

Deskryptory molekularne

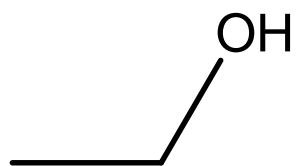
# Molekuła - definicje

- *Konstytucja* – sposób łączenia atomów
- *Stereochemia* – sposób rozmieszczenia atomów w przestrzeni (*konfiguracja*)
- *Izomery* – związki o tym samym wzorze cząsteczkowym różniące się konstytucją lub stereochemią

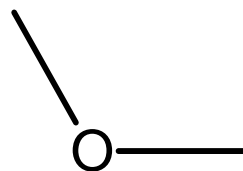
*IUPAC Recommendations 1994 [http://goldbook.iupac.org]*

---

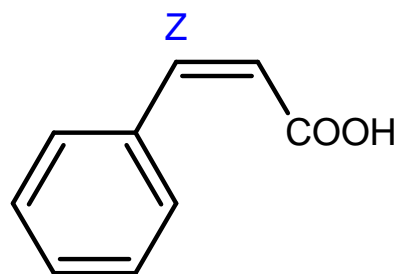
# Molekuła - izomery



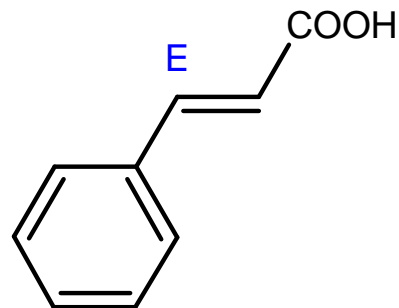
etanol



eter dietylowy



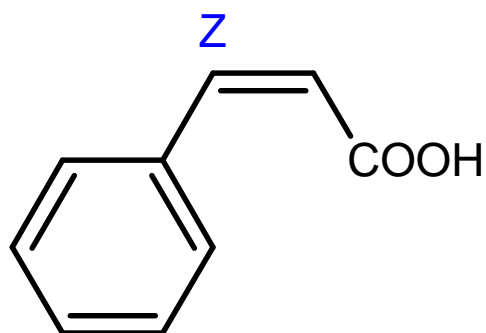
kwas (*Z*)-3-fenyloprop-2-enowy



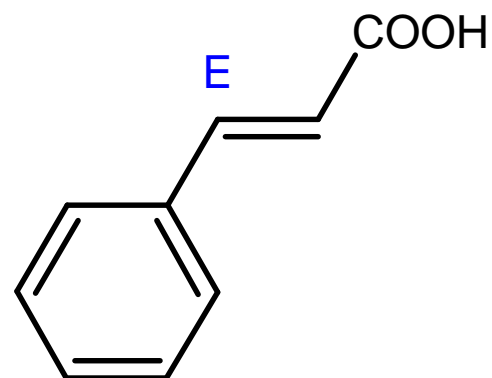
kwas (*E*)-3-fenyloprop-2-enowy

# Opis molekuł - topologia

statyczny opis opiera się na topologii (matematyka)

