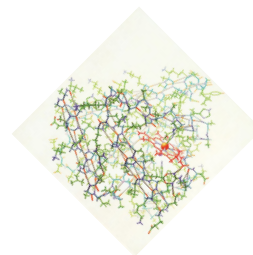


# INŻYNIERIA BIAŁEK TERAPEUTYCZNYCH



ANDRZEJ ŁYSKOWSKI, DR INŻ.

alyskowski@prz.edu.pl, H.237

Used with permission from the Howard Hughes Medical Institute ([www.hhmi.org](http://www.hhmi.org)). All rights reserved.

## Warunki zaliczenia

**Wykład:** 15 godz.  
**Laboratorium:** 15 godz.

### Wykład

Zaliczenie wystawiane jest studentom, którzy uzyskali min. 50% z puli wszystkich możliwych punktów z pisemnego zaliczenia.

### Projekt/Laboratoria/Ćwiczenia

Zaliczenie wystawiane jest studentom, którzy uczestniczyli we wszystkich zajęciach oraz zebrali min. 50% z puli wszystkich możliwych punktów z pisemnych zaliczeń i/lub złożyli w terminie pisemne raporty z wykonanych zadań.

### Ocena końcowa

Ocena końcowa z modułu jest pochodną liczby punktów z zajęć modułu:

$$K = \text{SUMA}(a_n * w * S_n)$$

$a_n$  -współczynnik wagowy dla składnika modułu  $S_n$ ,

w-współczynnik uwzględniający termin zaliczenia, odpowiednio 1,0, 0,9, 0,8 dla I, II, III terminu

# Zakres materiału

## WYKŁADY

- Rozumie cel inżynierii białek terapeutycznych oraz metody służące jego osiągnięciu.
- Zna i rozumie techniki bioinformatycznej charakterystyki białek i ich rekombinowanych pochodnych.
- Zna i rozumie techniki molekularne służące otrzymywaniu białek rekombinowanych.
- Zna i rozumie techniki izolowania, oczyszczania i charakteryzacji białek rekombinowanych.

## Laboratorium

Wykorzystanie metod *in-silico* w celu inżynierii białek w zakresie:

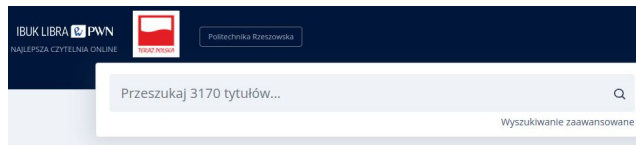
- analizy sekwencji DNA i białkowych,
- analizy proteomicznej,
- analizy strukturalnej.

# Laboratorium

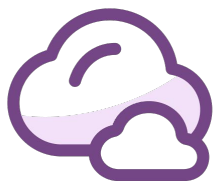
- Raport przedstawiający charakterystykę *in-silico* wybranego białka oraz wybrane aspekty jego inżynierii
- e-learning PRz  
Kurs: Inżynieria białek terapeutycznych (2023/24L)  
Hasło: IBT{1-2}
- Microsoft Sway: <https://sway.office.com/>

# Bibliografia

- <https://libra.ibuk.pl/>
  - Berg *et al.* Biochemia
  - InstantNotes:
    - Westhead *et al.* Bioinformatics
    - Hames *et al.* Biochemistry
  - Baxevanis *et al.* Bioinformatics
  - Doonan *et al.* Białka i peptydy
  - Liljas *et al.* Textbook of structural biology
  - Clegg *et al.* Crystal structure determination
  - Hore *et al.* Nuclear Magnetic Resonance
  - Vincent *et al.* Molecular symmetry and group theory
  - Rhodes *et al.* Crystallography made crystal clear
  - Jaskólski *et al.* Krystalografia dla biologów
  - Bojarski *et al.* Krystalografia
- 
- Aktualne publikacje naukowe



