

KOMPUTERY BIOMOLEKULARNE OPARTE NA BRAMKACH LOGICZNYCH ZA POMOCĄ DNA. WPŁYW KOMPUTERÓW BIOMOLEKULARNYCH NA ASPEKTY EKONOMICZNE

MGR MIŁOSZ GRUCHAŁA

Uniwersytet Łódzki
Wydział Matematyki i Informatyki
e-mail: mgruchala@math.uni.lodz.pl

SŁOWA KLUCZOWE

DNA, dane, komputery biomolekularne, bramki logiczne za pomocą DNA, przechowywanie danych

ABSTRAKT

Celem artykułu jest przedstawienie modelu komputera biomolekularnego dzięki wykorzystaniu bramek logicznych zbudowanych z łańcuchów DNA. Praca opisuje również, w jaki praktyczny sposób można wykorzystać model komputera biomolekularnego do przechowywania danych.

Wprowadzenie

Informatykę można definiować, jako naukę o przetwarzaniu, przechowywaniu i przekazywaniu informacji. Środowisko, w jakim przechowujemy dane może być oparte na elektronice, ale również i genetyce molekularnej. Biokomputery używają materiałów biologicznych jako narzędzi do pełnienia funkcji obliczeniowych. Biokomputer składa się z szeregu reakcji biochemicznych z materiałami biologicznymi, które są zaprojektowane w taki sposób, aby utworzyć skalowalny układ. Otrzymany ciąg reakcji biochemicznych, zachodzących w trakcie działania takiego układu, generuje wyjście lub też sygnał wyjściowy. Cały proces może być interpretowany jako postać analizy komputerowej.

Obecnie zauważalny jest duży nacisk w badaniach nad nowymi technologiami, idący w kierunku możliwości zastąpienia tradycyjnych komputerów oraz pokonania ograniczeń obliczeniowych, zarówno technicznych jak i fizycznych, znanych nam maszyn. Nowe technologie koncentrują się wokół komputerów kwantowych oraz biomolekularnych. Pomimo wczesnego etapu prac nad rozwiązaniami problemów z użyciem komputerów biomolekularnych, wydają się

one bardzo obiecujące. Właściwości kodowania oraz zarządzanie informacją za pomocą DNA posiada duży potencjał, między innymi dzięki małym rozmiarom przechowywania danych za ich pomocą.

Rozwój komputerów biomolekularnych zapoczątkowany został w sposób znaczący w 1994 roku przez Leonarda Adleman'a z University of Southern California. Adleman zaproponował zastosowanie obliczeń DNA do rozwiązania problemu ścieżki Hamiltona z użyciem siedmiu wierzchołków. Rozwiązanie amerykańskiego profesora było spektakularnym sukcesem, inspiracją do dalszych badań nad tematyką komputerów DNA, takich jak rozwiązanie problemu SAT, czy też zapisanie za pomocą DNA książki przez zespół naukowców Akademii Medycznej Harvard (Church, Gao, Kosuri, 2012).

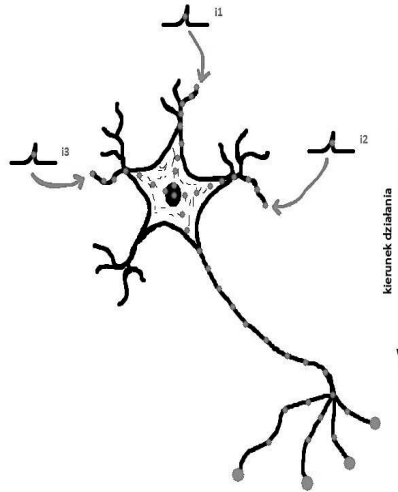
Innymi ciekawymi badaniami mogą być prace małżeństwa, zatrudnionego w Kalifornijskim Instytucie Technologii, profesora Erika Winfree oraz doktor Lulu Qian. Opublikowali oni na łamach czasopisma *Science* wyniki badań nad działaniem układu skalowalnego za pomocą bramek DNA (Qian, Winfree, 2011). Ich praca była również inspiracją dla mnie, a zrozumienie zasady działania oraz jego właściwości stały się początkiem mojej pracy naukowej.