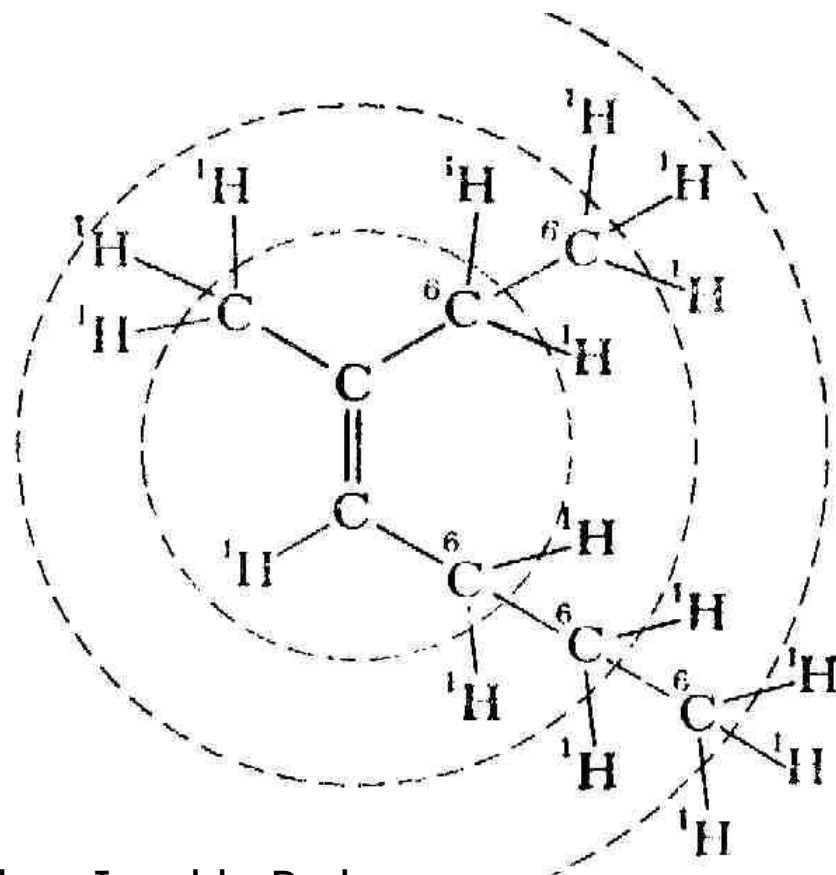


kwask (*Z*)-3-fenyloprop-2-enowy



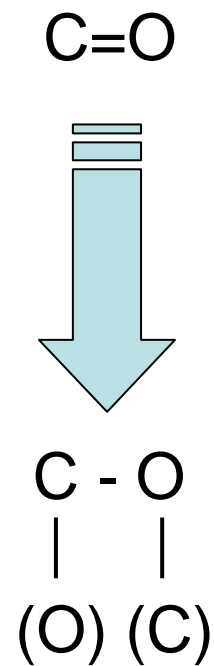
kwask (*E*)-3-fenyloprop-2-enowy

# Izomery E, Z - algorytm CIP

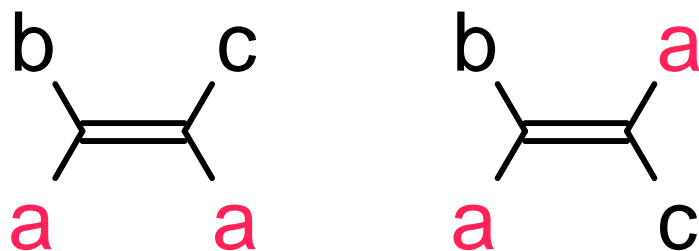


CIP = Cahn-Ingolda-Preloga

wiązania wielokrotne

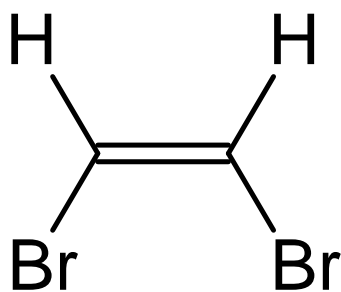


# Izomery E, Z - algorytm CIP

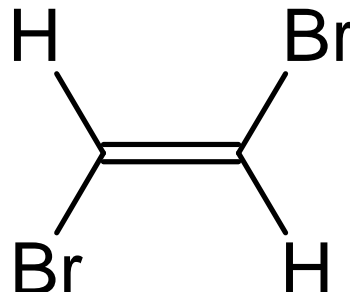


Z

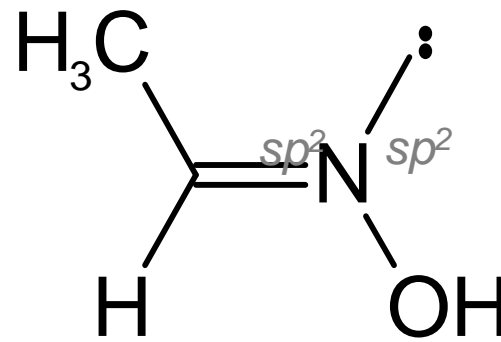
E



Z (cis)

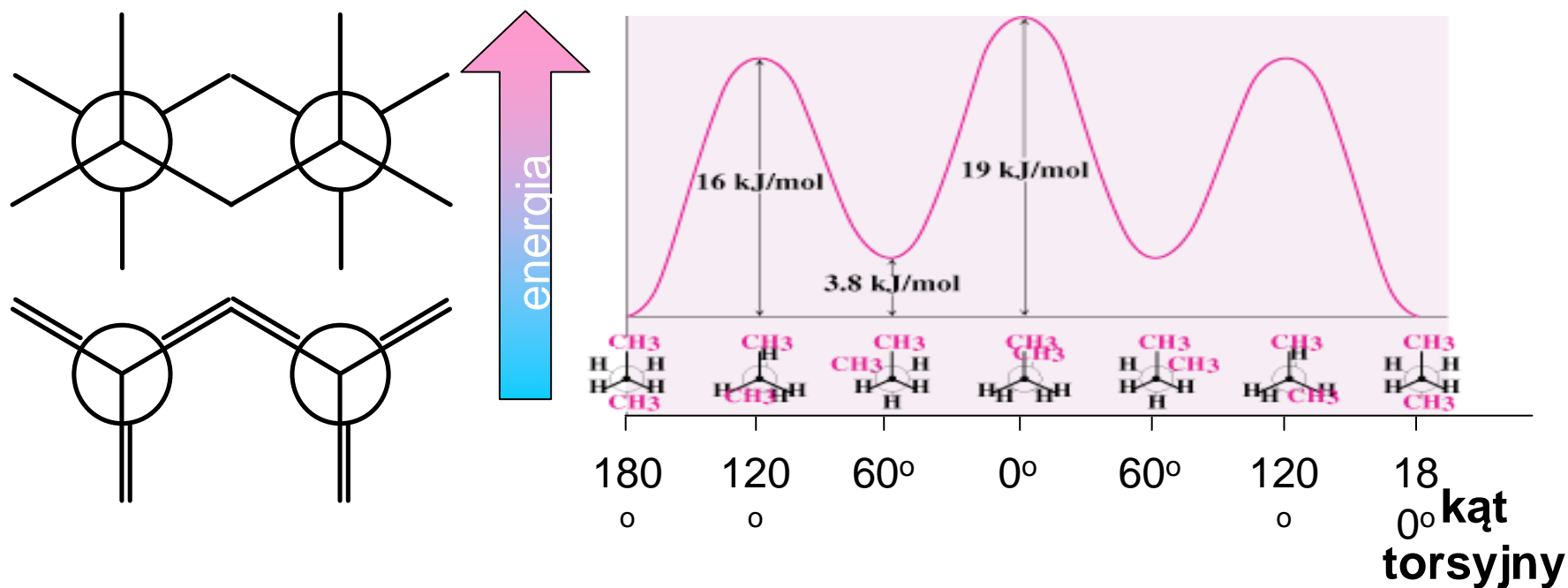


E (trans)



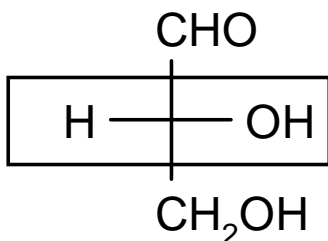
E (syn)

# Substancja chemiczna – komplikacja problemu molekuły

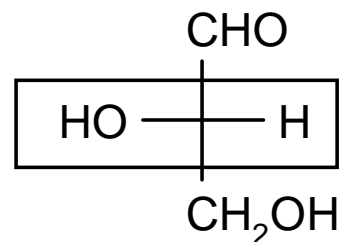


konformery - populacja kontrolowana przez energię molekuł

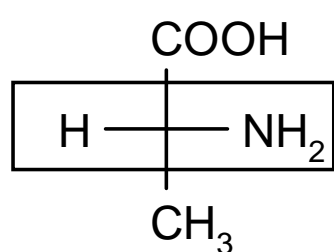
# Właściwość *P* cecha molekuły czy substancji?