slido



## Inżynieria białek...

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

## Human diseases 6,523 results

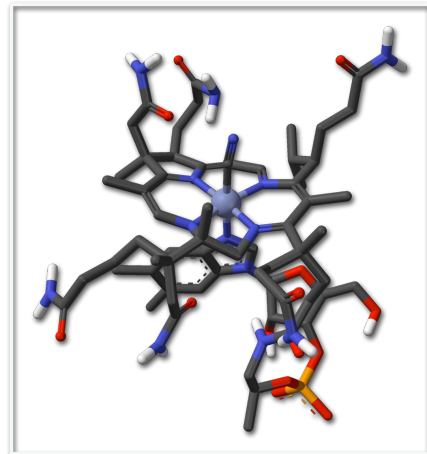
Download View: Cards Table Customize columns Share

ID	Name	Keywords
<input type="checkbox"/> DI-00001	HSD10 mitochondrial disease	Neurodegeneration
<input type="checkbox"/> DI-00002	3-alpha-hydroxyacyl-CoA dehydrogenase deficiency	
<input type="checkbox"/> DI-00003	3-hydroxy-3-methylglutaryl-CoA lyase deficiency	
<input type="checkbox"/> DI-00004	3-methylglutaconic aciduria 1	
<input type="checkbox"/> DI-00005	Barth syndrome	Cardiomyopathy
<input type="checkbox"/> DI-00006	3-methylglutaconic aciduria 3	
<input type="checkbox"/> DI-00007	3-methylglutaconic aciduria 5	Cardiomyopathy
<input type="checkbox"/> DI-00008	Rickets vitamin D-dependent 1B	
<input type="checkbox"/> DI-00009	3-ketothiolase deficiency	
<input type="checkbox"/> DI-00010	Phosphoserine phosphatase deficiency	
<input type="checkbox"/> DI-00011	3M syndrome 1	Dwarfism
<input type="checkbox"/> DI-00012	Aarskog-Scott syndrome	
<input type="checkbox"/> DI-00013	ABCD syndrome	Albinism Deafness Hirschsprung disease
<input type="checkbox"/> DI-00014	Abetalipoproteinemia	
<input type="checkbox"/> DI-00015	Thyroid hormone metabolism, abnormal, 1	
<input type="checkbox"/> DI-00016	Acatalasiaemia	

```

MQEKTRIIK EDIEAVRAYS DCFDVEMPDL DENGVEIGLP APYPREVAGT 50
VRSGYRIYDL AKKAKERGWP IQNPILGRNT AEETYGESQE MYAYADKFDE 100
TLFHFVHAEA TRHIDPLKGR ELINQSRGKG GITPIGEREF IAMGGGSKHP 150
VRINATGDTP HLSIINALIA GFDGTDIGPV IHVHFGGRGI HDYKTKVVNG 200
YKAIQICAE NIFVQLDSHK HLNNIGGTDG MALAMCLLSE GLAVHAGLPW 250
ELSAIQMNVA GINIYADLAV MRAFRKACHS KSIIVAVPETF QNPPGNLVAE 300
AAHFSRMAVT AKLGGADFYR PKAAESVGIP TGDSMGQAIW GTEDVFGHVV 350
NPDIQSPVID AREAIIIDEA LAVLEATLHL EGLTLEAMTD DFWKQWSDEA 400
LIDLIVAAGK AGVLDSQRAA GWDLKRHVVV NRDKDGITRY VKGYTPLGVD 450
ASRCAQSDED VEVHVEKAPT RPEKIVLATV GADAHVNGIN VIREAFQDAG 500
YDVVYLRGMN LPESVAEVAA EVGADAVGVS NLLGLGMELF PRVSKRLEEL 550
GLRDKMVVCA GGRIAEKEEE HRQFEKIQQ EGSAFMGMGD FFGPGSSPED 600
CVKIIGDMIN AKKA 614

