

# Bioinformatyka

Andrzej Łyskowski, dr inż.  
Katedra Biotechnologii i Bioinformatyki

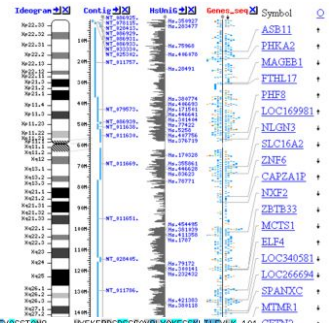
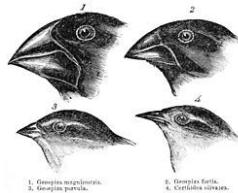
andrzej.lyskowski@prz.edu.pl  
H-237

## Bioinformatyka

Kombinacja informacji biologicznych i technik informacyjnych.  
Analiza bioinformatyczna opiera się na analizie informacji zebranych w bazach danych w celu wyjaśnienia obserwowanych faktów doświadczalnych

Rezultatem analizy są:

- Określenie wzajemnych zależności,
- Pokrewieństwa sekwencji kodujących,
- Podobieństwa struktur,
- Profile ekspresji,
- Profile szlaków metabolicznych.



Accession	Seq ID	Sequence	Gene	Start	End
AF0663.1	14	ETKLPPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	101	101	101
84F917.1	13	STKLPPSSESTINLVOMTNNLST...EDFSDH...VQLGKDEHENKTEEL...CFALQDE...HFREEPDQSSSAWLVAKETSDHMLVEV	100	100	100
AF51V2.1	23	VKLPPSSESTINLVOMTNNLST...AFESD...FRELHRODNRHLEEV...VDFGAE...HDSGGDIT...SISAWV...HSHMLVETL	109	109	109
896307.1	13	ETKLPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	100	100	100
DB0056.1	30	SFSLPPTQTRDRVVRVLDLGG...DILCKR...VAVPRAADCPARGLER...DFQDRA...SGEAAATRSVEITGILLVSKVSRHLLDFV	120	120	120
Q00422.2	46	SLSLPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	138	138	138
89M08.1	56	SFSLPPTQTRDRVVRVLDLGG...DILCKR...VAVPRAADCPARGLER...DFQDRA...SGEAAATRSVEITGILLVSKVSRHLLDFV	141	141	141
Q00422.2	29	SEHAPPTQTRDRVVRVLDLGG...DILCKR...VAVPRAADCPARGLER...DFQDRA...SGEAAATRSVEITGILLVSKVSRHLLDFV	123	123	123
89M46.1	13	STKLPPSSESTINLVOMTNNLST...EDFSDH...VQLGKDEHENKTEEL...CFALQDE...HFREEPDQSSSAWLVAKETSDHMLVEV	100	100	100
Q9C500.1	97	SLKPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	142	142	142
Q2617.1	26	IVYLPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	110	110	110
Q9M93.1	28	SLKPPSSTITLIVLVEPHNNLST...PFIYTK...VSLSKERENKQTEEV...RSTFQND...RYVCEPDGCGGSHWVWES...SKLLELQV	119	119	119
Q9M796.1	29	SFSLPPTQTRDRVVRVLDLGG...DILCKR...VAVPRAADCPARGLER...DFQDRA...SGEAAATRSVEITGILLVSKVSRHLLDFV	110	110	110
Q9LEB2.1	14	SVHPPSSESTINLVOMTNNLST...EDFSDH...VQLGKDEHENKTEEL...CFALQDE...HFREEPDQSSSAWLVAKETSDHMLVEV	101	101	101
Q9M93.1	13	STKLPPSSESTINLVOMTNNLST...EDFSDH...VQLGKDEHENKTEEL...CFALQDE...HFREEPDQSSSAWLVAKETSDHMLVEV	100	100	100
89R748.1	48	SLSLPPTQTRDRVVRVLDLGG...DILCKR...VAVPRAADCPARGLER...DFQDRA...SGEAAATRSVEITGILLVSKVSRHLLDFV	133	133	133

# Rola komputerów

Źródłem danych są organizmy żywe.

Każdy gatunek stanowi niepowtarzalny i unikalny zbiór danych, którego wielkość jest różna w zależności.

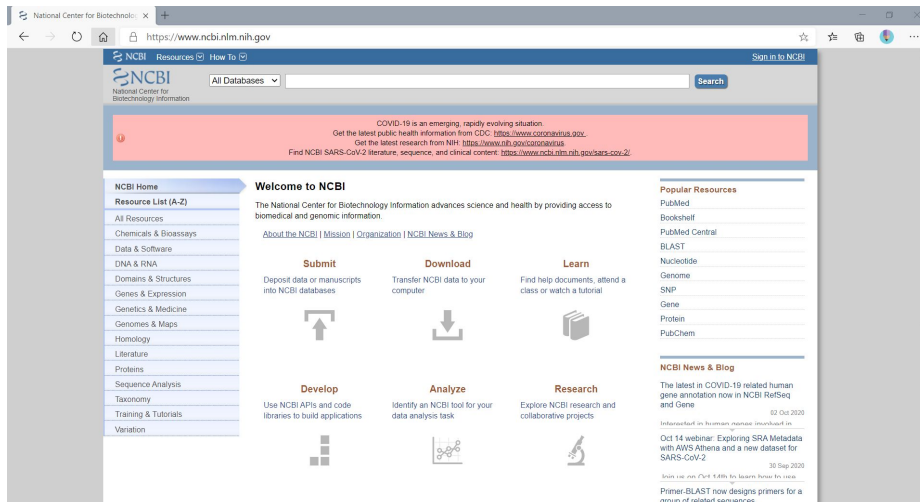
Każda naturalna wariacja w ramach gatunku stanowi nowy zbiór danych.

Komputery są niezbędne do gromadzenia, zarządzania i przetwarzania danych w akceptowalnym zakresie czasowym.

GENBANK AND WGS STATISTICS

Release	Date	GenBank		WGS	
		Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
14	Nov 1983	2274029	2427		
20	May 1984	3002088	3665		
129	Apr 2002	19072679701	16769983	692266338	172768
130	Jun 2002	20648748345	17471130	3267608441	397502
131	Aug 2002	22616937182	18197119	3848375582	427771
237	Apr 2020	415770027949	216531829	7788133221338	1267547429
238	Jun 2020	427823258901	217122233	8114046262158	1302852615
239	Aug 2020	654057069549	218642238	8841649410652	1408122887

# National Center for Biotechnology Information



# EMBL European Bioinformatics Institute

The screenshot shows the EMBL-EBI website homepage. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, and About us. The main header features the EMBL-EBI logo and the tagline "The home for big data in biology". Below this, a search bar is visible with the text "Explore dozens of biological data resources with our search service." and a search button. A prominent yellow banner in the center reads "EMBL-EBI response to COVID-19" and provides information about the institute's closure on March 18th. Below the banner, there are sections for "Featured topic" (highlighting the COVID-19 Data Portal and SARS-CoV-2) and "Latest news" (dated 23 Sep 2020).

8 /

# UniProt

The screenshot shows the UniProt website homepage. The navigation bar includes links for UniProtKB, BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. The main content area is titled "The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information." Below this, there are several key sections: UniProtKB (with counts for Swiss-Prot and TrEMBL), UniRef (UniProt Reference Clusters), UniParc (comprehensive and non-redundant database), and Proteomes (set of proteins for a species). A "Supporting data" section lists various resources like literature citations, taxonomy, and subcellular locations. A "News" section features a prominent red banner for "New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle." and other recent updates. The footer includes "Getting started", "UniProt data", and "Protein spotlight" sections.

