

wykład 2

Zadania bioinformatyki

dr Jacek Śmietański

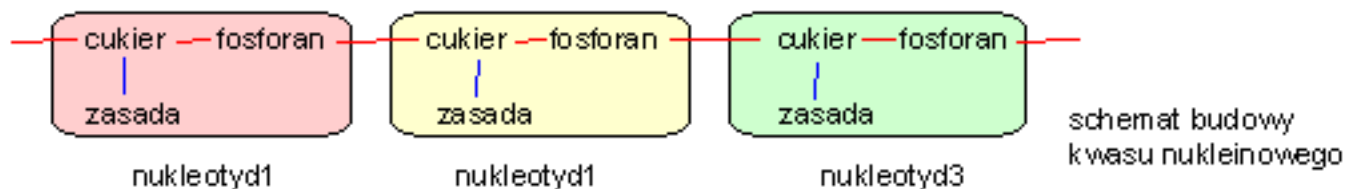
jacek.smietanski@ii.uj.edu.pl

<http://jaceksmietanski.net>

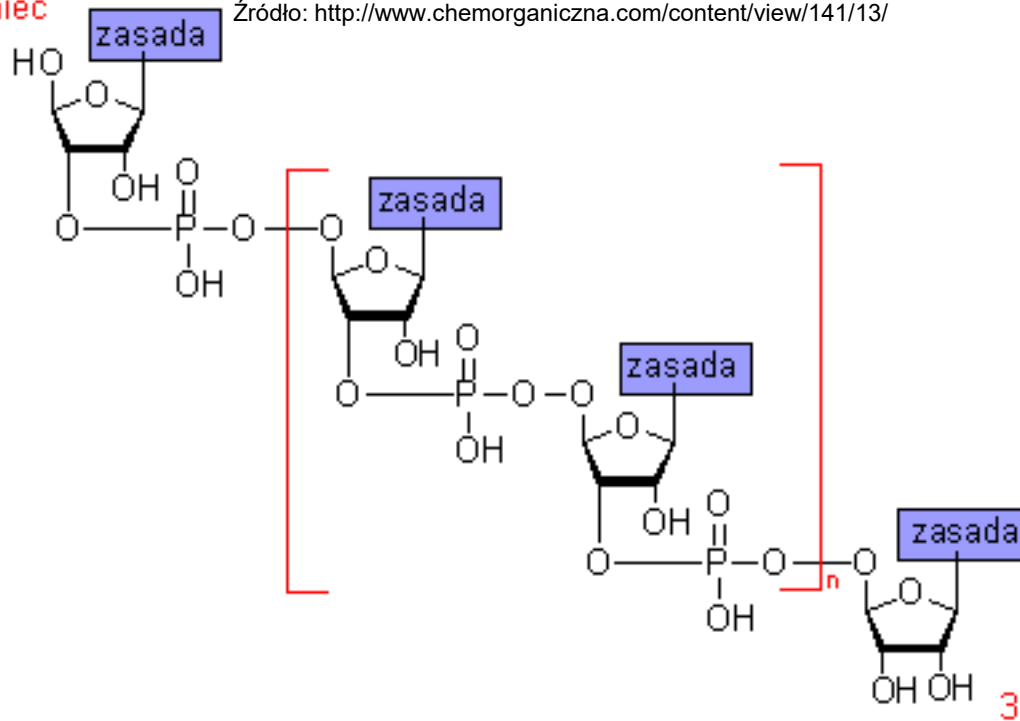
1. Odrobina biologii
2. Najważniejsze zadanie bioinformatyki
3. Narzędzia programistyczne (biblioteki Bio*)

Odrobina biologii

Kwasy nukleinowe (DNA i RNA) - budowa



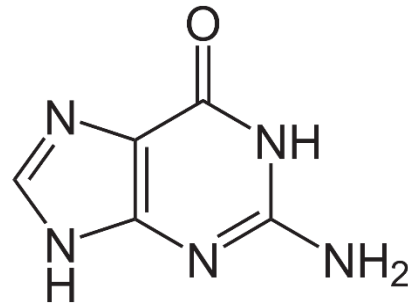
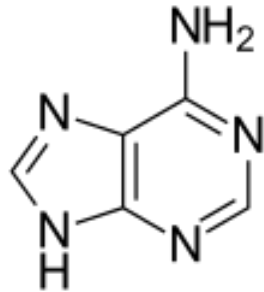
5' koniec



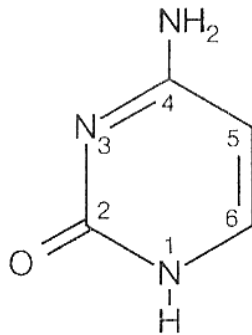
DNA od RNA odróżnia budowa rdzenia cukrowego oraz rodzaje zasad (azotowych).

Obrazowo: każda cząsteczka jest „nitką” (rdzeń cukrowo-fosforanowy), na którą nawleczone są koraliki czterech różnych rodzajów (zasady azotowe).

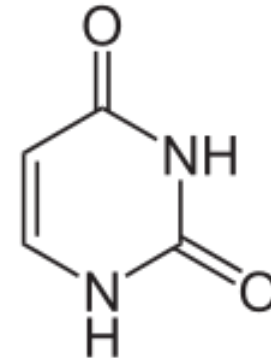
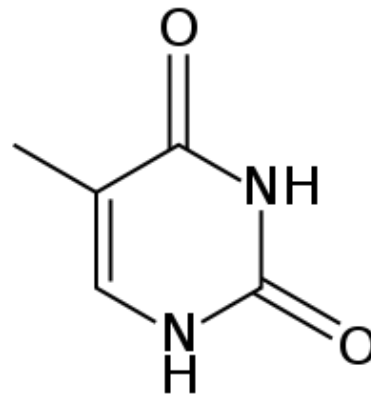
Zasady purynowe: adenina (A), guanina (G)



Zasady pirymidynowe: cytozyna (C), tymina (T), uracyl (U)



Cytozyna (C)

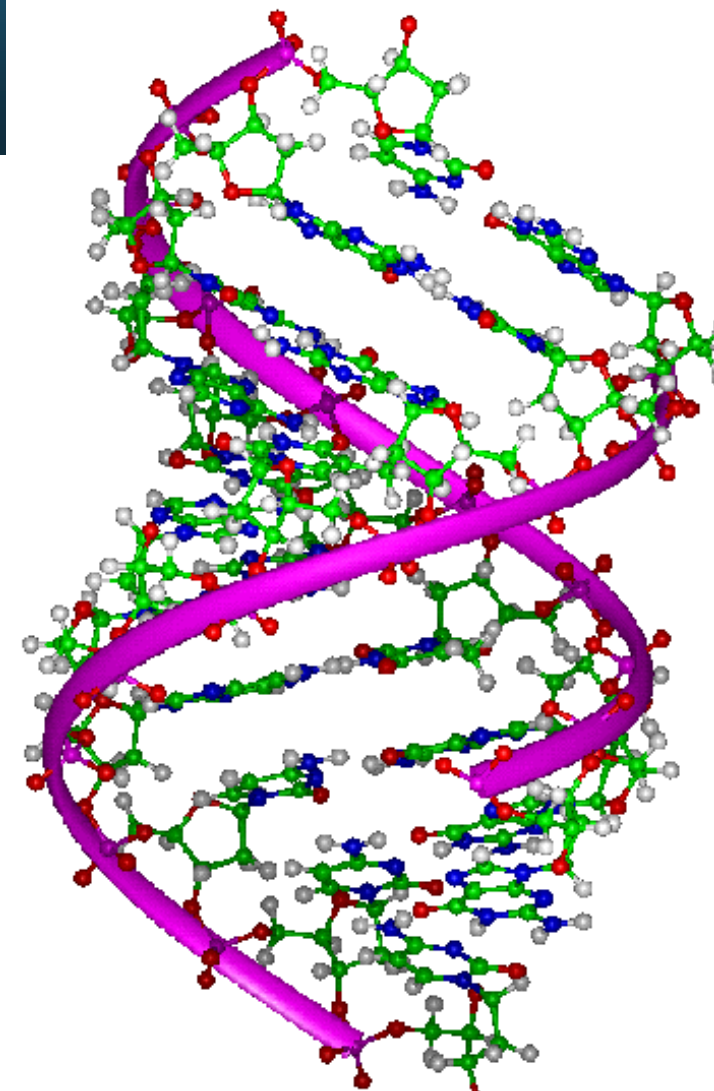
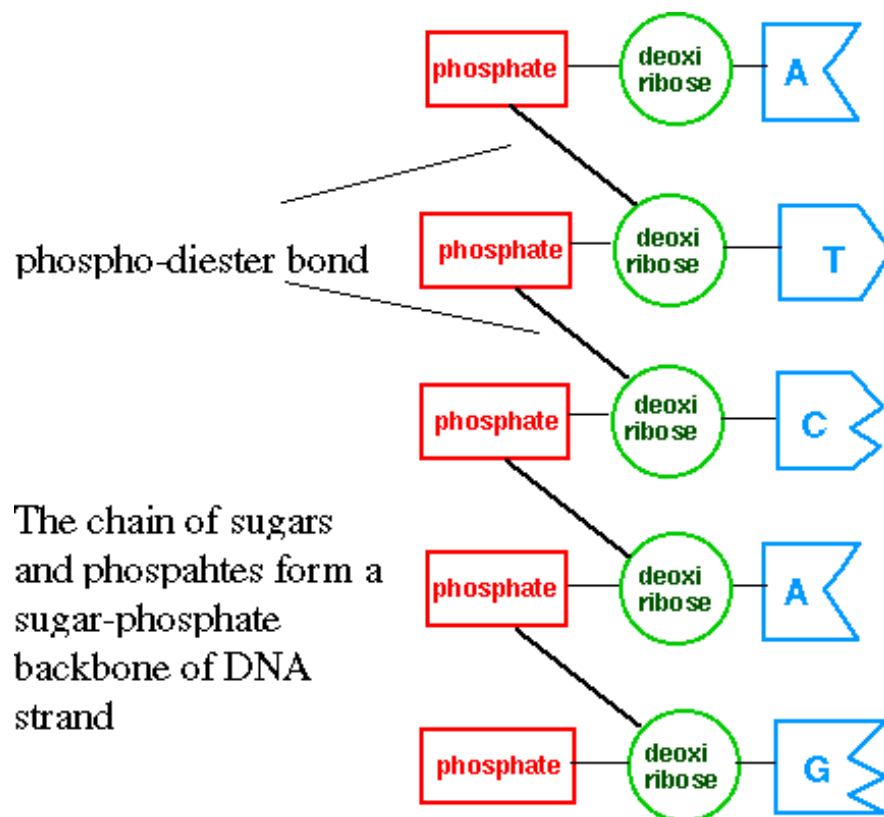


DNA - kwas deoksyrybonukleinowy

Pojedynczy element nici:

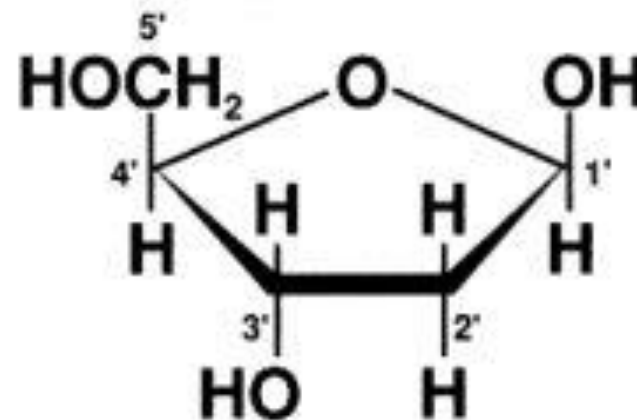
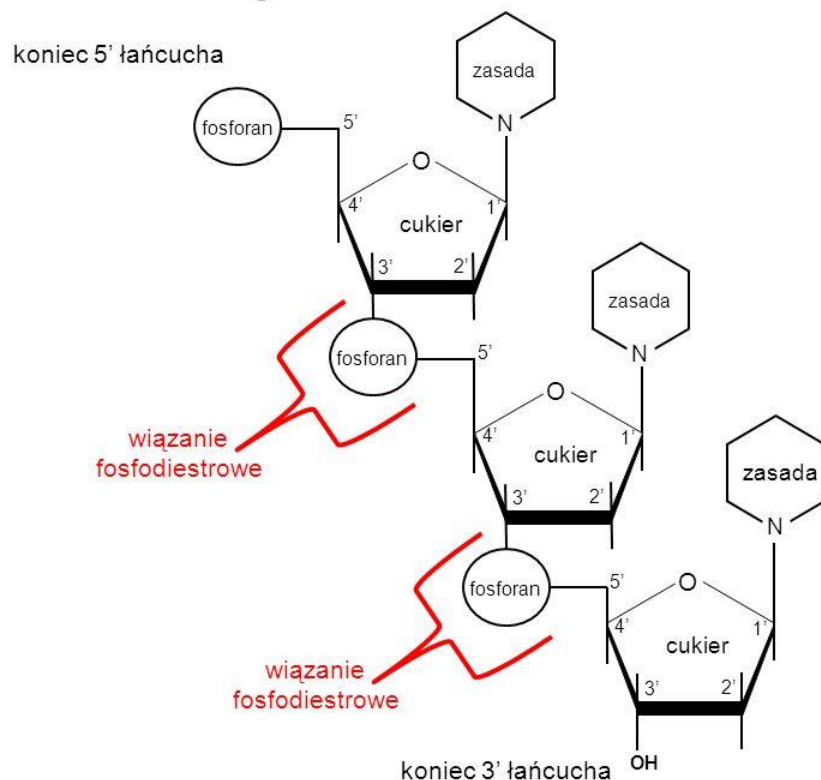
deoksyryboza + fosforan + zasada azotowa (A,T,G,C)

DNA tworzy strukturę helisy, złożoną z dwóch komplementarnych do siebie nici.



