

Uczenie maszynowe w projektowaniu leków - wprowadzenie

Dr hab. Sabina Podlewska

**Zakład Chemii Leków, Instytut Farmakologii im.
Jerzego Maja PAN,
ul. Smętna 12, 31-343 Kraków**



Plan przedmiotu

1. Wykład – 45 h, stacjonarnie (w wyjątkowych przypadkach zdalnie), częściowo w formie konsultacji do projektu
2. Ćwiczenia: 15 h
3. Zaliczenie: projekt

Przykłady zrealizowanych projektów



JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Article

Cite This: *J. Chem. Inf. Model.* 2019, 59, 4974–4992

pubs.acs.org/jcim

Development of New Methods Needs Proper Evaluation—Benchmarking Sets for Machine Learning Experiments for Class A GPCRs

Damian Leśniak,[†] Sabina Podlewska,^{*,‡,§} Stanisław Jastrzębski,[†] Igor Sieradzki,[†] Andrzej J. Bojarski,[§] and Jacek Tabor[†]

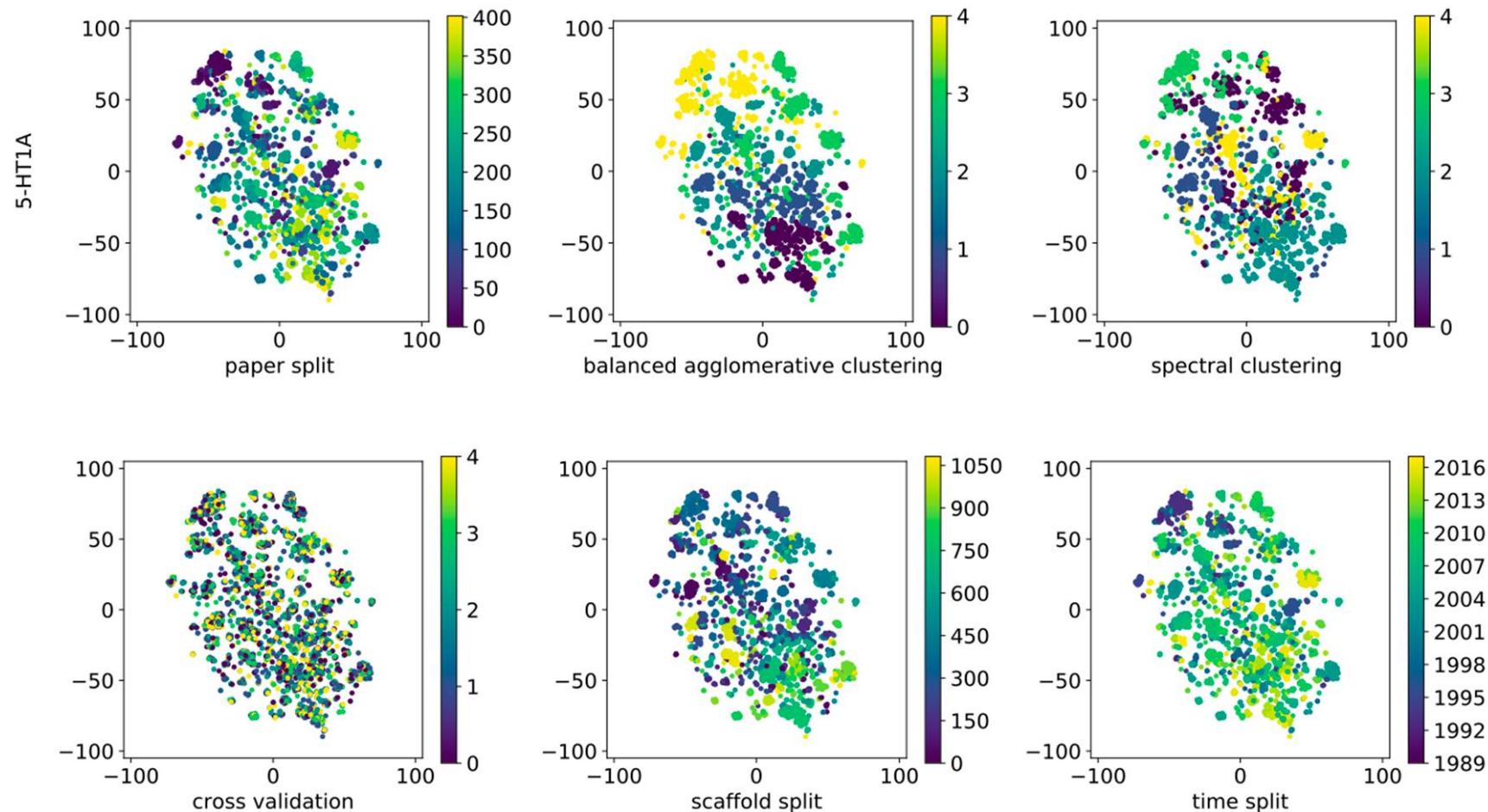
[†]Faculty of Mathematics and Computer Science, Jagiellonian University, 6 Łojasiewicza Street, 30-348 Kraków, Poland

[‡]Department of Technology and Biotechnology of Drugs, Jagiellonian University Medical College, 9 Medyczna Street, 30-688 Kraków, Poland

[§]Maj Institute of Pharmacology, Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland

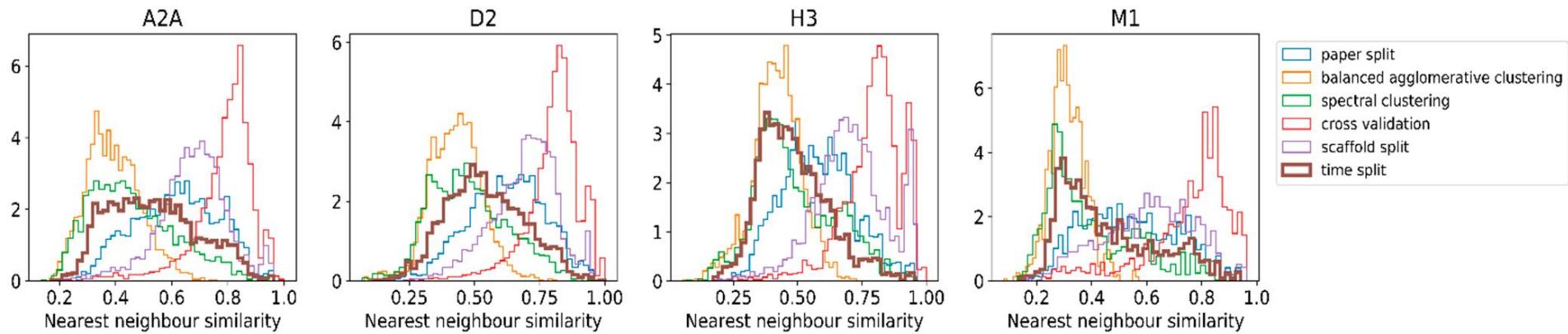
- Przygotowanie zestawu eksperymentów referencyjnych pozwalających na ocenę skuteczności predykcji przez powstające narzędzia
- Kompleksowe zbadanie wpływu poszczególnych etapów konstrukcji zbioru uczącego na skuteczność przewidywań aktywności

Przykłady zrealizowanych projektów – benchmarki (Leśniak et al.)



tSNE visualization of the chemical spaces of ligands for various splitting approaches

Przykłady zrealizowanych projektów – benchmarki (Leśniak et al.)



Analysis of similarity coefficients between training and test data for different splitting approaches.

Przykłady zrealizowanych projektów



Article

How Sure Can We Be about ML Methods-Based Evaluation of Compound Activity: Incorporation of Information about Prediction Uncertainty Using Deep Learning Techniques

Igor Sieradzki ¹, Damijan Leśniak ¹ and Sabina Podlewska ^{2,3,*}

¹ Faculty of Mathematics and Computer Science, Jagiellonian University, 6 S. Łojasiewicza Street, 30-348 Cracow, Poland; igor.sieradzki@doctoral.uj.edu.pl (I.S.); damian.lesniak@doctoral.uj.edu.pl (D.L.)

² Department of Technology and Biotechnology of Drugs, Jagiellonian University, Medical College, 9 Medyczna Street, 30-688 Cracow, Poland

³ Maj Institute of Pharmacology, 12 Smętna Street, 31-343 Cracow, Poland

* Correspondence: smusz@if-pan.krakow.pl; Tel.: +48-1266-23-301

Academic Editors: Igor F. Tsigelny and Bernard Maigret

Received: 27 December 2019; Accepted: 22 March 2020; Published: 23 March 2020



Przykłady zrealizowanych projektów

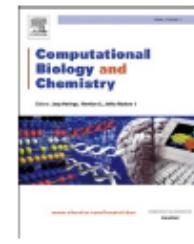
Computational Biology and Chemistry 88 (2020) 107367



Contents lists available at [ScienceDirect](#)

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac



Similar, or dissimilar, that is the question. How different are methods for comparison of compounds similarity?

Krzysztof Rajda^a, Sabina Podlewska^{b,c,*}

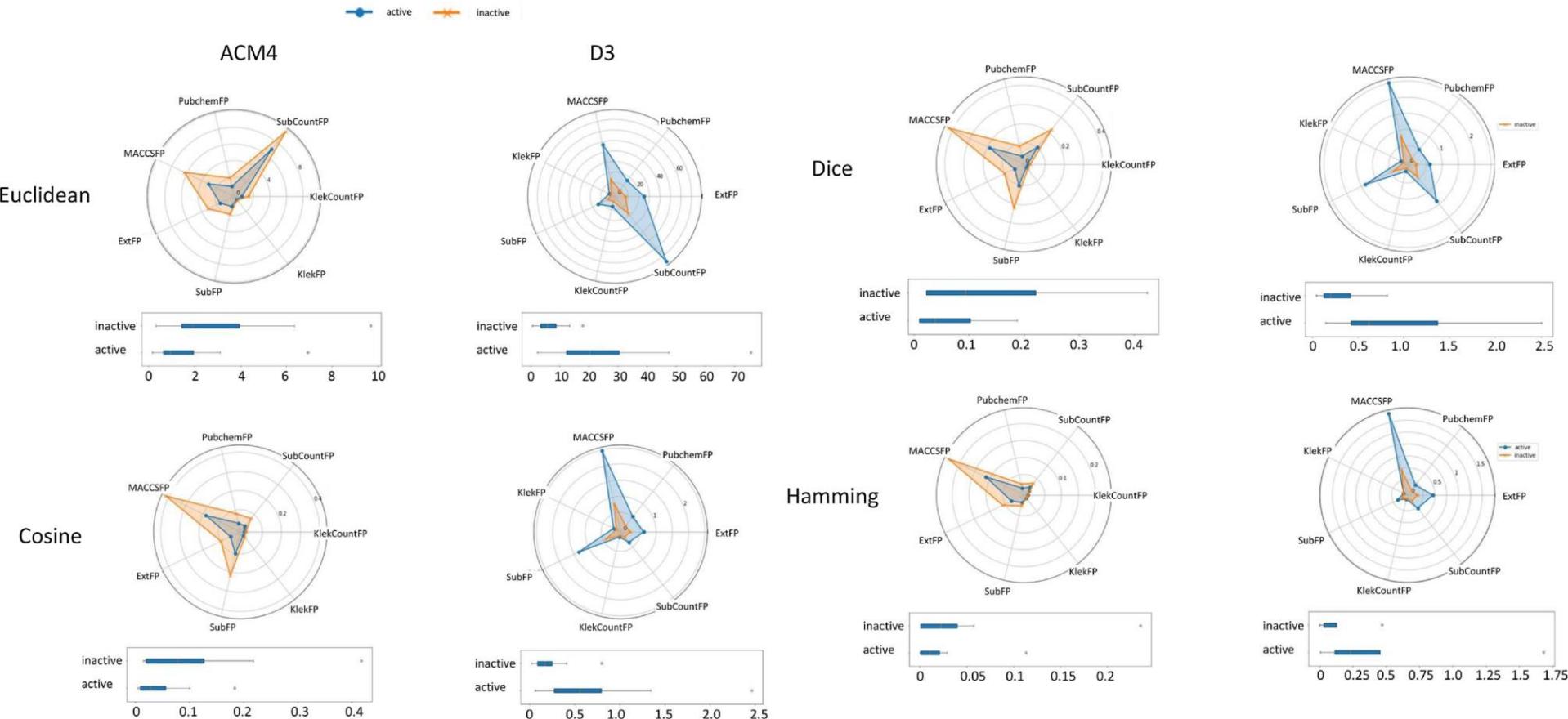
^a Wrocław University of Science and Technology, Faculty of Computer Science and Management, 50-371 Wrocław, 1. Łukasiewicza Street 5, Poland

^b Jagiellonian University Medical College, Department of Technology and Biotechnology of Drugs, 30-688 Kraków, 9 Medyczna Street, Poland

^c Maj Institute of Pharmacology, Polish Academy of Sciences, Department of Medicinal Chemistry, 31-343 Kraków, Smętna Street 12, Poland



Przykłady realizowanych projektów – metryki podobieństwa (Rajda et al.)



Analysis of mean normalized distances between all pairs of compounds for different metrics and compound representation for ACM4 and D3 receptors with box-whiskers plots indicating the results variation.

Przykłady projektów



pubs.acs.org/jcim

Application Note

ADMET-PrInt: Evaluation of ADMET Properties: Prediction and Interpretation

Ewelina Jamrozik, Marek Śmieja, and Sabina Podlewska*



Cite This: <https://doi.org/10.1021/acs.jcim.3c02038>

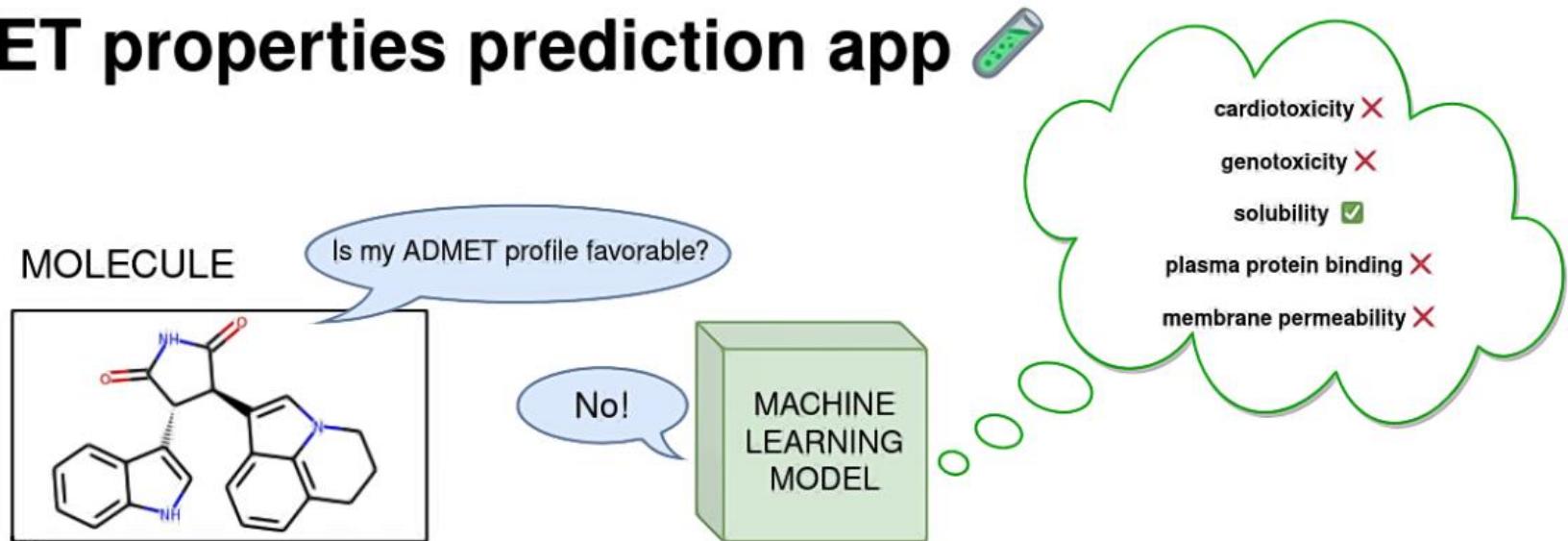


Read Online

- Serwis on-line do predykcji wybranych własności fizykochemicznych i ADMET związków
- Uwzględnienie metod wyjaśnialności ułatwiające przeprowadzenie procesu optymalizacji struktury związku chemicznego pod kątem danego parametru

Przykłady projektów: ADMETPrInt (Jamrozik et al.)

