

Bioinformatyka

Andrzej Łyskowski, dr inż.
Katedra Biotechnologii i Bioinformatyki
andrzej.lyskowski@prz.edu.pl
H-237

Warunki zaliczenia

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.

Uwagi:

Średnia może być wyciągnięta tylko jeżeli wszystkie kryteria zaliczenia dla poszczególnych składowych modułu są spełnione.

Wartości współczynników podane są do wiadomości studentów na pierwszych zajęciach.

Zakres materiału

Obszar zainteresowania badań bioinformatycznych.

Źródła danych.

Bioinformatyczne bazy danych.

Wyszukiwanie i analiza sekwencji.

- Dopasowanie sekwencji.
- Drzewa filogenetyczne.

Bioinformatyka strukturalna.

Szlaki metaboliczne i biologia systemów

slido



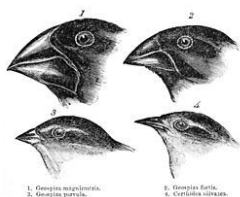
Bioinformatyka to...

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

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.



1. *Ammodramus melanoleucus*
 2. *Ammodramus melanoleucus*
 3. *Ammodramus melanoleucus*
 4. *Ammodramus melanoleucus*



5

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

6

Podstawowe zasoby internetowe

Bazy danych bioinformatycznych dostępne są online.

Większość kluczowych zasobów ma charakter otwarty – są one dostępne bezpłatnie dla badaczy z całego świata.

Bazy danych można podzielić ze względu na charakter zgromadzonych danych na:

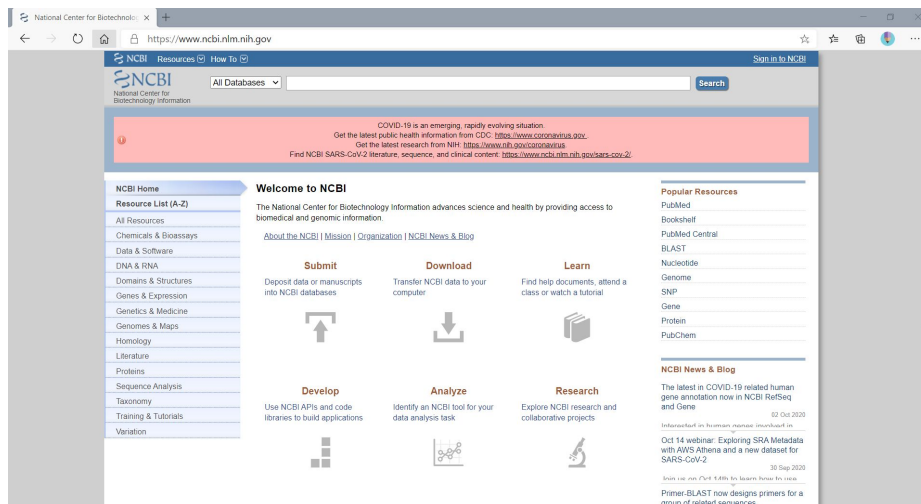
- Ogólne,
- Szczegółowe.

Wiele projektów badawczych generuje wysoce specjalistyczne bazy danych, których wykorzystanie wymaga fachowej wiedzy i wykorzystania szczegółowych narzędzi

Bazy danych globalnych są zarządzane przez konsorcja i ich poprawność weryfikowana w miarę napływających danych.

7 /

National Center for Biotechnology Information



8 /

EMBL European Bioinformatics Institute

The screenshot shows the EMBL-EBI website homepage. At the top, there's a navigation bar with links to 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 featured topic section highlights "COVID-19 Data Portal" and "SARS-CoV-2". A "Latest news" section is also present on the right side of the featured topics.

EMBL-EBI response to COVID-19

To help protect staff and visitors from the coronavirus outbreak, EMBL-EBI closed its premises on 18 March at 17:00. Our data resources and tools will continue to function as normal. [Read more](#)

Featured topic

COVID-19 Data Portal

SARS-CoV-2

Latest news

29 Sep 2020

9 /

UniProt

The screenshot shows the UniProt website homepage. At the top, there's a navigation bar with links to UniProtKB, BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. The main header features the UniProt logo and a search bar. Below this, a mission statement is provided: "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." The page is divided into several sections: UniProtKB (Swiss-Prot and TrEMBL), UniRef, UniParc, and Proteomes. A "Supporting data" section lists various resources like literature citations, taxonomy, and subcellular locations. A "News" section on the right provides updates on UniProt releases and protein spotlights.

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.

UniProtKB

UniProt Knowledgebase

Swiss-Prot (563,552)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (195,104,019)

Automatically annotated and not reviewed.

Records that await full manual annotation.

UniRef

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt knowledgebase (including isoforms) and selected UniParc records.

UniParc

UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes

A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

Supporting data

Literature citations

Taxonomy

Subcellular locations

Cross-ref. databases

Diseases

Keywords

News

New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

[View SARS-CoV-2 Proteins and Receptors](#)

Getting started

UniProt data

Download latest release

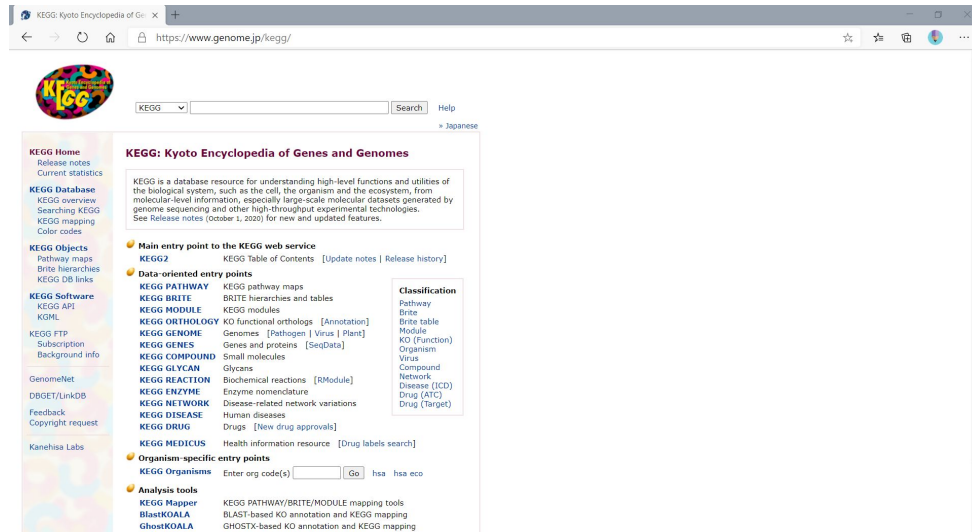
Protein spotlight

Either You, Or Me

10 /

KEGG

Kyoto Encyclopedia of Genes and Genomes



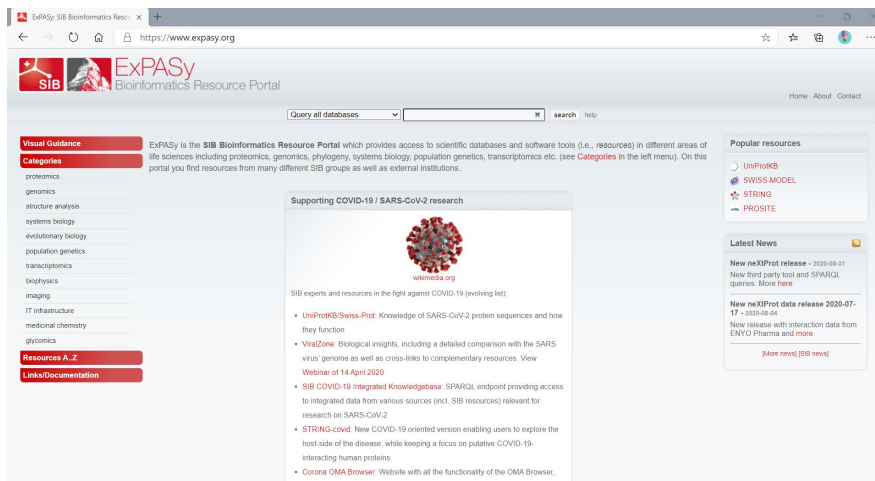
The screenshot shows the KEGG website homepage. The header includes the KEGG logo and a search bar. The main content area is divided into several sections:

- KEGG Home:** Release notes, Current statistics.
- KEGG Database:** KEGG overview, Searching KEGG, KEGG mapping, Color codes.
- KEGG Objects:** Pathway maps, Brite hierarchies, KEGG DB links.
- KEGG Software:** KEGG API, KEGG FTP, Subscription, Background info.
- GenomeNet:** DBGET/LinkDB, Feedback, Copyright request, Kanehisa Labs.
- KEGG: Kyoto Encyclopedia of Genes and Genomes:** A detailed description of KEGG as a database resource for understanding high-level functions and utilities of the biological system, 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 Table of Contents [Update notes | Release history].
- Data-oriented entry points:**
 - KEGG PATHWAY:** KEGG pathway maps.
 - KEGG BRITE:** BRITE hierarchies and tables.
 - KEGG MODULE:** KEGG modules.
 - KEGG ORTHOLOGY:** KO functional orthology [Annotation].
 - KEGG GENOME:** Genomes [Pathogen | Virus | Plant].
 - KEGG GENES:** Genes and proteins [SeqData].
 - KEGG COMPOUND:** Small molecules.
 - KEGG GLYCAN:** Glycans.
 - KEGG REACTION:** Biochemical reactions [RModule].
 - KEGG ENZYME:** Enzyme nomenclature.
 - KEGG NETWORK:** Disease-related network variations.
 - KEGG DISEASE:** Human diseases.
 - 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.
- Classification:** Pathway, Brite, Brte table, Module, KO (Function), Organism, Virus, Compound, Network, Disease (ICD), Drug (ATC), Drug (Target).

11 /

ExPASy

Swiss Institute of Bioinformatics

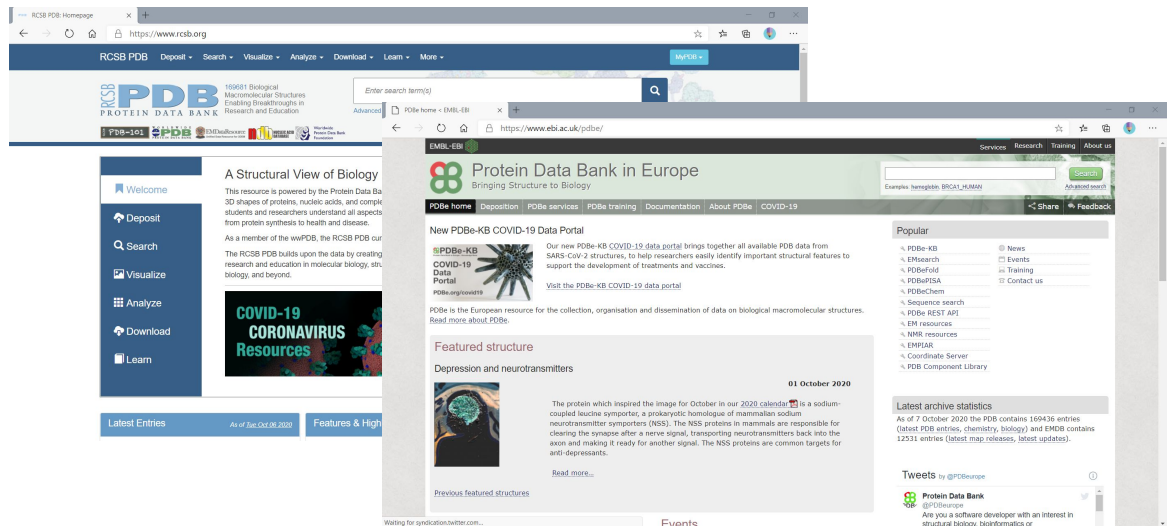


The screenshot shows the ExPASy website homepage. The header includes the ExPASy logo and a search bar. The main content area is divided into several sections:

- Visual Guidance:** ExPASy is the SIB Bioinformatics Resource Portal which provides access to scientific databases and software tools (i.e., resources) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.
- Categories:**
 - proteomics
 - genomics
 - structure analysis
 - systems biology
 - evolutionary biology
 - population genetics
 - transcriptomics
 - bioinformatics
 - imaging
 - IT infrastructure
 - medicinal chemistry
 - glycomics
- Resources A-Z**
- Links/Documentation**
- 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-conv: 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 now with COVID-19 related information. Visit this resource CAPED database. Read also about
- 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 - 2020-08-04: New release with interaction data from ENVO Pharma and more. [More news] [SIB news]

12 /

RCSB PDB | PDBe



13 /

Źródła danych...

DNA

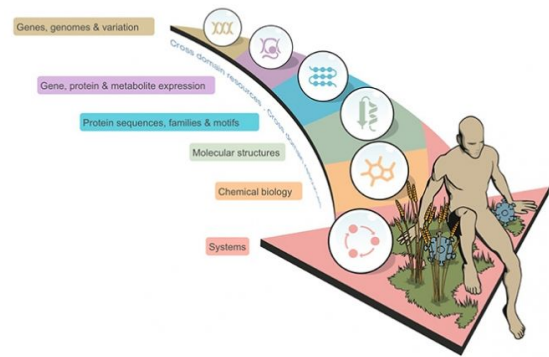
- Sekwencjonowanie DNA
- Sekwencjonowanie RNA
- Sekwencjonowanie mRNA
- Proteom
- Mikromacierze
- Elektryforeza wielowymiarowa ze spektrometrią masową
- Dane strukturalne
- Krystalografia
- NMR
- Mikroskopia elektronowa
- Metabolom
- Spektrometria MS/MS

14 /

Komputery w naukach biologicznych...

Wykorzystanie danych biologicznych:

- W medycynie
 - Informatyka medyczna
 - Wykorzystanie technik informatycznych w systemie służby zdrowia
 - Efektywne wykorzystanie danych biologicznych w medycynie
 - Aplikacja danych medycznych w celu rozwiązywania problemów medy
- W naukach przyrodniczych
 - Przetwarzanie danych z eksperymentów biologicznych
 - Faktycznych danych eksperymentalnych
 - Danych obrazowych
 - Danych omicznych



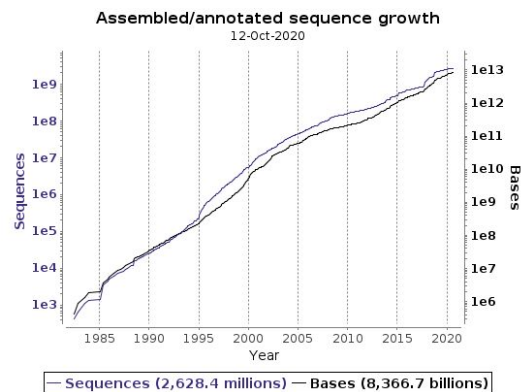
15 /

Biologia *in silico*

Biologia to nauka bazująca na obserwacjach.