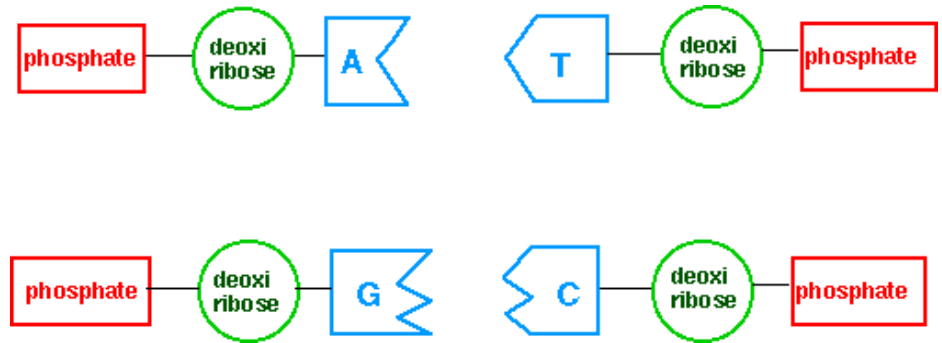
Nić DNA syntetyzowana jest w kierunku $5' \rightarrow 3'$. W tej też kolejności zapisywane są zasady azotowe.

Fragment łańcucha kwasu nukleinowego

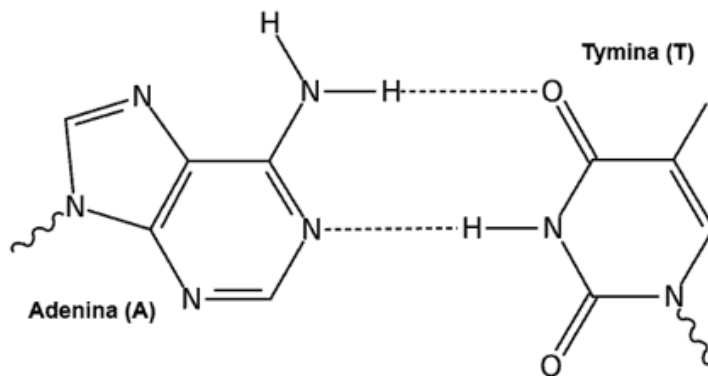


DNA – zasada komplementarności

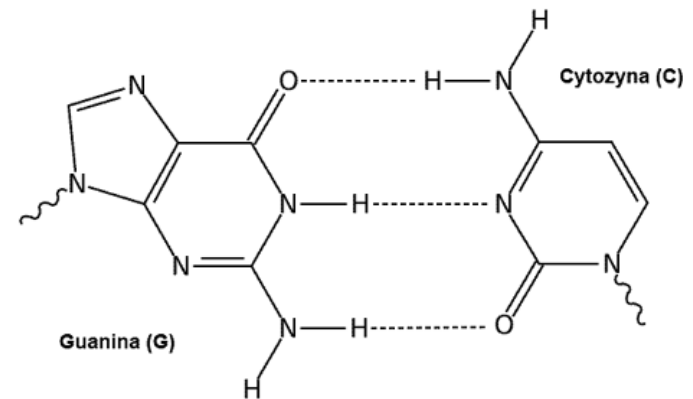
A – T
G – C



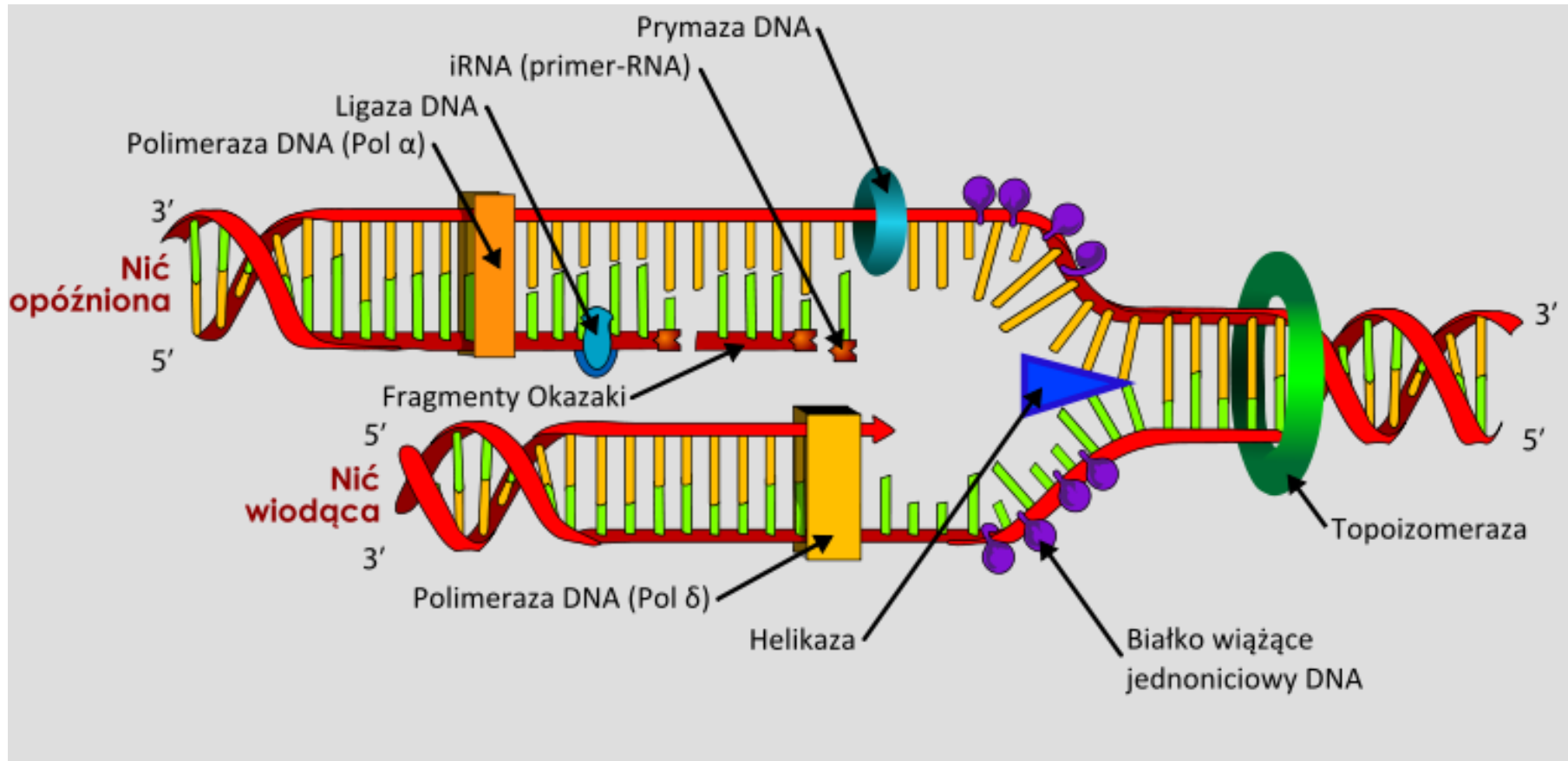
Każdej zasadzie na jednej nici odpowiada komplementarna jej zasada na drugiej nici.



(c) ebiolog.pl

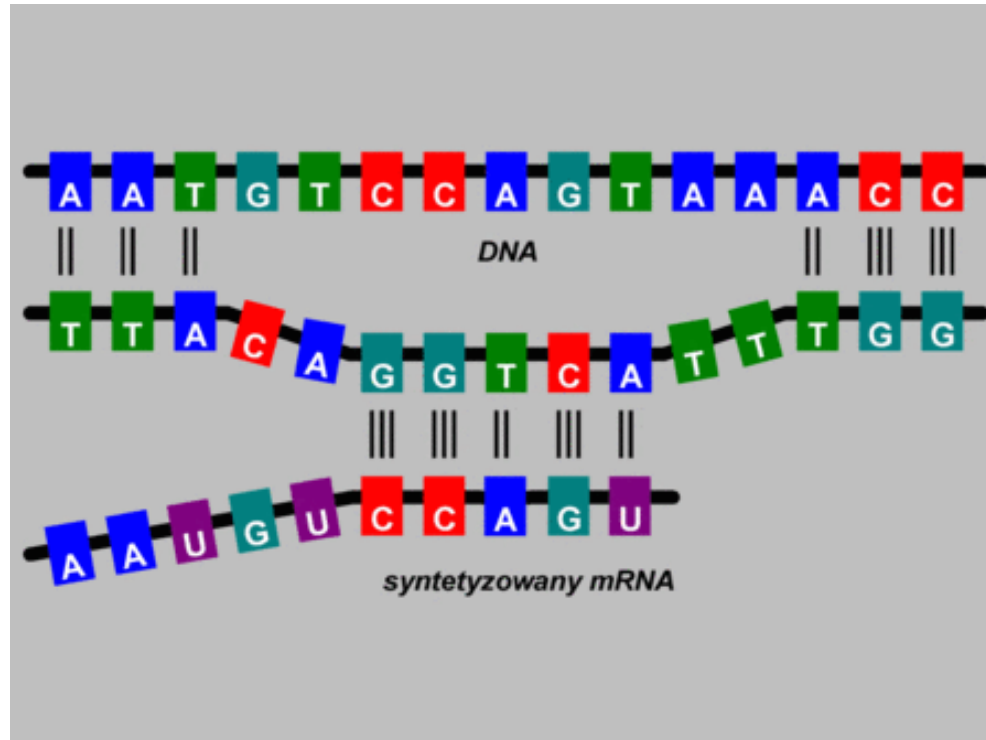


(c) ebiolog.pl



Tworzenie kopii nici DNA (podczas podziału komórki).
Proces bardzo dokładny.