ADMET properties prediction app



Przykłady projektów: ADMETPrInt (Jamrozik et al.)

Description info

Atom symbols:

A : Any valid periodic table element symbol
Q : Hetro atoms; any non-C or non-H atom
X : Halogens; F, Cl, Br, I
Z : Others; other than H, C, N, O, Si, P, S, F, Cl, Br, I

Bond types:

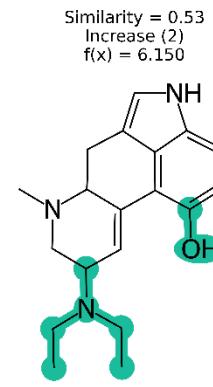
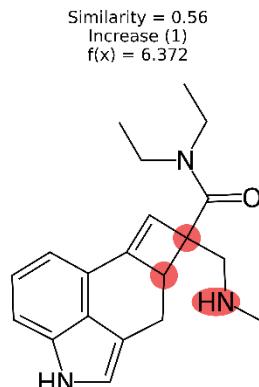
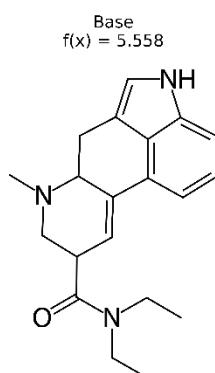
- : Single
= : Double
T : Triple
: Triple
~ : Single or double query bond
% : An aromatic query bond
None : Any bond type; no explicit bond specified
\$: Ring bond; before a bond type specifies ring bond
! : Chain or non-ring bond; before a bond type specifies chain bond
@ : A ring linkage and the number following it specifies the atoms position in the line, thus @1 means linked back to the f

Aromatic: Kekule or Arom5

Kekule: Bonds in 6-membered rings with alternate single/double bonds or perimeter bonds

Arom5: Bonds in 5-membered rings with two double bonds and a hetro atom at the apex of the ring.

Counterfactual analysis



Przykłady projektów - MedChemGame

MedChem Game: Gamification of Drug Design

Tomasz Danel,^{*,†,‡} Jan Łęski,[†] Sabina Podlewska,[¶] and Igor T. Podolak[†]

[†]*Faculty of Mathematics and Computer Science, Jagiellonian University, Łojasiewicza 6,
30-348 Kraków, Poland*

[‡]*Faculty of Chemistry, Jagiellonian University, Gronostajowa 2, 30-387 Kraków, Poland*

[¶]*Maj Institute of Pharmacology, Polish Academy of Sciences, Smętna 12, 31-343 Kraków,
Poland*

E-mail: tomasz.danel@uj.edu.pl

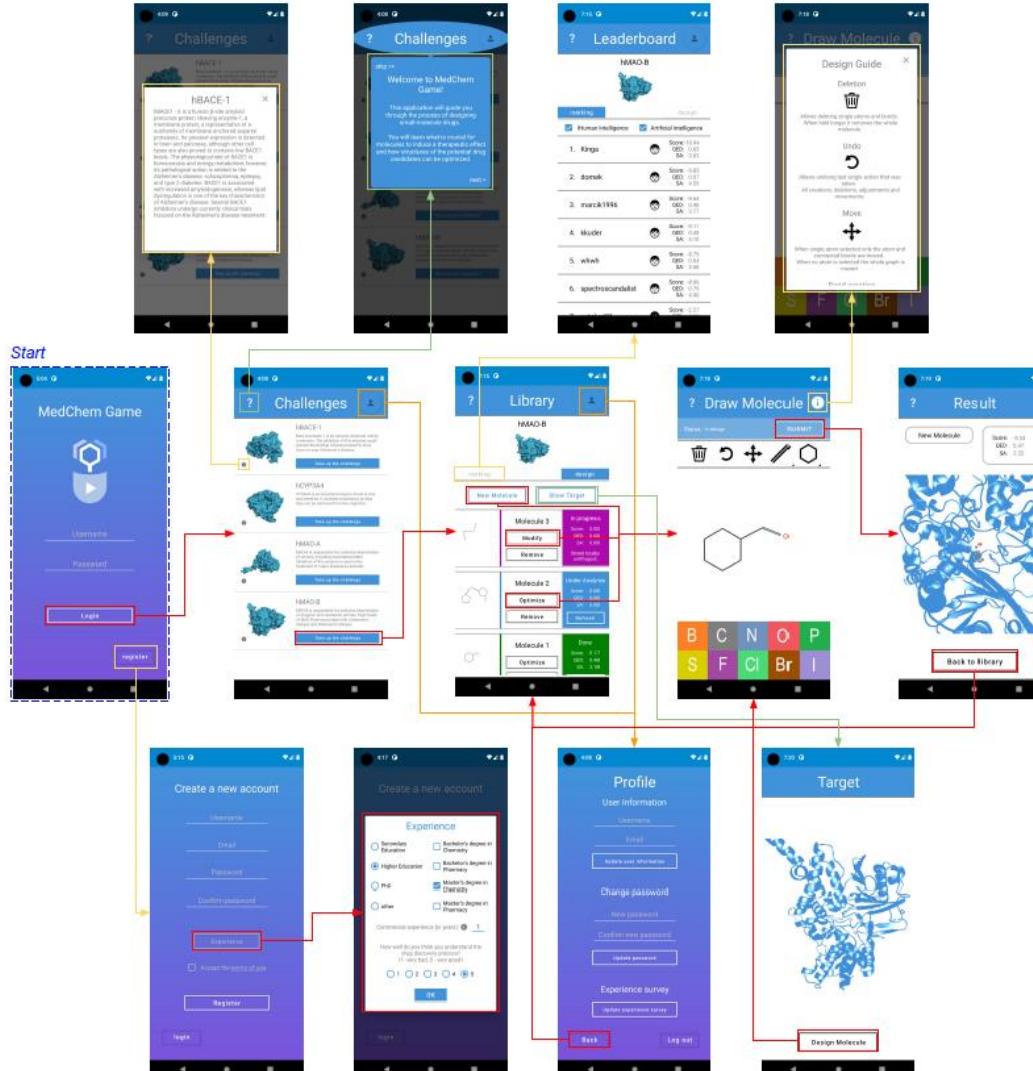
gra na telefony z systemem Android, w której uczestnicy tworzą związki chemiczne, które następnie poddawane są punktowaniu.

Przykłady projektów - MedChemGame



MedChem Game summarized in four steps: (1) target selection, (2) compound design, (3) inspection of the docked protein-ligand complex and results summary, and (4) participation in the global ranking and contribution to the ML model training.

Przykłady projektów - MedChemGame



Przykłady projektów - SERIFS (Rybka et al.)

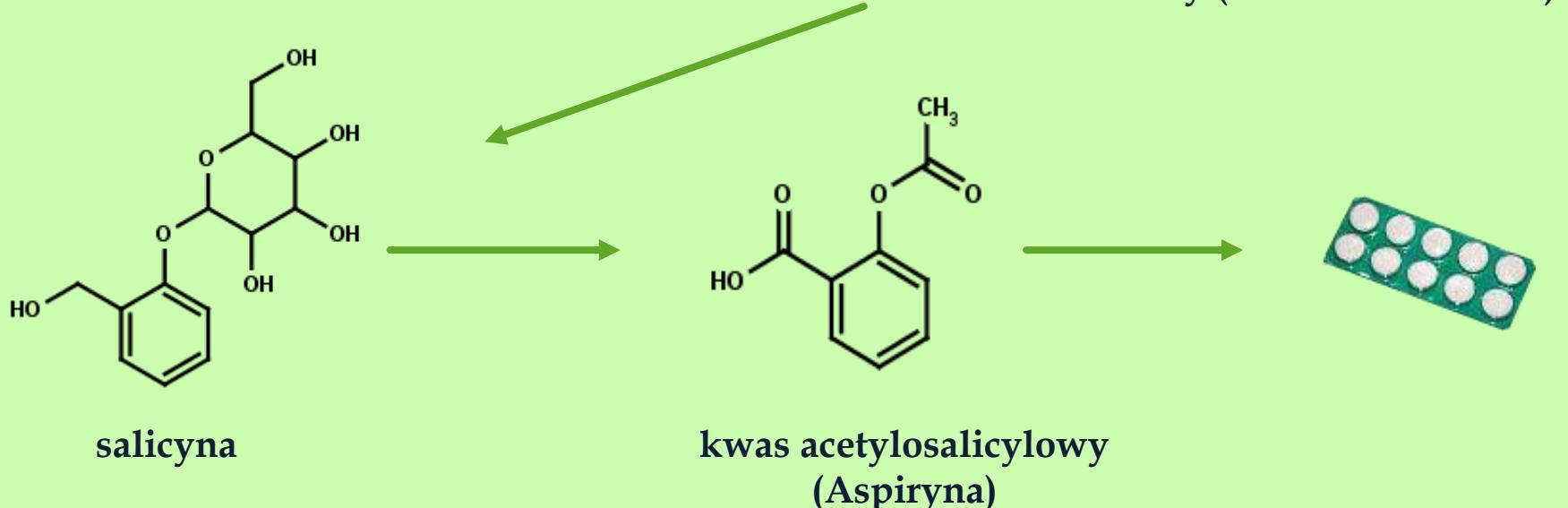
SERIFS: design of target-focused libraries by
exploring active ligand subspaces encoded
with fingerprints

Hubert Rybka,^{†,‡} Mateusz Iwan,[‡] Anton Siomchen,[‡] Tomasz Danel,^{*,‡,§} and
Sabina Podlewska^{*,¶,§}

Tworzenie nowych leków

Dawniej...

- Intuicja
- Przypadek
- Pochodzenie naturalne



Tworzenie nowych leków

i dziś

Racjonalne projektowanie leków



↓
Proces

• długotrwały:

~ 15 lat

• kosztowny:

~1,2 mld dolarów

Komputerowo wspomagane projektowanie leków



Projekty realizowane przez ZChL IF PAN

ProKog. Antagoniści receptora 5-HT6: Nowoczesne leki przeciwpsychotyczne o dodatkowym działaniu prokognitywnym.

The screenshot shows the homepage of the ProKog project website. At the top left is the logo 'PRO KOG' with a molecular structure icon. The main title 'ANTAGONIŚCI RECEPTORA 5-HT6: NOWOCZESNE LEKI PRZECIWPSYCHOTYCZNE O DODATKOWYM DZIAŁANIU PROKOGNITYWNYM' is centered above five circular images: a molecular model, two scientists in lab coats, a tray of test tubes, a mouse being tested, and a pile of blue and white capsules. Below the images is a dark navigation bar with links: Opis projektu, Zadania badawcze, Aktualności (highlighted in orange), Promocja, Galeria, Kontakt, Forum, and PL |.



Aktualności

Spotkanie Naukowe 12 października 2010

12 października 2010 odbyło się Pierwsze Spotkanie Naukowe wszystkich uczestników projektu, w którym uczestniczyli również zaproszeni eksperci: profesor Edmund Przegaliński, profesor Jerzy Vetulani, doc. Grażyna

Projekty realizowane przez ZChL IF PAN

ModAll. Modulacja allosteryczna – nowa strategia w farmakoterapii. Identyfikacja własności psychotropowych ligandów receptorów glutaminianergicznych III grupy.

The screenshot shows the homepage of the ModAll project website. At the top, there is a navigation bar with links: Opis Projektu, Zespoły badawcze, Aktualności, Promocja, Galerie, Linki, Kontakt, Zaloguj, Szukaj, and a search input field. Below the navigation bar, the project logo "ModAll" is displayed in large letters, followed by the subtitle "Modulacja allosteryczna - nowa strategia w farmakoterapii". A brief description of the project follows: "Identyfikacja własności psychotropowych ligandów receptorów glutaminianergicznych III grupy". The main visual element is a large, abstract blue and white graphic of a brain, with a diagonal slice showing internal structures. At the bottom left, there is a breadcrumb trail: Jesteś: Strona główna. The bottom section is divided into two columns: "Informacje o Projekcie" on the left and "Opis Projektu" on the right.

Informacje o Projekcie

Projekt ModAll jest przedsięwzięciem badawczo-rozwojowym realizowanym w okresie 04.2009 – 03.2013 samodzielnie przez Instytut Farmakologii PAN w Krakowie w ramach Programu Operacyjnego Innowacyjna Gospodarka współfinansowanego z Europejskiego Funduszu Rozwoju Regionalnego Unii Europejskiej.

Tematyka Projektu wpisuje się w Priorytet 1: Badania i rozwój nowoczesnych technologii, Działanie 1.3: Wsparcie projektów B+R na rzecz przedsiębiorców realizowanych przez jednostki naukowe oraz Poddziałanie 1.3.1: Projekty rozwojowe.

Opis Projektu

Program badawczy Projektu obejmuje zagadnienia z dziedziny psychofarmakologii i dotyczy metabotropowych receptorów glutaminianergicznych (mGluR). Receptory te szczególnie licznie występują w strukturach uważanych za kluczowe w patogenezie oraz leczeniu depresji i leku a kwas glutaminowy, będący endogennym ligandem tych receptorów jest najważniejszym aminokwasem pobudzającym ośrodkowego układu nerwowego (OUN).

Pierwszych przesłanek wskazujących na rolę układu glutaminianergicznego w patogenezie zaburzeń psychicznych dostarczyły wyniki badań

Projekty realizowane przez ZChL IF PAN

PLATFORMEX. Rozszerzenie badań Platformy akademickiej do poszukiwania związków o własnościach przeciwddepresyjnych

Strona główna Opis Partnerzy Aktualności Publikacje Konferencje Galeria Linki Kontakt

Strona Główna

Projekt Pol-Nor/198887/73/2013 pt.: „Rozszerzenie badań Platformy akademickiej do poszukiwania związków o własnościach przeciwddepresyjnych” jest kontynuacją badań prowadzonych w ramach grantu PNRF-103-AI-1/07 finansowanego w poprzedniej edycji Norweskiego Mechanizmu Finansowego.

Nowe serie związków zostały zaprojektowane w oparciu o wyniki otrzymane wcześniej i obejmują połączenia hybrydowe posiadające aktywność do wytypowanych receptorów serotoninowych oraz nowe modulatory receptorów glutamatergicznych. Ponadto rozszerzono zakres badań Platformy o poszukiwania modulatorów allosterycznych SERT oraz receptorów GABA B. Aktywność związków będzie oceniana w testach *in vitro* a ich własności terapeutyczne weryfikowane w behawioralnych modelach depresji.

POCZĄTEK PROJEKTU: 01.11.2013

ZAKOŃCZENIE PROJEKTU: 30.04.2016

Konsorcjum powołane do realizacji Projektu: Instytut Farmakologii PAN w Krakowie; department of Medical Biology, Faculty of Health Science of the University of Tromsø, Narodowy Instytut Leków.

Projekty zrealizowane w poprzedniej edycji kursu

- Rajda K.; Podlewska S. Similar, or dissimilar, that is the question. How different are methods for comparison of compounds similarity? *Comput. Biol. Chem.* 2020, 88, 107367.
- Sieradzki, I.; Lesniak, D.; Podlewska, S. How Sure Can We Be about ML Methods-Based Evaluation of Compound Activity: Incorporation of Information about Prediction Uncertainty Using Deep Learning Techniques. *Molecules*, 2020, 25, 1452.

Białka jako najczęstsze cele molekularne

Białko (proteina) – termin wprowadzony pod koniec XIXw. przez Jonsa Berzeliusa w celu podkreślenia znaczenia tej grupy związków.