```



2-methyleneglutarate = 2-methylene-3-methylsuccinate

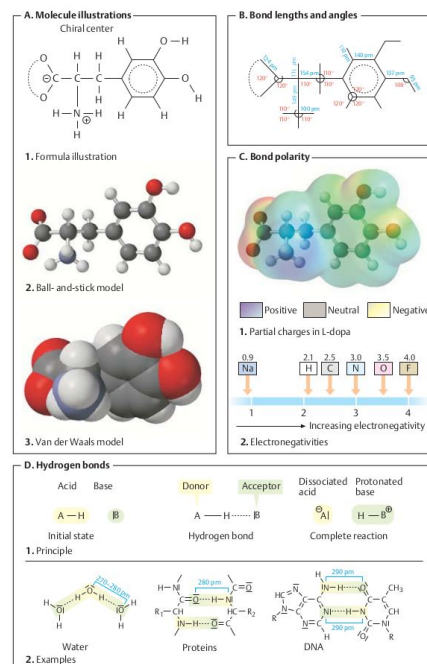
# Inżynieria...

**Działalność polegająca na projektowaniu, konstrukcji, modyfikacji i utrzymaniu efektywnych kosztowo rozwiązań dla praktycznych problemów, z wykorzystaniem wiedzy naukowej oraz technicznej.**

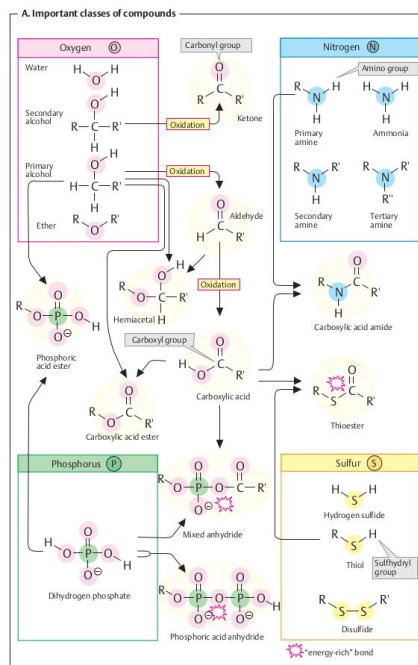
Działalność ta wymaga rozwiązywania problemów różnej natury oraz skali. Bardziej ogólnie, inżynieria zajmuje się też rozwojem technologii.

- '... projektowaniu, konstrukcji, modyfikacji' białek i enzymów.
- '... z wykorzystaniem wiedzy naukowej oraz technicznej', tj. metod opartych o np. narzędzia bioinformatyczne.
- '... rozwiązywania problemów różnej natury oraz skali', tj. utrzymanie integralności strukturalnej i/lub modyfikacja aktywności enzymatycznych.

