

Jan Barciszewski  
Instytut Chemii Bioorganicznej PAN  
Poznań

## Poznanie struktury genomu ludzkiego wielkim zadaniem biologii molekularnej

Poznanie mechanizmu powstawania raka w organizmie jest jednym z najważniejszych praktycznych i poznawczych celów badań w biologii molekularnej. Dotychczasowe doświadczenia w tej dziedzinie prowadzono wykorzystując kultury tkankowe oraz onkogenne wirusy *in vitro*. Przeprowadzone eksperymenty z kulturami tkankowymi pozwoliły na uproszczenie obiektu badań, a wykorzystanie wirusów na uniknięcie złożoności materiału genetycznego (genomu) organizmów wyższych. Coraz bardziej staje się oczywiste, że dla ostatecznego poznania mechanizmu powstawania raka konieczna jest znajomość struktury kwasu dezoksyrybonukleinowego, DNA. W każdym organizmie, od wirusów i bakterii do człowieka, dziedziczone instrukcje dla wzrostu i rozwoju zakodowane są w genach, które mogą być włączane i wyłączane (regulowane) w miarę rozwoju organizmu oraz zmieniających się warunków zewnętrznych. Zapis genetyczny w DNA przypomina długą taśmę, na której zapisane są kolejne słowa, zdania czy też rozdziały pełnej instrukcji tego, jak organizm będzie się tworzył i organizował.

Nowotwór wyodrębniony z organizmu pacjenta stanowi duży agregat komórek rakowych powstałych z pojedynczej komórki, która w przeszłości spełniała normalne funkcje. W pewnym okresie uległa ona zasadniczej przemianie, w wyniku której podział oraz proliferacja następowały na "własny" sygnał (a nie zewnętrzny, jak zwykle). Jakie są przyczyny "nienormalnego", nowotworowego rozwoju komórki?

W wyniku olbrzymiego postępu w biologii molekularnej i inżynierii genetycznej sytuacja zaczyna się wyjaśniać. Znalaziono bowiem geny powodujące rakowość w chromosomach komórek nowotworowych. Geny te uległy aktywacji i stanowią siłę powodującą transformację nowotworową. Stan ten może być scharakteryzowany na wiele możliwych sposobów. Zwykle najwyraźniejszy jest niekontrolowany wzrost. Również kształt komórek nowotworowych jest inny. Wiele z nich pobiera cukier w olbrzymich ilościach oraz rozwija się w oparciu o anaerobowy metabolizm (przemiany energetyczne niezależne od obecności tlenu). Zewnętrzna membrana komórek rakowych posiada antygeny nowotworowe, które mogą być łatwo rozróżniane metodami immunologicznymi. Obecnie coraz więcej jest danych wskazujących, że stan rakowy zakodowany jest w genach (DNA). Jest on wynikiem uszkodzenia systemu kontroli ekspresji informacji genetycznej (biosyntezy białka), zawartej w genach. Już ponad dwadzieścia lat temu Renato Dulbecco (Nagroda Nobla w 1975 r.) wykazał, że w warunkach laboratoryjnych możliwa jest transformacja komórek normalnych w nowotworowe. Przeniósł on wirusa *polyomy* do fibroplastów zarodka jeżowca

rosnących *in vitro* (na szalkach Petriego) i zauważył, że zakażone kultury tworzą stopy kolonii zmienionych komórek. U młodych szczurów zakażonych tymi właśnie komórkami stwierdzono stan nowotworowy. Eksperyment ten pokazał, że normalne komórki mogą ulec transformacji w wyniku prostych manipulacji. Stan taki można uzyskać stosując również czynniki chemiczne (kancerogeny), które modyfikują DNA powodując mutacje. Zanim jednak w 1978 r. Robert A. Weinberg pokazał, że informacja dla rakowatości zawarta jest w DNA, już w 1944 r. jeden ze współtwórców biologii molekularnej T. O. Avery stwierdził, że z dwóch badanych wówczas szczepów bakterii *Pneumococci*, tylko jeden powoduje ciężkie zapalenie płuc u myszy. Wyizolował on tzw. czynnik transformacyjny, który powodował przemianę niewirulentnych bakterii w stan patogeniczny. Czynnikiem tym okazał się DNA. Przeniesienie cząsteczki DNA za pomocą mikrokryształów fosforanu wapnia powoduje również transfer wirulentności z jednego szczepu bakterii do drugiego. Dalszym krokiem na tej drodze było wyodrębnienie niewielkiego fragmentu DNA powodującego transformację. Trzy grupy badawcze G.M. Coopera, R.A. Weinberga oraz M. Wiglera wyodrębniły za pomocą inżynierii genetycznej transformujący segment DNA. Jest to onkogen lub gen rakowy. Wiedząc, że aktywacja (utworzenie) onkogenu prowadzi do powstawania raka, należy zapytać o mechanizm jego funkcjonowania. Nauka nie posiada jednoznacznej odpowiedzi na ten temat.