Założeniem oraz wyzwaniem do moich dalszych badań jest chęć zbudowania biochemicznego mózgu złożonego z sieci neuronów, które zachowywałyby podobieństwo działania do ludzkiego mózgu. Każdy neuron posiada elektryczny potencjał, a jako komórka jest zdolny do przetwarzania i przewodzenia informacji w postaci sygnału elektrycznego. Neurony są podstawowym elementem układu nerwowego. Aby zrozumieć działanie bramki logicznej za pomocą DNA musimy wiedzieć, jak działają sygnały wejściowe działające na neuron.

Koncepcja budowy bramki logicznej za pomocą DNA

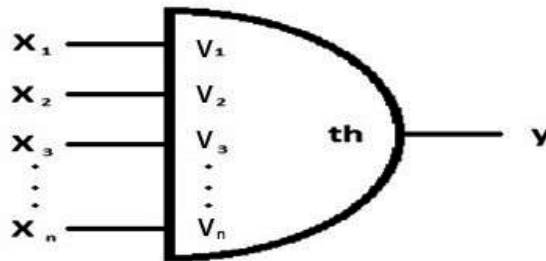
Ilustracja przedstawia działanie trzech sygnałów wejściowych (tzw. input): i_1 , i_2 oraz i_3 oddziałujących na neuron. Droga działania takiego sygnału zaczyna się w chwili oddziaływania na dendryty, następnie sygnał trafia do jądra neuronu wraz z jego potencjałem elektronicznym. W efekcie kumulacji generowany jest sygnał wyjściowy, jeśli istnieje (tzw. output). Potencjał elektroniczny sygnału wejściowego może być zarówno dodatni, jak i ujemny. Aby wystąpił sygnał wyjściowy potencjał elektroniczny musi przekroczyć wartość progu dla danego neuronu.

Oddziaływanie takich sygnałów wejściowych możemy potraktować również jako bramkę logiczną, która posiada sygnały wejściowe, ich wartość oraz wagę, a w środku takiej bramki istniałby próg. W zależności od wartości i wag wprowadzanych do bramki przez inputy, na wyjściu otrzymamy – lub nie – sygnał wyjściowy (do obliczeń używane jest oznaczenie 1 i 0, gdzie 1 oznacza wystąpienie sygnału wyjściowego, a 0 jego brak). Poniżej schemat działania bramki (z progiem).



Rysunek 1. Działanie neuronu

Źródło: opracowanie własne.

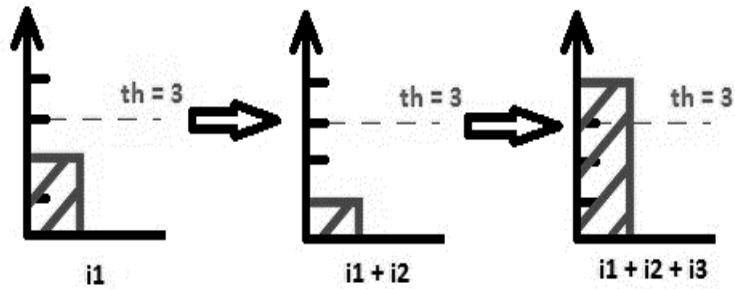


Rysunek 2. Schemat działania bramki

Źródło: opracowanie własne.

W zilustrowanym przykładzie mamy bramkę, na którą działają sygnały wejściowe x_1, x_2, \dots, x_n z wagami v_1, v_2, \dots, v_n . Widoczny jest próg th (threshold) oraz sygnał wyjściowy y .

Wystąpienie sygnału wyjściowego zależy jest od wartości wprowadzanych do bramki na wejściu. Aby przedstawić jak działa taka bramka, przyjmuję do obliczeń trzy sygnały wejściowe x_1, x_2, x_3 o wartości 1 oraz wagach $v_1=2, v_2=-1$ oraz $v_3=3$. Oddziaływanie sygnałów wejściowych możemy przedstawić krok po kroku na rysunku nr 3.



Rysunek 3. Wykres potencjału elektrycznego wraz z progiem o wartości $th=3$

Źródło: opracowanie własne.