Synteza RNA na matrycy DNA (zachowana jest zasada komplementarności)



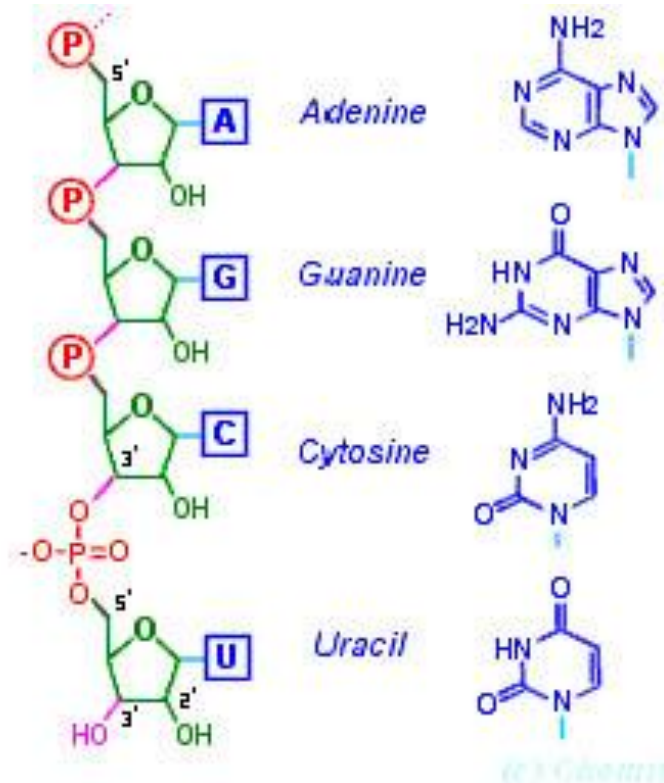
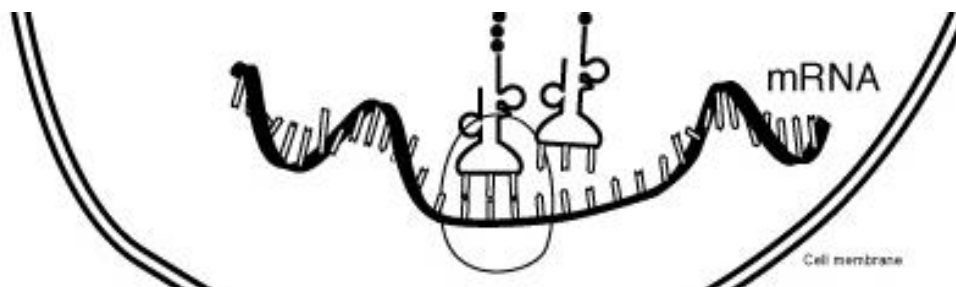
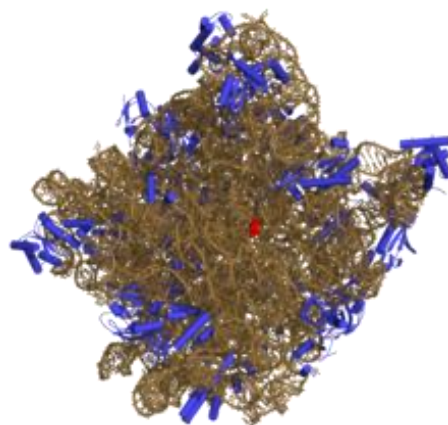
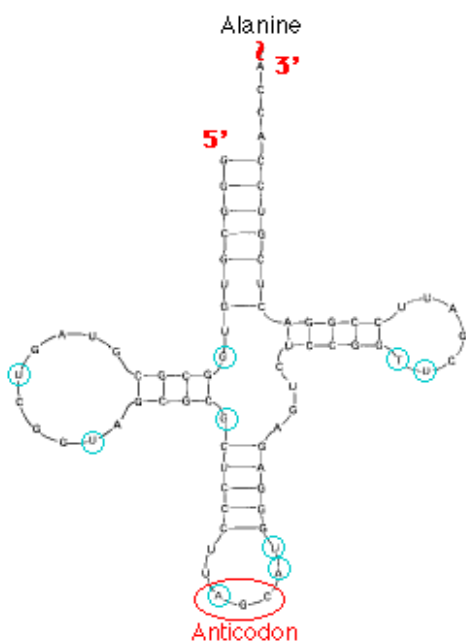
A	–	U
C	–	G
G	–	C
T	–	A

W RNA zamiast tyminy (T) jest uracyl (U).
Zakres błędów: 1 pomyłka na 10^4 – 10^5 nukleotydów.

Dlaczego proces ten może być mniej dokładny niż replikacja?

Zasady azotowe wchodzące w skład nici RNA:

- **A** (adenina)
- **G** (guanina)
- **C** (cytozyna)
- **U** (uracyl)



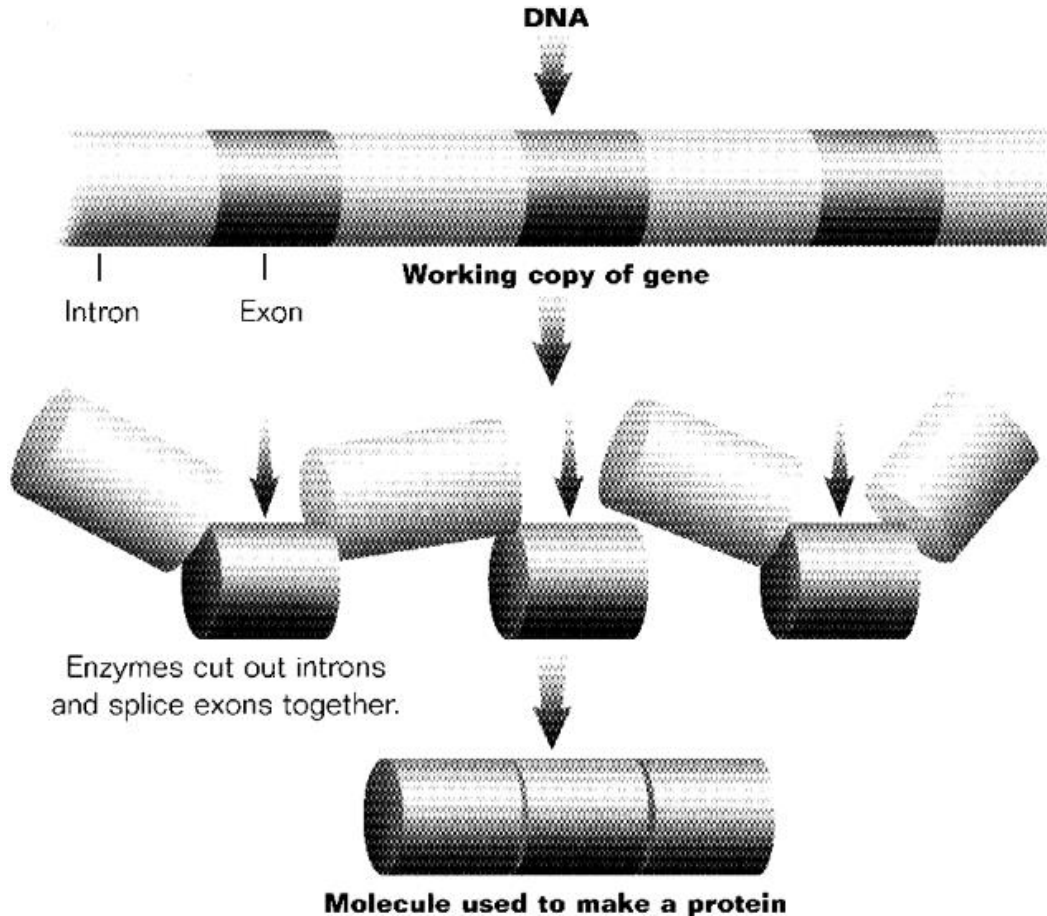
mRNA

matrycowy (informacyjny), nośnik przepisanej z DNA informacji o sekwencji aminokwasów w białku. Ma cechy umożliwiające przyłączanie się do rybosomów i udział w syntezie białka. Wielkość zależna od wielkości kodowanego polipeptydu. Zróżnicowana trwałość, raczej mało stabilny

ncRNA

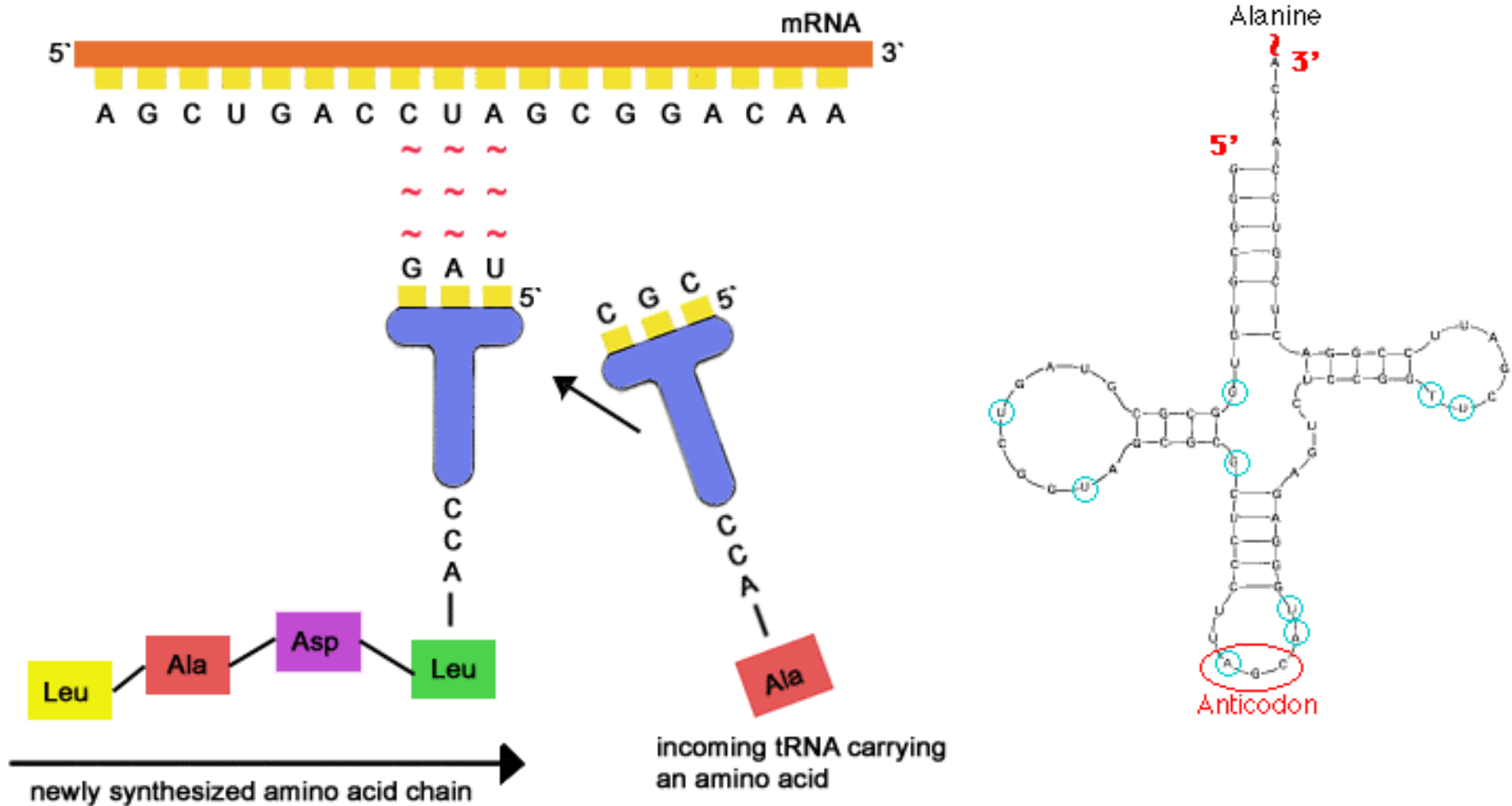
biorą udział w wielu procesach komórkowych, jak: regulacja transkrypcji, replikacji DNA, obróbki i modyfikacji innych cząsteczek RNA (transkryptów), np.:

- **rRNA** (rybosomowy, tworzy (wraz z białkami) rybosomy. Jeden z rRNA jest katalizatorem tworzenia wiązania peptydowego (rybozymbem). Różne rodzaje i wielkość (120-4700 zasad). rRNA eukariontów i prokariotów zasadniczo się różnią. Długożyjący (stabilny).
- **tRNA** (transportujący, mały (65-110 nt), przenosi zaktywowane aminokwasy do rybosomu)
- **snRNA** (małe jądrowe RNA) biorą udział w usuwaniu intronów i łączeniu egzonów;
- **miRNA** (mikro RNA) hamują translację;
- **siRNA** (małe interferencyjne RNA) ułatwiają degradację mRNA;
- ...



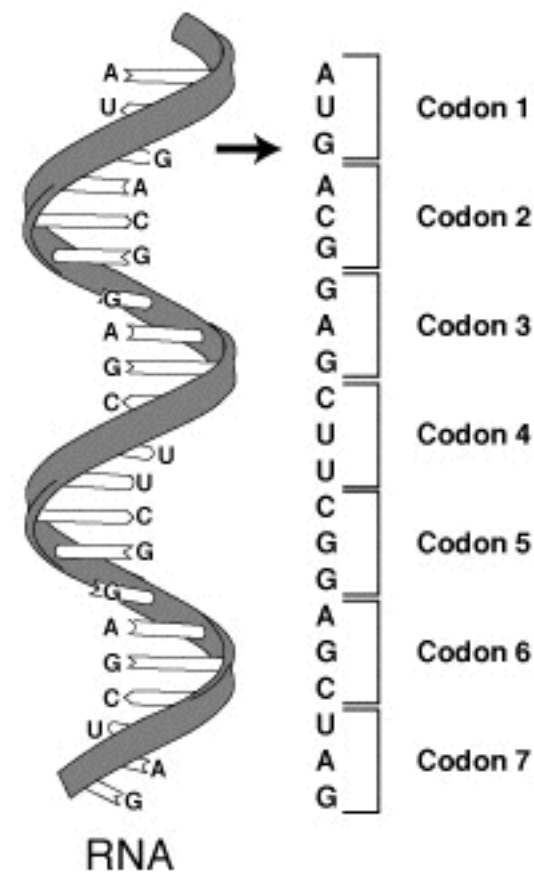
Proces usuwania intronów (fragmentów niekodujących) z pierwotnego transkryptu.

