

Ćwiczenie 2. Ilościowe zależności pomiędzy strukturą chemiczną a właściwościami chemicznymi (QSPR) dla nowych związków chemicznych

I. Cel ćwiczenia:

Celem ćwiczenia jest nabycie umiejętności wykonania ilościowej zależności pomiędzy strukturą chemiczną a właściwościami chemicznymi (QSPR) dla nowych związków chemicznych.

II. Zagadnienia do samodzielnego opracowania:

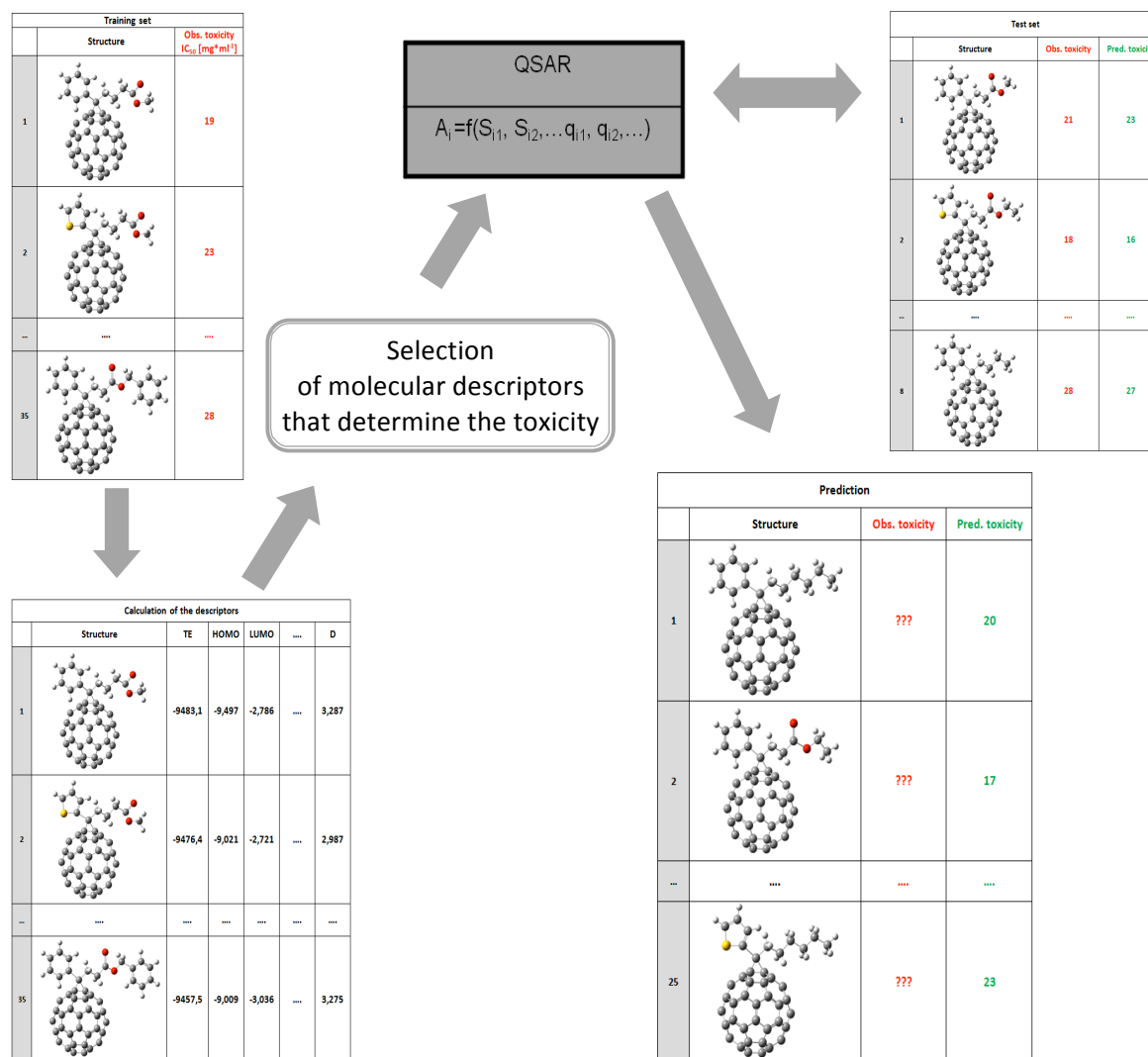
Idea ilościowego modelowania struktura-aktywność (QSAR) oraz struktura-właściwości (QSPR). Podstawy modelowania molekularnego: specyfikacja geometrii we współrzędnych wewnętrznych, znajdowanie przybliżenia równania Schrödingera w oparciu o półempiryczne metody chemii kwantowej, optymalizacja geometrii cząsteczki. Deskryptory molekularne w QSAR/QSPR, kalibracja modelu QSAR/QSPR w oparciu o metodę regresji wielorakiej. Walidacja modelu w oparciu o reguły OECD. Przykładowe modele QSPR służące do przewidywania współczynników podziału substancji chemicznych. Współczynniki podziału n-oktanol/woda ($\log K_{OW}$), n-oktanol/powietrze ($\log K_{OA}$) i powietrze/woda ($\log K_{AW}$) i metody ich eksperymentalnego oraz komputerowego wyznaczania. Znaczenie współczynników podziału w ocenie ryzyka dla nowych substancji chemicznych.

III. Wstęp teoretyczny

Idea przewidywania danych w oparciu o ilościowe zależności: struktura chemiczna – właściwości (QSPR)

Istnieje pilna konieczność wyznaczenia wartości kluczowych parametrów fizykochemicznych, dla wielu tysięcy chemikaliów określanych jako PBTs (z ang. *Persistent, Bioaccumulative, Toxic*) determinujących los tych zanieczyszczeń w środowisku oraz zagrożenie jakie mogą one stwarzać dla organizmów żywych. Niestety eksperymentalnie wyznaczone wartości wielu parametrów fizykochemicznych istnieją tylko dla nielicznych (zwłaszcza dla chloroorganicznych) związków. Wynika to z faktu iż doświadczenia eksperymentalne są czasochłonne, niezwykle kosztowne a ponadto niejednokrotnie trudności analityczne nie pozwalają na wyizolowanie „czystych” analogów.

