

Jadwiga Laska

ODCZYTAĆ KOD ŻYCIA



Wydział Fizyki i Informatyki Stosowanej AGH
11.01.2008 r.

Plan wykładu

1. Bioinformatyka – nowa dziedzina naukowa
2. Co to jest genom, gen, chromosom? Jaka jest budowa DNA i RNA?
3. Co to jest kod genetyczny?
4. W jaki sposób następuje przekaz informacji genetycznej?
5. W jaki sposób organizm syntezuje białka o określonej sekwencji?
6. Co wynika z poznania genomu?

Bioinformatyka – nowa dziedzina naukowa



Bioinformatyka – nowa dziedzina naukowa

1871 – Mischer – pierwsze wyodrębnienie kwasu nukleinowego w postaci ekstraktu z jądra komórkowego

Początek XX w – ponowne odkrycie praw Mendla (informacja jest w dyskretny sposób przenoszona z pokolenia na pokolenie)

Pierwsze ćwierćwiecze XX w – udowodniono, że podstawowym nośnikiem informacji na poziomie komórkowym są chromosomy

1944 – Avery – odkrycie, że podstawą tej informacji jest DNA

1950 – Crick, Watson, Wilkins – ustalenie budowy przestrzennej DNA (nagroda Nobla 1962)

2001 – zakończenie „roboczego opracowania” genomu człowieka (spec. Wyd. Nature i Science).

Co to jest genom?

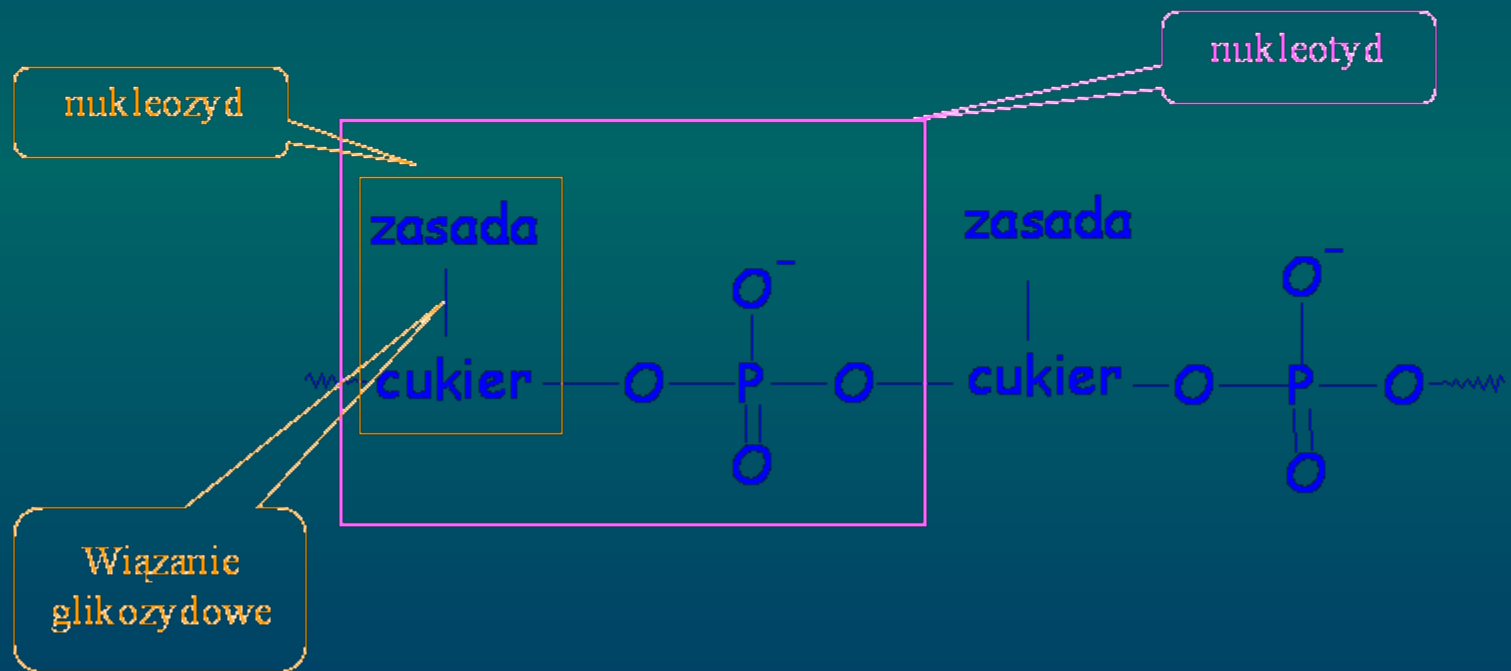
Genom termin zaproponowany pierwotnie przez Hansa Winklera, botanika z Uniwersytetu w Hamburgu, jako zbitka słów '**gen**' i '**chromosom**'.

Chromosom – forma organizacji materiału genetycznego wewnątrz komórki

Gen – odcinek DNA nadający komórce zdolność do tworzenia jakiegoś RNA

Kwasy nukleinowe

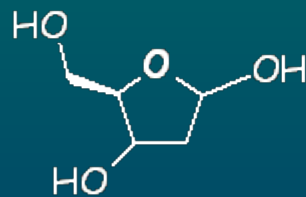
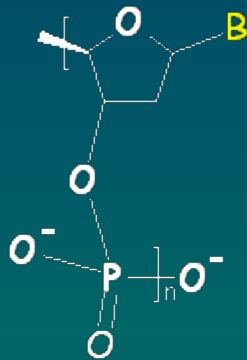
Kwasy nukleinowe są połączeniem **cukrów**, **zasad organicznych** i **grup fosforanowych**.



łańcuch polinukleotydowy
(poliester kwasu fosforowego i cukru)

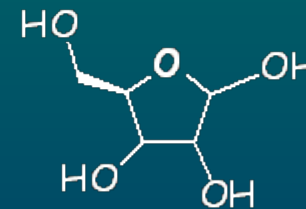
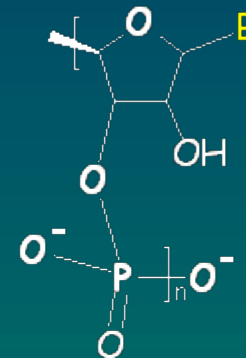
Jaka jest budowa DNA i RNA?

DNA



β -D-2-deoksyryboza

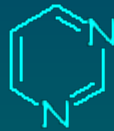
RNA



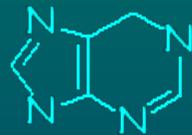
β -D-ryboza

B – zasady organiczne z grupy **pirymidyn** i **puryn**.

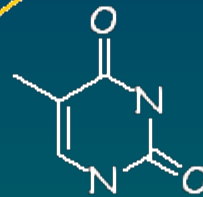
Zasady w kwasach nukleinowych



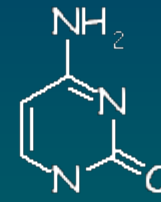
pirymidyna



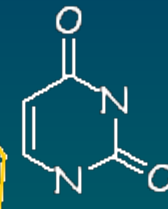
puryna



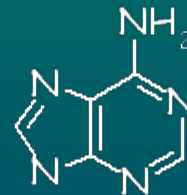
tymina (T)



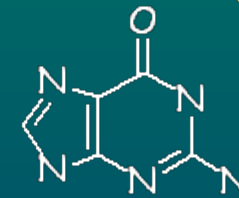
cytozyna (C)



uracyl (U)



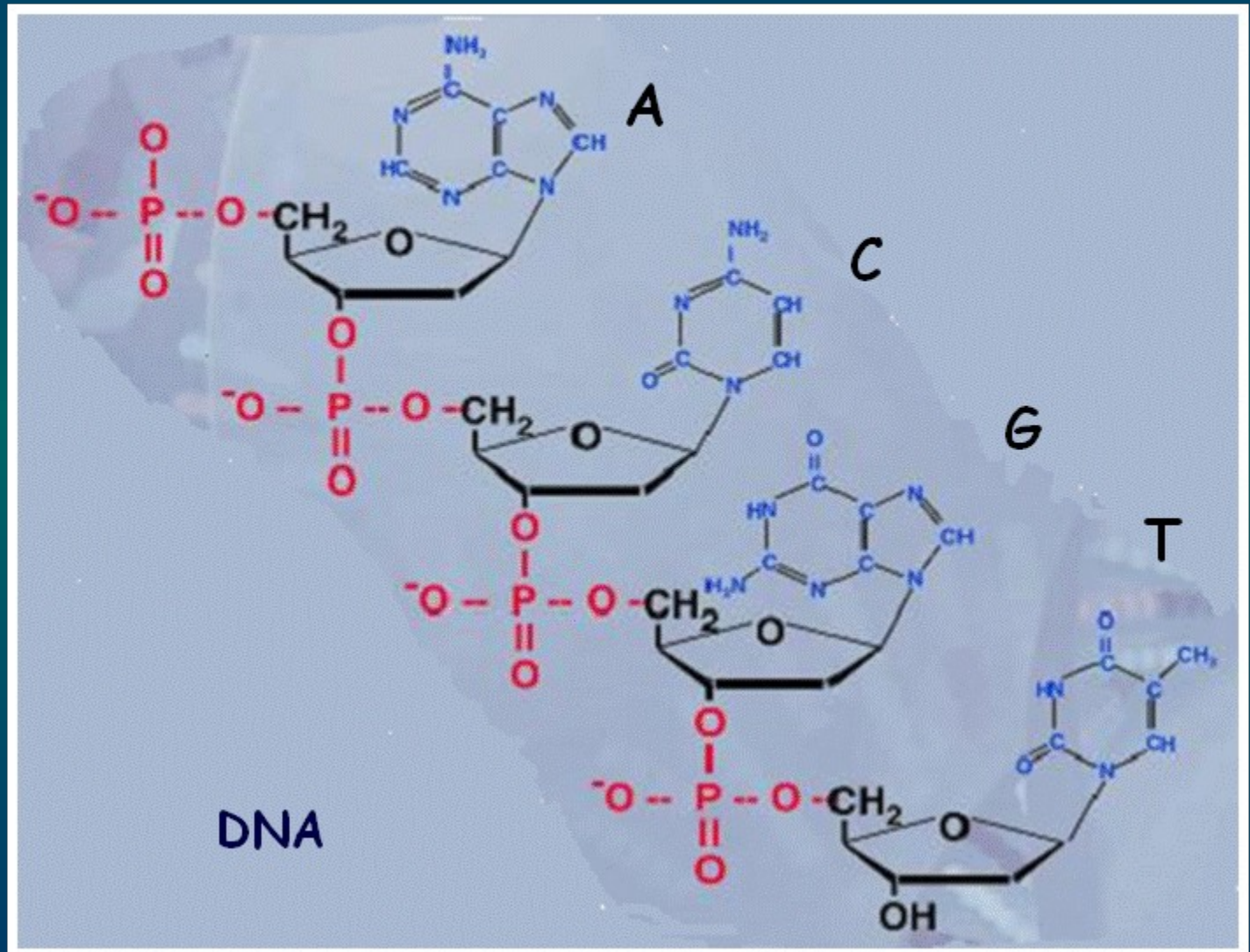
adenina (A)



