

Komputer z probówki

W 1994 roku amerykański uczyony Leonard Adleman z Uniwersytetu Południowej Kalifornii zaprezentował praktyczny schemat działania nowego typu komputera, w którym miejscem zapisu i przechowywania danych były cząsteczki DNA w probówce. Przetwarzanie danych było oparte o metody inżynierii genetycznej t.j. sposoby łączenia i dzielenia fragmentów DNA poza żywym organizmem. W tym momencie ziściła się idea Richarda Feynmana o submikroskopowym komputerze. Komputer z probówki (biologiczny) obliczył stosunkowo prosty problem i zajęło mu to 7 dni. Zanim jednak pokażemy co policzył pierwszy komputer biologiczny i jakie są jego potencjalne możliwości przyjrzyjmy się bliżej cząsteczce DNA.

DNA jest liniową, długą (do kilku centymetrów w komórkach człowieka) cząsteczką występującą we wszystkich żywych organizmach i niektórych wirusach. Fragment cząsteczki DNA przedstawia rys.1. Kulki odpowiadają atomom azotu, tlenu, wodoru, węgla i fosforu. Kolory mówią nam o tym czy dany atom "lubi" być w otoczeniu wody czy nie. Intensywne kolory (czerwony i niebieski) odpowiadają tym atomom które silnie przyciągają cząsteczki wody. Kolor biały odpowiada tym atomom, które wody "nie lubią". Te informacje są o tyle ważne, że wszelkie interesujące nas procesy związane z DNA zachodzą w wodzie.

W cząsteczce DNA jest zakodowana większość informacji o strukturze chemicznej białek, niemniej fakt ten nie wynika bezpośrednio z informacji o składzie chemicznym DNA, ale raczej z jego przestrzennej budowy. Zostało to odkryte w roku 1953 przez Jamesa Watsona i Francisa Cricka (*Podwójna Helisa*, Prószyński i Ska 1996 rok). Z analizy zdjęć rentgenowskich DNA, wykonanych przez Rosalind Franklin i Maurice Wilkinsa, badacze odtworzyli budowę przestrzenną tej cząsteczki. Schematyczny rysunek DNA, który pokazuje istotę kodu genetycznego jest przedstawiony na rys. 2. Nie widać już na nim poszczególnych atomów (rys.1),

ale grupy atomów (zasady) oznaczone literami A,T,G,C oraz dwie wstęgi (helisy) utworzone przez cukry. Litery oznaczają początek nazwy każdej zasady t.j. A (adeniny),T (tyminy) ,C (cytozyny) i G (guaniny). Adenina, tymina, cytozyna i guanina mają specyficzne własności, które powodują, że A zawsze łączy się z T a C łączy się z G. Oznacza to, że obie wstęgi wzajemnie się uzupełniają. Gdy komórka dzieli się na dwie potomne komórki materiał genetyczny jest kopiowany tak jak to pokazano na obrazku (rys. 3). Z jednej strony cząsteczka rozpada się na dwie wstęgi a następnie do każdej z nich jest dobudowywana nowa wstęga według schematu: A do T, C do G, T do A i G do C.

Aby jeszcze bardziej uprościć schemat cząsteczki rozwińmy wstęgi (rys.4). Co nam mówi rysunek 4? Zapomnijmy na chwilę o chemii i o tym co kryje się pod literami A,T,C,G. Wyobraźmy sobie, że mamy język złożony z tych czterech liter. W języku piszemy słowa a słowa składamy w zdania. W DNA w komórce słowa są trzyliterowe (każde trzy litery kodują jeden aminokwas a z 22 aminokwasów są zbudowane wszystkie białka w naszym organizmie). Tak więc mamy 4 litery, maksymalnie 64 słowa trzyliterowe i nieskończoną ilość zdań, które możemy z nich utworzyć. Tak naprawdę to możemy po prostu wykorzystać ten kod powstały w procesie ewolucji do zapisu takich informacji jakie chcemy i tak jak chcemy. Za pomocą tych czterech liter możemy zapisać wszystkie dzieła Szekspira, symfonie Bethovena (w końcu nuty to też pewien kod), wiersze Norwida, równania matematyczne fizyki i filozofię Arystotelesa. Jak to jest możliwe? No cóż, we współczesnym komputerze używamy tylko dwóch cyferek aby to wszystko zapisać (0,1), czyli używamy kodu binarnego. natomiast wykorzystując DNA będziemy używać czteroliterowego kodu genetycznego (A,T,C,G).

