo com qualquer uma das reivindicações 1 a 6, **caracterizado pelo** fato de que  $R^4$  é hidrogênio.

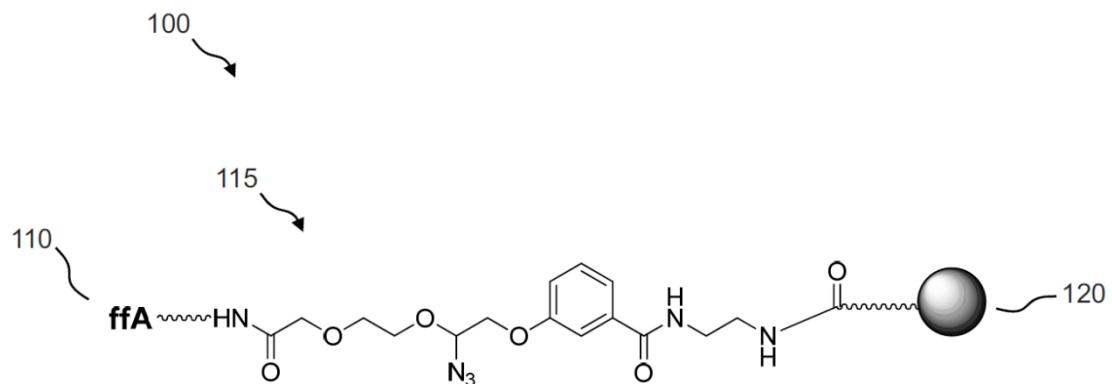
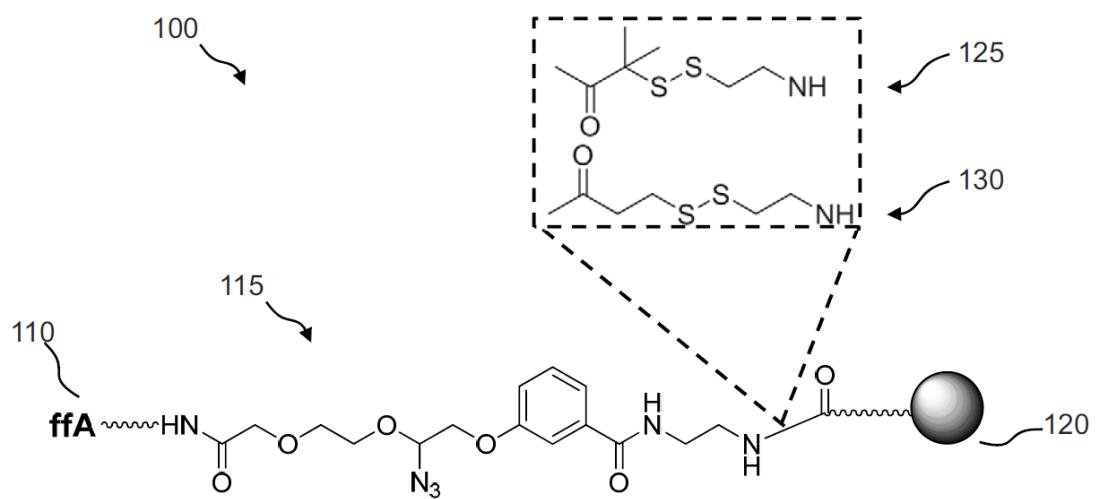
8. Nucleosídeo ou nucleotídeo, de acordo com qualquer uma das reivindicações 1 a 6, **caracterizado pelo** fato de que  $R^4$  é  $C_{1-6}$  alquil opcionalmente substituído ou metil.

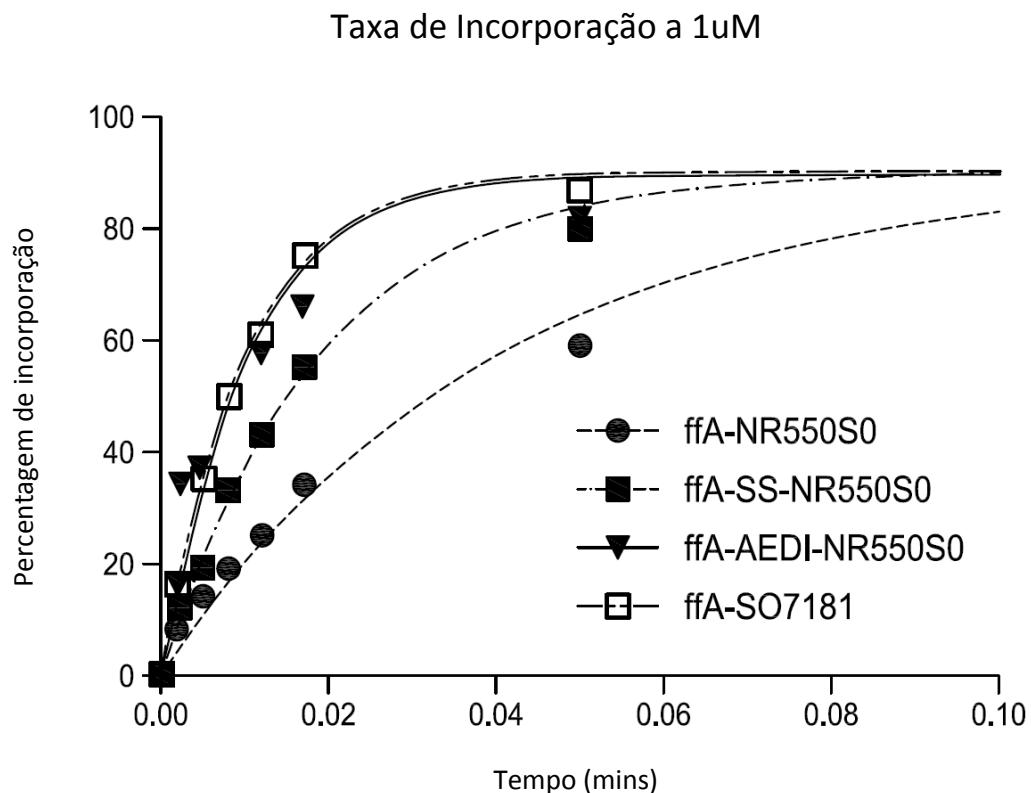
9. Nucleosídeo ou nucleotídeo, de acordo com qualquer uma das reivindicações 1 a 8, **caracterizado pelo** fato de que  $p$  é 1 ou 2.