aldehyd D-(+)-glicerynowy

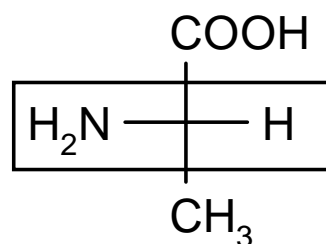


aldehyd L-(-)-glicerynowy



D-(-)-alanina

**R**

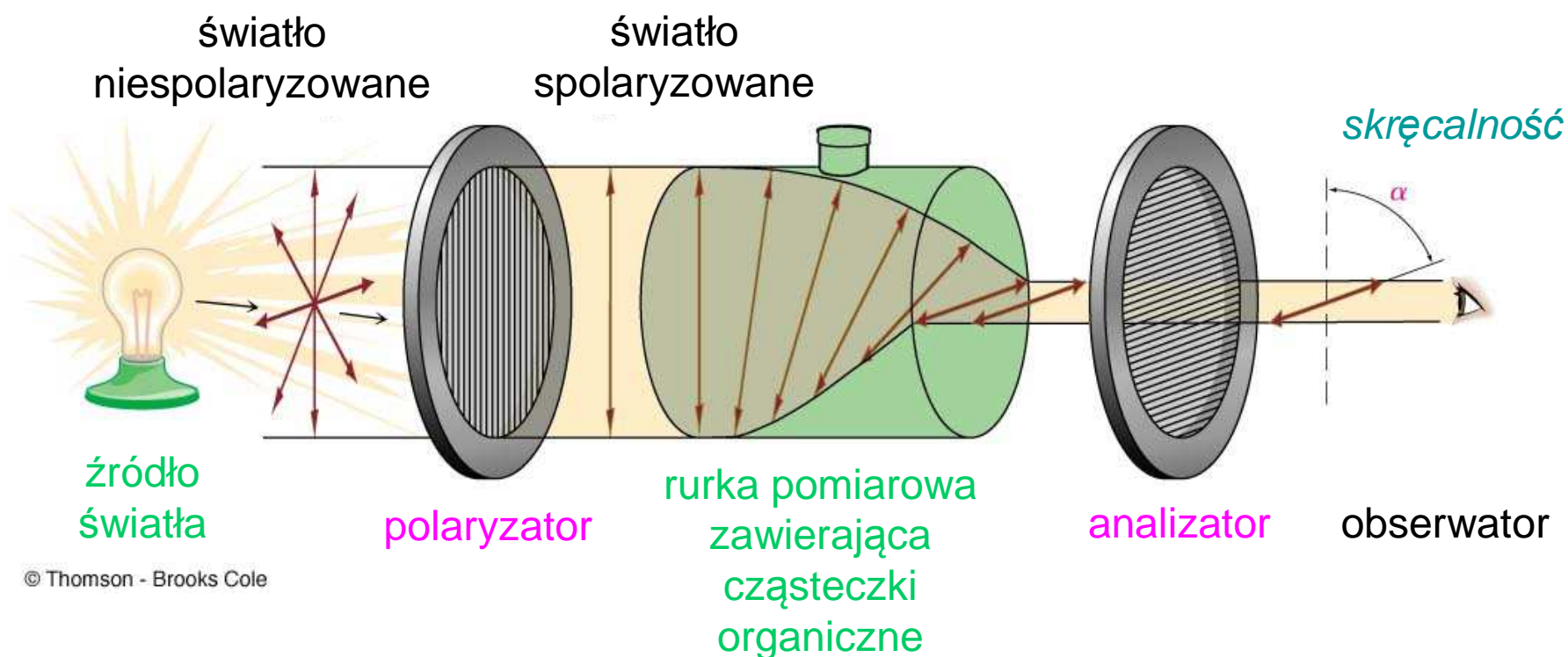


L-(+)-alanina

**S**



# Właściwość $P$ molekuły czy substancji?



McMurry Organic Chemistry 6th edition Chapter 9 (c) 2003

# Właściwości $P$ są w większości cechami substancji

Z obserwacji wynika, że:

- Nie ma jasnego związku między konfiguracją enancjomerów (R, S) a kierunkiem i wartością skręcania płaszczyzny światła spolaryzowanego (+, -) !!!
- A nawet w szczególności w różnych rozpuszczalnikach jeden enancjomer może skręcać płaszczyznę światła spolaryzowanego w dwóch różnych kierunkach !!!

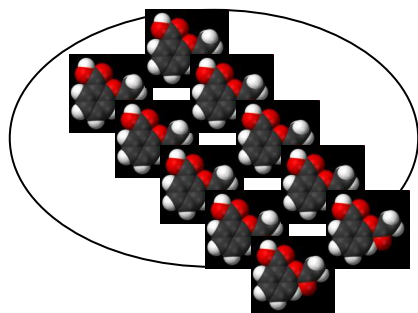
# Molekuła a substancja (związek chemiczny)

- Substancja chemiczna – zbiór molekuł *in vitro*
- **Cząsteczka (molekuła)** – obojętne elektrycznie indywiduum molekularne obejmujące więcej niż jeden atom ( $n > 1$ ). Formalnie cząsteczce musi odpowiadać minimum na powierzchni energii potencjalnej z co najmniej jednym stanem wibracyjnym
- **Indywiduum molekularne** (ang. *molecular entity*) – dowolny atom, molekuła, jon, para jonowa, rodnik, kompleks, konformer itp. wyróżniający się konstytucyjnie lub izotopowo, które posiada charakter odrębnej jednostki

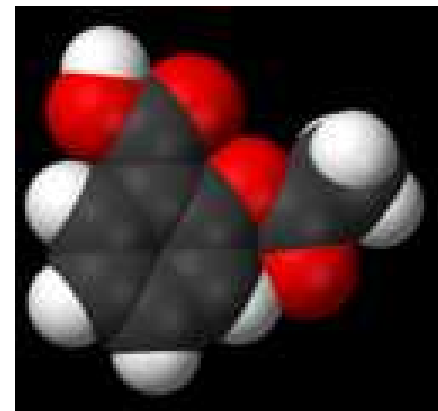
*IUPAC Recommendations 1994 [http://goldbook.iupac.org]*