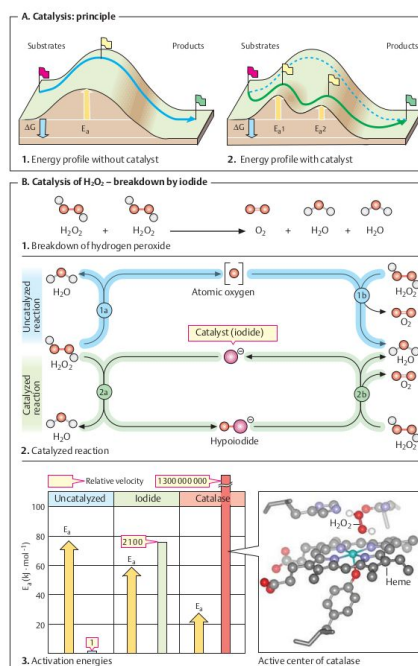
# Budowa molekularna



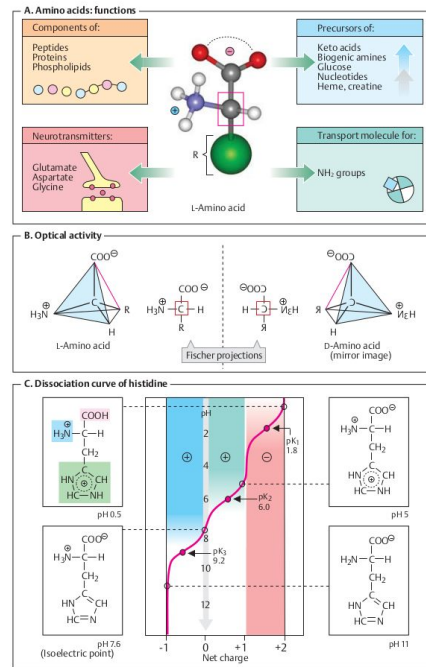
# Biocząsteczki i ich właściwości



# Biokataliza



# Aminokwasy



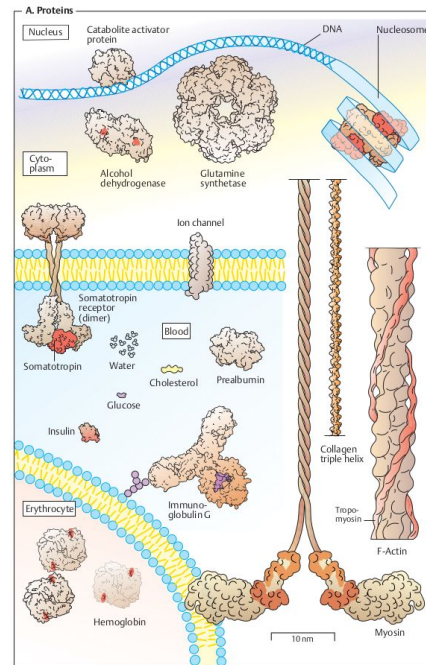
# Właściwości aminokwasów

**A. The proteinogenic amino acids**

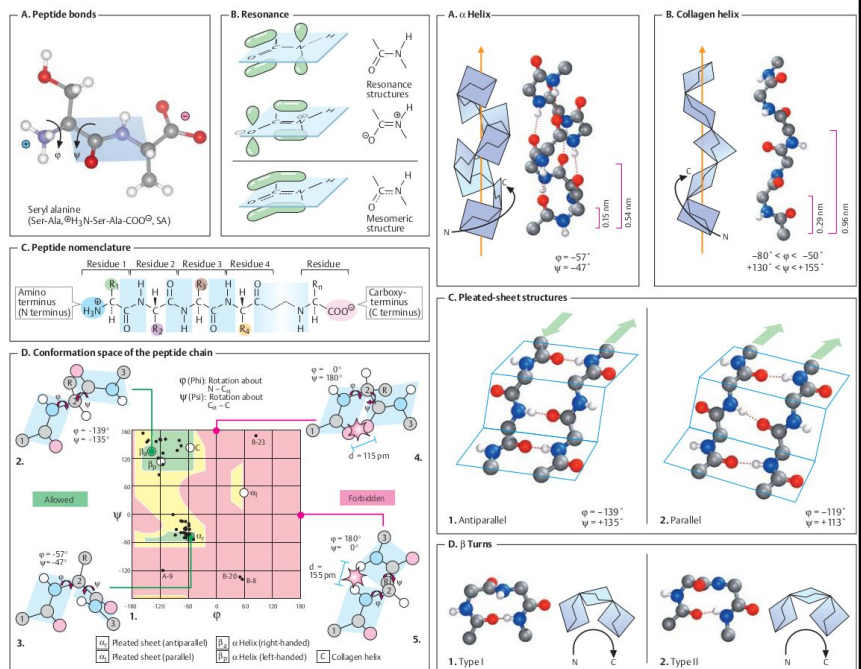
Aliphatic				Sulfur-containing		
Glycine (Gly, G)	Alanine (Ala, A)	Valine (Val, V)	Leucine (Leu, L)	Isoleucine (Ile, I)	Cysteine (Cys, C)	Methionine (Met, M)
H	CH <sub>3</sub>	H <sub>3</sub> C-CH(CH <sub>3</sub> )	H <sub>3</sub> C-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>3</sub>	H <sub>3</sub> C-CH(CH <sub>3</sub> )-CH <sub>2</sub> -CH <sub>3</sub>	CH <sub>2</sub> -SH	CH <sub>2</sub> -CH <sub>2</sub> -S-CH <sub>3</sub>
-2.4	-1.9	-2.0	-2.3	-2.2	-1.2	-1.5
Aromatic				Cyclic	Neutral	
Phenylalanine (Phe, F)	Tyrosine (Tyr, Y)	Tryptophan (Trp, W)	Proline (Pro, P)	Serine (Ser, S)	Threonine (Thr, T)	
CH <sub>2</sub> -C <sub>6</sub> H <sub>5</sub>	CH <sub>2</sub> -C <sub>6</sub> H <sub>4</sub> -OH	CH <sub>2</sub> -Indole ring	CH <sub>2</sub> -C <sub>5</sub> H <sub>9</sub> N	CH <sub>2</sub> -OH	H <sub>3</sub> C-CH(OH)-CH <sub>3</sub>	
+0.8	+6.1	+5.9	+6.0	+5.1	+4.9	
Neutral		Acidic		Basic		
Asparagine (Asn, N)	Glutamine (Gln, Q)	Aspartic acid (Asp, D)	Glutamic acid (Glu, E)	Histidine (His, H)	Lysine (Lys, K)	Arginine (Arg, R)
CH <sub>2</sub> -CONH <sub>2</sub>	CH <sub>2</sub> -CH <sub>2</sub> -CONH <sub>2</sub>	CH <sub>2</sub> -COO <sup>-</sup>	CH <sub>2</sub> -CH <sub>2</sub> -COO <sup>-</sup>	CH <sub>2</sub> -CH <sub>2</sub> -C(=NH)-NH <sup>+</sup>	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -NH <sub>3</sub> <sup>+</sup>	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -NH-C(=NH)-NH <sub>2</sub> <sup>+</sup>
+9.7	+9.4	+11.0	+10.2	+10.3	+15.0	+20.0