10. Nucleosídeo ou nucleotídeo, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a estrutura da fórmula (II) é também representada pelas fórmulas (IIb), (IIc), (IId), (IIe) or (IIIf):

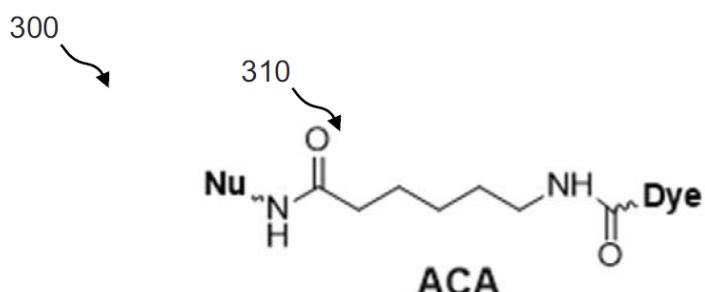
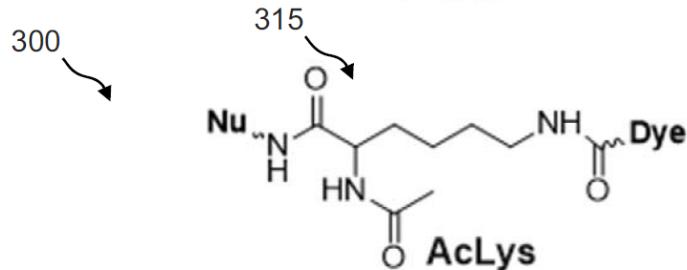
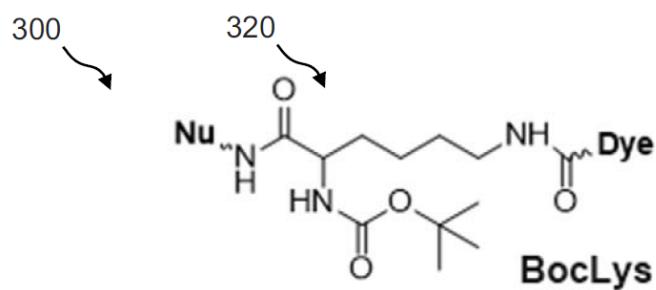
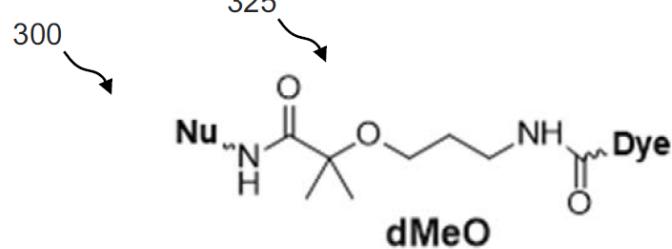
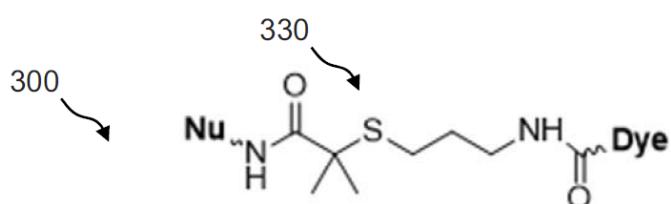
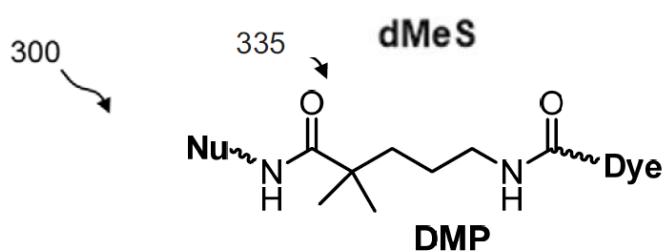


11. Kit **caracterizado pelo** fato de que compreende um nucleosídeo ou nucleotídeo conforme definido por qualquer uma das reivindicações 1 a 10.