Proteios (grec.) – pierwszorzędny, o głównym znaczeniu

Skład pierwiastkowy większości białek:

C – 50-55%

H – 6.5-7.5%

N – 15-19%

O – 19-21%

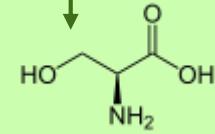
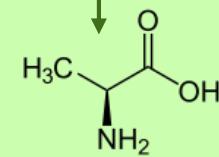
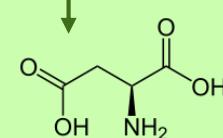
Budowa białek

Białka są zbudowane z aminokwasów



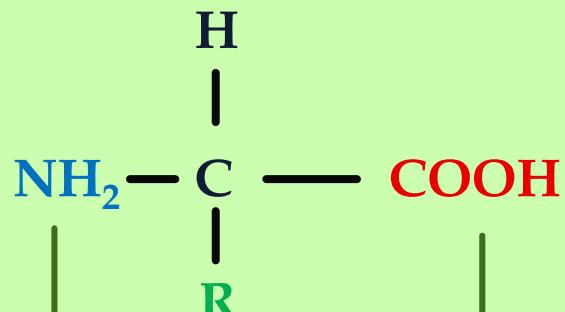
DVLACS

Asp Val Leu Ala Cys Ser



Aminokwasy

Budowa ogólna

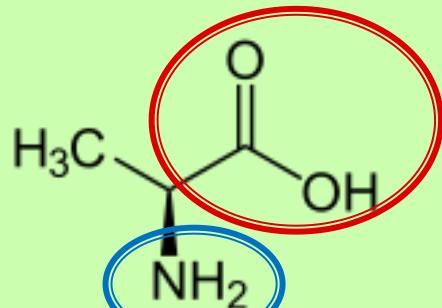


↓

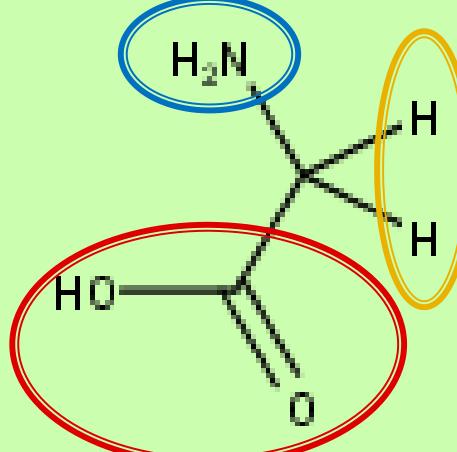
grupa
aminowa

↓

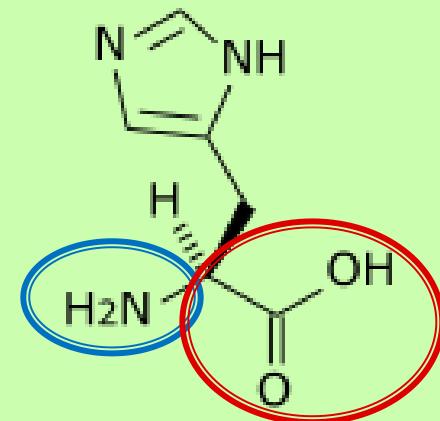
grupa
karboksylowa



Alanina



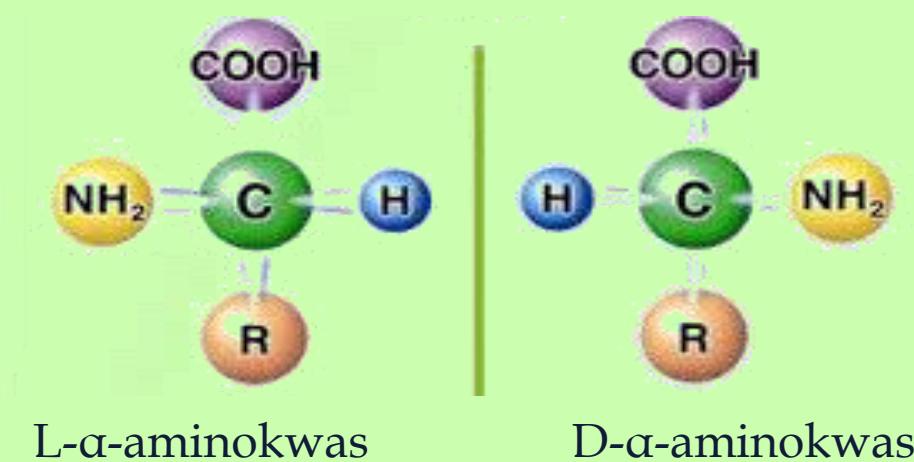
Glicyna



Histydyna

Czynność optyczna aminokwasów

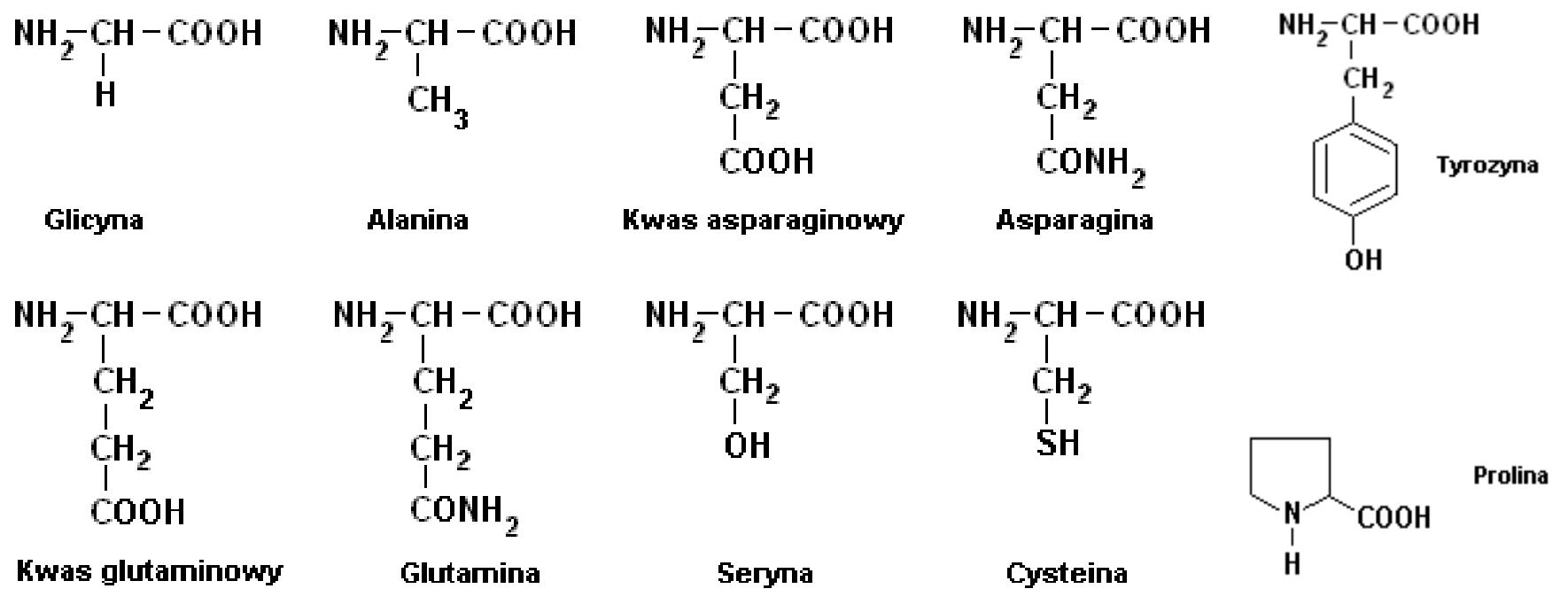
- Wszystkie aminokwasy, poza glicyną, są czynne optycznie



- Białka są zbudowane wyłącznie z aminokwasów o konfiguracji L
- 20 podstawowych aminokwasów wchodzących w skład wszystkich organizmów żywych

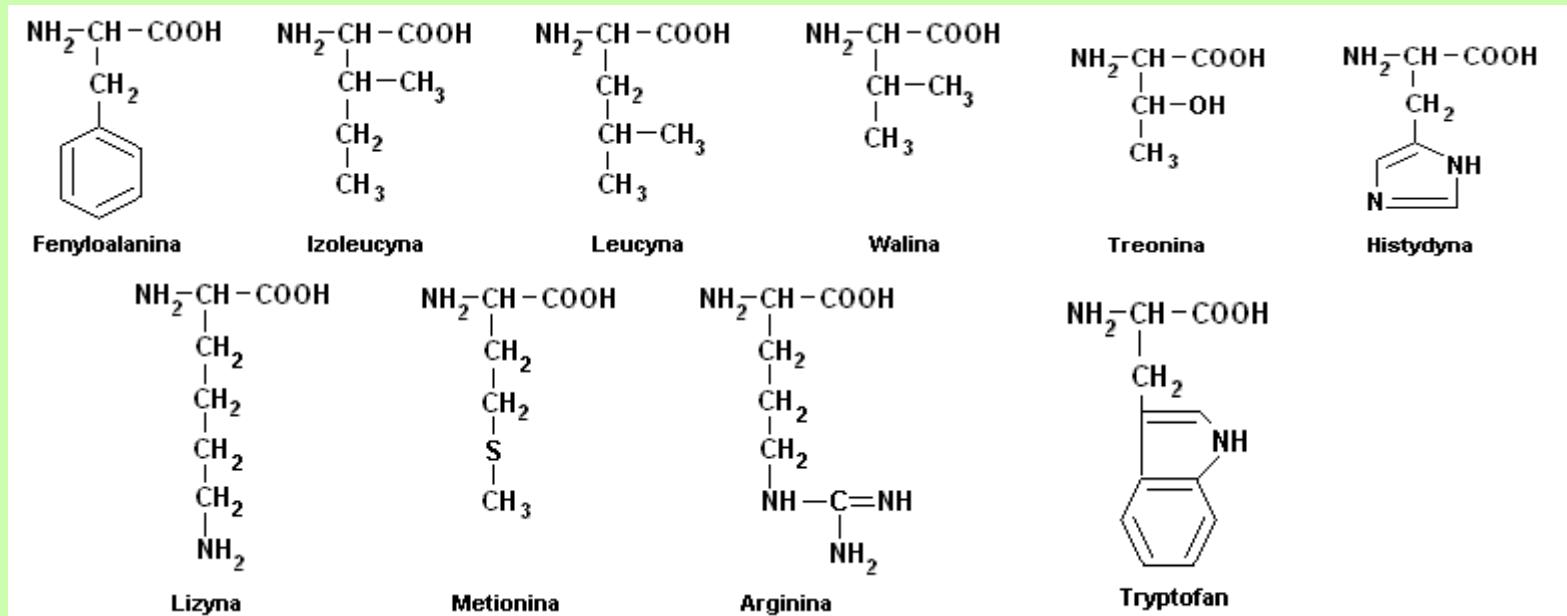
Aminokwasy endogenne

- Aminokwasy, które mogą być samodzielnie zsyntezowane przez organizm.



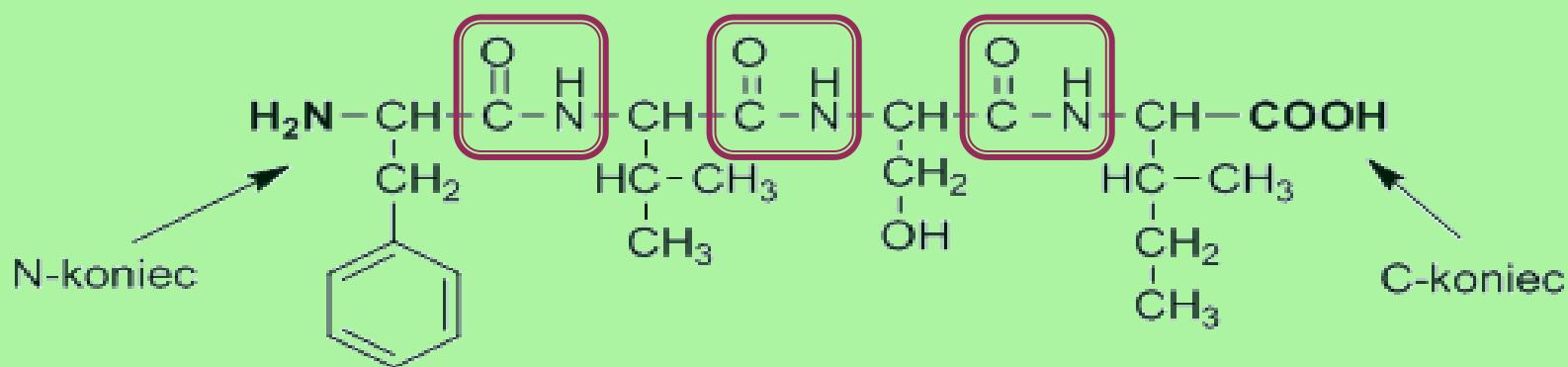
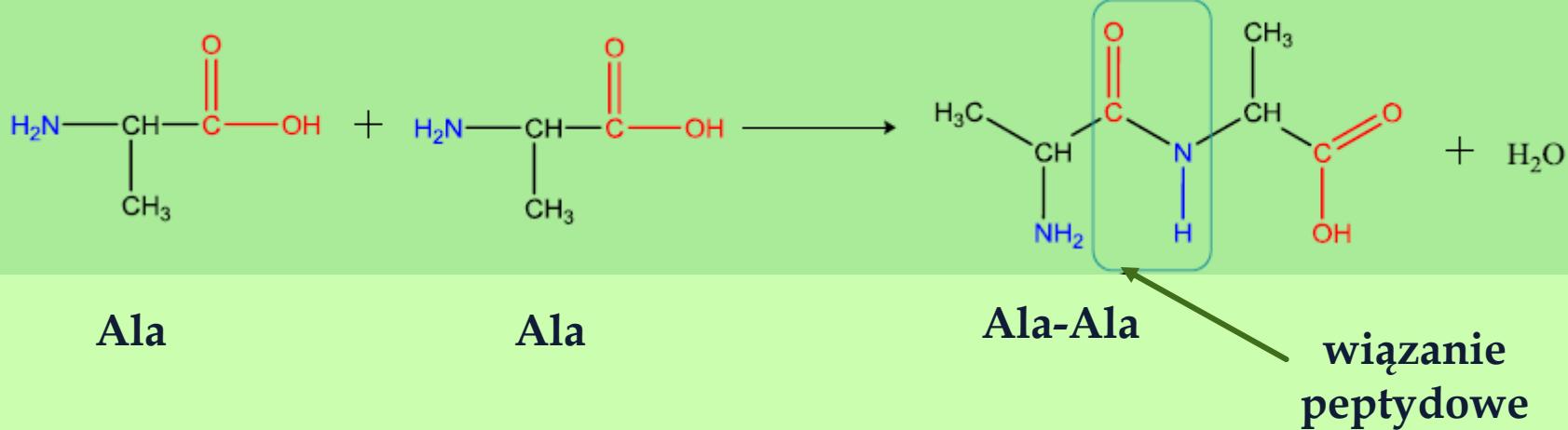
Aminokwasy egzogenne

- Aminokwasy, których organizm nie jest w stanie samodzielnie zsyntezować, dlatego muszą być one dostarczane w pożywieniu.