---

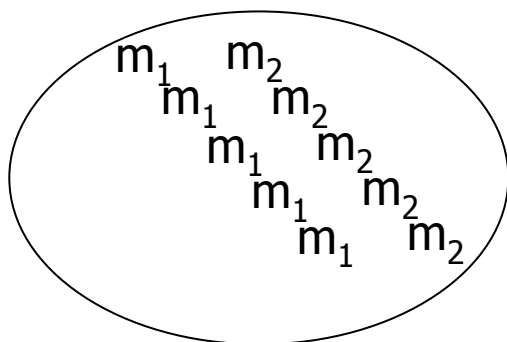
# Cząsteczka vs. związek chemiczny



chemical compound



a single molecule



a mixture of chemical compounds

*graphics: wikipedia*

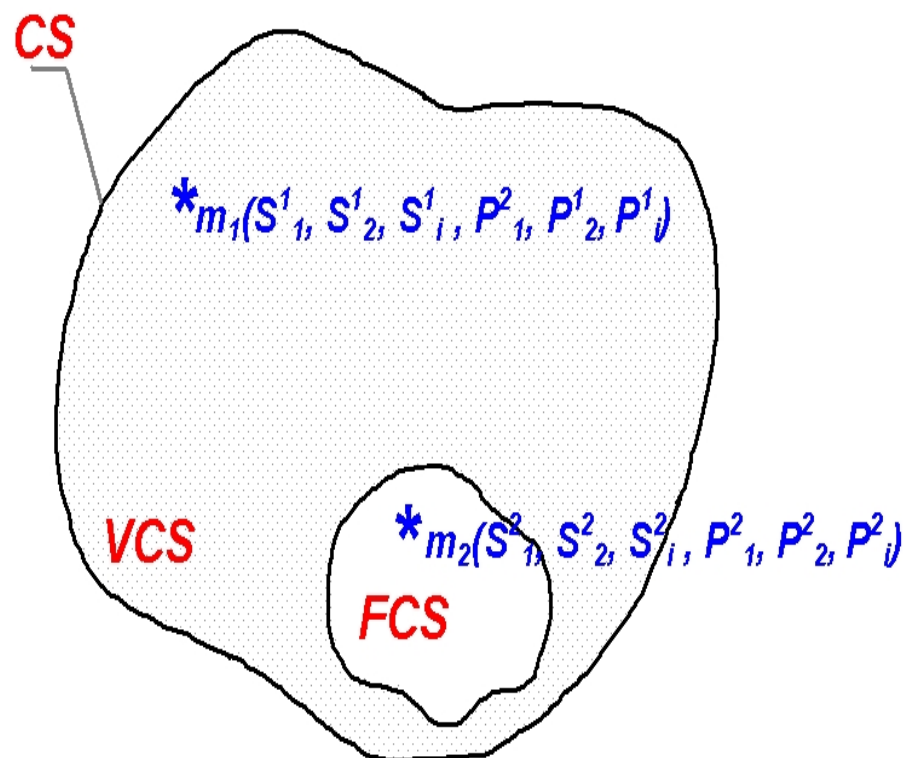
# Substancja chemiczna – komplikacja problemu molekuly

- **Substancja chemiczna** (*ang. chemical substance*) – materia o stałym składzie najlepiej charakteryzowanym przez jednostkowe molekuly, wzory, atomy, z których jest zbudowana. Substancji chemicznej przypisać można określone własności fizyczne takie jak gęstość, temperatura topnienia.
- **Odmiana chemiczna** (*ang. chemical species*) – Zbiór chemicznie identycznych indywiduów molekularnych, który może przyjmować wartości z tego samego zbioru poziomów energetycznych w skali czasowej eksperymentu. Termin ten stosuje się zarówno do chemicznie identycznych jednostkowych układów atomów jak i cząsteczek. Na przykład dwa izomery konformacyjne *konformery* mogą przekształcać się wzajemnie w siebie wystarczająco wolno, by można było ten efekt zaobserwować przez zróżnicowanie ich widm NMR.

*IUPAC Recommendations 1994 [http://goldbook.iupac.org]*

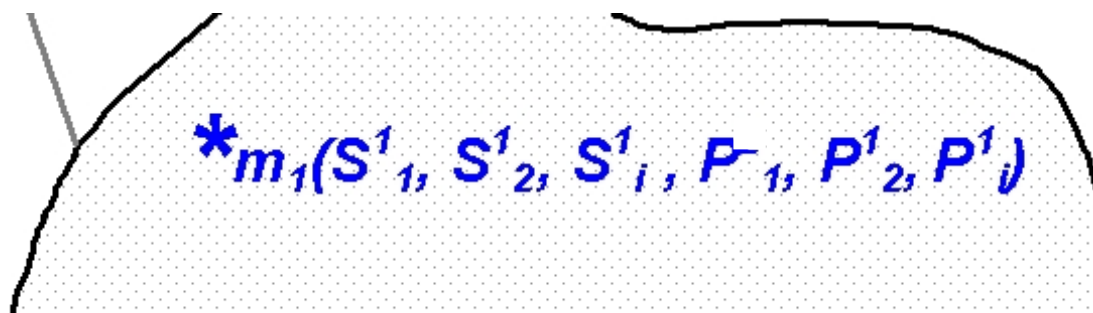
---

# Przestrzeń chemiczna CS



- chemical space CS  
virtual chemical space  
factual chemical space  
*Operacje in vitro lub in silico*  
= odwzorowania molekuł w CS