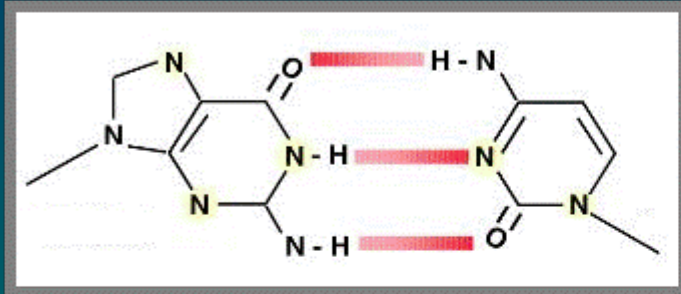
guanina (G)

W DNA występują tylko 4 zasady ACTG, w RNA również inne zasady, np. 2-tiocytosyna, 3-metylouracyl itp.. - różne typy RNA

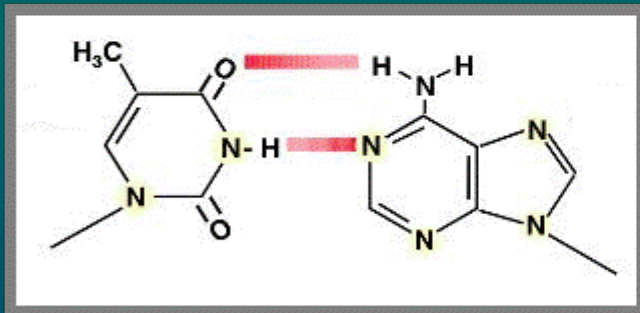
Zasady w kwasach nukleinowych



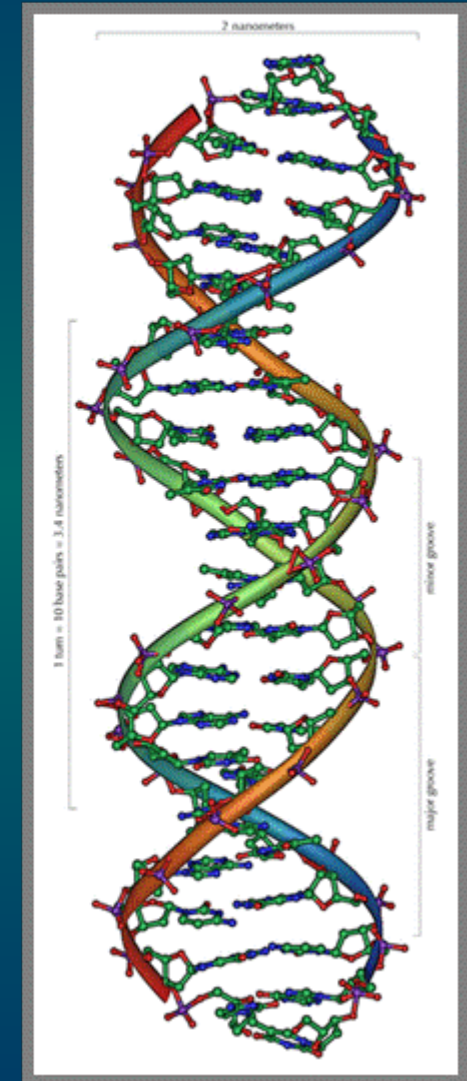
Struktura drugo- i trzecirzędowa DNA i RNA



G-C



T-A

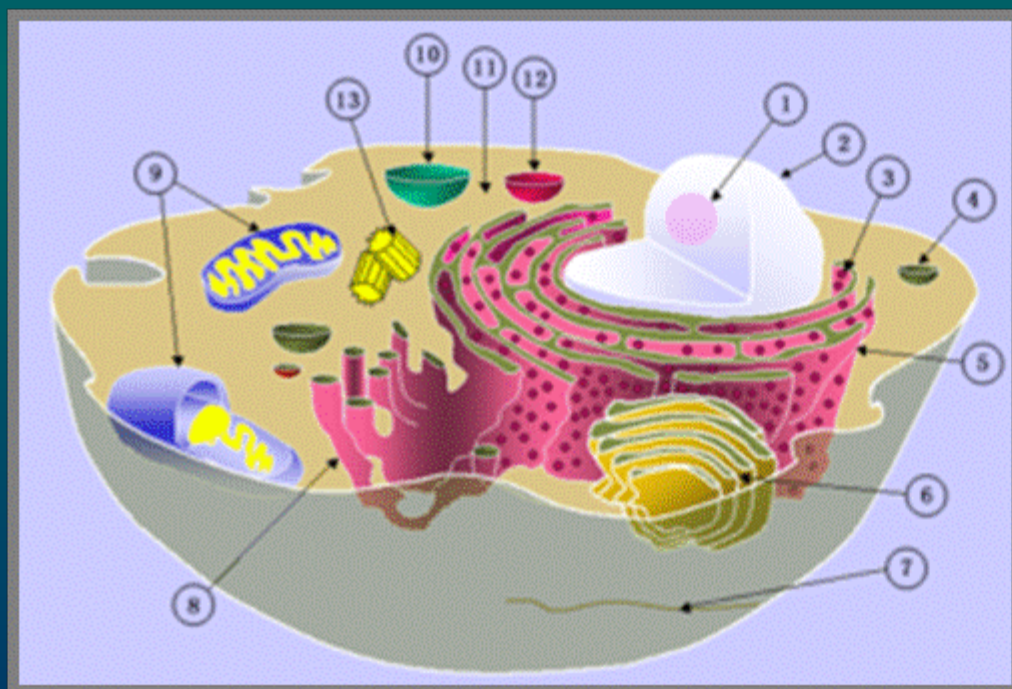


95°C - zniszczenie struktury
dwuniciowej - denaturacja
Proces odwracalny

Gdzie występują kwasy nukleinowe?

DNA na ogół występują wyłącznie w jądrze komórkowym.

RNA znajdują się w cytoplazmie i jądrze komórkowym



Budowa komórki zwierzęcej:
1 - jąderko; 2 - błona jądra komórkowego; 3 - rybosom; 4 - pęcherzyk; 5 - szorstkie retikulum endoplazmatyczne; 6 - aparat Golgiego; 7 - mikrotubule; 8 - gładkie retikulum endoplazmatyczne; 9 - mitochondrium; 10 - wakuole; 11 - cytoplazma; 12 - lizosom; 13 - centriola.

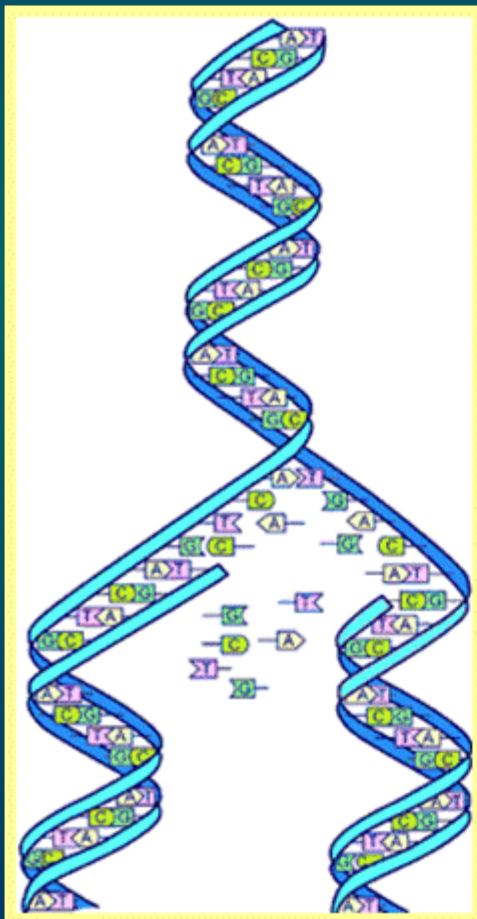
W jaki sposób następuje przekaz informacji genetycznej?

