NOTATKI DO CZĘŚCI BIOLOGICZNEJ

Jerzy Tiuryn
Instytut Informatyki
Uniwersytet Warszawski
rok akademicki 1999/2000

5 lutego 2001

Spis treści

1 Kwasy nukleinowe

Przechowują i przekazują informację w komórkach.

DNA – *kwas deoksyrybonukleinowy* (przechowuje zakodowaną informację, jest podstawowym materiałem genetycznym).

RNA – *kwas rybonukleinowy* (jest aktywnym czynnikiem w dekodowaniu informacji).

Obydwa kwasy są liniowymi polimerami składającymi się z monomerów połączonych w łańcuch. Te monomery (cegiełki do budowy kwasów) nazywa się *nukleotydami*.

Zarówno DNA jak i RNA jest zbudowane tylko z czterech rodzajów nukletydów. Trzy rodzaje występują w obydwu kwasach: *adenina* (A), *guanina* (G), *cytozyna* (C). Ponadto w DNA występuje *tymina* (T), a w RNA występuje *uracyl* (U).

Każdy nukleotyd składa się z trzech części:

- reszta fosforanowa
- cukier (oparty na pięciu węglach, tzw. *pentoza*). W DNA tym cukrem jest *deoksyryboza*, a w RNA *ryboza*.
- zasada azotowa (stąd nukleotyd bierze swoją nazwę).

Rysunki: schemat nukleotydu (4-1,L), struktura rybozy i deoksyrybozy (4-2,L).

Zasady występujące w nukleotydach dzielą się na dwie grupy:

- *puryny* (A, G)
- *pyrymidyny* (C, T, U).

Rysunki: przykładowy schemat nukletydu (A) (4-4,L), schemat łączenia polimerów (4-5,L).

Łańcuch nukletydów ma dwa różne końce: 5' (koniec fosfatowy) i 3' (koniec hydroksylowy) (liczby pochodzą od numeracji węgli). Zwykle sekwencje są zapisywane i czytane w kierunku od 5' do 3'.

1.1 DNA

W naturalnym stanie tworzy podwójną helisę (Watson i Crick, 1953) powstałą z połączenia dwóch łańcuchów nukleotydów (wiązania wodorowe). Kierunki łańcuchów w helisie są przeciwne. Zwykle helisa jest prawoskrętna.

Rysunki: schemat połaczeń w podwójnej helisie (4-6,L), (4-7,L). Modele podwójnej helisy (4-8,L), (4-9,L).

Pary komplementarne nukleotydów:

- A — T (słabsze wiązanie).
- G —- C (silniejsze wiązanie).

Możliwe są wyjątki, np. G — T.

Kształt: może być nić (podwójna) lub cykl.

Rozmiary: zasady są oddalone od siebie o 0.34 nm (licząc wzdłuż osi). Pełny obrót co 3.4 nm.

Denaturacja DNA – rozplątanie nici pod wpływem wysokiej temperatury (zwykle 80-90C, zależnie od kwaśności środowiska i składu nukleotydów). Słabsze wiązania rozpadają się szybciej. Po obniżeniu temperatury następuje ponowne łączenie w helisę (*renaturacja*).

1.2 RNA

Budowa chemiczna podobna do DNA, zamiast tyminy (T) jest uracyl (U). Jest hipoteza, że RNA pojawiło się na świecie przed DNA. RNA jest mniej stabilne niż DNA (np. rozpada się w alkalicznych roztworach). Podobnie jak DNA, tworzy pojedyncze nici, podwójne, liniowe lub cykliczne. Ważna jest trójwymiarowa struktura RNA (właściwości katalityczne lub łączenie się z białkami).

1.3 Centralny dogmat biologii molekularnej

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{białko}$$

- Przejście od DNA do RNA – *transkrypcja.*
- Przejście od RNA do białka – *translacja.*
- Białka biorą udział w wielu procesach (jako enzymy, składniki budulcowe, regulatory innych procesów).

Geny – minimalne fragmenty nici DNA zawierające informację odpowiedzialną za działanie każdej niezależnej funkcji w organizmie.

2 Zasady syntezy kwasów nukleinowych

1. Zarówno DNA jak i RNA są produkowane przez kopiowanie istniejącej wcześniej nici DNA (zgodnie z zasadą komplementarności Watsona-Cricka).