9 /

# KEGG Kyoto Encyclopedia of Genes and Genomes

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological systems, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See Release notes (October 1, 2020) for new and updated features.

**Main entry point to the KEGG web service**

**KEGG Z** KEGG Table of Contents [Update notes] [Release history]

**Data-oriented entry points**

Database	Description	Classification
KEGG PATHWAY	KEGG pathway maps	Pathway
KEGG BRITE	BRITE hierarchies and tables	Brize
KEGG MODULE	KEGG modules	Brize table
KEGG ORTHOLOGY	KO functional orthology [Annotation]	Module
KEGG GENOME	Genomes [Pathogen   Virus   Plant]	KO (Function)
KEGG GENES	Genes and proteins [SeqData]	Organism
KEGG COMPOUND	Small molecules	Virus
KEGG GLYCAN	Glycans	Compound
KEGG REACTION	Biochemical reactions [RModule]	Network
KEGG ENZYME	Enzyme nomenclature	Disease (ICD)
KEGG NETWORK	Disease-related network variations	Drug (ATC)
KEGG DISEASE	Human diseases	Drug (Target)
KEGG DRUG	Drugs [New drug approvals]	
KEGG MEDICUS	Health information resource [Drug labels search]	

**Organism-specific entry points**

KEGG Organisms Enter org code(s)  [Go] hsa hsa eco

**Analysis tools**

KEGG Mapper KEGG PATHWAY/BRITE/MODULE mapping tools  
 BlastKOALA BLAST-based KO annotation and KEGG mapping  
 GhostKOALA GHOSTX-based KO annotation and KEGG mapping

10 /

# ExPASy Swiss Institute of Bioinformatics

Supporting COVID-19 / SARS-CoV-2 research

SIB experts and resources in the fight against COVID-19 (evolving list):

- UniProtKB/Swiss-Prot: Knowledge of SARS-CoV-2 protein sequences and how they function
- ViralZone: Biological insights, including a detailed comparison with the SARS virus' genome as well as cross-links to complementary resources. View Webinar of 14 April 2020
- SIB COVID-19 integrated Knowledgebase: SPARQL endpoint providing access to integrated data from various sources (incl. SIB resources) relevant for research on SARS-CoV-2
- STRING-Covid: New COVID-19 oriented version enabling users to explore the host-side of the disease, while keeping a focus on putative COVID-19-interacting human proteins
- Corona OMA Browser: Website with all the functionality of the OMA Browser, but for 410 Microbiota-associated host proteins CAPS curated. Not active yet!

**Popular resources**

- UniProtKB
- SWISS-MODEL
- STRING
- PROSITE

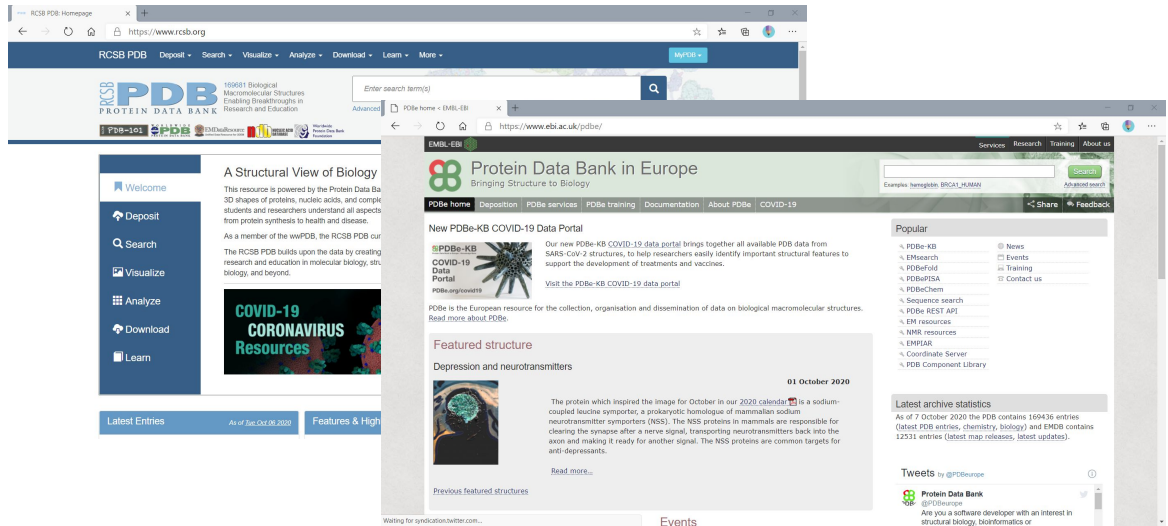
**Latest News**

- New UniProt release - 2020-08-31
- New third party tool and SPARQL queries. More here
- New UniProt data release 2020-07-17 - 2020-06-04
- New release with interaction data from ENVO Pharma and more

[More news] [SIB news]

11 /

# RCSB PDB | PDBe



12 /

## Źródła danych...

DNA

- Sekwencjonowanie DNA
- Sekwencjonowanie RNA
- Sekwencjonowanie mRNA
- Proteom
- Mikromacierze
- Ekelctroforeza wielowymiarowa ze spektrometria masową
- Dane strukturalne
- Krystalografia
- NMR
- Mikroskopia elektronowa
- Metabolom
- Spektrometria MS/MS

13 /

# Omica

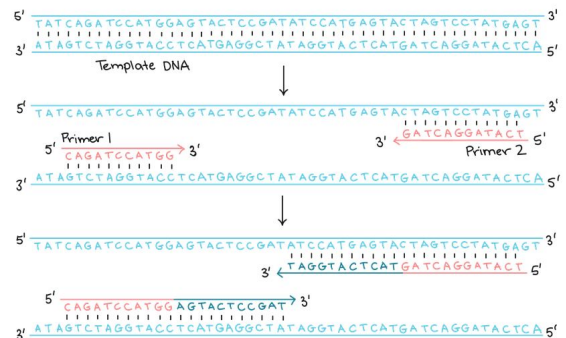
Dane omiczne to dane ze specyficznych eksperymentów naukowych padające poszczególne aspekty aktywności biologicznej na różnych jej poziomach molekularnych:

- Genomika: DNA
- Proteomika: białka
- Metabolomika: małowcząsteczkowe związki chemiczne (np. pochodzące ze szlaków metabolicznych)

# Genomika

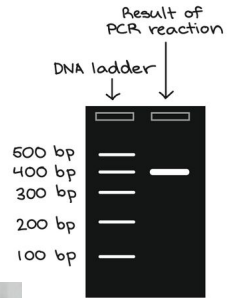
Klasyczne metody sekwencjonowanie DNA wykorzystuje reakcję PCR w celu powielenia matrycowego DNA.

- Denaturacja: Intensywne podgrzanie reakcji, aby rozdzielić, czyli zdenaturować nici DNA. W efekcie otrzymujemy jednoniciowy szablon dla następnego kroku.
- Przyłączenie: Ochłodzenie reakcji, aby startery mogły związać się z komplementarnymi sekwencjami na jednoniciowym matrycowym DNA.
- Wydłużanie: Podniesienie temperatury reakcji, aby polimeraza Taq wydłużała startery, syntezując nowe nici DNA.