Etapy przekazu informacji genetycznej:

1. **Replikacja** – reprodukcja DNA
2. **Transkrypcja** – przepisanie informacji genetycznej, zawartej w DNA, w postaci m-RNA
3. **Translacja** – synteza białek na matrycy RNA z wykorzystaniem 20 podstawowych aminokwasów

Replikacja – reprodukcja DNA

Polega na rozwinięciu podwójnej helisy i syntezie na powstałych fragmentach jednościachuchowych

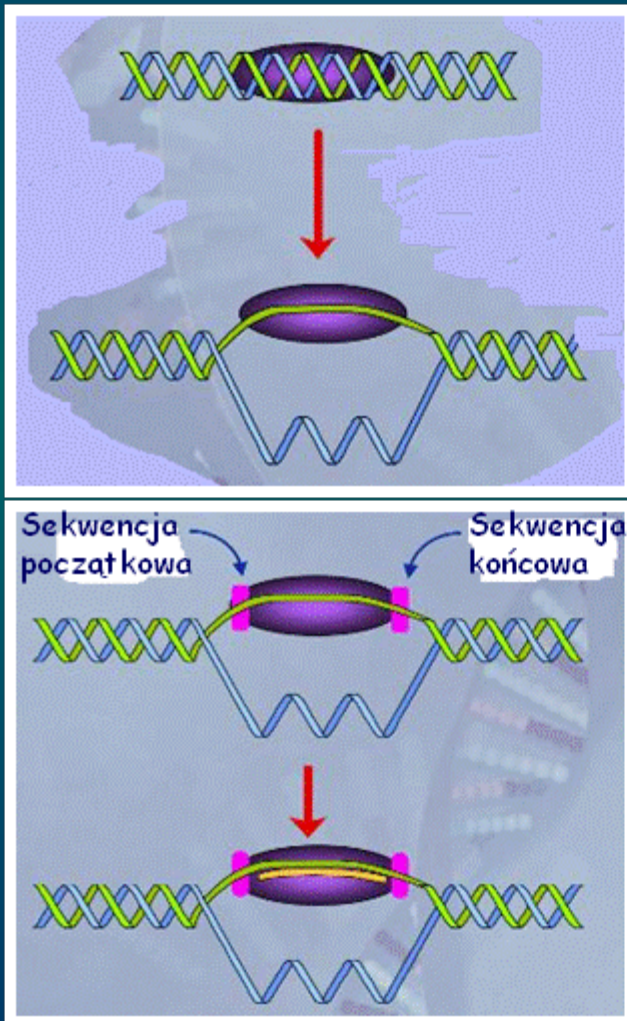


Dokładność procesu replikacji – 1 błąd na 10^8 przyłączeń

Replikacja ma miejsce tylko podczas podziału komórki

Transkrypcja

Jest to biosynteza RNA. Może przebiegać bez ograniczeń, w ciągu całego życia komórki.

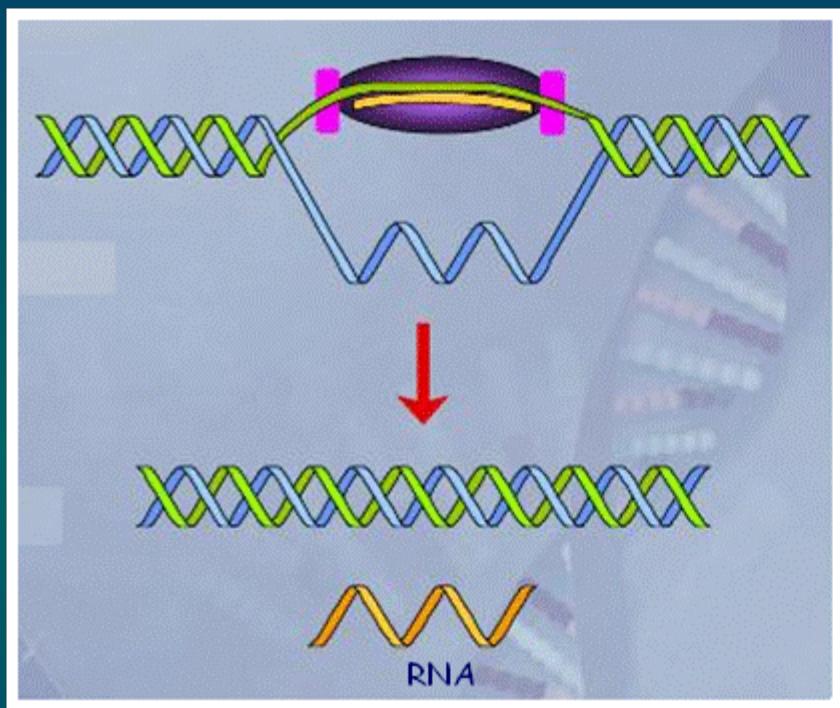


Enzym polimeraza przyłącza się do części jednej nici DNA

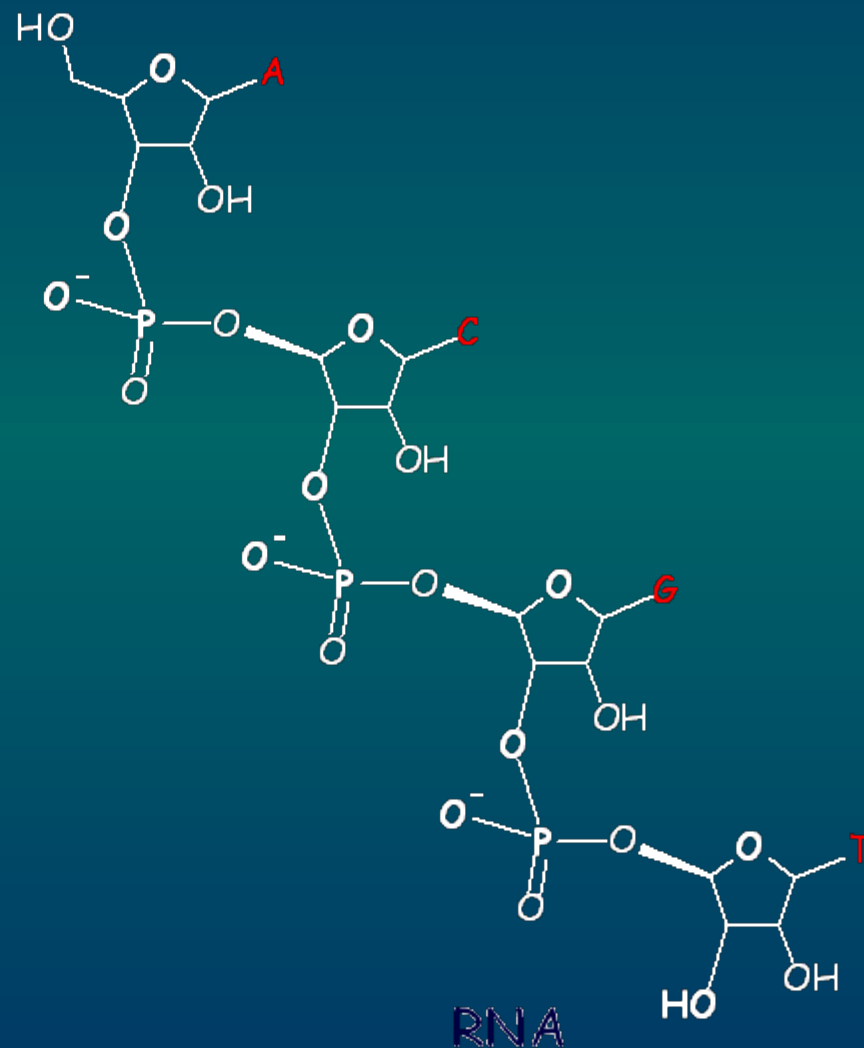
Następuje rozdzielenie nici.

Ściśle określone sekwencje zasad w DNA wskazują w którym miejscu zacznie, a w którym zakończy synteza RNA