☆ Essential amino acids      □ Chiral center

# Białka



# Właściwości białek





# Struktura białek

**A.  $\alpha$ -Keratin**

**B. Collagen**

1. Triple helix (section)

Gly	Arg	Hyp
Gly	Gln	Arg
Gly	Pro	Hyp
Gly	Pro	Gln
Gly	Ala	Arg

2. Typical sequence

3. Triple helix (view from above)

**A. Conformation-stabilizing interactions**

**B. Disulfide bonds**

S[C@@H](N)C(=O)O + S[C@@H](N)C(=O)O >> S[C@@H](N)C(=O)O + S[C@@H](N)C(=O)O

**C. Protein dynamics**

**D. Folding patterns**

1. Myoglobin

2. Estrogen receptor (domain)

3. Flavodoxin

4. Transducin ( $\beta$  subunit)

# Aktywność biologiczna

**A. Structure of insulin**

1. Primary structure

2. Secondary and tertiary structure

3. Quaternary structure

**B. Insulin (monomer)**

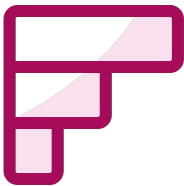
1. A-chain, B-chain

2. Invariant residue

3. Polar side chain, Apolar side chain, Involved in subunit interactions

Zasoby internetowe

slido



Zasoby internetowe

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

# UniProtKB statistics

## Introduction

This is release 2024\_01 of UniProtKB, published on Wed Jan 24 2024.

Previous release statistics are available from the UniProt FTP server.

Throughout this document, whenever a statistic has a corresponding query, a link has been provided. In some instances, due to the nature of the sta

### Total number of entries in this release of UniProtKB

Section	Number of entries in total	Number of entries with an annotation update	Number of entries with a sequence update
UniProtKB	250,322,721	161,118,372	1,287
Reviewed (Swiss-Prot)	570,830	433,476	169
Unreviewed (TrEMBL)	249,751,891	160,684,896	1,118

### Total number of new entries in this release of UniProtKB

Section	Number of new entries	Number of new sequences
UniProtKB	3,922,146	3,922,117
Reviewed (Swiss-Prot)	415	386
Unreviewed (TrEMBL)	3,921,731	3,921,731