Alternatywą dla wyznaczania właściwości fizykochemicznych i aktywności biologicznej drogą eksperymentalną, są matematyczne zależności między strukturą chemiczną reprezentowaną przez zbiór deskryptorów a właściwościami fizykochemicznymi (QSPR) lub aktywnością biologiczną (QSAR). U podstaw metody leży założenie, iż różnice pomiędzy właściwościami fizykochemicznymi poszczególnych związków wynikają z różnic w ich budowie, a zależności te można ilościowo opisać za pomocą odpowiednich równań matematycznych. Posiadając dane eksperymentalne dla części związków z danej grupy oraz znając wartości tzw. deskryptorów strukturalnych wszystkich związków z grupy, można przewidzieć właściwości chemikaliów, dla których danych eksperymentalnych brakuje. Idea metodologii QSAR/QSPR przedstawiona została na Rysunku 1.



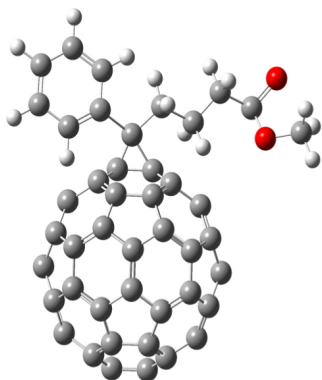
Metodologia QSAR obejmuje pięć kolejnych kroków:

- Zebranie danych eksperymentalnych i podział całego zbioru na:
 - Zbiór uczący (treningowy)
 - Zbiór testowy (walidacyjny)
- Obliczenie deskryptorów molekularnych
- Zbudowanie metody (kalibracja modelu)
- Testowanie metody (walidacja modelu)
 - Walidacja wewnętrzna (ang. internal validation)
 - Walidacja zewnętrzna (ang. external validation)
- Użycie najlepszego modelu do wyznaczenia brakujących wartości

Pierwszy etap badań stanowi zebranie dobrej jakości danych eksperymentalnych. Jest to etap kluczowy, mający zasadniczy wpływ na rezultaty modelowania QSAR/QSPR. Pod pojęciem dane wejściowe „dobrej jakości” rozumie się przede wszystkim dane uzyskane wyłącznie na drodze eksperymentalnej i zarejestrowane w takich samych warunkach laboratoryjnych. Następnie podział zebranych danych eksperymentalnych na:

- zbiór modelowy (70-75% związków), służący do kalibracji i wewnętrznej walidacji modelu
- testowy (30-25%), wykorzystywany do przeprowadzenia walidacji zewnętrznej (nie używany podczas kalibracji modelu).

Kolejnym etapem procedury QSAR/QSPR jest obliczenie tzw. deskryptorów strukturalnych (Rysunek 2), opisujących struktury molekularne wszystkich badanych związków chemicznych (zarówno tych, dla których są dostępne dane eksperymentalne, jak i tych, dla których danych takich brakuje).



- nAT – liczba atomów w cząsteczce
- MW – masa cząsteczkowa
- HOF – standardowe ciepło tworzenia
- EE – energia elektronowa
- Core – energia rdzenia
- TE – energia całkowita
- TE^+ – energia całkowita kationu
- HOFc – standardowe ciepło tworzenia w roztworze
- TEc – energia całkowita w roztworze
- VIP – wartościowy potencjał jonizacyjny