**FIG. 1A****FIG. 1B**

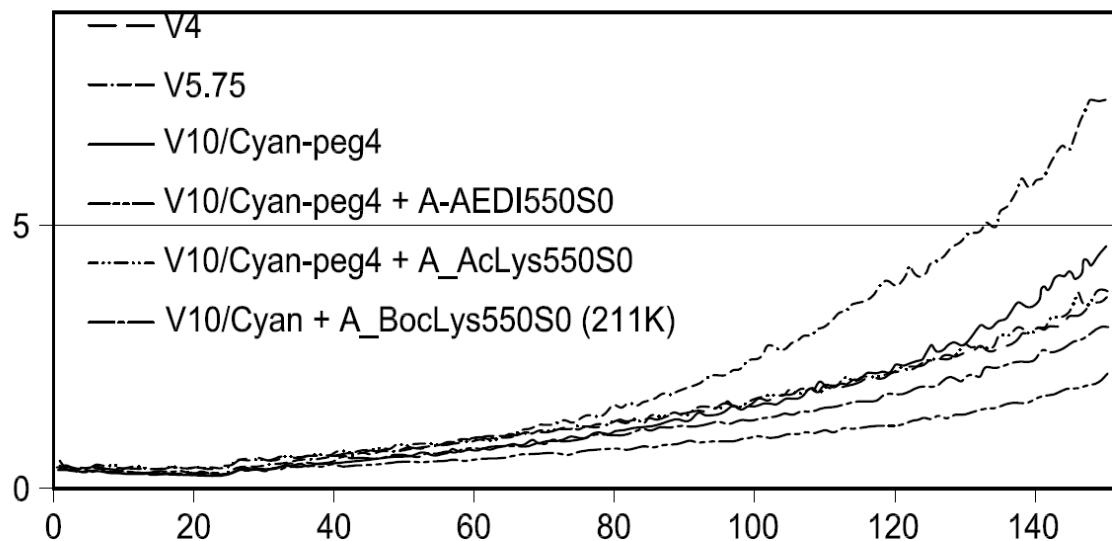
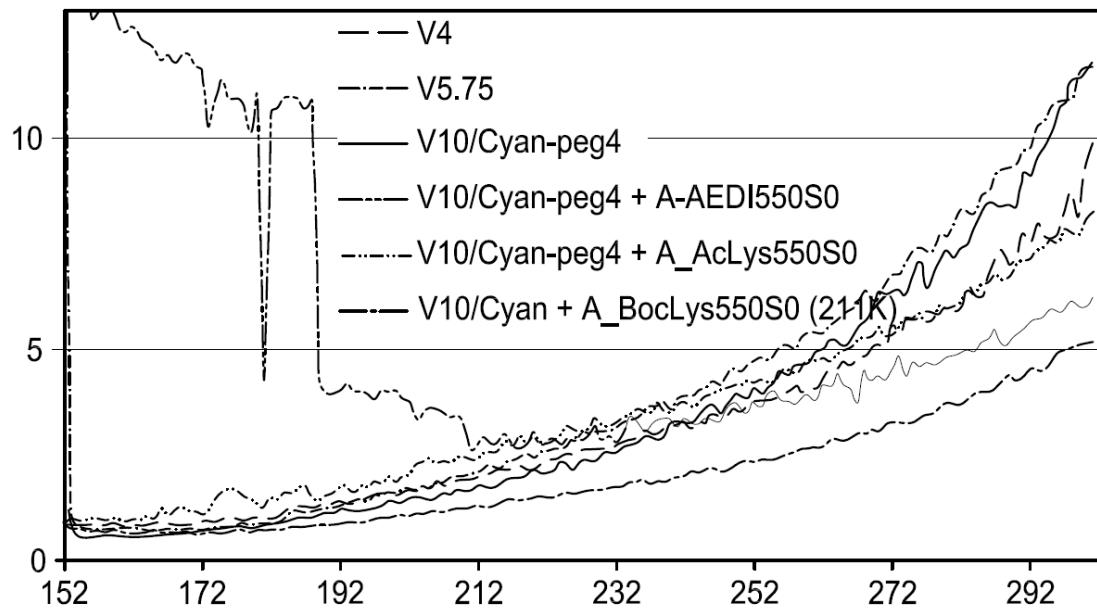