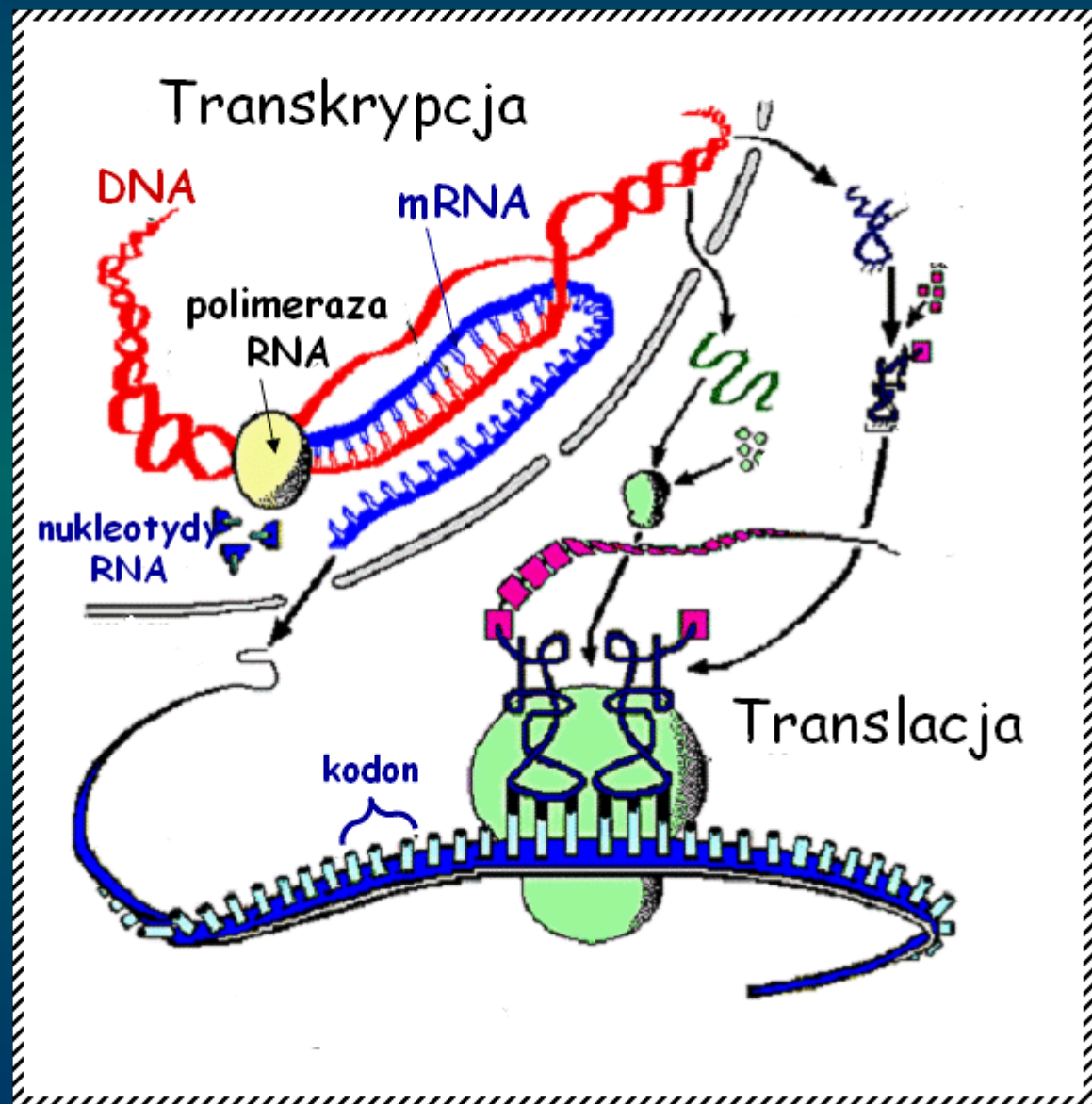
Transkrypcja c.d.



Gdy synteza RNA osiągnie końcową sekwencję zasad cząsteczka RNA i syntaza zostają uwolnione, a DNA ponownie zwiija się

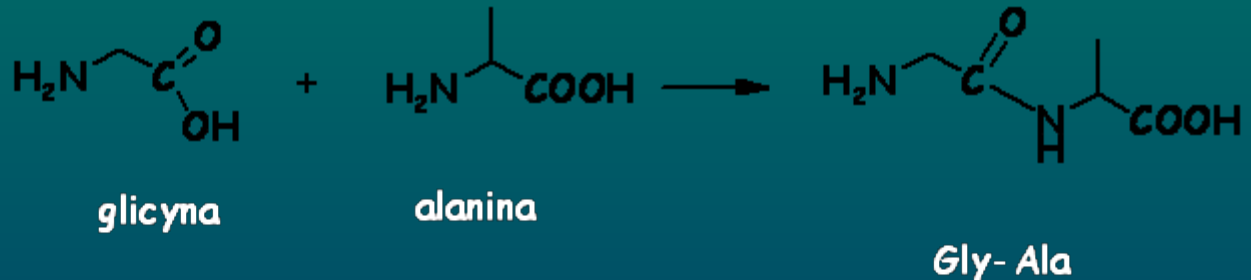
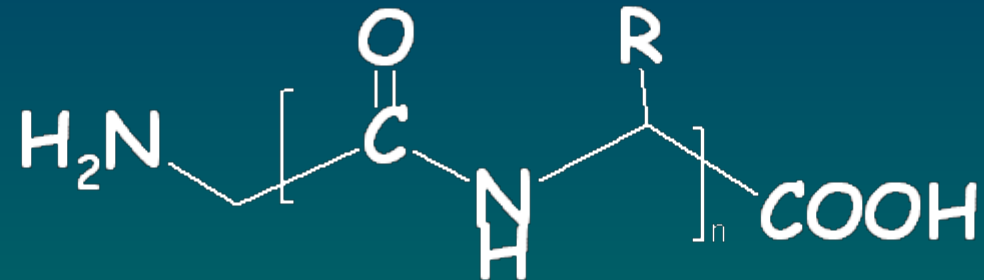


Synteza białek w komórce żywej



Białka

Sq poli(α-aminokwasami)



Tylko 20 (α-aminokwasów)!

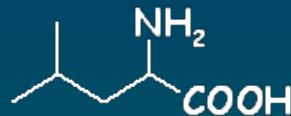
Co to są α-aminokwasy



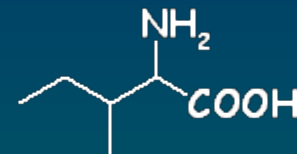
glicyna



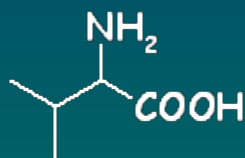
alanina



leucyna



izoleucyna



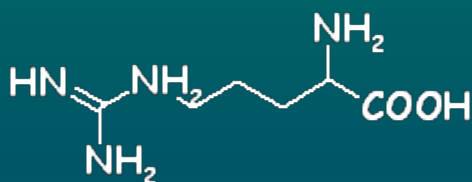
walina



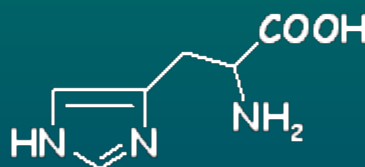
prolina



lizyna



arginina



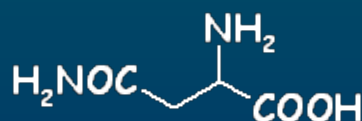
histydyna



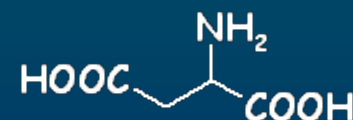
kwask glutaminowy



glutamina



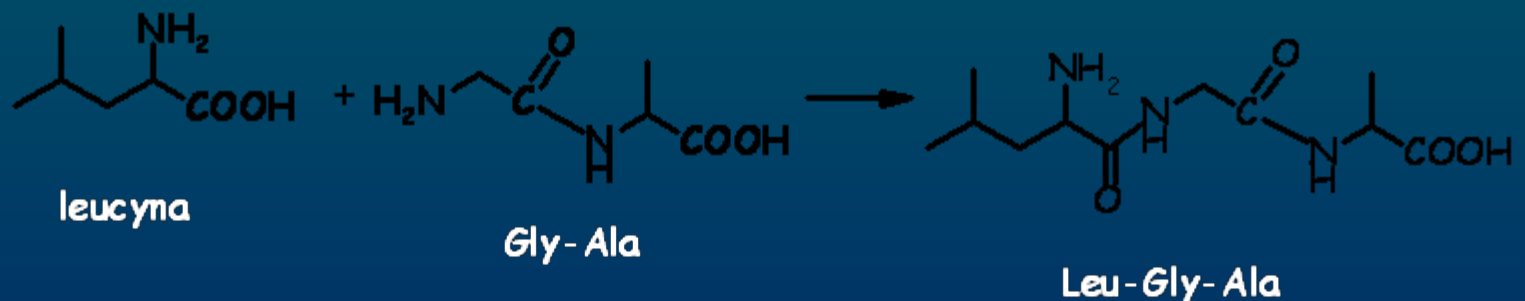
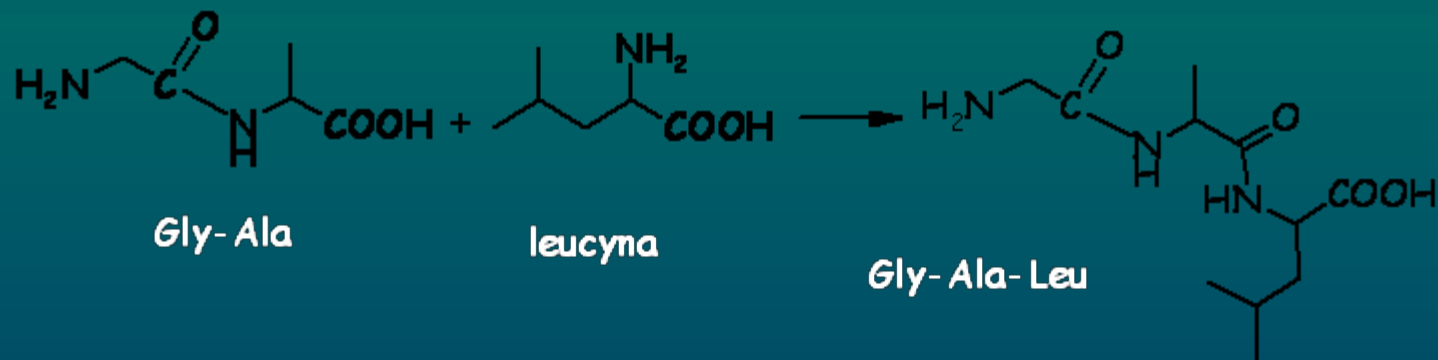
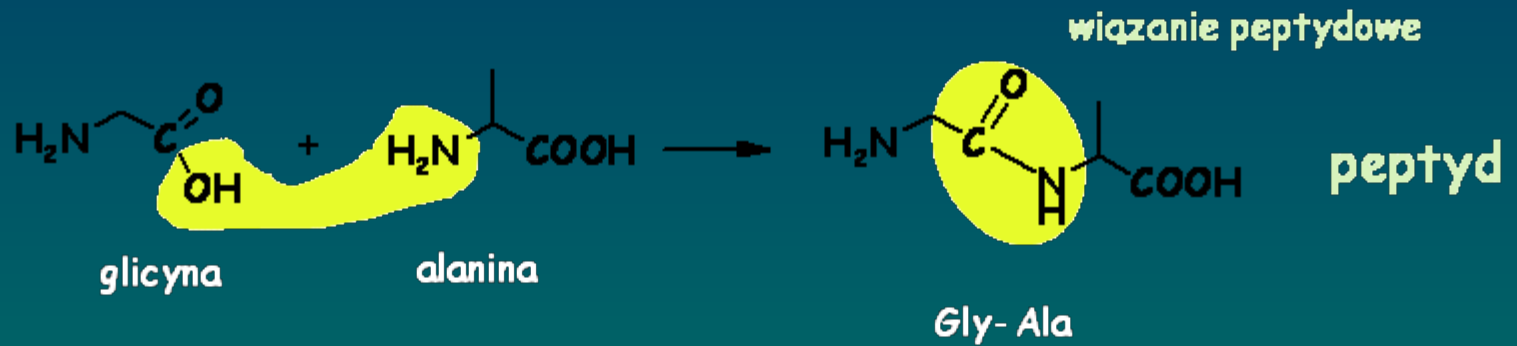
asparagina