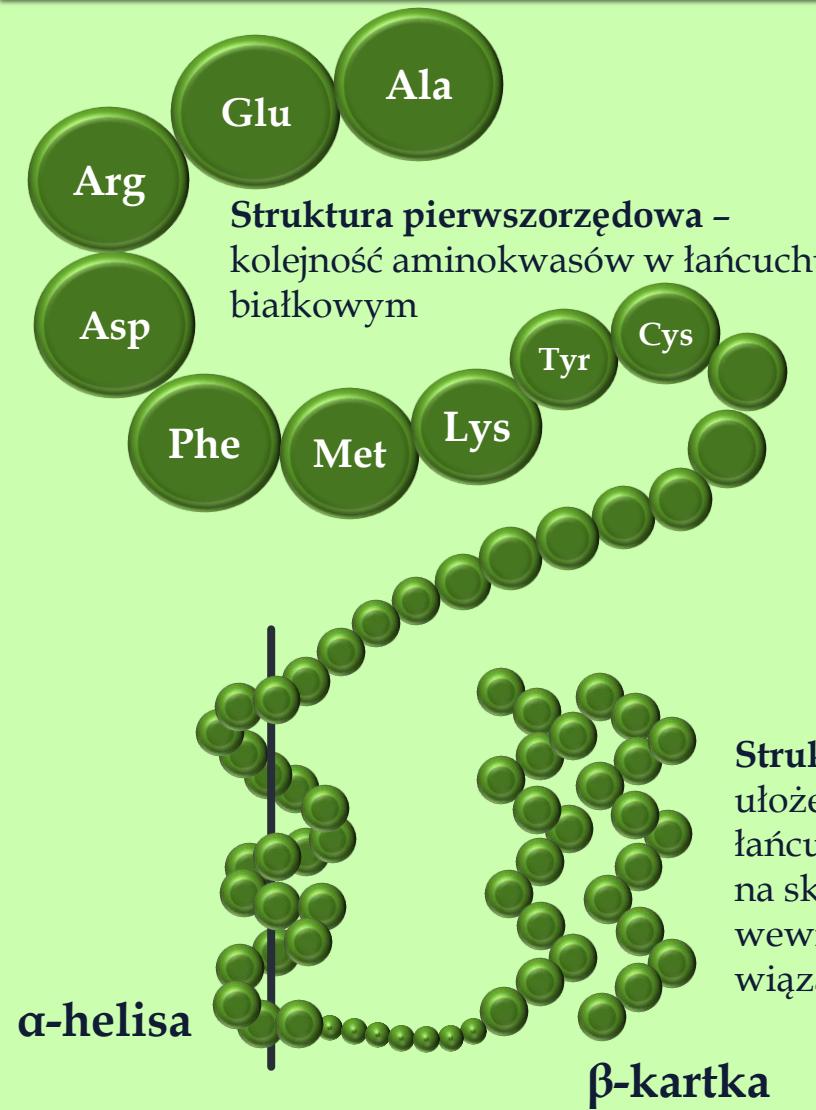


Łączenie aminokwasów – powstawanie peptydów

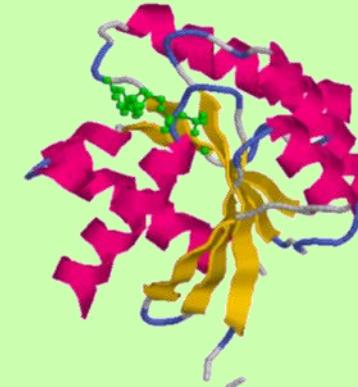
- Kondensacja aminokwasów prowadzi do powstawania peptydów oraz w dalszej kolejności białek.



Rzędowość struktury białkowej

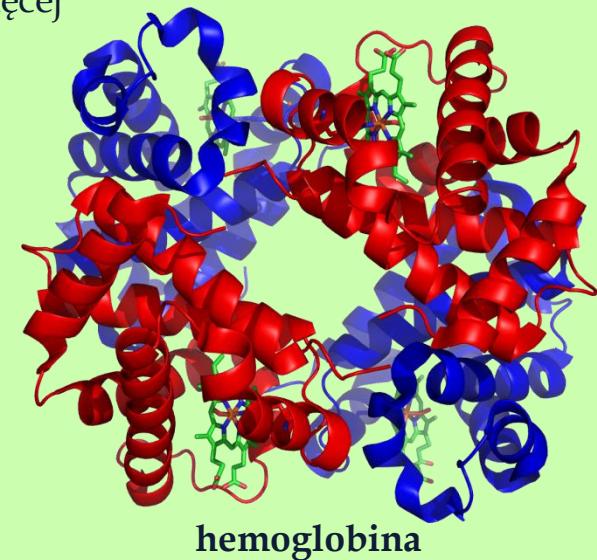


Struktura trzeciorzędowa – wzajemne położenie elementów struktury drugorzędowej



Struktura czwartorzędowa – sposób łączenia się struktur trzeciorzędowych w przestrzeni. Dotyczy jedynie białek zbudowanych z więcej niż jednego łańcucha polipeptydowego

Struktura drugorzędowa – ułożenie przestrzenne łańcucha polipeptydowego na skutek powstawania wewnętrzcząsteczkowych wiązań wodorowych



Nowych leków szukamy głównie w obszarze chemii organicznej

- Chemia organiczna: chemia związków węgla

Układ okresowy pierwiastków

LICZBA ATOMOWA (LICZBA PORZĄDKOWA)

NAZWA

SYMBOL PIERWIASTKA

MASA ATOMOWA

H Wodór 1.008	Be Beryl 9.01	B Bor 10.81	C Węgiel 12.01	N Azot 14.01	O Tlen 15.99	F Fluor 18.99											
Li Lit 6.94	Mg Magnez 24.30	Al Glin 26.98	Si Krzem 28.09	P Fosfor 30.97	S Słarka 32.07	Cl Chlork 35.45											
K Potas 39.12	Ca Wapń 40.08	Sc Skand 44.96	Ti Tytan 47.92	V Wanad 50.94	Cr Chrom 51.98	Mn Mangan 54.94	Fe Żelazo 55.85	Co Kobalt 58.93	Ni Nikiel 58.69	Cu Miedź 63.55	Zn Cynk 65.39	Ga Gal 69.72	Ge German 72.61	As Arsan 74.92	Se Selen 78.96	Br Brom 79.90	
Rb Rubid 85.47	Sr Stront 87.62	Y Iter 88.91	Zr Cyrron 88.22	Nb Niob 89.91	Mo Molibden 95.94	Tc Technet 98.01	Ru Ruten 101.07	Rh Rod 102.91	Pd Pallad 106.42	Ag Srebro 107.87	Cd Kadm 112.41	In Ind 114.82	Sn Cyra 118.71	Sb Antymon 121.75	Te Telur 127.69	I Jod 131.29	
Cs Cez 132.91	Ba Bar 137.33	La Lantan 138.91	Hf Hafn 178.49	Ta Tantal 180.95	W Wolfram 183.86	Re Ren 186.21	Os Osm 190.2	Ir Irid 192.22	Pt Platyna 195.08	Au Złoto 196.97	Hg Rtęć 200.59	Tl Tel 204.39	Pb Ołów 207.2	Bi Bismut 210.98	Po Polon 218.95	At Astat 220.09	Rn Radow 222.02
Fr Franc 223.02	Ra Rad 226.03	Ac Aktyn 227.03	Rf Rutherford 261.1	Db Dubn* 262.1	Sg Seaborg* 263.1	Bh Bohr* 262.1	Hs Has* 265.1	Mt Meltner* 266.1	Uun Ununnilium 268.1	Uuu Unununium 269.1	Uub Ununbium 270.1	Uuq Ununquadium 271.1		Uuh Ununhexium 272.1		Uuo Ununectium 273.1	

* Symbol chemiczny i angielska nazwa pierwiastka zostały zatwierdzone przez IUPAC w 1997 r.
Podano proponowaną nazwę polską.

Lantanowce

Ce Cer 140.12	Pr Przeodym 140.91	Nd Neodym 144.24	Pm Promet 144.91	Sm Samar 151.96	Eu Ewrop 151.96	Gd Gadolin 157.25	Tb Terb 158.93	Dy Dysproz 162.90	Ho Holm 164.93	Er Erb 167.26	Tm Tul 169.93	Yb Iterb 173.04	Lu Lutet 174.97
----------------------------	---------------------------------	-------------------------------	-------------------------------	------------------------------	------------------------------	--------------------------------	-----------------------------	--------------------------------	-----------------------------	----------------------------	----------------------------	------------------------------	------------------------------

Aktynowce

Th Tor 232.04	Pa Protaktyn 231.04	U Uran 238.03	Np Neptun 237.05	Pu Pułson 244.06	Am Ameryk 243.06	Cm Kur 247.07	Bk Berkel 247.07	Cf Kaliforn 251.08	Es Einstein 252.08	Fm Ferm 257.09	Md Mandskev 258.1	No Nobel 258.10	Lr Lorenz 260.11
----------------------------	----------------------------------	----------------------------	-------------------------------	-------------------------------	-------------------------------	----------------------------	-------------------------------	---------------------------------	---------------------------------	-----------------------------	--------------------------------	------------------------------	-------------------------------

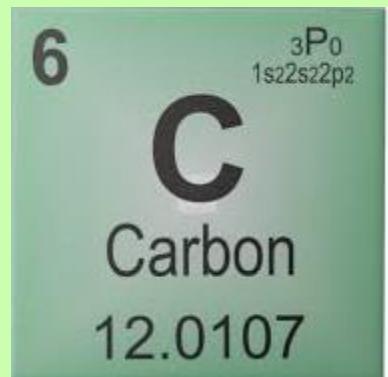
METALE

NIEMETALE

NIEMETALE – GAZY SZLACHETNE

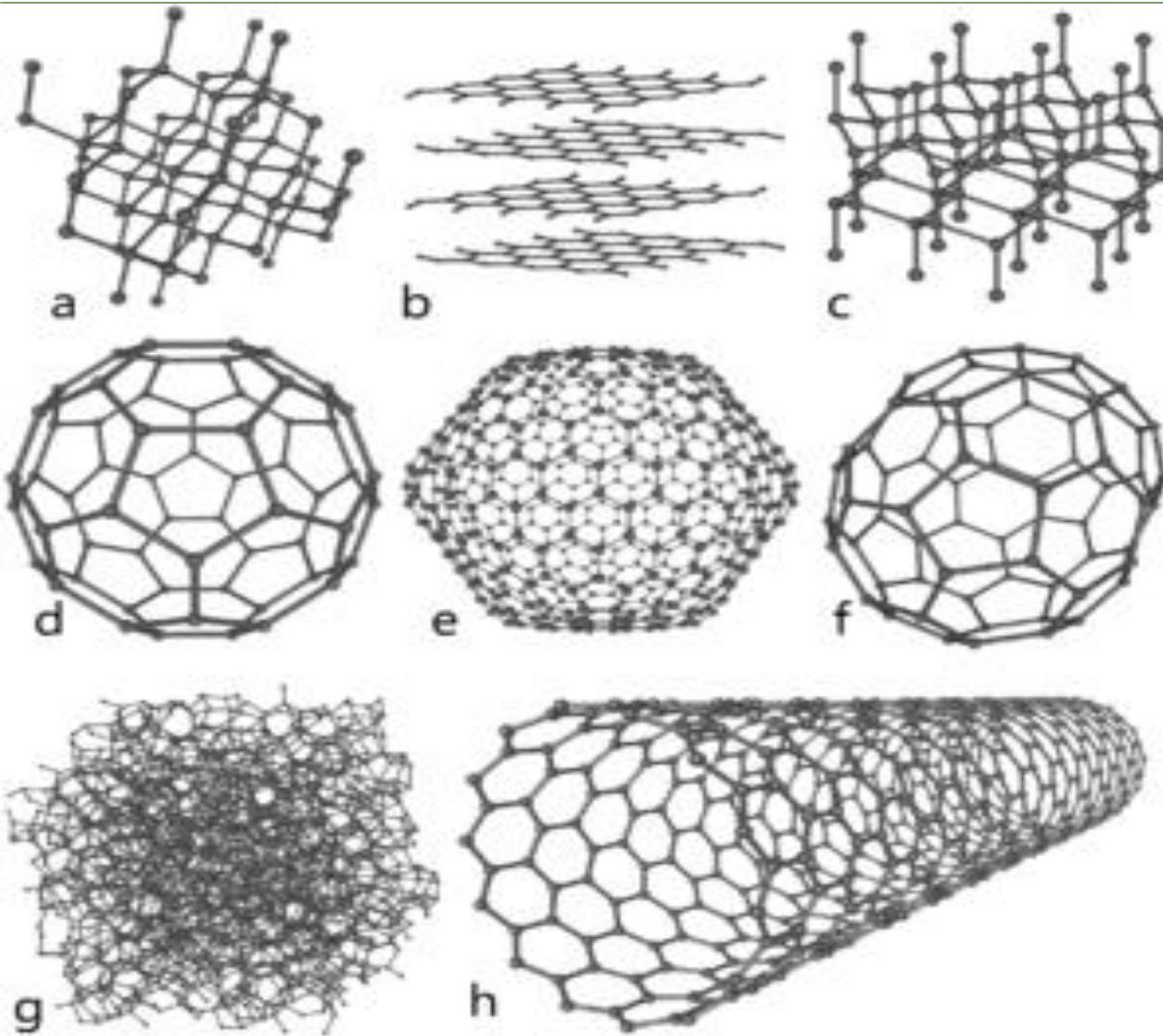
PÓŁMETALE

Węgiel - charakterystyka pierwiastka



- W stanie wolnym występuje w niewielkich ilościach – głównie jako grafit i diament
- w stanie związanym – m.in. buduje tkanki roślin i zwierząt oraz jest składnikiem paliw kopalnych; najważniejsze związki nieorganiczne: dwutlenek węgla (CO_2), tlenek węgla (CO), kwas węglowy (H_2CO_3) oraz jego sole
- Podstawowy pierwiastek organizmów żywych: składnik białek, kwasów nukleinowych, lipidów i innych związków

Węgiel – odmiany alotropowe



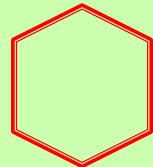
- a) Diament
- b) Grafit
- c) Lonsdaleit
- d) Fuleren C60
- e) Fuleren C540
- f) Fuleren C70
- g) Węgiel amorficzny
- h) Nanorurka

Węgiel - zastosowanie pierwiastka

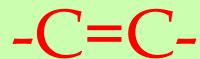
- Podstawa życia na Ziemi z uwagi na stanowienie podstawy budulcowej związków organicznych
- Paliwo kopalne, ważne źródło energii nieodnawialnej
- Izotop ^{14}C wykorzystywany w tzw. datowaniu radiowęglowym
- Grafit - poligrafia (produkcja ołówków), materiał do produkcji elektrod
- Diament - kamień szlachetny (jubilerstwo), z uwagi na wysoką twardość stosowany w urządzeniach pomiarowych, narzędziach do skrawania, szlifowania i wiercenia
- Węgiel bezpostaciowy - medycyna, zastosowanie w procesach filtracji i oczyszczania, katalizator/nośnik dla katalizatorów