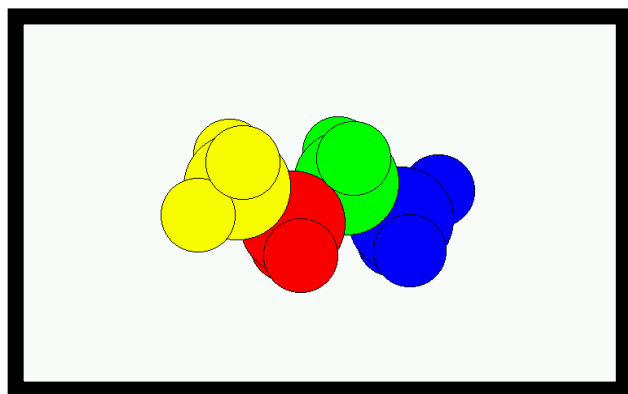
# Deskryptory vs właściwości



- deskryptory (S) – mogą być *obliczane* w operacjach wirtualnych [głównie dla cząsteczek]
- Właściwości (P) – można jedynie zmierzyć w eksperymencie [głównie związki chemiczne]

# Deskryptory molekularne $S_m$

*.... wynik operacji logicznej lub matematycznej, która przekształca informację chemiczną kodowaną w symbolicznej reprezentacji cząsteczki w jej postać numeryczną*



$$S = [s_1, s_2, s_3, \dots, s_i]$$

*Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors, Wiley-VCH: Weinheim, 2000.*



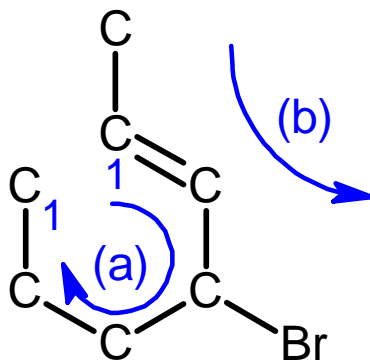
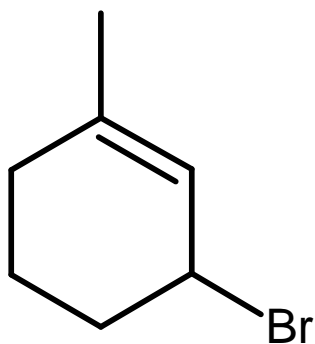
# Deskryptory molekularne $S_m$

- Kodujące deskryptory konstytucyjne
- Deskryptory daktyloskopowe cząsteczki
- Deskryptory obliczane na podstawie atomowej reprezentacji cząsteczki
- Deskryptory obliczane na podstawie fragmentów molekularnych
- Deskryptory topologiczne
- Proste deskryptory geometryczne
- Złożone deskryptory geometryczne
- Deskryptory pola oddziaływań cząsteczkowych
- Deskryptory profilu konformacyjnego
- Deskryptory wirtualnego miejsca receptorowego
- Deskryptory receptorowe
- Deskryptory złożonych systemów cząsteczkowych ligand-receptor
- Deskryptory skorelowane z właściwościami
- Korelaty właściwości globalnych
- Korelaty fragmentów molekularnych – stałe Hammetta
- Korelaty fragmentów molekularnych – stałe Hanscha

# Kod liniowy SMILES



*simplified molecular  
input line entry system*



(a) CC1=CC(Br)CCC1

(b) CC1=CC(CCC1)Br

Kod SMILES jest systemem notacji liniowej, gdzie do kodowania cząsteczek i reakcji chemicznych stosowane są znaki alfanumeryczne - *kody ASCII*.

# Kod liniowy SMILES



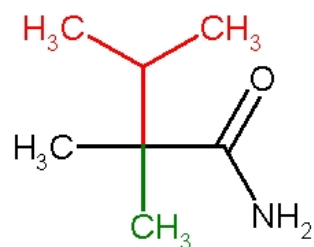
$\text{CH}_4$	Metan	C
$\text{NH}_3$	Amoniak	N
$\text{H}_2\text{O}$	Woda	O
Ag	Srebro	[Ag]
$\text{H}^+$	Proton	[H+]
$\text{Fe}^{3+}$	Kation żelaza	[Fe+3]=[Fe+++]
$\text{NH}_4^+$	Kation amoniowy	[NH4+]

# Kod liniowy SMILES

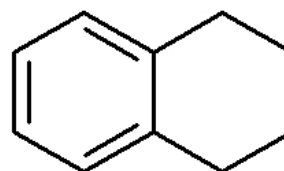


$\text{CH}_3\text{COOH}$	Kwas octowy	<chem>CC(=O)O</chem>
$\text{C}_6\text{H}_{10}$	Cykloheksen	<chem>C1=CCCCC1</chem>
$\text{C}_6\text{H}_5\text{NO}_2$	Nitrobenzen	<chem>c1ccccc1[N+](=O)[O-]</chem>
$\text{C}_6\text{H}_5\text{O}^-\text{Na}^+$	Fenolan sodu	<chem>[Na+].[O-]c1ccccc1</chem>
$\text{C}_{10}\text{H}_8$	Naftalen	<chem>c1cc2ccccc2cc1</chem>

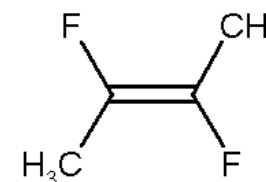
# Kod liniowy SMILES



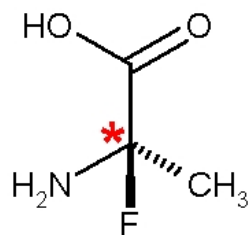
C(C(C)C)(C)C(=O)N



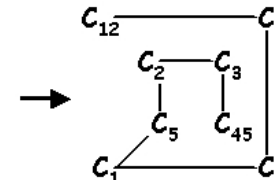
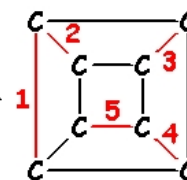
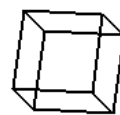
c1cc2CCCCc2cc1