Fullerene_C60-1h_1 - WordPad

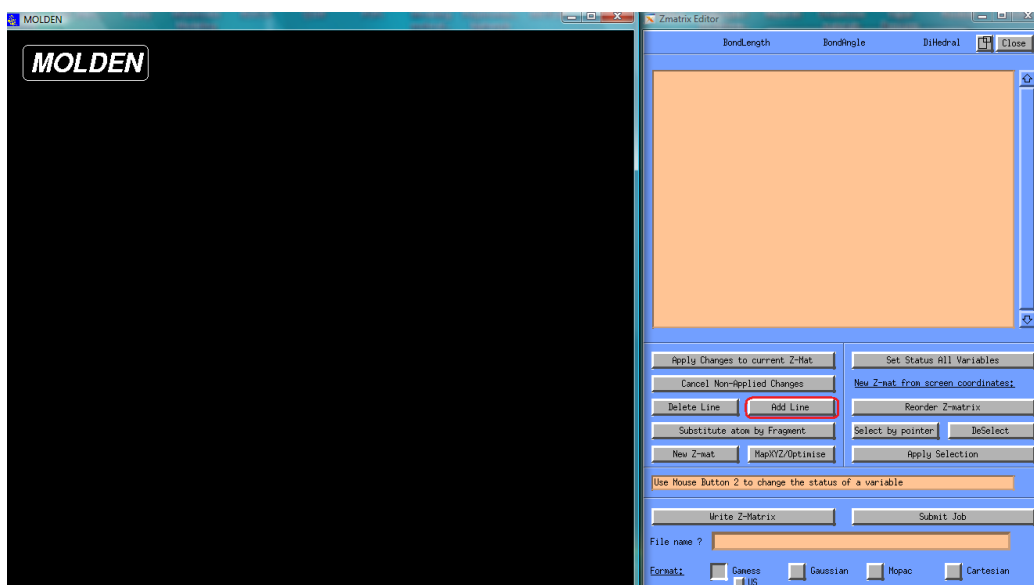
Plik Edycja Widok Wstaw Format Pomoc

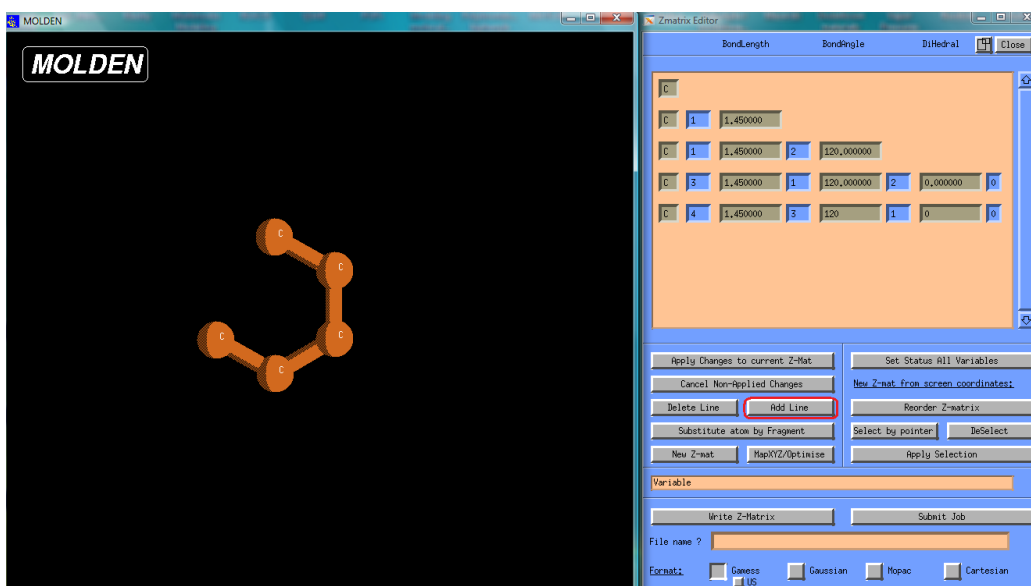
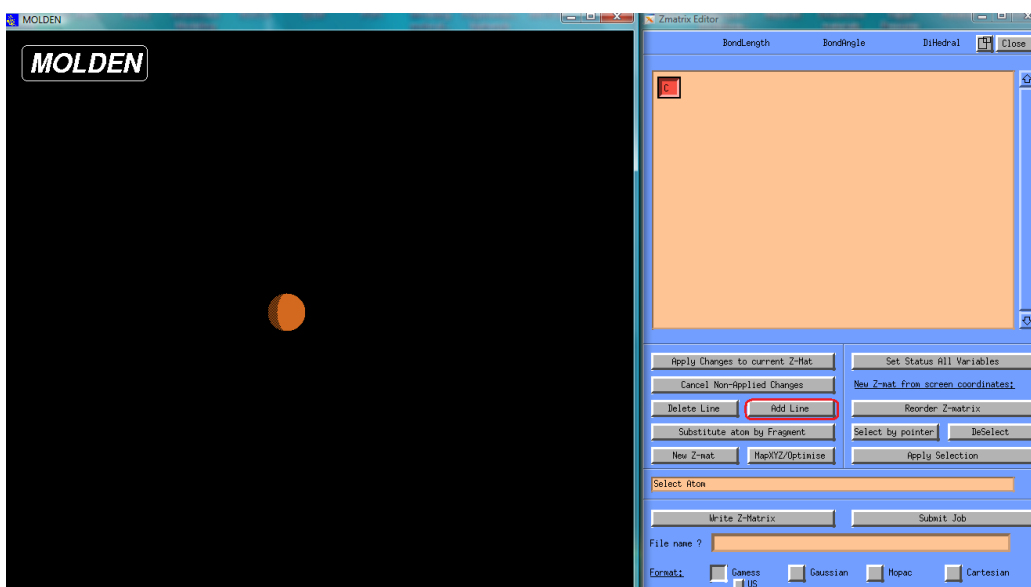
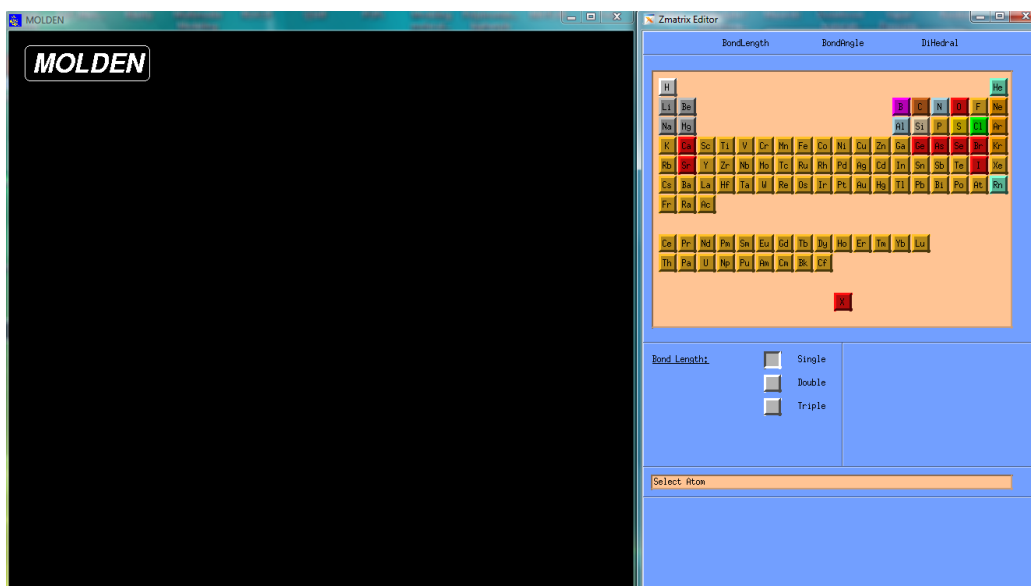
PM6 PRECISE

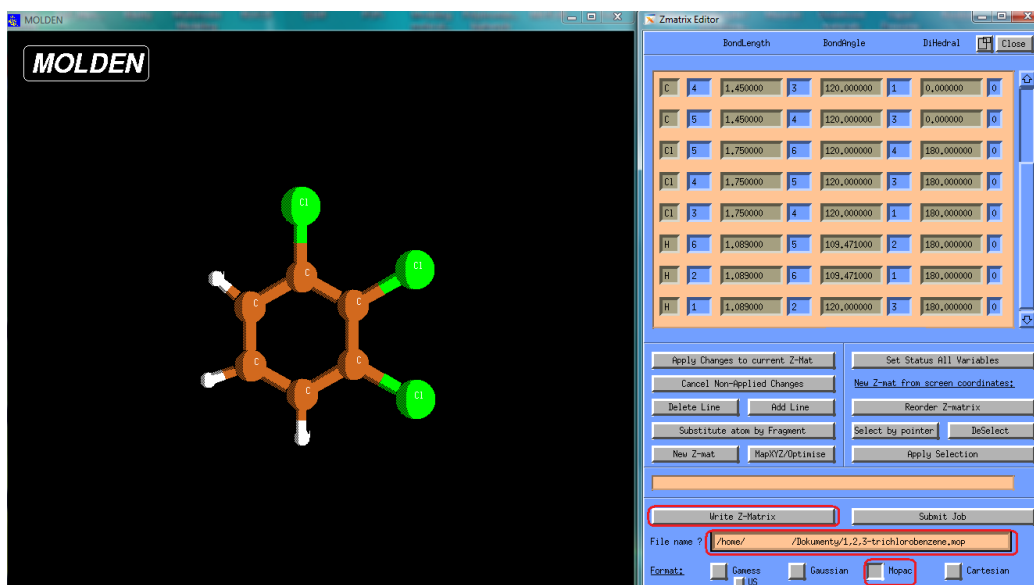
C	0.00000	1	0.00000	1	0.00000	1
C	1.45280	1	0.00000	1	0.00000	1
C	-0.69590	1	1.20410	1	0.00000	1
C	2.14800	1	1.20430	1	-0.00080	1
C	3.32300	1	1.36380	1	0.83760	1
C	1.90210	1	-1.09770	1	0.83890	1
C	3.02670	1	-0.94520	1	1.64100	1
C	3.75340	1	0.31290	1	1.64090	1
C	-2.29900	1	0.31330	1	1.64270	1
C	-1.57290	1	-0.94480	1	1.64280	1
C	-1.87100	1	1.36340	1	0.83970	1
C	-0.44870	1	-1.09790	1	0.83950	1
C	0.72730	1	-1.77700	1	1.35760	1
C	-1.57140	1	-1.46280	1	3.00000	1
C	-0.44680	1	-2.11240	1	3.49610	1
C	0.72800	1	-2.27260	1	2.65660	1
C	-2.29930	1	2.96930	1	2.65770	1
C	-2.74740	1	1.87100	1	3.49650	1
C	-1.87080	1	2.72040	1	1.35790	1
C	-2.74740	1	0.57250	1	3.00050	1
C	-2.29810	1	-0.52530	1	3.83930	1
C	-2.29700	1	2.13030	1	4.85440	1
C	0.03070	1	2.46220	1	-0.00010	1
C	1.45310	1	4.29730	1	1.64110	1
C	1.90300	1	4.55540	1	2.99780	1
C	3.32320	1	2.72090	1	1.35590	1
C	3.75430	1	2.96820	1	2.65460	1
C	3.02790	1	3.90670	1	3.49370	1
C	1.45650	1	2.44400	1	6.49370	1
C	0.03560	1	-0.01760	1	6.49500	1