Różnorodność związków organicznych - przyczyny

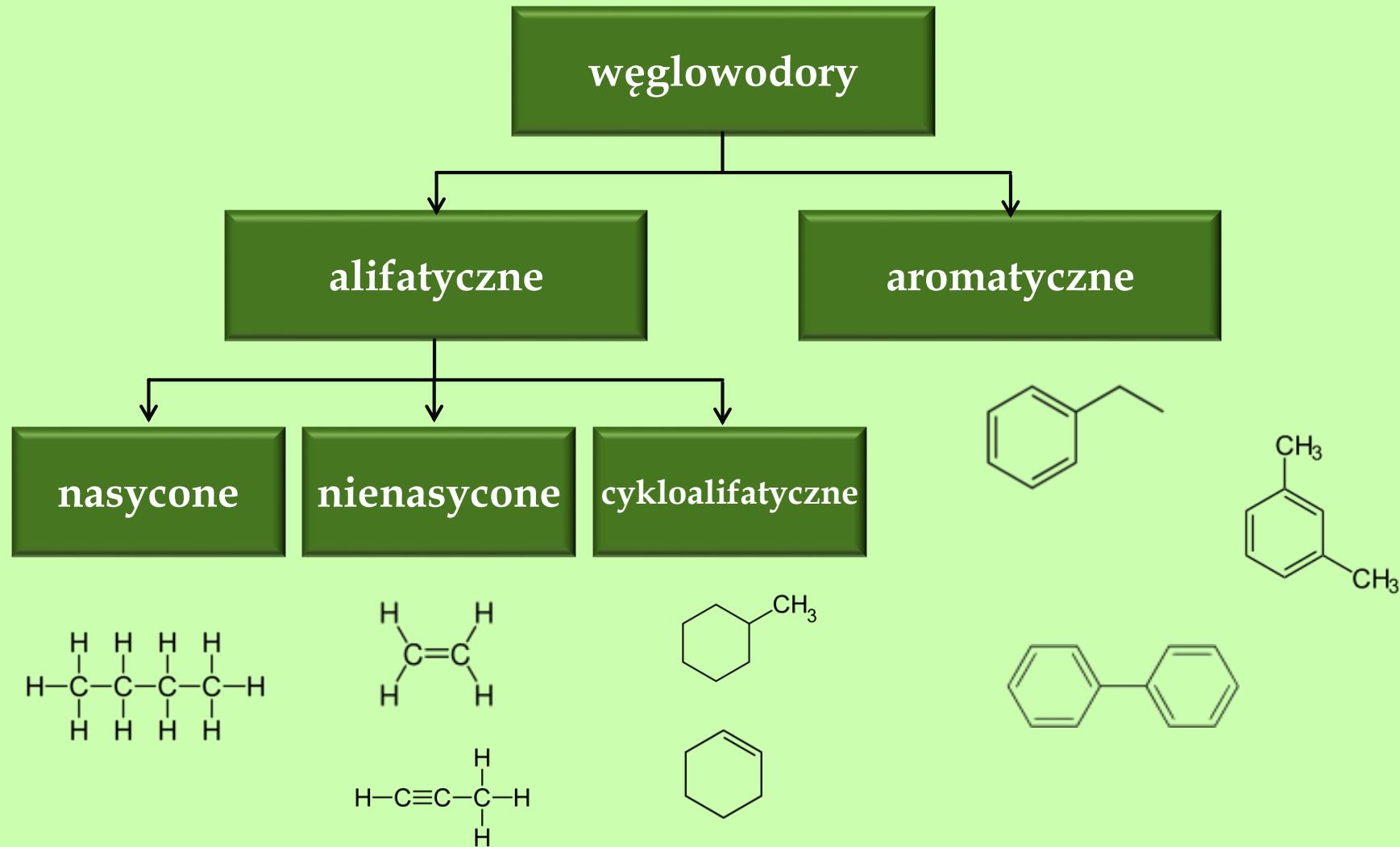
- Duża trwałość wiązań węgla z wodorem (średnia elektroujemność węgla)
- Zdolność do tworzenia połączeń pomiędzy atomami węgla - możliwość tworzenia długich łańcuchów i pierścieni



- Możliwość występowania wiązań wielokrotnych pomiędzy atomami węgla

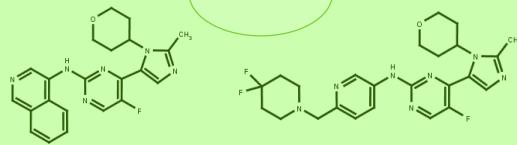
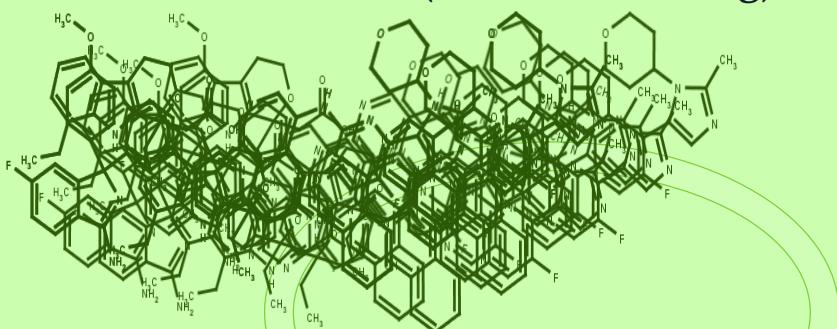


Węglowodory – związki węgla z wodorem

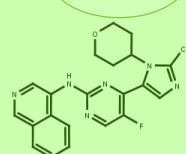
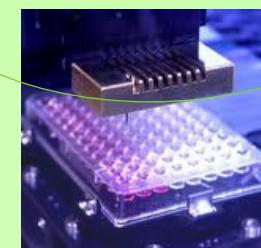
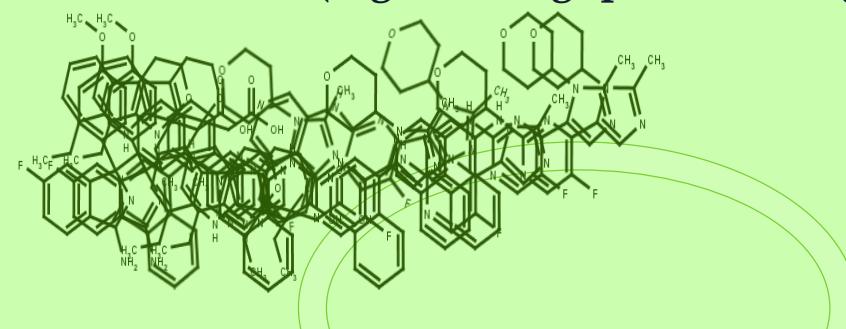


Wirtualny skrining

Wirtualny skrining
(virtual screening)



Wysokowydajne badanie przesiewowe
(high -throughput screening)



Bazy danych - ChEMBL



<https://www.ebi.ac.uk/chembl/>

Serotonin receptor 1A

Showing 1-11 out of 11 records

<input type="checkbox"/> ChEMBL ID	Search Hit	Name	UniProt Accessions	Type	Organism	Compounds	Activities
<input type="checkbox"/> CHEMBL3737		Serotonin 1a (5-HT1a) receptor	Q64264	SINGLE PROTEIN	Mus musculus	 127 By Mol. Wt.: 193 By Std. Type:	 193 By Std. Type:
<input type="checkbox"/> CHEMBL273		Serotonin 1a (5-HT1a) receptor	P19327	SINGLE PROTEIN	Rattus norvegicus	 5941 By Mol. Wt.: 8584 By Std. Type:	 8584 By Std. Type:
<input type="checkbox"/> CHEMBL214		Serotonin 1a (5-HT1a) receptor	P08908	SINGLE PROTEIN	Homo sapiens	 6460 By Mol. Wt.: 10016 By Std. Type:	 10016 By Std. Type:
<input type="checkbox"/> CHEMBL2111460		Dopamine D2 receptor and serotonin 1a receptor	P14416, P08908	SELECTIVITY GROUP	Homo sapiens	 10 By Mol. Wt.: 10 By Std. Type:	 10 By Std. Type:

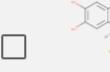
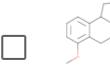
Bazy danych - ChEMBL



<https://www.ebi.ac.uk/chembl/>

ChEMBL

Search in ChEMBL

Molecule ChEMBL ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL Value	Comment	Assay ChEMBL ID	Assay Description
 CHEMBL288536	12a	Ki	>	5000.0	nM	No Data	No Data	CHEMBL616353	Binding affinity against hydroxytryptamine
 CHEMBL416266	EF-7412	Ki	=	27.0	nM	7.57	No Data	CHEMBL803256	Binding affinity toward HT1A receptor
 CHEMBL341326	12b	Ki	No Data	No Data	No Data	No Data	Not Determined	CHEMBL616388	Binding affinity toward hydroxytryptamine by the displacement of [3H]-DPAT in mouse hippocampus; Not
 CHEMBL10570	(+/-)-7d	Ki	>	333.0	nM	No Data	No Data	CHEMBL616562	Compound was evaluated for the binding affinity at [3H]-8-OH-DPAT-labeled hydroxytryptamine receptors in cloned CHC

Bazy danych - PDB



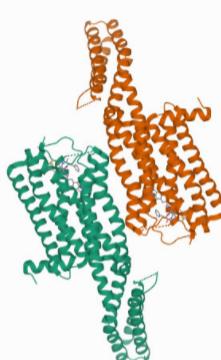
<https://www.rcsb.org/>

Baza struktur krystalicznych białek

Structure Summary 3D View Annotations Experiment Sequence Genome Ligands Versions

Display Files Download Files

Biological Assembly 1



4IAQ

Crystal structure of the chimeric protein of 5-HT1B-BRIL in complex with dihydroergotamine (PSI Community Target)

DOI: 10.2210/pdb4IAQ/pdb

Classification: Signaling Protein, Electron Transport

Organism(s): Homo sapiens, Escherichia coli

Expression System: Spodoptera frugiperda

Mutation(s): Yes

Membrane Protein: Yes

3D View | Electron Density | Ligand Interaction | Predict Membrane

Global Symmetry: Cyclic - C2 (3D View)

Global Stoichiometry: Homo 2-mer - A2

Find Similar Assemblies

Biological assembly 1 assigned by authors and

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.80 Å

R-Value Free: 0.257

R-Value Work: 0.236

R-Value Observed: 0.237

wwPDB Validation

3D Report | Full Report

Metric	Percentile Ranks	Value
Rfree	Worse	0.292
Clashscore	6	0.3%
Ramachandran outliers	9.3%	0.3%
Sidechain outliers	Worse	6.3%
RSRZ outliers	Percentile relative to all X-ray structures	0.292

Bazy danych - ZINC



<https://zinc.docking.org/>

Baza komercyjnie dostępnych związków chemicznych

ZINC 12

Not Authenticated — sign in

Active cart: Temporary Cart (0 items)

About Search Subsets Help Social 8+1 1 Quick Search Bar... Go

Tiles List Diversity Purchasable SC BB NP MOD Agents PBCs Annotated Collabocules BKS EFI

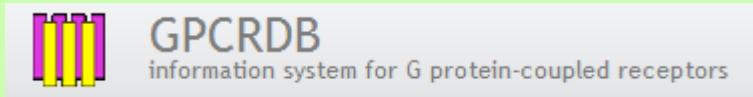
All Catalogs: Overview (295 in total)

Here are all source catalogs in ZINC in a compact tile view. For more information, to browse or download, click on the catalog. This includes both vendor catalogs containing molecules you may buy as well as annotated compounds that have biological activity but are not for sale.

These databases have been provided to us under collaborative agreements with each respective catalog owner. Catalogs have been filtered for drug discovery relevance, and thus compounds in the original catalog may not be in ZINC. Some documentation on ZINC subsets is available.

ZINC may be used free of charge for research by individuals and institutions. Whereas you are free to share the results of a ZINC search or a screen of molecules from ZINC, you may not redistribute major portions of ZINC without the express written permission of John Irwin.

Bazy danych - GPCRdb



<https://gpcrdb.org/>

Baza danych o receptorach sprzężonych z białkiem G

GPCRdb Drugs & ligands Structure constructs Biased signalling GproteinDb ArrestinDb Cite us Join us Quick search

Selected targets: Peptide receptors (Family)

Selected segments: 1x50 [GPCRdb] (Residue), 2x50 [GPCRdb] (Residue), 3x55 [GPCRdb] (Residue), 3x59 [GPCRdb] (Residue), 7x45 [GPCRdb] (Residue), 7x48 [GPCRdb] (Residue), 7x49 [GPCRdb] (Residue), 7x50 [GPCRdb] (Residue) + Save residue positions

Custom RESIDUE Z-SCALES

RESIDUE PROPERTIES & LENGTH

Alignments and Trees Segment and residue customized sequence alignments and phylogenetic trees.

GPCRdb contains reference data, interactive visualisation and experiment design tools for G protein-coupled receptors (GPCRs). GPCRdb curates sequence alignments, structures and receptor mutations from literature. Interactive diagrams visualise receptor residues (e.g. snakeplot and helix box plot) and relationships (e.g. phylogenetic trees). See the GPCRdb graphical overview, video, articles or documentation.

GproteinDb and ArrestinDb will be published as separate web resources.

GPCRdb is developed by the David Gloriam group, integrates data from the large GPCR community and maintains an active collaboration with the Guide To Pharmacology (IUPHAR), GPCRmd and ERNEST. If you use the GPCRdb, please remember to cite us.

GPCRdb statistics

424 Human proteins

Latest release

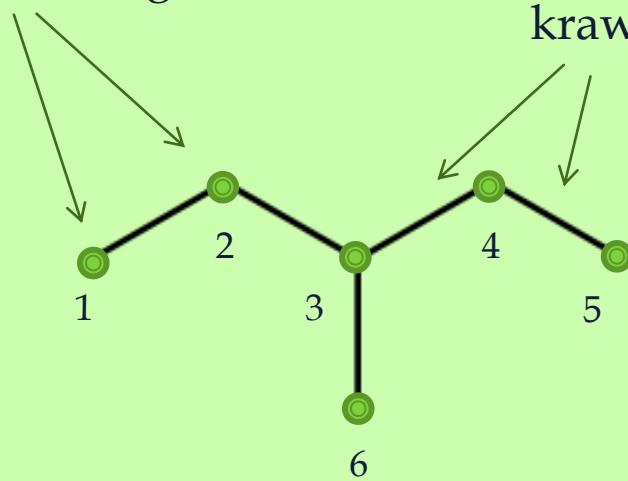
Sept. 10, 2021

Tweets by @gpcrdb

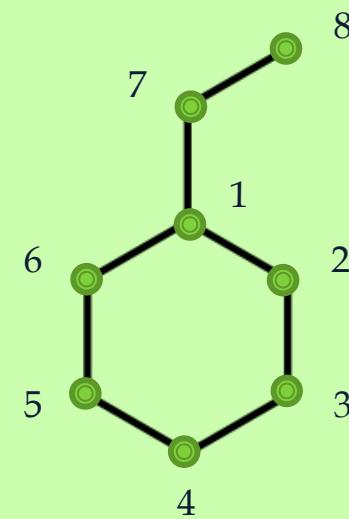
Kodowanie cząsteczek

- Grafy cząsteczek chemicznych

wierzchołki grafu

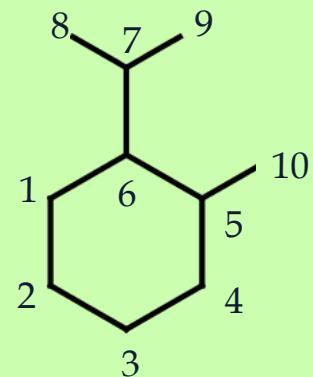
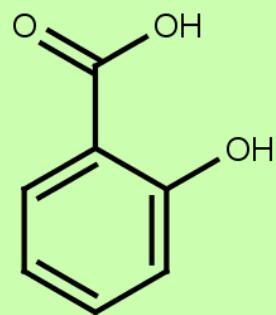


krawędzie grafu



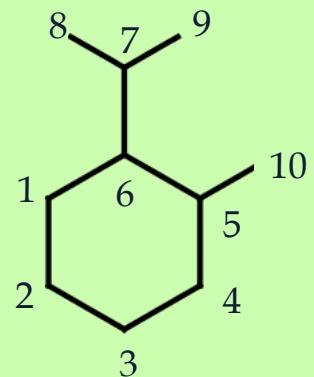
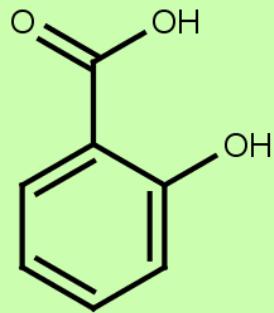
Kodowanie cząsteczki - systemy macierzowe

Macierz sąsiedztwa



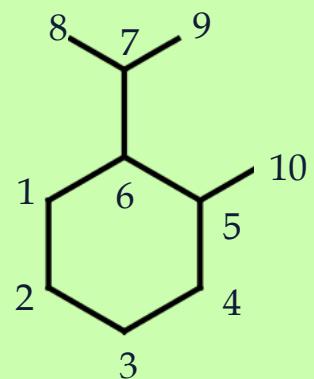
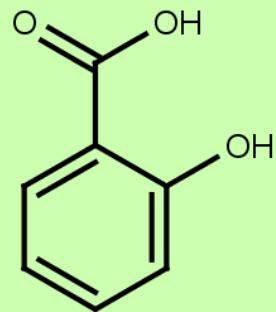
Kodowanie cząsteczki - systemy macierzowe

Macierz odległości topologicznej



Kodowanie cząsteczki – systemy macierzowe

Macierz wiązań



n	1	2	3	4	5	6	7	8	9	10
1	0	2				1				
2	2	0	1							
3		1	0	2						
4			2	0	1					
5				1	0	2				1
6	1				2	0	1			
7						1	0	1	2	
8							1	0	1	
9							2		0	
10					1					0