Funkcję liniową progu możemy przedstawić następującym wzorem:

$$f(x) = v_1 \times x_1 + v_2 \times x_2 + v_3 \times x_3.$$

Natomiast wynikiem działania bramki jest wartość y określona następująco:

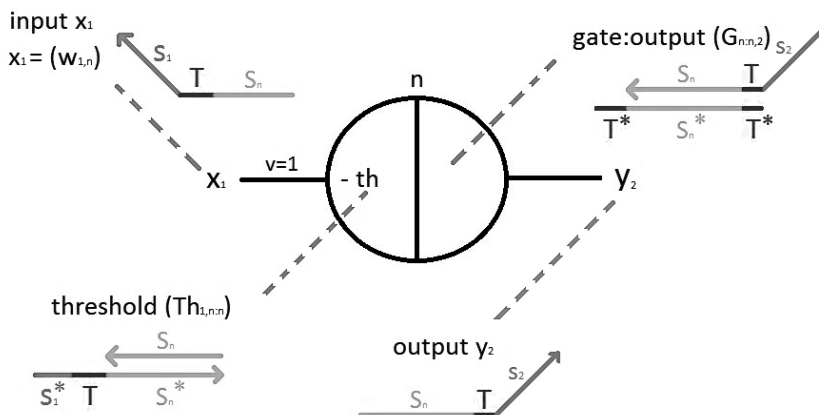
$$y = \begin{cases} 1 & \text{jeśli } f(x) > th \\ 0 & \text{jeśli } f(x) \leq th \end{cases}$$

Przy wprowadzonych wartościach otrzymujemy rezultat:

$$f(x) = 2 \times 1 + (-1) \times 1 + 3 \times 1 = 4,$$

$$y = 1 \text{ bo } f(x) > 3.$$

Schemat bramki molekularnej



Rysunek 4. Schemat bramki G_n

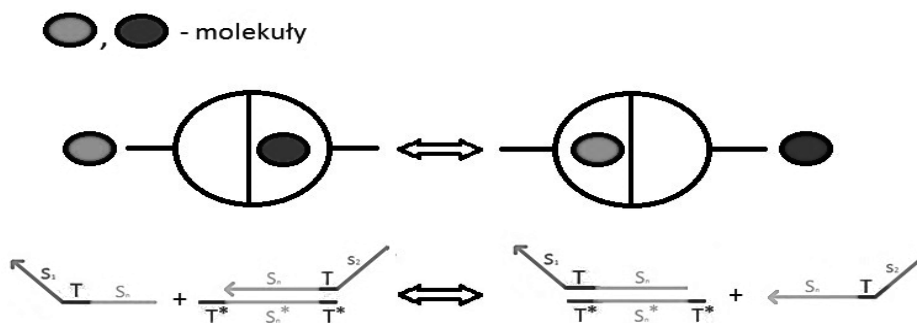
Źródło: opracowanie własne na podstawie Qian, Winfree, Scaling (2011).

Każda część odgrywa szczególną rolę w bramce i ma unikalną nazwę. Sygnał wejściowy to $x_1 = (w_{1,n})$, gdzie 1 pochodzi od molekuly S_1 , a n pochodzi od docelowego numeru bramki, do którego trafia sygnał wejściowy. Kolejne odcienie linii stanowią nici DNA oraz strzałki, które wskazują na kierunki nici od 3' do 5'. Przykładowe nici S_1, S_2, \dots, S_n są długie (15 nukleotydów) molekuly rozpoznawalne dla odpowiednich bramek 1, 2...n. T jest krótką (5 nukleotydów) molekula służącą do łączenia się poszczególnych nici. Każda nić z * jest jej komplementarnym uzupełnieniem. Przykładowo S_1^* jest komplementarną nicią do nici S_1 . Składowe bramki opisane powyżej nie muszą wszystkie pojawić się w każdej bramce.

Pierwsze bramki DNA skonstruowane zostały w 1997 roku przez badaczy Animesh'a Ray'a oraz Mitsu Ogihara'ego z uniwersytetu w Rochester (Ray, Ogihara, 1997). Wykrywały one na wejściu fragmenty materiału genetycznego, a następnie łącząc łańcuchy genetyczne tworzyły w rezultacie nowy łańcuch. Bramkę AND uzyskiwano przez połączenie dwóch łańcuchów (lepkimi końcami) przy pomocy ligazy. Wcześniej wspomniani Winfree oraz Qian bramkę logiczną AND oraz OR, w zależności od stężenia i wartości progu, opisują w trzech krokach. W pierwszym, sygnały wejściowe trafiają do bramki, otrzymując zsumowany sygnał. Krok drugi to wysłanie takiego sygnału, jako sygnału wejściowego do bramy z progiem. Przy przekroczeniu progu końcowym etapem jest raportowanie wyniku poprzez działanie tzw. reportera, który działa na zasadzie wystąpienia fluorescencji, jeśli sygnał wejściowy pojawił się na wejściu (Qian, Winfree, 2011).