**FIG. 2**

**FIG. 3A****FIG. 3B****FIG. 3C****FIG. 3D****FIG. 3E****FIG. 3F**

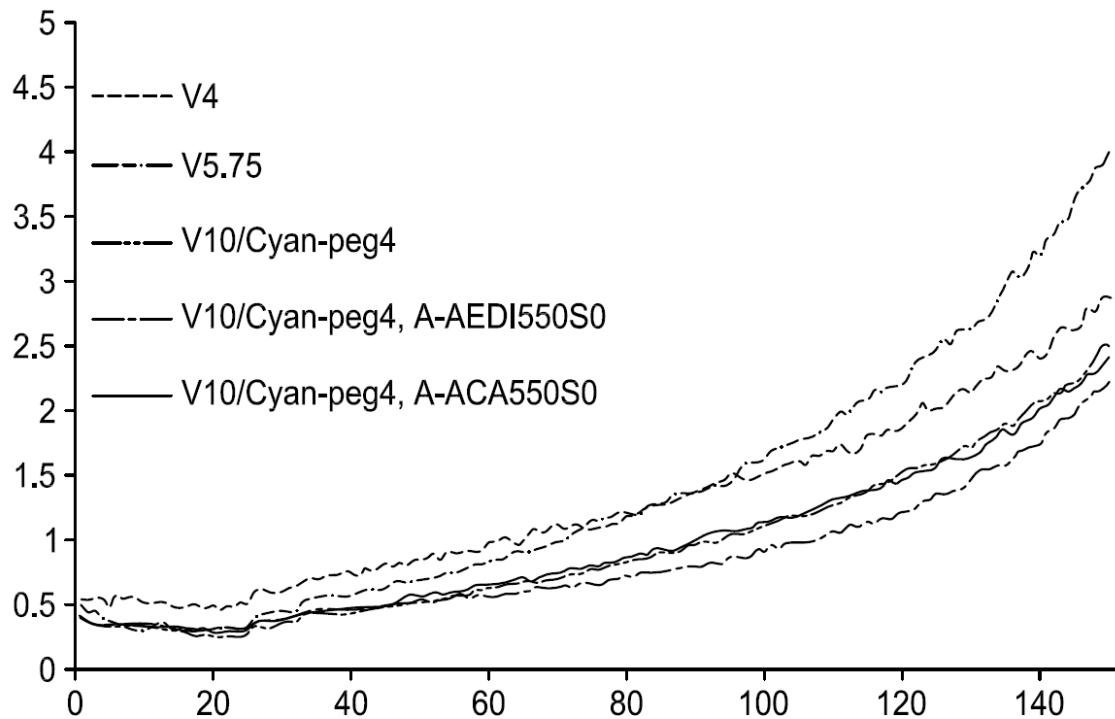
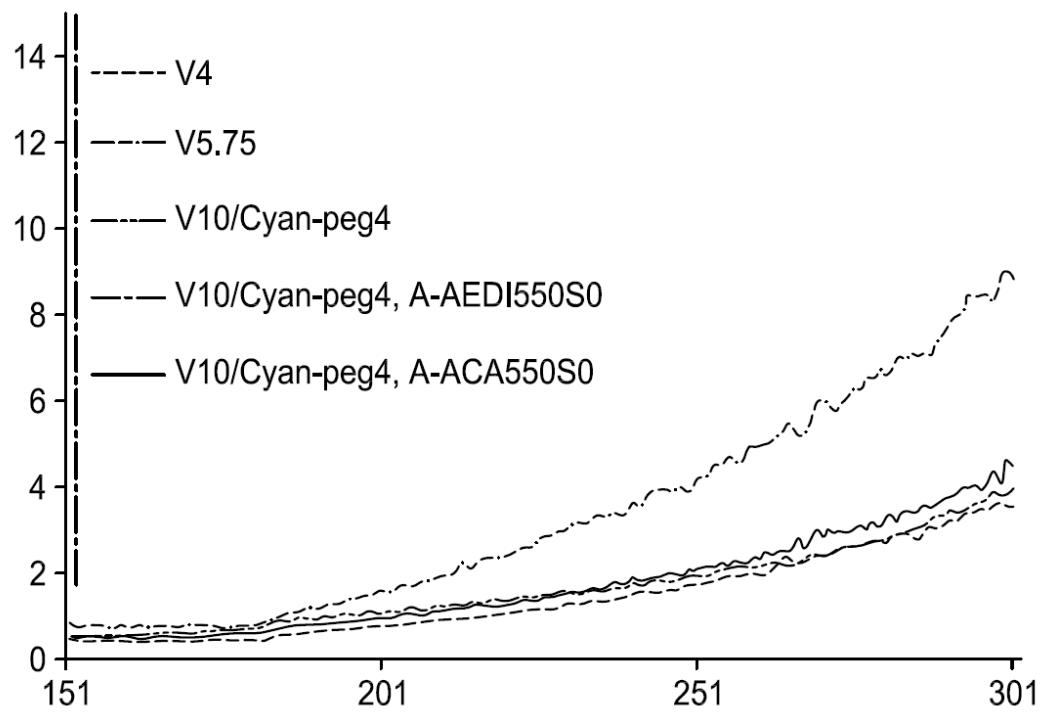
Dye set	Density	%PF	Ph R1	PPh R1	Ph R2	PPh R2	%Align R1	%Align R2	error R1	error R2
V4	230	79.6	0.205	0.111	0.35	0.149	82.2	79.3	1.41	3.19
V5.75	187	77.65	0.371	0.048	0.442	0.164	80.06	79.99	2.16	3.94
V10/cyan-peg4	222	79.88	0.305	0.084	0.53	0.179	81.89	79.45	1.42	3.48
V10/cyan-peg4, A-AEDI550S0	223	81.83	0.14	0.123	0.294	0.197	81.61	67.12	0.85	5.84
V10/cyan-peg4, A-AcLys550S0	227	78.52	0.2	0.128	0.319	0.272	82.23	78.52	1.44	3.48
V10/cyan, A-AcLys550S0	145	86.27	0.178	0.211	0.283	0.29	82.4	79.84	1.33	3.33

