MODELOWANIE W OCHRONIE ŚRODOWISKA

Instrukcja do ćwiczeń laboratoryjnych

Ćwiczenie 2

1. Cel ćwiczenia:

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

2. Zagadnienia do samodzielnego opracowania:

Idea ilościowego modelowania struktura-aktywność (QSAR) oraz struktura-właściwości (QSPR). Podstawy modelowania molekularnego: specyfikacja geometrii we współrzędnych wewnętrznych i kartezjańskich; optymalizacja geometrii cząsteczki. Deskryptory molekularne, kalibracja modelu QSAR/QSPR w oparciu o metodę regresji wielokrotnej. Walidacja modelu w oparciu o reguły OECD.

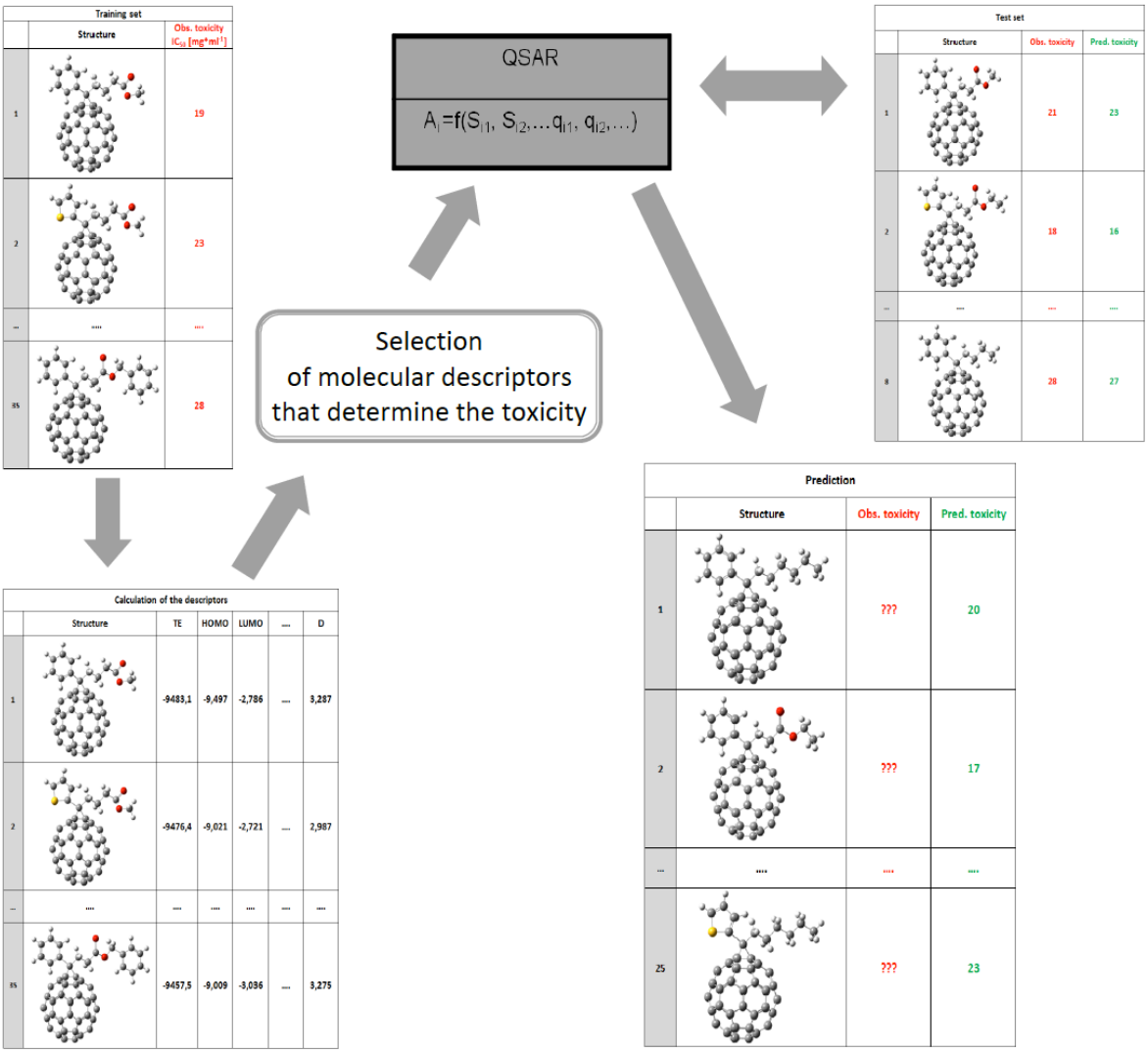
Trwałe zanieczyszczenia organiczne (POPs). Współczynniki podziału: n-oktanol/woda (K_{OW}), n-oktanol/powietrze (K_{OA}), powietrze/woda (K_{AW}) oraz metody ich eksperymentalnego oraz komputerowego wyznaczania. Znaczenie współczynników podziału w ocenie ryzyka dla nowych substancji chemicznych.

