

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4672021号
(P4672021)

(45) 発行日 平成23年4月20日(2011.4.20)

(24) 登録日 平成23年1月28日(2011.1.28)

(51) Int.Cl.

C 12 N 15/09 (2006.01)

F 1

C 12 N 15/00 Z N A A

請求項の数 12 (全 31 頁)

(21) 出願番号 特願2007-545384 (P2007-545384)
 (86) (22) 出願日 平成17年12月8日 (2005.12.8)
 (65) 公表番号 特表2008-522613 (P2008-522613A)
 (43) 公表日 平成20年7月3日 (2008.7.3)
 (86) 國際出願番号 PCT/KR2005/004207
 (87) 國際公開番号 WO2006/062369
 (87) 國際公開日 平成18年6月15日 (2006.6.15)
 審査請求日 平成19年7月12日 (2007.7.12)
 (31) 優先権主張番号 10-2004-0103283
 (32) 優先日 平成16年12月8日 (2004.12.8)
 (33) 優先権主張国 韓国 (KR)

(73) 特許権者 502077807
 バイオニア コーポレイション
 大韓民国 ダエジョン 306-220
 ダエデオクーラ ムーンピュングードング
 49-3
 (74) 代理人 100065215
 弁理士 三枝 英二
 (74) 代理人 100076510
 弁理士 掛樋 悠路
 (74) 代理人 100099988
 弁理士 斎藤 健治

最終頁に続く

(54) 【発明の名称】標的mRNAと相補的な塩基配列を有する siRNAを用いて標的mRNAの発現を抑制する方法

(57) 【特許請求の範囲】

【請求項1】

(1) 任意の標的mRNAに対し相補的なn個のヌクレオチドでなる全ての組合せのds(二本鎖)RNA配列を得る段階(nは整数である);

(2) 前記各組合せのdsRNA配列に対し、相補的に結合した部分の塩基配列のうち標的mRNAと相補的なアンチセンス鎖の5'-末端から数えて1~2番目の結合エネルギー(A区間)の平均値、3~7番目の結合エネルギー(B区間)の平均値、8~15番目の結合エネルギー(C区間)の平均値及び16~18番目の結合エネルギー(D区間)の平均値E_A、E_B、E_C及びE_Dをそれぞれ求める段階;

(3) 前記各組合せのdsRNA配列に対し、前記(A)~(D)の各区間に對し下記式によってY_(A-B)、Y_(B-C)、Y_(C-D)及びY_(A-D)値を割り当てる段階として、

i) -0.02<E_A-E_B<-0.38、-0.29<E_B-E_C<-0.01、0.00<E_C-E_D<0.35、0.07<E_D-E_A<0.37の範囲であるときY_(A-B)、Y_(B-C)、Y_(C-D)、Y_(A-D)はそれぞれ10点、

ii) -0.63<E_A-E_B<-0.21、0.05<E_B-E_C<0.44、-0.47<E_C-E_D<-0.09、-0.67<E_D-E_A<-0.23の範囲であるときY_(A-B)、Y_(B-C)、Y_(C-D)、Y_(A-D)はそれぞれ0点、

iii) i)とii)の範囲のいずれにも属しない場合はY_(A-B)、Y_(B-C)、Y_(C-D)およびY_(A-D)はそれぞれ5点を付与し;

(4) 前記各組合せのdsRNA配列に対し、下記数学式4によって相対結合エネルギー値Y値を割り当てる段階として、

【数1】

[数学式4]

$$Y = \frac{W_{(A-B)}Y_{(A-B)} + W_{(B-C)}Y_{(B-C)} + W_{(C-D)}Y_{(C-D)} + W_{(A-D)}Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

前記で、 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ は(A-B)、(B-C)、(C-D)及び(A-D)区間にに対する加重値として、それぞれ0.90~1.00、0.2~0.4、0.2~0.3及び0.7~0.9範囲であり、

(5) 前記各組合せのdsRNA配列に対し、下記数学式5によってZ値を割り当てる段階として、

【数2】

[数学式5]

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

前記で、 i は標的mRNAに対するsiRNAの抑制効率に影響を及ぼす因子の数を表わす整数であって、前記因子は相対的な結合エネルギーを必須因子に含み、3'-末端5個の塩基の中でA/Uの個数、1番位置のG/C存在の有無、19番位置のA/U存在の有無及びG/C含量程度で構成された群から選択される一つ以上の因子を選択的な因子として含み、

Z_i は各因子に付与された点数であって、

i) Z_1 は相対的な結合エネルギーの点数である前記Yであり、

但し、前記因子が

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 Z_1 は3'-末端の5個の塩基の中でA/U塩基の数であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 Z_1 は5'-末端の塩基がG/Cであれば1点で、そうでない場合は0点であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合は、 Z_1 は3'-末端の塩基がA/Uであれば1点で、そうでない場合は0点であり(i は1ではない)、及び

v) G/C含量の場合、 Z_1 はG/C含量が36~53%の範囲にある場合10点を与え、そうでない場合0点を与える(i は1ではない)、

M_i は各因子に割り当てられた所定の最高値で、

W_i は W_1 を基準に各因子に割り当てられた所定の加重値であり、

i) M_1 は相対的な結合エネルギーに割り当てられた最高値として100で、 W_1 は相対的な結合エネルギーに割り当てられた加重値として0.90であり、

但し、前記因子が、

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 M_i は5で、 W_i は0.07であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 M_i は1で、 W_i は0.15であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合、 M_i は1で、 W_i は0.19であり(i は1ではない)、及び

v) G/C含量の場合、 M_i は10で、 W_i は0.11である(i は1ではない)；

(6) 前記各組合せのdsRNA配列に対し、段階(5)で求めたZ値を高い順に配列した後、上位10%内にあたるZ値を有するdsRNA配列等を選択する段階；及び

(7) 前記各(6)から選択された配列のdsRNAを用いて標的mRNAの発現を抑制する段階を含む、siRNAを用いて標的mRNAの発現を抑制する方法。

10

20

30

40

50

【請求項 2】

第1項において、

前記siRNAはnが21である21 ヌクレオチドの二本鎖RNAであることを特徴とする方法。

【請求項 3】

第1項または第2項において、

前記siRNAは19 ヌクレオチドのdsRNA部分と両側3'-末端に1~3 ヌクレオチドのオーバーハンギング構造を有することを特徴とする方法。

【請求項 4】

第1項において、

段階(4)の加重値 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ はそれぞれ1.00、0.37、0.20及び0.90であることを特徴とする方法。 10

【請求項 5】

第1項において、

段階(5)の前記数学式5のi=5であり、

Z_1 =前記相対的な結合エネルギーの点数(Y)、 $Z_2=3'$ -末端5個の塩基の中でA/Uの個数に対し割り当てられた点数、 $Z_3=1$ 番位置のG/C存在の有無に対し割り当てられた点数、 $Z_4=19$ 番位置のA/U存在の有無に対し割り当てられた点数及び $Z_5=G/C$ 含量程度に対し割り当てられた点数であり；

$M_1 \sim M_5$ はそれぞれ100、5、1、1、10で、

$W_1 \sim W_5$ はそれぞれ0.90、0.07、0.15、0.19、0.11であることを特徴とする方法。 20

【請求項 6】

(1) 任意の標的mRNAに対し相補的なn個のヌクレオチドでなる全ての組合せのds(二本鎖)RNA配列を得る段階(nは整数である)；

(2) 前記各組合せのdsRNA配列に対し、相補的に結合した部分の塩基配列のうち標的mRNAと相補的なアンチセンス鎖の5'末端から数えて1~2番目の結合エネルギー(A区間)の平均値、3~6番目の結合エネルギー(B区間)の平均値、14~16番目の結合エネルギー(C区間)の平均値及び16~18番目の結合エネルギー(D区間)の平均値 E_A 、 E_B 、 E_C 及び E_D をそれぞれ求める段階；

(3) 前記各組合せのdsRNA配列に対し、前記(A)~(D)の各区間に對し下記式によって $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 及び $Y_{(A-D)}$ 値を割り当てる段階として、 30

i) $0.00 < E_A - E_B < 0.40$ 、 $-0.41 < E_B - E_C < -0.01$ 、 $0.07 < E_C - E_D < 0.39$ 、 $0.07 < E_D - E_A < 0.37$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ10点、

ii) $-0.63 < E_A - E_B < -0.21$ 、 $0.10 < E_B - E_C < 0.51$ 、 $-0.47 < E_C - E_D < -0.19$ 、 $-0.67 < E_D - E_A < -0.23$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ0点、

iii) i)とii)の範囲のいずれにも属しない場合は $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ および $Y_{(A-D)}$ はそれぞれ5点を付与し；

(4) 前記各組合せのdsRNA配列に対し、下記数学式4によって相対結合エネルギー値Y値を割り当てる段階として、

【数式 3】

[数学式 4]

$$Y = \frac{W_{(A-B)}Y_{(A-B)} + W_{(B-C)}Y_{(B-C)} + W_{(C-D)}Y_{(C-D)} + W_{(A-D)}Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

前記で、 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ は(A-B)、(B-C)、(C-D)及び(A-D)区間に對する加重値として、それぞれ0.5~0.7、0.3~0.5、0.3~0.5及び0.9~1.0範囲であり、

(5) 前記各組合せのdsRNA配列に対し、下記数学式5によってZ値を割り当てる段階として、

【数4】

[数学式5]

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

前記で、 i は標的mRNAに対するsiRNAの抑制効率に影響を及ぼす因子の数を表わす整数であって、前記因子は相対的な結合エネルギーを必須因子に含み、3'-末端5個の塩基の中でA/Uの個数、1番位置のG/C存在の有無、19番位置のA/U存在の有無及びG/C含量程度で構成された群から選択される一つ以上の因子を選択的な因子として含み、

Z_i は各因子に付与された点数であって、

i) Z_1 は相対的な結合エネルギーの点数である前記Yであり、

但し、前記因子が

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 Z_i は3'-末端の5個の塩基の中でA/U塩基の数であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 Z_i は5'-末端の塩基がG/Cであれば1点で、そうでない場合は0点であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合、 Z_i は3'-末端の塩基がA/Uであれば1点で、そうでない場合は0点であり(i は1ではない)、及び

v) G/C含量の場合、 Z_i はG/C含量が36~53%の範囲にある場合10点を与える、そうでない場合0点を与える(i は1ではない)、

M_i は各因子に割り当てられた所定の最高値で、

W_i は W_1 を基準に各因子に割り当てられた所定の加重値であり、

i) M_1 は相対的な結合エネルギーに割り当てられた最高値として100で、 W_i は相対的な結合エネルギーに割り当てられた加重値として0.90であり、

但し、前記因子が、

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 M_i は5で、 W_i は0.07であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 M_i は1で、 W_i は0.15であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合、 M_i は1で、 W_i は0.19であり(i は1ではない)、及び

v) G/C含量の場合、 M_i は10で、 W_i は0.11である(i は1ではない)；

(6) 前記各組合せのdsRNA配列に対し、段階(5)で求めたZ値を高い順に配列した後、上位10%内にあたるZ値を有するdsRNA配列等を選択する段階；及び

(7) 前記各(6)で選択された配列のdsRNAを用いて標的mRNAの発現を抑制する段階を含む、siRNAを用いて標的mRNAの発現を抑制する方法。

【請求項7】

第6項において、

前記siRNAはnが21である21ヌクレオチドの二本鎖RNAであることを特徴とする方法。

【請求項8】

第6項または第7項において、

前記siRNAは19ヌクレオチドのdsRNA部分と両側3'-末端に1~3ヌクレオチドのオーバーハンギング構造を有することを特徴とする方法。

【請求項9】

第6項において、

段階(4)の加重値 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ はそれぞれ0.65、0.48、0.48及び0.90であることを特徴とする方法。

10

20

30

40

50

【請求項 10】

第6項において、

段階(5)の前記数学式5の*i*=5であり、

Z_1 =前記相対的な結合エネルギーの点数(Y)、 $Z_2=3'$ -末端5個の塩基の中でA/Uの個数に対し割り当てられた点数、 $Z_3=1$ 番位置のG/C存在の有無に対し割り当てられた点数、 $Z_4=19$ 番位置のA/U存在の有無に対し割り当てられた点数及び $Z_5=G/C$ 含量程度に対し割り当てられた点数であり；