Synteza białka na matrycy dojrzałego mRNA.



Źródło: <http://library.thinkquest.org/C004535/media/translation.gif>

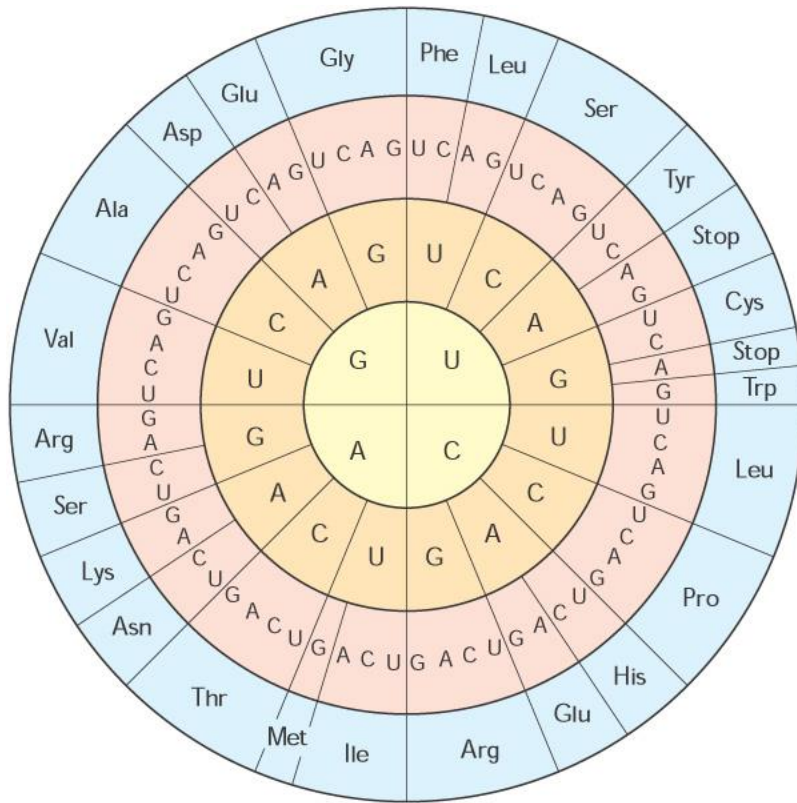
Każda trójka zasad koduje jeden określony aminokwas lub sygnał zakończenia translacji.



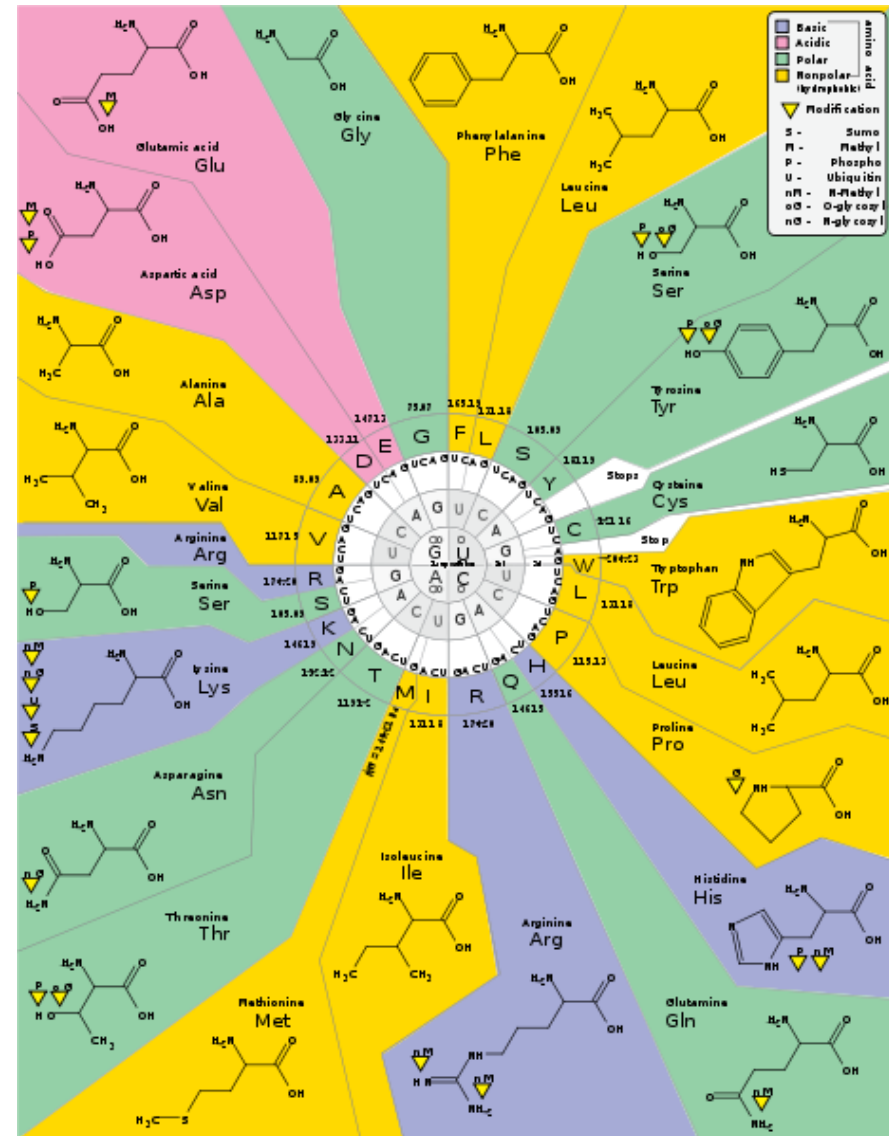
Ile jest różnych kodonów?

First Letter	Second Letter				Third Letter
	U	C	A	G	
U	phenylalanine	serine	tyrosine	cysteine	U
	phenylalanine	serine	tyrosine	cysteine	C
	leucine	serine	stop	stop	A
	leucine	serine	stop	tryptophan	G
C	leucine	proline	histidine	arginine	U
	leucine	proline	histidine	arginine	C
	leucine	proline	glutamine	arginine	A
	leucine	proline	glutamine	arginine	G
A	isoleucine	threonine	asparagine	serine	U
	isoleucine	threonine	asparagine	serine	C
	isoleucine	threonine	lysine	arginine	A
	(start) methionine	threonine	lysine	arginine	G
G	valine	alanine	aspartate	glycine	U
	valine	alanine	aspartate	glycine	C
	valine	alanine	glutamate	glycine	A
	valine	alanine	glutamate	glycine	G

Kod genetyczny (2)

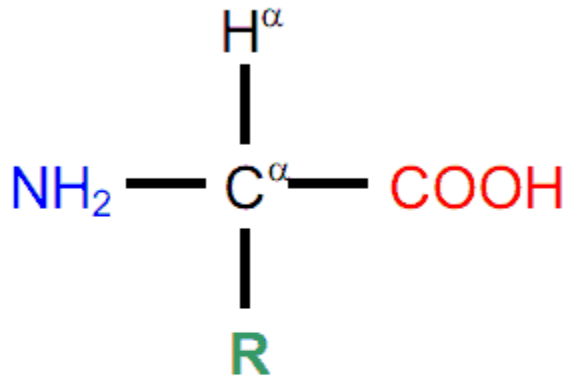


Zastanów się:
Jakie konsekwencje niesie za sobą
redundancja kodu genetycznego?

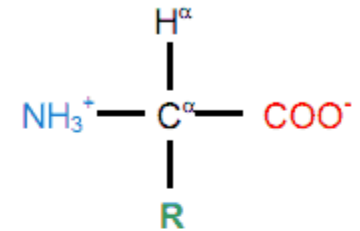


Źródło: <http://en.wikipedia.org/wiki/File:GeneticCode21-version-2.svg>

Ogólna budowa aminokwasów:



- w neutralnym pH



grupa aminowa - NH_2

grupa karboksylowa - COOH

Aminokwasy białkowe

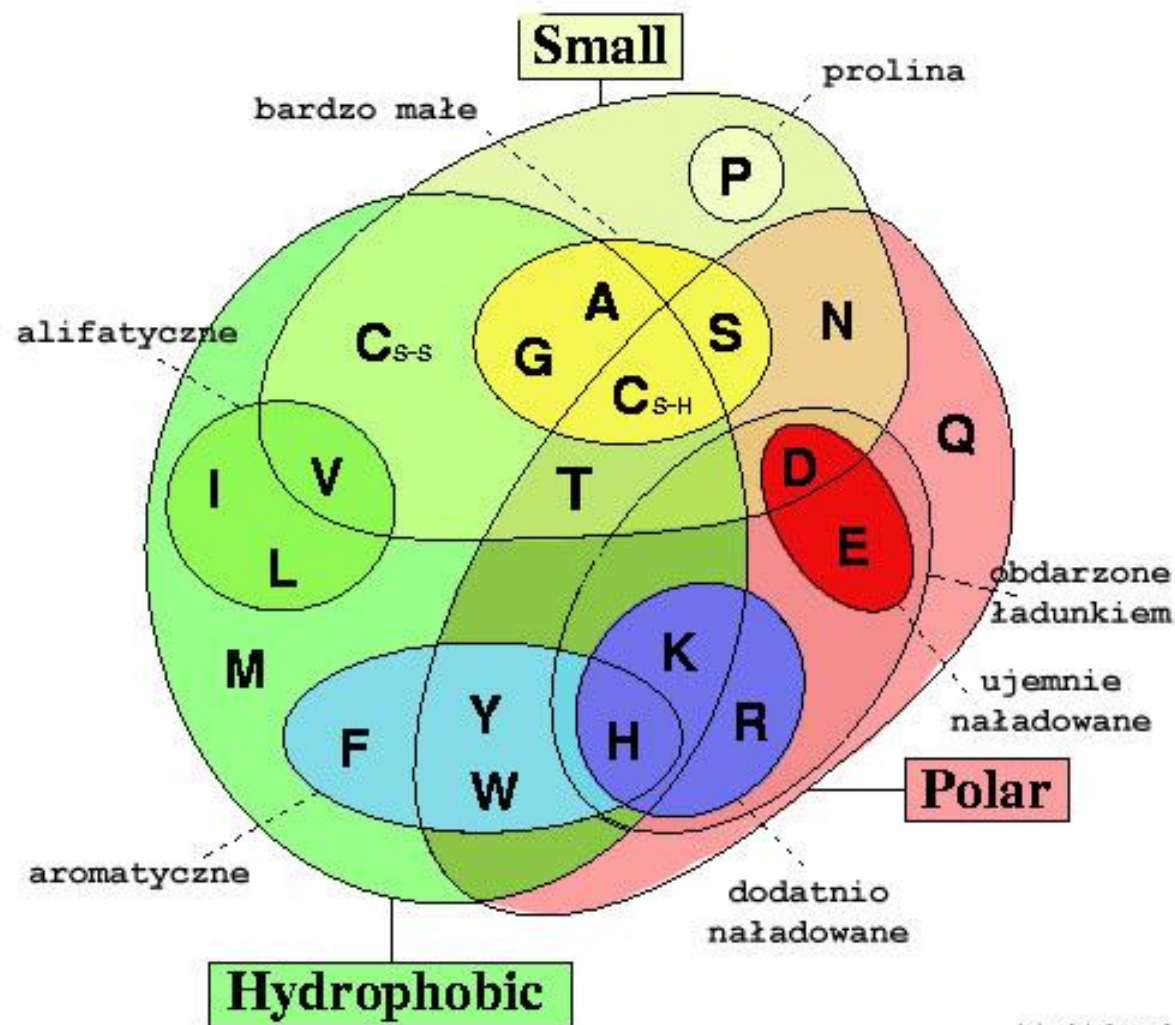
nazwa	symbol	skrót
alanina	A	Ala
arginina	R	Arg
asparagina	N	Asn
kw.asparaginowy	D	Asp
cysteina	C	Cys
glutamina	Q	Gln
kw.glutaminowy	E	Glu
glicyna	G	Gly
histydyna	H	His
izoleucyna	I	Ile

nazwa	symbol	skrót
leucyna	L	Leu
lizyna	K	Lys
metionina	M	Met
fenyloalanina	F	Phe
prolina	P	Pro
seryna	S	Ser
treonina	T	Thr
tryptofan	W	Trp
tyrozyna	Y	Tyr
walina	V	Val