Porównajmy możliwości i własności komputera tradycyjnego opartego na krzemie i układach scalonych z komputerem biologicznym opartym na cząsteczkach DNA

i inżynierii genetycznej. W komputerze krzemowym mamy binarny sposób zapisu z długością słowa 32 bity (już niedługo 64 na nowych Pentiumach), w komputerze biologicznym mamy kod czteroliterowy, chemiczny a długość słowa jest dowolna. W biologii jedno słowo to trzylitery, ale w naszym komputerze biologicznym nie jesteśmy tym ograniczeni. W tradycyjnym komputerze informacje zapisujemy magnetycznie na twardym dysku lub optycznie na płytach kompaktowych. W komputerze biologicznym zapisujemy informację chemicznie w cząsteczce DNA. Popatrzmy na gęstości zapisu informacji. Jeden bit informacji jest dany przez jedną literkę n.p. A. Ta literka zajmuje wzdłuż łańcucha DNA około jednej dziesięciomilionowej centymetra. Czyli w centymetrowej cząsteczce DNA można zapisać około 1 Mb informacji. Cząsteczka DNA jest 20 tys. razy cieńsza niż ludzki włos (mój włos ma grubość 50 mikrometrów), skąd wynika, że w łebku szpilki (o objętości 1 milimetra sześciennego) można zmieścić 10 miliardów takich cząsteczek, czyli w naszym zapisie biologicznym możemy pomieścić w łebku szpilki 100000 razy więcej informacji niż zmieściłoby się na dysku o pojemności 100 Gigabajtów (rys. 5 fotografia dysku 100 Gb i łebka szpilki). Jak widać pojemność pamięci biologicznej jest miliardy razy większa niż pamięci konwencjonalnej. Molekularny sposób zapisu informacji w cząsteczce DNA jaki został osiągnięty przez naturę przez miliardy lat ewolucji jest w chwili obecnej granicą górną pojemności pamięci i znacznie przewyższa pojemność pamięci ludzkiego mózgu.

Kopiowanie informacji w komputerze biologicznym odbywa się dokładnie tak samo jak pokazano na rys.3, czyli tak jak kopiuje się DNA przy podziale komórki. Przy czym w jednej probówce możemy mieć wiele milionów kopii tej samej cząsteczki czyli wiele kopii tej samej informacji. Istnieją proste urządzenia zwane PCR, od słów angielskich polymerase chain reaction, czyli łańcuchowa reakcja polimerazy (*Podstawy Inżynierii genetycznej*, Prószyński i Ska 1999) (fot. 6), które w krótkim

czasie wykonuje miliardy kopii tej samej cząsteczki. Przy tym kopiowanie informacji w PCR jest zautomatyzowane, tanie i łatwe. PCR zrewolucjonizował biologię i dlatego za jego wynalezienie przyznano nagrodę Nobla.

Organizacja pamięci na dysku w komputerze zwykłym jest inna niż w biologicznym. W zwykłym komputerze dysk jest podzielony na sektory i różne części dysku mają swoje adresy. Dana informacja zapisana pod danym adresem jest odczytana wtedy gdy się poda adres pod jakim jest zapisana. W zapisie biologicznym pamięć jest zorganizowana inaczej t.j. nie ma adresów jest tylko zawartość. Inaczej mówiąc pamięć jest adresowana jej zawartością. Jak to zrozumieć? Jako prostą analogię weźmy system pocztowy na świecie. Wysyłając list do Jana Kowalskiego musimy podać jego adres czyli kraj, miasto, ulicę, numer domu i numer mieszkania. Jeśli przypadkiem Pan Kowalski się wyprowadzi to i tak poczta dostarczy list pod podany adres, czyli list przeznaczony dla Pana Kowalskiego trafi do innej osoby mieszkającej pod tym adresem. Tak jest zorganizowana pamięć w komputerze tradycyjnym. W pamięci biologicznej adres jest związany z zawartością. Analogia pocztowa jest następująca. Wysyłamy list do Jana Kowalskiego i nie podajemy żadnego adresu. Na kopercie zamieszczamy tylko zdjęcie Jana Kowalskiego. List dochodzi do Jana Kowalskiego niezależnie od tego czy mieszka w Ameryce, w Polsce czy na Wyspie Króla Jerzego w pobliżu Antarktydy. Tak właśnie działa pamięć biologiczna i to całkiem niezłe. Pamięć w probówce wyglądałaby tak: w roztworze wodnym umieścilibyśmy wiele cząsteczek DNA, każda informacja miałaby wiele milionów kopii (prosta sprawa). Po podgrzaniu próbówki (tak jak to się robi w PCR) każde DNA rozpadłoby się na dwie wstęgi (rys. 3). Załóżmy, że na jednej z nich mamy zapisaną symfonię Bethovena. Żeby ją całą wydobyć z pamięci przygotowujemy krótki fragment DNA z zapisanym niepowtarzalnym kawałkiem tej symfonii. Niech to będzie zapis ATCGGCTAATCGGTCAT. Taki fragment pojedynczej he-