Opisując najogólniej rozwój i znaczenie badań genów oraz ich struktury warto przypomnieć, że w DNA oprócz zapisu o strukturze poszczególnych białek zawarta jest informacja o sposobie odczytywania zapisu genetycznego, który przypomina skomplikowany program komputerowy. Jednakże wnioskowanie, że jest jedynym nośnikiem informacji byłaby niedocenianiem jego funkcji. W żywej komórce informacja jest także zawarta w innych cząsteczkach i strukturach z nich zbudowanych.

Dla większości badaczy jest jasne, że biologia człowieka weszła obecnie w taki okres, że żaden jej aspekt nie może być rozwiązany lub zrozumiany bez wykorzystania genetyki molekularnej. W ciągu ostatnich 10 lat biologia molekularna wzbogaciła się o nowe techniki badawcze takie jak: manipulacja cząsteczkami DNA, analiza sekwencji (kolejność ułożenia nukleotydów) oraz szybka synteza chemiczna DNA. Centralnym narzędziem wydaje się klonowanie (powielanie) genów. Cząsteczka DNA może być specyficznie przecięta za pomocą wybranego enzymu restrykcyjnego, a określony jej fragment usunięty, zmodyfikowany, lub zastąpiony całkowicie innym, a następnie cząsteczka DNA może być zrekonstruowana, a także dalej powielona. Posiadając możliwość manipulacji materiałem genetycznym, biologowie mają zasadniczo dwie drogi rozpoznania mechanizmu powstawania raka, a mianowicie:

poszukiwania kolejnych genów odpowiedzialnych za proces nowotworowy albo sekwencjonowania całego genomu wybranego gatunku biologicznego. Poszukiwanie kolejnych genów wydaje się zadaniem pozornie łatwym, ale w dalszym ciągu niezwykle trudnym. Obecnie jest ono z mniejszym lub większym powodzeniem realizowane w kilku laboratoriach. Jednym z ważniejszych tutaj stawianych zagadnień jest to, na ile geny rakowe u różnych gatunków są zróżnicowane oraz jakie białka regulatorowe są przez nie kodowane.

Druga możliwość dotyczy określenia sekwencji nukleotydowej całego genomu komórkowego. Poznanie sekwencji umożliwiłoby przygotowanie odpowiednich znaczników dla poszczególnych genów oraz ich klasyfikację w zależności od ekspresji

w różnych typach komórek.

Posiadanie fizycznej mapy genetycznej umożliwiłoby identyfikację genów zaangażowanych w rozwój raka oraz innych schorzeń będących aktualnie w polu widzenia. Jeżeli nadrzędnym celem tego typu badań jest zrozumienie procesów nowotworowych u człowieka, oczywiste jest, że należałoby sekwencjonować genom ludzki. Ponieważ kontrola genetyczna tkanki rakowej wydaje się odmienna u różnych gatunków, poznanie genów zaangażowanych w powstawanie nowotworu stworzyłoby szansę na nowe podejście terapeutyczne, które mogłoby prowadzić do zatrzymania rozwoju stanu rakowego. Nie ma żadnych wątpliwości, że poznanie genomu oraz dostępność znaczników (sond) dla dowolnych genów byłoby istotnym czynnikiem w poznawaniu fizjologii i patologii człowieka, mechanizmu regulacji ekspresji genu i organizacji systemu nerwowego. Są również przesłanki, że wszystkie geny powstały z ograniczonej liczby genów rodowych (dziedzicznych), a zatem zyskalibyśmy nowe informacje na temat ewolucji i pochodzenia człowieka.