$M_1 \sim M_5$ はそれぞれ100、5、1、1、10で、

$W_1 \sim W_5$ のそれぞれ0.90、0.07、0.15、0.19、0.11であることを特徴とする方法。

【請求項 11】

10

(1) 任意の標的mRNAに対し相補的なn個のヌクレオチドでなる全ての組合せのds(二本鎖)RNA配列を得る段階(nは整数である)；

(2) 前記各組合せのdsRNA配列に対し、相補的に結合した部分の塩基配列のうち標的mRNA Aと相補的なアンチセンス鎖の5'-末端から数えて1~2番目の結合エネルギー(A区間)の平均値、3~7番目の結合エネルギー(B区間)の平均値、8~15番目の結合エネルギー(C区間)の平均値及び16~18番目の結合エネルギー(D区間)の平均値 E_A 、 E_B 、 E_C 及び E_D をそれぞれ求める段階；

(3) 前記各組合せのdsRNA配列に対し、前記(A)~(D)の各区間に對し下記式によって $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 及び $Y_{(A-D)}$ 値を割り当てる段階として、

i) $-0.02 < E_A - E_B < 0.38$ 、 $-0.29 < E_B - E_C < -0.01$ 、 $0.00 < E_C - E_D < 0.35$ 、 $0.07 < E_D - E_A < 0.37$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ10点、

ii) $-0.63 < E_A - E_B < -0.21$ 、 $0.05 < E_B - E_C < 0.44$ 、 $-0.47 < E_C - E_D < -0.09$ 、 $-0.67 < E_D - E_A < -0.23$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ0点、

iii) i)とii)の範囲のいずれにも属しない場合は $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ および $Y_{(A-D)}$ はそれぞれ5点を付与し；

(4) 前記各組合せのdsRNA配列に対し、下記数学式4によって相対結合エネルギー値 Y 値を割り当てる段階として、

【数 5】

[数学式 4]

20

$$Y = \frac{W_{(A-B)}Y_{(A-B)} + W_{(B-C)}Y_{(B-C)} + W_{(C-D)}Y_{(C-D)} + W_{(A-D)}Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

前記で、 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ は(A-B)、(B-C)、(C-D)及び(A-D)区間に對する加重値であって、それぞれ0.90~1.00、0.2~0.4、0.2~0.3及び0.7~0.9範囲であり、

(5) 前記各組合せのdsRNA配列に対し、下記数学式5によって Z 値を割り当てる段階として、

【数 6】

40

[数学式 5]

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

前記で、 i は標的mRNAに対するsiRNAの抑制効率に影響を及ぼす因子の数を表わす整数であって、前記因子は相対的な結合エネルギーを必須因子に含み、3'-末端5個の塩基の中で

50

A/Uの個数、1番位置のG/C存在の有無、19番位置のA/U存在の有無及びG/C含量程度で構成された群から選択される一つ以上の因子を選択的な因子として含み、

Z_i は各因子に付与された点数であって、

i) Z_1 は相対的な結合エネルギーの点数である前記Yであり、

但し、前記因子が

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 Z_i は3'-末端の5個の塩基の中でA/U塩基の数であり (iは1ではない)、

iii) 1番位置のG/C存在の有無の場合、 Z_i は5'-末端の塩基がG/Cであれば1点で、そうでない場合は0点であり (iは1ではない)、

iv) 19番位置のA/U存在の有無の場合、 Z_i は3'-末端の塩基がA/Uであれば1点で、そうでない場合は0点であり (iは1ではない)、及び

v) G/C含量の場合、 Z_i はG/C含量が36~53%の範囲にある場合10点を与え、そうでない場合0点を与える (iは1ではない)、

M_i は各因子に割り当てられた所定の最高値で、

W_i は W_1 を基準に各因子に割り当てられた所定の加重値であり、

i) M_1 は相対的な結合エネルギーに割り当てられた最高値として100で、 W_i は相対的な結合エネルギーに割り当てられた加重値として0.90であり、

但し、前記因子が、

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 M_i は5で、 W_i は0.07であり (iは1ではない)、

iii) 1番位置のG/C存在の有無の場合、 M_i は1で、 W_i は0.15であり (iは1ではない)、

iv) 19番位置のA/U存在の有無の場合、 M_i は1で、 W_i は0.19であり (iは1ではない)、及び

v) G/C含量の場合、 M_i は10で、 W_i は0.11である (iは1ではない)；

(6) 前記各組合せのdsRNA配列に対し、段階(5)で求めたZ値を高い順に配列した後、上位10%内にあたるZ値を有するdsRNA配列等を選択する段階を含むsiRNAデザインの最適化方法。

【請求項 1 2】

(1) 任意の標的mRNAに対し相補的なn個のヌクレオチドでなる全ての組合せのds(二本鎖)RNA配列を得る段階(nは整数である)；

(2) 前記各組合せのdsRNA配列に対し、相補的に結合した部分の塩基配列のうち標的mRNAと相補的なアンチセンス鎖の5'-末端から数えて1~2番目の結合エネルギー(A区間)の平均値、3~6番目の結合エネルギー(B区間)の平均値、14~16番目の結合エネルギー(C区間)の平均値及び16~18番目の結合エネルギー(D区間)の平均値 E_A 、 E_B 、 E_C 及び E_D をそれぞれ求める段階；

(3) 前記各組合せのdsRNA配列に対し、前記(A)~(D)の各区間に下記式によって $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 及び $Y_{(A-D)}$ 値を割り当てる段階として、

i) $0.00 < E_A - E_B < 0.40$ 、 $-0.41 < E_B - E_C < -0.01$ 、 $0.07 < E_C - E_D < 0.39$ 、 $0.07 < E_D - E_A < 0.37$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ10点、

ii) $-0.63 < E_A - E_B < -0.21$ 、 $0.10 < E_B - E_C < 0.51$ 、 $-0.47 < E_C - E_D < -0.19$ 、 $-0.67 < E_D - E_A < -0.23$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ はそれぞれ0点、

iii) i)とii)の範囲のいずれにも属しない場合は $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ および $Y_{(A-D)}$ はそれぞれ5点を付与し；

(4) 前記各組合せのdsRNA配列に対し、下記数学式4によって相対結合エネルギー値Y値を割り当てる段階として、

10

20

30

40

【数7】

[数学式4]

$$Y = \frac{W_{(A-B)}Y_{(A-B)} + W_{(B-C)}Y_{(B-C)} + W_{(C-D)}Y_{(C-D)} + W_{(A-D)}Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

前記で、 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 及び $W_{(A-D)}$ は(A-B)、(B-C)、(C-D)及び(A-D)区間にに対する加重値であって、それぞれ0.5~0.7、0.3~0.5、0.3~0.5及び0.9~1.0範囲であり、

(5) 前記各組合せのdsRNA配列に対し、下記数学式5によってZ値を割り当てる段階として、

10

【数8】

[数学式5]

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

20

前記で、 i は標的mRNAに対するsiRNAの抑制効率に影響を及ぼす因子の数を表わす整数であって、前記因子は相対的な結合エネルギーを必須因子に含み、3'-末端5個の塩基の中でA/Uの個数、1番位置のG/C存在の有無、19番位置のA/U存在の有無及びG/C含量程度で構成された群から選択される一つ以上の因子を選択的な因子として含み、

Z_i は各因子に付与された点数であって、

i) Z_1 は相対的な結合エネルギーの点数である前記Yであり、

但し、前記因子が

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 Z_1 は3'-末端の5個の塩基の中でA/U塩基の数であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 Z_1 は5'-末端の塩基がG/Cであれば1点で、そうでない場合は0点であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合、 Z_1 は3'-末端の塩基がA/Uであれば1点で、そうでない場合は0点であり(i は1ではない)、及び

v) G/C含量の場合、 Z_1 はG/C含量が36~53%の範囲にある場合10点を与え、そうでない場合0点を与える(i は1ではない)、

M_i は各因子に割り当てられた所定の最高値で、

W_i は W_1 を基準に各因子に割り当てられた所定の加重値であり、

i) M_1 は相対的な結合エネルギーに割り当てられた最高値として100で、 W_1 は相対的な結合エネルギーに割り当てられた加重値として0.90であり、

但し、前記因子が、

ii) 3'-末端5個の塩基の中でA/Uの個数の場合、 M_i は5で、 W_i は0.07であり(i は1ではない)、

iii) 1番位置のG/C存在の有無の場合、 M_i は1で、 W_i は0.15であり(i は1ではない)、

iv) 19番位置のA/U存在の有無の場合、 M_i は1で、 W_i は0.19であり(i は1ではない)、及び

v) G/C含量の場合、 M_i は10で、 W_i は0.11である(i は1ではない)；

(6) 前記各組合せのdsRNA配列に対し、段階(5)で求めたZ値を高い順に配列した後、上位10%内にあたるZ値を有するdsRNA配列等を選択する段階を含むsiRNAデザインの最適化方法。

【発明の詳細な説明】

40

50

【技術分野】

【0001】

本発明はsiRNAを用いて標的mRNAの発現を抑制する方法に関し、より詳しくは標的mRNAの活性を抑制する任意のsiRNA(small interfering RNA)塩基配列の隣接または非隣接区間の間の相対的な結合エネルギーパターンを分析することにより最適の抑制効率を示すものと予測されるsiRNAを選別した後、前記siRNAを用いて標的mRNAの発現を抑制する方法に関する。

【背景技術】

【0002】

RNA干渉(RNA interferenceまたはRNAi)は、二本鎖RNA(double-stranded RNAまたはdsRNA)によって同じ塩基配列を有する目標mRNAが細胞質から分解される現象を言う。1998年FireとMelloによってシー・エレガンス(*C.elegans*)(線虫)で初めて明らかになった以後、シヨウジョウ蝇(*Drosophila*)、トリパノソーマ(*Trypanosoma*、鞭毛虫の一種)、脊椎動物(*vertebrate*)などでもRNAi現象が生じるということが報告された(Tabara H, Grishok A, Mello CC, *Science*, 282(5388), 430-1, 1998)。人間の場合、dsRNAを細胞に取り入れると抗ウイルス性インターフェロン機作(antiviral interferon pathway)が誘発されてRNAi効果を見るのが困難であったが、2001年ElbashirとTuschlなどによって21 nt(スクレオチド)の小さなdsRNAを人間細胞に取り入れる場合にはinterferon pathwayが誘発されず、標的mRNAを特異的に分解させることができた(Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., *Nature*, 411, 494-498, 2001; Elbashir, S.M., Lendeckel, W., Tuschl, T., *Genes & Dev.*, 15, 188-200, 2001; Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T., *EMBO J.*, 20, 6877-6888, 2001)。以後、21 ntのdsRNAはスマールインターフィアリングRNA(siRNA)という名称で新しい機能遺伝体学(functional genomics)の道具として脚光を浴び始め、その重要性を認められて2002年度ScienceジャーナルでスマールインターフィアリングRNA(siRNAとmicroRNA)が今年の発見(Breakthrough of the year)1番に選定されることになった(Jennifer Couzin, BREAKTHROUGH OF THE YEAR:Small RNAs Make Big Splash, Jennifer Couzin, *Science* 20 December 2002: 2296-2297)。

【0003】

RNAiは既存のアンチセンスRNA(antisense RNA)技術に比べて機能遺伝体学(functional genomics)と治療(therapeutics)の手段として幾多の長所を有している。第一、アンチセンスRNAでは効率的な目標塩基配列を探すため多数のアンチセンスRNAを合成して多くの時間と経費をかけて実験しなければならない反面、siRNAの場合はいくつかのアルゴリズムを介しその効率がある程度予測可能であるので、より少ない数の実験を介しても効率の高いsiRNAを探すことができる。第二、siRNA(RNAi)はアンチセンスRNAよりさらに低い濃度で効率的に遺伝子発現を抑制させることができると知られている。これは研究用に用いられるときより少ない量を用いることができ、特に治療剤に用いられるとき非常に効果的であり得ることを意味する。第三、RNAiによる遺伝子発現抑制は生体内で自然に生じる機作でありながらその作用が非常に特異的である。

【0004】

RNAi実験は大きくsiRNAデザイン(target site selection)、細胞培養実験(cell culture assay、target mRNAの減少定量、効率が最も高いsiRNA選定)、動物実験(stability、modification、delivery、pharmacokinetics、toxicology)及び臨床実験に分けることができる、このうち最も重要なことが効率の高い目標塩基配列を選別する方法と、目的する組織にsiRNAを伝達(drug delivery)する方法であるということができる。効率の高い目標塩基配列を探さなければならない理由は塩基配列ごとにsiRNAの効率が異なり、特に高効率のsiRNA塩基配列を探してこそ実験結果が明らかであり、さらに治療剤としての使用が可能であるためである。目標塩基配列を探す方法にはコンピュータを利用した計算方法と実験的な方法があるが、実験的な方法は主に目標mRNAを試験管内転写により生成させて、これとよく結合する塩基配列を探すことになっている。しかし、このように試験管内で生成され