2. Kierunek kopiowania jest zawsze od 5' do 3' (tzn. powstająca nić jest budowana w kierunku od 5' do 3').

3. Przy kopiowaniu biorą udział specjalne enzymy, nazywane *polimerazami*. *DNA polimerazy* służą do budowania DNA, a *RNA polimerazy* służą do budowania RNA. Budowa RNA na podstawie matrycy DNA nazywa się *transkrypcją*. RNA polimerazy mogą zacząć budowę nowej nici (*de novo*). Natomiast DNA polimerazy nie mogą – muszą zaczynać od pewnego miejsca zwanego *starterem* lub *primerem* i doklejają nową nić z nukleotydów. Rysunek: transkrypcja DNA na RNA (4-20,L).

4. Synteza podwójnego DNA (*replikacja*) przebiega poprzez mechanizm zwany *widełkami replikacyjnymi*. Synteza wzdłuż jednej nici (*nić prowadząca, leading strand*) odbywa się w sposób ciągły. Synteza wzdłuż drugiej nici (*nić opóźniająca, lagging strand*) odbywa się kawałkami (zawsze od 5' do 3') zwanymi *fragmentami Okazaki*. Rysunek: schemat replikacji (4-22,L).

3 Trzy królestwa

3.1 Prokaryota

Organizmy, których komórki nie zawierają jądra, np. bakterie. (*To nie jest pełna definicja!*).

3.2 Eukaryota

Organizmy, których komórki zawierają jądro (np. drożdże, nicień, muszka owocowa, człowiek).

3.3 Archea

Bakterie żyjące w ekstremalnych warunkach (temperatura, ciśnienie), np. na dnach oceanów. Nie mają jądra, ale pod pewnymi względami (np. istnienie egzonów/intronów) przypominają Eukarionty.

4 Transkrypcja genów prokariontycznych

Geny odpowiedzialne za jeden cel metaboliczny (np. synteza jakiegoś aminokwasu) leżą obok siebie (na chromosomie) i ich transkrypcja jest wsólnie regulowana. Taki fragment nazywa się *operonem*. Transkrypcja przebiega według zwykłej zasady syntezy. Powstaje mRNA (*messanger RNA* lub *informacyjne RNA*). Jest to RNA zawierający wszystkie instrukcje potrzebne do syntezy białka.

Rysunek: schemat transkrypcji u prokarionta (4-23(a),L)

5 Transkrypcja genów eukariontycznych

U eukariontów geny są "w kawałkach" poprzedzielanych niekodującymi fragmentami (*intronami*). Kodujące części (*egzony*) oraz introny są początkowo kopiowane na pre-mRNA, które jest dualną kopią DNA (z zamianą T na U). Następnie dzięki procesowi wycinania intronów i sklejania pozostałych egzonów (*splicing*) powstaje ostateczne mRNA.

Rysunek: schemat transkrypcji u eukarionta (4-25,L). (*splicing*).

6 Białka

Białka są zbudowane z jednego lub kilku łańcuchów (*polipeptydów*). Polipeptydy są liniowymi polimerami cząstek nazywanych *aminokwasami*). Aminokwasy łączą się ze sobą przy pomocy silnych wiązań zwanych *wiązaniami peptydowymi*. Jest 20 rodzajów aminokwasów. W aminokwasach wyróżnia się dwa końce (podobnie jak w nukleotydach):

- koniec aminowy (N-koniec),
- koniec karboksylowy (C-koniec).

Rysunki: aminokwasy (2-15,B), schemat wiązania peptydowego (2-16,B).

Liczba aminokwasów występujących w białkach waha się od kilkudziesięciu do kilku tysięcy (są białka zawierające 30 000 aminokwasów). Białka spełniają wiele bardzo ważnych funkcji w działaniu organizmu (strukturalne, katalityczne (enzymy), regulacyjne (hormony), transportowe i magazynujące (hemoglobina), kontrolujące proces transkrypcji i translacji, układ odpornościowy (przeciwciała)). Struktura przestrzenna białek ma ogromny wpływ na ich funkcję.

Rysunek: różne kształty białek (3-1,L).