Jeżeli zagadnienie to ma tak istotne znaczenie jak napisano, to należy postawić pytanie, dlaczego nie jest ono dotąd realizowane? Co jest obecnie dostępne a czego brak? Wywołuje to od niedawna dyskusję wśród biologów molekularnych. Najwięcej uwagi tym kwestiom poświęca się w USA i Japonii. Podejmuje się tam pewne konkretne działania przygotowawcze. Zadanie, jakie tutaj pozostaje do rozwiązania polega na rozszyfrowaniu informacji genetycznej zakodowanej w 3 miliardach par zasad ( $3 \times 10^9$ ) znajdujących się w genomie ludzkim. W porównaniu z bakteriami i innymi komórkami prokariotycznymi jest to ponad 1000 razy więcej. Dzięki odpowiedniemu umieszczeniu w chromatynie łańcuch DNA o długości 1,5 m. zawierający  $3 \times 10^9$  par zasad, mieści się w komórce o średnicy około 20 mikronów. Dla wyjaśnienia pewnym przybliżeniem mogą być następujące porównania, a mianowicie: wyobraźmy sobie kartkę papieru formatu A-4, na której można zapisać 50 wierszy (pojedynczy odstęp na maszynie do pisania), każdy zawierający 100 znaków (nukleozydów). Na jednej zatem stronie można zapisać sekwencję – 5 tysięcy par zasad, np. cząsteczkę wirusa *φX174* mającego 5375 nukleozydów można zapisać w taki sposób na jednej stronie. Genom bakterii jest znacznie większy i potrzeba na ten cel 200 stron. Dla przykładu genom *Escherichia coli* (pałeczka okrężnicy) zawiera  $4,7 \times 10^6$  par zasad. Jak już wspomniano genom ludzki jest bez porównania większy i zawiera  $3 \times 10^9$  par zasad i dla jednego formalnego zapisu zużylibyśmy około 1 miliona kartek. Można przyjąć również, że przeciętna objętość książki stanowi 200 stron. Z naszych rozważań wynika, że wiedza o strukturze genomu ludzkiego zawiera się w około 10 000 tomów, a to przedstawia już sporą bibliotekę. Jeżeli nikogo nie przeraża liczba  $10^4$  tomów, to warto przypomnieć, że nie byłby to zwykły zbiór książek, a raczej winna to być encyklopedia. Przy założeniu, że do zakodowania białka składającego się z trzystu aminokwasów potrzeba genu złożonego z około 1000 par zasad, to do rozwiązania pozostaje nam 3 miliony genów. Zatem na jednej stronie w naszej encyklopedii znajduje się 3 – 5 genów. U stosunkowo dobrze rozpoznanych bakterii *E. coli* poznano dotychczas około 1000 genów, a ponad trzy tysiące pozostaje do analizy. Czytelnik łatwo zauważy, że długość łańcucha DNA albo genu określono tutaj liczbą nukleozydów lub par zasad. Wspomnieliśmy poprzednio o DNA jako o nośniku informacji genetycznej. Składa się on z dwóch komplementarnych łańcuchów nukleozydowych tworzących spiralę i w zasadzie każdy z pojedynczych łańcuchów DNA może

być łańcuchem kodującym syntezę białek.

Według pierwotnych obliczeń, uwzględniając poziom dostępnych technik, sekwencjonowanie 3 miliardów nukleozydów wymagać będzie 30 000 "osobo-lat" oraz kosztów około 2 miliardów dolarów. W sprzyjających okolicznościach może to być 300 osób, 20 lat oraz 1 miliard dolarów. Obecnymi metodami pojedyncza osoba może zsekwencjonować rocznie 100 000 par zasad przy średnich kosztach 1 dolara na zasadę. Jeżeli automatyzacja tego procesu posunie się nieco, wówczas liczba zsekwencjonowanych zasad może wzrosnąć o 1–2 rzędy, do wielkości  $10^6$ – $10^7$ . W takiej sytuacji dla 150 pracowników technicznych sekwencjonowanie genomu ludzkiego stanowiłoby zadanie na 20 lat. Dla porównania, koszt jednej stacji kosmicznej wynosi około 8 miliardów dolarów. W trakcie dyskusji zapoczątkowanej w połowie 1985 r. pojawiło się szereg pytań, na które trzeba znaleźć jasną i szczerą odpowiedź, zanim problem zostanie podjęty. Przypomnijmy w tym miejscu, że jeszcze około 10 lat temu określenie sekwencji DNA o długości 20 nukleotydów było zadaniem czasochłonnym, trwającym ponad rok. W dobrym laboratorium obecnie, można to zrealizować w ciągu jednego dnia. Koszty projektu są duże, ale możliwe do uzyskania. Zwykła ludzka ciekawość będzie również stymulowała działania w tym kierunku. Poważnym problemem jest brak dużej liczby biologów-specjalistów w zakresie sekwencjonowania, którzy mogliby temat ten podjąć.

Początkowo dyskusja wskazywała, że idea sekwencjonowania genomu ludzkiego uzyskała pełne poparcie naukowców. Jednakże w trakcie 2,5 rocznej dyskusji poglądy się zmieniały. Wielu doświadczonych biologów wskazuje, że znacznie bardziej interesujące byłoby jednoczesne posiadanie informacji strukturalnych (sekwencja) oraz danych genetycznych i biochemicznych. Oczywiście jest zatem, że początkowy entuzjazm w kierunku sekwencjonowania przesunął się w stronę bliższą i oczywistą, przynajmniej dla genetyki klasycznej, jak mapowanie genów. W pierwszym etapie należałoby otrzymać bibliotekę nakładających się kosmidów (wektory do klonowania, powielania dużych fragmentów DNA). Tutaj pojawiłaby się praktyczna korzyść przy diagnostyce chorób genetycznych. Dla genetyki klasycznej posiadanie mapy jest bardzo istotne, gdyż zawierałaby ona bardzo informatywne znaczniki. Obecnie istniejąca mapa przedstawia raczej niskorozdzielczy schemat i nie pozwala na dokładną analizę genomu. Ocenia się, że mapowanie powinno być wykonane w ciągu trzech lat przy nakładach 20 milionów dolarów. Przewidywanie trudności merytorycznych nie jest tematem łatwym, jeżeli w ogóle możliwym do rozwiązania, np. w przypadku, gdy w genomie znajdzie się znaczna liczba powtarzających się sekwencji. Złożenie poprawnej struktury może okazać się zadaniem wymagającym wielkiego wysiłku. Ważnym zagadnieniem wydaje się dokładność określania sekwencji. Jeden błąd może zmienić sens sekwencji kodującej określone białko. Dobrze byłoby w takim przypadku znać strukturę drugiej nici.