kwask asparaginowy

Jak powstają białka?

reakcja pomiędzy aminokwasami = wiązanie peptydowe = polimeryzacja = polipeptyd



Jak powstają białka?

Gly-Ala-Leu

Gly-Leu-Ala

Ala-Gly-Leu

Ala-Leu-Gly

Leu-Gly-Ala

Leu-Ala-Gly

3 aminokwasy - 6 peptydów

20 aminokwasów - 2 432 902 008 176 640 000 peptydów (10^{18})

Większość naturalnych polipeptydów składa się z mniejszej liczby różnych aminokwasów np. hormon przysadki mózgowej - oksytocyna - składa się z 9 aminokwasów.

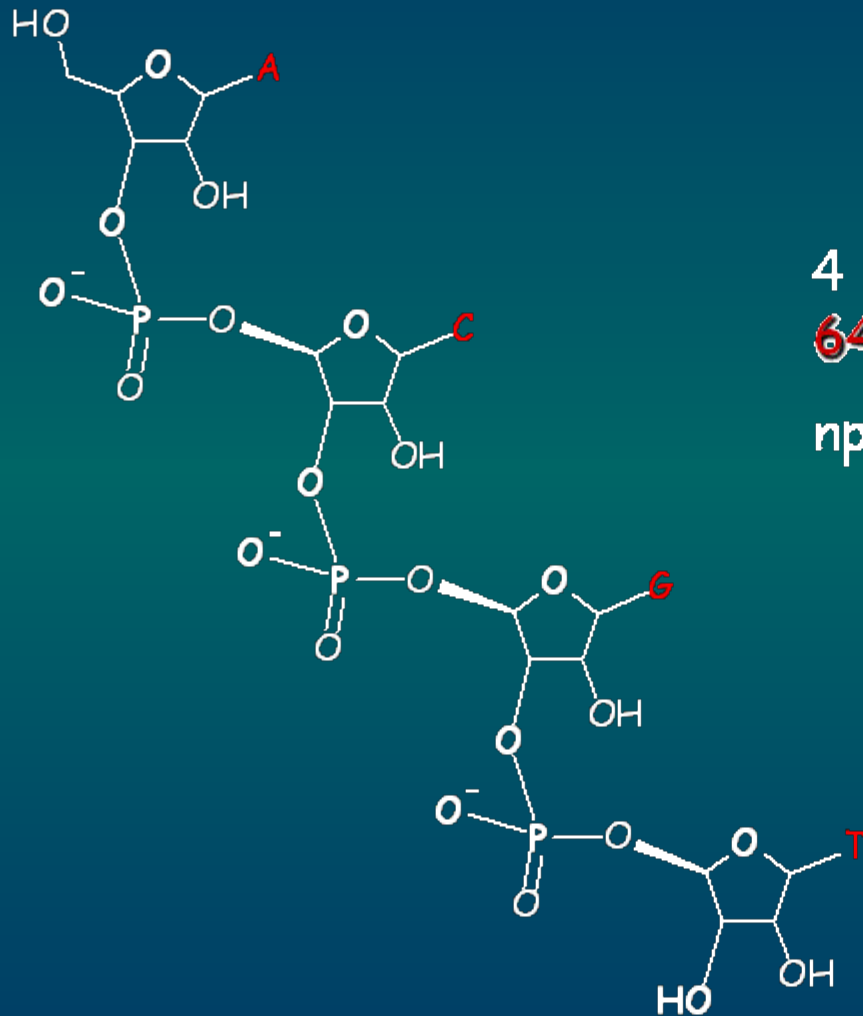
Aminokwasy te (reszty aminokwasowe) występują w danym polipeptydzie w określonej kolejności - tzw. **sekwencja białkowa**

Sekwencja ta jest zgodna z kodem genetycznym

Kolejność występowania aminokwasów w cząsteczce polipeptydu - to **struktura pierwszorzędowa białak**

Translacja

Jest to synteza białek z udziałem RNA.
Przebiega w rybosomach.



4 zasady można ułożyć trójkami na **64** różne sposoby

np. AAA, AGA, ACA, ATA, GAA, GGA, GCA...

Białka zbudowane mogą być z **20** różnych aminokwasów.

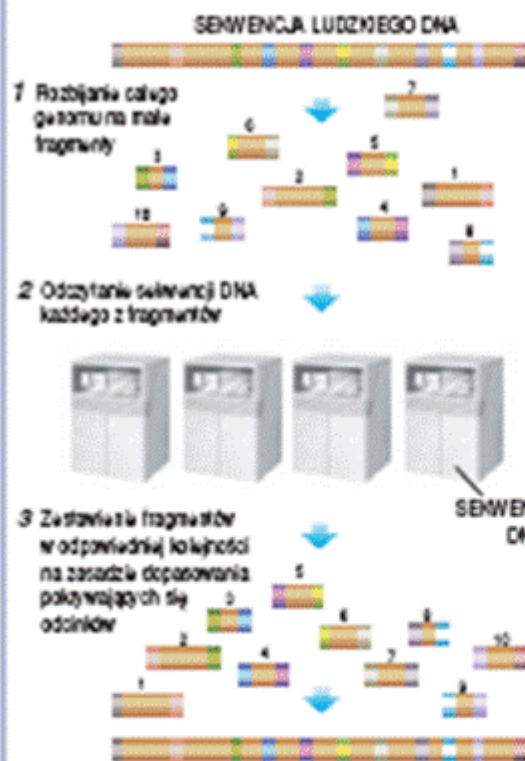
Każdy aminokwas określony jest **trójką sąsiednich zasad**.

Każdemu białku odpowiada oddzielny m-RNA.

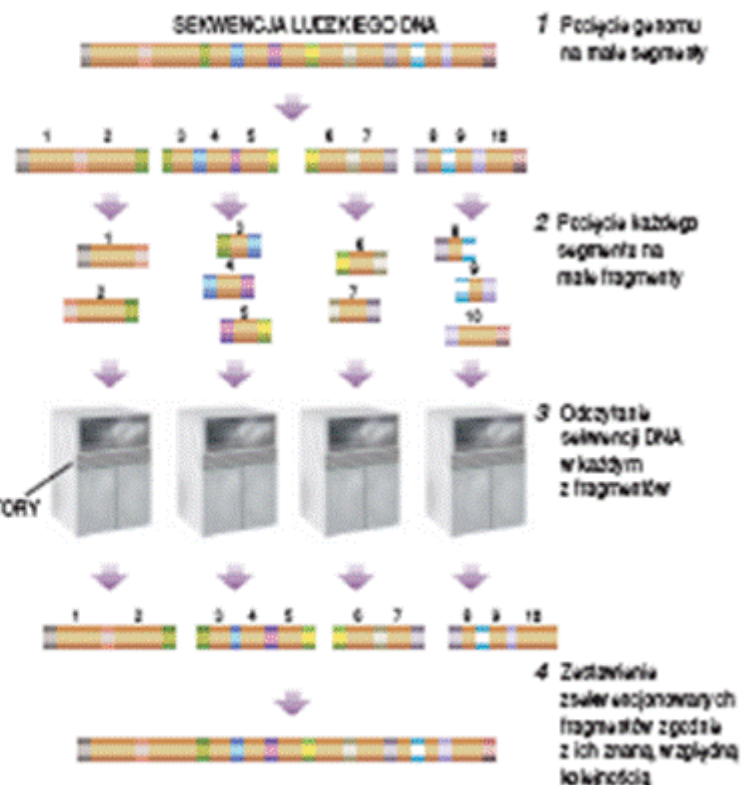
Sekwencjonowanie genomu

Dwie metody sekwencjonowania genomu

CELERA GENOMICS STRZAŁ NA SLEPO



PROJEKT POZNANIA LUDZKIEGO GENOMU METODA LOSOWEGO SEKWENCJONOWANIA



Całkowicie poznano genomy:

1. Kilkuset wirusów i wiroidów, plazmidów, organelli, bakterii, drożdży, rzodkiewnika, ryżu, nicienia oraz muszki owocowej i myszy
2. 2001 r. – ogłoszono rozszyfrowanie ludzkiego genomu (Celera i PPLG)



miliony danych do przetworzenia
w użyteczną wiedzę



bioinformatyka

Podstawowe zagadnienia bioinformatyki

1. katalogowanie informacji biologicznych (bazy danych, bazy danych sekwencji i wyszukiwanie sekwencji, anotacji, danych numerycznych w bazach danych)
2. analiza sekwencji DNA (składanie sekwencji, anotacja, wyszukiwanie sekwencji kodujących, regulatorowych i repetytywnych, motywów, markerów)
3. analiza sekwencji genomów, porównywanie genomów
4. ustalanie ewolucyjnych relacji pomiędzy zbiorami sekwencji / organizmów (drzewa filogenetyczne)
5. genotypowane (używane między innymi do wyszukiwania genów odpowiedzialnych za choroby genetyczne, w ustalaniu ojcostwa, kryminalistyce)

Podstawowe zagadnienia bioinformatyki

6. analiza ekspresji genów (głównie analiza danych z mikromacierzy)
7. analiza sekwencji białek, nazywana też proteomiką (porównywanie sekwencji, wyszukiwanie domen i motywów, przewidywanie własności fizyko-chemicznych, drugo- i trzecio-rzędowej struktury białka, lokalizacji w obrębie komórki, analiza danych z eksperymentów spektroskopowych)
8. katagolowanie funkcji genów/białek, analiza dróg metabolicznych (np metabolizm lipidów) oraz dróg sygnałowych (np od receptora na powierzchni komórki poprzez kaskadę kinaz do czynników transkrypcyjnych)

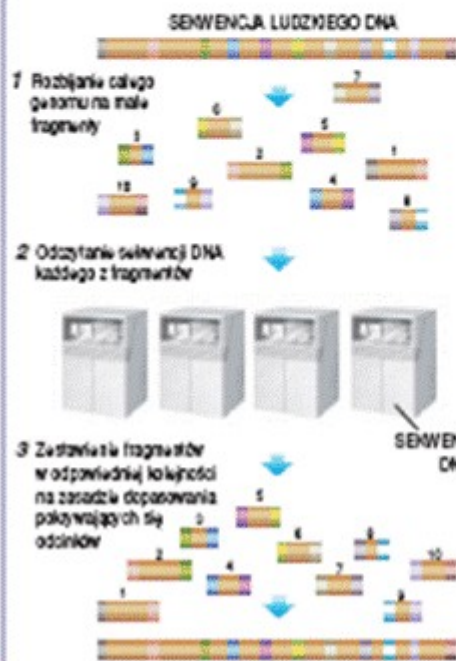
Podstawowe zagadnienia bioinformatyki

9. modelowanie układów biologicznych (np kinetyka szeregu reakcji enzymatycznych w komórce)
10. wirtualne dokowanie (ang. virtual docking) – np. używając trójwymiarowej struktury aktywnego centrum enzymu ("zamek" albo "kieszonka" ang. pocket) przeszukuje się w komputerze tysiące małych cząsteczek z których kilka-kilkanaście ('kluczy') będzie miało kształt mieszczący się w centrum aktywnym. Pierwszy krok w kierunku odkrywania nowych leków.
11. komputery DNA
12. morfometria / analiza obrazu

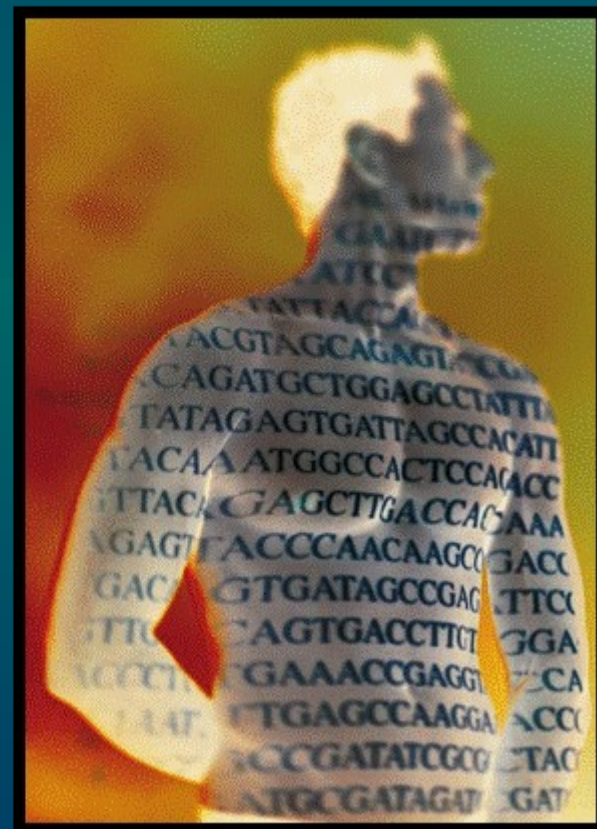
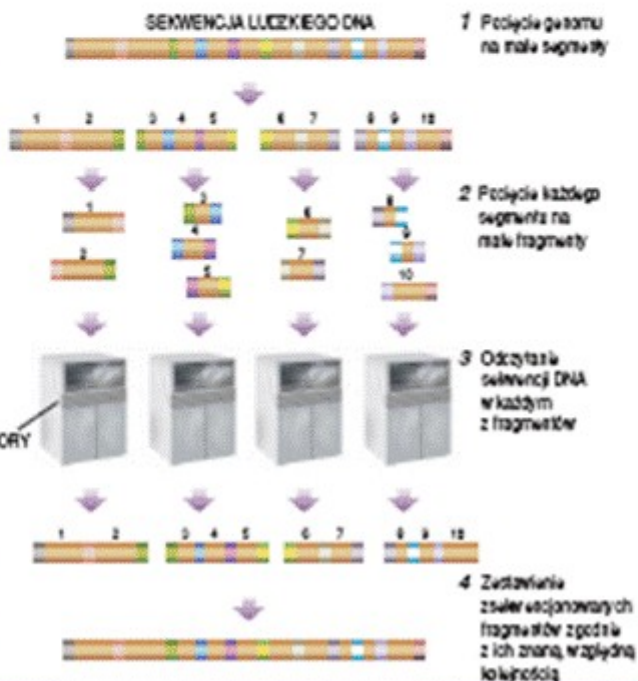
Sekwencjonowanie genomu

Dwie metody sekwencjonowania genomu

CELERA GENOMICS STRZAŁ NA SLEPO



PROJEKT POZNANIA LUDZKIEGO GENOMU METODA LOSOWEGO SEKWENCJONOWANIA



Bazy danych

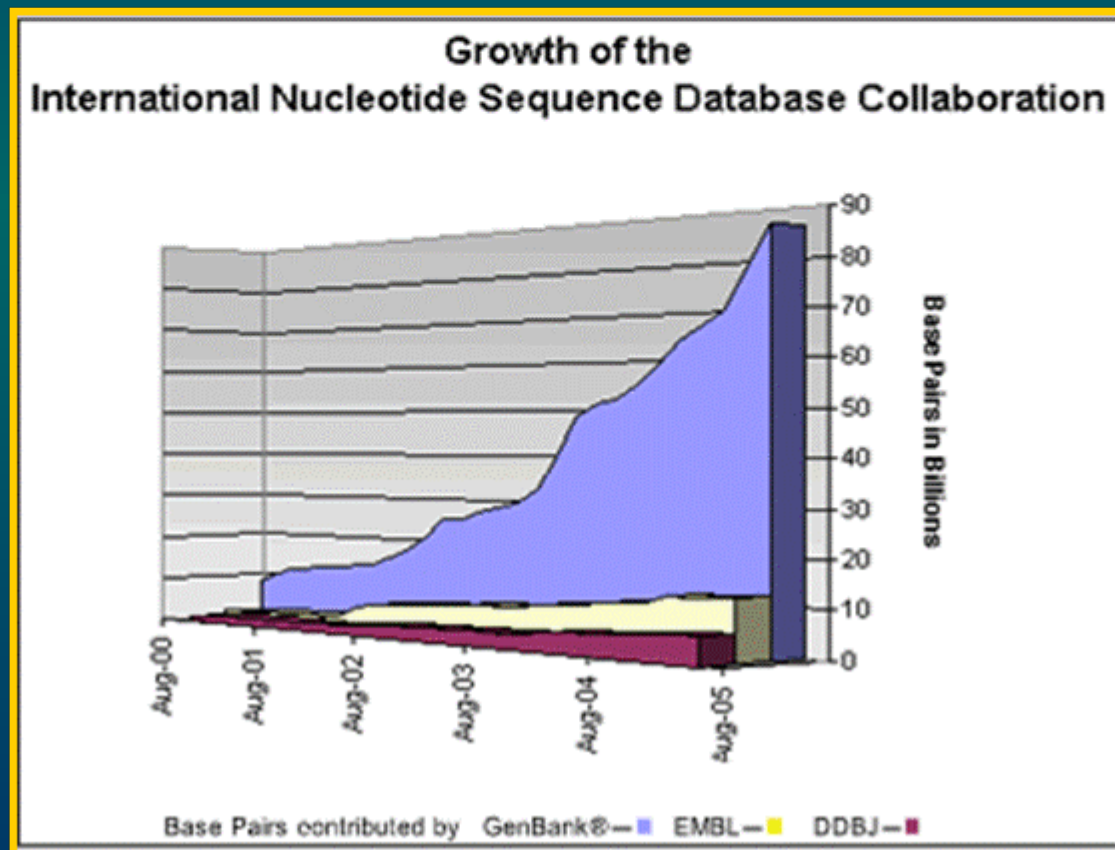
GenBank – największa baza danych genetycznych sekwencji nukleotydowych. Została stworzona w 1982 r. przez **National Center for Biotechnology Information** (National Institute of Health + National Library of Medicine), **Bethesda, USA**.