Przechowywanie informacji strukturalnej – format sdf

Mrv0541 01151319442D

```
10 10 0 0 0 0          999 V2000
 -1.6721 -0.1120 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.9576 -0.5245 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.9576 -1.3495 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -1.6721 -1.7620 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -2.3866 -1.3495 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -2.3866 -0.5245 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -1.6721 0.7130 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -2.3866 1.1255 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.9576 1.1255 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.2432 -0.1120 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1 2 1 0 0 0 0 0
 1 6 2 0 0 0 0 0
 2 3 2 0 0 0 0 0
 3 4 1 0 0 0 0 0
 4 5 2 0 0 0 0 0
 5 6 1 0 0 0 0 0
 1 7 1 0 0 0 0 0
 7 8 2 0 0 0 0 0
 7 9 1 0 0 0 0 0
 2 10 1 0 0 0 0 0
```

M END

> <CdId>

5

> <Mol Weight>

138.1207

> <Formula>

C7H6O3

} nagłówek

} blok atomów

} blok wiązań

} blok właściwości

\$\$\$\$

Kodowanie cząsteczek

Notacja SMILES (ang. *simplified molecular input entry system*)

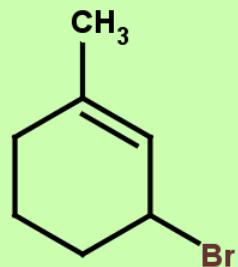
- Kodowanie cząsteczek związków chemicznych w postaci liniowej
związek chemiczny -> SMILES – **niejednoznaczne** (dla danego
związku chemicznego może istnieć więcej niż jeden poprawny
SMILES, np. CCO, OCC, C(O)C dla etanolu)
- Dokodowanie:
SMILES -> związek chemiczny – **jednoznaczne**
- Generowanie jednoznacznych ciągów SMILES – algorytmy c14n;
różne w zależności od wykorzystanego oprogramowania

Kodowanie cząsteczek

Notacja SMILES (ang. *simplified molecular input entry system*)

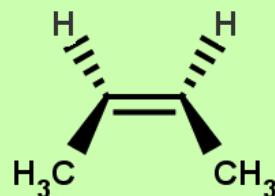
- atomy kodowane przy pomocy ich symboli atomowych
- atomy pierścieni aromatycznych zapisywane są z małej litery
- typy wiązań chemicznych – znaki specjalne -(pojedyncze), = (podwójne), # (potrójne), : (aromatyczne)
- rozróżnianie konformerów cis/trans: znaki specjalne / i \
- tetraedryczne centra chiralności: znaki @@ lub @ - wskazują kolejność podstawników zgodną i niezgodną z ruchem wskazówek zegara

Kodowanie cząsteczek – notacja SMILES



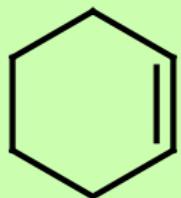
(C)C1=CC(Br)CCC1

3-bromo-1-metylocyloheksen



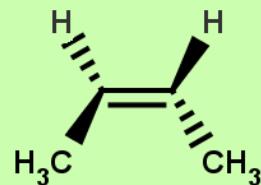
C/C=C\ C

cis-but-2-en



C1=CCCCCC1

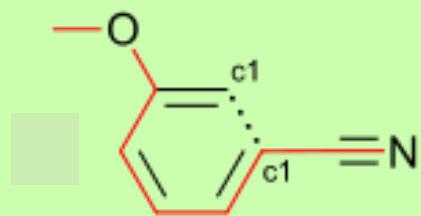
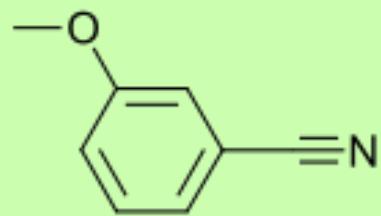
cykloheksen



C/C=C/C

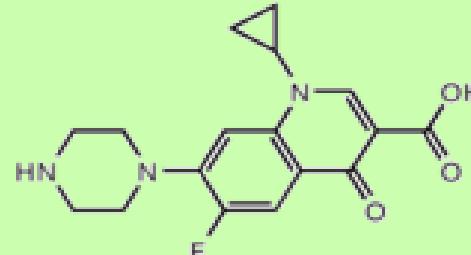
trans-but-2-en

Kodowanie cząsteczek – notacja SMILES

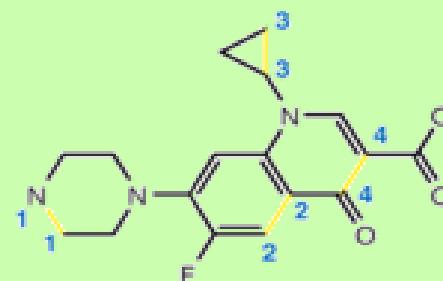


COc(c1)cccc1C#N

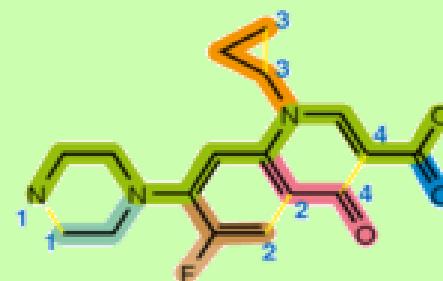
A



B



C



D

N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

Notacja SMARTS

- naturalne rozszerzenie notacji SMILES
- notacja dedykowana przeszukiwaniom podstrukturalnym
- wykorzystuje operatory logiczne do tworzenia bardziej ogólnych opisów

Symbol	Znaczenie
*	Dowolny atom
a	Atom należący do ugrupowania aromatycznego
A	Atom należący do ugrupowania alifatycznego
r<n>	Pierścień zbudowany z n atomów

Symbol	Znaczenie
-	Wiązanie pojedyncze
~	Dowolne wiązanie
@	Wiązanie w pierścieniu

Grafy orbitali molekularnych

- Bazują na grafach uwzględniających obecność atomów wodoru
- Każdy atom jest reprezentowany przez odpowiadające mu orbitale atomowe:

H: $1s^1$

C: $1s^2, 2s^2, 2p^2$ (trzy wierzchołki)

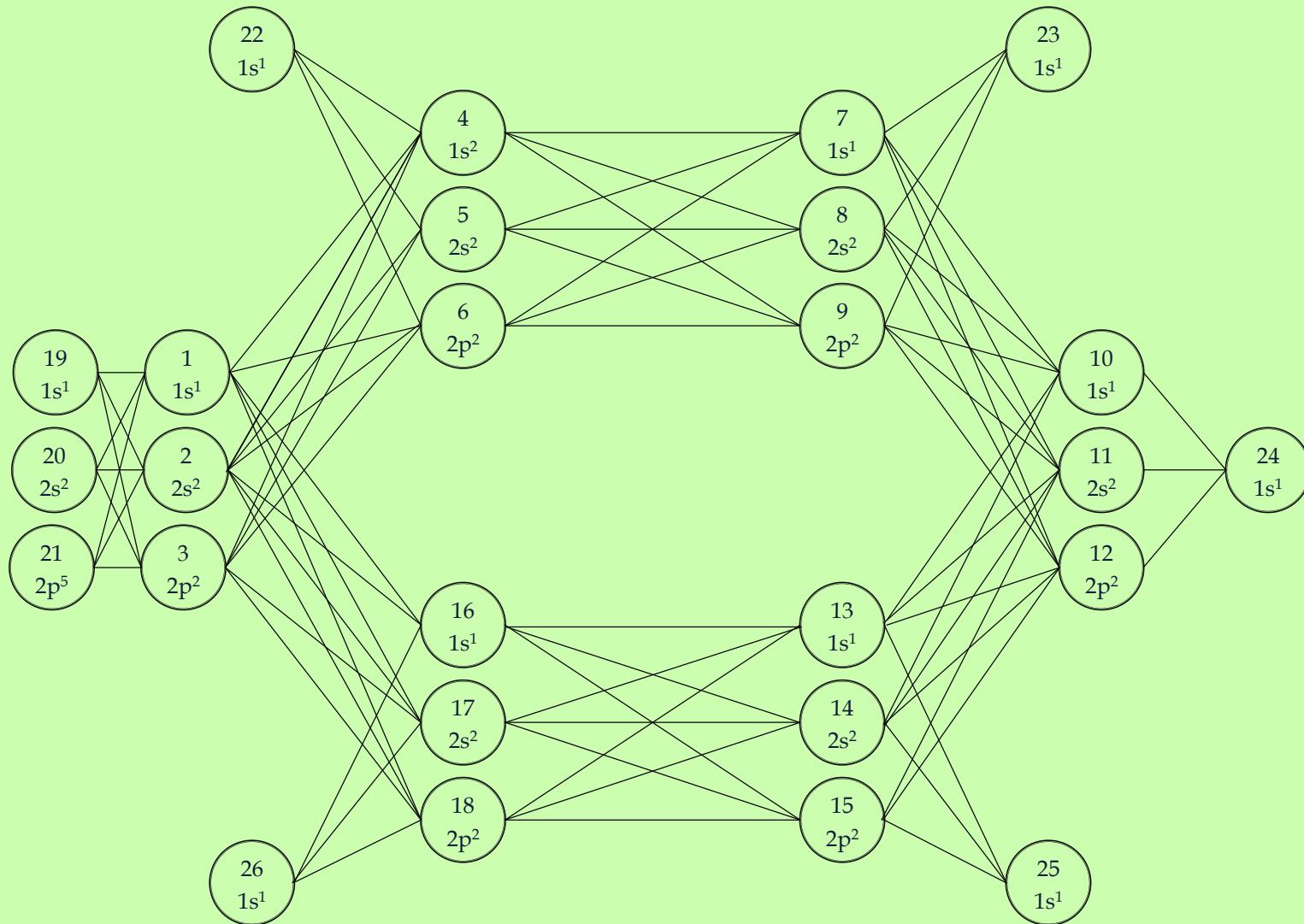
O: $1s^2, 2s^2, 2p^3$ (trzy wierzchołki)

N: $1s^2, 2s^2, 2p^4$ (trzy wierzchołki)

Cl: $1s^2, 2s^2, 2p^6, 3s^2, 3p^5$ (pięć wierzchołków)

Br: $1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^5$ (siedem wierzchołków)

Grafy orbitali molekularnych - przykład (fluorobenzen)

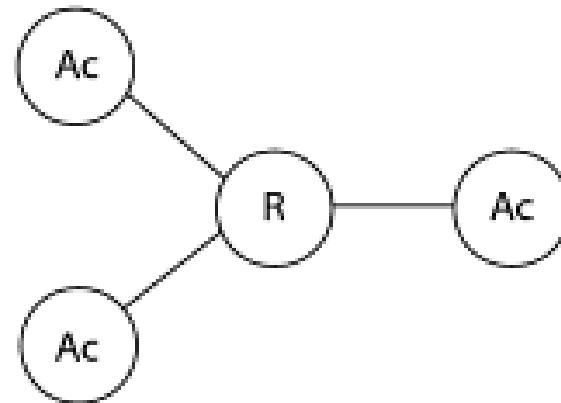
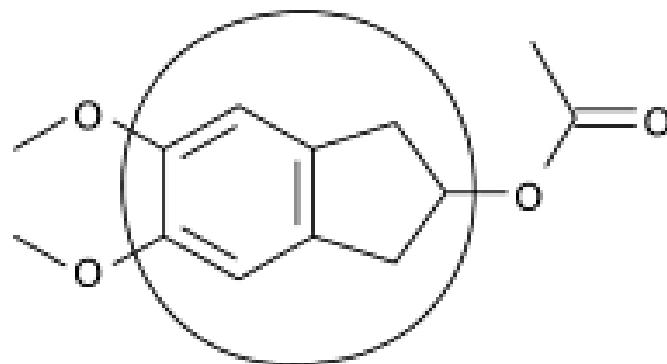


Reprezentacje grafowe - grafy zredukowane

- wierzchołek – grupa powiązanych atomów

Typ I:

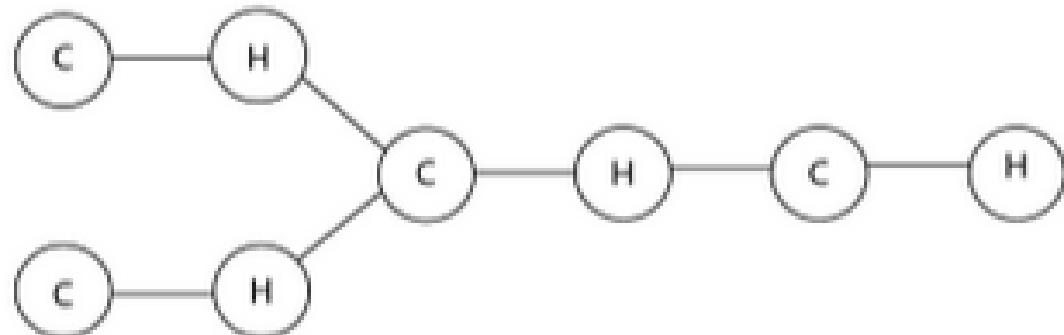
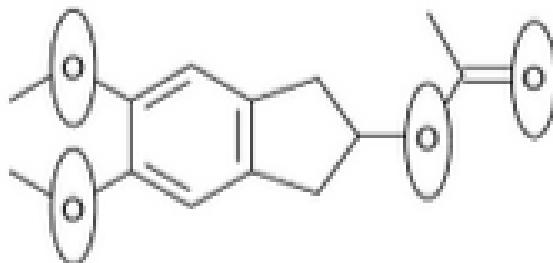
wierzchołki – układy pierścieniowe
(R)/przyłączone grupy acykliczne (Ac)



Reprezentacje grafowe - grafy zredukowane

Typ II:

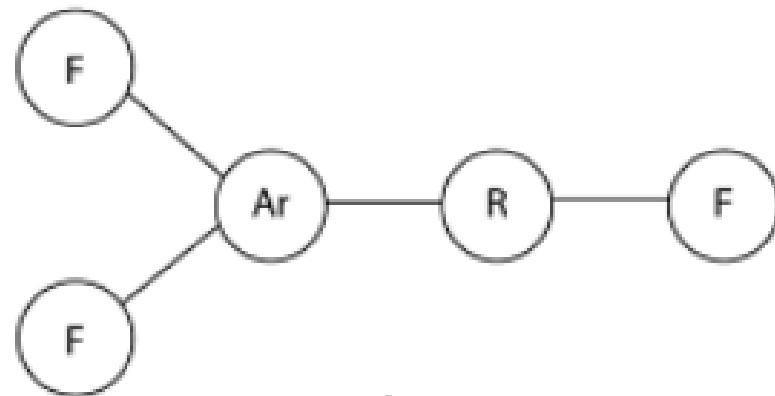
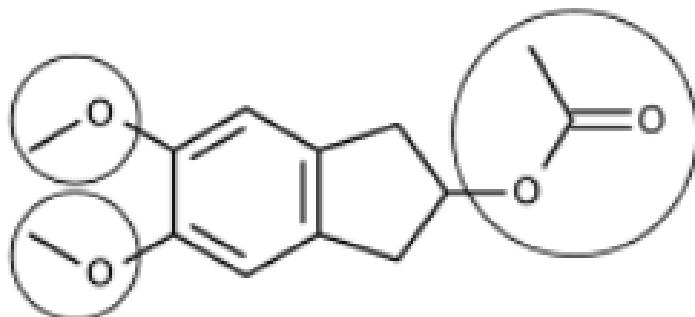
wierzchołki – ugrupowania zawierające węgiel (C)/heteroatomy (H)



Reprezentacje grafowe - grafy zredukowane

Typ III:

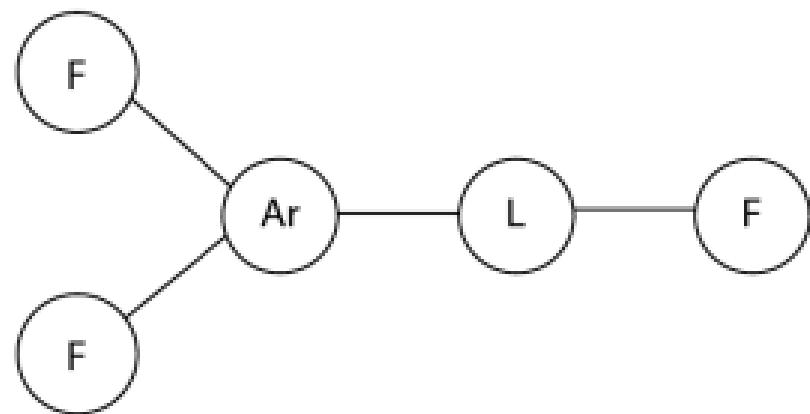
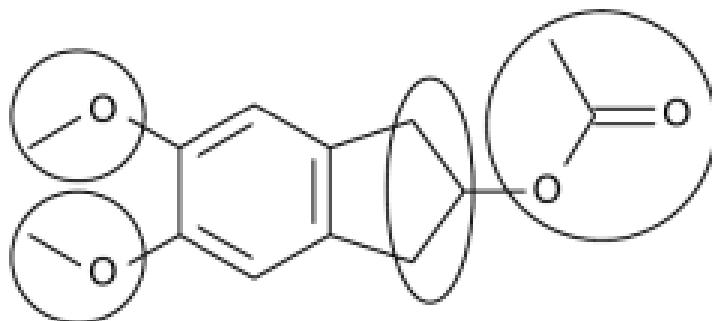
wierzchołki – pierścienie aromatyczne (Ar)/pierścienie alifatyczne (R), grupy funkcyjne (F)



Reprezentacje grafowe - grafy zredukowane

Typ IV:

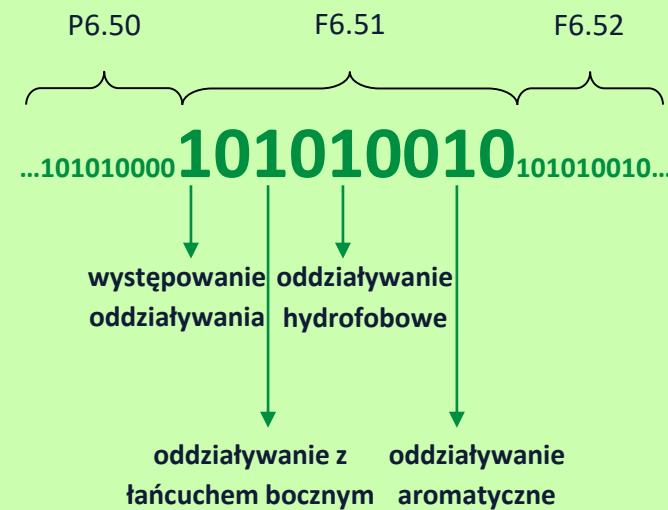
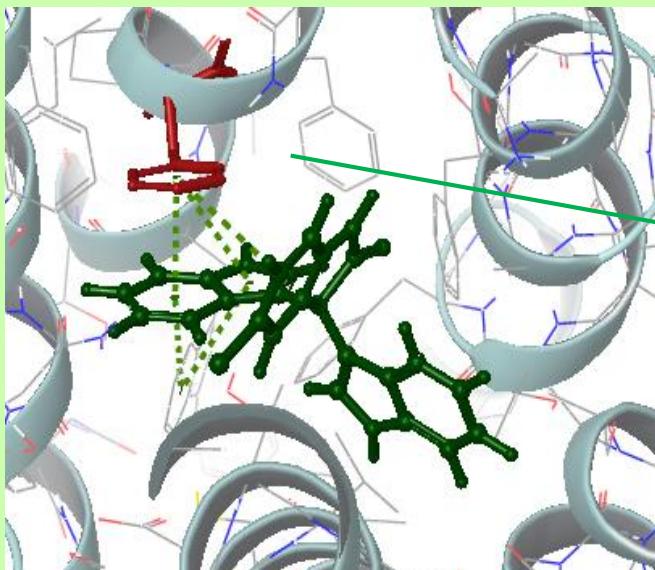
wierzchołki – pierścienie aromatyczne
(Ar)/grupy funkcyjne (F)/linkery (L)



Fingerprinty Oddziaływań Strukturalnych (Structural Interaction Fingerprints, SIFT)



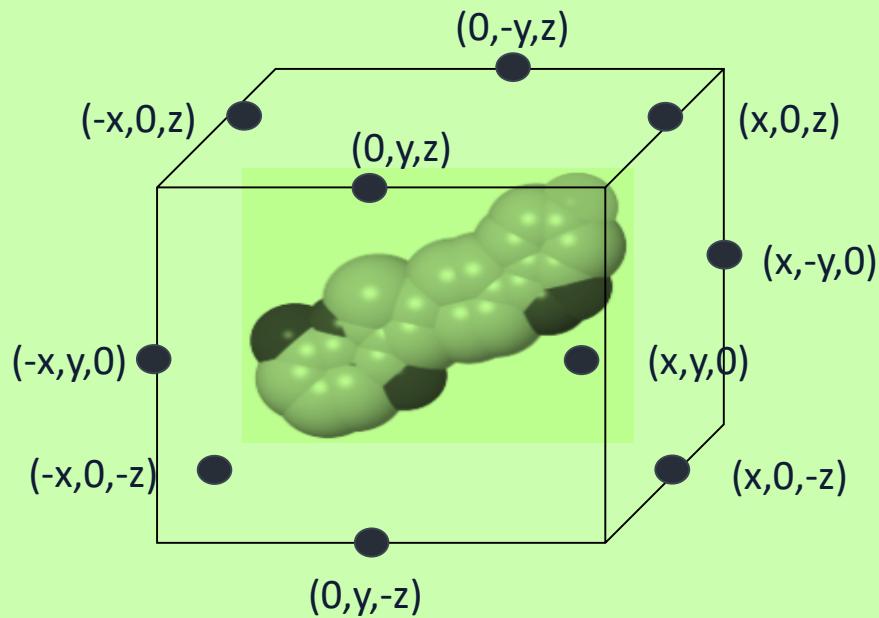
Fingerprinty Oddziaływań Strukturalnych (Structural Interaction Fingerprints, SIFT)



Deskryptory molekularne

- Przypisują różnym cechom analizowanych cząsteczek wartości numeryczne.
- Klasyfikacja:
 - a) w oparciu o wymiarowość:
 - deskryptory jednowymiarowe (1D) - generowane ze wzoru sumarycznego, np. masa molowa
 - deskryptory dwuwymiarowe (2D) - generowane z płaskiej struktury związku
 - deskryptory trójwymiarowe (3D) - generowane z trójwymiarowej struktury związku
 - b) w oparciu o odzwierciedlane właściwości (konstytucyjne, topologiczne, geometryczne, elektrostatyczne, kwantowo-mechaniczne)

Deskryptory typu Spectrophores

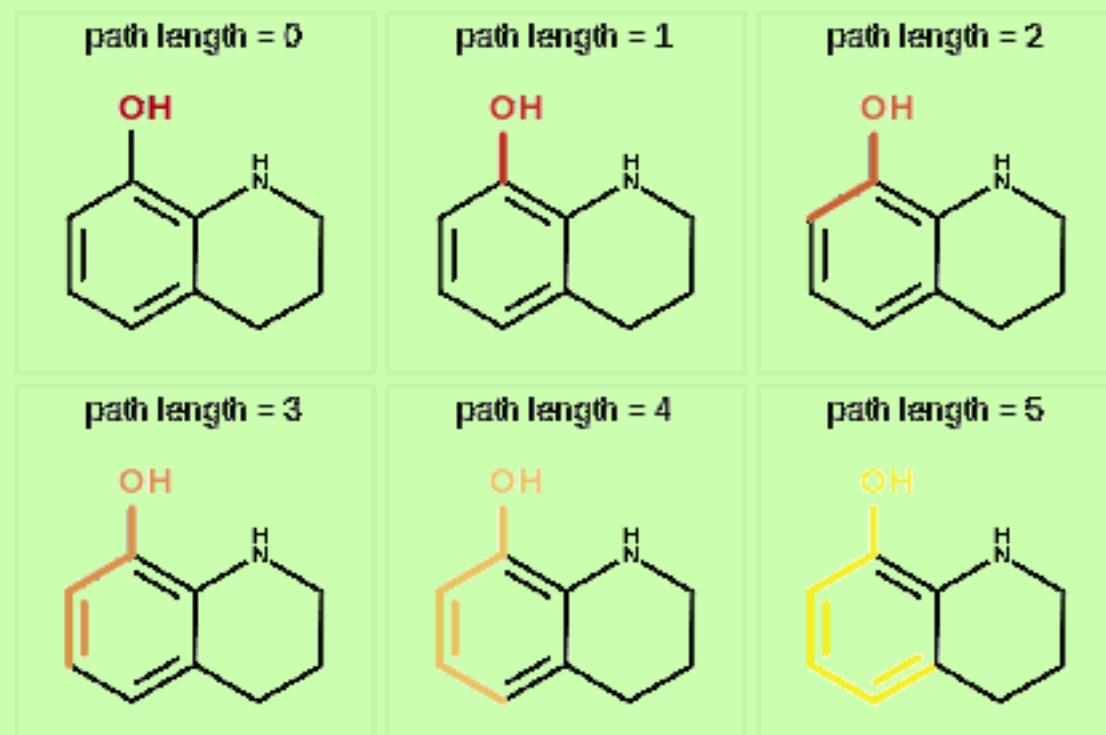
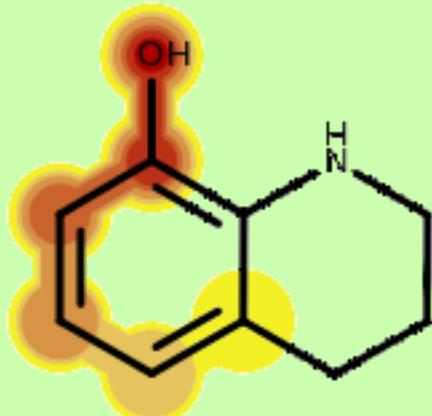


Spectrophores:

- atomic partial charges
- atomic lipophilicity properties
- atomic shape deviations
- atomic electrophilicity properties

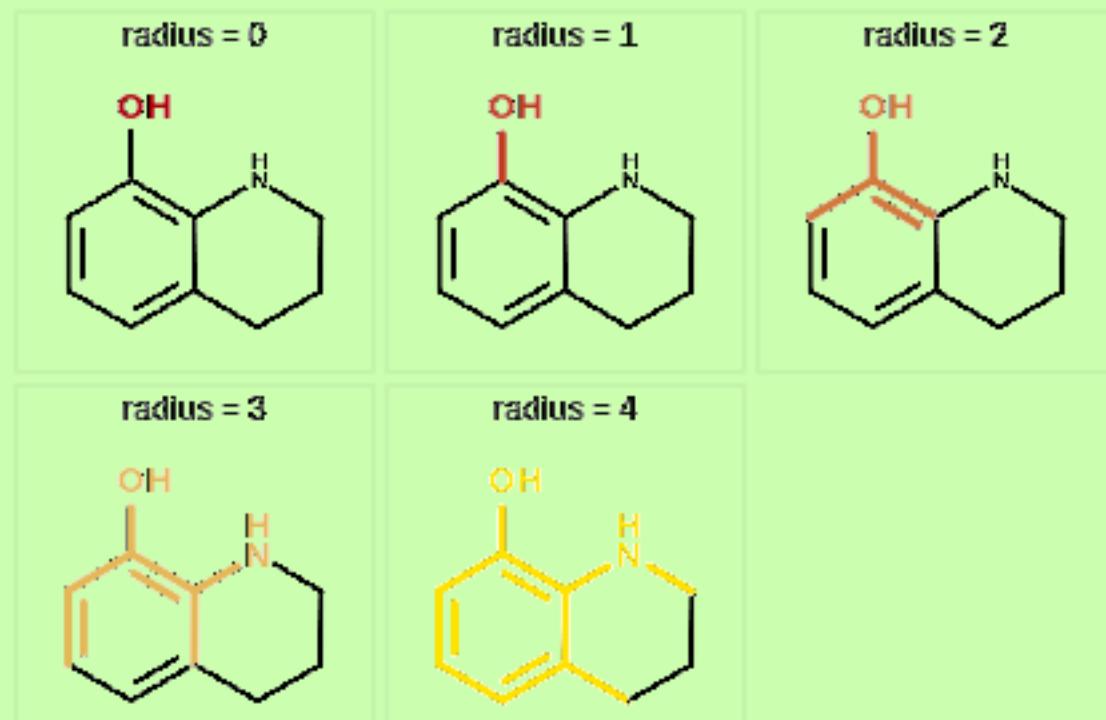
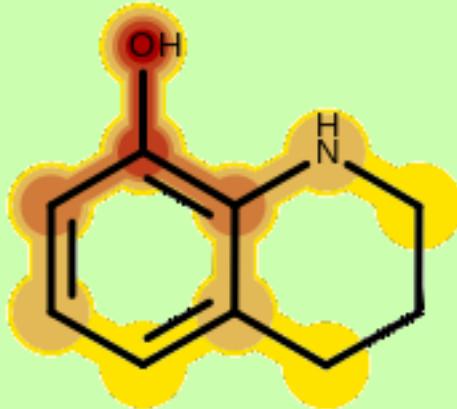
Fingerprinty haszowane

I. Oparte o generowanie ścieżek o określonej długości



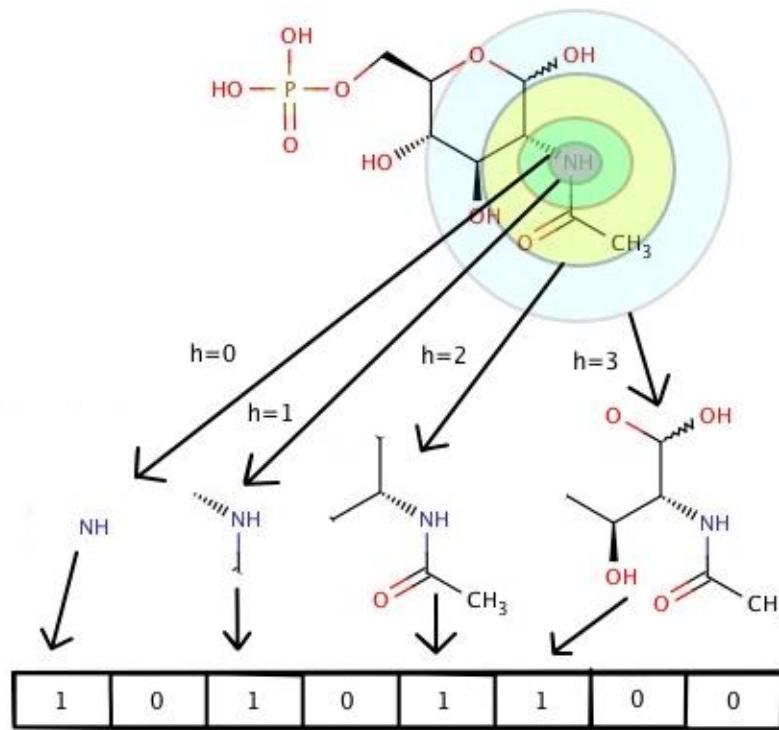
Fingerprinty haszowane

II. Oparte o generowanie fragmentów o początku w każdym ciężkim atomie, pokrywających obszar kołowy o rosnącym promieniu



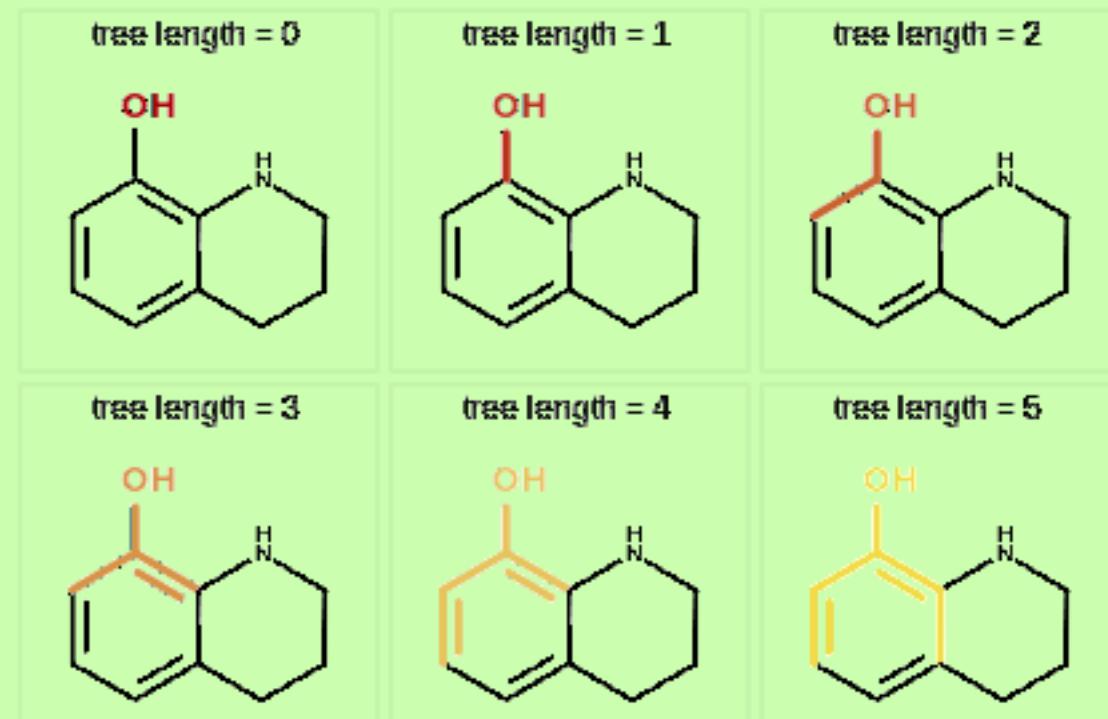
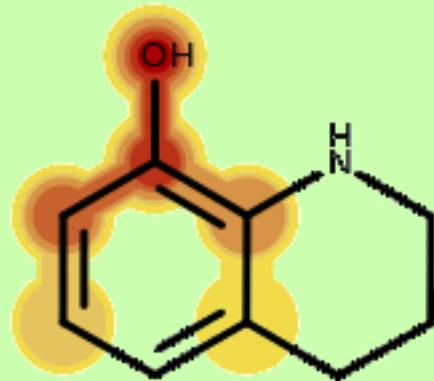
Fingerprinty haszowane