C/C(F)=C(F)/C

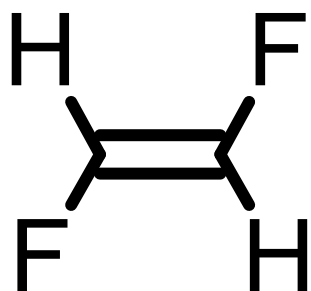


N[C@](F)(C(=O)O)C



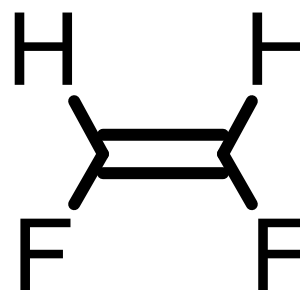
C12C3C4C1C5C2C3C45

# Kod liniowy SMILES



F/C=C/F

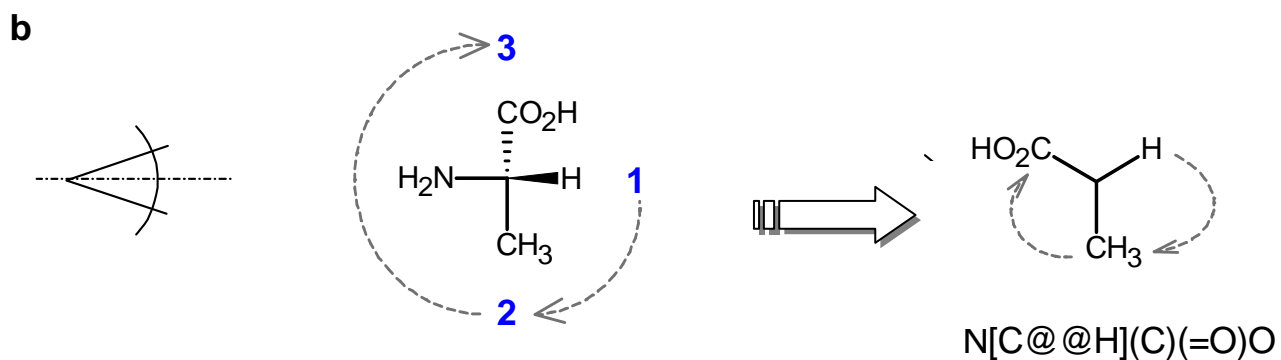
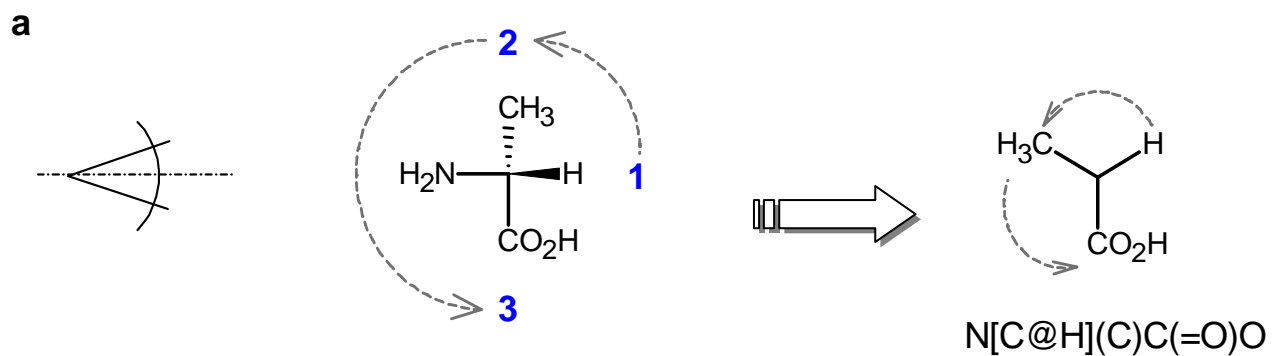
F\C=C\F



F/C=C\F

F\C=C/F

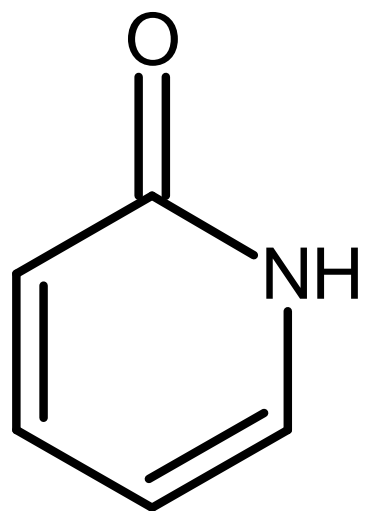
# Kod liniowy SMILES



# Kod liniowy SMILES

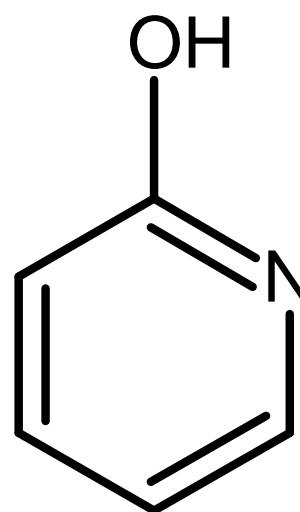


**a**



O=c1[nH]cccc1

**b**



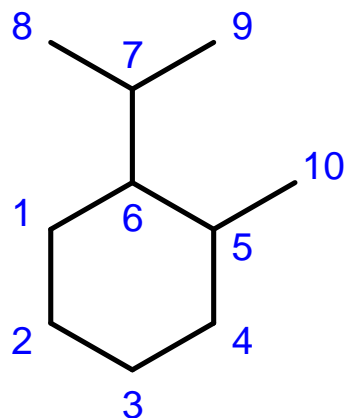
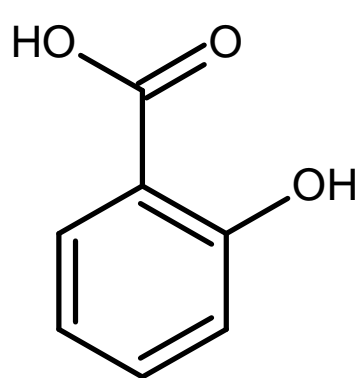
Oc1ncccc1



# Macierzowy zapis grafu

- Macierz sąsiedztwa
- Macierz odległości
- Macierz częstości
- Macierz wiązań
- Macierz elektronów wiążących

