



Urząd Patentowy
Rzeczypospolitej Polskiej

(21) Numer zgłoszenia: **335786**

(22) Data zgłoszenia: **30.09.1999**

(51) Int.Cl.
C12Q 1/68 (2006.01)
C12N 15/10 (2006.01)

(54) **Sposób sekwencjonowania kwasów nukleinowych**

(43) Zgłoszenie ogłoszono:
09.04.2001 BUP 08/01

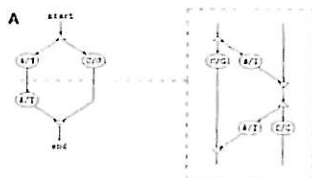
(45) O udzieleniu patentu ogłoszono:
29.02.2008 WUP 02/08

(73) Uprawniony z patentu:
**Instytut Chemii Bioorganicznej
PAN,Poznań,PL**

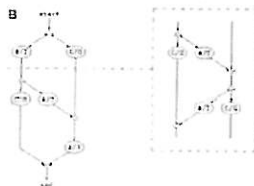
(72) Twórca(y) wynalazku:
Jacek Błażewicz,Poznań,PL
Piotr Formanowicz,Poznań,PL
Marta Kasprzak,Poznań,PL
Wojciech Tadeusz Markiewicz,Poznań,PL

(74) Pełnomocnik:
**Mirosław Szykuła, Biuro Usług
Prawno-Technicznych, "Dr A. Au & Co"**

- (57) 1. Sposób sekwencjonowania kwasów nukleinowych, poprzez prowadzoną w znany sposób hybrydyzację kwasu nukleinowego o nieznanej sekwencji z wykorzystaniem zestawu oligonukleotydów, **znamienny tym**, że tworzy i wykorzystuje się zestaw oligonukleotydów zawierający co najmniej 50% jakiejś biblioteki izorelacyjnej stanowiącej zbiór wszystkich oligonukleotydów spełniających relację taką, że $w_A x_A + w_T x_T + w_C x_C + w_G x_G = C_i$ gdzie w_A, w_C, w_G, w_T oznaczają wkłady wnoszone odpowiednio przez nukleotydy A, C, G, T, zmienne x_A, x_C, x_G, x_T oznaczają liczby odpowiednich nukleotydów w oligonukleotydzie, a C_i oznacza stałą dla danej biblioteki izorelacyjnej, z pominięciem sytuacji, gdy $w_A = w_C = w_G = w_T$.
2. Sposób według zastrz. 1, **znamienny tym**, że korzystnie bibliotekę izorelacyjną oligonukleotydów tworzy się jako izotermiczną, czyli taką, dla której parametry w_A, w_C, w_G, w_T pozostają między sobą w następujących relacjach takich, że: $w_A = w_T, w_C = w_G$ oraz $2w_A = w_C$, przy czym tę bibliotekę generuje się zgodnie ze schematem (A);



jeżeli C_i jest podzielne bez reszty przez $2w_A$ lub w przypadku przeciwnym zgodnie ze schematem (B.)



Opis wynalazku

Przedmiotem wynalazku jest sposób sekwencjonowania kwasów nukleinowych, przeznaczony do stosowania w biologii molekularnej jak i w biotechnologii. Kwasy nukleinowe pełnią w organizmach żywych wielorakie i podstawowe funkcje, takie jak przechowywanie informacji genetycznej, jej przekazywanie, odczytywanie i tłumaczenie na struktury białek strukturalnych i enzymatycznych. Znajomość sekwencji kwasów nukleinowych ma wielkie znaczenie praktyczne dla rozwoju nowoczesnych metod diagnostyki medycznej i rozwoju nowych biotechnologii. Sekwencjonowanie DNA, czyli zespół czynności mających na celu określenie kolejności występowania jednostek nukleotydowych w łańcuchu kwasu nukleinowego, stanowi jeden z najważniejszych problemów biologii molekularnej. Możliwość określenia sekwencji nukleotydów w DNA powinna przybliżyć molekularne rozpoznanie i kontrolę metabolizmu zarówno człowieka jak i zwierząt i roślin. Powinna ułatwić opanowanie chorób zarówno poprzez projektowanie nowych, skuteczniejszych leków jak i nowe terapie typu terapii genowych, itp. Postęp w rozwoju poznania genomu człowieka a także genomów ważnych roślin uprawnych, zwierząt hodowlanych i organizmów chorobotwórczych jest bardzo istotnie zależny od opracowania nowych metod sekwencjonowania kwasów nukleinowych. Poznanie sekwencji DNA organizmów żywych umożliwić powinno szybsze i tańsze wyprowadzenie nowych odmian roślin uprawnych, nowych ras zwierząt hodowlanych i skuteczniejsze zwalczanie chorób. Także określenie sekwencji RNA i analogów kwasów nukleinowych jest istotnym problemem zarówno w biologii molekularnej jak i biotechnologii.

Znane obecnie i stosowane sposoby sekwencjonowania DNA wykorzystują przede wszystkim różnice w mobilności elektroforetycznej fragmentów kwasów nukleinowych stanowiących krótsze podfragmenty sekwencjonowanego DNA/RNA, które są wytwarzane albo na drodze ich degradacji chemicznej lub enzymatycznej albo na drodze „niekompletnej” syntezy enzymatycznej. Znane są dwie metody sekwencjonowania DNA: metoda chemiczna i metoda enzymatyczna. Metoda chemiczna polega na przeprowadzeniu specyficznych reakcji chemicznych zasad heterocyklicznych wchodzących w skład sekwencjonowanego kwasu nukleinowego i poddaniu go następnie kontrolowanej degradacji. W wyniku tych procesów uzyskuje się mieszaninę krótszych fragmentów oligonukleotydowych, których rozdział elektroforetyczny pozwala na określenie sekwencji analizowanego kwasu nukleinowego (Maxam A.M., Gilbert W., A new method for sequencing DNA: *Proc. Natl. Acad. Sci. U.S.A.* 1977 74 560-564; Maxam A.M., Gilbert W.: Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.*, 1977 65 499-560.). Podobnie, druga metoda sekwencjonowania opiera się na wytworzeniu krótszych fragmentów oligonukleotydowych na matrycy analizowanego kwasu nukleinowego na drodze reakcji enzymatycznej polimeryzacji z wykorzystaniem odpowiednich terminatorów polimeryzacji (Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors: *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467.).