たmRNAの構造は細胞内における構造と異なることがあり、さらに細胞内ではmRNAに幾多のタンパク質等が結合する所以があるので、試験管内転写による実験で得られた結果が実際の結果と異なり得るとの可能性がある。よって、効率的なsiRNAを探すアルゴリズムの開発は非常に重要であり、これは非効率的なsiRNA塩基配列を除去させる幾多の変数等を考慮して開発していくことができる。

【0005】

伝統的に、siRNAデザインはTuschl rule等の方法(S.M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, Klaus Weber, T. Tuschl, *Nature*, 411, 494-498, 2001a; S.M. Elbashir, W. Lendeckel, T. Tuschl, *Genes & Dev.*, 15, 188-200, 2001b; S.M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, *EMBO J.*, 20, 6877-6888, 2001c)によって3'オーバーハング(overhang)の形態、GC含量、特定塩基の繰返し、塩基配列内のSNP(single nucleotide polymorphism)、RNA二次構造(secondary structure)、目標としないmRNA塩基配列との相同意性等を考慮して行なわれるのが一般であったが、最近はsiRNAの二本鎖を成す部分が如何なる結合エネルギー状態をしているのかを考慮してこれをsiRNAデザインに反映する傾向がある(Khvorova,A., Reynolds,A., Jayasena,S.D., *Cell*, 115(4), 505, 2003; Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S., Khvorova,A., *Nat. Biotechnol.*, 22(3), 326-330, 2004)。結合エネルギーの状態をsiRNAデザインに反映する最も代表的な例には、RISC(RNAi-induced silencing complex)がdsRNAであるsiRNAの二本のうち何れと結合するのかによってsiRNAの効率に決定的な影響を及ぼすことになることに着目し、5'末端と3'末端のエネルギー差をsiRNA効率の予測に取り入れたことを挙げることができる(Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD., *Cell*, 115(2), 199-208, 2003, 図1参照)。

【0006】

本発明者等は、その間一部分に対してのみ断片的に知られていたsiRNAの効率と結合エネルギー状態との間の相関関係をsiRNAの二本鎖を成す全部分に対し検討しており、統計的な方法を介し一層明確且つ精密に考察した。その結果、未知のsiRNAの相対的な結合エネルギーパターンの分析を介し標的mRNAに対する未知のsiRNAの抑制効率を予め予測することができることを確認し、このように選別された優れた抑制効率を有するsiRNAを用いて標的mRNAの発現を効果的に抑制することができることを明らかにすることにより本発明を完成した。

【発明の開示】

【発明が解決しようとする課題】

【0007】

本発明は、未知のsiRNAの相対的な結合エネルギーパターンを分析することにより実験を通じなくとも標的mRNAの発現を効果的に抑制することができるsiRNA等を選別することができることを確認し、このように選別されたsiRNAを用いて標的mRNAの発現を効果的に抑制することができる方法を提供することにその目的がある。

【0008】

以下、本発明を詳しく説明する。

【0009】

本発明のsiRNAを用いて標的mRNAの発現を抑制する方法は

(1)任意の標的mRNAに対し相補的なn個のヌクレオチドでなる全ての組合せのds(二本鎖)RNA配列を得る段階(nは整数である);

(2)前記各組合せのdsRNA配列に対し、相補的に結合した部分の塩基配列のうち1~2番目の区間(A)の平均結合エネルギー、3~7番目の区間(B)の平均結合エネルギー、8~15番目の区間(C)の平均結合エネルギー、及び16~18番目の区間(D)の平均結合エネルギー値 E_A 、 E_B 、 E_C 及び E_D をそれぞれ求める段階;

(3)前記各組合せのdsRNA配列に対し、前記(A)~(D)の各区間に對し下記式により $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 及び $Y_{(A-D)}$ 値を割り当てる段階として、

(A-B)区間に對し

10

20

30

40

50

i)

【0 0 1 0】

【数1】

$$E_{f(A-B)} - 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}} < X_{(A-B)} < E_{f(A-B)} + 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}}$$

【0 0 1 1】

であれば $Y_{(A-B)}$ =10点;

ii)

【0 0 1 2】

【数2】

$$E_{n(A-B)} - 1.96 \sqrt{\frac{S_{n(A-B)}}{N_n}} < X_{(A-B)} < E_{n(A-B)} + 1.96 \sqrt{\frac{S_{n(A-B)}}{N_n}}$$

【0 0 1 3】

であれば $Y_{(A-B)}$ =0点、

iii) i)とii)の範囲のいずれにも属しない場合は $Y_{(A-B)}$ =5点を付与し、前記と同じ方式で(B-C)、(C-D)及び(A-D)区間に対しそれぞれ $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 及び $Y_{(A-D)}$ 値を割り当て、前記で、 $E_i(A-B)$ は(A-B)区間の間の区間別平均エネルギーの差の平均値、

 $S_i(A-B)$ は前記 $E_i(A-B)$ の分散値、 N_i はそれぞれのsiRNA実験データの個数、

$X_{(A-B)}$ は区間(A)の平均結合エネルギー E_A と区間(B)の平均結合エネルギー E_B との間の差にあたる値であり、 $X_{(B-C)}$ 、 $X_{(C-D)}$ 、 $X_{(A-D)}$ の場合もこれと同じであり；

(4)前記各組合せのdsRNA配列に対し、下記数学式4によって相対結合エネルギー値 Y 値を割り当てる段階として、

【0 0 1 4】

【数3】

[数学式4]

30

$$Y = \frac{W_{(A-B)}Y_{(A-B)} + W_{(B-C)}Y_{(B-C)} + W_{(C-D)}Y_{(C-D)} + W_{(A-D)}Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

【0 0 1 5】

前記で、 $W_{(A-B)}$ は(A-B)区間にに対する加重値であり；(5)前記各組合せのdsRNA配列に対し、下記数学式5によって Z 値を割り当てる段階として

【0 0 1 6】

【数4】

[数学式5]

40

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

【0 0 1 7】

前記で、 i は標的mRNAに対するsiRNAの抑制効率に影響を及ぼす因子を表わす整数であつ

50

て、このうち少なくとも一つは前記siRNAの相対的な結合エネルギーであり、 Z_i は各因子に付与された点数であって、 Z_i は相対的な結合エネルギーの点数であるYであり、

M_i は各因子に割り当てられた所定の最高値であり、

W_i は W_1 を基準に各因子に割り当てられた所定の加重値であり；

(6)前記各組合せのdsRNA配列に対し、段階5)で求めたZ値を高い順に配列した後、上位の所定%内にあたるZ値を有するdsRNA配列等を選択する段階；及び

(7)前記各6)で選択された配列のdsRNAを用いて標的mRNAの発現を抑制する段階を含む。

【 0 0 1 8 】

前記で、siRNAは21～23個のヌクレオチド、望ましくは21個のヌクレオチドで構成されるdsRNAであって、19 nucleotideのdsRNA部分と両側3'-末端に1～3ヌクレオチド、望ましくは2ヌクレオチドのオーバーハング構造を有する形態をしている(図3参照)。

10

【 0 0 1 9 】

本発明では、特定の標的mRNAの発現を抑制するsiRNA等の相対的な結合エネルギーパターンを分析して任意の標的mRNAに対するsiRNAのデザインを最適化するため、siRNA構造上二本鎖を成す部分の相対的な結合エネルギーパターンに従いこれを点数化して体系化した。

【 0 0 2 0 】

先ず、ある未知のsiRNAが標的mRNAに対しどれほどの抑制効率を有するものかという問題を解決するため、本発明者等はsiRNAの結合エネルギー状態と抑制効率との間にどれほどの相関関係があるのかを調査した。ここで、本発明者等はsiRNA内に二本鎖を成す19nt部分のうち一部区間の絶対的な結合エネルギー値ではなく、あくまでも隣接または非隣接区間の間の相対的な結合エネルギーの変化量に焦点を合わせた(図2参照)。

20

【 0 0 2 1 】

本発明の望ましい具体例によれば、siRNAを利用した遺伝子発現抑制実験データは二つの海外ジャーナルに掲載された論文、すなわちKhvorovaの論文(Khvorova A, Reynolds A, Jayasena SD, Cell, 115(4), 505, 2003)とAmarzguiouiの論文(Amarzguioui M, Prydz H, Biochem. Biophys. Res. Commun., 316(4), 1050-8, 2004)から収集された。前記Khvorovaの論文では人間シクロフィリン(hCyPB)遺伝子の193～390番目の塩基配列にあたる配列番号1に記載される塩基配列と、ホタルルシフェラーゼ(pGL3)遺伝子の1434～1631番目の塩基配列にあたる配列番号2に記載される塩基配列、及び前記遺伝子を抑制するsiRNA等が開示されており、Amarzguiouiの論文では多様な遺伝子(AA)を抑制するsiRNA等が開示されている。収集されたデータからデータ分析に用いられたsiRNAの塩基配列とそのsiRNAがどれほどの遺伝子発現抑制の効果があるかという二つの情報を得た。表1はKhvorovaの論文から収集した実験データの一部である。このように得られた塩基配列の情報等をINN-HBニアリストネイバーモデル(nearest neighbor model)を用いて結合エネルギーに対するデータに作成した(Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH, Biochemistry, 37(42), 14719-35, 1998, 図3及び図4参照)。

30

【 0 0 2 2 】

【表1】

表1

遺伝子	位置	塩基配列*	配列番号	ノックダウン %
hCyPB	5 (+192)	CAAAACAGTGGATAATT	3	>90
M60857	27 (+192)	GGCCTTAGCTACAGGAGAG	4	>90
	35 (+192)	CTACAGGAGAGAAAGGATT	5	>90
	41 (+192)	GAGAGAAAGGATTTGGCTA	6	>90
	43 (+192)	GAGAAAGGATTTGGCTACA	7	>90
	45 (+192)	GAAAGGATTTGGCTACAAA	8	>90
	65 (+192)	ACAGCAAATTCCATCGTGT	9	>90
	69 (+192)	CAAATTCCATCGTGTAA	10	>90
	95 (+192)	TCATGATCCAGGGCGGAGA	11	>90
	99 (+192)	GATCCAGGGCGGAGACTTC	12	>90
	131 (+192)	GCACAGGAGGAAGAGCAT	13	>90
	139 (+192)	GGAAAGAGCATCTACGGT	14	>90
	159 (+192)	GCGCTTCCCCGATGAGAAC	15	>90
	7 (+192)	AAAACAGTGGATAATT	16	<50
	9 (+192)	AAACAGTGGATAATT	17	<50
	11 (+192)	CAGTGGATAATT	18	<50
	17 (+192)	ATAATTGTTGCGCTT	19	<50
	23 (+192)	TTGTGGCCTTAGCTACAGG	20	<50
	31 (+192)	TTAGCTACAGGAGAGAAAG	21	<50
	51 (+192)	ATTTGGCTACAAAACAGC	22	<50
	61 (+192)	AAAAACAGCAAATTCCATC	23	<50
	63 (+192)	AAACAGCAAATTCCATCGT	24	<50
	73 (+192)	TTCCATCGTGTAA	25	<50
	97 (+192)	ATGATCCAGGGCGGAGACT	26	<50
	101 (+192)	TCCAGGGCGGAGACTTCAC	27	<50
	103 (+192)	CAGGGCGGAGACTTCACCA	28	<50
	113 (+192)	ACTTCACCAGGGGAGATGG	29	<50
	115 (+192)	TTCACCAGGGGAGATGGCA	30	<50
	119 (+192)	CCAGGGGAGATGGCACAGG	31	<50
	149 (+192)	TCTACGGTGAGCGCTTCCC	32	<50
	151 (+192)	TACGGTGAGCGCTTCCC	33	<50
	171 (+192)	TGAGAACTTCAAAC	34	<50
	173 (+192)	AGAAACTTCAAAC	35	<50
	179 (+192)	TCAAAC	36	<50