**Proteins**  
UniProt Knowledgebase

Reviewed (Swiss-Prot)  
570,830

Unreviewed (TrEMBL)  
249,751,891

Protein sets for species with sequenced genomes from across the tree of life

Clusters of protein sequences at 100%, 90% & 50% identity

Non-redundant archive of publicly available protein sequences seen across different databases

## Supporting Data

- Taxonomy
- Keywords
- Literature Citations
- Human diseases
- Cross-referenced databases
- Subcellular locations
- Automatic annotations: UniRule & ARBA

- Entry
- Variant viewer
- Feature viewer**
- Genomic coordinates
- Publications
- External links
- History

## Feature viewer

Download

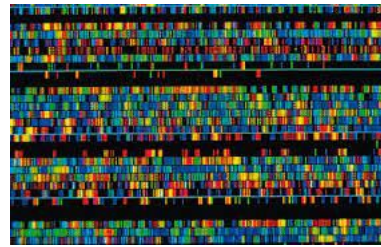


# Bioinformatyka

Część eksperymentalna projektów badawczych opiera się o eksperymenty prowadzone:

- *in vivo*  
wewnątrz żyjącego organizmu;
- *in vitro*  
w probówce (w szkle);
- *in silico*  
na komputerze (w krzemie);

Symulacje komputerowe w oparciu o bazy danych i analiza uzyskanych danych to główne obszary aktywności badań bioinformatycznych.



# Identyfikacja sekwencji

Homologia to podstawowe zjawisko odzwierciedlające wzajemną relację między dwoma lub większą liczbą jednostek.

Dwie cząsteczki są homologiczne gdy pochodzą od jednego przodka.

Homologi można podzielić na:

**Paralogi** czyli homologi występujące w obrębie jednego gatunku.

Paralogi często różnią się funkcjami biochemicznymi.

**Ortologi** czyli homologi występujące w różnych gatunkach.

Funkcja biochemiczna zostaje zachowana.

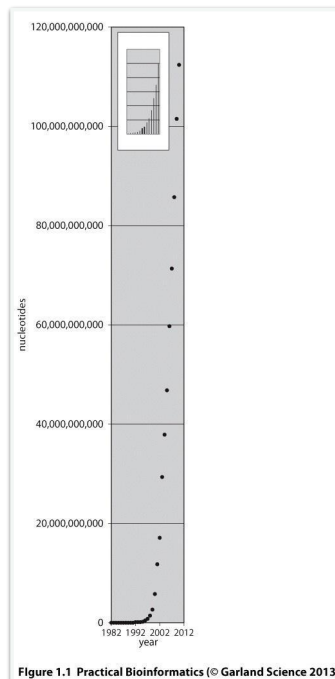


Figure 1.1 Practical Bioinformatics (© Garland Science 2013)

# Identyfikacja homologów

Homologia przejawia się często znaczącym podobieństwem w:

- sekwencji nukleotydowej;
- sekwencji aminokwasowej;
- strukturze przestrzennej.

*Problem: Identyfikacja w oparciu o którą cechę będzie najwydajniejsza?*

**Table 1.1 The size of genomes**

Species	Genome size (10 <sup>6</sup> nucleotides)	Number of genes
<i>Escherichia coli</i>	4.7	4300
<i>Saccharomyces cerevisiae</i>	12	6700
<i>Drosophila melanogaster</i>	169	13,900
<i>Danio rerio</i>	1500	26,000
<i>Homo sapiens</i>	3200	21,000
<i>Zea mays</i>	3200	63,000
<i>Oryza sativa</i>	488	57,000

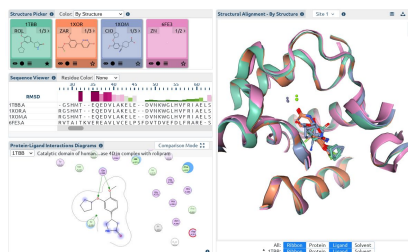
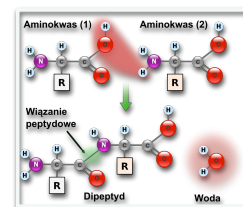
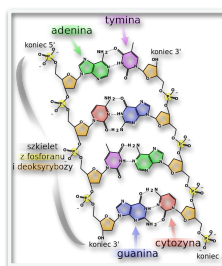
Source: The Ensembl Genome Browser ([www.ensembl.org](http://www.ensembl.org)) April 2012.