Działanie bramek logicznych jest skomplikowane, a budowane z nich układy mogą przyjmować bardzo duże rozmiary. Oprócz bramek AND i OR naukowcy badają również bramki logiczne, takie jak NOR, NOT, XOR i inne. Ciekawym obiektem studiów wydaje się dalsza analiza badań nad działaniem bramek logicznych DNA.

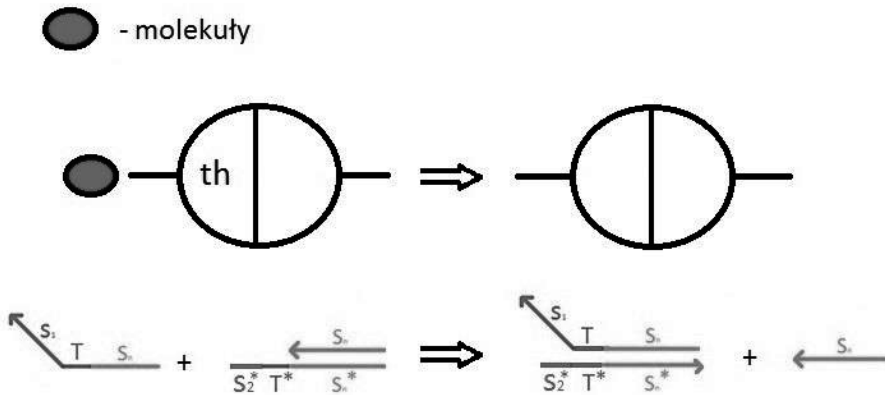
Trzy podstawowe mechanizmy reakcji na bramkach logicznych za pomocą DNA: *seesawing* (huśtanie/huśtawkowanie), *thresholding* (progowanie) i *reporting* (raportowanie).



Rysunek 5. Huśtanie (*seesawing*)

Źródło: opracowanie własne.

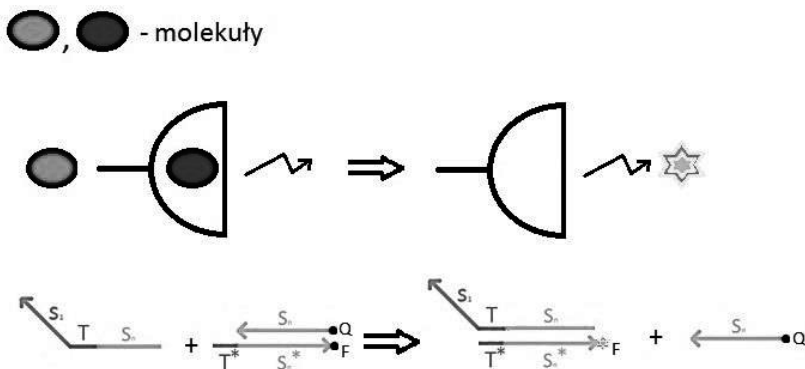
Rysunek przedstawia schemat procesu huśtania, z widoczną reakcją sygnału wejściowego, w tym przypadku jest to molekula S_1TS_n , która reaguje z bramą G_n , następnie zachodzi przesunięcie nici, potocznie zwane huśtaniem, a w rezultacie powstaje sygnał wyjściowy, konkretnie molekula S_nTS_2 .



Rysunek 6. Progowanie (*thresholding*)

Źródło: opracowanie własne.

Powyżej zilustrowane zostało progowanie. Widoczny jest sygnał wejściowy, który reaguje z bramą, posiadającą próg. Sygnał wejściowy posiada stężenie, opisane na schemacie bramy jako waga. Sygnał wyjściowy powstanie tylko w przypadku, gdy stężenie sygnału wejściowego będzie większe niż prog. Gdy sygnał raz zareaguje z bramą prog, to nigdy nie będzie mógł być użyty w reakcji, ponieważ wszystkie istotne punkty wyparcia są skutecznie zablokowane. Tylko wtedy, gdy wszystkie nici prog zostaną wykorzystane, pozostałe pasma sygnału wejściowego mogą reagować z oryginalną/pierwotną bramą.