* ; 配列番号1に記載される塩基配列の指定された位置から始めて21番目に
あたるヌクレオチドまでの塩基配列を表わす。

【0023】

40

図3に示されているように、siRNAには18個の結合エネルギーが存在する。段階(a)で収集した特定の塩基配列を有するsiRNAの18個の結合エネルギーパターンとその遺伝子発現抑制効率との相関関係を明らかにするためには、先ず前記18個の結合エネルギーを如何なる方式で区間を分けて全体的な結合エネルギーの形態を見るべきかということを決めなければならない。このため、先ず本発明者等は(a)で収集された140個のsiRNA遺伝子発現抑制実験データセットに対し、1番から18番の位置それぞれの結合エネルギーに対し平均値(mean)を求めた後、1番から18番までの位置をx軸、結合エネルギー(-G)をy軸にしてグラフを描いてみた。図5はその結果の一部である。

【0024】

18個の結合エネルギーの位置を何れの区間に分けるのかという問題を解決するため、本

50

発明者等が最大の基準に定めたのは、一区間とその隣接区間との平均結合エネルギーの差が効率的なsiRNA(90%以上遺伝子抑制)と非効率的なsiRNA(50%未満遺伝子抑制)との間で最も大きく逆転される現象を示すよう区間を設定することである。すなわち、区間を複数個、望ましくはA、B、C、Dの四つに分けてそのそれぞれの平均エネルギーを E_A 、 E_B 、 E_C 、 E_D とする場合、効率的なsiRNAと非効率的なsiRNAの各区間別平均結合エネルギーの差、すなわち E_A-E_B 、 E_B-E_C 、 E_C-E_D それぞれの値が0から最も遠く、変化が最も激しく表われるよう区間を設定しなければならない。

【0025】

このため、先ずsiRNA遺伝子発現抑制実験データを効率的なものと非効率的なものの二つの集団に分け、1番から18番の結合エネルギーの位置全部に対し各結合エネルギーの位置で二つの集団に差がないという帰無仮説を立てた後、これをt-テストを介し検証してみた。すなわち、ここでp-値が0.05未満で出る結合エネルギーの位置は前記の二つの集団に対し有意水準5%で結合エネルギーの差が生じる位置であることを意味する。図6はt-テストの結果をx軸を結合エネルギーの位置、y軸をp-値にして表わしたグラフであり、図7はx軸を結合エネルギーの位置、y軸をt-値にして柔らかな形の曲線で表わしたグラフである。前記t-値は下記数学式1によって計算される。

【0026】

【数5】

[数学式1]

$$(t-value) = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x}{N_x} + \frac{S_y}{N_y}}}$$

前記で、 \bar{X} ：効率的な集団の平均結合エネルギー；

\bar{Y} ：非効率的な集団の平均結合エネルギー；

S_x ：効率的な集団の分散；

S_y ：非効率的な集団の分散；

N_x ：効率的な集団の変量の個数；

N_y ：非効率的な集団の変量の個数。

【0027】

本発明の望ましい具現例では三種類のデータセットが用いられた。Khvorovaの論文から抜粋した二種類のデータセットはpGL3とhCyPBに対する遺伝子抑制実験結果を効率:90%以上抑制、非効率:50%未満に分類しておいたものであり、Amarzguiouiの論文から抜粋した一つのデータセットは幾種類の遺伝子に対し複合的に(AA)効率:70%以上抑制、非効率:70%未満に分類しておいたものである。Khvorovaの論文で遺伝子ホタルルシフェラーゼ(pGL3)に対する実験結果は効率的なものが40個、非効率的なものが20個であり、人間シクロフィリン(hCyPB)に対する実験結果は効率的なものが13個、非効率的なものが21個である。Amarzguiouiの論文での実験結果(AA)は効率的なものが21個、非効率的なものが25個である。

【0028】

一応、本発明者等は図7で三つのデータセットのt-値の変化形態が一致するパターンに表われることに注目した。また、Amarzguiouiの論文⁶⁾で得たデータセットは残りの二つのセットに比べて効率と非効率の区分が一層曖昧であろうとの予想どおり、t-値の変化幅が他のデータセット等に比べて少ないものに表われた。これは効率的なsiRNAと非効率的

10

20

30

40

50

なsiRNAとの間には結合エネルギーの形態に確かに特殊な区分があることを示唆するものと見ることができる。

【0029】

t-値が極大または極小値を有するところ、またはp-値が0に近くなるところは効率的なsiRNA集団と非効率的な集団との間の結合エネルギーの差が隣接した部分に比べて極端に大きい部分であると言える。すなわち、この部分を中心にして周辺近隣を一区間に取れば隣接区間等の間の結合エネルギーの偏差を極大化させることができる。さらに、t-値が極大や極小を有するが、二つの値の偏差が大きくなない点、すなわちp-値が留意するほどの水準で小さくない地点等はあまり弁別力が大きくなない点に取り扱い、前記の区間選定においてその候補から排除させることができる。

10

【0030】

本発明の望ましい具現例では、このような事項等に基づき図6のp-値値を用いて区間の中心となる位置等を選定した。このとき次のような基準を適用した：

- 1 Khovorovaの二つのデータセットのうち一つ以上のp-値が0.1以下の位置
- 2 Khovorovaの二つのデータセット全部が0.4以下の位置

1と2の基準に適した位置は全て次の4個が選定された：1番結合エネルギーの位置、5～6番結合エネルギーの位置、14番結合エネルギーの位置、17～18番結合エネルギーの位置。

【0031】

以下の過程ではKhovorovaの二つのデータセットのみを用いた。これはAmarzguiouiのデータセットの場合、グループを分ける基準がKhovorovaの二つのデータセットと異なるためでもあり、また本発明のsiRNAの効率を測定する採点方法が完成された後、その性能をテストするための目的で残しておいたものもある。

20

【0032】

次に、このように決定された四箇所の位置を中心にその近隣のどこまでを一区間に取るべきかを決定する。これを決定する基準は決められた区間の平均結合エネルギーを求め、隣接した他の区間の結合エネルギーとの差を求めた後、この差の変化を極大化させることができることを選択するようにした。望ましくは、これ以後の過程は次の二つに分けて進めることができる：

- (1) 隣接した区間の間に空の空間がなく連続的につながるように設定する場合
- (2) 隣接した区間の間に空の空間があり得るよう不連続的に設定する場合

30

この二つの場合、全て皆一長一短がある。(1)の方法は全ての結合エネルギーに対しその状態をよく見られるが、一部弁別力が劣る区間を含ませることによりその予測力を低下させ得るとの短所がある。一方、(2)の方法は弁別力のない区間を除外されることによりその予測力を極大化させることができるが、一部区間が除かれることによりその位置に対する評価が不可能になるとの短所がある。

【0033】

- (1)区間の設定は望ましくは次のようになされる：

1と2の基準を介し選定された四箇所の位置をそれぞれ含みながら他の位置の領域を侵犯しない範囲内で全体に亘って全ての結合エネルギーの位置が含まれるようA、B、C、D四つの区間に分けて表2で見られる20種の組合せを作る。

40

【0034】

【表2】

表2

区間 A	区間 B	区間 C	区間 D	区間 A	区間 B	区間 C	区間 D
1 ~ 2	3 ~ 7	8 ~ 14	15 ~ 18	1 ~ 3	4 ~ 7	8 ~ 14	15 ~ 18
1 ~ 2	3 ~ 8	9 ~ 14	15 ~ 18	1 ~ 3	4 ~ 8	9 ~ 14	15 ~ 18
1 ~ 2	3 ~ 9	10 ~ 14	15 ~ 18	1 ~ 3	4 ~ 9	10 ~ 14	15 ~ 18
1 ~ 2	3 ~ 10	11 ~ 14	15 ~ 18	1 ~ 3	4 ~ 10	11 ~ 14	15 ~ 18
1 ~ 2	3 ~ 11	12 ~ 14	15 ~ 18	1 ~ 3	4 ~ 11	12 ~ 14	15 ~ 18
1 ~ 2	3 ~ 7	8 ~ 15	16 ~ 18	1 ~ 3	4 ~ 7	8 ~ 15	16 ~ 18
1 ~ 2	3 ~ 8	9 ~ 15	16 ~ 18	1 ~ 3	4 ~ 8	9 ~ 15	16 ~ 18
1 ~ 2	3 ~ 9	10 ~ 15	16 ~ 18	1 ~ 3	4 ~ 9	10 ~ 15	16 ~ 18
1 ~ 2	3 ~ 10	11 ~ 15	16 ~ 18	1 ~ 3	4 ~ 10	11 ~ 15	16 ~ 18
1 ~ 2	3 ~ 11	12 ~ 15	16 ~ 18	1 ~ 3	4 ~ 11	12 ~ 15	16 ~ 18

【0035】

ここで効率的なsiRNAの個数を N_f 、非効率的なsiRNAの個数を N_n とし、効率がi(効率的なグループのsiRNAであれば‘f’、非効率的なグループのsiRNAであれば‘n’である)でj(1 ~ N_f または $1 \sim N_n$ 中の数を値として有する)番目のsiRNAが区間k(A、B、C、D)のうち一つの値を有する)で有する結合エネルギー一つ当たりの平均結合エネルギーを E_{ijk} に定義する。すなわち、効率的なグループの3番目のsiRNAの区間Bでの結合エネルギー一つ当たりの平均エネルギーは E_{f3B} に表示される。それぞれの E_{ijk} を実験データを用いて求める。

【0036】

前記で求めたそれぞれの E_{ijk} を用いて区間A ~ B($E_{i(A-B)}$)、B ~ C($E_{i(B-C)}$)、C ~ D($E_{i(C-D)}$)の間の代表になる平均結合エネルギーの変化量を下記数学式2によって求める。

【0037】

【数6】

[数学式2]

$$E_{i(A-B)} = E_{iA} - E_{iB} = \frac{1}{N_i} \sum_j (E_{ijA} - E_{ijB})$$

【0038】

前記数学式2を利用すれば $E_{i(B-C)}$ と $E_{i(C-D)}$ も求めることができるはずである。ここで $E_{f(A-B)}$ の意味は効率的なグループのsiRNA等の区間AとBでの結合エネルギーの位置一つ当たりの結合エネルギーを代表する値と言え、 $E_{n(A-B)}$ の場合は非効率的な場合のそれと言えるはずである。すなわち、 $E_{f(A-B)} - E_{n(A-B)}$ の絶対値が大きくなるよう区間を取れば、区間Aと区間Bで効率的なsiRNA集団と非効率的なsiRNA集団との平均結合エネルギーの差を大きくすることができ、これを用いて区間を選定することができる。これはB ~ C、C ~ Dにも同様に適用される。これを用いて本発明者等は $E_{f(A-B)} - E_{n(A-B)}$ 、 $E_{f(B-C)} - E_{n(B-C)}$ 、 $E_{f(C-D)} - E_{n(C-D)}$ の絶対値が全て0.1以上の区間の組合せ等のみを選定した。本発明の望ましい具現例では全て四つの区間が選定されており、選定された区間にに対する情報は表3の通りである。

【0039】

10

20

30

40

【表3】

表3

区間 A	区間 B	区間 C	区間 D
1 ~ 2	3 ~ 7	8 ~ 15	16 ~ 18
1 ~ 2	3 ~ 8	9 ~ 15	16 ~ 18
1 ~ 3	4 ~ 7	8 ~ 15	16 ~ 18
1 ~ 3	4 ~ 8	9 ~ 15	16 ~ 18

10

【0040】

選定された四つの区間に對し $E_{f(A-B)}$ と $E_{n(A-B)}$ 、 $E_{f(B-C)}$ と $E_{n(B-C)}$ 、 $E_{f(C-D)}$ と $E_{n(C-D)}$ の間で t-テストをして t-値と p-値を求めてみた。この過程を介し、最終的に効率的な siRNA 集団と非効率的な siRNA 集団を最もよく区分することができる一つの区間を、遺伝子 hCypB、pGL3 の全ての区間で p-値 < 0.05、t-値 > 2 の水準で選定した。選定された区間は A(1 ~ 2)、B(3 ~ 7)、C(8 ~ 15)、D(16 ~ 18) 区間である。この区間に對する各種の情報は図8に示した。

【0041】

一方、(2)の区間の設定は望ましくは次のようになされる：

基本的には(1)とほとんど同じ方法を用いる。ただ(1)と異に不連續的で区間等同士の重畳を許容するはずであるがため、区間の幅を定めることにおいて別の方法を用いる。一応、1 と 2 の基準を介し選定された4箇所の結合エネルギーの位置を含みながらその位置で ± 2 結合エネルギーの位置内で作ることができる全ての区間の組合せを作っており、その結果は表4の通りである。