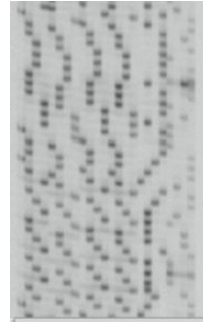
# PCR

Jak z reakcji, która generuje jeden specyficzny produkt otrzymać wiele produktów a z ich wielkości odczytać sekwencję badanego łańcuch DNA?



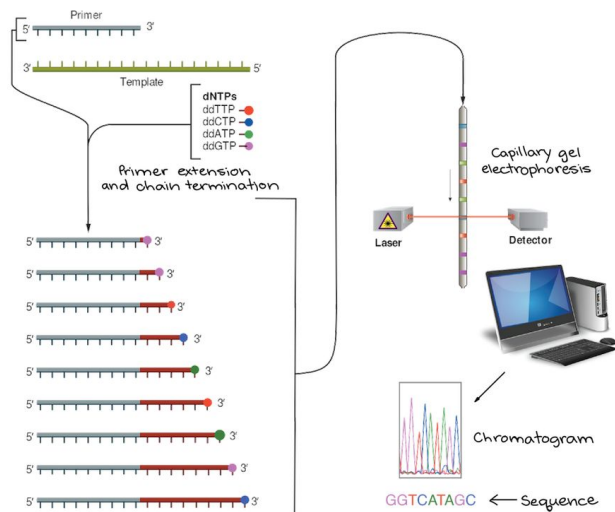
## Sekwencjonowania Sangera

- to proces określania sekwencji (kolejności) nukleotydów (A, T, C i G) w fragmencie DNA.
- docelowy DNA jest kopiowany wiele razy, w wyniku czego powstają fragmenty o różnej długości.
- Fluorescencyjne nukleotydy „terminujące” (kończące wzrost łańcucha) wyznakowują końce fragmentów i umożliwiają określenie sekwencji.



21 /

# Sekwencjonowanie Sangera



23 /

## Dla kogo?



**Leon, postdoc**  
**Goal:** to understand what makes a normally harmless bacterium pathogenic in the lungs of people with cystic fibrosis.  
**Tasks:** "I'm using a combination of transcriptomics, proteomics and metabolomics to understand these pathogenic changes better."

**Problem:**  
zrozumieć czynniki powodujące zmianę bakterii probiotycznej w patogeną u osób ze zdiagnozowanym nowotworem.

**Wymagane dane:**  
-genom bakterii  
-chory i zdrowy proteom  
-chory i zdrowy metabolom

**gDNA, cDNA, rDNA**



**Barend, plant geneticist**  
**Goal:** to identify new crop strains resistant to drought, salt and fungal diseases.  
**Tasks:** "We're doing linkage studies to find out which genes are involved in resistance to different types of stress. We've got genomic and expression QTLs that we need to map on to well-characterised plants."

**Problem:**  
identyfikacja nowego zboża odpornego na suszę, zasolenie i choroby.

**Wymagane dane:**  
-proteom i genom  
-identyfikacja markerów poszukiwanych cech

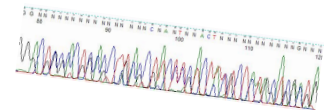


**Ola, clinician-scientist**  
**Goal:** to identify proteomics-based biomarkers in urine for the early detection of bladder cancer  
**Tasks:** "I do mass spectrometry of samples from patients coming in for biopsies. I've found a phosphoprotein that seems to be upregulated in some patients."

**Problem:**  
diagnostyka pacjentów we wczesnym etapie choroby nowotworowej  
**Wymagane dane:**  
-proteom i metabolom pacjenta zdrowego  
-proteom i metabolom pacjenta chorego

27 /

## Kontrola jakości



Garbage In, Garbage Out, GIGO (pol. śmieci na wejściu – śmieci na wyjściu) – w obszarze technologii informacyjnych angielski zwrot mówiący o tym, że nawet gdy program lub procedura przetwarzania były poprawne, wyniki przetwarzania błędnych danych będą błędne.

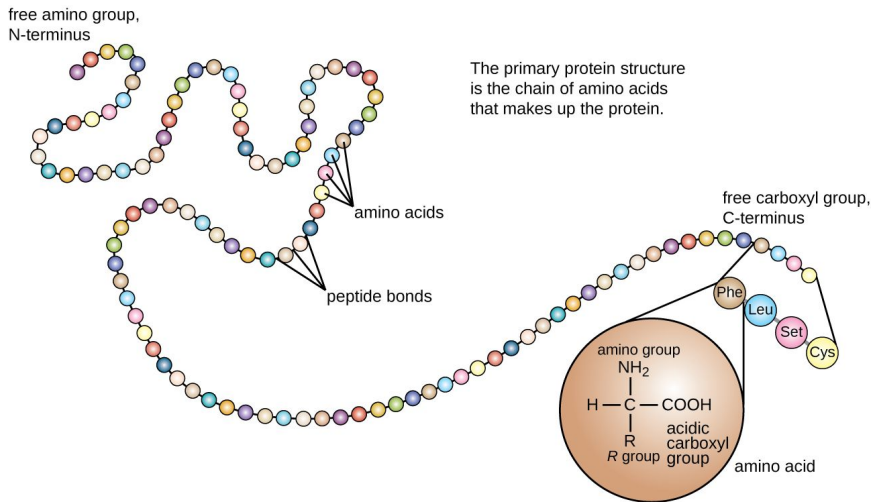
Sekwencjonowanie genomów:

- Pre-assembly processing czyli analiza wstępna
- Base calling czyli identyfikacja właściwej ścieżki odczytu sekwencji
- Genome assembly składanie genomu

29 /

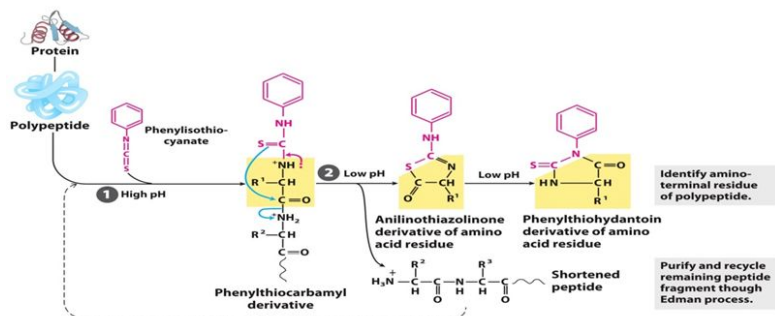


# Sekwencjonowanie białek



Nie ma metody powieliania białka analogicznej do PCR!

# Degradacja Edmana



Procedura wyznaczania sekwencji białka od jego końca N (N-terminal) do końca C (C-terminal).

- Wymaga odpowiednich ilości materiału wyjściowego – białka.
- Materiał ulega zniszczeniu i nie może być ponownie wykorzystany – metoda destrukcyjna.

# Proteom

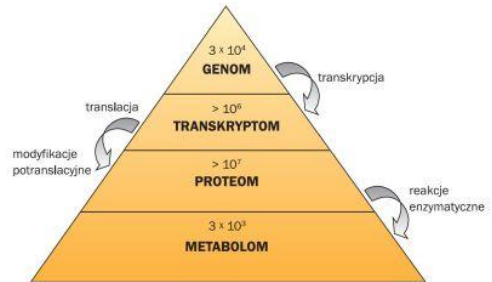
Proteom  
component białkowy komórek kodowany  
przez genom.