3. 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 miedzy strukturą chemiczną reprezentowaną przez zbiór deskryptorów a właściwościami fizykochemicznymi (QSPR) lub aktywnością biologiczną (QSAR). U podstaw metody leży założenie, iż różnice pomiędzy właściwościami fizyko-chemicznymi 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 tych dla których danych eksperymentalnych brakuje. Idea metodologii QSAR/QSPR przedstawiona została na Rysunku1.



Rys. 1 Idea metodologii QSAR/QSPR

Metodologia QSAR obejmuje pięć kolejnych kroków:

i. Zebranie danych eksperymentalnych i podział całego zbioru na:
Zbiór uczący (treningowy)
Zbiór testowy (walidacyjny)
ii. Obliczenie deskryptorów molekularnych
iii. Zbudowanie metody (kalibracja modelu)
iv. Testowanie metody (walidacja modelu)
Walidacja wewnętrzna (ang. Internal validation)
Walidacja zewnętrzna (ang. External validation)
v. 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 wewnętrznej walidacji modelu
- testowy (30-25%), wykorzystywany do przeprowadzenia walidacji zewnętrznej (nieużywany podczas kalibracji modelu).

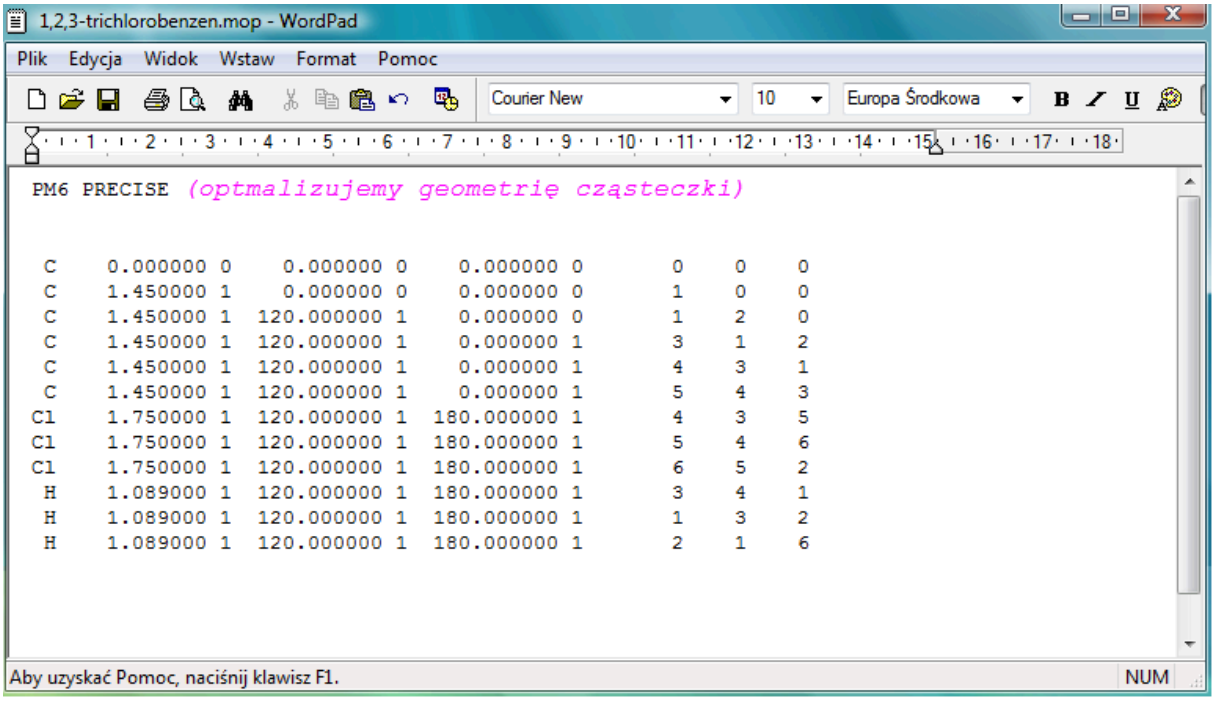
Kolejnym etapem procedury QSAR/QSPR jest obliczenie tzw. deskryptorów strukturalnych opisujących struktury molekularne wszystkich badanych związków chemicznych (zarówno tych, dla których są dostępne dane eksperymentalne, jak i tych, dla których danych takich brakuje).