Metoda ta doczekała się opracowania wielu wariantów i jest stosowana obecnie najpowszechniej. Jest to jednak metoda droga i stosunkowo wolna. Cel sekwencjonowania genomów dużych organizmów przekracza możliwości pojedynczych laboratoriów i jest w praktyce podejmowany tylko w ramach międzynarodowych wieloletnich i bardzo kosztownych programów badawczych angażujących dziesiątki ośrodków naukowych.

Szybkość i koszt sekwencjonowania z wykorzystaniem powyższych metod stanowią poważne ograniczenie w praktycznym wykorzystaniu informacji zawartej w sekwencji DNA. Od szeregu lat trwają usilne poszukiwania nowych podejść metodycznych do problemu sekwencjonowania DNA. Poszukuje się metod szybszych i tańszych. Wśród nowych metod sekwencjonowania kwasów nukleinowych należy wymienić zmierzające do wykorzystania detekcji pojedynczych znakowanych bądź nie znakowanych nukleotydów przez enzym o właściwościach egzonukleolitycznych z pojedynczej, unieruchomionej na podłożu stałym nici sekwencjonowanego DNA (Jett, J.H.; Keller, R.A.; Martin, J.C.; Marro-ne, B.L.; Moyzis, R.K.; Ratliff, R.L.; Seitzinger, N.K.; Shera, E.B.; Stewart, C.C.: High-Speed DNA Sequencing: An Approach Based Upon Fluorescence Detection of Single Molecules. *J. Biomolec. Struct. Dynam.* 1989;7:301-309.) Metody sekwencjonowania pojedynczych nici kwasów nukleinowych są we wczesnej fazie ich opracowywania i nie mają obecnie praktycznego znaczenia. Pośród nowych metod sekwencjonowania DNA, których opracowaniu poświęcono dużo uwagi, należy wymienić tzw. sekwencjonowanie poprzez hybrydyzację (ang. *sequencing by hybridization*, SBH). Odwracalny proces asocjacji kwasów nukleinowych i ich fragmentów (oligonukleotydów) o komplementarnych sekwencjach ma zasadnicze znaczenie dla biologii molekularnej, ale także dla nowych biotechnologii. Jednym z zastosowań asocjacji kwasów nukleinowych i komplementarnych do nich fragmentów kwa-

sów nukleinowych czyli oligonukleotydów jest analiza obecności określonych sekwencji oligonukleotydowych na podstawie wyniku eksperymentu hybrydyzacji. Doświadczenie hybrydyzacji można przeprowadzić stosując albo znakowany radioizotopowo lub nieradioizotopowo kwas nukleinowy albo znakowany komplementarny oligonukleotyd. Znakowanie pozwala na przeprowadzenie pomiaru z bardzo dużą czułością i stwierdzenie faktu utworzenia komplementarnego kompleksu (dupleksu).

Tak więc metoda hybrydyzacji kwasów nukleinowych i ich fragmentów (oligonukleotydów) o komplementarnych sekwencjach może służyć do uzyskania wiedzy co do obecności bądź nieobecności w analizowanym kwasie nukleinowym określonych sekwencji oligonukleotydowych. Jeśli takie doświadczenia hybrydyzacji zostałyby przeprowadzone dla wszystkich możliwych kombinacji sekwencyjnych, to po ich przeprowadzeniu uzyskano by informację na temat wszystkich sekwencji zawartych w sekwencjonowanym kwasie nukleinowym.

Jednym ze sposobów przeprowadzenia sekwencjonowania kwasów nukleinowych przez hybrydyzację jest przeprowadzenie procesu hybrydyzacji kwasu nukleinowego o nieznanej sekwencji z wykorzystaniem pełnego zestawu oligonukleotydów o określonej długości, zwanego biblioteką standardową.

Na podstawie danych sekwencjonowania SBH można określić sekwencję nieznanego kwasu nukleinowego stosując odpowiednie procedury obliczeniowe. Opublikowano kilka doniesień literaturowych na ten temat. Jednocześnie wiadomo, że dane pomiarowe mogą zawierać błędy w ocenie obecności sekwencji oligonukleotydowych w analizowanym DNA. Źródła tych błędów mogą być różnorodne. Jednym z nich jest występowanie sekwencji powtórzonych w obrębie sekwencji nieznanego kwasu nukleinowego. Inne błędy mają swoje źródło w naturze zjawiska hybrydyzacji komplementarnych kwasów nukleinowych. Pojawienie się błędów w zestawie danych powoduje, że znalezienie na drodze obliczeń właściwej sekwencji kwasu nukleinowego staje się problemem obliczeniowo trudnym. Oznaczać to może znaczne przedłużenie etapu obliczeniowego analizy sekwencyjnej za pomocą SBH, a w niektórych przypadkach znalezienie jednego prawidłowego rozwiązania może nie być możliwe. W niektórych przypadkach dane mogą zawierać tak wiele błędów pozytywnych, że ich obecność i rodzaj całkowicie uniemożliwiają znalezienie prawidłowego rozwiązania. Z drugiej strony przeprowadzenie pomiaru hybrydyzacji w warunkach, które doprowadzą do zmniejszenia liczby błędów pozytywnych może doprowadzić do uzyskania danych o zwiększonej liczbie błędów negatywnych. Również i w takim przypadku znalezienie prawidłowego rozwiązania może nie być możliwe, gdyż pewne części analizowanej sekwencji DNA nie zostaną w ogóle odczytane.