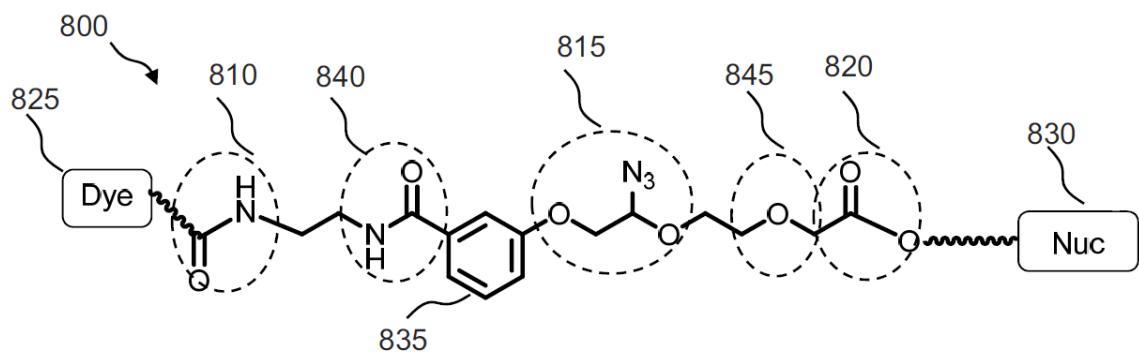
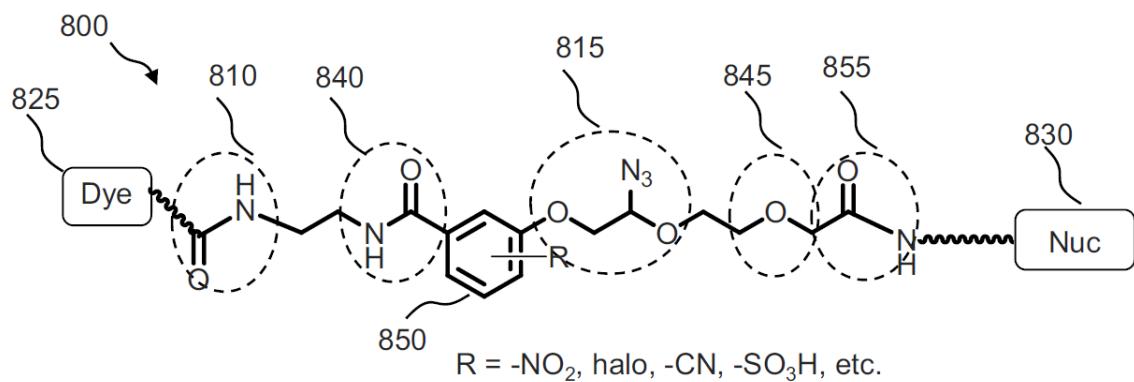
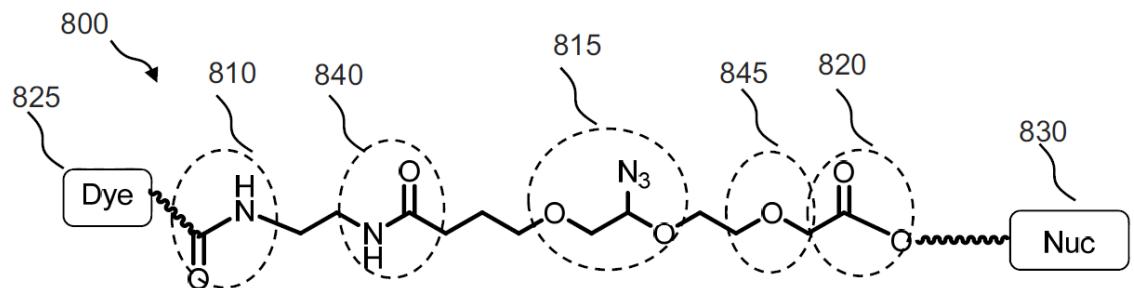
FIG. 4

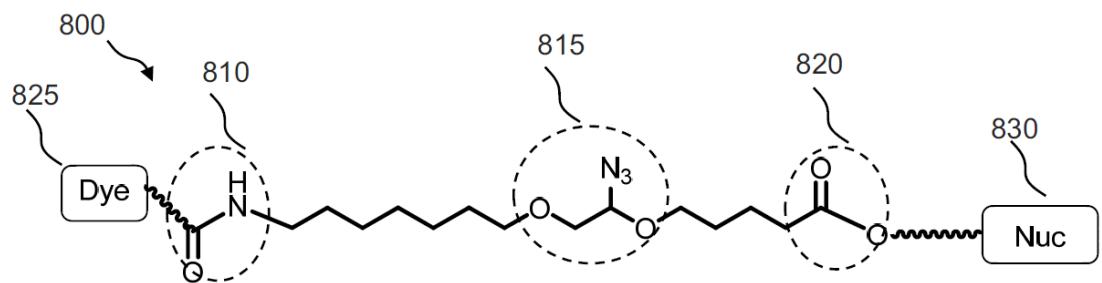
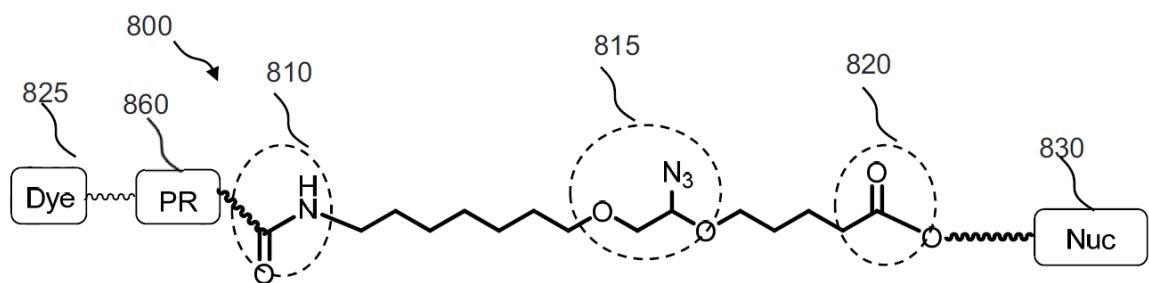
**FIG. 5A****FIG. 5B**

Dye set	Density	%PF	Ph R1	PPh R1	Ph R2	PPh R2	%Align R1	%Align R2	error R1	error R2
V4	191	82.62	0.132	0.151	0.242	0.163	81.91	79.59	1.34	2.43
V5.75	235	79.02	0.212	0.05	0.304	0.187	82.54	78.74	1.4	4.2
V10/Cyan-peg4	213	84.7	0.2	0.063	0.348	0.179	83.3	81.19	0.96	2.02
V10/Cyan-peg4	220	83.03	0.188	0.115	0.317	0.207	82.92	80.9	1.11	2.16
V10/Cyan-peg4, A-AEDI-550S0	252	76.37	0.102	0.097	0.132	0.412	80.1	34.9	0.84	25.58
V10/Cyan-peg4, A-ACA550S0	204	85.88	0.145	0.076	0.271	0.222	83.21	80.58	0.96	2.39