W przyrodzie występuje stosunkowo niewiele białek (jak na liczbę możliwych kombinacji aminokwasów) – znakomita większość kombinacji jest niestabilna. Obecnie znanych jest ponad 100 000 białek.

7 Kod genetyczny

Każdemu aminokwasowi odpowiada sekwencja trzech nukleotydów (*kodon*), która go koduje. Często więcej niż jedna sekwencja koduje ten sam aminokwas, np. leucyna ma 6 kodonów.

tablice kodonów: (4-3,L), (4-4,L).

Kodon AUG (*metionina*) koduje początek polipeptydu. Są trzy kodony stopu (nie kodują żadnego aminokwasu): UAA, UGA, UAG. Tak więc łącznie 61 kodonów reprezentuje 20 aminokwasów.

Ciąg nukleotydów (RNA) wyznacza pewien ciąg aminokwasów. W zależności gdzie zaczniemy czytać, możemy dostać trzy różne ciągi aminokwasów. *Ramka odczytu* to ciąg nukletydów nie przerwany kodonem stopu.

8 Translacja

Jest to proces przetwarzania zakodowanej (przy pomocy kodonów) informacji w mRNA na białko. Jest to bardzo skomplikowany proces, w którym biorą udział następujące kwasy rybonukleinowe:

- mRNA, informacyjne RNA, zawiera zakodowaną informację o białku.
- tRNA, *transportujący RNA* (ang. *transfer RNA*). Długośc": 70-90 nukleotydów. Ustala odpowiedniość pomiędzy aminokwasem i kodonem. Łączenie 'pustego' tRNA z odpowiadającym aminokwasem jest regulowane przez enzym. Powyższa odpowiedniość jest osiągnięta poprzez tzw. *antykodon* czyli kodon komplementarny. Są różne tRNA dla różnych aminokwasów. Bakterie mają około 30-40 różnych rodzajów tRNA, a zwierzęta i rośliny około 50.
  Rysunek: schemat tRNA (4-31(a),L).
- rRNA (*rybosomowe RNA*). Rybosomy są kompleksami składającymi się z wielu podjednostek białkowych i rRNA. Przeprowadzają one właściwą syntezę białek.

8.1 Proces translacji

1. (*Faza inicjacji*) tRNA niosący metioninę, wraz z innymi białkami (inicjującymi) i z mniejszą częścią rybosomu (rybosom składa się z dwóch nierównych części) łączą się z mRNA w pozycji start (AUG). Wówczas dołącza się większa część rybosomu.
   Rysunek: faza inicjacji (4-42(a),L).

2. (*Faza wydłużania*) Zostaje pobrany następny tRNA, którego antykodon pasuje do kolejnego kodonu w ciągu mRNA. Polipeptyd wyprodukowany dotąd 'wisi' przy poprzednim tRNA. Nowy aminokwas zostaje dołączony do polipeptydu (na końcu karboksylowym) i większy polipeptyd 'wisi' teraz przy ostatnim tRNA. Poprzednie tRNA (puste) zostaje wydalone z rybosomu. Wydłużanie polipeptydu jest przeprowadzane od końca aminowego (N) w kierunku końca karboksylowego (C). Następnie rybosom przesuwa się o kolejne trzy pozycje wzdłuż mRNA (w kierunku od 5' do 3').

3. (*Terminacja*) Po napotkaniu przez rybosom kodonu stopu następuje uwolnienie zbudowanego polipeptydu, uwolnienie ostatniego pustego tRNA i odłączenie rybosomu od mRNA. Rysunek: schemat fazy wydłużania i terminacji (4-42(b,c),L).

Kilka rybosomów (po kolei) może pracować w tym samym czasie na jednym mRNA. Prędkość budowania polipeptydu przez rybosom to 3-5 aminokwasów na sekundę. Zatem białko o 100-200 aminokwasach jest produkowane przez ok. 1 minutę. Duże białka (np. titina, 30 000 aminokwasów) są budowane przez 2-3 godziny.

9 Geny, chromosomy, genomy

9.1 Co to jest gen?

Jest to taki fragment ciągu nukleotydów DNA, który jest niezbędny do syntezy funkcjonalnego polipeptydu lub cząsteczki RNA. Tak więc, oprócz niezbędnej informacji kodującej ciąg aminokwasów w białku lub funcjonalne RNA (takie jak tRNA lub rRNA) geny zawierają jeszcze inne ciągi nukleotydów DNA, które są niezbędne do przeprowadzenia transkrypcji. Takie ciągi mogą leżeć daleko (50 kbp lub więcej) od regionu kodującego.