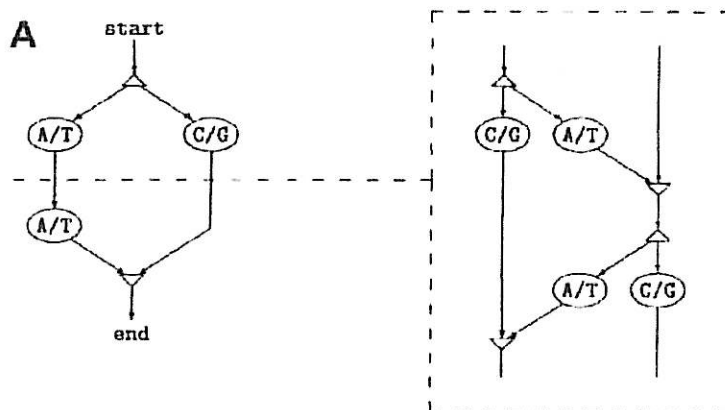
Opisane sposoby sekwencjonowania kwasów nukleinowych i ich analogów poprzez SBH lub warianty SBH są realizowane z wykorzystaniem do hybrydyzacji z kwasem nukleinowym o nieznanej sekwencji oligonukleotydów o jednakowej długości. Jednym z wariantów takiego sposobu realizacji SBH jest wykorzystanie oligonukleotydów przyłączonych do podłoża stałego albo poprzez 3' albo poprzez 5' koniec lub związanych z podłożem stałym w inny sposób (patenty: Fodor, S. P. A.; Aldwin L.; Ross D. A.; Winkler J. L.; Buchko C. J., Very large scale immobilized polymer synthesis using mechanically directed flow, Patent US5885837, 1999, Bradley J.-C.; Pirrung M. C.; Shuey S. W., 5' to 3' nucleic acid synthesis using 3'-photoremovable protecting group, Patent US5908926, 1999) zgodnie z wiedzą znaną specjalistom w dziedzinie.

Wiadomo od dawna, że duplekсы DNA bogate w pary C/G są trwalsze niż duplekсы zawierające dużo par A/T. Ta właściwość kwasów nukleinowych stanowi istotną przeszkodę doświadczalną we wszystkich metodach wykorzystujących zjawisko hybrydyzacji. Może ona być źródłem licznych błędów pozytywnych w SBH. Modyfikacja warunków prowadzenia hybrydyzacji mająca na celu zmniejszenie liczby nieprawidłowych dupleksów z udziałem oligonukleotydów bogatych w nukleotydy C/G może prowadzić do zmniejszenia liczby poprawnie odczytanych prawidłowych dupleksów oligonukleotydów bogatych w nukleotydy A/T. Innymi słowy taka modyfikacja warunków hybrydyzacji, która zmniejsza liczbę błędów pozytywnych może prowadzić do zwiększenia liczby błędów doświadczalnych negatywnych. Zależność tworzenia się dupleksu od składu nukleotydowego oligonukleotydów można zmniejszyć prowadząc hybrydyzację w obecności soli chaotropowych takich jak chlorek tetrametyloamonowy (Maskos, U.; Southern, E.M., A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support, *Nucleic Acids Research* 1993, 21, 4663-4669). Mimo tej możliwości wciąż poszukuje się nowych rozwiązań techniczno-doświadczalnych, które pozwoliłyby zmniejszyć liczbę błędów w SBH i jej wariantach, aby rozwinąć te metody sekwencjonowania kwasów nukleinowych i uczynić z nich metody o znaczeniu praktycznym. Wśród prób modyfikacji SBH należy w tym miejscu wymienić metody modyfikacji struktury oligonukleotydów zmniejszające zależność tworzenia się dupleksu od jego składu nukleotydowego poprzez przyłączenie dodatkowych funkcji (mody-

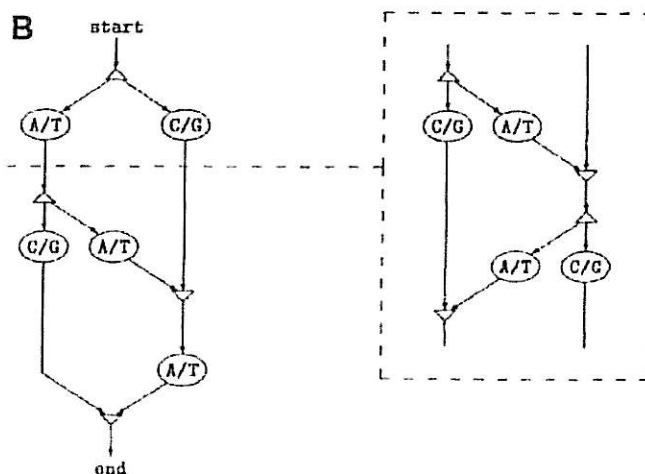
fikacji cząsteczkowych) zwiększających trwałość bogatych w pary A/T. Przykładem takiego rozwiązania są koniugaty oligonukleotydów i związków o właściwościach wiązania kompleksowego w rejonie małej bruzdy duplexu DNA (Kutyavin I.V., Lukhtanov E. A., Gamper H. B., Meyer R. B., Covalently Linked Oligonucleotide Minor Groove Binder Conjugates, Patent US5801155, 1998). Rozwiązania techniczne w tym względzie są wciąż w fazie intensywnych poszukiwań chemiczno-technicznych i ze względu na znaczenie problemu jakie ma sekwencjonowanie kwasów nukleinowych istnieje duże zapotrzebowanie na nowe rozwiązania w tym zakresie. Istotne jest także ścisłe sprzężenie prac chemiczno-technicznych z opracowaniem nowych analiz teoretycznych powstających na gruncie nauk informatyczno-biologicznych (bioinformatyczno-technicznych).