Aby uzyskać Pomoc, naciśnij klawisz F1.

Do zbudowania niezbędnych struktur można użyć dostępne programy komputerowe np. Molden. Po zapisaniu struktury każdego związku w postaci tzw. macierzy współrzędnych wewnętrznych („z-macierzy”), kombinatorycznym wygenerowaniu struktur wszystkich możliwie istniejących kongenerów każdej badanej grupy związków chemicznych i optymalizacji geometrii każdej z nich, obliczane są odpowiednie deskryptory molekularne.







Niezbędne obliczenia mogą zostać przeprowadzone na różnych poziomach chemii teoretycznej. Z ostatnich badań wynika, iż kwantowo-mechaniczne deskryptory molekularne wyznaczone za pomocą nowych metod półempirycznych (PM6, RM1) pozwalają na uzyskanie modeli QSAR/QSPR o jakości zbliżonej do modeli zbudowanych w oparciu o deskryptory pochodzące nawet z metod Funkcjonału Gęstości Elektronowej DFT w znacznie krótszym czasie.

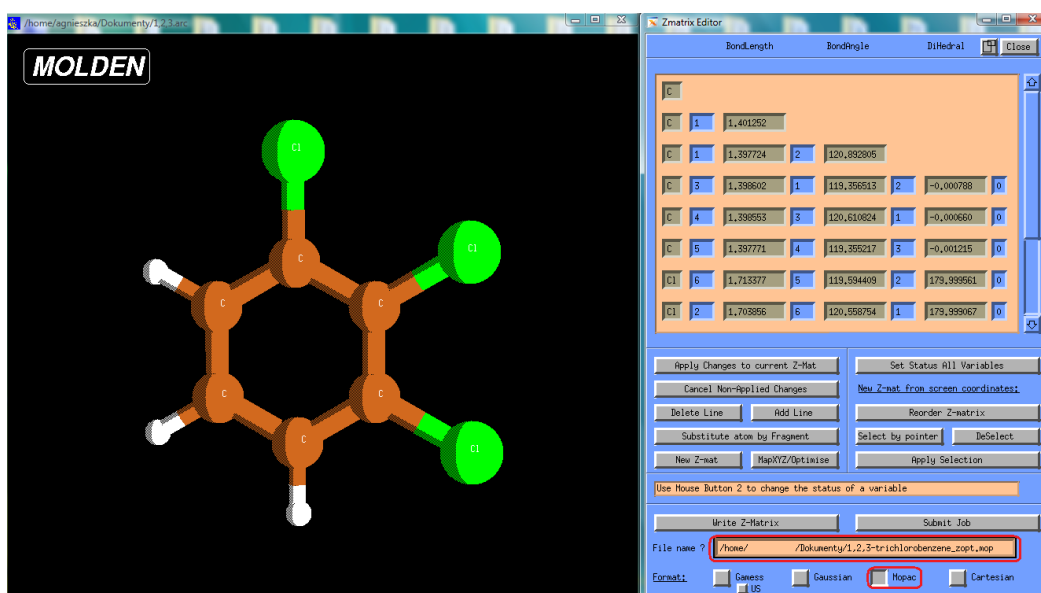
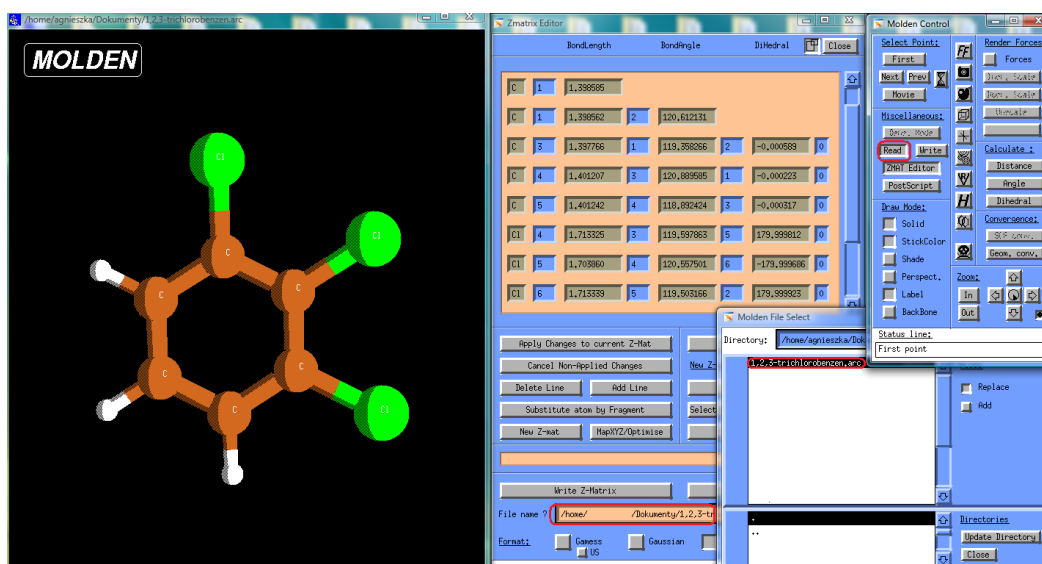
Utworzoną w Moldenie i zapisaną w formacie .mop cząsteczkę, otwieramy jako plik tekstowy a następnie przygotowujemy plik wejściowy (tak zwany "input") wpisując w pierwszej linii słowa kluczowe, w dwóch kolejnych komentarz (ewentualnie te dwie linie pozostawia się puste), a następnie w czwartej linii podawana jest geometria związku w postaci macierzy współrzędnych wewnętrznych.

```

PM6 PRECISE (optmalizujemy geometrię cząsteczki)

C 0.000000 0 0.000000 0 0.000000 0 0 0 0
C 1.450000 1 0.000000 0 0.000000 0 1 0 0
C 1.450000 1 120.000000 1 0.000000 0 1 2 0
C 1.450000 1 120.000000 1 0.000000 1 3 1 2
C 1.450000 1 120.000000 1 0.000000 1 4 3 1
C 1.450000 1 120.000000 1 0.000000 1 5 4 3
Cl 1.750000 1 120.000000 1 180.000000 1 4 3 5
Cl 1.750000 1 120.000000 1 180.000000 1 5 4 6
Cl 1.750000 1 120.000000 1 180.000000 1 6 5 2
H 1.089000 1 120.000000 1 180.000000 1 3 4 1
H 1.089000 1 120.000000 1 180.000000 1 1 3 2
H 1.089000 1 120.000000 1 180.000000 1 2 1 6
  