Gwałtowny wzrost ilości danych o sprawił iż obserwacje biologiczne stały się przeliczalne.

- 1953 model struktury DNA (Watson-Crick)
- 1981 sekwencjonowanie mitochondrialnego DNA człowieka: 16 568 bp
- 1990 rozpoczęcie International Human Genome Project (15 lat trwania)
- 1996 mapa genetyczna genomu człowieka wysokiej rozdzielczości (markery co 600 000 bp)
- 2000 udostępnienie schematu sekwencji genomu człowieka



16 /

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."



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."



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:
zrozumieć czynniki
powodujące zmianę bakterii
probiotycznej w patogenną
u osób ze zdiagnozowanym
nowotworem.
Wymagane dane:
-genom bakterii
-chory i zdrowy proteom
-chory i zdrowy metabolom

Problem:
identyfikacja nowego zboża
odpornego na suszę, zasolenie i
choroby.
Wymagane dane:
-proteom i genom
-identyfikacja markerów
poszukiwanych cech

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

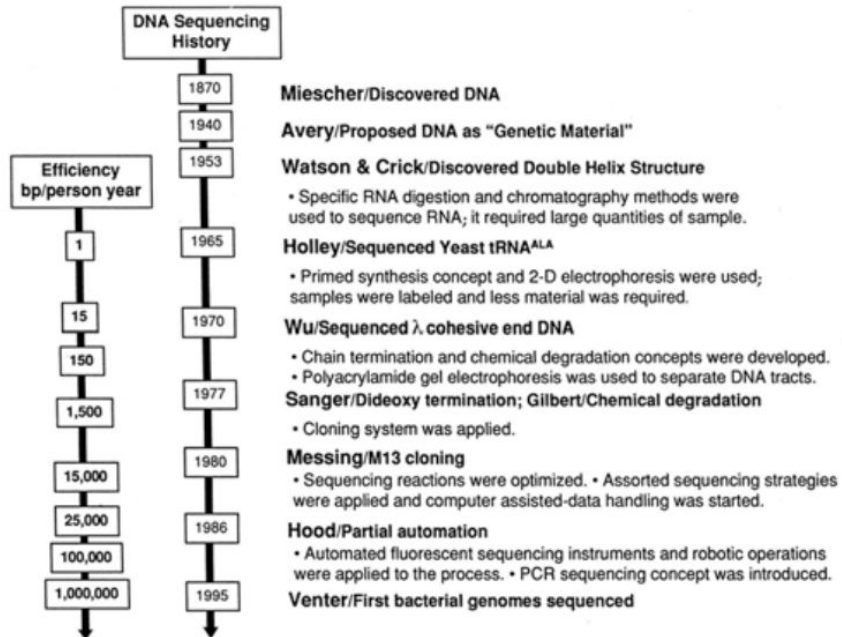
17 /

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)

18 /

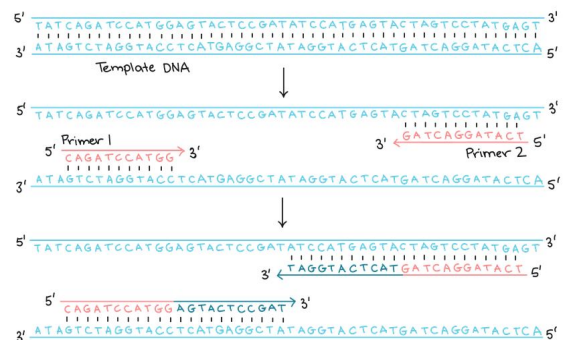


19 /

Genomika

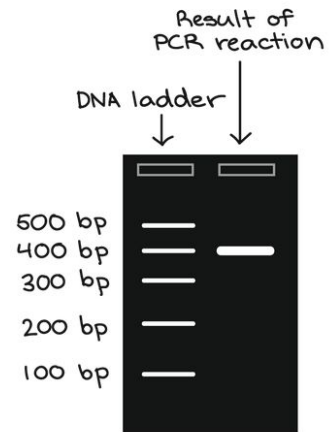
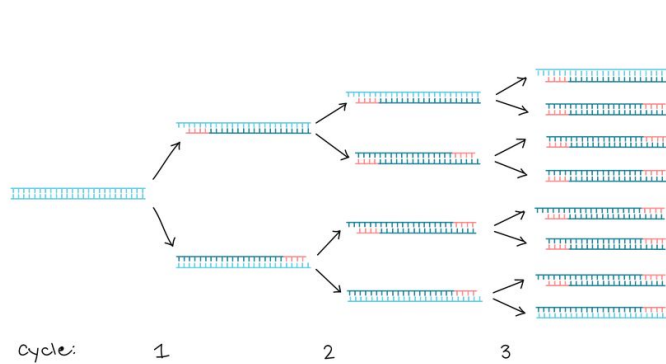
Klasyczne metody sekwencjonowania DNA wykorzystuje reakcję PCR w celu powielenia matrycowego DNA.

- 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.



20 /

PCR



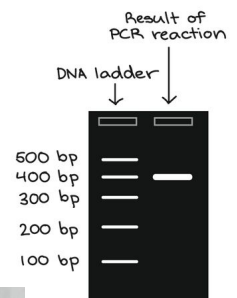
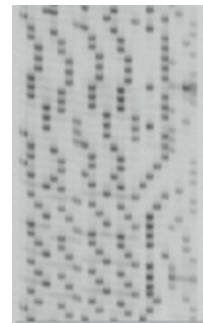
21 /

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) we 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.

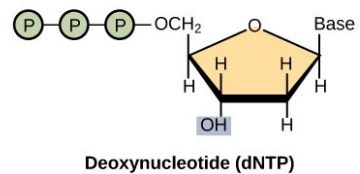
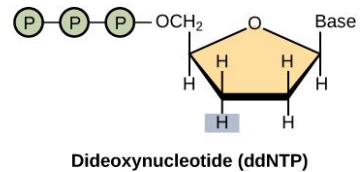


22 /

Sekwencjonowanie Sangera

Reakcja PCR

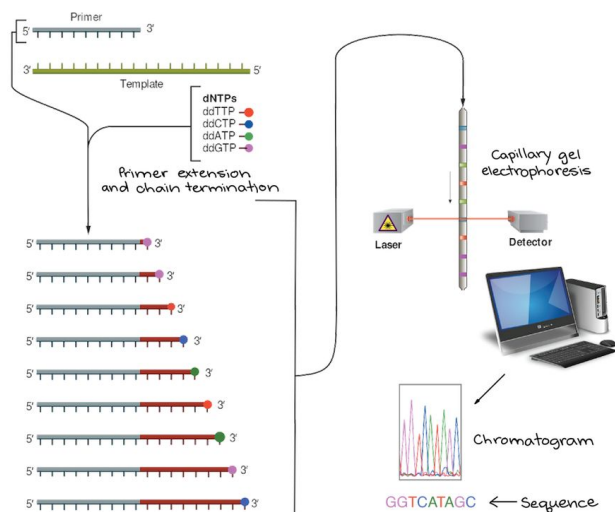
- Enzym polimeraza DNA
- Primer, który jest krótkim fragmentem pojedynczego DNA wiążącego się z DNA matrycowym i działającym jako „starter” polimerazy
- Cztery nukleotydy DNA (dATP, dTTP, dCTP, dGTP)
- DNA matrycowe do sekwencjonowania