W sposobie sekwencjonowania kwasów nukleinowych według wynalazku, poprzez prowadzoną w znany sposób hybrydację kwasu nukleinowego o nieznannej sekwencji z wykorzystaniem zestawu oligonukleotydów, istota rozwiązania polega na tym, że tworzy i wykorzystuje się zestaw oligonukleotydów zawierający co najmniej 50% jakiejś biblioteki izorelacyjnej stanowiącej zbiór wszystkich oligonukleotydów spełniających relację taką, że $w_A x_A + w_T x_T + w_C x_C + w_G x_G = C_i$, gdzie w_A, w_C, w_G, w_T oznaczają wkłady (inkrementy) wnoszone odpowiednio przez nukleotydy A, C, G, T, zmienne x_A, x_C, x_G, x_T oznaczają liczby odpowiednich nukleotydów w oligonukleotydzie, a C_i oznacza stałą dla danej biblioteki izorelacyjnej, z pominięciem sytuacji, gdy $w_A = w_C = w_G = w_T$.

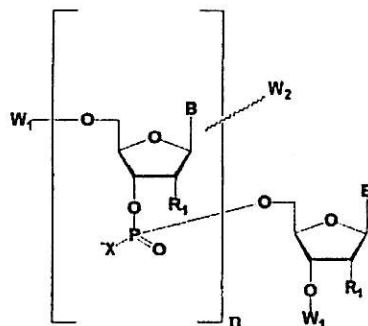
Korzystnie bibliotekę izorelacyjną oligonukleotydów tworzy się jako izotermiczną, czyli taką, dla której parametry w_A, w_C, w_G, w_T pozostają między sobą w następujących relacjach takich, że: $w_A = w_T, w_C = w_G$ oraz $2w_A = w_C$, przy czym tę bibliotekę generuje się zgodnie ze schematem A:



Jeżeli C_i jest podzielone bez reszty przez $2w_A$ lub w przypadku przeciwnym zgodnie ze schematem B



Korzystnie jest również, gdy oligonukleotyd izotermicznej biblioteki oligonukleotydowej konstruuje się z jednostek nukleotydowych zgodnie z wzorem:



gdzie:

n jest liczbą całkowitą od 1 do 100

B jest aglikonem spośród zasad heterocyklicznych występujących w naturalnych kwasach nukleinowych lub inną zasadą heterocykliczną, której oddziaływanie z komplementarnymi zasadami heterocyklicznymi jest znane i odpowiada oddziaływaniom międzycząsteczkowym, w których biorą udział zasady heterocykliczne występujące w DNA: adenina (Ade), cytozyna (Cyt), guanina (Gua) i tymina (Thy);

X jest O lub S

R_1 jest H, OH, lub grupą alkoksylową albo alkenylową albo F

W_1 jest H, $PO(OH)_2$ lub jej solą lub łącznikiem z podłożem stałym

W_2 jest nieobecne albo jest łącznikiem z podłożem stałym gdy W_1 jest H, $PO(OH)_2$ lub jej solą.

Fakt, że duplekisy DNA bogate w pary C/G są trwalsze niż duplekisy zawierające dużo par A/T, stanowi istotną przeszkodę doświadczalną we wszystkich metodach sekwencjonowania kwasów nukleinowych wykorzystujących zjawisko hybrydyzacji, i może być źródłem licznych błędów pomiarowych. Tworzenie dupleksu zależy od składu nukleotydowego, ale i od jego długości. We wcześniejszych badaniach z wykorzystaniem allelo-specyficznych oligonukleotydów w badaniach mutacji DNA korzystano z prostej zależności opisującej temperatury topnienia (mięknięcia) dupleksów przyjmując wkład 4°C (stopni) dla pary C/G i 2°C (stopni) dla pary A/T (Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H., and Itakura K., The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit beta-globin DNA. Nucleic Acids Research 9, 879-894, 1981). Tak więc, formalnie biorąc, można skompensować mniejszą trwałość dupleksów bogatych w pary A/T poprzez zwiększenie ich długości. Wiadomo, że powyższy opis nie jest bardzo dokładny, ale odzwierciedla ogólną zależność trwałości od składu nukleotydowego i długości dupleksu dość dobrze. Lepszego opisu stabilności dupleksów dostarcza tzw. model najbliższych sąsiadów (Breslauer K.J. Frank R. Blöcker H. Marky LA. Predicting DNA duplex stability from the base sequence. Proc. National Acad. Sciences USA, 83,3746-3750, 1986). Nieoczekiwanie okazało się, że opisany wyżej prosty sposób obliczenia stabilności dupleksów z wykorzystaniem wyłącznie liczby par A/T i C/G stwarza nowe możliwości rozwiązywania problemów sekwencjonowania kwasów nukleinowych przez hybrydyzację.

Zgodnie ze sposobem według wynalazku uzyskuje się zestaw oligonukleotydów spełniających relację: $w_A x_A + w_T x_T + w_C x_C + w_G x_G = C_i$, gdzie w_A, w_C, w_G, w_T oznaczają wkłady (inkrementy) wnoszone odpowiednio przez nukleotydy A, C, G, T, zmienne x_A, x_C, x_G, x_T oznaczają liczby odpowiednich nukleotydów w oligonukleotydzie, a C_i oznacza stałą dla danej biblioteki, przy czym zestawy wszystkich oligonukleotydów spełniających powyższą relację nazwano *izorelacyjnymi bibliotekami oligonukleotydowymi*. W przypadku, gdy $w_A = w_C = w_G = w_T$, mamy bibliotekę, odpowiadającą standardowej bibliotece zawierającej oligonukleotydy o równej długości, nie objętej istotą sposobu według wynalazku.