Table 1.1 Practical Bioinformatics (© Garland Science 2013)

# Dopasowanie sekwencji

Klasyfikacja przyrównań (ang. sequence alignment):

- nukleotydowej: cztery elementy porównawcze w tripletach.
- sekwencji aminokwasowej: 20 jednostek porównawczych.
- strukturze przestrzennej: struktura I, II, III, IV rzędowa definiująca przestrzenne ułożenie atomów.



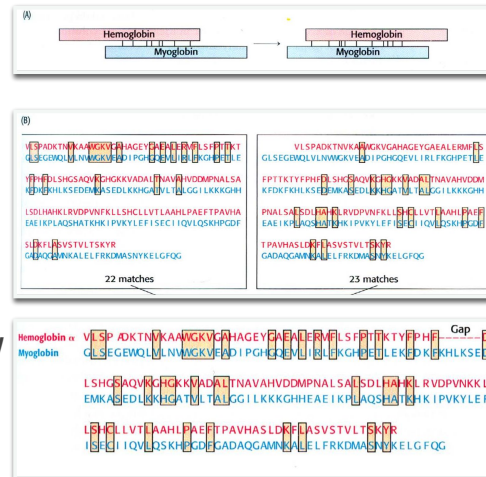
# Przyrównanie doskonałe

1:1

Przesunięcie sekwencji względem siebie i obliczenie punktacji przyrównania (*ang. score*).

1:1 + przerwa (*ang. gap*)

Celem lepszego dopasowania wprowadza się w sekwencję przerwy celem kompensacji insercji lub delecji nukleotydów.



# Zmiany ewolucyjne

Mutacje (zmiana tożsamości, usunięcie lub dodanie aminokwasu/domeny) na poziomie genetycznym zachodzą tylko w określonych warunkach i gdy nie wpływają na strukturę lub funkcję kodowanego białka

*Problem: jak ocenić podobieństwo sekwencji uwzględniając zmiany ewolucyjne?*

# Kod DNA

## Cechy kodu DNA:

- Trójkowy
- Niezachodzący
- Bezprzecinkowy
- Zdegenerowany  
różne kodony umożliwiają kodowanie tego samego aminokwasu, tzn. prawie wszystkie aminokwasy mogą być zakodowane na kilka sposobów.
- Część zmian w informacji genetycznej w wyniku mutacji nie znajduje swojego odbicia w sekwencji aminokwasów.
- Jednoznaczny
- Kolinearny
- Uniwersalny

AMINO ACIDS AND THEIR SYMBOLS	CODONS	MEMORY AID
A Ala Alanine	GCA GCC GCG GCT	Alanine
C Cys Cysteine	TGC TGT	Cysteine
D Asp Aspartic acid	GAC GAT	Aspartic acid
E Glu Glutamic acid	GAA GAG	Glutamic acid
F Phe Phenylalanine	TTC TTT	Phenylalanine
G Gly Glycine	GGA GGC GGG GGT	Glycine
H His Histidine	CAC CAT	Histidine
I Ile Isoleucine	ATA ATC ATT	Isoleucine
K Lys Lysine	AAA AAG	K is adjacent to Lysine in the alphabet
L Leu Leucine	TTA TTG CTA CTC CTG CTT	Leucine
M Met Methionine	ATG	Methionine
N Asn Asparagine	AAC AAT	Asparagine
P Pro Proline	CCA CCC CCG CCT	Proline
Q Gln Glutamine	CAA CAG	Glutamine
R Arg Arginine	AGA AGG CGA CGC CGG CGT	Arginine
S Ser Serine	AGC AGT TCA TCC TCG TCT	Serine
T Thr Threonine	ACA ACC ACG ACT	Threonine
V Val Valine	CTA CTC GTG GTT	Valine
W Trp Tryptophan	TGG	Tryptophan
Y Tyr Tyrosine	TAC TAT	Tyrosine
X Any amino acid		