9.2 Chromosom

Jest to cząsteczka DNA zawarta w komórce (spójny kawałek). Otaczają tę cząsteczkę białka odpowiedzialne za ekspresję genów. Liczba różnych chromosomów (oraz ich wielkość) są specyficzne dla każdego gatunku. U prokariontów jest jeden chromosom (jest on zwykle cykliczną cząsteczką DNA). U eukariontów jest wiele chromosomów. Zawarte są one w jądrach komórek. Chromosomy eukariontów są liniowymi podwójnymi nićmi DNA.

DNA w chromosomach jest bardzo mocno skondensowane (zwinięte). Np. kolisty chromosom DNA *E. coli* ma po rozplecieniu długość ok. 1.4mm, ale mieści się w komórce o długości 0.002 mm i średnicy 0.001 mm. W komórce ludzkiej DNA o łącznej długości ok. 2 m (tak! 2000 mm) mieści się w jądrze komórki o średnicy 0.005 mm.

U większości eukariontów chromosomy występują parami identycznych (*homologicznych*) chromosomów. Przykładowe liczby chromosomów (*liczby haploidalne*):
człowiek: 23
rezus: 21
pies: 39
kot: 19
koń: 32
karp: 52
drożdż: 16
ziemniak: 24
cebula: 8
muszka owocowa: 4

Czasem zdaża się dodatkowy chromosom. Zwykle powoduje to objawy chorobowe. Np. u człowieka trzeci chromosom nr.21 powoduje wystąpienie zespołu Downa.

9.3 Genomy

Genom jest to całkowity DNA zawarty w chromosomach danego gatunku. Ponieważ chromosomy eukariontów występują zwykle w homologicznych parach, to wystarczy podać sekwencję połowy chromosomów dla określenia całego genomu. Nazywa się to tzw. *haploidalnym genomem.*

Przykładowe rozmiary haploidalnych genomów:
wiroid: 0,38 kbp
wirus, fag λ: 50 kbp
Escherichia coli: 4 000 kbp
drożdż: 14 000 kbp
nicień: 80 000 kbp
muszka owocowa: 170 000 kbp
kura: 1 200 000 kbp
człowiek: 3 500 000 kbp
kukurydza: 5 000 000 kbp.

W genomach eukariontów większość DNA nie ma (poznanego) funkcyjnego znaczenia — nie koduje białek ani RNA (tzw. *junk DNA*). Szacuje się, że ponad 90% genomu człowieka nie ma funcyjnego znaczenia. Nie wiadomo jaka jest rola tej części DNA.

9.4 Historia poznawania sekwencji genomów

- Bakteriofag $\phi X174$ (5,4 kbp) – pierwszy całkowicie zsekwencjonowany w całości materiał genetyczny (Sanger, 1972).
- Fag λ (48,5 kbp), Sanger 1982.
- Bakteria *Haemophilus influenze* (1 830 kbp), Venter 1995.
- Bakteria *Mycoplasma genitalium* (580 kbp), Venter 1995.
- Drożdże piekarskie *Saccharomyces cerevisiae* (13 500 kbp) – pierwszy genom eukarionta, (współpraca międzynarodowa), 1996.
- Bakteria *Escherichia coli* (4 000 kbp), 1997.
- Nicień *Caenorhabditis elegans* (100 000 kbp) 1998.

Trwają prace:

- Muszka owocowa *Drosophila melanogaster* (160 000 kbp).
- Mysz *Mus musculus* (3 500 000 kbp) szacuje się, że w 75% genom myszy jest podobny do genomu człowieka.
- Rzodkiewnik *Arabidopsis thaliana* (100 000 kbp)

9.5 Projekt poznania genomu człowieka (Human Genome Project)

Zapis całego genomu zajmie (używając alfabetu A,C,G,T) około 750 Mb pamięci.

Inicjatywa projektu powstała w USA pod koniec 1984. Uruchomienie projektu nastąpiło w 1988. Główną rolę przy realizacji przewidziano dla NIH (National Institutes of Health), ale również mocno zaangażowany jest Departament Energii.