Sekwencjonowania Sangera:

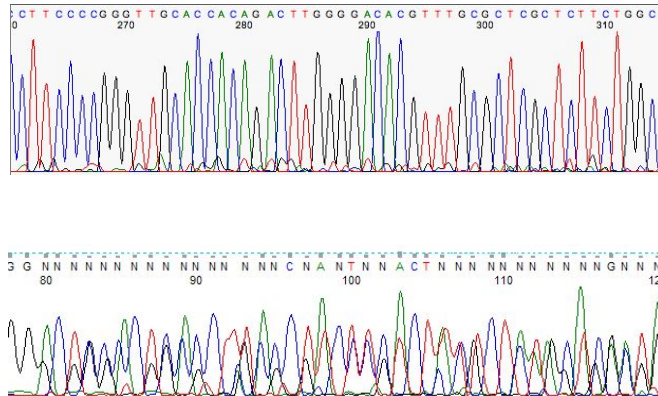
23 /

Sekwencjonowanie Sangera



24 /

Sekwencjonowanie Sangera



25 /

Sekwencjonowanie nowej generacji

- Sekwencjonowanie Sangera zapewnia sekwencjonowanie wysokiej jakości dla stosunkowo długich odcinków DNA (do około 900 par zasad). Zazwyczaj stosuje się je do sekwencjonowania poszczególnych fragmentów DNA, takich jak plazmidy bakteryjne lub DNA skopiowane w PCR.
- Jednak sekwencjonowanie Sangera jest kosztowne i nieefektywne w przypadku projektów na większą skalę, takich jak sekwencjonowanie całego genomu lub metagenomu („genomu zbiorowego” populacji drobnoustrojów). W przypadku takich zadań nowe techniki sekwencjonowania na dużą skalę są szybsze i tańsze.
- Wysoce równoległe: wiele reakcji sekwencyjnych ma miejsce w tym samym czasie
- Mikroskala: reakcje są prowadzone w małych objętościach i można ich prowadzić wiele jednocześnie na chipie.
- Szybkie: ponieważ reakcje są wykonywane równoległe, wyniki są gotowe o wiele szybciej
- Niski koszt: sekwencjonowanie genomu jest tańsze niż sekwencjonowanie Sangera
- Krótsze długości: odczytują typowy zakres 50 -700 nukleotydów

26 /

Rodzaje sekwencji genetycznych

- Genomowe DNA
pobrane bezpośrednio z genomu żywego organizmu i zawiera geny w ich naturalnym stanie.
- Kopiowane DNA (cDNA)
materiał przygotowany w procesie odwrotnej transkrypcji informacyjnego RNA (ang. messenger RNA, mRNA).
- Rekombinowane DNA
np. z plamidów, modyfikowanych wirusów i innych wykorzystywanych w biologii molekularnej sekwencji genetycznych.

Jakie wady i zalety z punktu widzenia analizy ma każdy z rodzajów sekwencji genetycznych?

27 /

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."



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."



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:
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

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

Wymagane dane:

- proteom i genom
- identyfikacja markerów poszukiwanych cech

Problem:
diagnostyka pacjentów we wczesnym etapie choroby nowotworowej

Wymagane dane:

- proteom i metabolom pacjenta zdrowego
- proteom i metabolom pacjenta chorego

gDNA, cDNA, rDNA

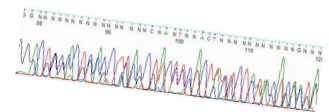
28 /

Strategie sekwencjonowania genomów

- Shotgun sequencing
Duże cząsteczki DNA dzielone są na mniejsze części i każda z nich sekwencjonowana jest indywidualnie.
 - Relatywnie krótkie i losowe sekwencje genomu muszą zostać złożone.
- Contig cloning
Fragmenty DNA poddawanego analizie są klonowane i dopiero te fragmenty są sekwencjonowane.
 - Całe DNA genomowe musi być sklonowane co znacznie wydłuża proces sekwencjonowania.

29 /

Kontrola jakości



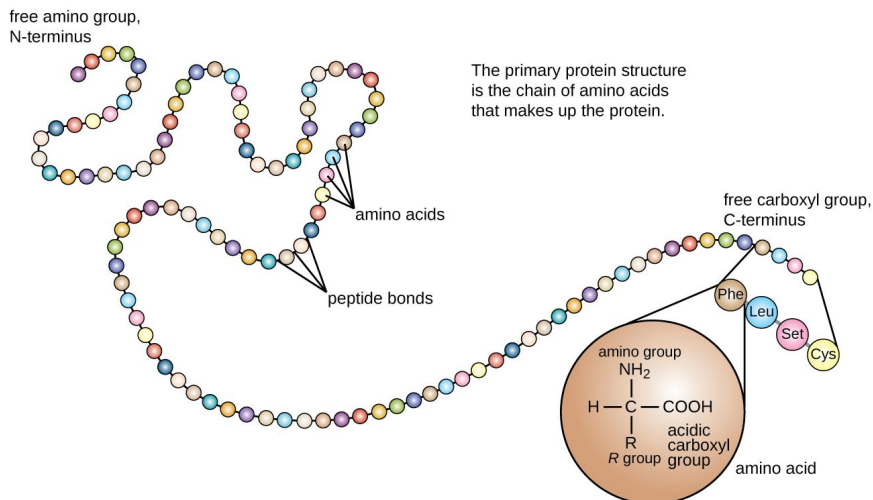
Garbage In, Garbage Out, GIGO (pol. śmieci na wejściu – śmieci na wyjściu) – w obszarze technologii informacyjnych angielski zwrot mówiący o tym, że nawet gdy program lub procedura przetwarzania były poprawne, wyniki przetwarzania błędnych danych będą błędne.

Sekwencjonowanie genomów:

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

30 /

Sekwencjonowanie białek



Nie ma metody powielania białka analogicznej do PCR!

31 /

Degradacja Edmana

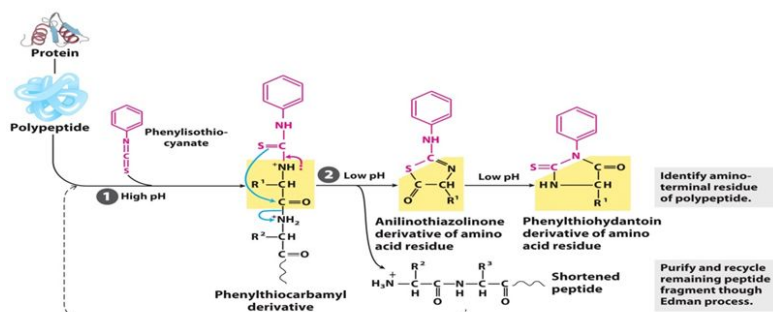


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

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.

32 /

Proteom

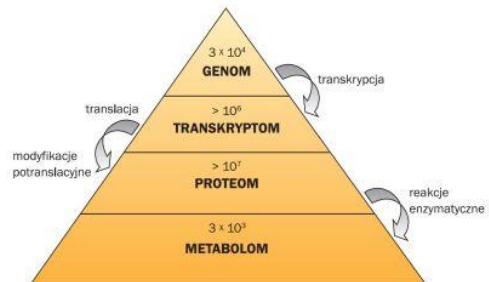
Proteom
component białkowy komórek kodowany
przez genom.