Stężenie białek w komórce...

- nie jest prostą funkcją ekspresji genów
- podlega licznym modyfikacjom, które decydują o końcowych właściwościach.

W połączeniu z genomem pozwala:

- identyfikować szlaki metaboliczne zaangażowane w badane procesy
- monitorować procesy patologiczne na poziomie komórkowym



# Profil białkowy

Genom	Proteom	Metabolom
stacyczny	dynamiczny	dynamiczny
możliwość amplifikacji	brak	brak
jednorodny	niejednorodny	niejednorodny
stałe stężenie	zmienne stężenie	zmienne stężenie

Analiza proteomiczna  
musi uwzględnić jego dynamiczny charakter.

- zmiany na poziomie komórkowym, np.: fazy rozwoju komórki;
- zmiany w funkcji czasu;
- zmiany w otoczeniu zewnątrzkomórkowym;
- zmiany w środowisku zewnątrzkomórkowym;

Analiza musi uwzględniać:

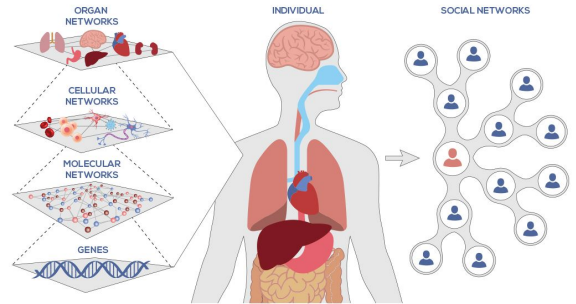
- jednoczesne porównanie dwóch lub więcej próbek;
- przygotowanych w podobny sposób;

Marker proteomiczny (lub profil proteomiczny) to jedna lub więcej cząsteczek charakterystycznych dla badanego procesu.

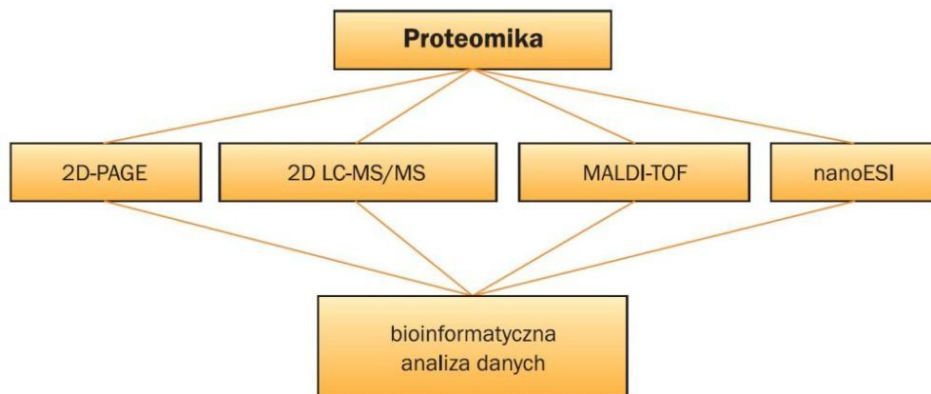
## Biologia systemów

...to całościowe, holistyczne, spojrzenie na funkcjonowanie komórek czy organizmów opierające się na analizie wycinkowych prac badawczych umożliwiające modelowanie układów biologicznych.

Biologia systemów wykorzystuje w równym stopniu informacje genetyczne, proteomiczne i metabolomiczne poprzez analizę funkcjonalną (analizę funkcji cząsteczek).



## Strategie badań proteomicznych



## Strategia identyfikacji białek

Metody proteomiczne...

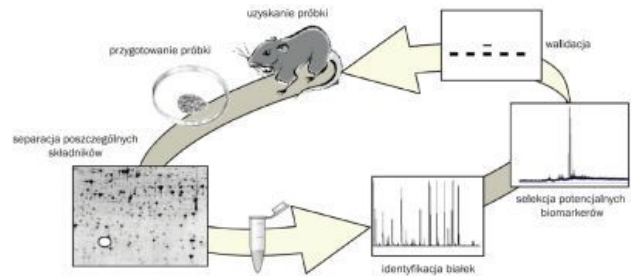
-oczyszczanie i separacja (materiału biologicznego)

-identyfikacja (komponentów)

-analiza i interpretacja wyników

-weryfikacja zebranych danych...

...w oparciu o unikalny, selektywny test (np.: ELISA, reakcja enzymatyczna, badania receptorowe, itp.)

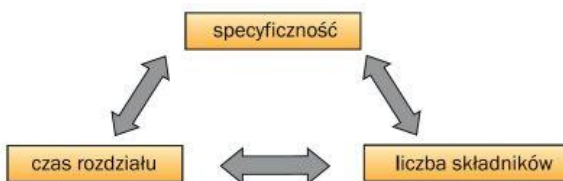


## Wiarygodność analizy proteomicznej

Etapy analizy są potencjalnym źródłem błędów...

-przygotowanie powtarzalnych próbek (wielu) do analizy;

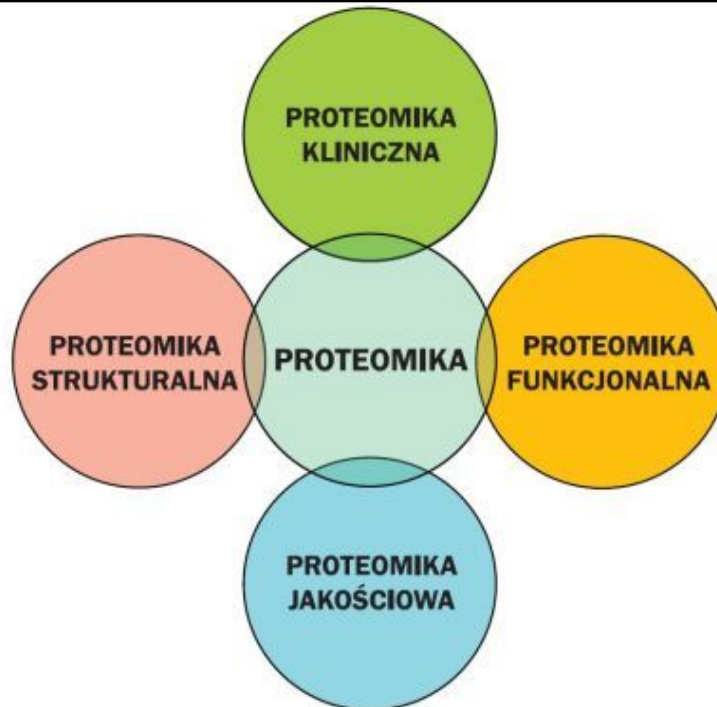
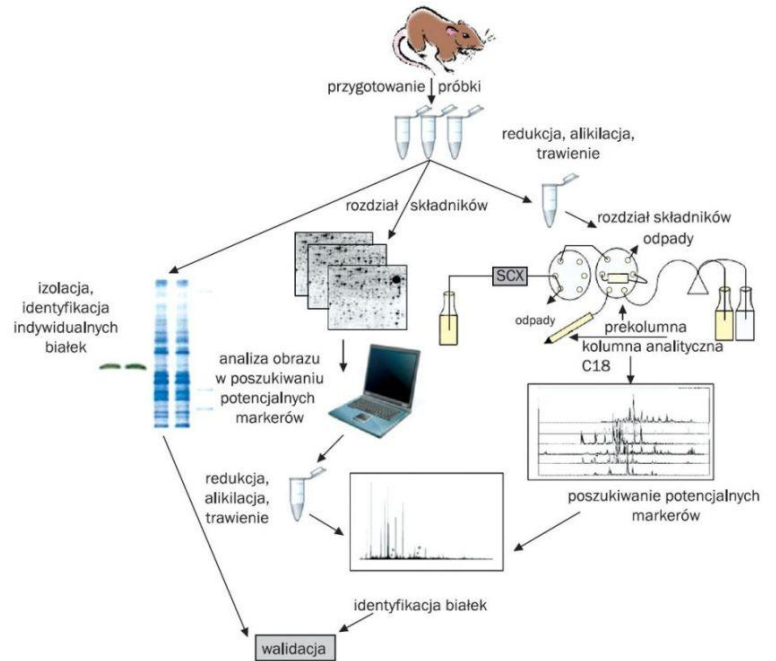
Skrócenie czasu separacji mieszaniny wpływa na selektywność oznaczeń.