Projekt zaplanowano na 15 lat i łączny budżet 3 mld USD (po 200 mln USD na rok). Jest to becnie bardzo duże międzynarodowe przedsięwzięcie, w którym udział biorą m.in.: USA, Francja, Japonia, Niemcy, Rosja, Wielka Brytania.

10 Pewne metody rekombinacji DNA

10.1 Klonowanie

10.1.1 Enzymy restrykcyjne

Tną dwuniciowe DNA w specyficznym miejscu wyznaczonym przez charakterystyczny dla tego enzymu ciąg nukleotydów. Np. enzym EcoRI wytwarzany przez *E. coli* tnie pomiędzy G oraz A (w nici wiodącej) oraz pomiędzy A oraz G (w nici opóźniającej), jeśli napotka następującą sekwencję:
5' ... GAATTC ... 3'
3' ... CTTAAG ... 5'

(Tab. s.65,W) – przykłady enzymów restrykcyjnych.

Zakończenie dwuniciowego DNA o nierównych końcach nazywa się *lepkim końcem* (sticky end). Lepkie końce można enzymatycznie dołączać do "tępych" końców.

10.1.2 Wektory plazmidowe

Plazmid to mała kolista cząstka DNA, która nie jest związana z żadnym chromosomem. Może być wprowadzona do komórki bakterii, od której się wywodzi. Często używa się plazmidów *E. coli*. Plazmid musi zawierać *miejsce startu replikacji* (aby mógł się replikować, gen lub kilka genów odpowiedzialnych za odporność na pewne anybiotyki. Pozostaje jeszcze pewna część w DNA plazmidu, w którą można wstawić fragment obcego DNA (o długości do 15-20 kbp). Natępnie hoduje się bakterie, dodając antybiotyk. Te bakterie, które zawierają plazmidy z genem odporności na ten antybiotyk przeżywają. Powstaje w ten sposób *kolonia klonów*.

Rysunek: (5-11,W) – klonowanie DNA w plazmidzie.

10.1.3 Komplementarne DNA (cDNA)

Jest to DNA otrzymane z mRNA przy pomocy enzymu zwanego *odwrotną tranzkrypktazą* (pochodzi od retrowirusów). Zaletą pracy z cDNA zamiast z genomowym DNA jest to, że ten pierwszy nie zawiera już intronów tylko czysty ciąg nukleotydów kodujący białko. Każda cząsteczka cDNA jest kopią pojedynczego mRNA. Jak odróżnić mRNA od innych kwasów nukleinowych? Na końcu 3' jest długi łańcuch A, tzw. poly(A).

10.1.4 Wektory λ-fagów

λ-fagi są wirusami atakującymi bakterie E.coli. Ich genom ma rozmiar $\sim$50 kbp, z czego $\sim$25 kpb można usunąć, wymieniając na obcy DNA, bez wpływu na proces infekowania bakterii. λ-fag namnaża się w zainfekowanej komórce, produkując nawet więcej niż 100 kopii w jednej komórce. Zabija przy tym bakterię, uwalniając zreplikowane nowe kopie wirusa. Przygotowanie zmutowanych λ-fagów można przeprowadzać *in vitro*.

Rysunek: (7-11,L) – budowa λ-faga.

10.1.5 Biblioteki genomowe

Pobiera się genomowe DNA, tnie na kawałki $\sim$20 kbp przy pomocy enzymu, którym się również tnie genom λ-faga. Następnie tworzy się zrekombinowane λ-fagi i zaraża nimi bakterie aby wytworzyć dużą liczbę kopii. Tak otrzymane kolonie nazywają się *bibliotekami genomowymi*. Zamiast λ-faga używa się również plazmidów czy też sztucznych chromosomów.

10.1.6 Inne wektory

Kosmidy: połączenie techniki wykorzystania plazmidów z techniką λ-fagów (tworzy się duże plazmidy, które są wprowadzane do bakterii przy pomocy "obudowy" od λ-fagów). Dzięki temu pomysłowi można tworzyć wstawki o długości $\sim$45 kbp.
Wektory P1: $\leq$ 100 kbp.
YAC: (sztuczny chromosom drożdża) $\leq$ 1000 kbp.