20

【0042】

【表4】

表4

区間 A	1	1~2	1~3					
区間 B	3~6	4~6	5~6	3~7	4~7	5~7	3~8	4~8
区間 C	12~14	13~14	14	12~15	13~15	14~15	12~16	13~16
区間 D	15~18	16~18	17~18					

30

【0043】

表4で区間 A、B、C、D の中で一つずつを選べば必要な区間の組合せが成り立つ。全て 729 (=3 × 9 × 9 × 3) 種の組合せが可能である。729 種の組合せ全てに對し数学式2の方法と t-テストを介しただけ一つの区間の組合せを選択するということは少なからずの無理があるので、望ましくは新しい変数 R(robustness の略字)を取り入れる。R は区間内に 1 と 2 の基準によって選定された4箇所の結合エネルギー以外に追加的に幾箇所の結合エネルギーがあるのかを表わす数字である。例えば、区間 A を 1 ~ 2 に定めて区間 B を 4 ~ 7 に取れば、区間 A の R 値は 1 であり区間 B の R 値は 2 である。さらに、区間 A(1 ~ 2) と区間 B(4 ~ 7) で (1) の $E_{f(A-B)}$ のように二つの区間に對する R 値を考慮しなければならない場合、二つの区間それぞれの R 値を合算して A ~ B 区間に對する R 値は 3 に選定される。

40

【0044】

表4で見られる A、B、C、D 区間の全ての組合せに對し (1) で言及した E_{ijk} をそれぞれ求めた。数学式2から計算される $E_{i(A-B)}$ 、 $E_{i(B-C)}$ 、 $E_{i(C-D)}$ 値を表4を介し可能な全ての組合せに對し求めており、それぞれに對し t-テストを実施して t-値と p-値を求めた。ここに前記で言及した R 値を適用した。図9は、特定 R 値を有する A ~ B、B ~ C、C ~ D 区間の組合せ等のうち p-値が 0.05 未満のものらの比率をグラフで示したものである。R 値が増加するに伴い p-値が減少する傾向があるので、p-値の減少が急激に起こる前までの R 値を求めるこ

50

により所望の水準のp-値を有しながら最大限広い範囲を含ませる区間を算出し出すことができる。図9の結果を見れば、R値が3または4以下の値を有するときp-値<0.05である区間の比率が高いことが分かる。よって、本発明の望ましい具現例ではR=3または4の値を有する区間等だけを選んで選定される区間の候補に含ませた。

【0045】

最終的な区間の決定はR値とt-テストの結果を介しながら選定される。二つの区間でR値が3または4でなければならないので、両方に区間追加が行なわれる区間Bと区間Cは2個の結合エネルギーの位置を加え、一側に区間追加が行われる区間Aと区間Dは1個の結合エネルギーの位置を加えた。結果的にA～BでR=3、B～CでR=4、C～DでR=3の値を有することになる。この条件を満足する区間等の全ての組合せを作った後、この組合せ等に対しt-テストを行いこの組合せ等の中でp-値が特に低い一つの区間組合せを選定した。選定された区間はA(1～2)、B(3～6)、C(14～16)、D(16～18)である。これに対する情報は表5に示されている。

【0046】

【表5】

表5

		区間 A-B	区間 B-C	区間 C-D
		1～2	3～6	14～16
		3～6	14～16	16～18
hCyPB	t-値	3.175553	-3.4246	5.915552
	p-値	0.00165	0.000853	0.000001
pGL3	t-値	2.68004	-2.32939	3.217273
	p-値	0.004783	0.011671	0.001059
AA	t-値	1.887835	-0.89566	1.266718
	p-値	0.032827	0.18765	0.10596

10

20

30

【0047】

本発明の望ましい具体例で、(1)と(2)を介し選定された二つの区間(図10参照)は隣接区間との相対的な結合エネルギーパターンのみを判別することにより選定された。しかし、非隣接区間の間にも結合エネルギーの差が充分に生じ得るため、これをもう少し拡大してA、B、C、D四つの区間の差で可能な全ての組合せ、A-B、B-C、C-D、A-C、A-D、B-Dの六種の組合せに対し全部t-テストを再度行っており、その結果は表6の通りである。

【0048】

【表6】

表6

区間 A	区間 B	区間 C	区間 D				
1~2	3~7	8~15	16~18				
		区間 A-B	区間 B-C	区間 C-D	区間 A-C	区間 A-D	区間 B-D
hCyPB	t-値	3.15303	-2.25399	3.27599	1.38792	5.40182	1.00611
	p-値	0.00175	0.01559	0.00127	0.08737	0.00000	0.16095
pGL3	t-値	2.42243	-2.40223	2.13573	0.42633	2.31082	-0.15585
	p-値	0.00928	0.00976	0.01847	0.33572	0.01221	0.43834
AA	t-値	1.87483	-1.02960	1.09863	1.41229	1.94585	0.22186
	p-値	0.03373	0.15441	0.13895	0.08245	0.02904	0.41273
区間 A	区間 B	区間 C	区間 D				
1~2	3~6	14~16	16~18				
		区間 A-B	区間 B-C	区間 C-D	区間 A-C	区間 A-D	区間 B-D
hCyPB	t-値	3.16461	-3.42274	5.92078	0.65134	5.40182	0.82726
	p-値	0.00340	0.00172	0.00000	0.51948	0.00001	0.41421
pGL3	t-値	2.69174	-2.32867	3.20424	0.17064	2.31082	-0.32109
	p-値	0.00464	0.01169	0.00110	0.43255	0.01221	0.37465
AA	t-値	1.89671	-0.91889	1.27660	1.29998	1.94585	0.16337
	p-値	0.03222	0.18158	0.10422	0.10019	0.02904	0.43549

【0049】

表6で見られるように、A-C、B-Dの区間では互いに大きい差が存在しなかった。非隣接区間でp-値<0.05の条件を満足するのはA-Dの組合せであったが、ここで区間Aは5'末端、区間Bは3'末端でこの二つの区間の結合エネルギーの差がsiRNAの効率に影響を及ぼすとのことは既に他の実験等を介してもよく知られている事実である(Schwarz,D.S., Hutvagner,G., Du,T., Xu,Z., Aronin,N., Zamore,P.D., Cell, 115(2), 199-20, 2003)。

【0050】

本発明者等は、未知のsiRNAの相対的な結合エネルギーを点数化するため前記で収集した実験データと選定された区間等を利用した。先ず採点システムの構築のため前記で収集したデータのうちKhvorovaの論文から抜粋した二種類のデータセット、すなわちホタルルシフェラーゼ(pGL3)と人間シクロフィリン(hCyPB)に対する二つの実験結果を合わせて一層大きいデータセットを作ってこれを利用した。Amarzguiouiの論文から抜粋した一つのデータセットは遺伝子発現抑制の効率を70%を基準にして分けたものであり、90%以上を効率的、50%以下を非効率的に見たKhvorovaの論文のデータとその分類基準が違うとの点を勘案して採点システムを、構築のためのデータから排除させた。このように得られたデータを効率的なグループ(遺伝子発現抑制効率90%以上、functional、またはf)と非効率的なグループ(遺伝子発現抑制効率50%未満、nonfunctional、またはn)の二つの互いに異なる

10

20

30

40

50

集団に分類した。

【0051】

このように得られたデータ等を前記過程を介し得られた区間等に分け、数学式2から $E_{i(A-B)}$ 、 $E_{i(B-C)}$ 、 $E_{i(C-D)}$ 、 $E_{i(A-D)}$ 値等を求めた。この値等は、各区間等の間の区間別平均エネルギーの差等に関する値等をグループ別に束ねて平均を出したエネルギー値を意味する。この過程でそれぞれは分散値を有することになるが、これを $S_{i(A-B)}$ 、 $S_{i(B-C)}$ 、 $S_{i(C-D)}$ 、 $S_{i(A-D)}$ に定義する。そしてそれぞれのsiRNA実験データの個数を N_i に定義する。

このとき前の過程から得られたデータ等の $E_{i(A-B)}$ 、 $E_{i(B-C)}$ 、 $E_{i(C-D)}$ 、 $E_{i(A-D)}$ 値と $S_{i(A-B)}$ 、 $S_{i(B-C)}$ 、 $S_{i(C-D)}$ 、 $S_{i(A-D)}$ 値、 N_i 値を求め、t-テストを介しt-値とp-値を求めてみれば、表7のような値を有する。

【0052】

【表7】

表7

区間 A	区間 B	区間 C	区間 D		
1~2	3~7	8~15	16~18		
		区間 A-B	区間 B-C	区間 C-D	区間 A-D
	平均(Ef)	0.18	-0.15	0.18	0.22
効率的	分散(Sf)	0.55	0.28	0.41	0.32
Nf=53	標準偏差	0.74	0.53	0.64	0.57
	Nf	53	53	53	53
	平均(Ef)	-0.42	0.25	-0.28	-0.45
非効率的	分散(Sf)	0.49	0.43	0.4	0.53
Nn=41	標準偏差	0.7	0.65	0.63	0.73
	Nn	41	41	41	41
	T	4.026342	-3.16981	3.489798	4.826898
	P	0.000058	0.001036	0.000372	0.000003
<hr/>					
区間 A	区間 B	区間 C	区間 D		
1~2	3~6	14~16	16~18		
		区間 A-B	区間 B-C	区間 C-D	区間 A-D
	平均(Ef)	0.2	-0.21	0.23	0.22
効率的	分散(Sf)	0.56	0.57	0.34	0.32
Nf=53	標準偏差	0.75	0.75	0.59	0.57
	Nf	53	53	53	53
	平均(Ef)	-0.42	0.3	-0.33	-0.45
非効率的	分散(Sf)	0.47	0.45	0.21	0.53
Nn=41	標準偏差	0.69	0.67	0.46	0.73
	Nn	41	41	41	41
	T	4.166805	-3.49839	5.207057	4.826898
	P	0.000035	0.000362	0.000001	0.000003

【0053】

表7で見られるように、このデータセットは全ての区間でp-値<0.05であるので、効率的

10

20

30

40

50

なsiRNAと非効率的なsiRNAを分離し出す採点システムに用いるのに大きい無理がないものと見られる。

【0054】

効率的なsiRNAグループ内の特定siRNAの区間Aと区間Bとの間の平均結合エネルギーの差を $X_{f(A-B)}$ とすれば、p-値<0.05の有意水準でXは下記数学式3と同じ範囲内にあると言える。

【0055】

【数7】

[数学式 3]

10

$$E_{f(A-B)} - 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}} < X_{f(A-B)} < E_{f(A-B)} + 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}}$$

【0056】

数学式3は $X_{i(A-B)}$ 、 $X_{i(B-C)}$ 、 $X_{i(C-D)}$ 、 $X_{i(A-D)}$ 値等の全てに対し適用することができ、これを介しそれぞれの $X_{i(A-B)}$ 、 $X_{i(B-C)}$ 、 $X_{i(C-D)}$ 、 $X_{i(A-D)}$ 値等が取ることができる範囲等を求めることができる。この範囲等を図式化したのが図11である。

【0057】

これまでの結果等を総合して未知のsiRNAの効率を相対的な結合エネルギーの形態を介し採点する方式は次の通りである：

20

1) 未知のsiRNAの区間A-B、B-C、C-D、A-Dでの平均結合エネルギー値、すなわち $X_{(A-B)}$ 、 $X_{(B-C)}$ 、 $X_{(C-D)}$ 、 $X_{(A-D)}$ を求める。

【0058】

2) $X_{(A-B)}$ の値が次のうち何れの範囲に属するのかを判別して次のように点数を付与する：

i)

【0059】

【数8】

$$E_{f(A-B)} - 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}} < X_{(A-B)} < E_{f(A-B)} + 1.96 \sqrt{\frac{S_{f(A-B)}}{N_f}}$$

30

【0060】

であれば10点を付与し；

ii)

【0061】

【数9】

$$E_{n(A-B)} - 1.96 \sqrt{\frac{S_{n(A-B)}}{N_n}} < X_{(A-B)} < E_{n(A-B)} + 1.96 \sqrt{\frac{S_{n(A-B)}}{N_n}}$$