Szczególnym przypadkiem oligonukleotydowej biblioteki izorelacyjnej jest *izotermiczna biblioteka oligonukleotydowa* powstająca wtedy, gdy wkład (inkrement) nukleotydu C(G) jest dwa razy większy niż nukleotydu A(T). Suma wkładów (inkrementów) C oligonukleotydu jest charakterystyczna dla danej biblioteki izotermicznej i nazywana jest *temperaturą biblioteki* (T_i).

Oligonukleotydy zawarte w takiej bibliotece powinny tworzyć duplekisy z komplementarnymi kwasami nukleinowymi w węższym zakresie warunków doświadczalnych (temperatura, stężenie soli, itp.) niż zakres charakterystyczny dla kompletnej biblioteki zawierającej oligonukleotydy jednakowej

długości. Zatem, doświadczenia hybrydyzacyjne (pomiar) przeprowadzone z wykorzystaniem bibliotek izotermicznych powinny dostarczyć danych zawierających mniejszą liczbę błędów doświadczalnych. Ta właściwość izotermicznych bibliotek oligonukleotydowych była impulsem do przeprowadzenia analizy ich przydatności do sekwencjonowania kwasów nukleinowych przez hybrydyzację (SBH). Jak powiedziano powyżej, zastosowanie oligonukleotydowych bibliotek izotermicznych w doświadczeniach hybrydyzacyjnych powinno dostarczyć danych pomiarowych obarczonych dużo mniejszą liczbą błędów. Przeprowadzona analiza matematyczna problemu przydatności oligonukleotydowych bibliotek izotermicznych do SBH i wariantów wykazała, że dwie biblioteki izotermiczne są wystarczające dla pełnego odzwierciedlenia wszelkich możliwych sekwencji kwasów nukleinowych, gdy ich temperatury różnią się o jeden wkład A lub T. Przeprowadzona analiza wykazała, że jedna biblioteka izotermiczna nie jest wystarczająca dla SBH. Mimo to pomiary hybrydyzacyjne z użyciem jednej biblioteki izotermicznej mogą w przypadkach szczególnych być wystarczające dla sekwencjonowania określonych sekwencji kwasów nukleinowych. Ponadto lepsza jakość pomiarowych danych z doświadczeń hybrydyzacyjnych z użyciem pojedynczej biblioteki izotermicznej może być wykorzystana do odtworzenia sekwencji badanego kwasu nukleinowego ze względu na nadmiarowość danych prowadząc do rozwiązań pewnych lub bardzo prawdopodobnych z nie gorszą skutecznością niż SBH z wykorzystaniem standardowych bibliotek oligonukleotydowych.

Przeprowadzona analiza problemu sekwencjonowania z wykorzystaniem bibliotek izotermicznych doprowadziła do jednoznacznego wniosku, że dwie biblioteki izotermiczne różniące się temperaturą o jeden wkład $A(T)$: T_I i T_I+2 są wystarczające do sekwencjonowania kwasu nukleinowego o dowolnej sekwencji.

Opracowany został również sposób przechodzenia na drodze matematyczno-obliczeniowej od zestawu oligonukleotydów zidentyfikowanych podczas pomiarów hybrydyzacyjnych za pomocą biblioteki (bibliotek) izotermicznych do sekwencji analizowanego kwasu nukleinowego.

Istotne znaczenie ma także mniejsza liczba oligonukleotydów w izotermicznych bibliotekach oligonukleotydowych w porównaniu z liczbą oligonukleotydów w standardowych bibliotekach oligonukleotydowych. Jak wiadomo liczba oligonukleotydów w standardowych bibliotekach oligonukleotydowych rośnie wykładniczo wraz z długością oligonukleotydu i w przypadku standardowych fragmentów DNA zbudowanych z czterech nukleotydów A, C, G i T wynosi:

$$4^k, \text{ gdzie } k - \text{długość oligonukleotydu}$$

W przypadku bibliotek izotermicznych ich licznosc opisana jest dwoma wzorami, jednym dla bibliotek o temperaturze T_I podzielnej bez reszty przez $2w_A$:

$$\sum_{i=0}^{\frac{T_I}{2w_A}} \left[\left(\frac{T_I}{2w_A} + i \right) \cdot 2^{\frac{T_I}{2w_A} + i} \right]$$

i drugim, dla bibliotek o temperaturze T_I niepodzielnej bez reszty przez $2w_A$:

$$\sum_{i=0}^{\left\lfloor \frac{T_I}{2w_A} \right\rfloor} \left[\left(\left\lfloor \frac{T_I}{2w_A} \right\rfloor + i + 1 \right) \cdot 2^{\left\lfloor \frac{T_I}{2w_A} \right\rfloor + i + 1} \right]$$

Długość oligonukleotydów (l) wchodzących w skład odpowiedniej biblioteki izorelacyjnej można określić z relacji podstawowej, pamiętając o prostej zależności:

$$D = x_A + x_B + x_C + x_D$$

Przedmiot wynalazku został bliżej przedstawiony na przykładach realizacji sposobu według wynalazku, objaśnionym rysunkiem, na którym na którym na fig. 1 i 2 uwidoczono schematy generacji izotermicznych bibliotek oligonukleotydowych.

W pierwszym przykładzie realizacji sposobu sekwencjonowania kwasów nukleinowych według wynalazku, zestaw oligonukleotydów wchodzących w skład biblioteki izotermicznej stanowi scaloną bibliotekę oligonukleotydów. Nomenklatura w zakresie bibliotek została przedstawiona wcześniej (Markiewicz W. T., Andrych-Rożek K., Markiewicz M., Żebrowska A., Astriab A., Synthesis of oligonucleotides permanently linked with solid supports for use as synthetic oligonucleotide combinatorial libraries. *Innovations in solid phase synthesis*. W Epton R. (red.), Biological and Biomedical Applications, Mayflower Worldwide, 1994, Birmingham, 339-346). W scalonej bibliotece oligonukleotydowej oligonukleotydy są przyłączone do powierzchni materiału stałego w taki sposób, że każdy oligonukleotyd z izotermicznej biblioteki oligonukleotydowej zajmuje określone pole powierzchni. Do określonego pola scalonej biblioteki oligonukleotydowej przyłączona jest co najmniej jedna cząsteczka oligonukleotydu zdolna do wzięcia udziału w procesie hybrydyzacji z sekwencjonowanym kwasem nukleinowym lub jego fragmentem.