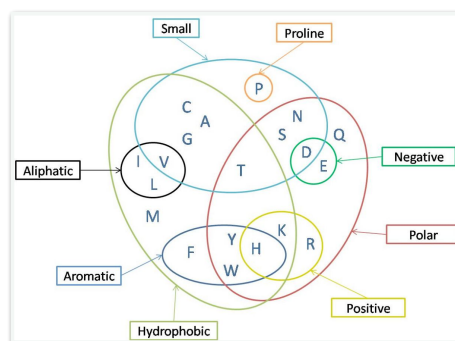
Figure 4-4 Practical Bioinformatics (© Garland Science 2013)

# Ewolucja białek

## Zmiany ewolucyjne mogą zachodzić w dwóch obszarach sekwencji:

- *kluczowym*, czyli obszarze odpowiedzialnym za konkretną funkcję biologiczną;
- *powłoce*, czyli obszarze odpowiedzialnym za uformowanie struktury białka ale nie biorącym bezpośredniego udziału w katalizie.

Zachodzące mutacje w obszarze kluczowym mają tendencję do zamiany aminokwasów na inne o podobnych właściwościach fizykochemicznych.  
Np. zamiana hydrofobowych aminokwasów w rdzeniu białka (Leu, Ile, Val).



# Macierze substytucji

*Problem: Jak przyrównać sekwencje zawierające mutacje?*

Ocenie jakości przyrównani sekwencji białek zawierających mutacje służą macierze substytucji (*ang. substitution matrices*).

Służą one do oceny podobieństwa pomiędzy aminokwasami i obliczenia punktacji dla konkretnego wariantu przyrównania.

*Uwaga: przerwa w sekwencji ma zawsze ujemną wartość punktową!*

A	4																		
R	-1	5																	
N	-2	0	6																
D	-2	-2	1	6															
C	0	-3	-3	-3	9														
Q	-1	1	0	0	-3	5													
E	-1	0	0	2	-4	2	5												
G	0	-2	0	-1	-3	-2	-2	6											
H	-2	0	1	-1	-3	0	0	-2	8										
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4									
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4								
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5							
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5						
F	-2	-3	-3	-3	-2	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	4	7				
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4			
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	
Y	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	0	-3	-1	4
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

Figure 4.9 Practical Bioinformatics (© Garland Science 2012)

# Typowe macierze

*PAM (ang. Accepted Point Mutation)*

PAM250 to macierz zawierająca punktację dla drogi ewolucyjnej równej 250 akceptowalnym mutacjom punktowym na 100 reszt aminokwasowych.

Macierze PAM oblicza się przez przyrównanie i porównanie zmian ewolucyjnych w blisko spokrewnionych białkach a następnie ekstrapoluje się te wartości na dalsze 'odległości' ewolucyjne.



# Typowe macierze

BLOSUM (*ang. BLOcks SUBstitutio Matrix*)

Punktację macierzy BLOSUM ustala się w oparciu o przyrównanie i porównanie segmentów sekwencji ewolucyjnie spokrewnionych białek. Im mniejszy numer identyfikacyjny macierzy BLOSUM tym dłuższa odległość ewolucyjna.

## PAM vs. BLOSUM

PAM250, PAM100

BLOSUM62, BLOSUM50

# Zasoby internetowe

## Portale:

- European Bioinformatics Institute (EBI)  
<http://www.ebi.ac.uk/>
- National Center for Biotechnology Information (NCBI)  
<http://www.ncbi.nlm.nih.gov/>
- Expert Protein Analysis System (ExpASy)  
<http://expasy.org/>