40

【0062】

であれば0点を付与する。

【0063】

iii) i)とii)の範囲のいずれにも属しない場合は5点を付与する。

【0064】

$X_{(B-C)}$ 、 $X_{(C-D)}$ 、 $X_{(A-D)}$ に対しても同じ方式で点数を付与する。

【0065】

それぞれの点数を $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ という。

50

【 0 0 6 6 】

図11を参照すれば、連続的な区間において $-0.02 < X_{(A-B)} < 0.38$ 、 $-0.29 < X_{(B-C)} < -0.01$ 、 $0.00 < X_{(C-D)} < 0.35$ 、 $0.07 < X_{(A-D)} < 0.37$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =10点を付与し、 $-0.63 < X_{(A-B)} < -0.21$ 、 $0.05 < X_{(B-C)} < 0.44$ 、 $-0.47 < X_{(C-D)} < -0.09$ 、 $-0.67 < X_{(A-D)} < -0.23$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =0点を付与し、それ以外の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =5点を付与する。

【 0 0 6 7 】

不連続的な区間において $0.00 < X_{(A-B)} < 0.40$ 、 $-0.41 < X_{(B-C)} < -0.01$ 、 $0.07 < X_{(C-D)} < 0.39$ 、 $0.07 < X_{(A-D)} < 0.37$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =10点を付与し、 $-0.63 < X_{(A-B)} < -0.21$ 、 $0.10 < X_{(B-C)} < 0.51$ 、 $-0.47 < X_{(C-D)} < -0.19$ 、 $-0.67 < X_{(A-D)} < -0.23$ の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =0点を付与し、それ以外の範囲であるとき $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ =5点を付与する。

【 0 0 6 8 】

3) $Y_{(A-B)}$ 、 $Y_{(B-C)}$ 、 $Y_{(C-D)}$ 、 $Y_{(A-D)}$ の加重値をそれぞれ $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 、 $W_{(A-D)}$ とするとき、下記数学式4を利用して相対的な結合エネルギー形態の点数Yを100点満点に換算して求める。

【 0 0 6 9 】

【数 1 0】

[数学式 4]

20

$$Y = \frac{W_{(A-B)} Y_{(A-B)} + W_{(B-C)} Y_{(B-C)} + W_{(C-D)} Y_{(C-D)} + W_{(A-D)} Y_{(A-D)}}{10(W_{(A-B)} + W_{(B-C)} + W_{(C-D)} + W_{(A-D)})} \times 100$$

【 0 0 7 0 】

siRNAの結合エネルギー形態の点数化はもう一つの問題だけを残している。 $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 、 $W_{(A-D)}$ に命名された各区間の点数に対する加重値を如何に設定するかという問題である。加重値の組合せを最適化するため、各加重値の値を0から1まで0.01単位に増加させながらこのときの効率的なsiRNAグループと非効率的なsiRNAグループとの間のt-値を調査した。図12は、調査した加重値組合せ等をt-値によって降順に整理した後、そのうち最上位の100個を取ってこの100個のうち各加重値の値に従って数個ずつの組合せが表われるのかの分布を描いたものである。分布を見れば、各加重値ごとに効率的なsiRNAグループと非効率的なsiRNAグループとの間のt-値を極大化させることができる、すなわち二つのグループ間の結合エネルギー変化量の差を極大化させることができる位置を見付けることができる。二つのグループ間のt-値を極大化させた $W_{(A-B)}$ 、 $W_{(B-C)}$ 、 $W_{(C-D)}$ 、 $W_{(A-D)}$ の組合せは、連続的な区間の組合せでは $0.90 \sim 1.00$ 、 $0.2 \sim 0.4$ 、 $0.2 \sim 0.3$ 及び $0.7 \sim 0.9$ であり、望ましくは 1.00 、 0.37 、 0.20 、 0.90 で、不連続的な区間の組合せでは $0.5 \sim 0.7$ 、 $0.3 \sim 0.5$ 、 $0.3 \sim 0.5$ 及び $0.9 \sim 1.0$ であり、望ましくは 0.65 、 0.48 、 0.48 、 0.90 である。各場合において臨界値を外れることになればt-値が急激に低下することになり、採点方法自体の弁別力が別に意味のない水準に低下することになる。

30

【 0 0 7 1 】

最後の段階でこのように得られた相対的な結合エネルギー形態の点数を他の因子等(GC含量、 T_m 、絶対的な結合エネルギーの点数等、他のmRNAとの相同性、RNA二次構造など)と如何なる方法を介し結合し、siRNAの効率を総合的に予測することができるシステムを作るのかを考慮した。基本的に相対的な結合エネルギー形態の点数化と同じ方式で

40

【 0 0 7 2 】

【数 1 1】

$$S_t = \sum_i W_i S_i$$

50

【0073】

形態の線形方程式を採点方式に用いた。それぞれの因子に対し付けられた点数を Z_i (Z_1 、 Z_2 、 Z_3 、 \dots 、 Z_n)、それぞれの因子点数の満点を M_i (M_1 、 M_2 、 M_3 、 \dots 、 M_n)、各因子の効率、各点数等に対する加重値を W_i (W_1 、 W_2 、 W_3 、 \dots 、 W_n)とすれば、我々が望むsiRNAの効率を代表する点数 Z は、次の式のように100点満点で表現することができる。

【0074】

【数12】

[数学式5]

10

$$Z = 100 \times \frac{\sum_i W_i \frac{Z_i}{M_i}}{\sum_i W_i}$$

【0075】

前記で、 i は1～ n の自然数であり、 Z_i には標的mRNAに対する抑制程度に影響を及ぼす多様な因子等が適用でき、このとき前記で考慮した相対的な結合エネルギーを必須因子に含み、3'-末端5個塩基のうちA/Uの個数、1番位置のG/C存在の有無、19番位置のA/U存在の有無、G/C含量程度、 T_m 、RNA二次構造、他のmRNAとの相同性などで構成された群から選択される一つ以上の因子を選択的な因子に含むことができる。前記選択的な因子等は Z 値を割り当てるにおいて必ず含まれなければならない要素ではなく、相対的な結合エネルギーデータとともに考慮するとき、より良好な予測程度を導き出すことができる因子等を制限なく含むことができ、その因子等の組合せにおいても特別な制限があるものではない。本発明の望ましい具現例では、 Z_i に下記のような因子等を選定した: Z_1 -相対的な結合エネルギー形態の点数(Y)、 Z_2 -3'末端5個塩基のうちA/Uの個数、 Z_3 -1番の位置にG/C存在の有無、 Z_4 -19番の位置にA/U存在の有無、 Z_5 -G/C含量の点数。このとき、 M_i 値はそれぞれ次の通りである: $M_1=100$ 、 $M_2=5$ 、 $M_3=1$ 、 $M_4=1$ 、 $M_5=10$ 。

20

【0076】

本発明の望ましい具現例では Z_1 は前記で計算した点数Yであり、 Z_2 は3'末端の5個の塩基のうちA/U塩基の数で、 Z_3 は5'末端の塩基がG/Cであれば1点で、そうでない場合は0点を与え、 Z_4 は3'末端の塩基がA/Uであれば1点で、そうでない場合は0点を与え、 Z_5 であるG/C含量の場合は36～53%の範囲にある場合10点を与え、そうでない場合0点を与えた。

30

【0077】

図13は、相対的な結合エネルギー形態の点数化の場合と同じ方法で各点数等に対する加重値 W_i を最適化するため、図12のような形態のグラフを描いてみたものである。このような過程を介し最適化された W_1 、 W_2 、 W_3 、 W_4 、 W_5 の組合せは0.9～1.0、0.0～0.2、0.1～0.3及び0.0～0.2で、望ましくは0.90、0.07、0.15、0.19、0.11である。

【0078】

前記のような過程等を介し得られた Z 値は、未知のsiRNAが如何なる相対的な結合エネルギーパターンを有するのか判別することができる指標となることができ、これは塩基配列を分析することだけで結合エネルギーの状態を評価してこれを最適化することができるようにより、siRNAの設計及び製作効率を極大化させることができる。

40

【0079】

本発明の方法を介し標的mRNAに対する未知のsiRNAの抑制効率がどの程度になるのかを予測することが可能であり、抑制効率が優れるものと予想される選別されたsiRNA、望ましくは上位10%内の Z 値を有する選別されたsiRNAを用いて公知された方法によって標的mRNAに処理することにより、標的mRNAの発現を効果的に抑制することができる。前記数値は任意的な値として、候補siRNA群の標本の大きさ、実験条件等によって弾力的に適用することができる。

50

【発明を実施するための最良の形態】

【0080】

以下、本発明を実試例によって詳しく説明する。

【0081】

ただ、下記実試例は本発明を例示するためのものであるだけで、本発明の内容が下記の具体例により限定されるものではない。

【0082】

<実施例1> 従来のsiRNAデザイン方法との比較

本発明の相対的な結合エネルギー形態判別を適用したsiRNAデザイン最適化方法がどれほどの性能を発揮するかをテストするため、従来のsiRNAデザイン方法に関するWO2004/045543号特許(Functional and Hyperfunctional siRNA、2004年6月3日公開)に開示されている採点方法と比べてみた。前記特許内の幾多のアルゴリズム中に開示されているsiRNA効率採点方式は下記数学式6の通りである。

[数学式6]

$$\text{siRNAの相対的な機能性} = -(GC/3) + (AU_{15-19}) - (Tm_{20}) \times 3 - (G_{13}) \times 3 - (C_{19}) + (A_{19}) \times 2 + (A_3) + (U_{10}) + (A_{13}) - (U_5) - (A_{11})$$

Khvorovaの論文とAmarzguiouiの論文から得られた三つのデータセットの中で相対的な結合エネルギー形態の点数化の具現に用いたKhvorovaの論文から抜粋した二つのデータセットを除いて残りの一つのAmarzguiouiの論文から抜粋したデータセットをテストセットにして二つの採点方式の予測力を比較した。先ず、二つの採点方式を利用して効率的/非効率的な二つのグループに属するそれぞれのsiRNAの点数を計算した。そして、LDA(Lineal discriminant analysis)及びQDA(Quadratic discriminant analysis)を介し任意のsiRNAが効率的なのか非効率的なのかを如何によく当てるのかを計算してみた。前記値は、望ましくは統計プログラムR(<http://www.R-project.org>)を利用して求めることができる([1] Richard A. Becker, John M. Chambers, and Allan R. Wilks. *The New S Language*. Chapman & Hall, London, 1988; [2] John M. Chambers and Trevor J. Hastie. *Statistical Models in S*. Chapman & Hall, London, 1992; [3] John M. Chambers. *Programming with Data*. Springer, New York, 1998. ISBN 0-387-98503-4; [4] William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S*. Fourth Edition. Springer, 2002. ISBN 0-387-95457-0; [5] William N. Venables and Brian D. Ripley. *S Programming*. Springer, 2000. ISBN 0-387-98966-8; [6] Deborah Nolan and Terry Speed. *Stat Labs: Mathematical Statistics Through Applications*. Springer Texts in Statistics. Springer, 2000. ISBN 0-387-98974-9; [7] Jose C. Pinheiro and Douglas M. Bates. *Mixed-Effects Models in S and S-Plus*. Springer, 2000. ISBN 0-387-98957-0; [8] Frank E. Harrell. *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression*. Springer, 2001. ISBN 0-387-95232-2; [9] Manuel Castejon Limas, Joaquin Ordieres Mere, Fco. Javier de Cos Juez, and Fco. Javier Martinez de Pison Ascacibar. *Control de Calidad. Metodologia para el analisis previo a la modelizacion de datos en procesos industriales. Fundamentos teoricos y aplicaciones con R*. Servicio de Publicaciones de la Universidad de La Rioja, 2001. ISBN 84-95301-48-2; [10] John Fox. *An R and S-Plus Companion to Applied Regression*. Sage Publications, Thousand Oaks, CA, USA, 2002. ISBN 0761922792; [11] Peter Dalgaard. *Introductory Statistics with R*. Springer, 2002. ISBN 0-387-95475-9; [12] Stefano Iacus and Guido Masirotto. *Laboratorio di statistica con R*. McGraw-Hill, Milano, 2003. ISBN 88-386-6084-0; [13] John Maindonald and John Braun. *Data Analysis and Graphics Using R*. Cambridge University Press, Cambridge, 2003. ISBN 0-521-81336-0; [14] Giovanni Parmigiani, Elizabeth S. Garrett, Rafael A. Irizarry, and Scott L. Zeger. *The Analysis of Gene Expression Data*. Springer, New York, 2003. ISBN 0-387-95577-1; [15] Sylvie Huet, Annie Bouvier, Marie-Anne Gruet, and Emmanuel Jolivet. *Statistical Tools f*