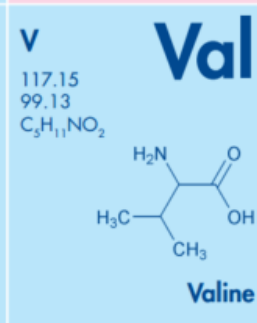
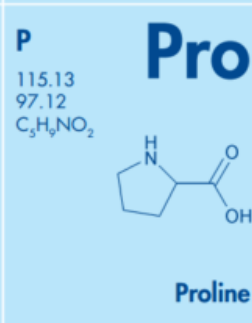
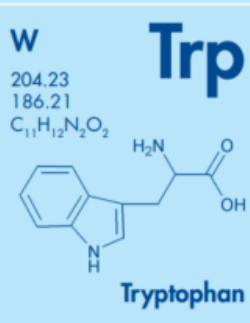
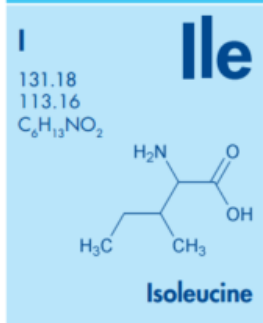
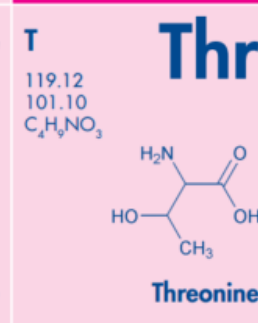
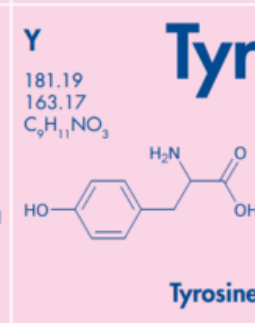
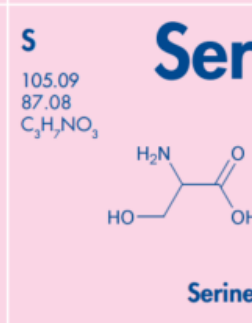
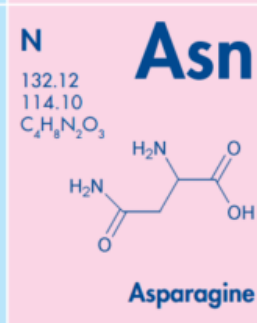
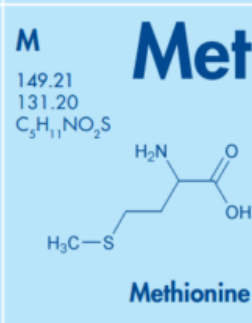
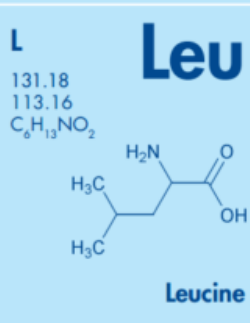
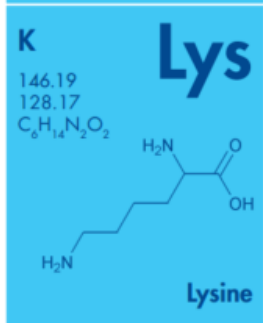
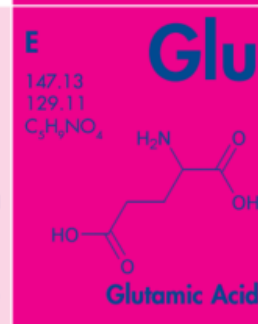
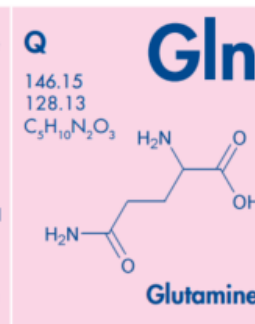
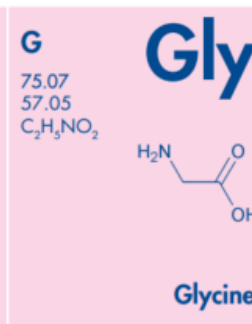
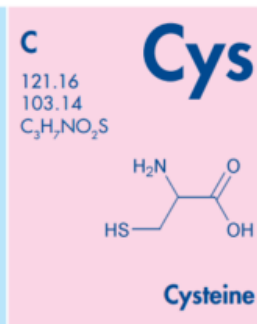
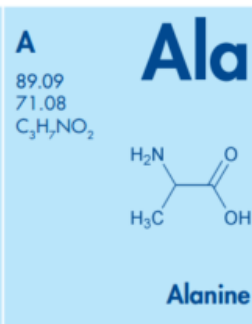
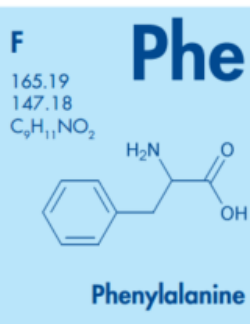
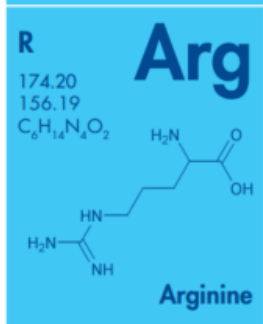
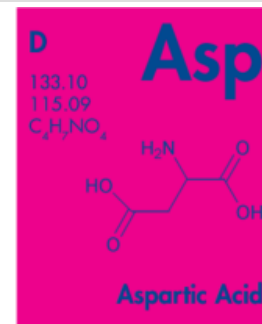
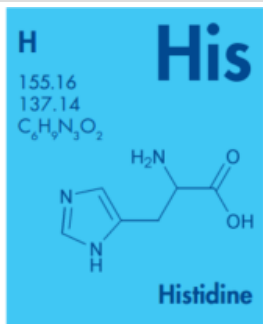
Np. wielkość,
ładunek,
hydrofobowość,
aromatyczność, ...

Powyższe właściwości
można zilustrować
na diagramie
Venna

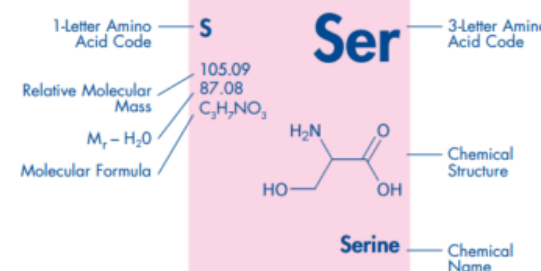


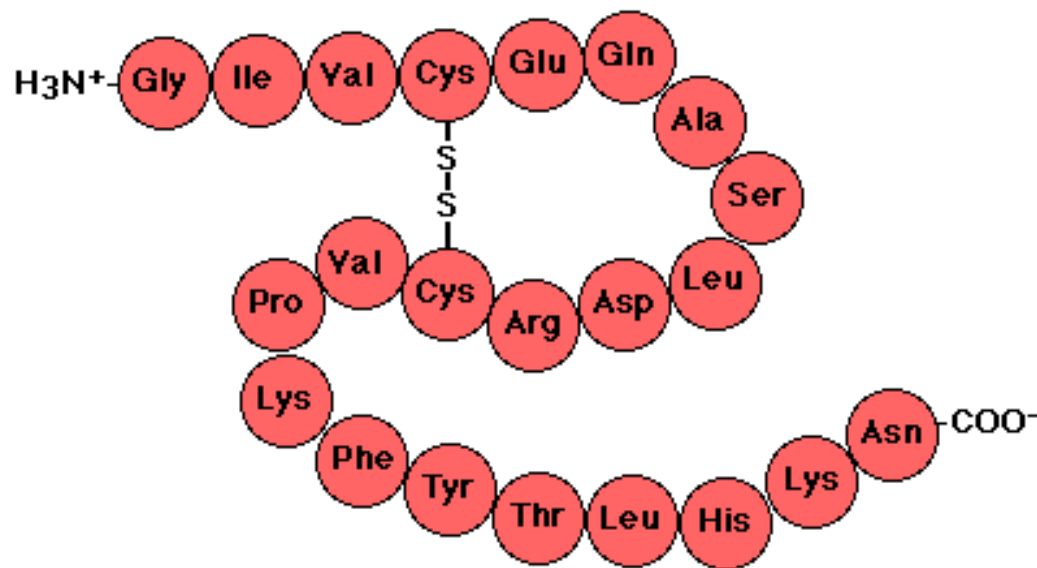
Periodic Chart of Amino Acids

www.bachem.com



Basic
Non-polar (hydrophobic)
Polar, uncharged
Acidic



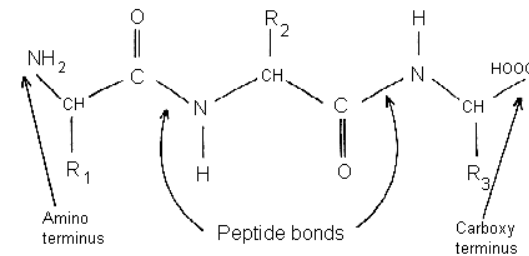


N-terminus

C-terminus



Podstawowa jednostka budulcowa i funkcjonalna organizmu. Ciąg aminokwasów połączonych wiązaniami peptydowymi.



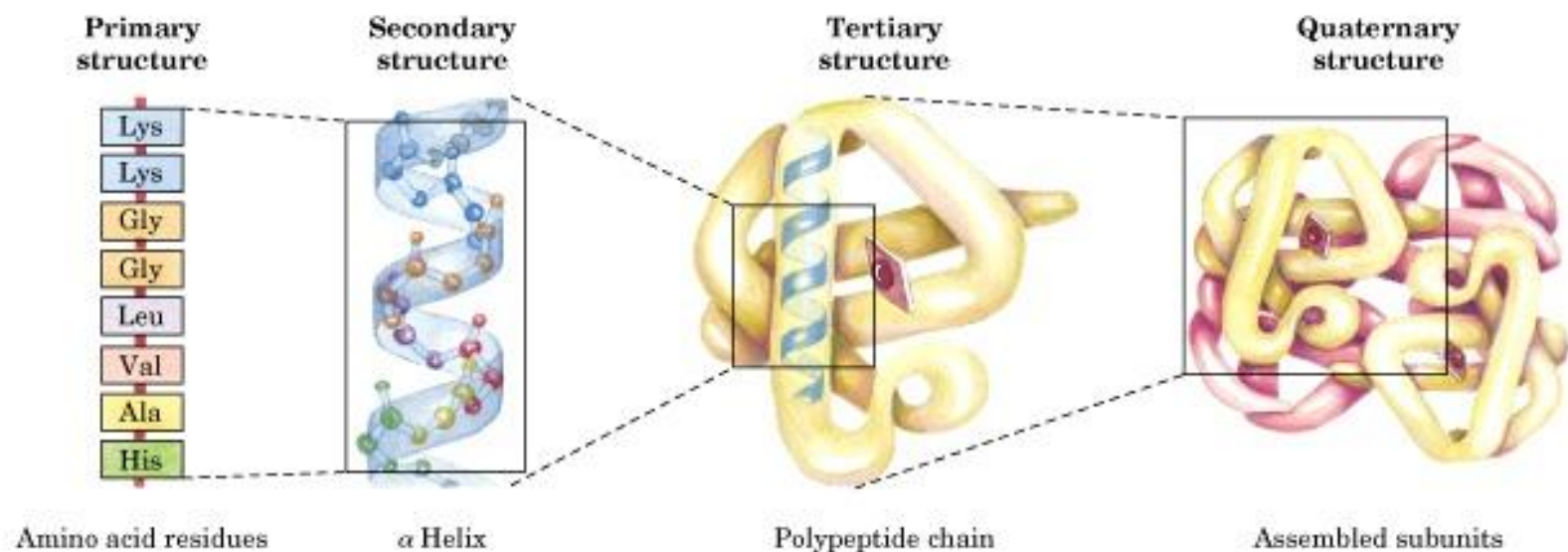
Poziomy przestrzennej organizacji białek:

I rządowa – liniowa sekwencja aminokwasów

II rządowa – opisuje lokalne pofałdowanie (α -helisy, β -kartki)