EMBL Nucleotide Sequence Database – baza danych genetycznych sekwencji nukleotydowych EBI (**European Bioinformatics Institute, Cambridge, UK**). Gromadzi dane od indywidualnych badaczy, dane z patentów i projektów badawczych.

DNA Data Bank of Japan (DDBJ) **Mishima, Japonia**

Bazy danych

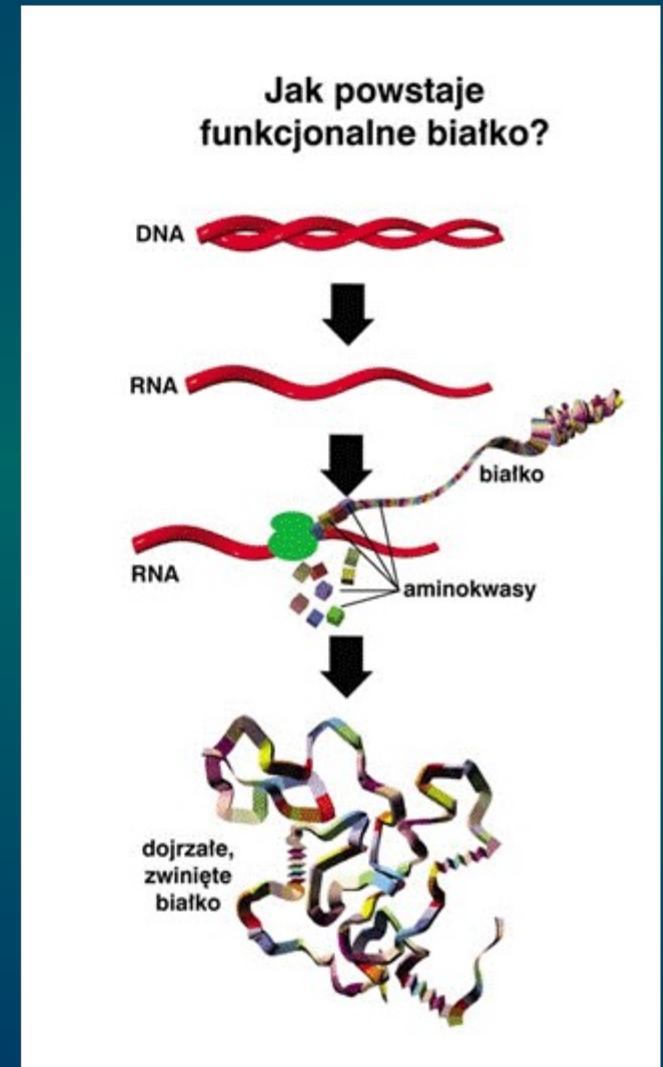
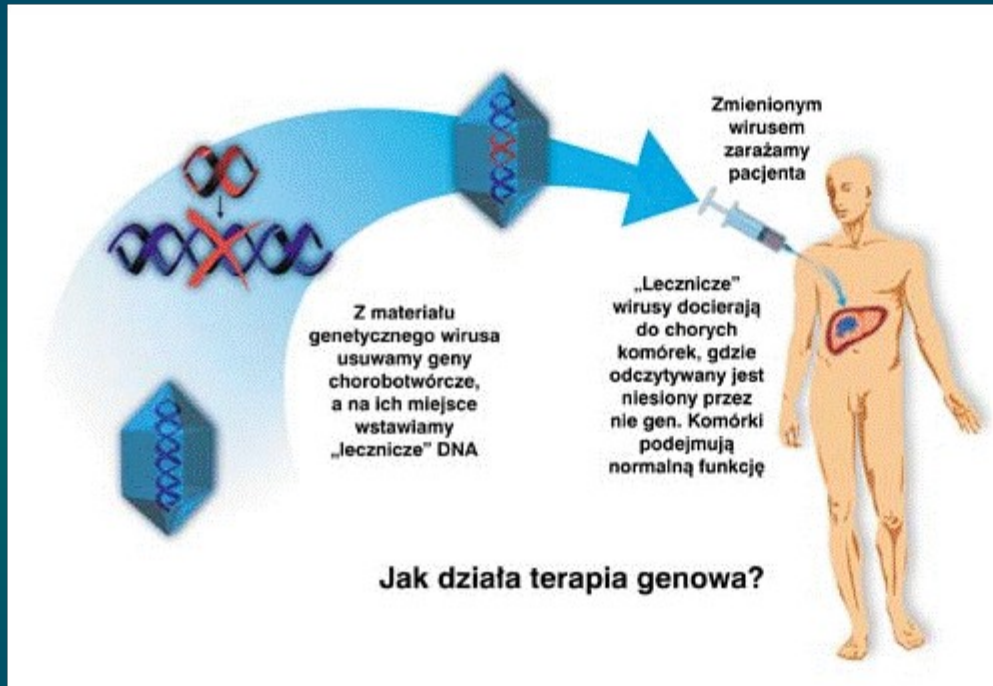
W 2005 roku liczba sekwencji znajdujących się w GenBank przekroczyła 100 miliardów par zasad azotowych). Jest dostępna dla każdego.



Instytucje zajmujące się bioinformatyką w Polsce

1. Pracownia Bioinformatyki i Inżynierii Białka w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej – przewidywanie i analiza struktury i funkcji białek
2. Pracownia Bioinformatyki, Wydział Biologii, Uniwersytet im. Adama Mickiewicza – analiza RNA
3. Pracownia Teorii Biopolimerów, Uniwersytet Warszawski – bioinformatyka, modelowanie białek
4. Zakład Bioinformatyki i Telemedycyny przy Collegium Medicum Uniwersytetu Jagiellońskiego w Krakowie
5. Centrum Bioinformatyczne ICM

Terapia genowa



Komputerowe projektowanie leków

Wspomaganie Komputerowe Projektowania Leków, nazywane jest w skrócie **CAAD**. System ten jest wykorzystywany między innymi przez GenBank. Jest to specjalistyczna dziedzina, która wykorzystuje metody komputerowe w procesie stymulacji interakcji między receptorem a lekiem. CAAD zależy w głównej mierze od zastosowanych narzędzi bioinformatycznych, aplikacji oraz wykorzystywanych baz danych. Istnieje związek pomiędzy badaniami stosowanymi w CADD, a badaniami w dziedzinie bioinformatyki.



LUDZKI CHROMOSOM 3

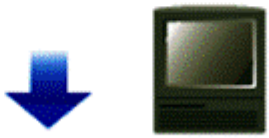


GEN *MLH1*
(na prążku 21.3)

1 WYZOLOWANIE SEKWENCJI LUDZKIEGO DNA

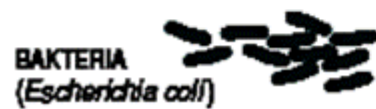
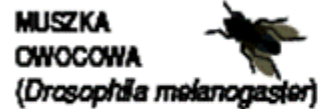
... G A G A A C T G T T T A G A T G C A A A A T C C A C A A G T ...

2 TRANSLACJA SEKWENCJI ZASAD W DNA
NA SEKWENCJĘ AMINOKWASÓW
(jednostek budulcowych białka) Z ZASTOSOWANIEM
PROGRAMU KOMPUTEROWEGO



... E N C L D A K S T S ... SEKWENCJA AMINOKWASÓW CZŁOWIEKA

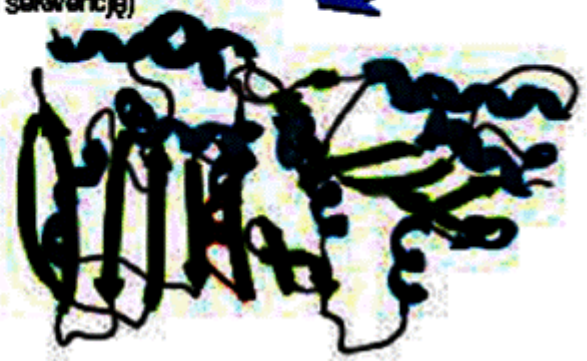
3 POSZUKIWANIE SEKWENCJI HOMOLOGICZNYCH
W BAZACH DANYCH, ZAWIERAJĄCYCH
INFORMACJĘ O BIAŁKACH ORGANIZMÓW
MODELOWYCH (zielony oznacza duże różnice,
pomarańczowy – mniejsze)



CZŁOWIEK	...	E	N	C	L	D	A	K	S	T	S	...
MUSZKA OWOCOWA (<i>Drosophila melanogaster</i>)	...	E	N	S	L	D	A	Q	S	T	H	...
NICIEŃ (<i>Caenorhabditis elegans</i>)	...	E	N	S	L	D	A	G	A	T	E	...
DROZDZE PIEKARNICZE (<i>Saccharomyces cerevisiae</i>)	...	E	N	S	I	D	A	N	A	T	M	...
BAKTERIA (<i>Escherichia coli</i>)	...	E	N	S	L	D	A	G	A	T	R	...