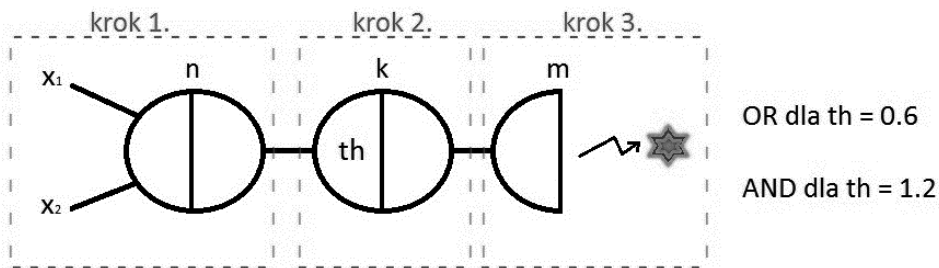


Rysunek 7. Raportowanie (*reporting*)

Źródło: opracowanie własne.

Aby uzyskać widoczne efekty prac laboratoryjnych wprowadzono reporter. Jest to pomocna rzecz, aby zobaczyć ludzkim okiem efekt działania bez użycia żadnych narzędzi laboratoryjnych. Powyższy schemat przedstawia reporter: F i Q oznaczają odpowiednio fluorofor oraz jego wygaszacz. Gdy komplementarne nici oddzieli się, wygaszacz przestaje absorbować fluorofor, co powoduje, że cząsteczka świeci.

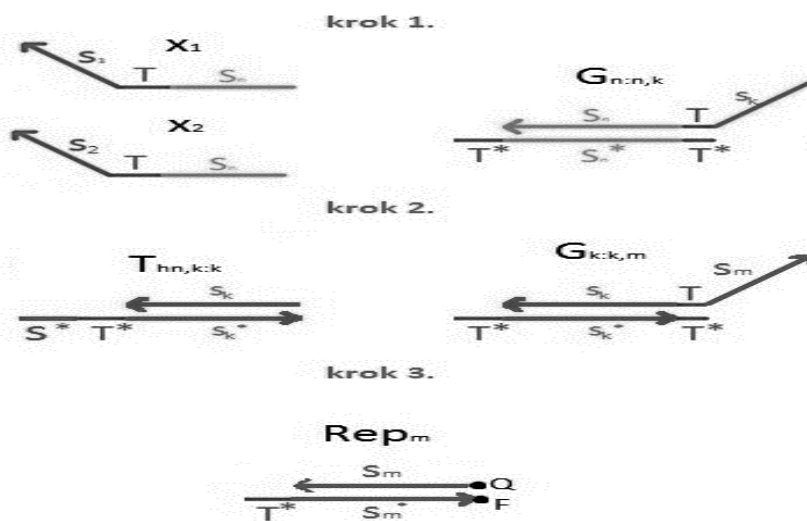
Cyfrowe bramki logiczne realizowane motywem DNA



Rysunek 8. Schemat bramki logicznej OR albo AND

Źródło: opracowanie własne.

Sygnaly wejściowe x_1 i x_2 są sumowane w fazie pierwszej poprzez huśtawkowanie w bramce $G_{n,n,k}$, następnie trafiają do bramy $G_{k:k,m}$ z progiem $Th_{n,k:k}$. Kolejnym etapem jest progowanie. Jeżeli suma sygnałów wejściowych przekroczy próg, generowany jest sygnał wyjściowy, który jest następnie raportowany w trzeciej fazie przez fluorofor w reporterze Rep_m .



Rysunek 9. Nici DNA użyte w każdym z kroków

Źródło: opracowanie własne na podstawie: Qian, Winfree, Scaling (2011).

Wpływ i możliwości powszechnego stosowania komputerów biomolekularnych na aspekty ekonomiczne

Technologie komputerowe szybko się rozwijają, a ich błyskawiczny rozwój możemy obserwować w ostatnich dwóch dekadach. W całym świecie biznesu zarówno komercyjnym, jak i niehandlowym, istnieje potrzeba wypracowania rozwiązania, które pomoże przechwycić dane pochodzące z różnych źródeł, analizować wiadomości, odkrywać nowe informacje o ważnym znaczeniu, powiązać je, a na ich podstawie umożliwić podjęcie racjonalnych decyzji. Technologia powinna umożliwiać inteligentny dostęp do danych i możliwość interakcji, a nie prowadzić do nadmiaru i chaosu informacyjnego.