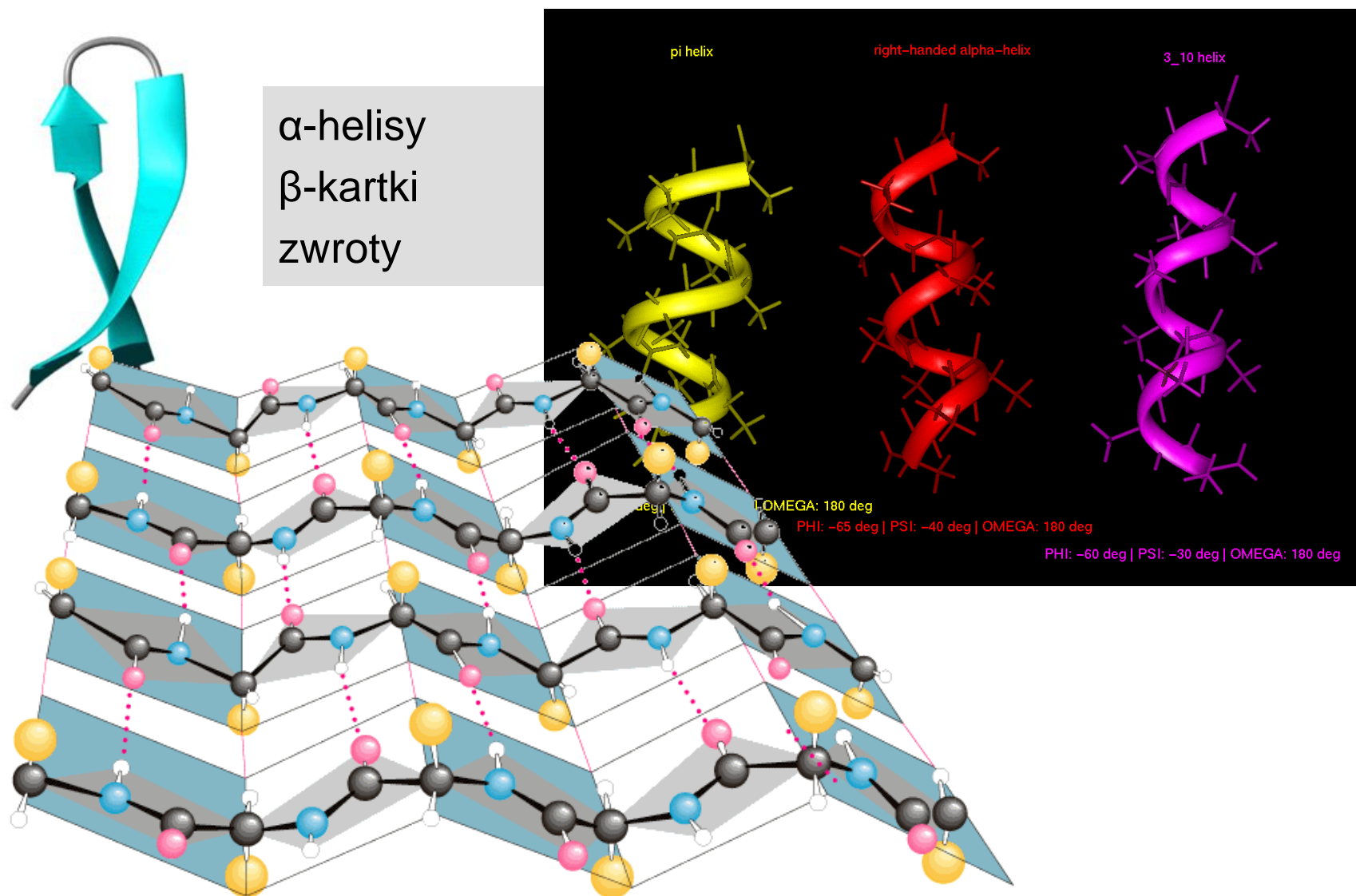
III rządowa – struktura 3D pojedynczego łańcucha

IV rządowa – struktura 3D całego białka (połączone wszystkie łańcuchy)



Zwijanie białka (film): <http://www.youtube.com/watch?v=fvBO3TqJ6FE>

Struktura 2-rzędowa



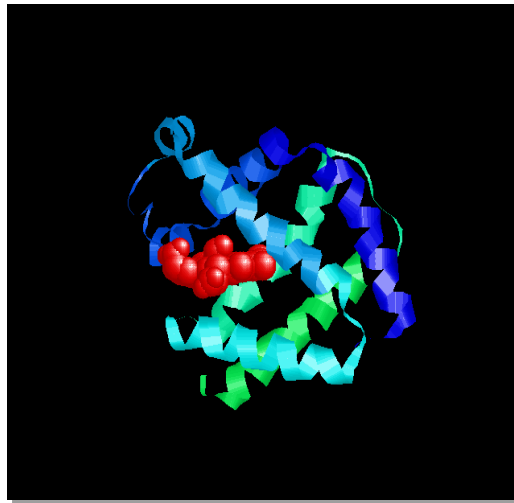
DNA → RNA → Białko

Sekwencja → Struktura → Funkcja → Fenotyp

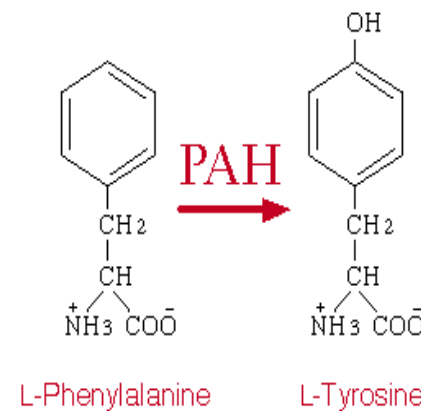
Informacja
genetyczna

MVHLTPEEKT
AVNALWGKVN
VDAVGGEALG
RLLVVYPWTQ
RFFESFGDLS
SPDAVMGNPK
VKAHGKKVLG
AFSDGLAHL
DLNLTGTFSQ
LS ELHCDKLH
VD PENFRLLG
NV LVCVLARN
FG KEFTPQM
QAA YQKVVAG
VAN ALAHKYH

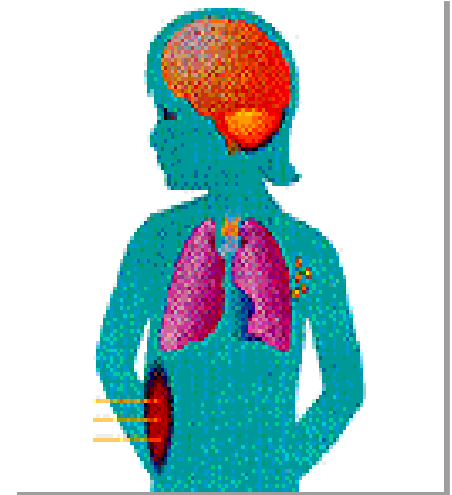
Struktura
molekularna

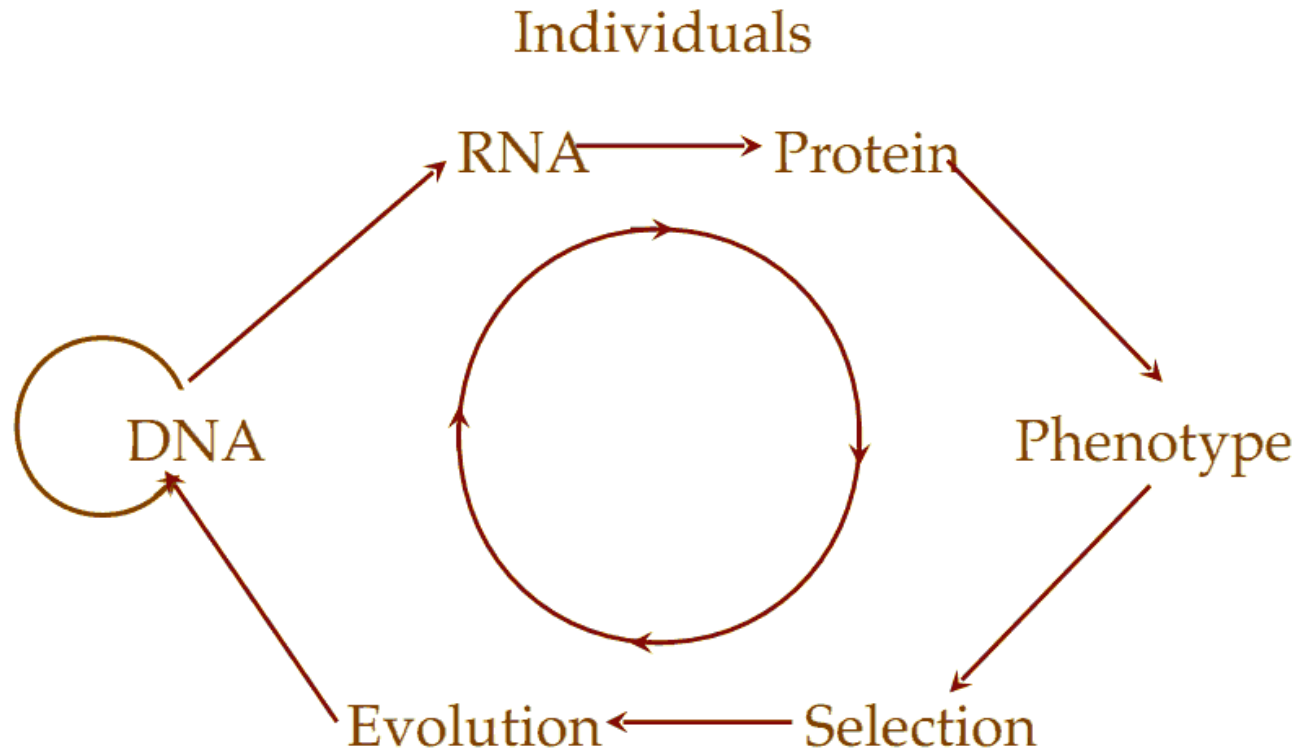


Funkcja
biochemiczna



Fenotyp
(objawy)





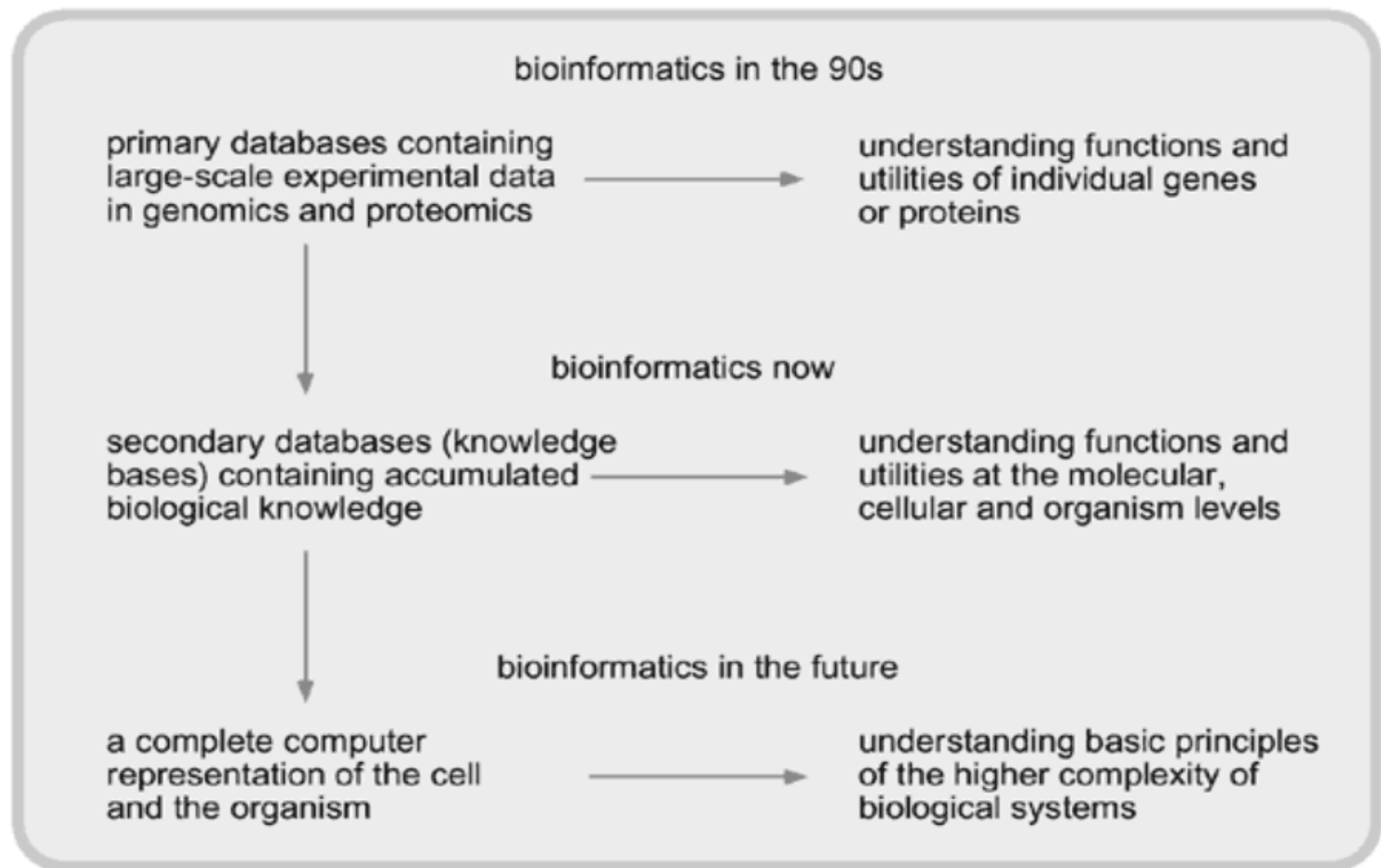
Biological Information

Przepływ informacji na poziomie całej populacji
w szerokiej przestrzeni czasowej

- Informacja genetyczna jest redundantna
różne geny mogą kodować to samo białko
- Informacja strukturalna jest redundantna
różne białka mogą mieć tę samą funkcję
- Jeden gen może mieć wiele funkcji
- Geny są jednowymiarowe, ale ich funkcja zależy od struktury trójwymiarowej kodowanego białka