**FIG. 6**

**FIG. 7A****FIG. 7B**

**FIG. 8A****FIG. 8B****FIG. 8C**

**FIG. 8D****FIG. 8E**

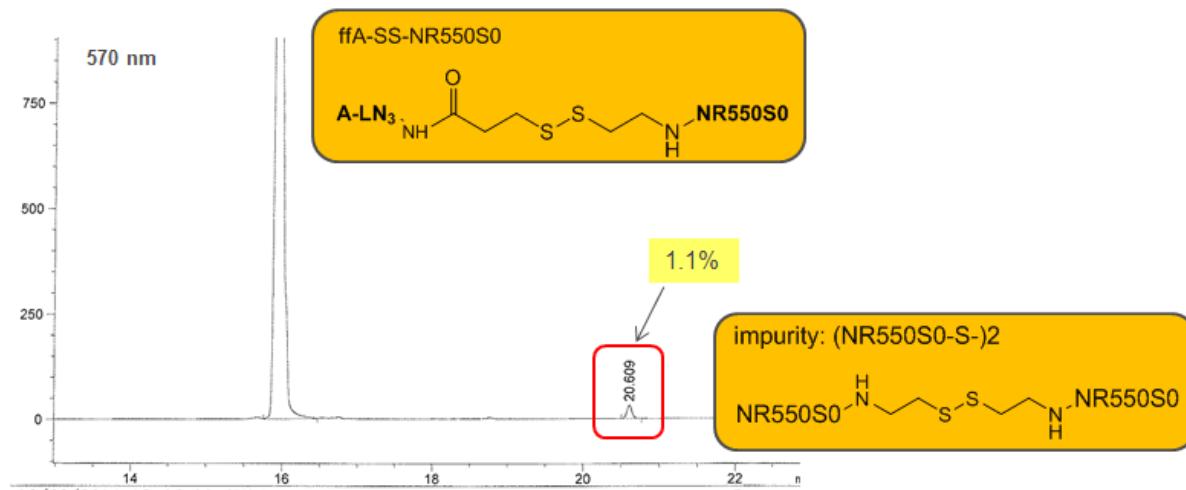


FIG 9A

Tampão		Resultados *	
	21 hours @ 60 °C	ffA-SS-NR550S0	ffA-AEDI-NR550S0
1	PR2 buffer	~ 50% disulfide side-product	~ 6 % disulfide side-product
2	Incorporation buffer (pH 9.9) (no Mg <sup>2+</sup> , no Chaps)	~ 10% disulfide side-product	No disulfide side-product
3	IMX	~ 38% disulfide side-product	No disulfide side-product
4	IMX + Pol 812	~ 34% disulfide side-product	No disulfide side-product
5	SRE	No disulfide side-product	No disulfide side-product
6	TEAB (0.1 M) : CH <sub>3</sub> CN (1:1)	~ 20% disulfide side-product	No disulfide side-product