Do zbudowania niezbędnych struktur można użyć dostępne programy komputerowe np. Avogadro, Molden. Po zapisaniu struktury każdego związku w postaci tzw. Macierzy współrzędnych wewnętrznych („z-macierzy") 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, PM7) 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.

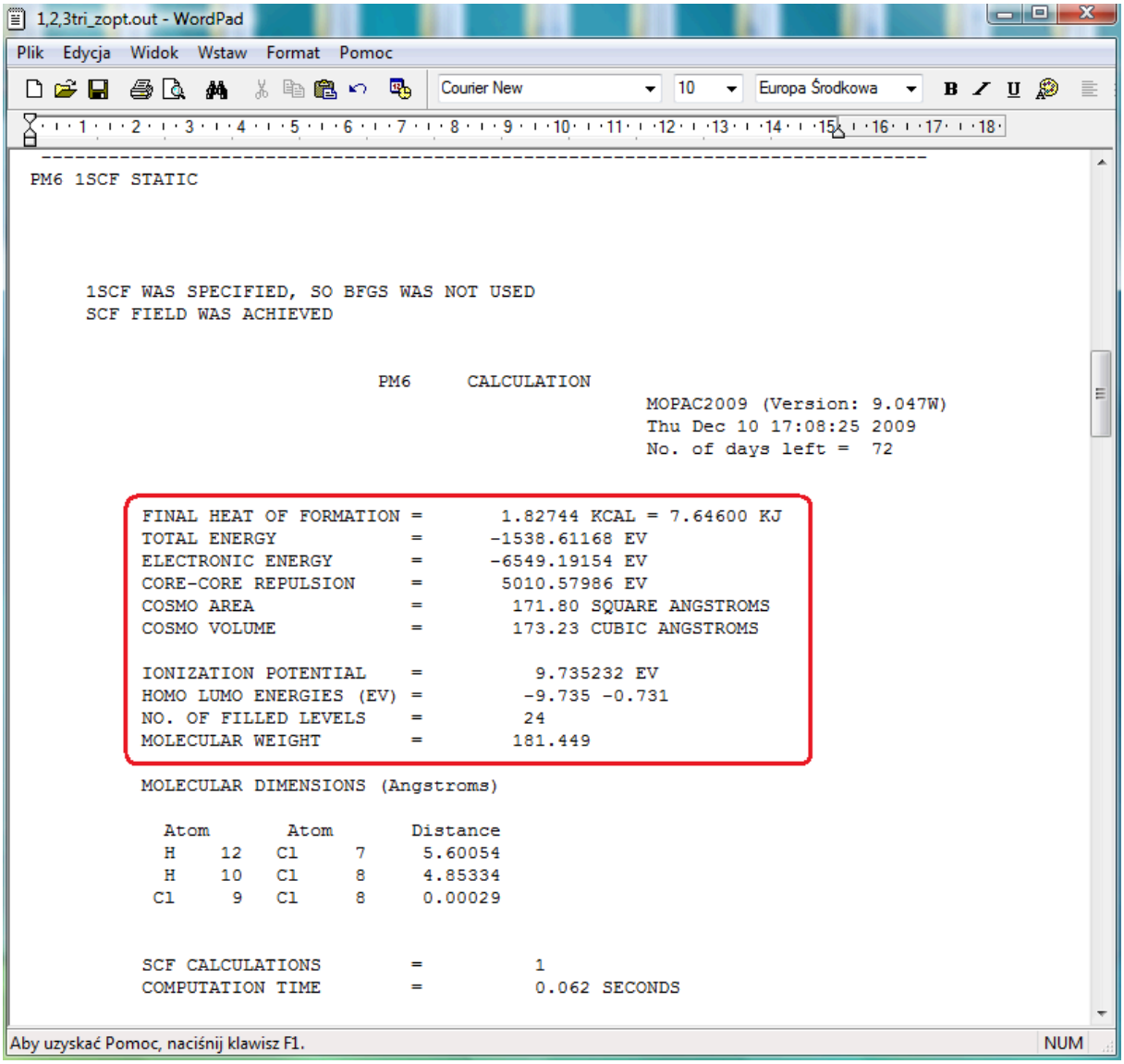
Utworzoną w Avogadro i zapisaną w formacie .mop cząsteczkę, otwieramy jako plik tekstowy a następnie przygotowujemy plik wejściowy (tak zwany "input") wpisując w pierwszej linii słowa kluczowe, w dwóch kolejnych komentarz (ewentualnie te dwie linie pozostawia się puste), a następnie w czwartej linii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