Aspekt przechowywania danych w przypadku komputerów biomolekularnych jest ich największym atutem, gdyż molekuly DNA są niezwykle małe, a potrafią przechowywać ogromne ilości danych. Dla porównania, komputer biomolekularny o wielkości zwykłej stacji końcowej, mógłby wykonywać więcej operacji niż wszystkie komputery elektroniczne razem wzięte. Obrazowo skalę tego zjawiska najlepiej pokazuje sytuacja z 2012 roku, naukowiec George Church przechowywał 70 miliardów egzemplarzy swojej książki „Regenesis: How Synthetic Biology Will Reinvent Nature and Ourselves in DNA” na niciach DNA, a ich rozmiar równał się ludzkiemu paznokciu.

Problemem czasów współczesnych staje się przechowywanie danych, które rosną wykładniczo, a pojemność istniejących nośników danych okazuje się niewystarczająca. Archiwizacja danych za pomocą DNA jest więc atrakcyjną alternatywą. Zaletą przechowywania danych dzięki DNA jest to, iż ich struktura jest bardzo gęsta i może przechowywać aż 1 eksabajt / mm³, co jest równoznaczną wartością do 10⁹ GB. Dodatkową wartością jest długotrwałość, która według badań może wynosić ponad 500 lat.

Grupa badawcza z Uniwersytetu w Waszyngtonie opracowała architekturę opartą na DNA służącą do przechowywania danych. Początkiem badań było oszacowanie, jak wielką ilość danych będziemy przechowywać na świecie w 2017 roku (około 16. zettabajtów). Nawet przy uwzględnieniu prognozowanych usprawnień w magazynowaniu danych ich ilość przekroczy zdolności przechowywania danych w obecnej technologii.

Większość danych na świecie w tym momencie, przechowywanych jest na magnetycznych oraz optycznych nośnikach danych. Dodatkowo, ulepszona technologia składowania danych na taśmach umożliwia przechowywanie do 185 TB danych na jednym kartridżu i jest najgęstszą formą przechowywania danych, co daje wynik 10 GB/ mm³. Najnowsze badania wykazują możliwości przechowywania na dyskach optycznych danych o pojemności 1 petabajta, z gęstością 100 GB/ mm³. Pomimo udoskonaleń, które niedługo wejdą w życie, przechowywanie zettabajtów danych zapełni miliony nośników oraz wiele fizycznego miejsca. Kolejnym problemem jest trwałość nośników, tj. 3–5 lat dla dysków i 10–30 lat dla taśm. Obecnie długoterminowe przechowywanie archiwaliów to rozwiązania wymagające „odświeżenia”, by usunąć uszkodzone dane lub je zastąpić. Zważając na powyższe aspekty, należy szukać znaczących postępów zarówno

jeśli chodzi o gęstość, jak i trwałość przechowywania danych. Syntetyczne sekwencje DNA od dawna uważane są za potencjalny nośnik danych cyfrowych.

Proces zapisywania danych polega na przechowywaniu danych cyfrowych w postaci sekwencji nukleotydowych, które łączą się w cząsteczki DNA. Czytanie danych obejmuje sekwencjonowanie cząsteczek DNA i dekodowanie informacji z powrotem na oryginalne cyfrowe dane. Zarówno synteza, jak i sekwencjonowanie cząsteczek DNA są standardową praktyką w biotechnologii.

Postępy w przechowywaniu danych za pomocą DNA są bardzo szybkie. W 1999 roku naukowcy zapisali i odczytali 23-znakową wiadomość, natomiast już w 2013 naukowcy byli w stanie zakodować wiadomość o wielkości 739 kilobajtów. W tym momencie ilość danych, które mogą być przechowywane za pomocą DNA jest ograniczona jedynie przez koszty.

Do sekwencjonowania DNA istnieje kilka wysokowydajnych metod, które wykorzystują enzymy polimerazy DNA i są powszechnie określane jako „sekwencjonowanie przez syntezę”. Co ważne, podczas procesu syntezy używa się nukleotydów fluorescencyjnych, które emitują różne kolory, a odczytanie ich sekwencji możliwe jest optycznie.