10.2 Metody sekwencjonowania

10.2.1 Metoda chemiczna Maxama-Gilberta (1977)

Przeprowadza się cztery reakcje chemiczne, które tną nić DNA w miejscach G, A+G (puryny), C+T (pirymidyny), C. Reakcja jest tak przeprowadzana aby każdą nić przeciąć tylko w jednym miejscu. Następnie, używając elektroforezy na żelu, porównuje się długości pociętych fragmentów (radioaktywnie zaznacza się jeden koniec nici i przeprowadza autoradiografię prążków powstałych w żelu). Ta metoda jest pracochłonna i kosztowna.

Rysunek: (5-3,W) – ilustracja metody Maxama-Gilberta.

10.2.2 Metoda enzymatyczna Sangera (1977)

Matrycą jest jednoniciowe DNA. Do niej dołącza się starter (tzw. *primer*), jest on syntetycznie wyprodukowany. Jego koniec 5' jest oznakowany radioaktywnie. Proces przeprowadza się w 4 probówkach. W każdej z nich znajduje się matryca ze starterem, nukleotydy A, C, G, T oraz jeden z czetrech rodzajów nukleotydów pozbawiony grupy 3'-hydroksylowej (tzw. *dideoksy*

ATP, CTP, GTP, TTP). Do tego dodaje się enzym DNA polimerazy, który powoduje wydłużanie nici zaczynającej sie starterem. Dołaczenie w danym momencie dideoksy powoduje, że proces wydłużania zostaje zatrzymany (bo nie ma końca 3'). Następnie denaturuje się podwójne nici, wykonuje elektroforezę dla porównania długości i przeprowadza autoradiografię prążków powstałych w żelu.

Rysunek: (5-4, W) – ilustracja metody Sangera.

10.3 Technika PCR

PCR, reakcja łańcuchowa polimerazy (*Polimerase Chain Reaction*) odkryta przez Kary Mullis ($\sim$ 1985) jest używana do szybkiego kopiowania materiału genetycznego. Proces ma charakter wzrostu wykładniczego.

Podstawowe kroki procesu:

Załóżmy, że chcemy powielić materiał genetyczny zawarty w pewnym regionie DNA. Wytwarzamy syntetycznie startery (dwa rodzaje), których pozycje będą ograniczały ten region z obu stron. Dajemy bardzo dużo starterów oraz wolnych nukleotydów do materiału, który mamy powielić. Dodajemy enzym polimerazy (zwykle *Taq* polimerazę z bakterii *Thermus aquaticus*, która żyje w gorących źródłach – może wydłużać nici w temperaturze do 72C). Następnie powtarzamy cyklicznie następujące trzy kroki:

1. Podgrzewamy roztwór do temperatury 95C. Natępuje denaturacja nici materiału genetycznego.

2. Ochładzamy do 60C. Wówczas startery łączą się z pojedynczymi nićmi. Ponieważ materiału genetycznego jest mało w porównaniu z liczbą starterów to szansa połączenia nici materiału genetycznego ze sobą jest mała.

3. Natępuje faza budowania nowych (komplementarnych) nici przez enzym polimerazy. Po zakończeniu tej fazy cykl powtarzamy tak długo aż uzyskamy odpowiednią ilość materiału genetycznego.

Po każdym cyklu PCR ilość materiału genetycznego podwaja się. W ten sposób można powielać (*amplifikować*) fragmenty o długości do 20 kbp. Przykładowe zastosowanie: wykrywanie obecności wirusa HIV w bardzo wczesnym stadium zarażenia (przed wystąpieniem objawów).

Rysunek: (7-27,L) – ilustracja techniki PCR.

Literatura (do części biologicznej):

1. *"Język genów"*, P. Berg, M. Singer, Prószyński i S-ka, 1997.

2. *"Genom człowieka"*, red. W. Krzyżosiak, PWN, Warszawa, 1997.

3. *"Genetyka molekularna"*, red. P. Węgleński, PWN, Warszawa, 1998.

4. *"Recombinant DNA"*, J.D. Watson, M. Gilman, J. Witkowski, M. Zoller, Scientific American Books, 1992.

5. *"Molecular Cell Biology"*, H. Lodish, D. Baltimore, A. Berk, ..., Scientific American Books, 1995.