## Wybór strategii

Materiał biologiczny i cel analizy determinuje strategię badawczą.

Typowe materiały do badań:  
-surowica/plazma krwi  
12 białek (albumina, immunoglobuliny, itp.) stanowią 96% wszystkich białek!  
-mocz  
-ślina  
-płyn mózgowo-rdzeniowy  
-...inne roztwory biologiczne.



## Strategie badawcze

Bottom-up...

analizie poddawane są sekwencje peptydowe uzyskane w wyniku enzymatycznego trawienia białka.

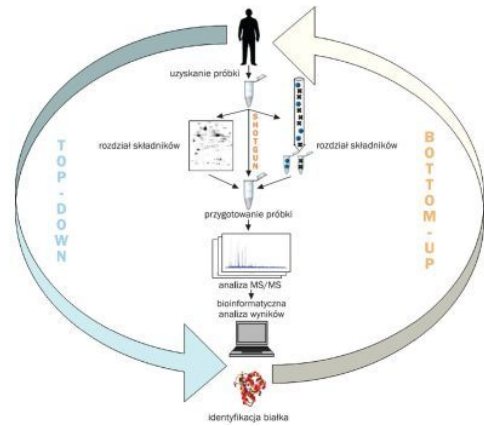
Wykorzystuje tandemową spektrometrię mas (MS/MS), lub mapy peptydowe (PMF, ang. *peptide mass fingerprinting*).

Shotgun...

analizowana próbka poddawana jest trawieniu za pomocą wybranych enzymów proteolitycznych.

Top-down...

natywne białka poddaje się analizie bez ich uprzedniej fragmentacji.



## Przygotowanie materiału do badań proteomicznych

Zadania:

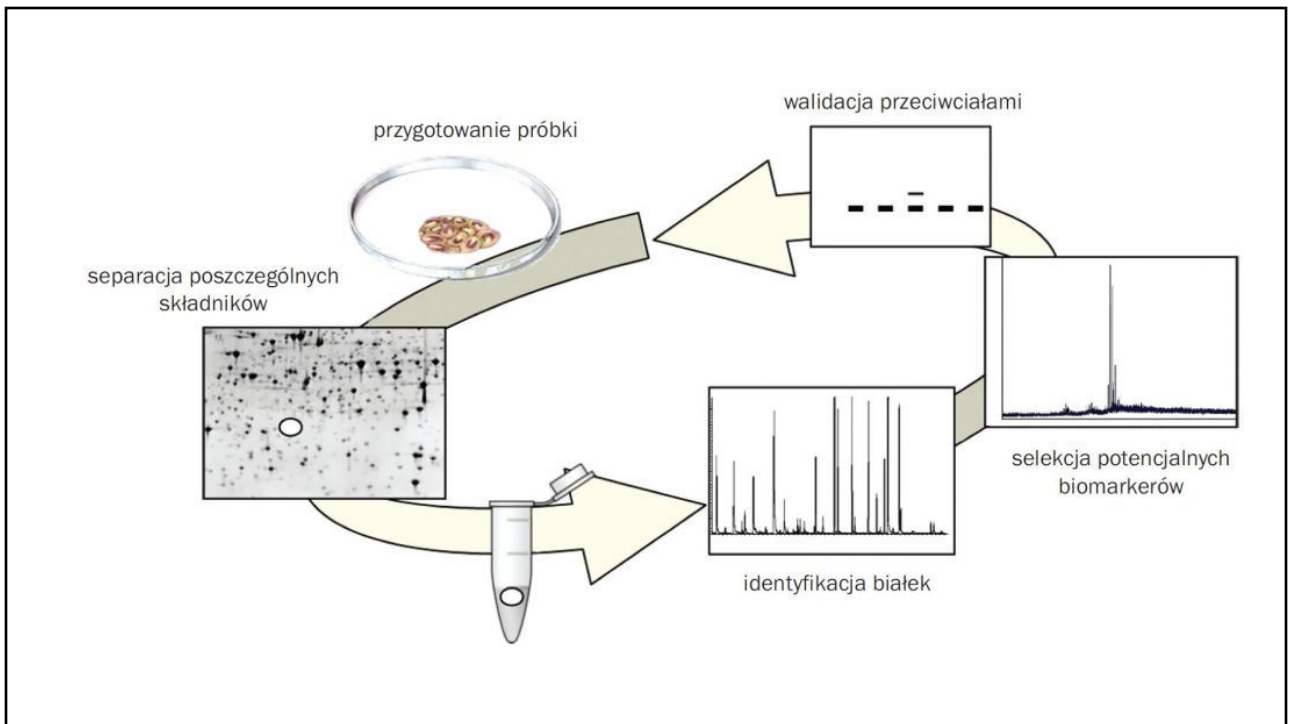
- Redukcja objętości
- Eliminacja zanieczyszczeń i niepożądanych składników

Cele:

- Jednorodny
- Homogeniczny
- Stabilny  
...materiał do analizy

Materiał pobrany należy przechowywać w temperaturach kriogenicznych!





## Źródła próbek biologicznych

**Tkanki zwierzęce** – minimalizacja heterogenności poprzez usunięcie różnicowych elementów tkankowych. Homogenizacja mechaniczna i wirowanie różnicowe pozwala na usunięcie organelli komórkowych.

**Tkanki roślinne** – wielokrotna liza z wykorzystaniem zamrażania i ultradźwięków. Strącanie komponentów białkowych w acetonie lub 10% TCA.

**Hodowle komórkowe** – usunięcie pożywki.

**Płyny ustrojowe** – redukcja objętości, usunięcie zanieczyszczeń za pomocą chromatografii.

**Bakterie** – mechaniczna degradacja w celu usunięcia ściany komórkowej, buforów do lizy.

**Wirusy** – wirowanie różnicowe, rozpuszczenie i homogenizacja.

# Rozpuszczanie białek

**Czynniki chaotropowe** – przeciwdziałają tworzeniu się agregatów (redukcja wiązań wodorowych i oddziaływań hydrofilowych).  
Mocznik, tiomocznik...