```

Po zoptymalizowaniu geometrii cząsteczki, otrzymujemy trzy pliki typu output: .arc, .mop oraz .out. Plik z rozszerzeniem .arc należy otworzyć przy użyciu programu Molden i zapisać zoptymalizowaną geometrię jako .mop. Tak zapisaną geometrię wykorzystuje się w obliczeniach kwantowo-mechanicznych wprowadzając odpowiednie słowa kluczowe (Tabela 1).



1,2,3tri_zopt.mop - WordPad

Plik Edycja Widok Wstaw Format Pomoc

Courier New 10 Europa Środkowa

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

PM6 1SCF STATIC

```

C      0.000000  0      0.000000  0      0.000000  0      0      0      0
C      1.401252  1      0.000000  0      0.000000  0      1      0      0
C      1.397724  1     120.892805  1      0.000000  0      1      2      0
C      1.398602  1     119.356513  1     -0.000788  1      3      1      2
C      1.398553  1     120.610824  1     -0.000660  1      4      3      1
C      1.397771  1     119.355217  1     -0.001215  1      5      4      3
Cl     1.713377  1     119.594409  1     179.999561  1      6      5      2
Cl     1.703856  1     120.558754  1     179.999067  1      2      6      1
Cl     1.713343  1     119.598785  1     179.998920  1      1      3      2
H      1.091179  1     120.565585  1    -179.999482  1      5      4      6
H      1.091183  1     119.702152  1    -179.999214  1      4      5      3
H      1.091191  1     120.559416  1     179.999895  1      3      4      1
  
```

Aby uzyskać Pomoc, naciśnij klawisz F1. NUM

W wyniku obliczeń, otrzymujemy pliki wyjściowe typu .out (tak zwane "outputy"). Pliki te otwieramy jako pliki tekstowe, a następnie odczytujemy z nich wartości poszczególnych deskryptorów.

```

1,2,3tri_zopt.out - WordPad
Plik  Edycja  Widok  Wstaw  Format  Pomoc

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

-----
PM6 1SCF STATIC

1SCF WAS SPECIFIED, SO BFGS WAS NOT USED
SCF FIELD WAS ACHIEVED

PM6      CALCULATION

MOPAC2009 (Version: 9.047W)
Thu Dec 10 17:08:25 2009
No. of days left = 72

FINAL HEAT OF FORMATION = 1.82744 KCAL = 7.64600 KJ
TOTAL ENERGY           = -1538.61168 EV
ELECTRONIC ENERGY      = -6549.19154 EV
CORE-CORE REPULSION     = 5010.57986 EV
COSMO AREA              = 171.80 SQUARE ANGSTROMS
COSMO VOLUME            = 173.23 CUBIC ANGSTROMS

IONIZATION POTENTIAL    = 9.735232 EV
HOMO LUMO ENERGIES (EV) = -9.735 -0.731
NO. OF FILLED LEVELS    = 24
MOLECULAR WEIGHT        = 181.449

MOLECULAR DIMENSIONS (Angstroms)

Atom      Atom      Distance
H   12   Cl   7   5.60054
H   10   Cl   8   4.85334
Cl   9   Cl   8   0.00029

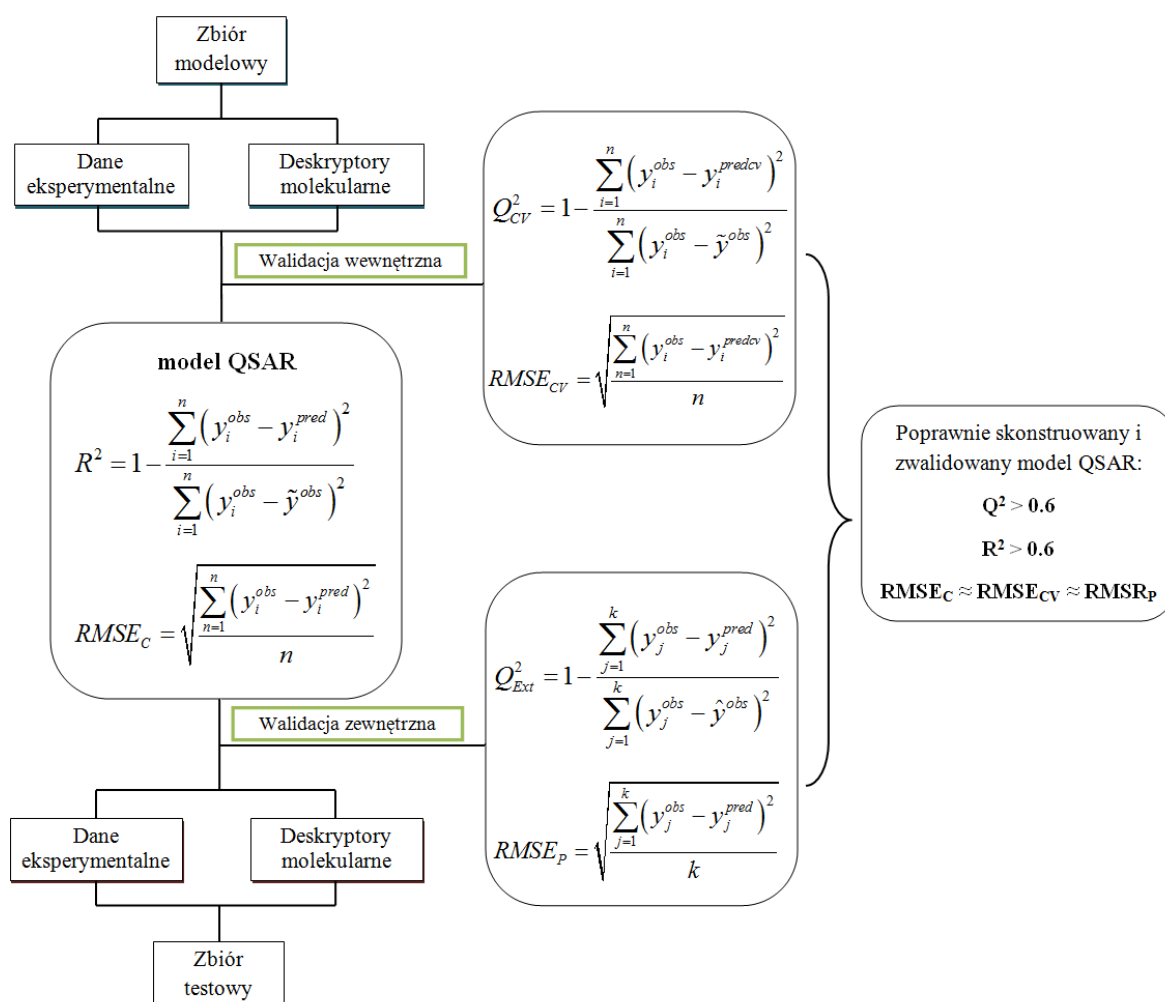
SCF CALCULATIONS      = 1
COMPUTATION TIME      = 0.062 SECONDS

Aby uzyskać Pomoc, naciśnij klawisz F1.
NUM
  