Podsumowanie

Przechowywanie danych za pomocą DNA może okazać się innowacyjnym rozwiązaniem, które może zrewolucjonizować światowy biznes. Operacje na DNA posiadają potencjał do ostatecznego rozwiązania problemu przechowywania olbrzymich danych na świecie, a ich główne zalety to niezwykła gęstość i trwałość. Biorąc pod uwagę ograniczenia technologii krzemowych, warto inwestować w hybrydowe systemy biochemiczne. Nadszedł czas, by architekci komputerowi rozważyli praktyczne możliwości komputerów biomolekularnych. Przemysł biotechnologiczny może mieć znaczący wpływ na kształtowanie się przyszłościowych rozwiązań komputerowych, a sam aspekt przechowywania danych za pomocą DNA daje nam wachlarz rozwiązań do efektywnego zarządzania funduszami organizacyjnymi.

Skonstruowanie komputera biomolekularnego, opartego na działaniu bramek logicznych za pomocą DNA, to nieodległa przyszłość. Faza testów, a co za tym idzie sprawdzenie potencjału wprowadzenia biotechnologii do naszego życia spodziewane jest w ciągu najbliższego dziesięciolecia. Głównym torem kolejnych etapów badań będą prace nad szybkością i wydajnością zbudowanego komputera oraz możliwościami wprowadzenia zastosowań do przechowywania danych za pomocą DNA.

Literatura

- Bornholt, J., Lopez, R., Carmean, D.M., Ceze, L., Seelig, G., Strauss, K. (2016). *A DNA-Based Archival Storage System*. University of Washington, Microsoft Research. Pobrane z: http://www.dnakomputer.pl/index.php?option=com_content&view=article&id=3:dnaobliczenia&catid=3:ogolne (10.08.2017).
- Church, G.M., Gao, Y., Kosuri, S. (2012). Next-Generation Digital Information Storage in DNA. Pobrane z: <https://www.ncbi.nlm.nih.gov/pubmed/22903519> (10.02.1017).
- IDC Research (2013). *Where in the world is storage*. Pobrane z: http://www.idc.com/downloads/where_is_storage_infographic_243338.pdf (10.10.2017).
- Kraśniński, T., Sakowski, A.S. (2008). A review of models and practical implementations of DNA computation. *Studia Informatica*, 29 (80), 5–26.
- Miller, R. Miller (2013). *Facebook builds exabyte data centers for cold storage*. Pobrane z: <http://www.datacenterknowledge.com/archives/2013/01/18/facebook-builds-newdata-centers-for-cold-storage/> (10.10.2017).
- Palmer, J. (2011). *DNA computer "calculates square roots"*. Pobrane z: <http://www.bbc.com/news/science-environment-13626583> (10.10.2017).
- Plafke, J. (2013). *ExtremeTech. New optical laser can increase DVD storage up to one petabyte*. ExtremeTech. Pobrane z: <http://www.extremetech.com/computing/159245-new-optical-laser-can-increase-dvd-storage-up-to-one-petabyte> (10.10.2017).
- Qian, L., Erik Winfree., E. Scaling (2011). Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332 (6034), 1196–1201.
- Qian, L., Winfree, E. (2009). A simple DNA gate motif for synthesizing large-scale circuits. *Journal of the Royal Society Interface*, 8 (62), 1281–1297.
- Qian, L., Winfree, E., Bruck, J. (2011). Neural network computation with DNA strand displacement cascades. *Nature*, 475, 368–372.
- A. Ray, M. Ogihara (1997). Executing parallel logical operations with DNA
- Sony (2014). *Sony develops magnetic tape technology with the world's highest recording density*. Pobrane z: <http://www.sony.net/SonyInfo/News/Press/201404/14-044E> (10.10.2017).

BIOMOLECULAR COMPUTERS BASED ON DNA LOGICAL GATES. INFLUENCE OF BIOMOLECULAR COMPUTERS ON ECONOMIC ASPECTS

KEYWORDS

DNA, data, biomolecular computers, DNA-based logical gates, storage

ABSTRACT

The purpose of the article is to present a biomolecular computer model, based on logical gates built with DNA chains. Article also describes how a biomolecular computing can be used for data storage in practical way.

Translated by Mitosz Gruchala