**Detergenty** – niwelują oddziaływania hydrofobowe.  
-jonowe, np.: SDS

-niejonowe, np.: Triton X-100, NP-40

-amfoteryczne, np.: CHAPS, CHAPSO, ASB-14

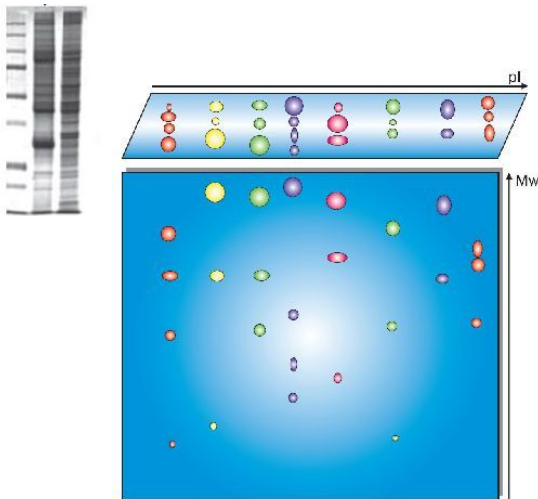
**Odczynniki redukujące** – redukcja grup sulfhydrylowych.

-DDT (ditiotrietol), DTE (1,4-ditioerytrol), 2-merkaptoetanol, tributylfosfina (TBP), tris(2-karboksyetylo)fosfina (TECEP)

**Inhibitory proteaz** – niekontrolowana aktywność proteaz endogennych komplikuje profil białkowy.

-fluorek fenylometylosulfonowy (PMSF), fluorek aminoetylobenzylsulfonowy (AEBSF), keton chlorometylowy N-tosylo-L-fenylalaniny (TLCK), benzamidyna, EDTA, amastatyna

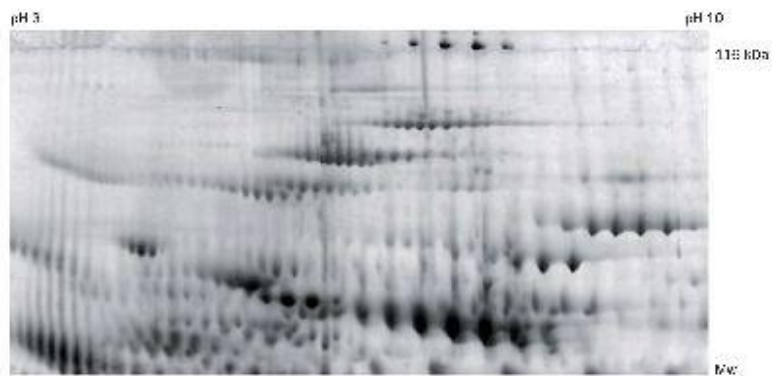
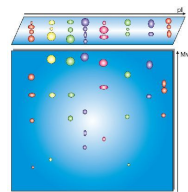
# Elektroforeza dwuwymiarowa (2-DE)



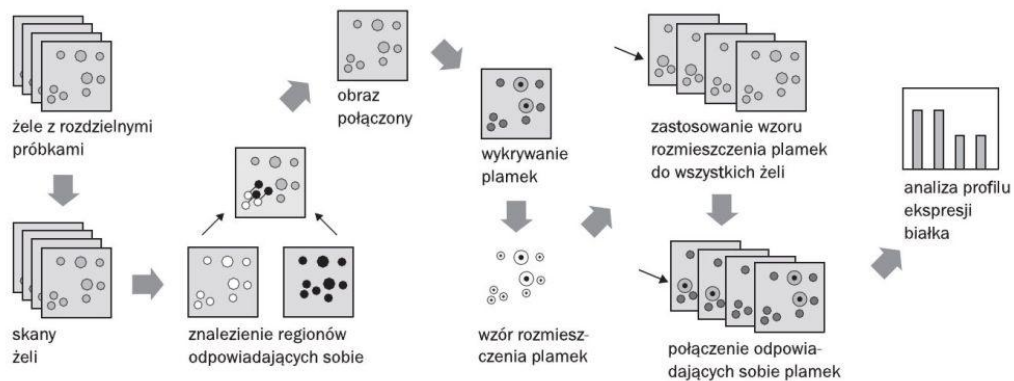
Rozdzielanie mieszanin białkowych wymach zastosowania specjalistycznych metod elektroforetycznych i chromatograficznych.

Celem jest identyfikacja \_pojedynczego\_ białka pochodzącego z mieszaniny.





## Analiza porównawcza żeli



## Identyfikacja białek w mieszaninach

Całościowa analiza proteomu wiąże się z koniecznością \_jednoczesnego\_  
-rozdziatu,  
-identyfikacji,  
-oznaczenia wszystkich białek badanego układu biologicznego.

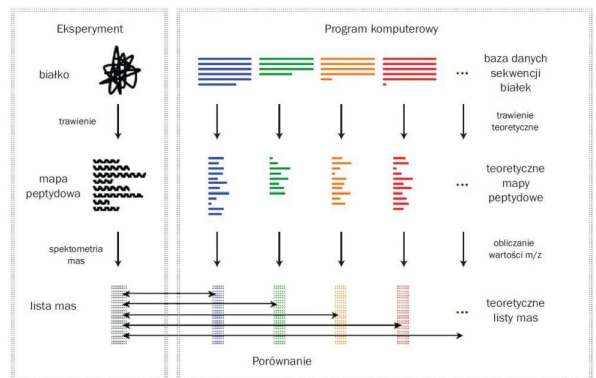
Metody analizy:

- wysokorozdzielcze w aspekcie separacji,
- wydajne i skuteczne w zakresie identyfikowanych składników białkowych.

## „Odcisk palca” mapy peptydowej

Metoda odcisku palca mapy peptydowej  
(PMF, ang. *peptide mass fingerprinting*)

Każde białko ma unikatową sekwencję aminokwasową.  
Wykorzystanie unikalnych, charakterystycznych cech danej sekwencji pozwala na jej identyfikację.



**MASCOT Peptide Mass Fingerprint**

Your name  Email

Search title

Database(s)  Enzyme   
    Allow up to  missed cleavages

Taxonomy

Fixed modifications

Variable modifications

Protein mass  kDa Peptide tol.  Da

Mass values  MH+  CM  M-H  Monoisotopic  Average

Data file  No file chosen

Query

Data input

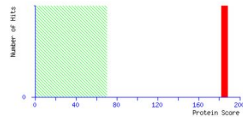
Decoy  Report top  hits

**MASCOT Search Results**

User :  
 Email :  
 Search title : Peptide Mass Fingerprint Example  
 Database : SwissProt 2019\_10 (561336 sequences; 20185828 residues)  
 Timestamp : 9 Jan 2020 at 11:21:29 GMT  
 Top Score : 185 for **PHL\_HUMAN**, Protein PHL OS=Homo sapiens OX=9606 GN=PHL PE=1 SV=3

**Mascot Score Histogram**

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event. Protein scores greater than 70 are significant ( $p < 0.05$ ).