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
statyczny	dynamiczny	dynamiczny
możliwość amplifikacji	brak	brak
jednorodny	niejednorodny	niejednorodny
stałe stężenie	zmienne stężenie	zmienne stężenie

Analiza proteomiczna
musi uwzględniać 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:

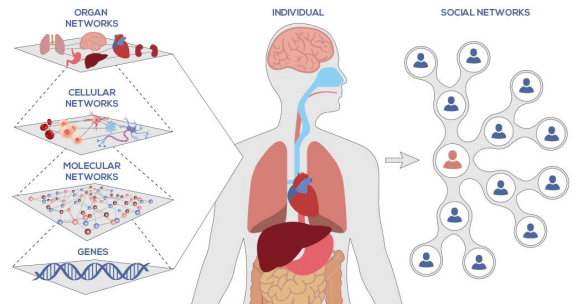
- jednoczesne porównanie dwóch lub więcej próbek;
- przygotowanych w podobny sposób;

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

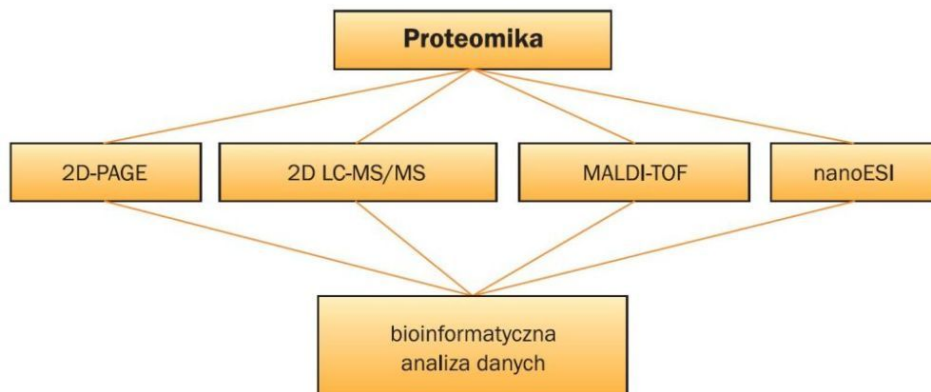
Biologia systemów

...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



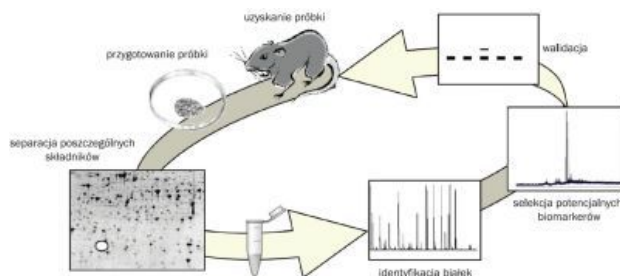
Strategia identyfikacji białek

Metody proteomiczne...

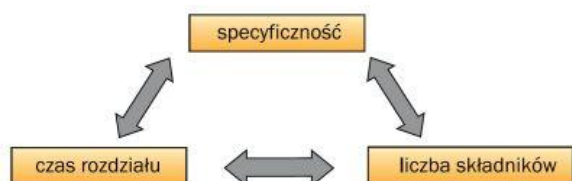
- oczyszczanie i separacja (materiału biologicznego)
- identyfikacja (komponentów)
- analiza i interpretacja wyników

-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...

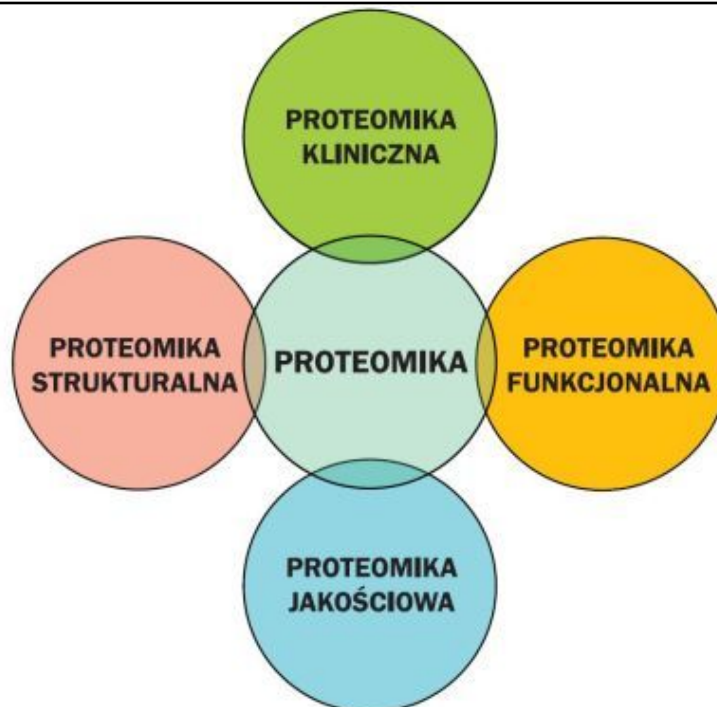
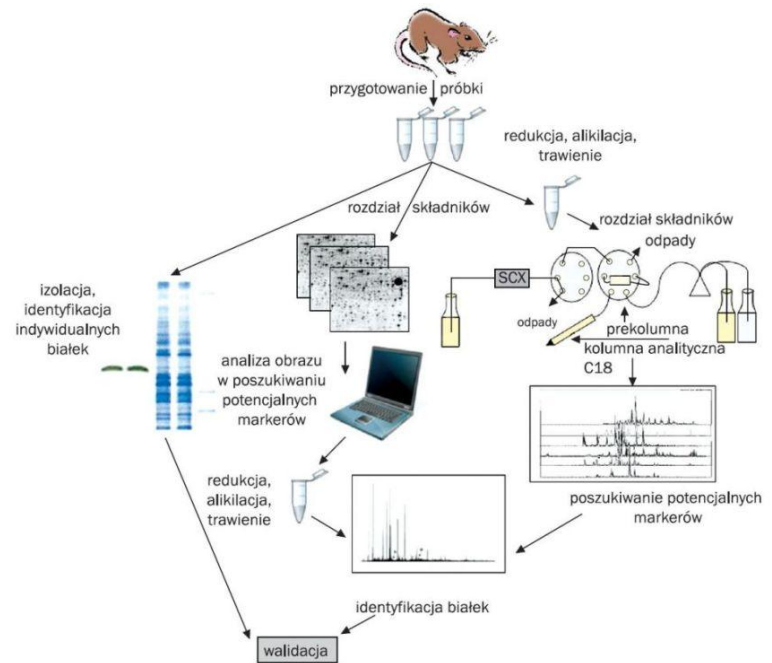
-przygotowanie powtarzalnych próbek (wielu) do analizy;

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...