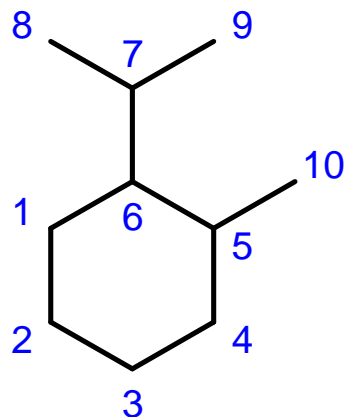
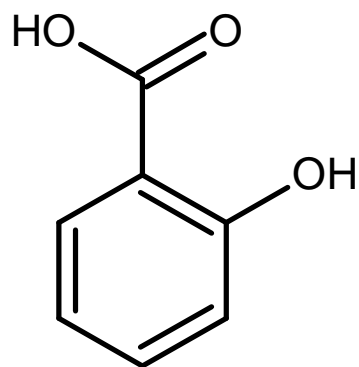
# Macierzowe systemy kodowania molekuł - macierz sąsiedztwa



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

$$[A(G)]_{ij} = \begin{cases} 1 & \text{dla } i \neq j \text{ oraz } e_{ij} \in E(G) \\ 0 & \text{dla } i = j \text{ lub } e_{ij} \notin E(G) \end{cases}$$

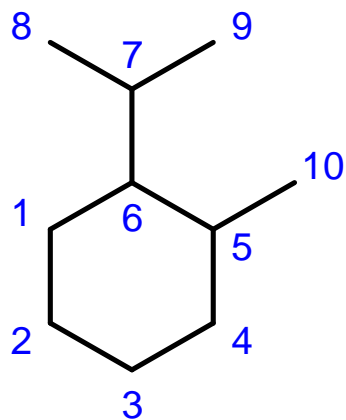
# Trójkątna macierz sąsiedztwa



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

$$[A(G)]_{ij} = \begin{cases} 1 & \text{dla } i \neq j \text{ oraz } e_{ij} \in E(G) \\ 0 & \text{dla } i = j \text{ lub } e_{ij} \notin E(G) \end{cases}$$

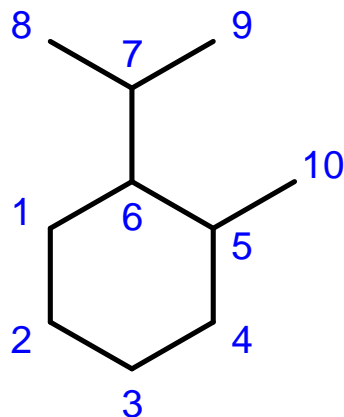
# Macierz odległości topologicznej



$D_D^2$	1	2	3	4	5	6	7	8	9	10
1	0	1	2	3	2	1	2	3	3	3
2		0	1	2	3	2	3	4	4	4
3			0	1	2	3	4	5	5	3
4				0	1	2	3	4	4	2
5					0	1	2	3	3	1
6						0	1	2	2	2
7							0	1	1	3
8								0	2	4
9									0	4
										0

$$[D(G)]_{ij} = \begin{cases} d_{ij} & \text{dla } i \neq j \\ 0 & \text{dla } i = j \end{cases}$$

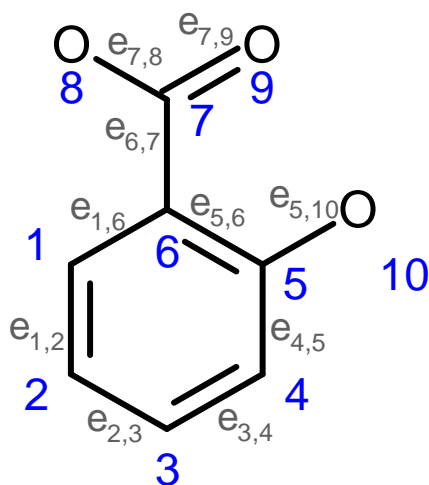
# Macierz odległości geometrycznej



	1	2	3	4	5	6	7	8	9	10
1	0	1,38	2,40	2,78	2,42	1,40	2,49	2,85	3,59	3,66
2		0	1,39	2,40	2,78	2,40	3,75	4,23	4,73	4,13
3			0	1,38	2,40	2,77	4,25	5,06	4,97	3,64
4				0	1,39	2,41	3,76	4,86	4,18	2,38
5					0	1,40	2,49	3,71	2,79	1,36
6						0	1,47	2,44	2,33	2,39
7							0	1,35	1,21	2,83
8								0	2,22	4,17
9									0	2,50
10										0

$$[D(G)]_{ij} = \begin{cases} d_{ij} & \text{dla } i \neq j \\ 0 & \text{dla } i = j \end{cases}$$

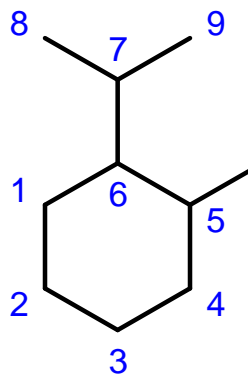
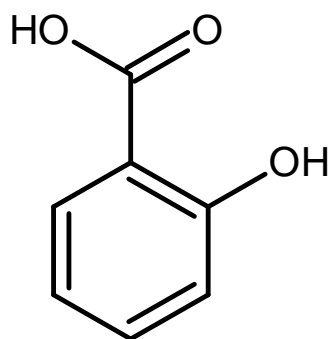
# Macierz incydencji



	1	2	3	4	5	6	7	8	9	10
$e_{1,2}$	1	1								
$e_{2,3}$		1	1							
$e_{3,4}$			1	1						
$e_{4,5}$				1	1					
$e_{5,6}$					1	1				
$e_{5,10}$					1					1
$e_{1,6}$	1					1				
$e_{6,7}$						1	1			
$e_{7,8}$							1	1		
$e_{7,9}$							1		1	

W macierzy incydencji o wymiarach  $n \times m$  kolumny reprezentują wierzchołki grafu molekularnego (atomy), wiersze zaś odpowiadają jego krawędziom (wiązaniami chemicznym). Wartość 1 występująca na przecięciu  $i$ -tego wiersza i  $j$ -tej kolumny wskazuje, iż wierzchołek  $j$  jest końcem krawędzi  $i$ .