4 MODEL LUDZKIEGO BIAŁKA
STWORZONY W OPACIU
O ZNANĄ STRUKTURĘ
PODOBNEJ PROTEINY
WYSTĘPUJĄCEJ W ORGANIZMIE
MODELOWYM (czerwonobrazowy
obszar jest kodowany przez
ukazaną tu sekwencję)

5 ODNALEZIENIE ZWIĄZKU,
KTÓRY PRZYŁĄCZA SIĘ
DO SKONSTRUOWANEGO
WIRTUALNIE BIAŁKA

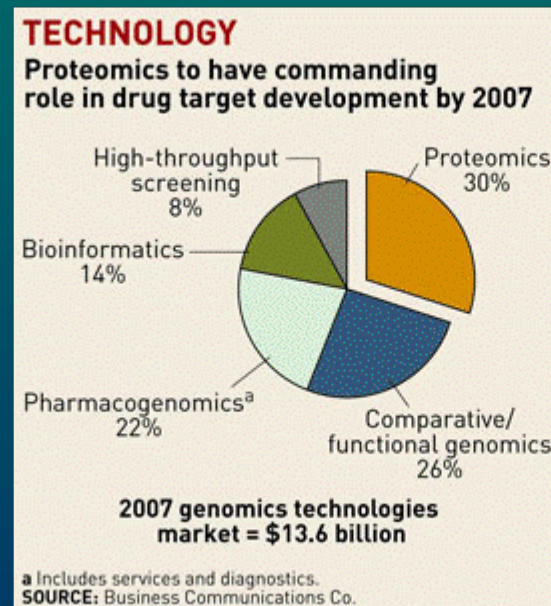


Komputerowe projektowanie leków

Wśród korzyści wynikających ze stosowania metody CADD można wymienić redukcję kosztów. Jest to jeden z najważniejszych czynników, który wpływa na zastosowanie metod komputerowych w przemyśle farmaceutycznym. Kolejną zaletą jest wybór najlepszego z możliwych medykamentów. Stosownie metod bioinformatycznych w tej dziedzinie pozwala odnaleźć możliwe interakcje, co bez wątpienia jest bardzo dużym atutem. Często, gdy naukowcy tworzą przy pomocy komputera związek, składający się z dwóch molekuł, który jest potencjalnym lekiem, okazuje się, że w momencie symulacji pojawiają się nowe koncepcje, dotyczące ulepszenia takiego specyfiku.

Komputerowe projektowanie leków

Metoda CADD razem z bioinformatyką tworzą bardzo ważną i potężną kombinację w przypadku badań nad nowymi lekami, a jedynym problemem jest znalezienie odpowiednich, doświadczonych ludzi, którzy będą wprowadzać udoskonalenia a także tworzyć nowe programy dla przemysłu farmaceutycznego



Bioinformatyka – koszty

Określenie sekwencji ludzkiego genomu trwało ponad 10 lat i kosztowało 300 mln \$.

2007 – Watson – ujawnił pełny swój kod genetyczny (CD).

„Projekt Jim” trwał 67 dni i kosztował 1 mln \$.

Powiązanie niektórych chorób z defektami genetycznymi (Cancer Genom Atlas: National Cancer Institute i National Human Genome Research Institute) – opracowanie szybszych i tańszych metod sekwencjonowania

„Wąskie gardło” – metoda Sanger: cięcie DNA, amplifikacja przez wielokrotne kopiowanie, sekwencjonowanie na wielu aparatach równocześnie (mimo iż stosowana od 30 lat – wciąż badana na poziomie akademickim)

Bioinformatyka – koszty

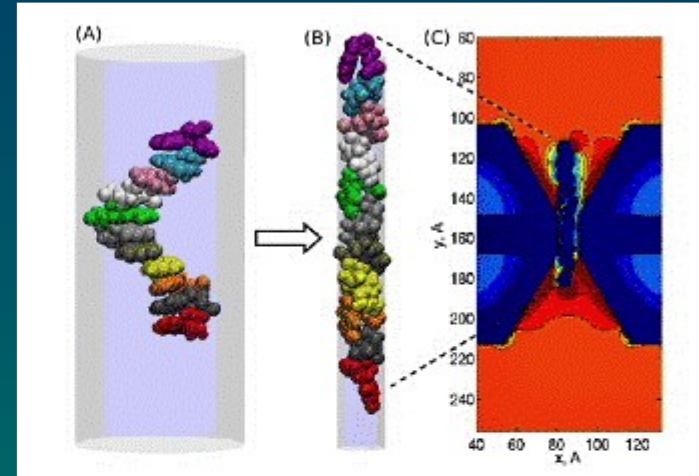
„Wąskie gardło” – metoda Sanger: cięcie DNA, amplifikacja przez wielokrotne kopiowanie, sekwencjonowanie na wielu aparatach równocześnie (mimo iż stosowana od 30 lat – wciąż badana na poziomie akademickim)

Zastosowanie mikrocieczy, elektroforezy kapilarnej, polimerazowe cięcie nici DNA

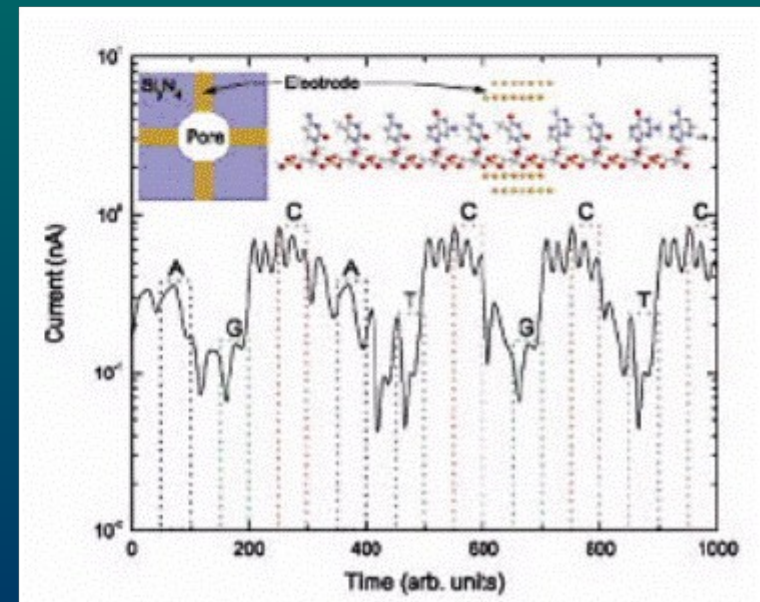
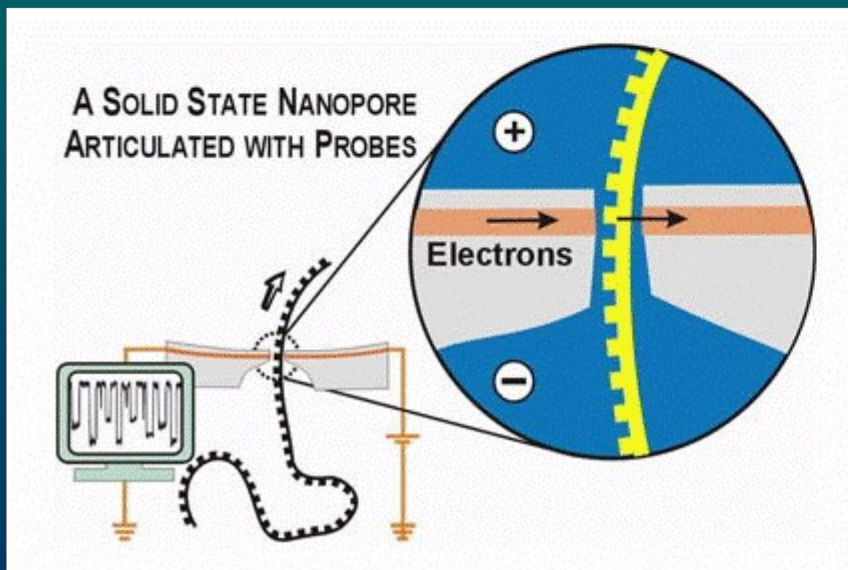
Marzenie: „przeciąganie” pojedynczych cząsteczek DNA przez nanopory i odczyt sekwencji przy pomocy nanoelektroniki

Bioinformatyka - koszty

Pomiar momentu dipolowego
(LeBurton, Univ. Illinois)



Zakłócenia natężenia prądu
(Branton, Harvard Univ.)



Bioinformatyka – koszty

Problemy w realizacji:

Szumy w sygnale elektrycznym powodowane przez skręcanie i zginanie cząsteczki DNA

DNA przechodzi przez por za szybko (poszczególne zasady przesuwają się co 30 ns)

Stosunkowo małe różnice w budowie chemicznej zasad – zakłócenia pola elektrycznego bardzo podobne

(Massimiliano Di Silva, San Diego)

Co daje nam poznanie genomu?

- ✂ Zrozumienie samego zjawiska życia
- ✂ Jakie procesy są konieczne, aby ono powstało?
- ✂ Jak przebiegała ewolucja organizmów żywych?
- ✂ Jak była chronologia zaludniania poszczególnych części Ziemi?
- ✂ Czy możliwe jest stworzenie syntetycznej formy życia?
- ✂ Czy istnieje zestaw białek wspólnych dla wszystkich organizmów?

Co daje nam poznanie genomu?

1. Zrozumienie samego zjawiska życia
2. Jakie procesy są konieczne, aby ono powstało?
3. Jak przebiegała ewolucja organizmów żywych?
4. Jak była chronologia zaludniania poszczególnych części Ziemi?
5. Czy możliwe jest stworzenie syntetycznej formy życia?
6. Czy istnieje zestaw białek wspólnych dla wszystkich organizmów?

Dziękuję
za
uwagę