II. Oparte o generowanie fragmentów o początku w każdym ciężkim atomie, pokrywających obszar kołowy o rosnącym promieniu



Fingerprinty haszowane

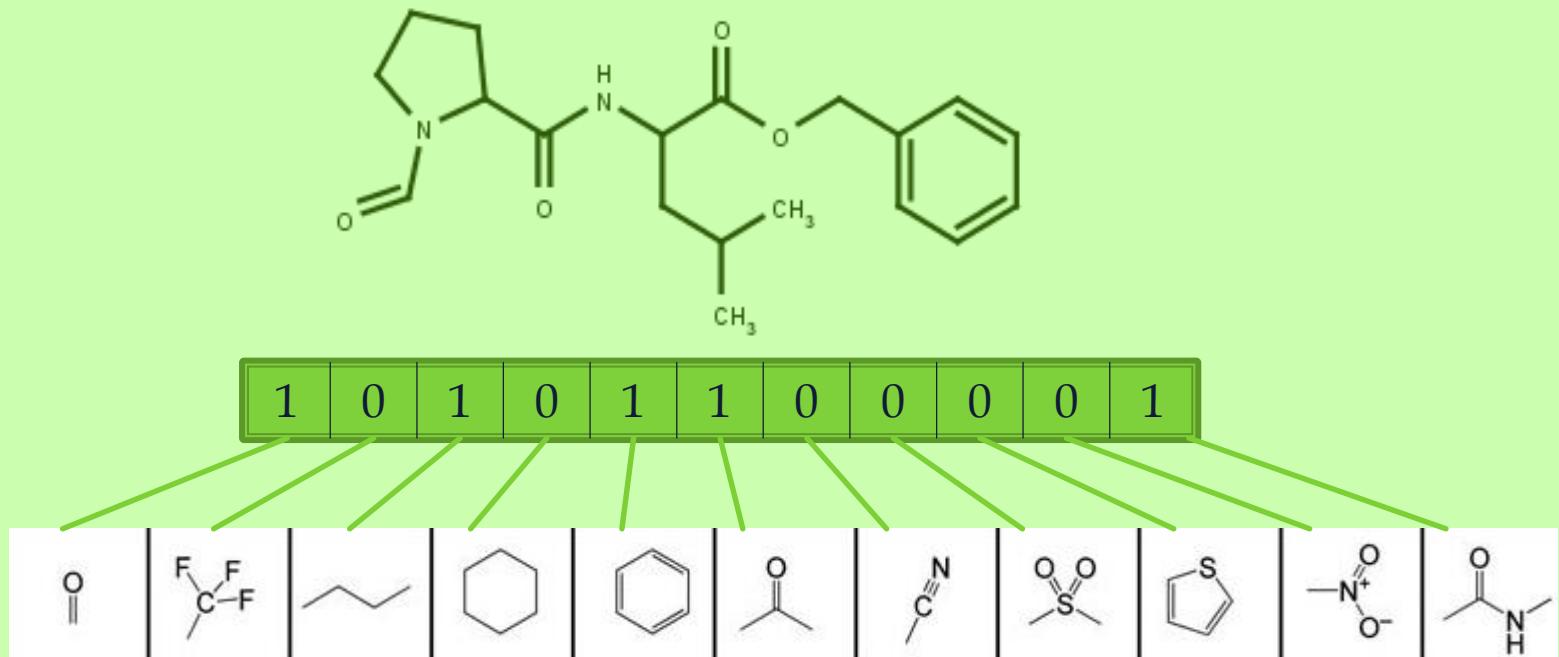
III. Oparte o generowanie wszystkich drzew grafu



Fingerprinty oparte na kluczach

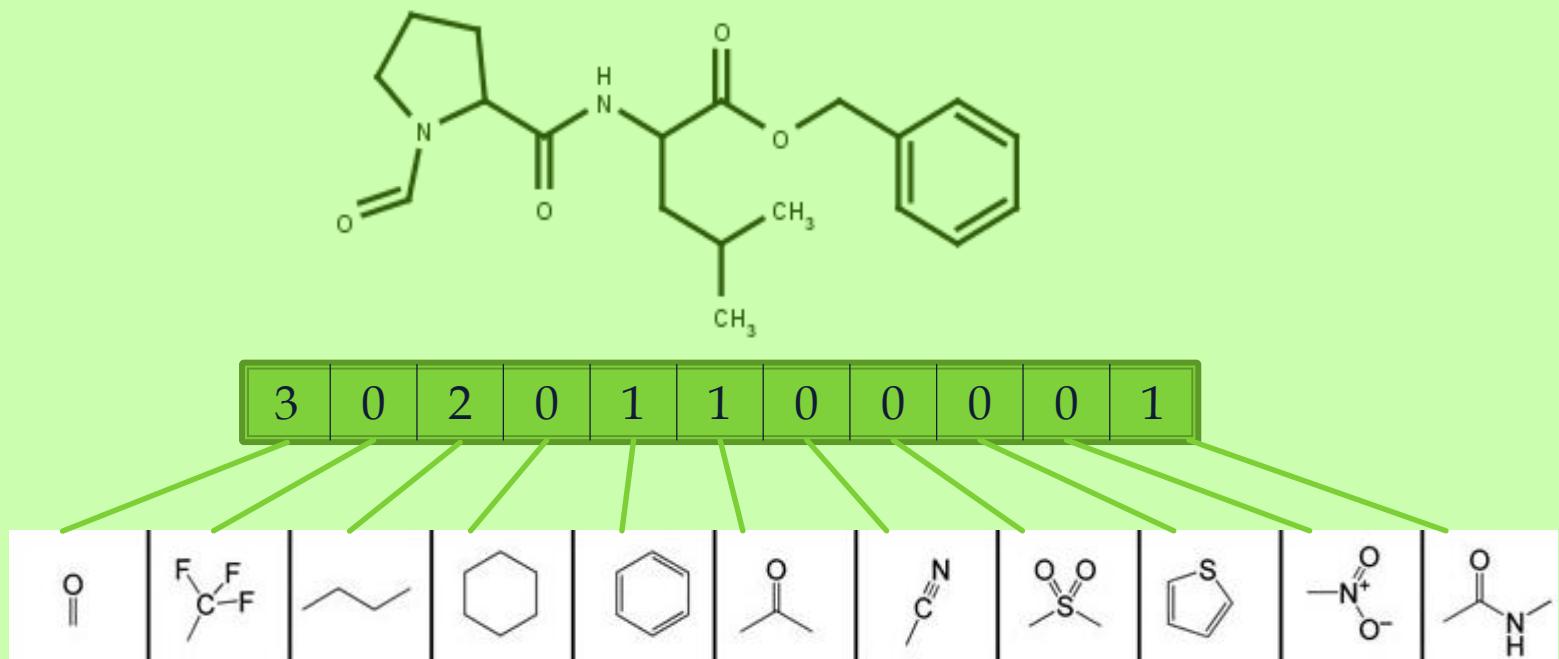
- Każda pozycja reprezentuje obecność lub brak określonej cechy w molekule

SCHEMAT POWSTAWANIA FINGERPRINTU PODSTRUKTURALNEGO:



Fingerprinty podstrukturalne „zliczeniowe”

- Podają ilość wystąpień określonej podstruktury w cząsteczce



Atom pairs/topological torsion fingerprints

- Koduję wszystkie możliwe pary atomów wraz z liczbą wiązań, która je rozdziela (atom pairs)/grupy „czteroatomowe” tworzące kąt torsyjny (topological torsion)

Atomtype—Distance—Atomtype

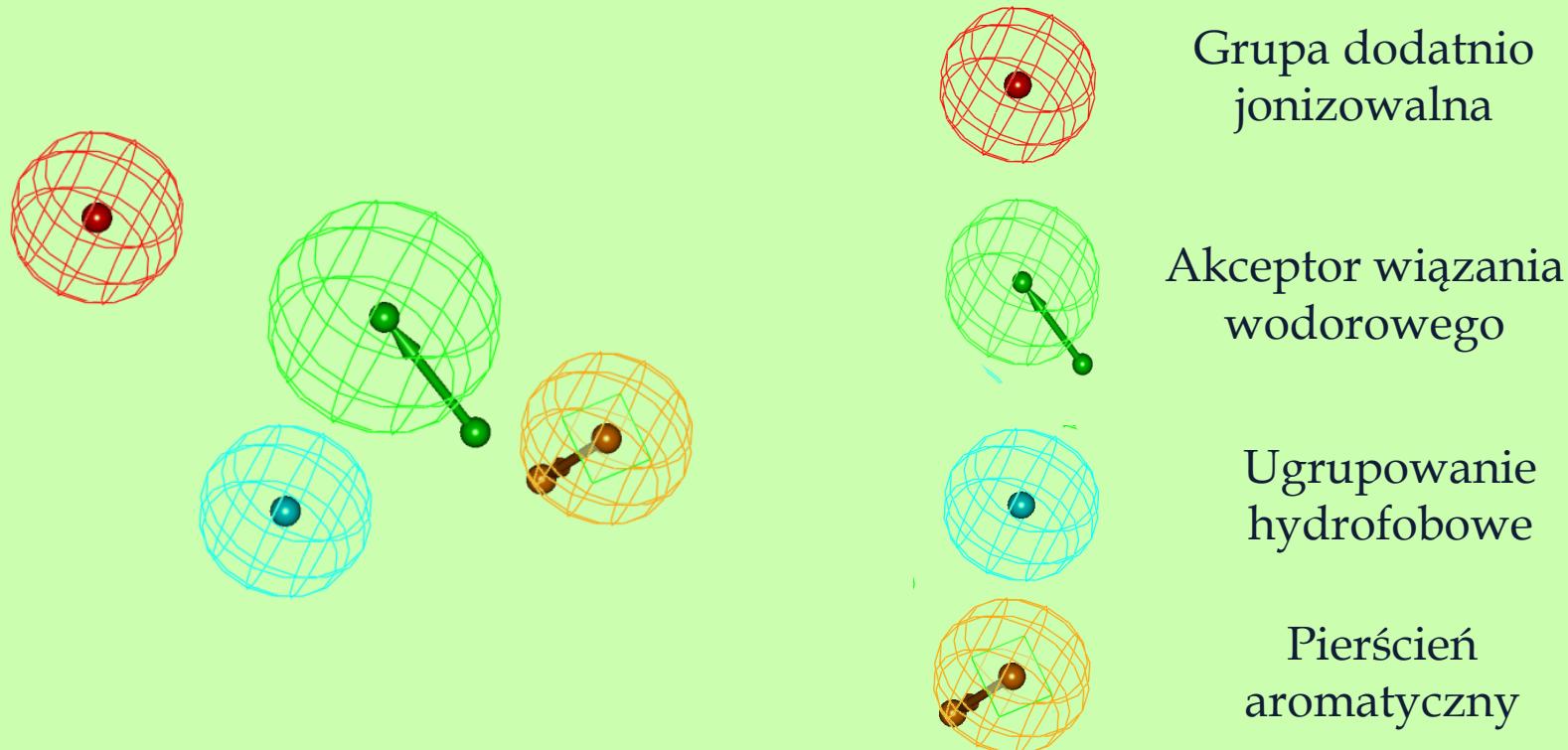
Atom Pairs

Atomtype—Atomtype—Atomtype—Atomtype

Topological Torsions

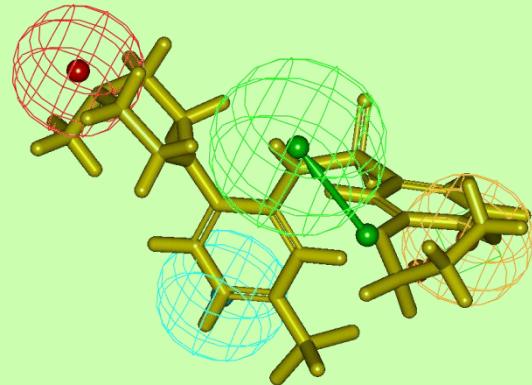
Fingerprinty farmakoforowe

- Farmakofor – zestaw cech opisujących cząsteczkę wraz z zależnościami geometrycznymi pomiędzy nimi



Fingerprinty farmakoforowe

Mapowanie cząsteczki na model farmakoforowy



- Fingerprinty farmakoforowe - najczęściej atom-based - cechy przypisywane poszczególnym atomom; obliczana odległość topologiczna pomiędzy kodowanymi cechami

Fingerprinty - modyfikacje

- Tworzenie reprezentacji hybrydowych



FP1



FP2



FP3



Dwuwymiarowe fingerprints podstrukturalne

- **Cel:**

Dostarczanie informacji nie tylko o obecności określonych podstruktur, lecz również o występowaniu połączeń pomiędzy nimi

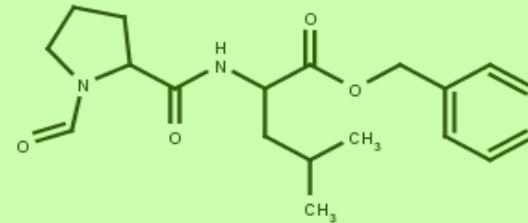
- **Problem:**

Stosunkowo duża ilość podstruktur w definicji fingerprintu prowadzi do otrzymywania macierzy o dużych wymiarach

- **Rozwiązanie:**

Selekcja podstruktur

					
	0	0	0	0	1
	0	0	0	1	0
	0	0	0	0	0
	0	1	0	0	0
	1	0	0	0	0



Po co to wszystko?

- Przeszukiwanie baz związków w oparciu o różnorodne współczynniki podobieństwa
- Poszukiwanie aktywnych związków przy wykorzystaniu metod uczenia maszynowego
- Klasteryzacja

⁹⁰Th
THORIUM
232.0381

⁷N
NITROGEN
14.00674

¹⁹K
POTASSIUM
39.0983

³⁹Y
YTTRIUM
88.90585

⁸O
OXYGEN
15.9994

⁹²U
URANIUM
238.0289