Scalone biblioteki oligonukleotydowe tworzą albo poprzez syntezę oligonukleotydów na drodze chemicznej syntezy kombinatorycznej tak, jak to opisano w literaturze dla tzw. chipów DNA lub zestawów DNA (ang. DNA chip lub DNA array) (Fodor S.P.A., Pirrung M.C., Stryer L., Read J.L., Very large scale immobilized polymer synthesis, Patent US5424186; Pirrung M.C., Shuey S.W., Bradley J.-C., 5' to 3' nucleic acid synthesis using 3'-photoremovable protecting Group, Patent US5908926), lub poprzez oddzielne otrzymanie oligonukleotydów na drodze chemicznej i ich następne przyłączenie do wybranych pól scalonej biblioteki oligonukleotydowej (Timofeey E., Mirzabekov A., Sundberg S., Fujimoto D., McGall G., Schullek J., Yang H., Holmes C.P., Jacobs J.W., Derivatization of solid supports and methods for oligomer synthesis, Patent US5919523). Ten sam sposób postępowania stosuje się do scalonej izotermicznej biblioteki oligonukleotydowej. Oligonukleotydy są przyłączone do podłoża stałego poprzez ich 3' lub 5'-koniec albo poprzez inną grupę funkcyjną nie znajdującą się na 3' lub 5'-końcu oligonukleotydu w taki sposób, że możliwa jest hybrydyzacja z sekwencjonowanym kwasem nukleinowym lub jego fragmentami. Oligonukleotyd może być albo zbudowany z naturalnych deoksynukleotydów lub deoksynukleotydów pochodnych zasad heterocyklicznych takich jak uracyl lub inne zasady heterocykliczne, których oddziaływania z komplementarnymi zasadami heterocyklicznymi są znane i odpowiadają oddziaływaniom międzycząsteczkowym, w których biorą udział zasady heterocykliczne występujące w DNA: adenina (Ade), cytozyna (Cyt), guanina (Gua) i tymina (Thy).

Oligonukleotyd może również zawierać inne modyfikacje nie zmieniające specyficzności jego oddziaływania z komplementarnym kwasem nukleinowym lub jego fragmentami a jedynie modyfikującymi siłę tych oddziaływań, przy czym modyfikacje te mogą być dołączone do oligonukleotydu wchodzącego w skład izotermicznej biblioteki oligonukleotydowej albo na 3' albo na 5' końcu albo w innej części jego łańcucha w taki sposób, że możliwa jest hybrydyzacja z sekwencjonowanym kwasem nukleinowym lub jego fragmentami.

Syntezę kombinatoryczną scalonej biblioteki izotermicznej realizuje się stosując opisane znane sposoby z użyciem odpowiednio aktywowanych monomerów nukleotydowych, które zawierają grupy fotolabilne lub inne grupy ochronne selektywnie odślaniane bądź za pomocą światła bądź innych właściwych czynników działających lokalnie na wybranym polu otrzymywanej izotermicznej biblioteki scalonej. Jako grupy ochronne usuwane za pomocą światła nie naruszającego budowy oligonukleotydu można wykorzystać pochodne typu o-nitrobenzylowego lub inne opisane w literaturze do syntezy scalonych bibliotek oligonukleotydowych. Odpowiednie maski stosowane na poszczególnych etapach kombinatorycznej syntezy scalonej izotermicznej biblioteki oligonukleotydowej, których kształt i kolejność użycia wynikają ze struktury scalonej biblioteki izotermicznej, zaprojektowano wykorzystując schematy A, B przedstawione na fig. 1 i 2. W przypadku syntezy biblioteki izotermicznej, której temperatura T_i jest podzielna przez $2w_A$ bez reszty, stosuje się schemat A, przedstawiony na fig. 1, a w przypadku przeciwnym stosuje się schemat B przedstawiony na fig. 2. Oznaczenia na fig. 1 mają następujący sens: Δ oznacza rozdzielanie, a ∇ łączenie etapów, A/T oznacza czynność rozdzielania na dwie części, przyłączania do jednej A do drugiej T i ponowne połączenie, a znak C/G ma to samo znaczenie jak dla A/T, ale dla etapów przyłączania jednostek wbudowujących C i G.

W przypadku, gdy T_i biblioteki izotermicznej wynosi $2w_A + 2tw_A$ należy t razy wstawić schemat A przedstawiony w prawej części fig. 1 do schematu z jej lewej części, w zaznaczonym miejscu. W przypadku bibliotek, dla których $T_i/2w_A$ nie jest liczbą całkowitą i resztę równą w_A i $T_i = 3w_A + 2tw_A$, należy do schematu B postępowania pokazanego na fig. 2 z lewej strony włączyć w zaznaczonym miejscu t razy schemat z prawej części fig. 2.