współrzędnych wewnętrznych.



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 i tak zapisaną geometrię wykorzystuje się w obliczeniach kwantowo-mechanicznych wprowadzając dodatkowe słowa kluczowe (Tabela1 - załącznik).

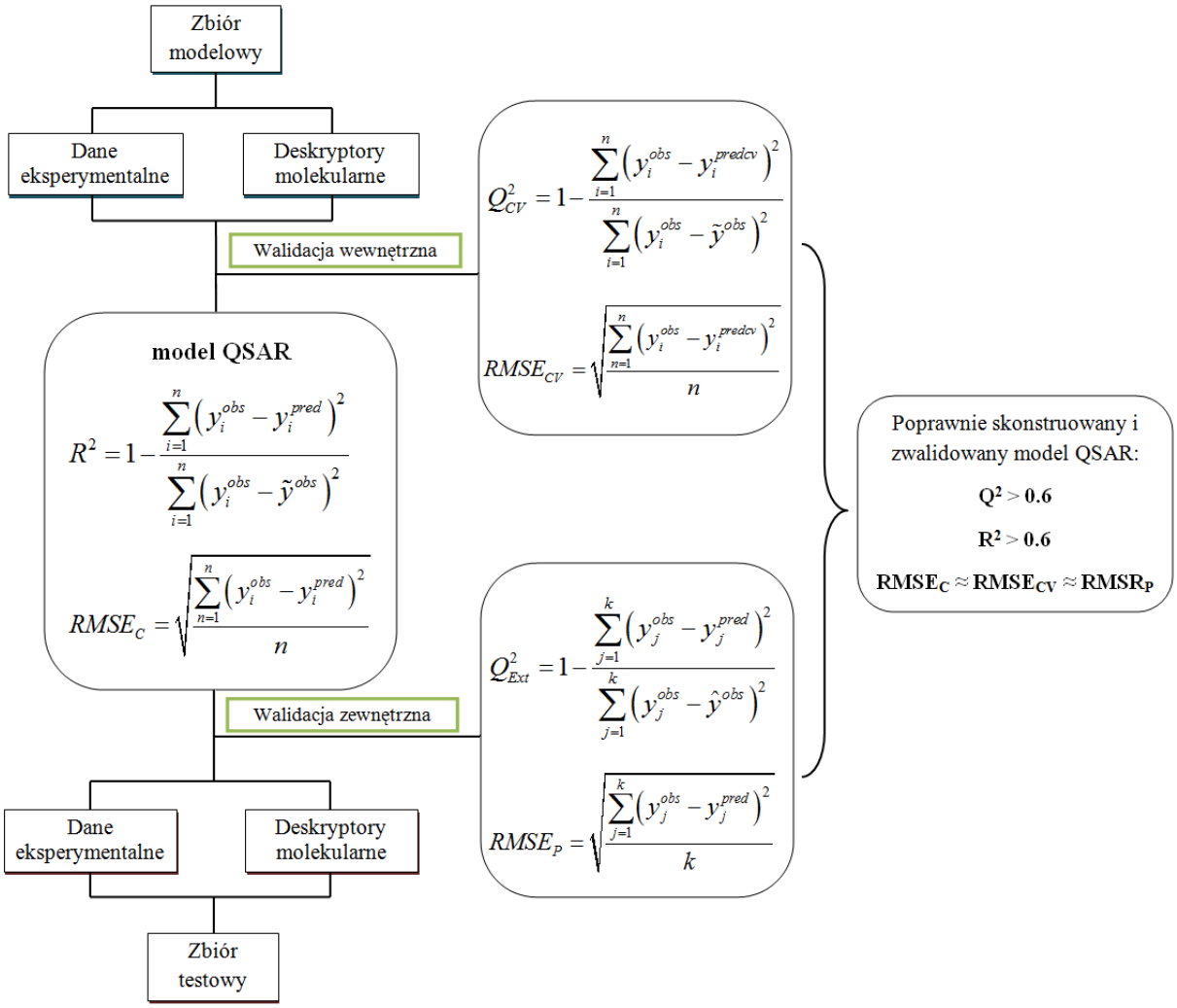
W wyniku obliczeń, otrzymujemy pliki wyjściowe typu .out (tak zwane "outputy"). Pliki te otwieramy jako pliki tekstowe, a następnie odczytujemy z nich wartości poszczególnych deskryptorów.