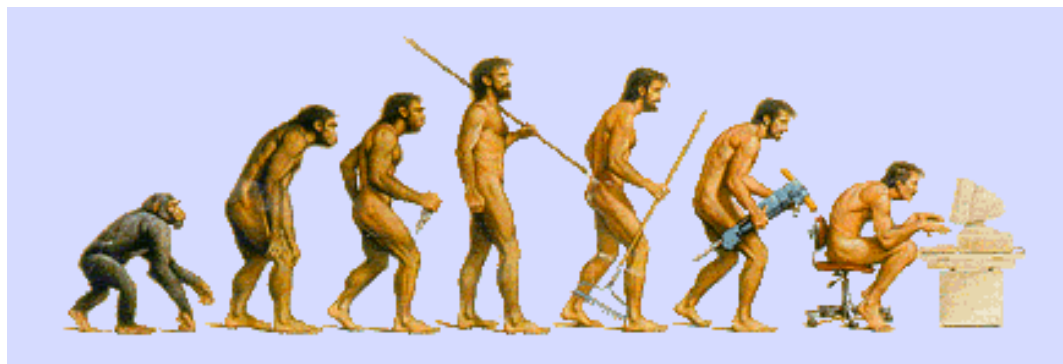
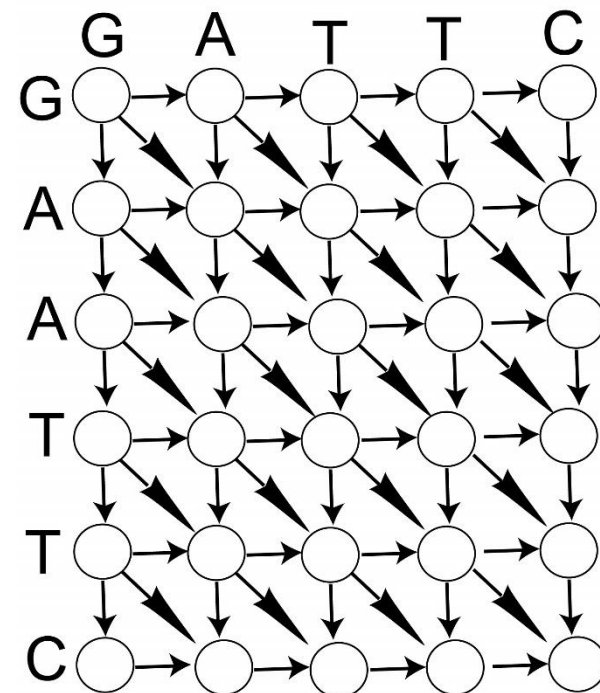
Najważniejsze zadania bioinformatyki



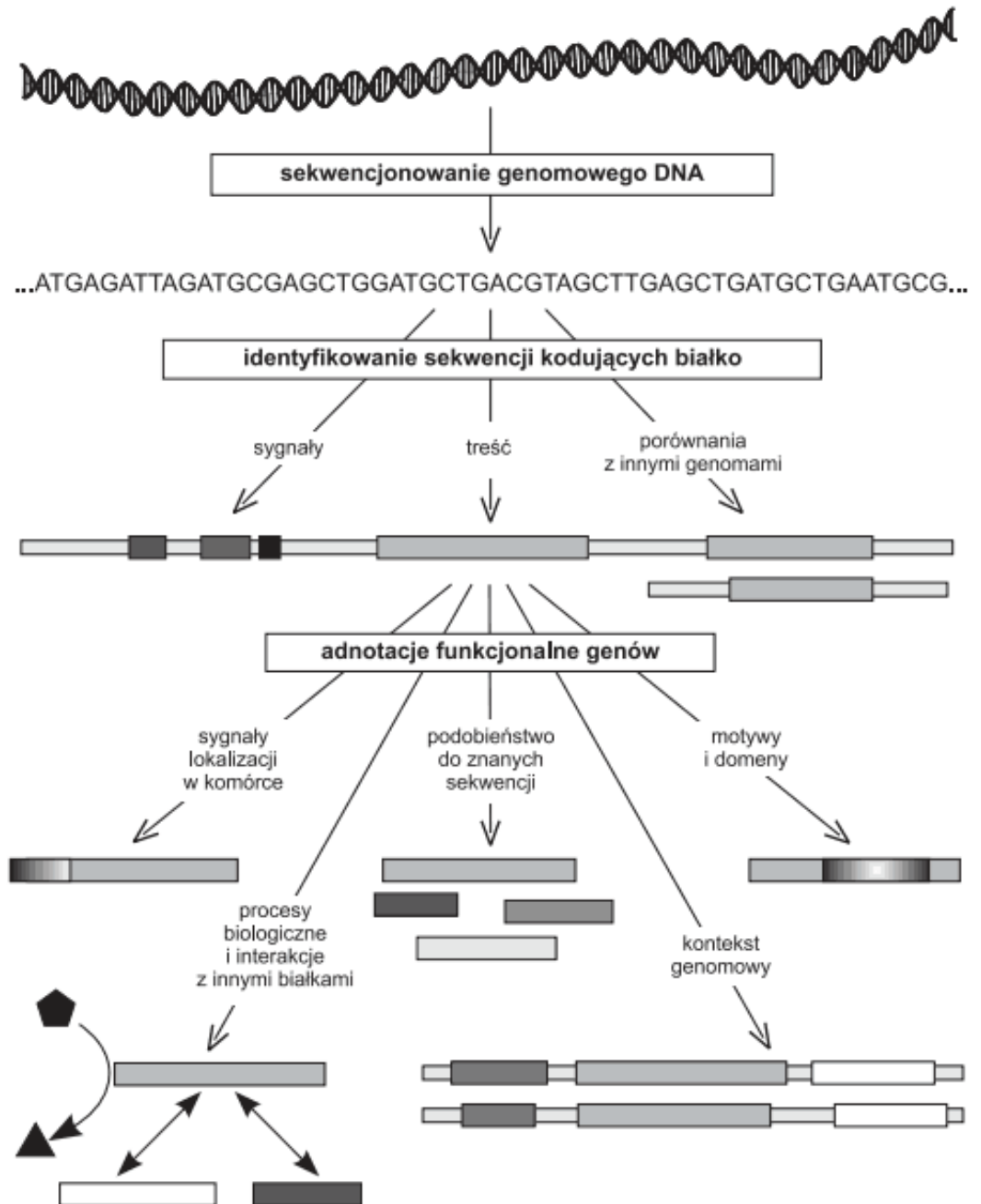
„Tak jak wiek XIX był wiekiem węgla, pary i mechaniki,
a wiek XX: wiekiem atomu i fizyki, tak wszystko wskazuje na to, że
wiek XXI będzie wiekiem biologii i informatyki”

(Prof. Jacek Błazewicz)

- Sekwencjonowanie DNA i RNA
- Mapowanie genomów
- Analiza i porównywanie sekwencji
- Zarządzanie dużymi bazami danych (np. GenBank, EMBL, DDBJ)
- Algorytmy i miary podobieństwa (BLAST itp.)
- Biologia ewolucyjna

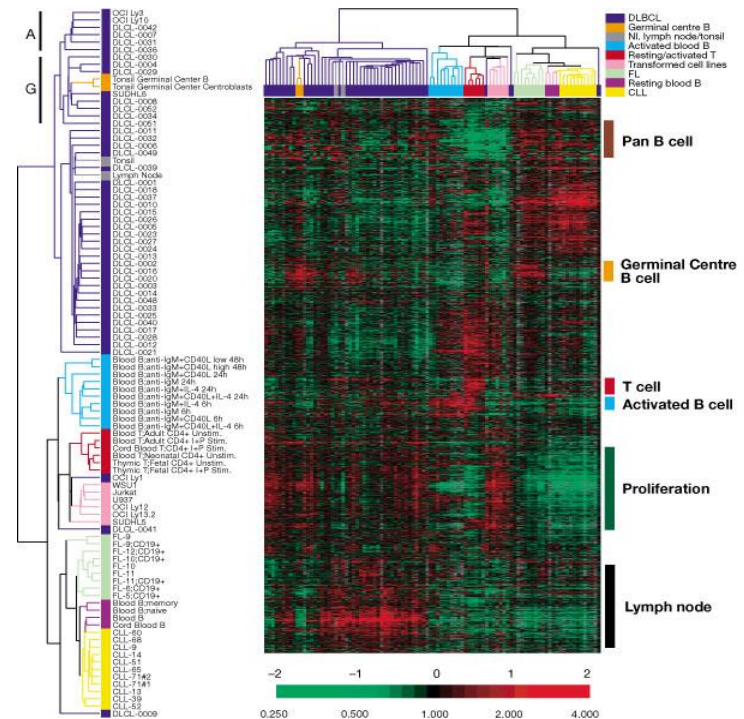
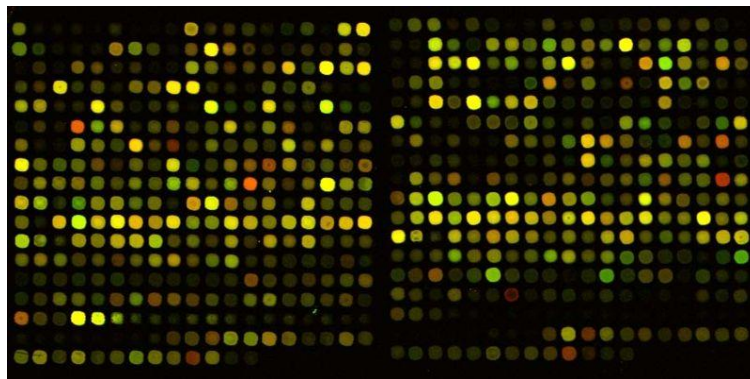