Otrzymaną bibliotekę izotermiczną poddano hybrydyzacji z wyznakowanym sekwencjonowanym kwasem nukleinowym lub wyznakowanymi fragmentami tego kwasu nukleinowego bądź wyznakowanymi odpowiednikami oligorybonukleotydowymi fragmentów sekwencjonowanego kwasu nukleinowego, a następnie zgodnie z istniejącą wiedzą w zakresie hybrydyzacji kwasów nukleinowych mierzone za pomocą odpowiednich urządzeń pomiarowych zajście procesu hybrydyzacji dla każdego z pól scalonej biblioteki izotermicznej. Uzyskane dane pomiarowe wskazujące na występowanie sekwencji komplementarnej do każdego zidentyfikowanego oligonukleotydu wchodzącego w skład scalonej biblioteki izotermicznej wykorzystane następnie do odtworzenia sekwencji badanego kwasu nukleinowego przy zastosowaniu odpowiedniej procedury matematyczno-obliczeniowej.

W drugim przykładzie realizacji sposobu sekwencjonowania kwasów nukleinowych według wynalazku utworzono zestaw oligonukleotydów wchodzących w skład biblioteki izotermicznej stanowiący rozproszoną bibliotekę oligonukleotydów. W rozproszonej bibliotece oligonukleotydów są one albo pojedynczymi cząsteczkami albo są przyłączone do drobin materiału stałego, przy czym do drobin materiału stałego przyłączona jest przynajmniej jedna cząsteczka oligonukleotydu zdolna do wzięcia udziału w procesie hybrydyzacji z sekwencjonowanym kwasem nukleinowym lub jego fragmentami.

Rozproszone biblioteki oligonukleotydowe otrzymano na drodze kombinatorycznej syntezy chemicznej, przy czym możliwe jest także wykorzystanie w tym celu znanego powszechnie szerokiego zakresu grup ochronnych i znanych metod syntezy. Sposoby przyłączenia oligonukleotydów do drobin podłoża stałego mogą być również takie jak omówione w przykładzie dla scalonych bibliotek izotermicznych. To samo dotyczy jednostek nukleotydowych tworzących oligonukleotydy bibliotek izotermicznych. Sposób prowadzenia chemicznej syntezy kombinatorycznej rozproszonych izotermicznych bibliotek oligonukleotydowych jest analogiczny do sposobu projektowania etapów syntezy kombinatorycznej scalonych bibliotek izotermicznych, które ilustrują fig. 1 i fig. 2. W wyniku syntezy kombinatorycznej otrzymano rozproszoną izotermiczną bibliotekę oligonukleotydową.

Rozproszoną izotermiczną bibliotekę oligonukleotydową można też otrzymać poprzez nie kombinatoryczną, sekwencyjną syntezę wszystkich oligonukleotydów wchodzących w jej skład i następne ich zmieszanie w jednakowych proporcjach.

Sekwencjonowanie z wykorzystaniem rozproszonej izotermicznej biblioteki oligonukleotydowej prowadzono realizując hybrydyzację z odpowiednio wyznakowanym kwasem nukleinowym lub jego wyznakowanymi fragmentami lub wyznakowanymi odpowiednikami oligorybonukleotydowymi fragmentów sekwencjonowanego kwasu nukleinowego. Jako grupy znakujące stosowano różnorodne znane grupy markerowe pozwalające na przeprowadzenie detekcji za pomocą promieniowania elektromagnetycznego albo poprzez wysoce selektywną asocjację z innymi czynnikami molekularnymi. Wśród takich markerów można wymienić bądź grupy o właściwościach fluorescencyjnych, bądź białynę oddziałującą bardzo specyficznie i silnie z białkiem streptawidyną.

Po przeprowadzeniu hybrydyzacji zgodnie z wiedzą w tym zakresie prowadzono selekcję tych elementów rozproszonej izotermicznej biblioteki oligonukleotydowej, które zawierały sekwencje komplementarne do analizowanego kwasu nukleinowego lub jego fragmentów czy odpowiedników oligorybonukleotydowych. Etap ten można przeprowadzić stosując selekcję mechaniczną lub selekcję za pomocą urządzeń typu automatycznych sorterów komórkowych (typu FACS) lub selekcję masywną za pomocą podłoży magnetycznych (Markiewicz, M.; Astriab, A., Markiewicz W.T., Application of a Synthetic Oligonucleotide Combinatorial Library in Triplex DNA Studies, Collection. Czechoslov.Chem.Comm. 61, Spec. Issue, Str. 120-121, 1996; Siebenkotten G., Koester M., Miltenyi S., Micro column system for magnetic separation, Patent EP0941766).

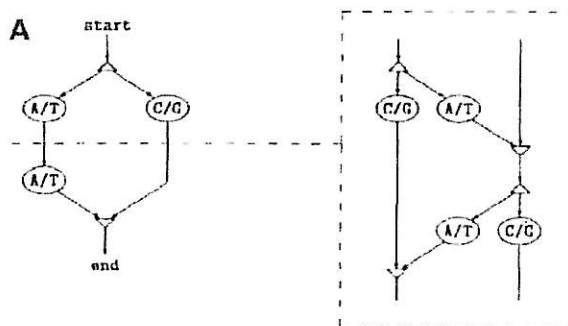
Wybrane drobinę niosące unikalne sekwencje oligonukleotydowe spośród oligonukleotydów rozproszonej izotermicznej biblioteki oligonukleotydowej, a jednocześnie komplementarne do badanego kwasu nukleinowego lub jego fragmentów, czy odpowiedników oligorybonukleotydowych należy poddać analizie za pomocą innych znanych metod sekwencyjnych. Wśród metod ustalenia sekwencji oligonukleotydów wyselekcjonowanych z rozproszonej biblioteki izotermicznej należy wymienić spektrometrię mas w wariacie MALDI-TOF i inne znane metody sekwencjonowania łącznie z metodami enzymatycznymi.

Uzyskane wyniki pomiaru hybrydyzacji sekwencjonowanego kwasu nukleinowego wykorzystano następnie do odtworzenia jego sekwencji stosując odpowiednie procedury matematyczno-obliczeniowe identyczne jak w opisanym wcześniej pierwszym przykładzie realizacji sposobu według wynalazku.