FIG 9B

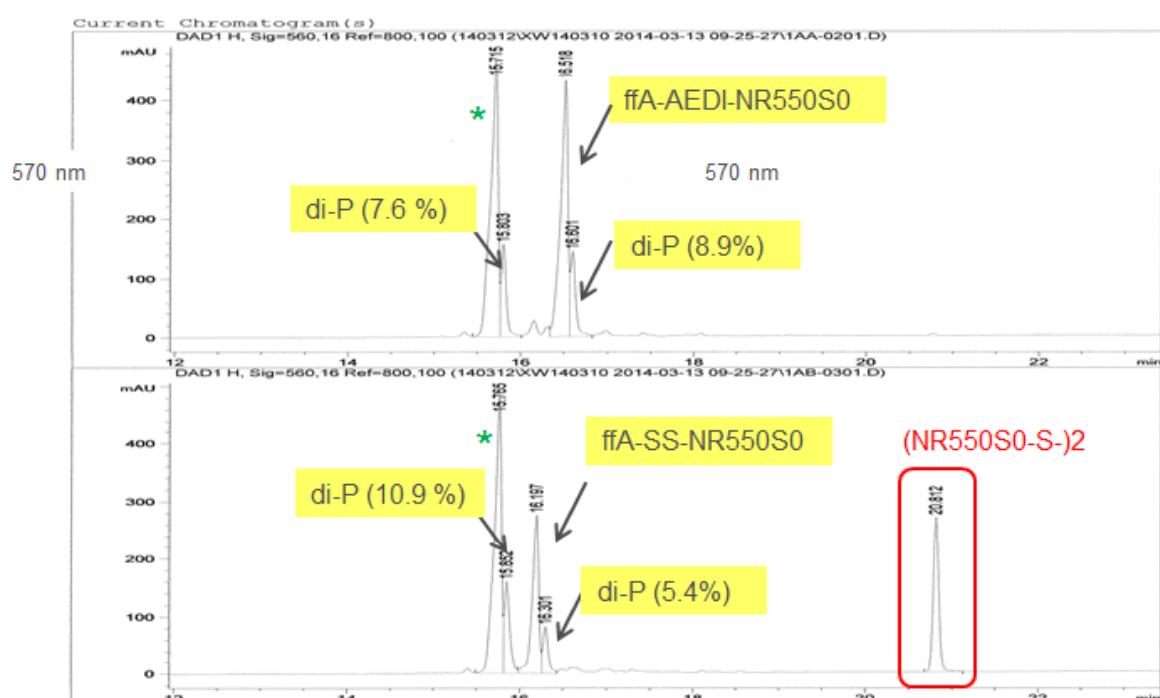
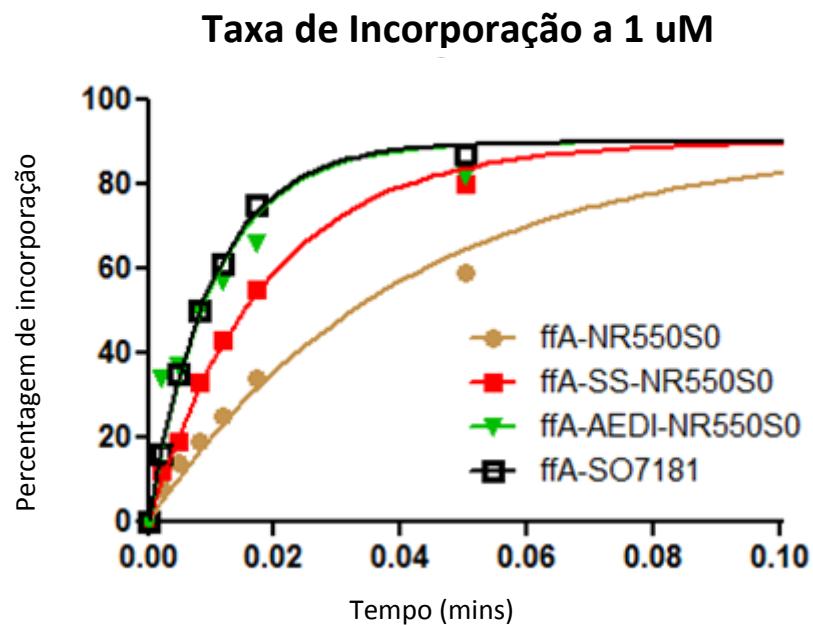
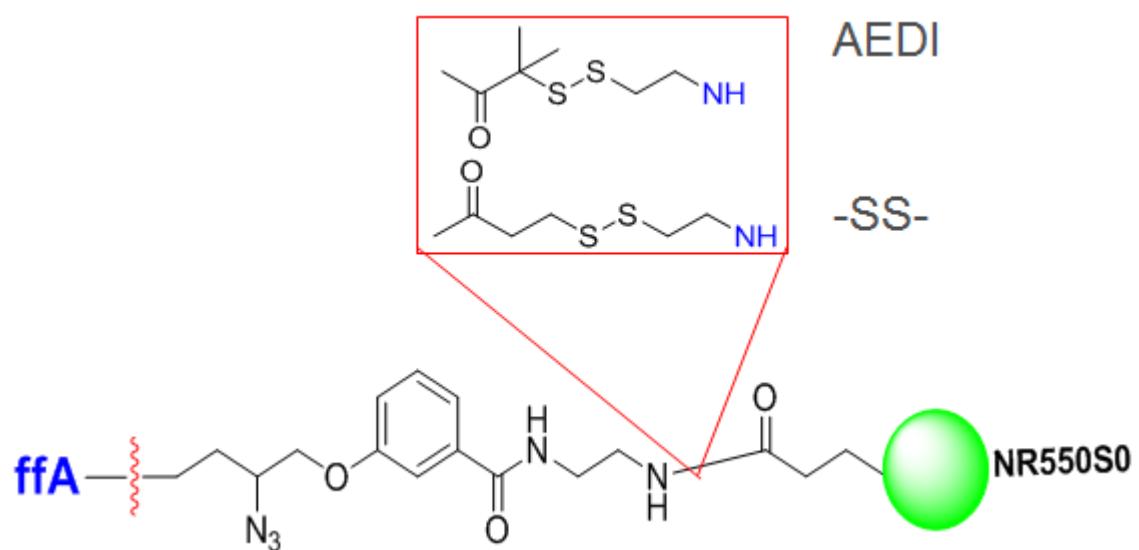


FIG 9C

**FIG 10A**

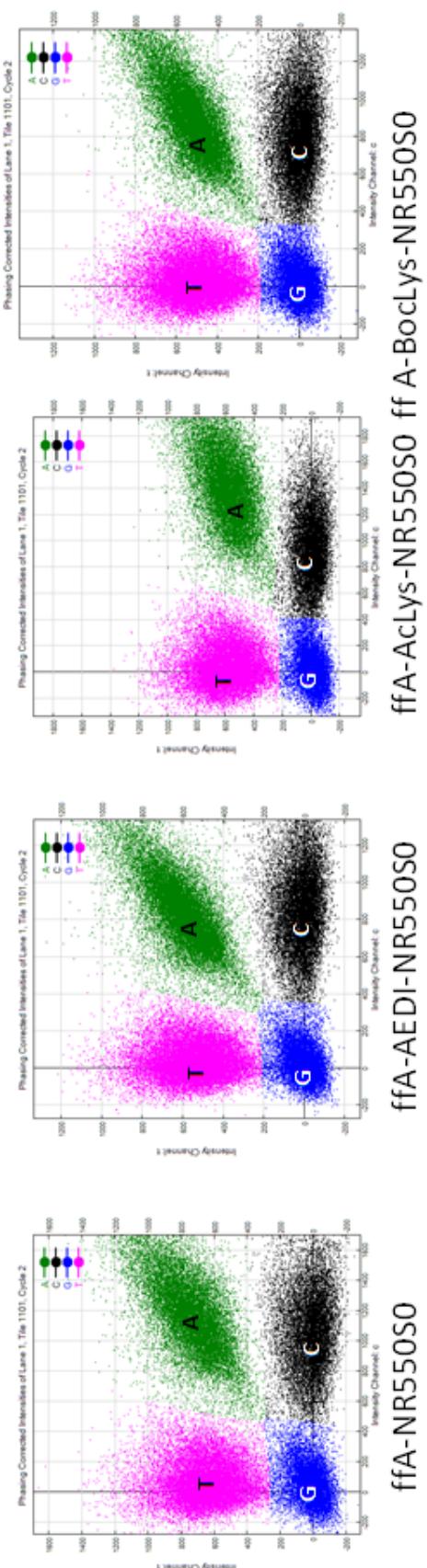
ffA	K (uM/min)
NR550S0	22 (1X)
-SS-NR550S0	54 (2X)
-AEDI-NR550S0	97 (4X)
Std -SO7181	99 (4X)