10

20

30

40

50

or Nonlinear Regression. Springer, New York, 2003. ISBN 0-387-40081-8; [16] S. Mase, T. Kamakura, M. Jimbo, and K. Kanefuji. Introduction to Data Science for Engineers- Data analysis using free statistical software R (in Japanese). Suuri-Kogaku-sha, Tokyo, April 2004. ISBN 4901683128; [17] Julian J. Faraway. Linear Models with R. Chapman & Hall/CRC, Boca Raton, FL, 2004. ISBN 1-584-88425-8; [18] Richard M. Heiberger and Burt Holland. Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS. Springer Texts in Statistics. Springer, 2004. ISBN 0-387-40270-5; [19] John Verzani. Using R for Introductory Statistics. Chapman & Hall/CRC, Boca Raton, FL, 2005. ISBN 1-584-88450-9; [20] Uwe Ligges. Programmieren mit R. Springer-Verlag, Heidelberg, 2005. ISBN 3-540-20727-9, in German; [21] Fionn Murtagh. Correspondence Analysis and Data Coding with JAVA and R. Chapman & Hall/CRC, Boca Raton, FL, 2005. ISBN 1-584-88528-9; [22] Paul Murrell. R Graphics. Chapman & Hall/CRC, Boca Raton, FL, 2005. ISBN 1-584-88486-X; [23] Michael J. Crawley. Statistics: An Introduction using R. Wiley, 2005. ISBN 0-470-02297-3; [24] Brian S. Everitt. An R and S-Plus Companion to Multivariate Analysis. Springer, 2005. ISBN 1-85233-882-2; [25] Richard C. Deonier, Simon Tavare, and Michael S. Waterman. Computational Genome Analysis: An Introduction. Springer, 2005. ISBN: 0-387-98785-1; [26] Robert Gentleman, Vince Carey, Wolfgang Huber, Rafael Irizarry, and Sandrine Dudoit, editors. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health. Springer, 2005. ISBN: 0-387-25146-4; [27] Terry M. Therneau and Patricia M. Grambsch. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health. Springer, 2000. ISBN: 0-387-98784-3)。Amarzguiouiの論文から抜粋したデータセットは、Khvorovaの論文のそれとは別に効率的/非効率的な二つのグループを発現抑制効率70%を基準に分けておいた。すなわち、このデータセットで二つの採点方式の予測成功率を比べると、その差を一層明らかに見られると期待される。結果は表8の通りである。

【0083】

【表8】

表8

	相対的な結合エネルギー 形態判別	ダーマコン(Dharmacon)
LDA	0.652	0.586
QDA	0.657	0.521

【0084】

表8の結果を見れば、LDAとQDAの二つの場合全部従来のsiRNA効率採点方式より本発明の相対的な結合エネルギー形態採点方式が10%ほどその予測成功率が一層高く出るということが分かる。

【0085】

<実施例2> サバイビン(Survivin)遺伝子の発現抑制実験

本発明の相対的な結合エネルギー形態判別を適用したsiRNAデザイン最適化方法を介しサバイビン遺伝子の発現を抑制することができる36個のsiRNAをデザインした後、実際にサバイビン遺伝子の発現抑制実験を行なった。このように得られたデータセットを発現抑制効率75%を基準に効率的/非効率的な二つのグループに区分した。Khvorovaの論文とAmarzguiouiの論文から得られた三つのデータセットをトレインセットにしてサバイビンデータセットをテストセットにして実施例1と同じ方式でsiRNAの点数を採点した後、統計プログラムRを利用してLDA(Linear discriminant analysis)、QDA(Quadratic discriminant analysis)を介し任意のsiRNAが効率的なのか非効率的なのかをどれほどよく予測するのか

10

20

30

40

50

を計算してみた。その結果、LDA、QDAの二つの場合全部予測成功率が0.64で、実施例1で見せたものと殆ど同じ水準の結果を見せた(表 9)。

【0086】

【表9】

表9

番号	Exp. ID number	塩基配列 (3' overhang: TT)	配列番号	ノックダウン(%)	Z点数	正確な予測可否
1	570 (D)	GCAAUGUCUUAGGAAAGGA	37	>90	62.83	0
2	1106 (D)	AGAAUAGCACAAACUACAA	38	>90	53.31	0
3	1189 (G)	GAGACAGAAUAGAGUGAUA	39	>90	72.15	0
4	1212 (Q)	GCGUCUGGCAGAUACUCU	40	>90	68.48	0
5	299 (AS)	UGCGCUUUCCUUCUGUCA	41	75-90	40.89	
6	319 (G)	GAAGCAGUUUGAAGAAUUA	42	75-90	64.37	0
7	574 (Q) 572	UGUCUUAGGAAAGGAGAUC	43	75-90	50.92	0
8	783 (Q)	GGCAGUGUCCUUUUGCUA	44	75-90	57.52	0
9	1099 (AS)	AAUUCACAGAAUAGCACAA	45	75-90	46.80	
10	1133 (D)	AAGCACAAAGCCAUUCUAA	46	75-90	53.35	0
11	1305 (Q)	GGCAGUGGCCUAAAUCUU	47	75-90	69.63	0
12	1480 (G)	GGCUGAAGUCUGGCGUAAG	48	75-90	50.20	0
13	1481 (G)	GCUGAAGUCUGGCGUAAGA	49	75-90	45.91	
14	1585 (G)	CGGCUGUUCUCUGAGAAAUA	50	75-90	72.72	0
15	92 (D)	AAGGACCACCGCAUCUCA	51	50-75	41.57	0
16	94 (Q) 92	GGACCACCGCAUCUCA	52	50-75	71.82	
17	294 (G)	CCGGUUGCGCUUCCUUC	53	50-75	44.18	0
18	693 (D)	GCUGCUUCUCUCUCUCU	54	50-75	63.54	
19	1021 (G)	GUGAUGAGAAUGGAGAC	55	50-75	57.86	
20	1188 (G)	GGAGACAGAAUAGAGUGAU	56	50-75	57.44	
21	1394 (Q)	CCUUCACAUUCUGUCACGUU	57	50-75	57.48	
22	1546 (G)	GAUUGUUACAGCUUCUGCUG	58	50-75	57.37	
23	90 (AS)	UCAAGGACCACCGCAUCUC	59	<50	29.75	0
24	95 (G)	GACCACCGCAUCUCAUCAU	60	<50	55.86	
25	284 (Q) 282	AAGCAUUCGUCCGGUUGCG	61	<50	18.86	0
26	289 (D)	UUCGUCCGGUUGCGCUUUC	62	<50	39.01	0
27	428 (Q) 426	ACUGCGAAGAAAGUGCGCC	63	<50	23.96	0
28	780 (Q) 778	GAAGGCAGUGUCCUUUUG	64	<50	56.04	
29	807 (G)	GACAGCUUJGUUCGCGUGG	65	<50	43.89	0
30	846 (Q)	UGUGUCUGGACCUAUGUU	66	<50	47.41	0
31	1130 (Q)	ACUAAGCACAAAGCCAUUC	67	<50	47.75	0
32	1141 (Q)	AGCCAUUCUAAGUCAUJGG	68	<50	33.49	0
33	1142 (Q)	GCCAUUCUAAGUCAUUGGG	69	<50	37.58	0
34	1236 (D)	CACUGCUGUGAUUAGAC	70	<50	35.92	0
35	1325 (D)	UUAAAUGACUUGGCUUGAU	71	<50	52.86	
36	1390 (G)	CCAACCUUCACAUUCUGUCA	72	<50	63.50	
		総予測度 (23個/36個)=64%				23

【産業上の利用可能性】

【0087】

前記で検討してみたように、本発明の方法を利用すれば研究者や実験者が実際に実験し

てみなくとも未知のsiRNAの塩基配列に対する相対的な結合エネルギーのパターンを分析することにより、前記siRNAが効率的なのかまたは非効率的なのかの可否を速かに判別することができるので、siRNAの設計及び製作効率を極大化させることができ、このように選別された標的mRNAに対する効率の優れたsiRNAを用いて前記標的mRNAの発現を効果的に抑制することができる。

【図面の簡単な説明】

【0088】

【図1】図1は、RISC酵素の結合形態に従いsiRNAの遺伝子発現抑制効率が変わることを示す概略図である。

【図2】図2は、siRNAの遺伝子発現抑制効率と結合エネルギーとの間の相関関係を点数化する方法を示す概略図である。 10

【図3】図3は、INN-HBニアリストネイバーモデルにおけるsiRNAの結合エネルギーの分布を示す概略図である。

【図4】図4は、INN-HBニアリストネイバーモデルにおける結合エネルギー値を示す。

【図5】図5は、収集されたsiRNAデータの位置別結合エネルギーの平均値(mean)を示すグラフである： X軸；1番から18番までの位置、Y軸；結合エネルギー(-G)の平均値、 実線；遺伝子発現抑制効率が90%以上の場合、 点線；遺伝子発現抑制効率が50%以下の場合。

【図6】図6は、収集されたsiRNAデータの位置別結合エネルギーのt-テスト結果を示すグラフである： X軸；1番から18番までの位置、Y軸；p-値、 点線；pGL3遺伝子、 実線；hCyPB遺伝子、 半点線；Amarzguiouiの論文から抜粋した複合遺伝子。 20

【図7】図7は、収集されたsiRNAデータの位置別結合エネルギーのt-テスト結果を示すグラフである： X軸；1番から18番までの位置、Y軸；t-値、 点線；pGL3遺伝子、 実線；hCyPB遺伝子、 半点線；Amarzguiouiの論文から抜粋した複合遺伝子。

【図8】図8は、(1)の過程を介し結合エネルギーのデータを分析して選定された区間であるA(1~2)、B(3~7)、C(8~15)及びD(16~18)に対する各種の情報を示すグラフである。

【図9】図9は、特定のR値を有するA~B、B~C、C~D区間の組合せ等のうちp-値が0.05未満のものなどの比率分布を示すグラフである。

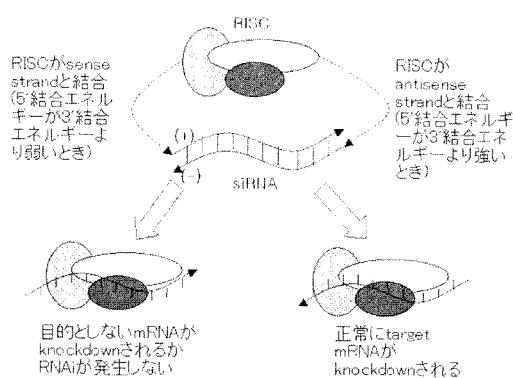
【図10】図10は、(1)と(2)の過程を介し選定された区間を示す概略図である。

【図11】図11は、(1)の過程を介し選定された区間等の組合せであるA~B、B~C、C~D及びA~Dで非効率的なsiRNAと効率的なsiRNAが有することができる平均結合エネルギーの相対的な差の信頼区間を示したグラフ(A)及び(2)の過程を介し選定された区間等の組合せであるA~B、B~C、C~D及びA~Dで非効率的なsiRNAと効率的なsiRNAが有することができる平均結合エネルギーの相対的な差の信頼区間を示したグラフ(B)である。 30

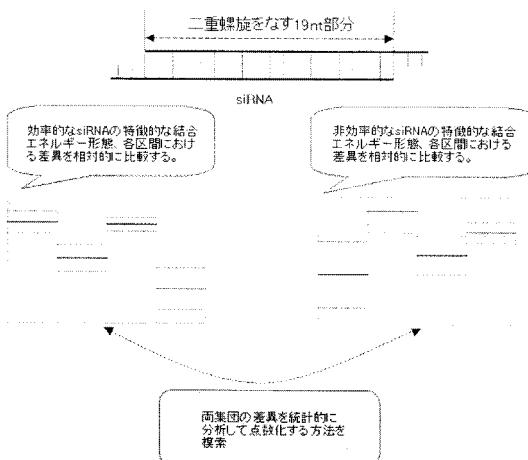
【図12】図12は、相対的な結合エネルギー形態の点数において加重値(weighting factor)とt-値との関係を示すグラフであって、加重値等の組合せをt-値に従い降順に整理した後そのうち最上位100個を選択してこれらが各区間で有する加重値の値等の個数をグラフで示したものである。Aは連続的な区間組合せにおける、Bは不連続的な区間組合せにおける加重値の分布である。