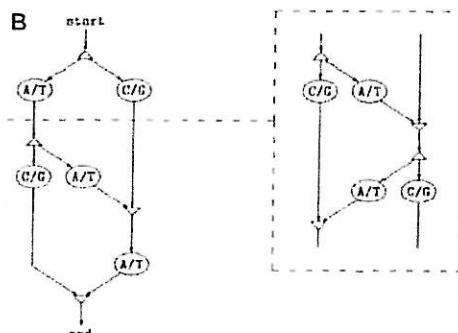
Zastrzeżenia patentowe

1. Sposób sekwencjonowania kwasów nukleinowych, poprzez prowadzoną w znany sposób hybrydyzację kwasu nukleinowego o nieznanej sekwencji z wykorzystaniem zestawu oligonukleotydów, **znamienny tym**, że tworzy i wykorzystuje się zestaw oligonukleotydów zawierający co najmniej 50% jakiegś biblioteki izorelacyjnej stanowiącej zbiór wszystkich oligonukleotydów spełniających relację taką, że $w_A x_A + w_T x_T + w_C x_C + w_G x_G = C_i$, gdzie w_A, w_C, w_G, w_T oznaczają wkłady wnoszone odpowiednio przez nukleotydy A, C, G, T, zmienne x_A, x_C, x_G, x_T oznaczają liczby odpowiednich nukleotydów w oligonukleotydzie, a C_i oznacza stałą dla danej biblioteki izorelacyjnej, z pominięciem sytuacji, gdy $w_A = w_C = w_G = w_T$.

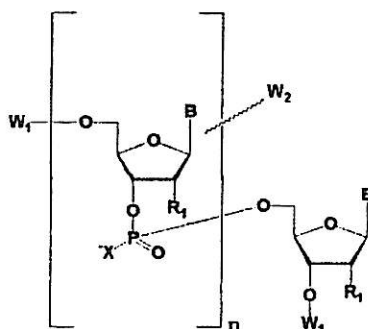
2. Sposób według zastrz. 1, **znamienny tym**, że korzystnie bibliotekę izorelacyjną oligonukleotydów tworzy się jako izotermiczną, czyli taką, dla której parametry w_A, w_C, w_G, w_T pozostają między sobą w następujących relacjach takich, że: $w_A = w_T, w_C = w_G$ oraz $2w_A = w_C$, przy czym tę bibliotekę generuje się zgodnie ze schematem (A):



jeżeli C_i jest podzielne bez reszty przez $2w_A$ lub w przypadku przeciwnym zgodnie ze schematem (B).



3. Sposób według zastrz. 2, **znamienny tym**, że oligonukleotyd izotermicznej biblioteki oligonukleotydowej konstruuje się z jednostek nukleotydowych zgodnie z wzorem:



gdzie:

n jest liczbą całkowitą od 1 do 100

B jest aglikonem spośród zasad heterocyklicznych występujących w naturalnych kwasach nukleinowych lub inną zasadą heterocykliczną, której oddziaływanie z komplementarnymi zasadami heterocyklicznymi jest znane i odpowiada oddziaływaniom międzycząsteczkowym, w których biorą udział zasady heterocykliczne występujące w DNA: adenina (Ade), cytozyna (Cyt), guanina (Gua) i tymina (Thy);

X⁻ jest O lub S

R₁ jest H, OH, lub grupą alkoksylową albo alkenylową albo F

W₁ jest H, PO(OH)₂ lub jej solą lub łącznikiem z podłożem stałym

W₂ jest nieobecne albo jest łącznikiem z podłożem stałym gdy W₁ jest H, PO(OH)₂ lub jej solą.

Rysunki

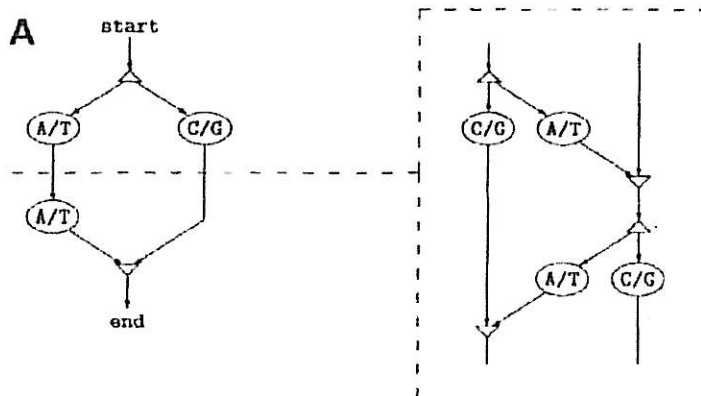


Fig.1

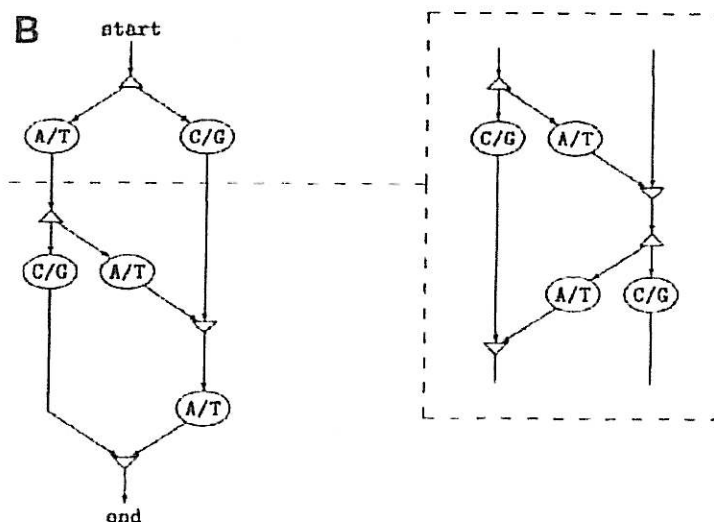


Fig. 2