**Concise Protein Summary Report**

Format As

Significance threshold  $p < 0.05$  Max number of hits

Preferred taxonomy

- PHL\_HUMAN** Mass: 97489 Score: 185 Expect: 1.8e-13 Matches: 16  
 Protein PHL OS=Homo sapiens OX=9606 GN=PHL PE=1 SV=3

**SLC4A3\_BACUL** Mass: 37935 Score: 49 Expect: 6.7 Matches: 5  
 Protein RecA OS=Hoplodactylus castertonii (strain DSM 13941 / MCB) OX=383372 GN=recA PE=3 SV=1

**IF5A\_P2910** Mass: 14588 Score: 47 Expect: 11 Matches: 4  
 Translation Initiation Factor 5A OS=Proteobium neutropilum (strain DSM 2330 / JCM 9278 / V245ta) OX=444157 GN=IF5A PE=3 SV=1

**MSDC\_C4\_L2** Mass: 22438 Score: 44 Expect: 20 Matches: 4  
 Probable nicotinate-nucleotide adenyltransferase OS=Chlorobium limicola (strain DSM 245 / MRC 183883 / 6330) OX=299315 GN=msdD PE=3 SV=1

**MSL8\_MOUSE** Mass: 23928 Score: 42 Expect: 39 Matches: 4  
 Inactive ribonuclease-like protein 18 OS=Equus caballus OX=9796 GN=HNA5E18 PE=2 SV=2

**IL3C\_ID110** Mass: 52994 Score: 42 Expect: 39 Matches: 5  
 IIP- $\alpha$ -cytotoxic- $\gamma$ -alanine ligase OS=Ictioneira ichthyensis (strain ATCC BAA-735 / DSM 15497 / L2-Tr) OX=283942 GN=msrC PE=3 SV=1

**TR6B\_HUMAN** Mass: 13327 Score: 41 Expect: 41 Matches: 3  
 T cell receptor gamma variable 8 OS=Homo sapiens OX=9606 GN=TR6B PE=1 SV=1

**CLP81\_BIFID** Mass: 23812 Score: 40 Expect: 54 Matches: 4  
 ATP-dependent Clp protease proteolytic subunit 1 OS=Bifidobacterium longum (strain NCC 2785) OX=208672 GN=clpP PE=3 SV=1

**SLC4A3\_BACUL** Mass: 64475 Score: 40 Expect: 59 Matches: 6  
 Na<sup>+</sup>/surface glycoprotein 6 OS=Human metoreovirus (isolate Canada goose/Minnesota/15a/2001) OX=652954 GN=SLC4A3 PE=3 SV=1

**ISPF\_C4102** Mass: 41632 Score: 40 Expect: 69 Matches: 5  
 Bifunctional enzyme IspD/IspF OS=Campylobacter jejuni subsp. jejuni serotype O:6 (strain 81116 / NCTC 11828) OX=407148 GN=IspDF PE=3 SV=1

**ISPF\_C4101** Mass: 41664 Score: 40 Expect: 69 Matches: 5  
 Bifunctional enzyme IspD/IspF OS=Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11368) OX=192222 GN=IspDF PE=1 SV=1

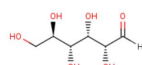
**ISPF\_C4103** Mass: 41663 Score: 40 Expect: 69 Matches: 5  
 Bifunctional enzyme IspD/IspF OS=Campylobacter jejuni (strain RM1221) OX=195989 GN=IspDF PE=3 SV=1

# Metabolomika

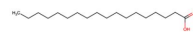
Zbiór danych obejmujących wszechstronną (całościową) analizę jakościową i ilościową, charakterystykę i identyfikację endogennych, drobnocząsteczkowych związków chemicznych.

Human Metabolome Database  
(<https://hmdb.ca/>)

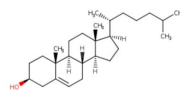
- 114,215 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant ( $> 1 \mu\text{M}$ ) or relatively rare ( $< 1 \text{nM}$ ).



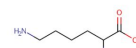
glucose  
(a sugar)



stearic acid  
(a fatty acid)



cholesterol  
(a lipid)



lysine  
(an amino acid)



# HMDB

- 1) Dane chemiczne - chemical data,
- 2) Dane kliniczne - clinical data,
- 3) Dane biochemiczne/biologii molekularnej - molecular biology/biochemistry data

Każda zindeksowana cząsteczka posiada przypisane właściwości:

- Chemiczne,
- Strukturalne,
- Biologiczne.

Dla każdej cząsteczki przypisany jest jeden lub więcej enzymów i białek transportowych.

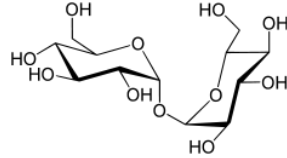




## Metabolom jest dynamiczny

Skład metabolomu jest zmienny i zależy:

- Od stanu komórki,  
*Odmienne od genomu, podobnie do proteomu.*
- Specyficznej aktywności enzymów,
- Różnorodności metabolicznej organizmów,  
*Trehaloza – organiczny związek chemiczny z grupy cukrów, disacharyd złożony z dwóch cząsteczek glukozy połączonych wiązaniem O-glikozydowym ( $\alpha$ ,D-glukopiranozylo-(1 $\rightarrow$ 1)- $\alpha$ ,D-glukopiranozyd). Jest głównym cukrem hemolimfy owadów, znajduje się także w grzybach i drożdżach.*
- Aktualnego stanu metabolicznego komórki.



## Metabolom ma charakter...

...hierarchiczny

Metablom komórki jest fragmentem metabolomu organizmu.



...sekwencyjny

W badaniach szlaku metabolicznego wystarczy określenie jednego bądź kilku metabolitów kluczowych odpowiedzialnych za daną reakcję.



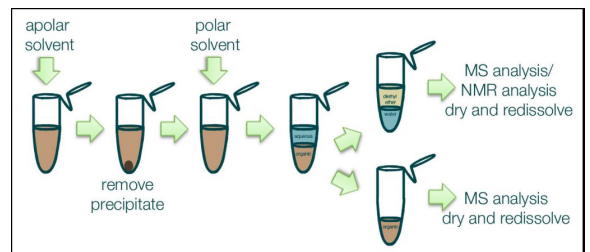
## Strategie analityczne

1. Analiza wybranego metabolitu lub grupy metabolitów (*ang. targeted analysis*).  
*Wysoka czułość i selektywność.*
2. Analiza ilościowa grupy metabolitów (*ang. metabolite profiling*).  
*Wysoka czułość, niska selektywność.*
3. Metabolomika – kompleksowa ilościowa analiza organizmu lub jego wycinka.  
*Niska czułość, niska selektywność.*
4. Analiza „odcisku palca” (*ang. metabolite profiling*).  
*Niska czułość, niska selektywność. Umożliwia klasyfikację całego profilu a nie pojedynczych związków.*

## Techniki analityczne

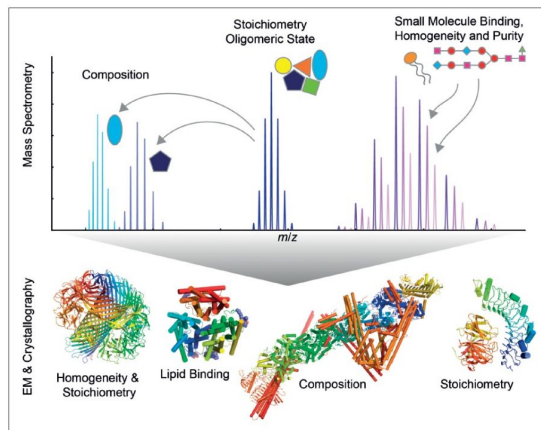
Rodzaj próbki determinuje strategię identyfikacyjną i przygotowawczą.  
Przygotowanie próbki może wymagać np.:

- Precypitacji białek
  - Ultrafiltracji
  - Ekstrakcji
  - Frakcjonowania
  - ...innych.
- 
- Spektrometria mass wspomagana technikami separacji
  - NMR
  - Detekcja elektrochemiczna
  - Spektroskopia IR
  - Spektroskopia UV
  - Spektroskopia fluorescencyjna
  - ...inne metody analityczne



	Nuclear magnetic resonance (NMR)	Mass spectrometry (MS)
<b>Sensitivity</b>	Low	High
<b>Reproducibility</b>	Very high	Average
<b>Number of detectable metabolites</b>	30-100	300-1000+ (depending on whether GC-MS or LC-MS is used)
<b>Targeted analysis</b>	Not optimal for targeted analysis	Better for targeted analysis than NMR
<b>Sample preparation</b>	Minimal sample preparation required	More complex sample preparation required
<b>Tissue extraction</b>	Not required – tissues can be analysed directly	Requires tissue extraction
<b>Sample analysis time</b>	Fast – the entire sample can be analysed in one measurement	Longer than NMR – requires different chromatography techniques depending on the metabolites analysed
<b>Instrument Cost</b>	More expensive and occupies more space than MS	Cheaper and occupies less space than NMR
<b>Sample Cost</b>	Low cost per sample	High cost per sample

## Spektometria mass

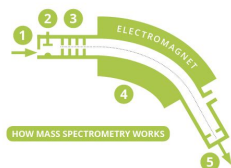




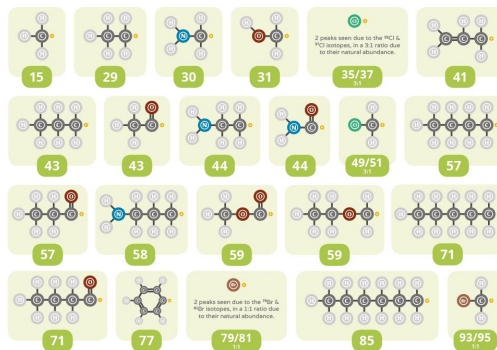
# Spektometria mass

## A GUIDE TO INTERPRETING MASS SPECTRA

Mass spectrometry is an analytical technique that allows us to measure the masses of atoms and molecules. The most important peak in a mass spectrum is the molecular ion peak, which can be used to determine the mass of the molecule, but fragment ions can also provide information on chemical structure.

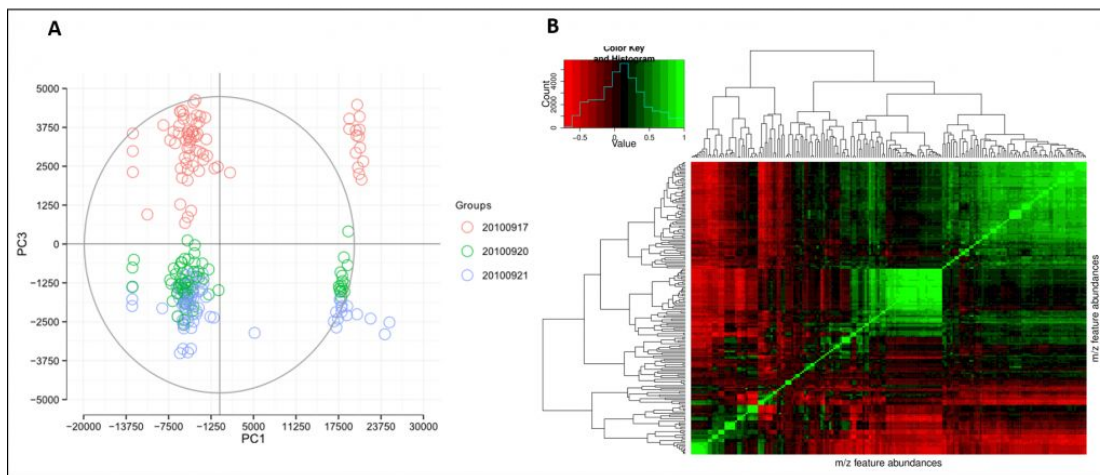


- 1 The sample is introduced to the mass spectrometer. Only very small samples are required. A heater is often present to vapourise the sample.
- 2 An electron gun ionises molecules in the sample by knocking out electrons, producing positive ions. Some molecules break into smaller ions & fragments.
- 3 The positive ions generated are passed through an electric field which accelerates them into a magnetic field generated by an electromagnet.
- 4 As the positive ions pass through the magnetic field, they are deflected. Lighter ions are deflected more than heavier ions, as are those with higher charges.
- 5 The positive ions hit charged plates & accept electrons, creating a signal. The more ions that hit, the greater the signal. The output is a complex stick diagram.



Above are shown a selection of common fragment ions seen in mass spectra, along with their masses. Note that the structures shown are general representations, and it can also be possible for isomeric structures (those with the same constituent atoms, but a different structure) to cause the peaks in spectra. There are also many more fragments possible than those shown, but knowledge of these should suffice to interpret spectra of most simple molecules.

© COMPOUND INTEREST 2015 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem  
 This graphic is shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.



# Zastosowanie danych metabolicznych

- Rolnictwo  
*Identyfikacja nowych, wydajnych odmian roślin uprawnych.*
- Medycyna  
*Małoinwazyjna kompleksowa ocena stanu pacjenta i ukierunkowane leczenie*
  - Diagnostyka
  - Medycyna spersonalizowana

# Zasoby internetowe

The screenshot displays the MetaboLights website. At the top, there is a navigation bar with links for Home, Create Studies, Create Compounds, Create Species, Download, Help, Get us feedback, and About. A search bar is located in the top right corner. The main content area features a large heading 'MetaboLights' and a brief description: 'MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolite experiments. MetaboLights is the recommended metabolomics repository.' Below this, there are three main sections: 'Study' (Browse, Quick Search, MetaboLight Labs), 'Compound Library' (Compounds, Species), and 'Training' (Training Online, Quick Tour). On the right side, there is a 'Tweets by @metabolights' section showing three tweets from the account @metabolights. At the bottom of the page, there is a footer with a privacy notice: 'This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice.' and a 'Log out, distrust this browser' link.

# Zasoby internetowe

The screenshot shows the ChEBI (Chemical Entities of Biological Interest) website. The header includes the EMBL-EBI logo and navigation links for Home, Advanced Search, Browse, Documentation, Download, Tools, and About ChEBI. A search bar is prominently displayed with a search button. Below the search bar, there are sections for Documentation, Downloads, and News. The 'Entity of the month' section features a chemical structure of Metabdehyde and a tweet from ChEBI. A footer notice states: 'This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice and Terms of Use. I agree, dismiss this banner.'

# Zasoby internetowe

The screenshot shows the Reactome website homepage. At the top, there is a search bar with the text 'Find Reactions, Proteins and Pathways' and a search button. Below the search bar, there are four main service icons: Pathway Browser, Analyze Data, ReactomeFIViz, and Documentation. Each icon has a brief description of its function. Below these icons, there is a section titled 'USE REACTOME GRAPH DATABASE IN YOUR PROJECT' with a 'LEARN MORE' button. This section contains two columns of text: 'Why Reactome' and 'Tweets'. The 'Why Reactome' section describes the database as a free, open-source, curated and peer-reviewed pathway database. The 'Tweets' section features a tweet from Reactome highlighting their collaboration with the #COVID19 Disease Map group.