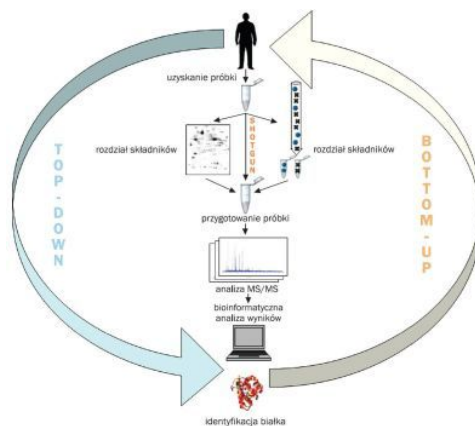
analizie poddawane są sekwencje peptydowe uzyskane w wyniku enzymatycznego trawienia białła. Wykorzystuje tandemową spektrometrię mas (MS/MS), lub mapy peptydowe (PMF, ang. *peptide mass fingerprinting*).

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:

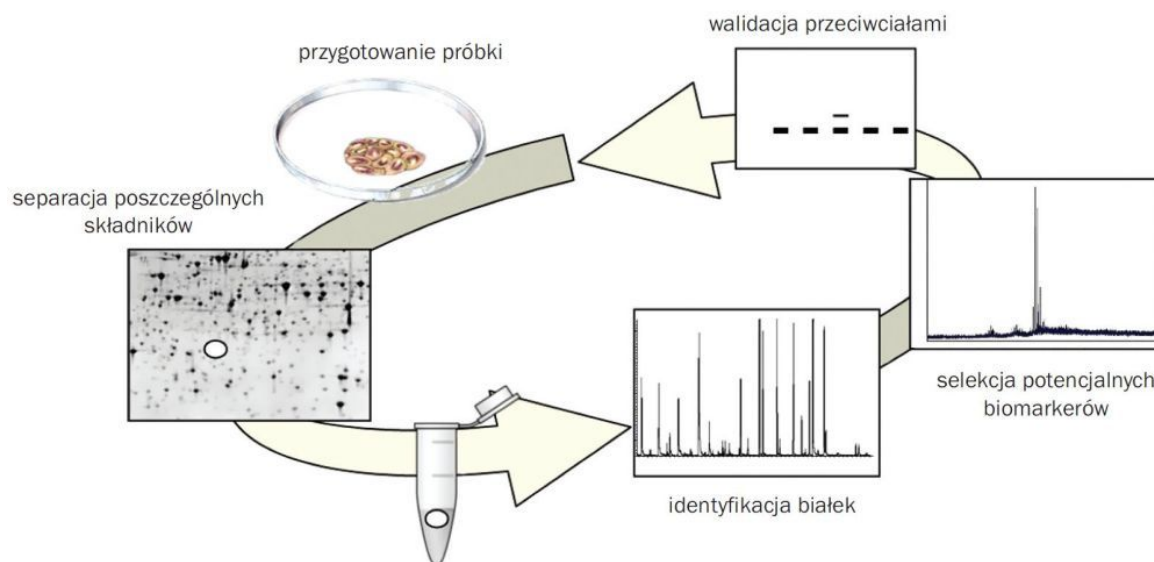
- Redukcja objętości
- Eliminacja zanieczyszczeń i niepożądanych składników

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, bufory 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

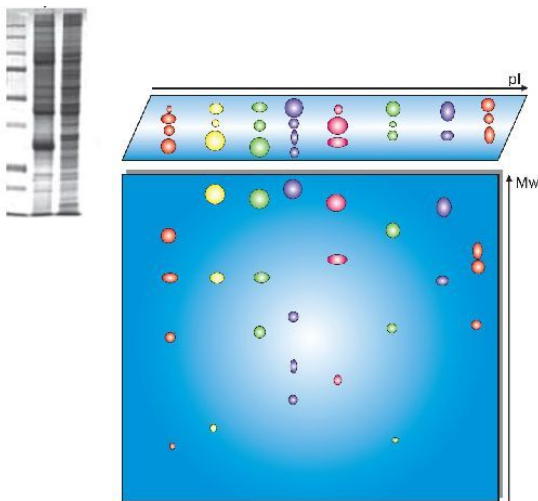
Odczynniki redukujące – redukcja grup sulfhydrylowych.

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

Inhibitory proteaz – niekontrolowana aktywność proteaz endogennych komplikuje profil białkowy.

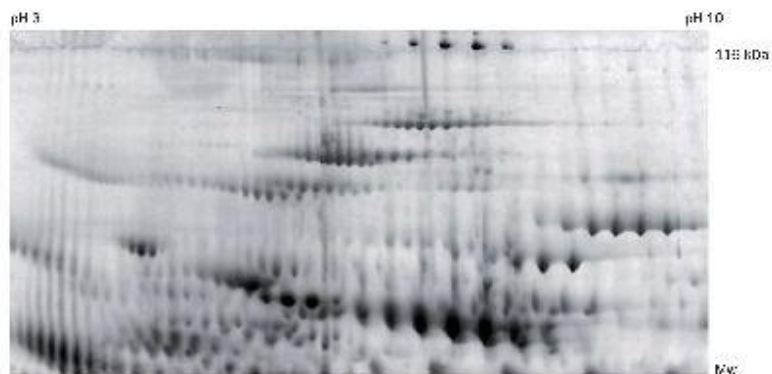
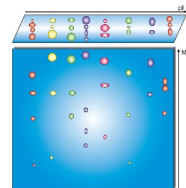
-fluorek fenylometylosulfonowy (PMSF), fluorek aminoetylobenzylsulfonowy (AEBSF), keton chlorometylowy N-tosylo-L-feniloalaniny (TLCK), benzamidyna, EDTA, amastatyna

Elektroforeza dwuwymiarowa (2-DE)

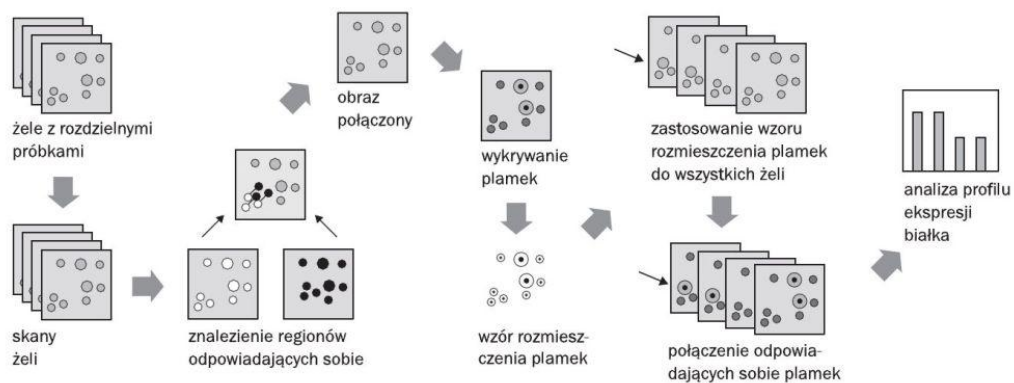


Rozdzielanie mieszanin białkowych wymach zastosowania specjalistycznych metod elektroforetycznych i chromatograficznych.

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.