【図13】図13は、相対的な結合エネルギー形態の点数化の場合と同じ方法で各点数等に対する加重値W_iを最適化するため、図12のような形態のグラフを描いてみたものである。 40

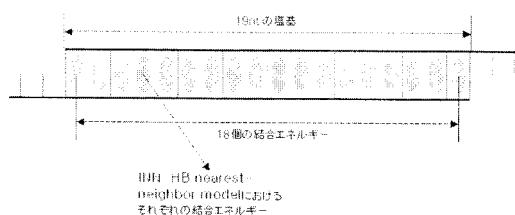
【図1】



【図2】



【図3】

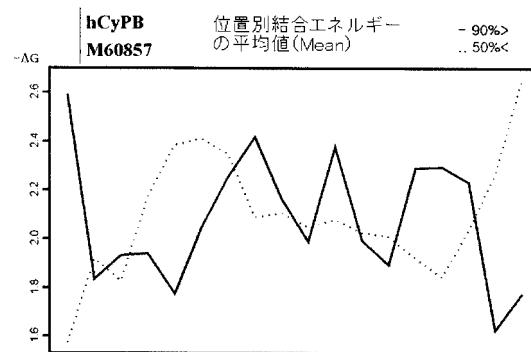


【図4】

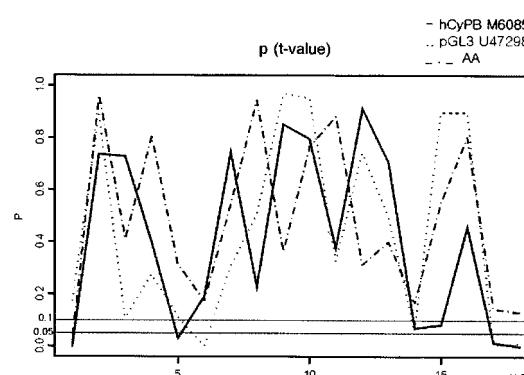
INN-HB ニアリストネイバーモデルに対するRNA熱力学パラメーター、1M NaCl, pH7 ^a			
parameters	$\Delta G_{37}^{\text{ex}}$ (kcal/mol)	ΔH^{p} (kcal/mol)	ΔS^{p} (eu)
5'AA ^{3'}	-0.93(0.03)	-6.82(0.09)	-19.0(2.5)
3'U ^{5'}			
5'AU ^{3'}	-1.10(0.08)	-9.38(1.68)	-26.7(5.2)
3'AU ^{5'}			
5'UA ^{3'}	-1.33(0.09)	-7.69(2.02)	-20.5(6.3)
3'AU ^{5'}			
5'CU ^{3'}	-2.08(0.06)	-10.48(1.24)	-27.1(3.8)
3'GA ^{5'}			
5'CA ^{3'}	-2.11(0.07)	-10.44(1.28)	-26.9(3.9)
3'GU ^{5'}			
5'GU ^{3'}	-2.24(0.06)	-11.40(1.23)	-29.5(3.9)
3'CA ^{5'}			
5'GA ^{3'}	-2.35(0.06)	-12.44(1.20)	-32.5(3.7)
3'CU ^{5'}			
5'CG ^{3'}	-2.36(0.09)	-10.64(1.65)	-26.7(5.0)
3'GC ^{5'}			
5'GG ^{3'}	-3.26(0.07)	-13.39(1.24)	-32.7(3.8)
3'CC ^{5'}			
5'GC ^{3'}	-3.42(0.08)	-14.98(1.58)	-36.9(4.9)
3'CG ^{5'}			
開始 ^c	4.09(0.22)	3.61(4.12)	-1.5(12.7)
末端AU当り ^d	0.45(0.04)	3.72'(0.83)	10.5'(2.6)
対称矯正(自体相補的)	0.43	0	-1.4
対称矯正(非自体相補的)	0	0	0

^a括弧内の数字はパラメーターに対する不確実性(uncertainty)である。
^b $\Delta G_{37}^{\text{ex}}$ 及び ΔH^{p} はニアリストネイバーパラメーターで計算する材料及び方法を参照。
^c潜在的GC及び効果を含む。
^dパラメーターは末端AU対当りである。

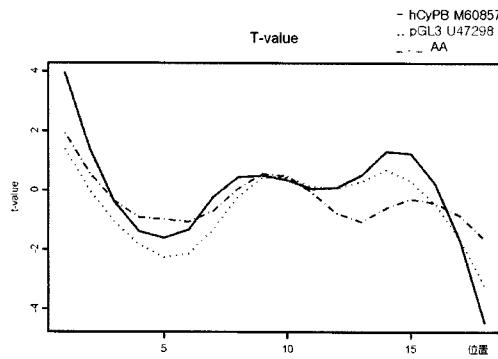
【図5】



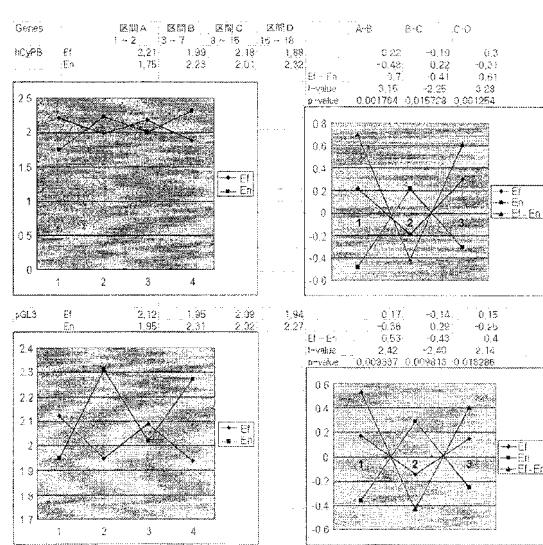
【図6】



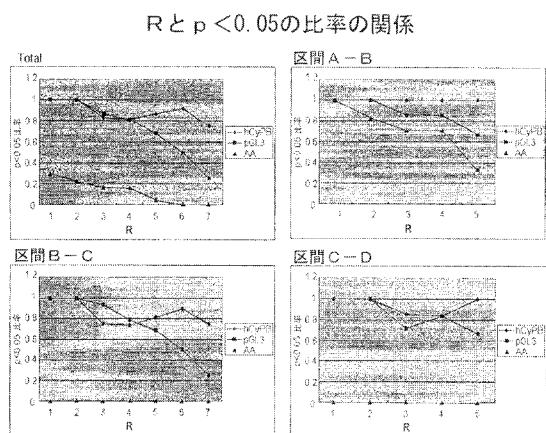
【図7】



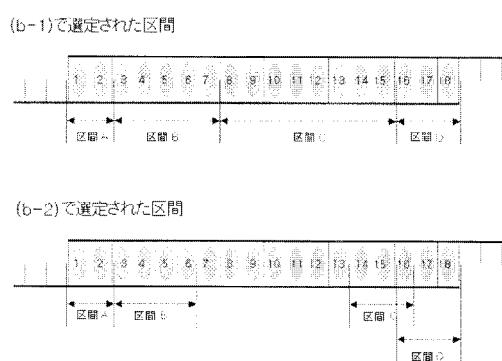
【図8】



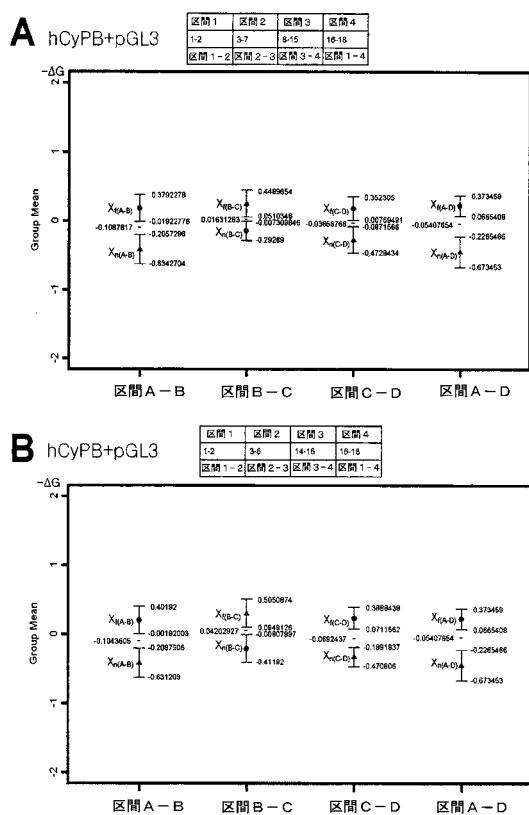
【図9】



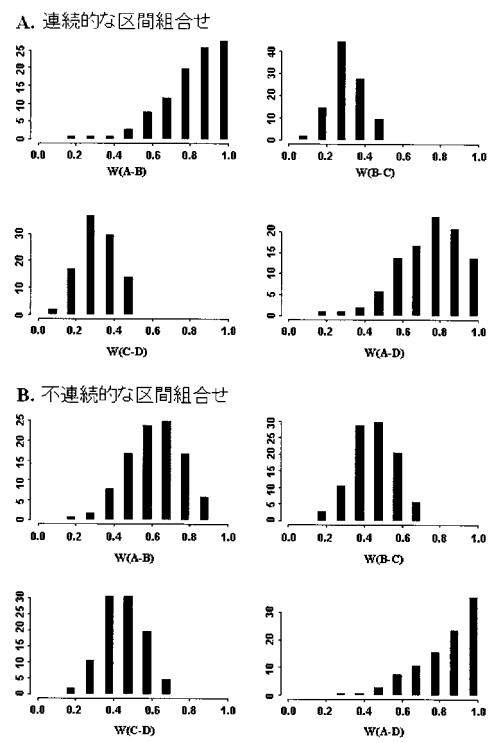
【図10】



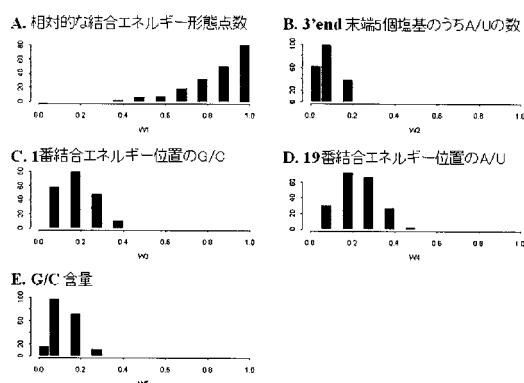
【図1-1】



【図1-2】



【図1-3】



【配列表】

0004672021000001.app

フロントページの続き

- (72)発明者 チョイ ヨング - チュル
大韓民国 305-761 ダエジョン ユースング - グ ジョンミン - ドング エキスポ アパートメント 508-401
- (72)発明者 パク ハン オー
大韓民国 305-761 ダエジョン ユースング - グ ジョンミン - ドング エキスポ アパートメント 208-601
- (72)発明者 チョング ソリミ
大韓民国 302-773 ダエジョン セオ - グ 1 - ドング ダンサン ハンマル アパートメント 105-1104
- (72)発明者 キム ヨング ジョー
大韓民国 305-761 ダエジョン ユースング - グ ジョンミン - ドング エキスポ アパートメント 102-1206
- (72)発明者 キム サング ソー
大韓民国 302-782 ダエジョン セオ - グ サムシェオン - ドング ゴークワ アパートメント 505-802
- (72)発明者 パク セオン - ミン
大韓民国 305-805 ダエジョン ユースング - グ シンセオン - ドング 201-65310
- (72)発明者 キム サン - チョル
大韓民国 136-861 ソウル セオングブク - グ ヨンガルム 1 - ドング 62 105
- (72)発明者 ヨーン ギュマン
大韓民国 305-805 ダエジョン ユースング - グ シンセオン - ドング 209-20101
- (72)発明者 チョイ キョング オーク
大韓民国 780-932 ギョンサンブク - ド ギョンジュ - シ ノセオ - ドング 162-18
- (72)発明者 カン ヒョ ジン
大韓民国 305-338 ダエジョン ユースング - グ グセオング - ドング 373-1 ドクシンリョウ 2-307

審査官 飯室 里美

(56)参考文献 Oligonucleotides , 2004年 , Vol.14, No.3 , p.227-232
Biochemical and Biophysical Research Communications , 2004年 , Vol.319 , p.264-274

(58)調査した分野(Int.Cl. , DB名)
C12N 15/09
BIOSIS/MEDLINE/WPIDS(STN)
JSTPlus/JMEDPlus/JST7580(JDreamII)