lisy wpuszczamy do próbówki za pomocą magnesu, t.j. do końca cząsteczki przy-
czepiony jest mikroskopowy magnesik złożony z cząsteczki chemicznej. Tak to robił
Leonard Adleman w swoim urządzeniu. Teraz nasza cząsteczka próbna podłączy się
do helisy która ma sekwencję komplementarną do tej podanej wyżej, czyli TAGC-
CGATTAGCCAGTA (pamiętamy, że zawsze A łączy się z T a C łączy się z G).
Ta sekwencja literek oczywiście stanowi malutką część całej symfonii Bethovena,
niemniej jest ona wystarczająca, żeby z banku pamięci biologicznej wyciągnąć tę
właściwą cząsteczkę kodującą symfonię Bethovena. **(tutaj można dać rysunek
nici z literkami łączącej się z próbnikiem i wyciąganej za pomocą mag-
nesiku z próbówki)**

Jak wyglądałoby porównanie szybkości komputera Pentium III 800 MHz i kom-
putera biologicznego. W ciągu 1 sekundy nasz Pentium wykona maksymalnie 800
milionów operacji. W tym czasie na jednej cząsteczce DNA zostanie wykonana w
naszym komputerze biologicznym jedna operacja. Na pierwszy rzut oka wygląda na
to, że komputer biologiczny jest strasznie wolny. Ale nie możemy zapomnieć, że w
jednej próbówce możemy zmieścić astronomiczną liczbę cząsteczek DNA (n.p. 10^{20}
lub więcej). W czasie tej jednej sekundy na każdej z tych cząsteczek jest wykonana
jedna operacja. Czyli w sumie w 1 sekundzie nasz komputer wykona 10^{20} oper-
acji podczas gdy najnowszy Pentium III tylko 10^9 operacji. Oznacza to, że nasz
komputer bilogiczny jest 10^{11} (100 miliardów) razy szybszy niż Pentium III.

Powstaje pytanie, jakie problemy może rozwiązywać komputer biologiczny, z
którymi nie radzi sobie komputer tradycyjny. Jak wskazują na to uczeni zajmujący
się informatyką, są to wszystkie problemy, w których należy szybko sprawdzić as-
tronomiczną ilość możliwości. Takim problemem jest n.p. łamanie szyfrów. Rząd
amerykański korzysta z systemu szyfrowania DES (digital encryption standard czyli
cyfrowy system kodowania). W tym systemie informacja jest szyfrowana za pomocą

jednego z 2^{56} kluczy kodujących (ogólnie dostępnych). Aby odczytać zaszyfrowaną wiadomość bez znajomości klucza należy sprawdzić wszystkie 2^{56} kluczy. Dla jednego komputera jest to zadanie, które zajęłoby ogromnie dużo czasu. Kilku kryptografów wpadło na pomysł, żeby złamać DES wykorzystując Internet. Zaprośili do zabawy setki tysięcy internautów i po kilku miesiącach pracy setek tysięcy komputerów kod został złamany. W naszym komputerze biologicznym po zapisaniu 2^{56} kluczy w cząsteczkach DNA łamanie kodu odbywałoby się w sekundę, ponieważ nasz komputer równocześnie sprawdzałby wszystkie klucze.

Na koniec muszę niestety wylać kubel zimnej wody na wszystkie rozpalone głowy miłośników szybkich komputerów. Pierwszy komputer biologiczny policzył przez 7 dni problem, który dziecko rozwiązałoby w kilkanaście sekund a komputer krzemowy jeszcze szybciej. W luźnym tłumaczeniu problem ten wyglądał następująco (problem komiwojażera): masz siedem miast ponumerowanych od 1 do 7 i położonych na jednej linii. Narysuj najkrótszą drogę, która zaczyna się w mieście 1, kończy w mieście 7 i przez każde miasto przechodzi tylko raz. (rysunek 7). Przykład ten pokazuje, że upłynie jeszcze wiele czasu zanim komputer biologiczny stanie na półkach ze sprzętem komputerowym, niemniej chyba już dziś warto wiedzieć co czeka nas w przyszłości.

Robert Hołyst

Instytut Chemii Fizycznej PAN