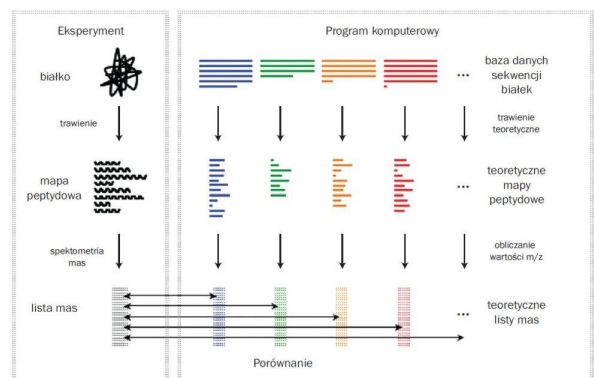
Metody analizy:

- 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 Monoisotopic Average

Data file Choose File No file chosen

Query

Data input

Decoy Report top hits

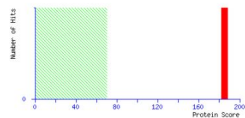
Start Search ... Reset Form

Mascot Search Results

User :
Email :
Search title : Peptide Mass Fingerprint Example
Database : SwissProt 2019_10 (561356 sequences; 20185828 residues)
Timestamp : 9 Jan 2020 at 11:11:29 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 Concise Protein Summary Help

Significance threshold $p < 0.05$ Max number of hits

Preferred taxonomy

Re-Search All Search Unmatched

- 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

BLCA_B0663 Mass: 37035 Score: 49 Expect: 6.7 Matches: 5
Protein Reck OS=Homo sapiens OX=9606 GN=RECK PE=3 SV=1

IFSA_P280V Mass: 14588 Score: 47 Expect: 11 Matches: 4
Translation initiation factor 58 OS=Homo sapiens OX=9606 GN=IFSA PE=3 SV=1

ISL2_C412 Mass: 22438 Score: 44 Expect: 20 Matches: 4
Probable nicotinate-nucleotide adenylyltransferase OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_H056 Mass: 22928 Score: 42 Expect: 39 Matches: 5
Inactive ribonuclease-like protein 10 OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_I0110 Mass: 52994 Score: 42 Expect: 39 Matches: 5
ISL2 OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_H056 Mass: 13327 Score: 41 Expect: 41 Matches: 3
T cell receptor gamma variable 8 OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_H056 Mass: 23812 Score: 40 Expect: 54 Matches: 4
ATP-dependent Clp protease proteolytic subunit 1 OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_H056 Mass: 64475 Score: 40 Expect: 59 Matches: 6
Protein surface glycoprotein 6 OS=Homo sapiens OX=9606 GN=ISL2 PE=3 SV=1

ISL2_H056 Mass: 41632 Score: 40 Expect: 60 Matches: 5
Bifunctional enzyme IspB/IspP OS=Campylobacter jejuni serotype O:1 (strain 81116 / NCTC 11628) OX=407148 GN=IspB PE=3 SV=1

ISL2_H056 Mass: 41664 Score: 40 Expect: 60 Matches: 5
Bifunctional enzyme IspB/IspP OS=Campylobacter jejuni serotype O:2 (strain ATCC 700819 / NCTC 11668) OX=407148 GN=IspP PE=1 SV=1

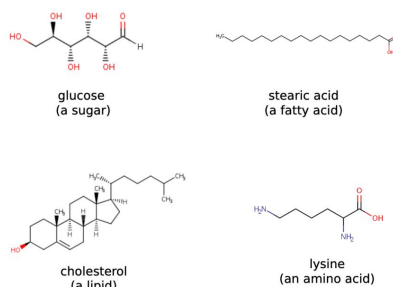
ISL2_H056 Mass: 41663 Score: 40 Expect: 60 Matches: 5
Bifunctional enzyme IspB/IspP OS=Campylobacter jejuni (strain 81121) OX=407148 GN=IspP PE=3 SV=1

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 ($< 1 \text{ nM}$).



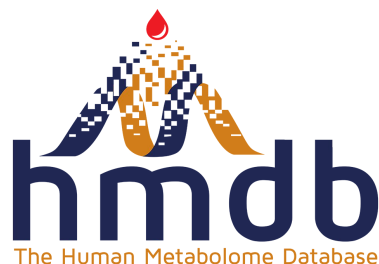
HMDB

- 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.

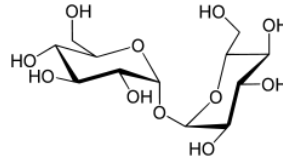
Dla każdej cząsteczki przypisany jest jeden lub więcej enzymów i białek transportowych.



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 (α ,D-glukopiranozylo-(1 $\rightarrow$ 1)- α ,D-glukopiranozyd). Jest głównym cukrem hemolimfy owadów, znajduje się także w grzybach i drożdżach.
- Aktualnego stanu metabolocznego 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.



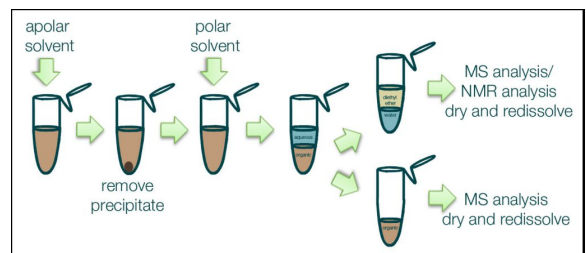
Strategie analityczne

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

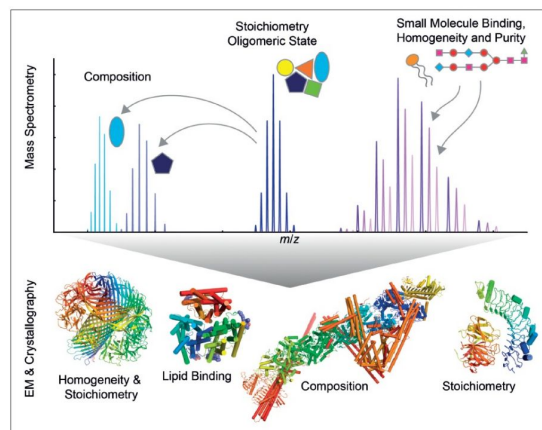
Rodzaj próbki determinuje strategię identyfikacyjną i przygotowawczą.
Przygotowanie próbki może wymagać np.:

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

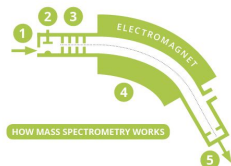
Spektometria mass



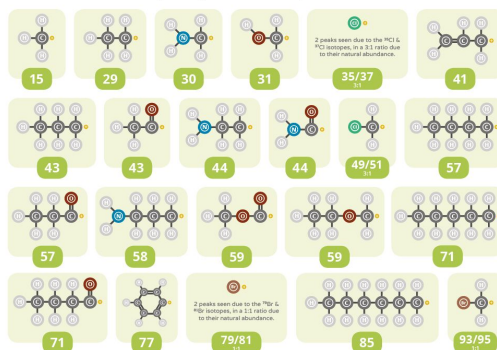
Spektometria mass

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.