```

Mając z jednej strony konsystentnie wyznaczone dane eksperymentalne a z drugiej poprawnie wyznaczone deskryptory, następnym etapem modelowania jest kalibracja finalnego modelu QSAR/QSPR. Etap ten polega na zdefiniowaniu zależności matematycznej pomiędzy deskryptorami strukturalnymi poszczególnych kongenerów a ich aktywnością biologiczną/właściwościami fizykochemicznymi, poszukując najlepszego równania modelu. Istnieje wiele metod chemometrycznych wykorzystywanych do budowy modeli QSAR/QSPR, począwszy od najprostszych metod regresji prostej i regresji wielokrotnej (MLR; ang. *multilinear regression*), poprzez bardziej skomplikowane metody regresji głównych składowych (PCR; ang. *principal components regression*) oraz częściowych najmniejszych kwadratów (PLS; ang. *partial least squares*) i ich warianty aż po metody takie jak sztuczne sieci neuronowe (NN; ang. *neural networks*). Do wyboru najbardziej optymalnego zestawu deskryptorów wykorzystuje się między innymi algorytmy genetyczne (GA, ang. *genetic algorithms*). Następnie dokonujemy walidacji wewnętrznej (krosswalidacji). Ostatnim etapem, poprzedzającym właściwe przewidywanie toksyczności dla nowych związków chemicznych, jest walidacja zewnętrzna modelu.

Walidacja powinna być przeprowadzana zgodnie z procedurami rekomendowanymi przez OECD, polegającymi na ocenie:

- Czy model jest dobrze dopasowany do punktów pomiarowych? – miarą jakości dopasowania modelu są: współczynnik determinacji modelu R^2 oraz średni błąd kwadratowy zbioru modelowego $RMSE_C$ (ang. *root mean square error of calibration*).
- Czy model jest wystarczająco stabilny? – stabilność modelu potwierdzona zostanie z wykorzystaniem walidacji wewnętrznej – walidacji krzyżowej typu „wyrzucić jeden obiekt”. Miarą stabilności modelu są: współczynnik kroswalidacji Q_{CV} oraz średni błąd kwadratowy kroswalidacji $RMSE_{CV}$ (ang. *root mean square error of cross-validation*).
- Czy model posiada wystarczające zdolności prognostyczne? – zdolności prognostyczne modelu zweryfikowane zostaną za pomocą walidacji zewnętrznej, czyli oceny zdolności predykcyjnych dla zbioru próbek, który nie był użyty do konstrukcji modelu. Miarą zdolności prognostycznych są: współczynnik walidacji Q_{Ext} oraz średni błąd kwadratowy przewidywania $RMSE_{Ext}$ (ang. *root mean square error of prediction*).
- Czy wszystkie obiekty, dla których przewidujemy odpowiedź należą do dziedziny modelu? – Dziedzina modelu zostanie wyznaczona w oparciu o: (i) wartości współczynnika dźwigni (ang. *leverages*) oraz (ii) wykres powierzchni błędu modelu w zależności od wartości deskryptorów.



gdzie: y_i^{obs} – eksperymentalna (obserwowana) wartość dla i -tego związku; y_i^{pred} – przewidywana wartość dla i -tego związku; \bar{y}^{obs} – średnia eksperymentalna wartość dla związków zbioru modelowego; n – liczba związków w zbiorze modelowym, y_i^{predcv} – przewidywana wartość dla i -tego związku tymczasowo wyłączonego ze zbioru modelowego w metodzie walidacji krzyżowej;

y_j^{obs} – eksperymentalna (obserwowana) wartość dla j -tego związku; y_j^{pred} – przewidywana wartość dla j -tego związku; \hat{y}^{obs} – średnia eksperymentalna wartość dla związków zbioru testowego; k – liczba związków w zbiorze testowym.