# Macierz wiązań

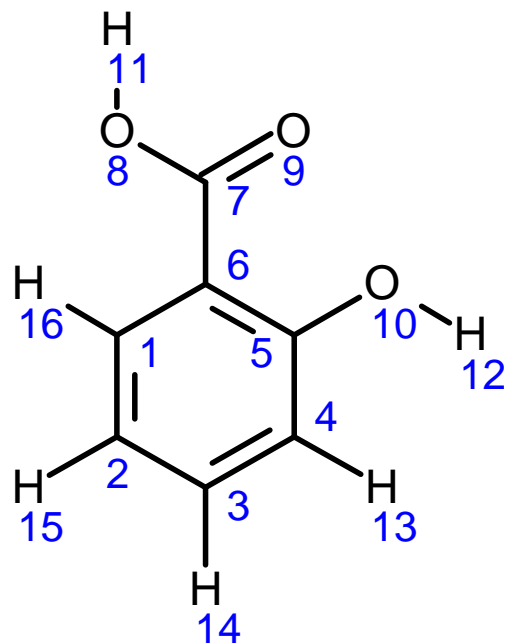


	1	2	3	4	5	6	7	8	9	10
1	0	2				1				
2	2	0	1							
3		1	0	2						
4			2	0	1					
5				1	0	2				1
6	1				2	0	1			
7						1	0	1	2	
8							1	0	1	
9							2		0	
10					1					0

$$[B(G)]_{ij} = \begin{cases} b_{ij} & \text{dla } i \neq j \\ 0 & \text{dla } i = j \end{cases}$$

b – numeryczna reprezentacja rodzaju wiązania

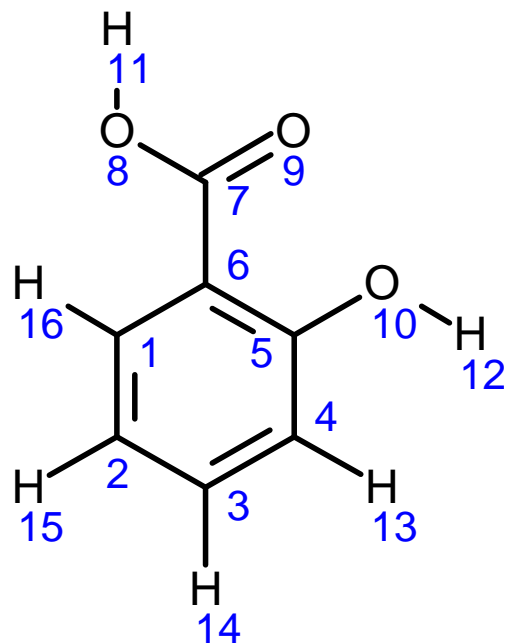
# Macierz elektronów wiążących



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	suma wierszy	atom
1		2				1										1	4	C
2	2		1												1		4	C
3		1		2										1			4	C
4			2		1								1				4	C
5				1		2				1							4	C
6	1				2		1										4	C
7						1		1	2								4	C
8							1	4			1						6	O
9							2		4								6	O
10					1					4		1					6	O
11								1									1	H
12										1							1	H
13				1													1	H
14			1														1	H
15		1															1	H
16	1																1	H
suma kolumn	4	4	4	4	4	4	4	6	6	6	1	1	1	1	1	1	<b>104</b>	
suma	8	8	8	8	8	8	8	8	8	8	2	2	2	2	2	2		



# Tablica połączeń



Lista atomów	
1	C
2	C
3	C
4	C
5	C
6	C
7	C
8	O
9	O
10	O
11	H
12	H
13	H
14	H
15	H
16	H

Lista wiązań		
A 1	A 2	W
1	2	2
1	6	1
1	16	1
2	3	1
2	15	1
3	4	2
3	14	1
4	5	1
4	13	1
5	6	2
5	10	1
6	7	1
7	8	1
7	9	2
8	11	1
10	12	1

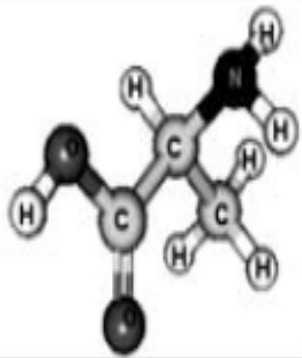
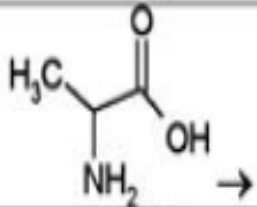
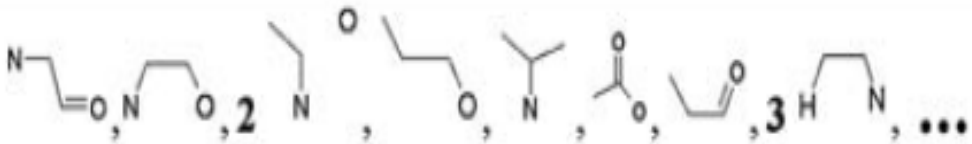
# Simplex 1D

- 1D

$$\text{liczba} = \frac{a!}{(a-i)! * i!}$$

kombinacja bez powtórzeń i-elementowa  
ze zbioru wszystkich atomów cząsteczki  
(a)

# Simplex 2D

Level	Structure	Simplex generation
		
1D	$C_3H_7O_2N \rightarrow$	6 CCNO, 42 CNOH, 63 CNHH, 21 CCNH, 42 NOHH, 7 CCCH, 35 NHHH, ...
2D		

VICTOR E. KUZ'MIN et al., VIRTUAL SCREENING AND MOLECULAR DESIGN BASED ON HIERARCHICAL QSAR TECHNOLOGY, in: Recent Advances in QSAR Studies, Springer 2010