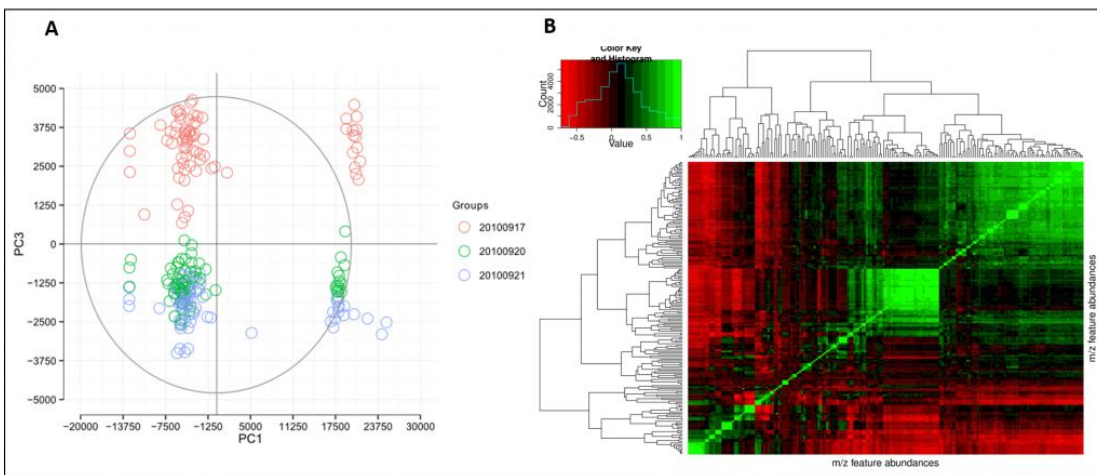
- 1 The sample is introduced to the mass spectrometer. Only very small samples are required. A heater is often present to vapourise the sample.
- 2 An electron gun ionises molecules in the sample by knocking out electrons, producing positive ions. Some molecules break into smaller ions & fragments.
- 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 a charged plate & 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.



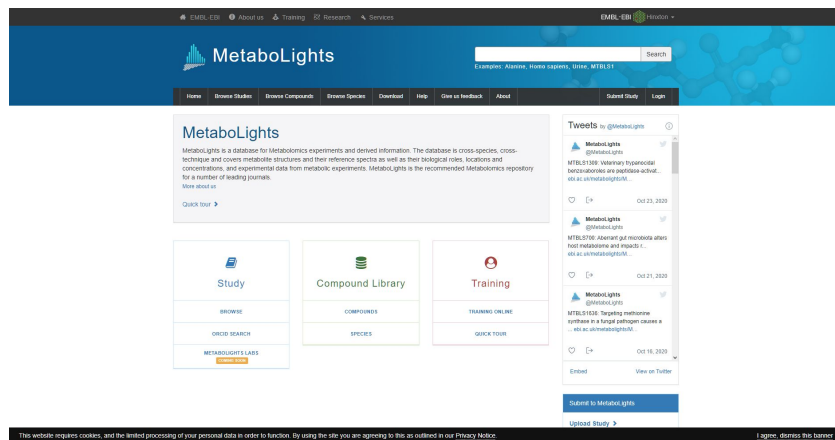
© COMPOUND INTEREST 2015 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem
This graphic is shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.



Zastosowanie danych metabolicznych

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

Zasoby internetowe



Zasoby internetowe

ChEBI

Home **Advanced Search** **Browse** **Documentation** **Download** **Tools** **About ChEBI**

Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on "small chemical compounds".

Example test: `KCO1=1SCN0C(=O)N2CN(C)CC1`, caffeine
[Advanced Search](#) | [Browse ChEBI](#)

Documentation

User Manual. Learn more about the data fields in the ChEBI and data sources for ChEBI. If you are unable to access the online Google document, please use the PDF links to view our manual.

[Train Online Tutorial](#). An online quick tour (0.3 h) and a more detailed tour (1 h) covers much of the material in the User Manual. Can be found on the EBI Train Online site here.

[Statistics](#). Graphs showing the growth of ChEBI, the numbers of curated, submitted and uncurated entries, the number of times for other resources and the frequency information present in ChEBI.

[Read more...](#)

Downloads

SDF files. ChEBI provides its chemical structures and additional data in structured data file (SDF) format.

[Ontology files](#). ChEBI ontology is provided in the W3C standard [RDF](#) ([OWL](#)) and [GML](#) formats.

Database files. ChEBI is stored in a relational database and we currently provide the ChEBI table in [flat file](#), [file delimited format](#), as an [Oracle](#) tablespace and a generic [SQL script](#) for MySQL and PostgreSQL databases.

[Read more...](#)

News

Tweets @chbi

[@chbi](#) [Open](#)

Stop kit now on loaned in the UK. Read more... [UK NPL&P facilitates bioscience sharing through tissue biobanking](#)

[@chbi](#) [Open](#)

Stop kit now on loaned in the UK. Read more... [UK NPL&P facilitates bioscience sharing through tissue biobanking](#)

[@chbi](#) [Open](#)

Oct 1, 2020

[@chbi](#) [Open](#)

ChEBI release 102.1 new analysis, vbl 6428. Lots entered directly, no via entity of the month. Entailed ENH742GP

[Enrich](#) [View on Twitter](#)

Entity of the month

2nd October 2020
Methylcobalamin

Stops cause significant damage to plants and crops, particularly cereals, potatoes, and all seed types and are regarded as arborescence's number one pest [1]. Methylcobalamin (CHEBI:51313) is a cyclic holotranscobalamin derivative that was first isolated and determined in 1938 [2]. It initially found use as a solid fat-soluble vitamin, sold as Water-Sol but nowadays, it is used as a feed additive for the control of slug and snail in agricultural fields and domestic gardens. It kills slugs and snails by disrupting their ability to produce mucus thereby reducing their digestion and mobility thus making them susceptible to dehydration [3].

[Read more...](#)

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

Pathway Browser

Visualize and interact with Reactome biological pathways

Analyze Data

Merges pathway identifier mapping over-representation, and expression analysis

ReactomeFIViz

Designed to find pathways and network patterns related to cancer and other types of diseases

Documentation

Information to browse the database and use its principal tools for data analysis

USE REACTOME GRAPH DATABASE IN YOUR PROJECT

[LEARN MORE](#)

Why Reactome

Reactome is a free, open-source, curated and peer-reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genomic analysis, modeling, systems biology and education.

If you use Reactome in R, we suggest using our Chordal server site at reactome.graph.rn

The development of Reactome is supported by grants from the UK Science Institutes of Health (UHS H080773) and the European Molecular Biology Laboratory.

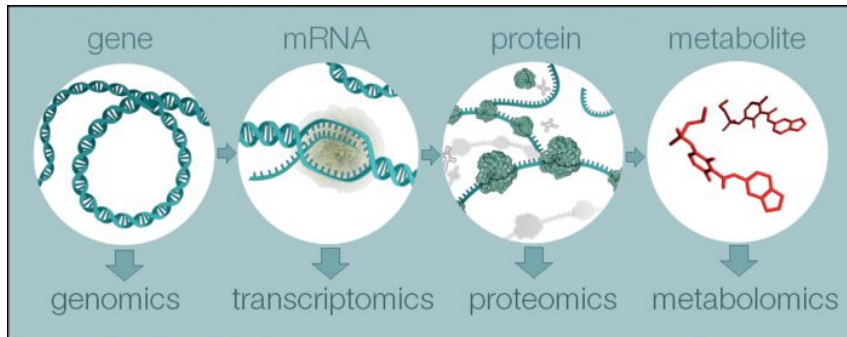
Tweets

Reactome Retweeted

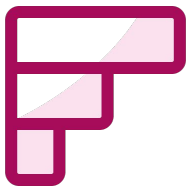
EMBL-EBI @europeanbioinformatics

Reactome in GenesData

Our @Reactome colleagues and collaborators fast-tracked the annotation of the [H3K9me3](#) infection pathway in collaboration with the [ePROVE](#) (Epidemic Map group) @Week #SupportOIA #COVID19ways reactome.org/about/news/761



slido



Z praktycznego punktu widzenia (łatwości, szybkości przetwarzania danych), który rodzaj informacji genetycznej będzie najlepszy?

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