Przykład: Przewidywanie wartości współczynników podziału n-oktanol/woda ($\log K_{\text{OW}}$), n-oktanol/powietrze ($\log K_{\text{OA}}$) i powietrze/woda ($\log K_{\text{AW}}$)

W poniższym przykładzie model QSPR skonstruowany został przy wykorzystaniu metody regresji wielokrotnej (MLR). Zmienne niezależne (deskrytory molekularne: moment dipolowy, D oraz powierzchnia molekuly dostępna dla rozpuszczalnika, SAS) wybrano arbitralnie w oparciu o sugestie literaturowe oraz intuicję chemiczną, zgodnie z którymi lipofilność ($\log K_{\text{OW}}$) związków organicznych maleje ze wzrostem polarności cząsteczki oraz rośnie wraz ze zwiększaniem się powierzchni molekuly. Otrzymano model opisany następującym równaniem:

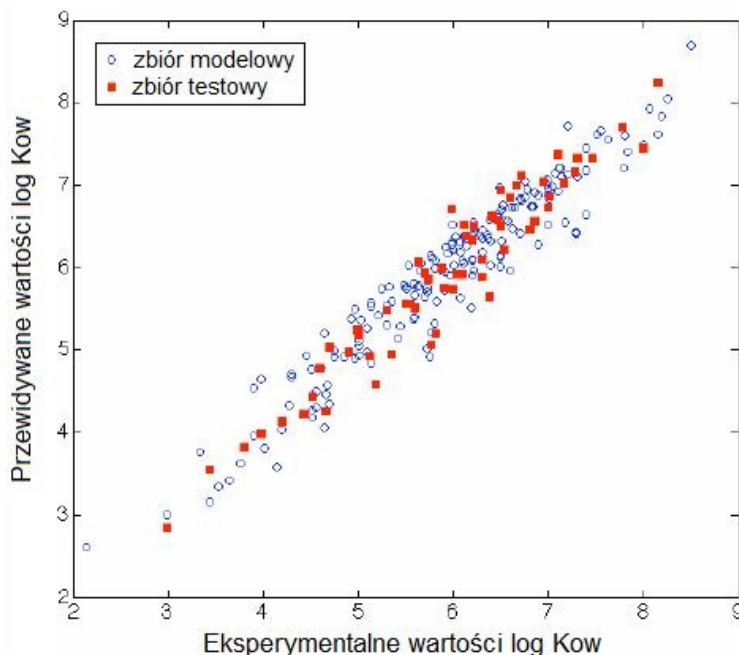
$$\log K_{\text{OW}} = -0,3587 - 0,1220 D + 0,0247 SAS$$

$$n=178; n_{\text{test}}=59; R^2=0,920;$$

$$Q^2_{\text{CV}}=0,918; Q^2_{\text{Ext}}=0,924; \text{RMSE}_C=0,315; \text{RMSE}_{\text{CV}}=0,320; \text{RMSE}_P=0,302$$

Należy zauważyć, że wartości parametrów opisujących dopasowanie, elastyczność oraz zdolności predykcyjne modelu są zbliżone a korelacja pomiędzy $\log K_{\text{OW}}$ obserwowanymi a przewidywanymi z modelu (Rysunek 3) wysoka. Rezultaty walidacji upoważniają więc do prowadzenia przewidywania wartości współczynnika podziału dla nowych związków.

Rys. 1. Korelacja pomiędzy obserwowanymi i przewidywanymi wartościami $\log K_{\text{OW}}$



Porównanie obliczonych wartości współczynników podziału n-oktanol/woda ($\log K_{\text{OW}}$), n-oktanol/powietrze ($\log K_{\text{OA}}$) i powietrze/woda ($\log K_{\text{AW}}$) z danymi wyznaczonymi eksperymentalnie oraz z modelami oceny właściwości fizyko-chemicznych obliczonymi przez Amerykańską Agencję Ochrony Środowiska:

Postępując zgodnie z opisaną powyżej procedurą dla cząsteczki **2,3,3',4',6-pentachlorobifenylu** (PCB-110) wyznaczono wartości dwóch deskryptorów molekularnych:

1. moment dipolowy, **D=2.77** [Debaj]

2. powierzchnia molekuly dostępna dla rozpuszczalnika, **SAS=275.05 Å²**

Wartości uzyskanych deskryptorów posłużyły do wyznaczenia współczynnika podziału n-oktanol/woda z wykorzystaniem modelu QSPR:

$$\log K_{OW} = -0.3587 - 0.1220 D + 0.0247 SAS$$

$$\log K_{OW} = 6.10$$

Uzyskany wynik można porównać do wartości wyznaczonej na drodze eksperymentalnej jak również do danych generowanych przez pakiety rekomendowane przez Amerykańską Agencję Ochrony Środowiska (ang. *The Environmental Protection Agency U.S. EPA*).

IUPAC	Związek	Smiles	log K _{ow}		
			Exp	QSAR	EPISuite
PCB-110	2,3,3',4',6-pentachlorobifenyl	c1(Cl)ccc(Cl)c(Cl)c1c2cc(Cl)c(Cl)cc2	6.30	6.10	6.98

EPI Suite - Welcome Screen

Input CAS #

Input Smiles:

Input Chem Name:

Name Lookup

Henry LC: atm-m³/mole

Water Solubility: mg/L

Melting Point: Celsius

Vapor Pressure: mm Hg

Boiling Point: Celsius

Log Kow:

River Lake

Water Depth: meters


Wind Velocity: meters/sec

Current Velocity: meters/sec

The Estimation Programs Interface (EPI) Suite™ was developed by the US Environmental Protection Agency's Office of Pollution Prevention and Toxics and Syracuse Research Corporation (SRC). It is a screening-level tool, intended for use in applications such as to quickly screen chemicals for release potential and "bin" chemicals by priority for future work. Estimated values should not be used when experimental (measured) values are available.

EPI Suite™ cannot be used for all chemical substances. The intended application domain is organic chemicals. Inorganic and organometallic chemicals generally are outside the domain.

Important information on the performance, development and application of EPI Suite™ and the individual programs within it can be found under the Help tab. Copyright 2000-2008 United States Environmental Protection Agency for EPI Suite™ and all component programs except BioHCwin and KOAWIN.