**FIG 10B**

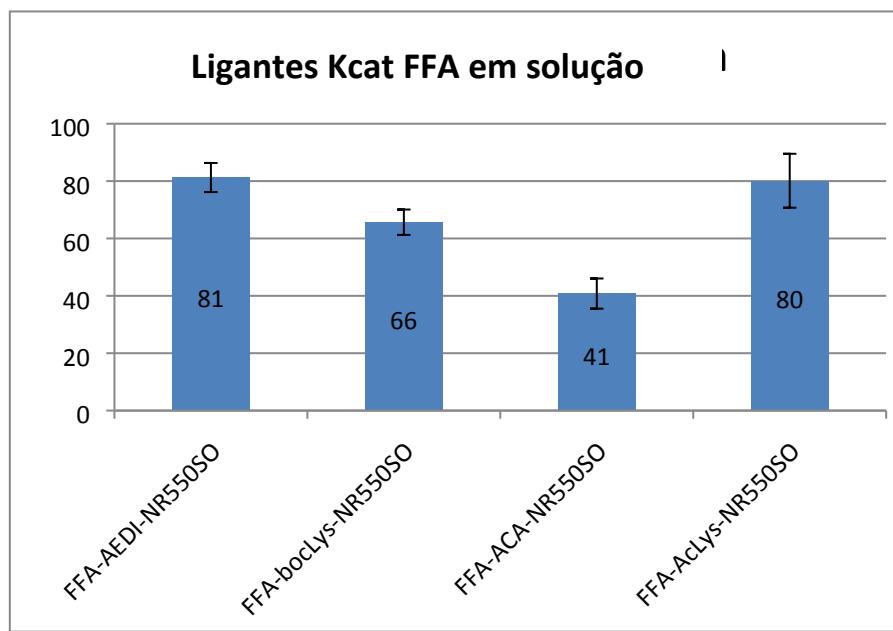


**FIG 10C**

Gráfico de dispersão de Combição V10 com diferente A-550SO (mesma concentração) para título de 1

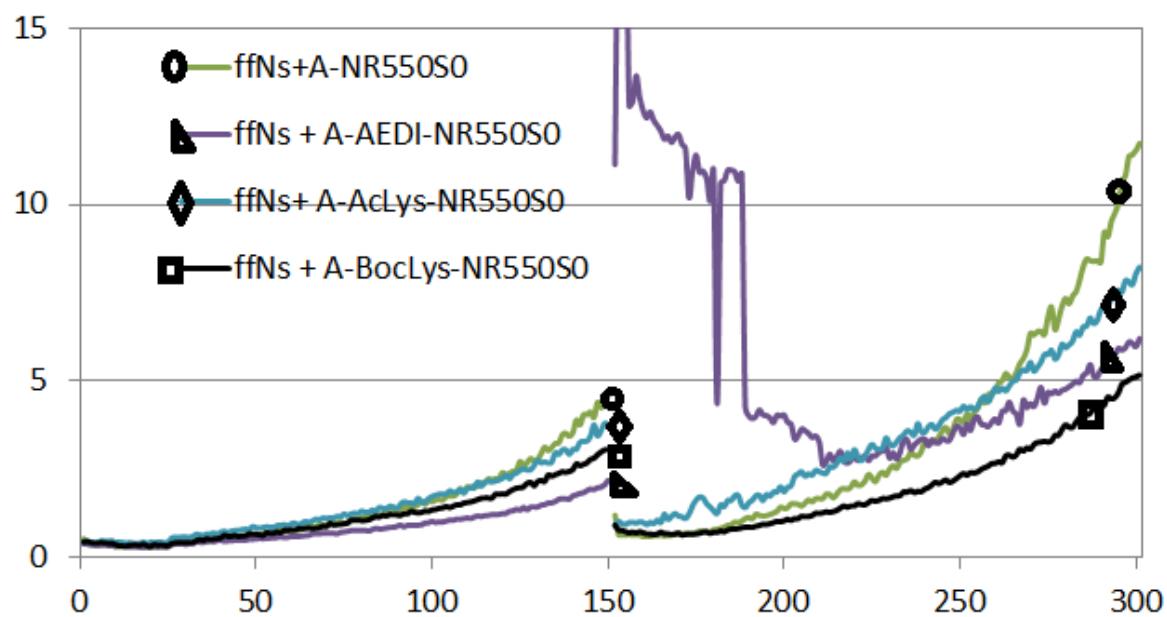


**FIG 11A**

**FIG 11B**

Dye set	Density	%PF	Ph R1	PPh R1	Ph R2	PPh R2	%Align R1	%Align R2	error R1	error R2
ffNs + ffA-NR550S0	222	79.88	0.305	0.084	0.53	0.179	81.89	79.45	1.42	3.48
ffNs + ffA-AEDI-NR550S0	223	81.83	0.14	0.123	0.294	0.197	81.61	67.12*	0.85	5.84*
ffNs + ffA-AcLys-NR550S0	227	78.52	0.2	0.128	0.319	0.272	82.23	78.52	1.44	3.48
ffNs + ffA-BocLys-NR550S0	211	82.97	0.126	0.146	0.226	0.249	82.54	80.45	1.17	1.99

FIG 12A

**FIG 12B**