Przykład: Etapy analizy genomowego DNA

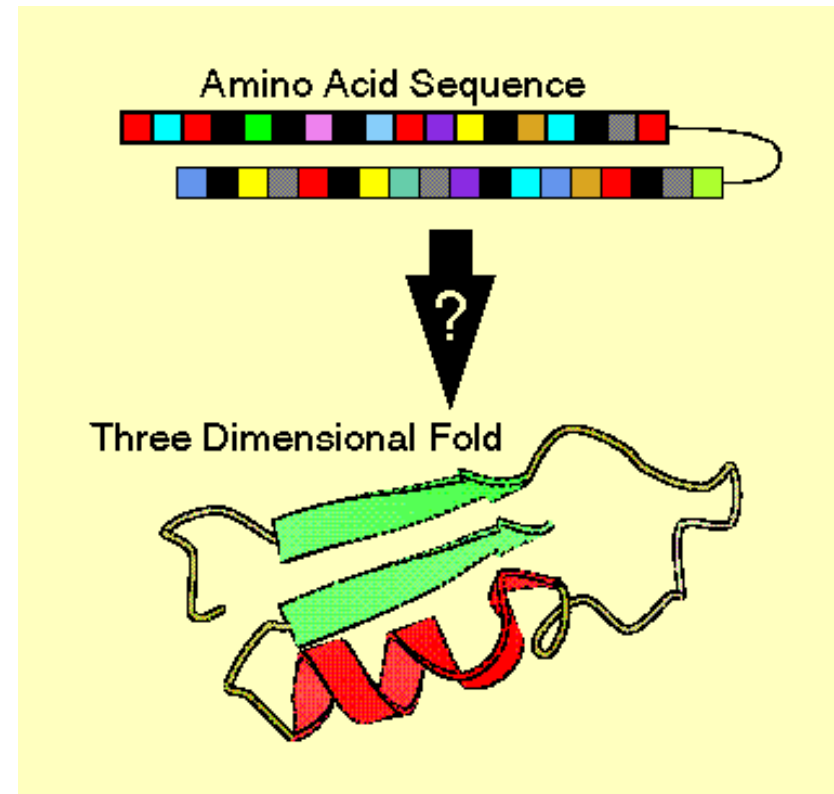
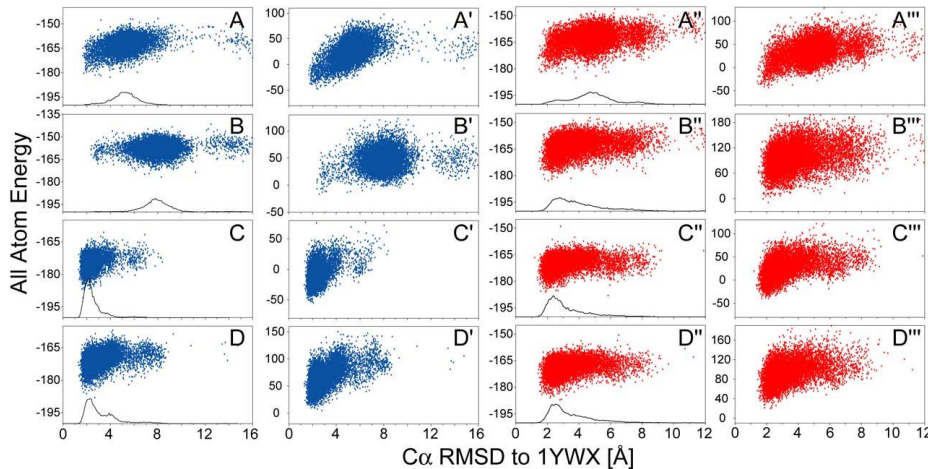
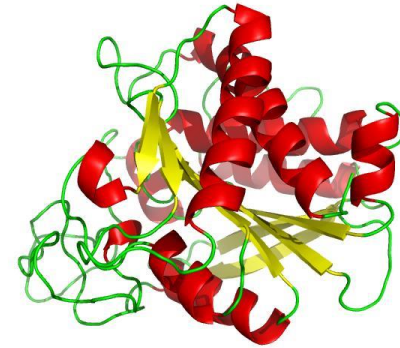


Źródło: Biotechnologia, 3(70):2005, 24

- Sekwencjonowanie transkryptomów
- Analiza mikromacierzy
- Poznawanie funkcji genów
- Badanie interakcji RNA-białko
- Rola RNA różnych typów

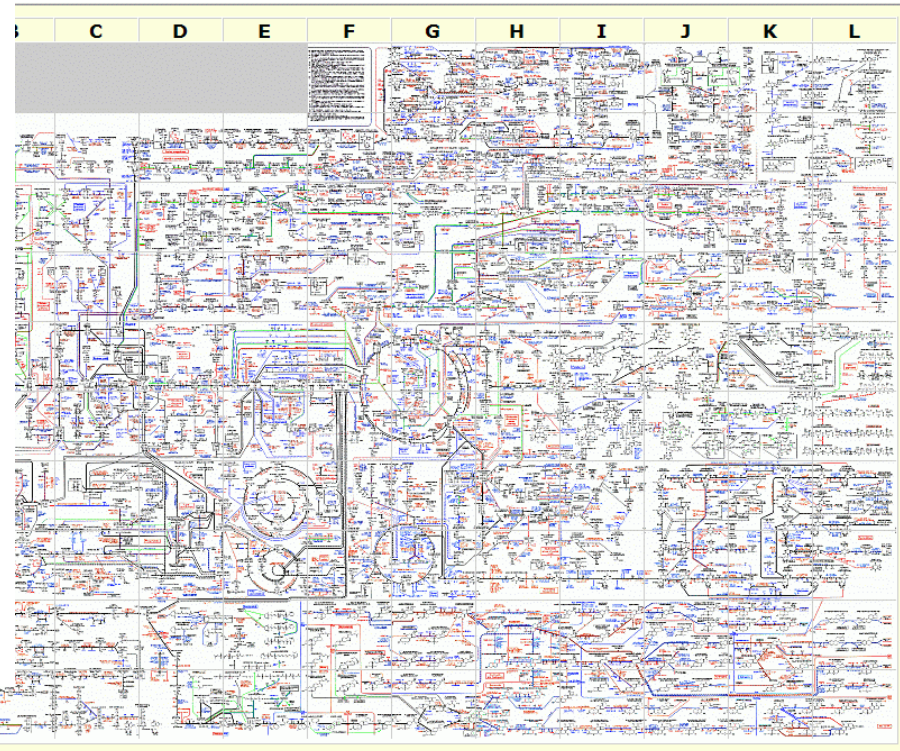
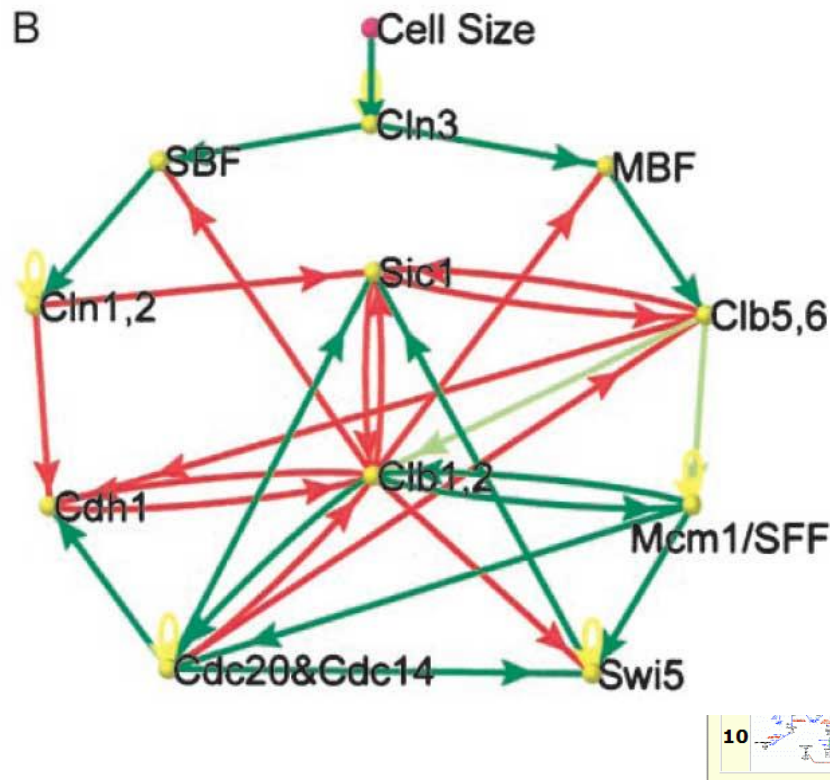


- Klasyfikacja białek
- Przewidywanie struktury białek
- Przewidywanie funkcji białek
- Poszukiwanie miejsc wiążących
- Modelowanie molekularne
- Projektowanie leków (CADD)

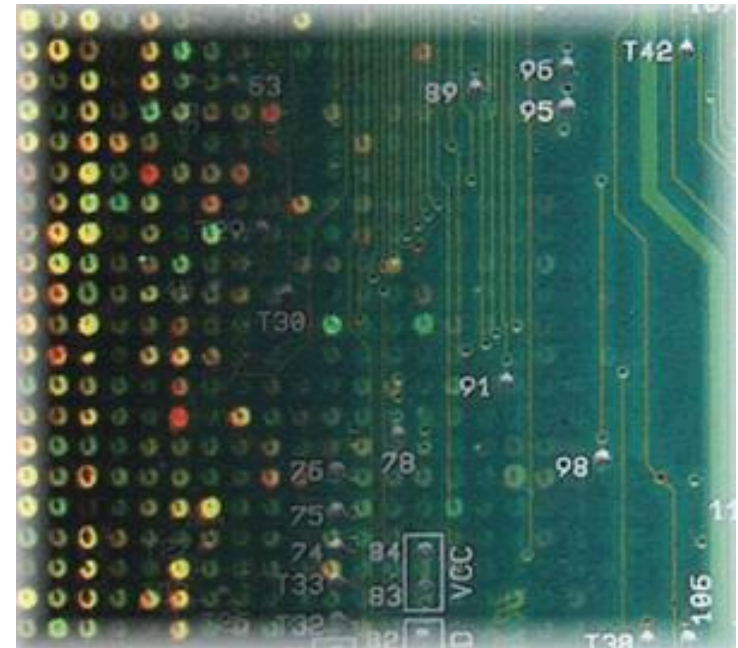
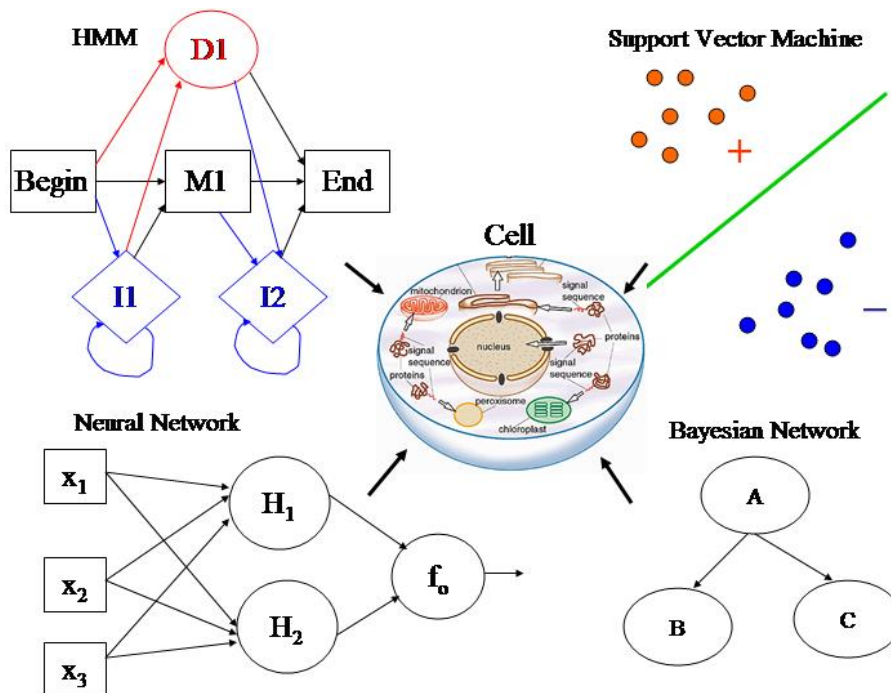


czyli biologia systemów:

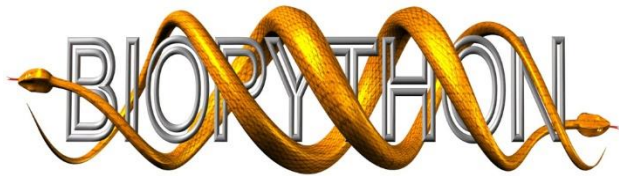
- Modelowanie interakcji w złożonych systemach biologicznych (szlaki metaboliczne);
- Wizualizacja, grafy, sieci zależności



- **genomika funkcjonalna**
(część transkryptomiki; badanie funkcji genów)
- **bioinformatyka strukturalna**
(analiza struktur i interakcji: RNA, białka, kompleksy)



Narzędzia programistyczne (biblioteki Bio*)



Biopython

<http://biopython.org>



BioJava

<http://biojava.org>



BioPerl

<http://www.bioperl.org>



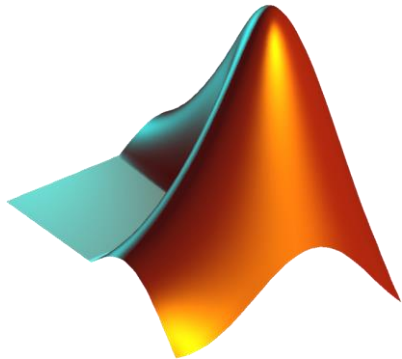
BioRuby

<http://bioruby.open-bio.org/>



R

<http://www.r-project.org/>



MatLab

<http://www.mathworks.com/products/matlab/>

Narzędzia do modelowania molekularnego

np. gromos: <http://www.gromos.net/>