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. Multi linear 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 (ANN; ang. Artificial Neural networks). Do wyboru najbardziej optymalnego zestawu deskryptorów wykorzystuje się między innymi algorytmy genetyczne (GA, ang. genetic algorithms). Następnie dokonujemy walidacji wewnętrznej (ang. cross-validation). Ostatnim etapem, poprzedzającym właściwe przewidywanie toksyczności dla nowych związków chemicznych, jest walidacja zewnętrzna modelu.

Walidacja powinna być przeprowadza na 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 R2 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 „wyrzuć jeden obiekt". Miarą stabilności modelu są: współczynnik walidacji wewnętrznej Q_{CV} oraz średni błąd kwadratowy walidacji wewnętrznej $RMSE_{CV}$ (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 zewnętrznej QExt oraz średni błąd kwadratowy przewidywania $RMSE_{Ext}$ (ang. Root mean square error of prediction).
- Czy wszystkie obiekty, dla których przewidujemy odpowiedz 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.

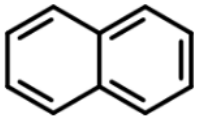


gdzie: y_i^{obs} –eksperymentalna (obserwowana) wartość dla i-tego związku; y_i^{pred}– przewidywana wartość dla i-tego związku; $\tilde{y}^{obs}$ – średnia eksperymentalna wartość dla związków zbioru modelowego; n – liczba związków w zbiorze modelowym, y_i^{predcv} – przewidywana wartość dla i-tego 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.

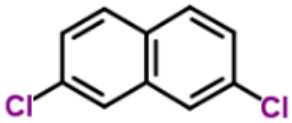
3. Przebieg ćwiczenia:

Za pomocą programu Avogadro proszę zbudować modele molekularne sześciu wymienionych cząsteczek:

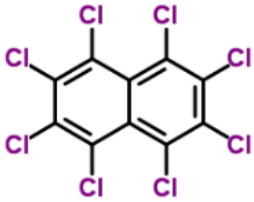
- Naphthalene



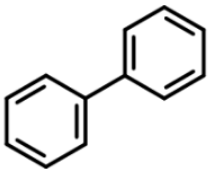
- 2,7-dichloronaphthalene



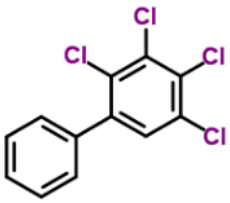
- Octachloronaphthalene



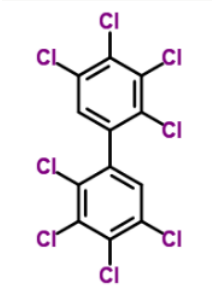
- Biphenyl



- 2,3,4,5-tetrachlorobiphenyl



- 2,3,4,5,2',3',4',5'-octachlorobiphenyl



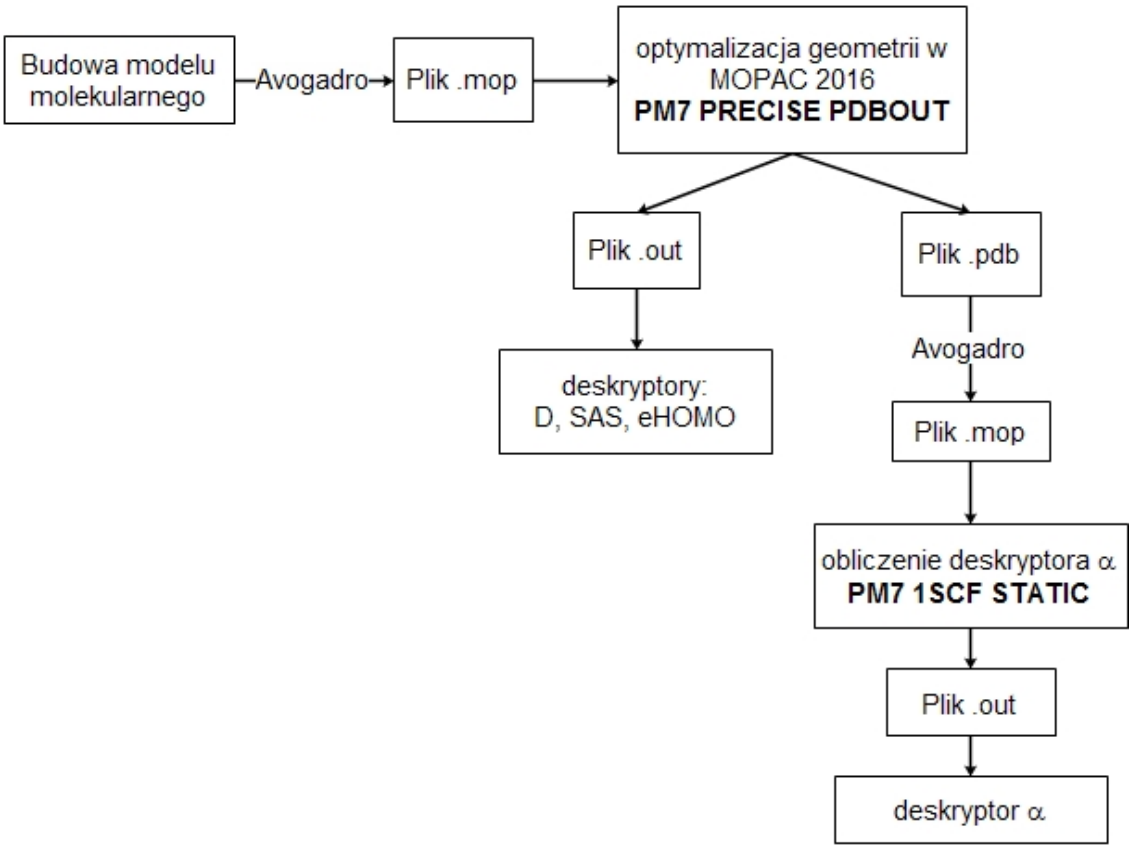
Następnie (zgodnie ze schematem) korzystając z programu MOPAC2016 zoptymalizuj geometrię zbudowanych cząsteczek za pomocą metody PM7. Wyznacz poniższe deskryptory molekularne:

- ✓ Moment dipolowy, **D**
- ✓ Powierzchnia molekuły dostępna dla rozpuszczalnika, **SAS**
- ✓ Energia najwyższego obsadzonego orbitalu molekularnego, **εHOMO**
- ✓ Polaryzowalność, **α**

Na podstawie otrzymanych wartości deskryptorów proszę obliczyć współczynniki podziału log K_{OW} , log K_{OA} oraz log K_{AW} wykorzystując poniższe równania:

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

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



Ponadto proszę porównać uzyskane wyniki z wartościami eksperymentalnymi (Tabela 1) oraz wartościami wyznaczonymi za pomocą programu ACD/Lab oraz programu Amerykańskiej Agencji Ochrony Środowiska EPISuite umieszczonymi w bazie www.chemspider.com .

4. Sprawozdanie:

Sprawozdanie w wersji elektronicznej powinno zawierać wstęp, w którym opiszą Państwo przyczynę i cel wykonanych analiz; uzyskane rezultaty; dyskusję wyników oraz wnioski.

5. Literatura

- Puzyn, T.; Mostrąg-Szlichtyng, A.; Suzuki, N.; Harańczyk, M., Metody chemometryczne w ocenie ryzyka: Ilościowe zależności pomiędzy strukturą chemiczną, a właściwościami (QSPR) dla nowych rodzajów zanieczyszczeń chemicznych., *Chemometria w nauce i praktyce*, Wydawnictwo Instytutu Ekspertyz Sądowych, Kraków, 2009; pp 61-71.

- Puzyn, T.; Suzuki, N.; Haranczyk, M., *Environ. Sci. Technol.* **2008,** *42*, 5189-5195.

- http://wwwóchemspiderócom/

- http://openmopac.net/manual/

6. Oprogramowanie

- Avogadro (http://avogadroócc/wiki/Main_Page)
- MOPAC2016 (http://openmopac.net/)
- Padel (http://www.yapcwsoftócom/dd/padeldescriptor/)