Wspomniano już o finansowaniu tego przedsięwzięcia. Warto tutaj dodać, że w istocie rzeczy zagadnienie to polega na konstrukcji nowego narzędzia badawczego jakim będzie znana struktura genomu człowieka. Dla tego celu potrzebne są nowe, w dużym stopniu automatyczne urządzenia laboratoryjne. Ostatnie doniesienia wskazują na możliwość 5–krotnego obniżenia kosztów po wprowadzeniu robotyzacji na niektórych etapach. W takiej sytuacji koszt określenia jednej zasady zmalałby do 20 centów.

Na pytanie, jaki genom powinien być poddany analizie brak absolutnie odpowiedzi. Wiadomo jednakże, że konieczne jest posiadanie genomu haploidalnego, w takiej sytuacji można przyjąć założenie wykluczające kobiety jako dawcy genomu diploidalnego.

Jedną z przyczyn pojawiających się często wątpliwości jest brak pewności, czy zadanie jest wykonalne. Jeśli przy obecnych technikach sekwencjonowania możliwa jest analiza  $10^6$  nukleotydów na 1 rok, nie należy się dziwić, że perspektywa 3000 lat oczekiwania na ostateczny wynik nie jest zachęcająca. W związku z tym podejmowane są realne wysiłki mające na celu przyspieszenie tego procesu. Istota analizy struktury genomu polega na wykonaniu serii specyficznych fragmentów oraz wielu etapów rozdziału. Analiza genomu ludzkiego mogłaby przebiegać według następujących etapów: rozdział chromosomów, restrykcja chromosomalnego DNA, klonowanie fragmentów restrykcyjnych, wykorzystując kosmidy, dalsze specyficzne zmniejszanie wielkości fragmentów DNA oraz klonowanie i sekwencjonowanie. Realizując te główne etapy, liczba fragmentów DNA przeznaczonych do sekwencjonowania rośnie aż do około  $10^7$ , gdyż oligonukleotydy wielkości  $10^2$  potrafimy łatwo analizować. Aby to wykonać proces ten musi być zautomatyzowany. Obecnie testowany robot firmy SEIKO może sekwencjonować 300 000 zasad na 1 dzień, czyli 1 zasadę w ciągu 0,28 sekund. Dużym udoskonaleniem jest metoda pulsowej elektroforezy, która pozwala rozdzielać fragmenty DNA od 10 000 do 1 miliona zasad. Dotychczasowe techniki umożliwiały rozdział fragmentów o długości maksymalnej do 20 000 zasad. Techniczne problemy związane z produkcją cienkich żeli sekwencyjnych (podstawą rozdziału jest wielkość, długość cząsteczki) wydają się bliskie pokonania. Firma FUJI PHOTO FILM jest przygotowana do produkcji 1000 żeli o wymiarach 40 cm x 20 cm x 0,2 cm w ciągu 1 dnia. Inny robot opracowano w Kalifornijskim Instytucie Technologii. Do sekwencjonowania wykorzystuje się cztery specyficzne reakcje enzymatyczne przebiegające w obecności startera (specyficzna sekwencja nukleotydów) znakowanego jednym z czterech barwników. Każda specyficzna reakcja kodowana jest odpowiednim kolorem. Fragment DNA po wykonaniu reakcji sekwencjonowania rozdziela się na kolumnie wypełnionej żel. Znajdujący się u wylotu kolumny laser monitoruje wypływ i odczytuje sekwencję analizowanego fragmentu.

Na zakończenie należy powiedzieć, że o ile to gigantyczne zadanie zostanie podjęte to nie będzie ono realizowane przez pojedynczy zespół czy grupę zapaleńców. Problem jest olbrzymiej wagi. Jeśli ma być on rozwiązany to powinna być podjęta szeroka współpraca międzynarodowa. W tym wysiłku powinniśmy brać również udział jako kraj, niezależnie od bieżących trudności i niedostatków.