 **ChemSpider**
Building community for chemists

Search ChemSpider

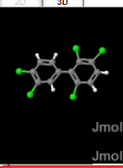
About Login Search Services Support Contact Us

Top
Inherent Properties
Predicted Properties
Focused Libraries
Data Sources
Synonyms
Database IDs
Miscellaneous
LASSO
Spectra
Images
CIFs
Feedback
History

INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES

ChemSpider ID: 34850
Empirical Formula: C₁₂H₅Cl₅
Molecular Weight: 326.4331
Nominal Mass: 324 Da
Average Mass: 326.4331 Da
Monoisotopic Mass: 323.883389 Da

Quick Links: [Permalink](#) [Similar](#) [Isomers](#) [Wikibox](#)

 Jmol
Jmol 1,2,4-trichloro-3-(3,4-dichlorophenyl)benzene

SMILES: Clc2c(c1ccc(Cl)c(Cl)c1)c(Cl)ccc2Cl
InChI: InChI=1/C12H5Cl5/c13-7-2-1-6(5-10(7)16)11-8(14)3-4-9(15)12(11)17/h1-5H
InChIKey: ARXHJUMGSIYYRZ-UHFFFAOYAT
Std. InChI: InChI=1S/C12H5Cl5/c13-7-2-1-6(5-10(7)16)11-8(14)3-4-9(15)12(11)17/h1-5H
Std. InChIKey: ARXHJUMGSIYYRZ-UHFFFAOYSA-N

NAMES AND SYNONYMS

Validated by Experts, Validated by Users, Non-Validated, Removed by Users, Redirected by Users, Redirect Approved by Experts

2,3,3',4',6-Pentachloro-1,1'-biphenyl
2,3,3',4',6-Pentachloro-1,1'-biphenyl
2,3,3',4',6-pentachlorobiphenyl
2,3,6,3',4'-pentachlorobiphenyl
38380-03-9 [R/N]
1,1'-Biphenyl, 2,3,3',4',6-pentachloro-
Chlorobiphenyl 110

DATABASE ID(S)

PREDICTED PROPERTIES

LogP:	ACD/LogP: 6.32 XLogP: 7.10 ALOGPS: 6.78	# of Rule of 5 Violations:	1
ACD/LogD (pH 5.5):	6.32	ACD/LogD (pH 7.4):	6.32
ACD/BCF (pH 5.5):	37665.81	ACD/BCF (pH 7.4):	37665.81
ACD/KOC (pH 5.5):	65621.57	ACD/KOC (pH 7.4):	65621.57
#H bond acceptors:	0	#H bond donors:	0
#Freely Rotating Bonds:	1	Polar Surface Area:	0 Å ²
Index of Refraction:	1.619	Molar Refractivity:	75.31 cm ³

EPI Suite

File Edit Functions Batch Mode Show Structure Output Fugacity STP Help

EPI Suite - Welcome Screen

PhysProp Previous Get User Save User Search CAS **Calculate** Clear Input Fields

Input CAS #:
Input Smiles: Clc2c(c1ccc(Cl)c(Cl)c1)c(Cl)ccc2Cl
Input Chem Name: 2,3,3',4',6-pentachlorobiphenyl
Name Lookup

Henry LC: atm·m /mole Water Solubility: mg/L
Melting Point: Celsius Vapor Pressure: mm Hg
Boiling Point: Celsius Log Kow:

River Lake
Water Depth: meters
Wind Velocity: meters/sec
Current Velocity: meters/sec

Output
☐ Full
☒ Summary

The Estimation Programs Interface (EPI) Suite™ was developed by the US Environmental Protection Agency's Office of Pollution Prevention and Toxics and Syracuse Research Corporation (SRC). It is a screening-level tool, intended for use in applications such as to quickly screen chemicals for release potential and "bin" chemicals by priority for future work. Estimated values should not be used when experimental (measured) values are available.

EPI Suite™ cannot be used for all chemical substances. The intended application domain is organic chemicals. Inorganic and organometallic chemicals generally are outside the domain.

Important information on the performance, development and application of EPI Suite™ and the individual programs within it can be found under the Help tab. Copyright 2000-2008 United States Environmental Protection Agency for EPI Suite™ and all component programs except BioHCWIN and KOAWIN.

IV. Przebieg ćwiczenia:

Za pomocą programu Molden proszę zbudować cząsteczkę octachloronaphthalene (CAS 2234-13-1), a następnie półempiryczną metodą PM6 proszę zoptymalizować jej strukturę oraz wyznaczyć poniższe deskryptory molekularne:

- 1) moment dipolowy, **D**
- 2) powierzchnia molekuly dostępna dla rozpuszczalnika, **SAS**
- 3) energia najwyższego obsadzonego orbitalu molekularnego, **ϵ_{HOMO}**
- 4) polaryzowalność, **α**

Na podstawie otrzymanych wartości proszę obliczyć współczynniki podziału: n-oktanol/woda ($\log K_{\text{OW}}$), n-oktanol/powietrze ($\log K_{\text{OA}}$) powietrze/woda ($\log K_{\text{AW}}$) wykorzystując poniższe równania:

$$\log K_{\text{OW}} = -0.3587 - 0.1220 D + 0.0247 SAS$$

$$\log K_{\text{OA}} = 7.3108 + 0.7408 \epsilon_{\text{HOMO}} + 0.2862 \alpha$$

Ponadto proszę porównać uzyskane wyniki z wartościami eksperymentalnymi ($\log K_{\text{OW}}=6.42$, $\log K_{\text{OA}}=11.48$, $\log K_{\text{AW}}=-5.06$) oraz w wartościami wyznaczonymi za pomocą programu Amerykańskiej Agencji Ochrony Środowiska (EPI Suite).

V. Sprawozdanie:

Sprawozdanie w wersji elektronicznej powinno zawierać wstęp, w którym opiszą Państwo (i) przyczynę i cel wykonywanej analiz; (ii) uzyskane rezultaty (iii) dyskusję oraz **wnioski**.

Obowiązuje **odpowiednia forma** sprawozdania. Sprawozdania niezgodne z poniższym formatem **nie zostaną** sprawdzone.

Format: **.doc**. Nazwa pliku według schematu: **Nazwisko_numer.indeksu_ćw2_QSPR_spr.doc**

VI. Literatura

- T. Puzyn, A. Mostrąg-Szlichtyng, N. Suzuki, M. Harańczyk, Metody Chemometryczne w ocenie ryzyka: ilościowe zależności pomiędzy strukturą chemiczną a właściwościami (QSPR) dla nowych rodzajów zanieczyszczeń chemicznych. Chemometria w nauce i praktyce. Kraków, 2009.
- T. Puzyn, N. Suzuki, M. Haranczyk, How do the partitioning properties of polyhalogenated POPs change when chlorine is replaced with bromine? *Environ. Sci. Technol.* 2008, 42 (14), 5189–5195
- <http://www.chemspider.com>
- <http://openmopac.net/manual/index.html>

VII. Oprogramowanie:

- Molden 4.7 for Windows
<http://www.cmbi.ru.nl/molden/windowsnt95.html>
- UWAGA: do użytkowania programu Molden pod systemem Windows dodatkowo potrzebny jest program Xming
<http://sourceforge.net/projects/xming/>
- OpenMOPAC:
<http://openmopac.net/downloads.html>
- Estimation Program Interface (EPI) Suite:
<http://www.epa.gov/oppt/exposure/pubs/episuite.htm>