

**Recenzja rozprawy doktorskiej mgr inż. Joachima Eichenlauba
pt. „Wykorzystanie modelowania QSPR w celu przewidywania właściwości
fizykochemicznych cieczy jonowych”**

Praca wykonana została pod kierunkiem promotora prof. dra hab. inż. Adama Kloskowskiego w trakcie kształcenia w Szkole Doktorskiej w Politechnice Gdańskiej.

Przedmiotem badań Doktoranta jest modelowanie QSPR właściwości chemicznych. Modelowaną właściwością była rozpuszczalność ditlenku węgla w cieczach jonowych. Doktorant jako właściwość charakteryzującą rozpuszczalność przyjął stałą Henry’ego. Praca doktorska porusza istotne problemy współczesnej chemii. W związku z uwarunkowaniami środowiskowymi zagadnienia związane z ditlenkiem węgla są obecnie przedmiotem gwałtownego wzrostu zainteresowania chemii fizycznej oraz chemii środowiska. O ile przedmiotem badań Doktoranta jest ditlenek węgla w cieczach jonowych, metodą badawczą jest QSPR Quantitative Structure–Activity Relationship, która wywodzi się od QSAR, Quantitative Structure–Activity Relationship. Pierwszy model QSAR Hanscha, powstał w latach 1950’. Był pierwszym algorytmem który pozwalał na eksplorację danych aktywności biologicznych (A-activity) jako funkcji strukturalnej budowy cząsteczek. Model Hanscha operował w szczególności stałą podziału pomiędzy wodę a n-oktanol, logP, jako główną zmienną niezależną (predyktor), która miała modelować transport leku przez błony komórkowe. Przez lata powstało wiele odmian QSAR, które stopniowo zwiększały zakres analizowanych predyktorów, modelując zwykle pojedynczą zmienną zależną. Taka nierównowaga zmiennej mierzonej oraz obliczanych, tzw. deskryptorów cząsteczkowych, prowadziła do problemów wiarygodności statystycznej uzyskiwanych modeli. Stopniowo zorientowano się, że typowa statystyka nie jest odpowiednia dla badań tego typu. Z jednej strony klasyczny QSAR przynosi istotny przełom w zrozumieniu potencjalnych uproszczonych modeli mechanizmów oddziaływania leku, z drugiej strony, cały szereg problemów. Jednym z ważniejszych jest dychotomia między zdolność modelowania danych vs. prognozowania nowych obiektów. Z reguły QSAR wielowymiarowy stosuje modelowanie PLS, połączone z walidacją krzyżową LOO (leave-one-out), tworzeniem zbiorów testowych, w których umieszcza się szereg badanych obiektów, które nie są wykorzystywane w etapie treningowym. Z jednej strony QSAR był tak nośną koncepcją, że przez lata termin ten był synonimem projektowania leków. Z drugiej zaś, metody QSAR nigdy nie osiągnęły pokładanych w nich nadziei, to znaczy efektywnego opracowywania rzeczywistych leków. Porównaj np. publikacja [Johnson, S.R. The Trouble with QSAR or How I Learned To Stop Worrying and Embrace Fallacy. *J. Chem. Inf. Model.* **2008**, 48, 25–26, doi:10.1021/ci700332k].

Doktorant jako cele badawcze stawia sobie: (i) zbudowanie modelu QSPR w celu przewidywania wartości stałej Henry’ego ditlenku węgla w cieczach jonowych, (ii)

przeprowadzenie analizy wpływu wolnej objętości w cieczach jonowych na proces absorpcji ditlenku węgla, (iii) analizę wpływu zmian środowiska oraz sposobu optymalizacji struktur cieczy jonowych na uzyskiwane modele QSPR.

W wielowymiarowym QSAR liczba zmiennych niezależnych może dochodzić nawet do kilkudziesięciu tysięcy. Najważniejszą sprawą staje się redukcja zmiennych. Doktorat nie stosuje takich ekstremalnych modeli. Jako zmienne niezależne wykorzystuje standardowe deskryptory molekularne, które można obliczyć dostępnymi programami komputerowymi. Wykorzystując różne formalizmy statystyczne (PLS, MLR, LR) modeluje zależność pomiędzy stałą Henry'ego, a tak obliczonymi deskryptorami. W wyniku obliczeń uzyskuje statystycznie poprawne modele QSPR. Najcenniejszym wynikiem jest interpretacja tak uzyskanych modeli, pozwalająca na identyfikację cech strukturalnych cieczy jonowych, które wpływają na rozpuszczalność CO₂. Pozwala ona zidentyfikować, wolną objętość w cieczach jonowych, zależną głównie od struktury kationu jako czynnik wpływający na rozpuszczalność CO₂. Okazuje się, że *deskryptory molekularne 3D wykazują znaczną podatność na zmiany w sposobie optymalizacji struktur cieczy jonowych*. Ciekawym efektem obserwowanym przez Autora jest fakt, że *wartości deskryptorów zmieniają się z zachowaniem tego samego trendu dla każdej z metod, co w rezultacie jedynie nieznacznie przekłada się na wpływ na parametry otrzymanych modeli QSPR*. To znany efekt opisany w modelowaniu QSAR. Obserwowałem go też w wielu moich pracach. Zmiany metody nie prowadzą do zmian statystyki modeli dla danego szeregu związków i właściwości. Zgadzam się z tezą Autora, że efekt wynika z faktu, iż *wartości deskryptorów zmieniają się z zachowaniem tego samego trendu dla każdej z metod*. Deskryptory dekodują informacje, eksplorując zawsze te same dane molekularne. W pewnym sensie, kiedy człowiek analizuje struktury chemiczne przez ich proste *oglądanie*, też potrafi znaleźć różne aspekty ich podobieństwa i różnic.

Warto podkreślić, że Doktorant prawidłowo stosuje formalizmy statystyczne, używając poprawnych standardowo przyjętych statystyk. Poza standardowym R^2 stosuje Q^2 , R^2_{pred} oraz Y_{rand} , a uzyskiwane przez niego wyniki są też prawidłowo interpretowane. Tutaj natomiast skrytykować muszę braki w informacji o stosowanej metodzie, nie podaje bowiem informacji, czy stosuje walidację krzyżową LOO (leave-one-out) czy też z wieloma opuszczeniami? Więcej, w metodzie podziału na zbiór treningowy i testowy nie podaje ich liczebności. Tę informację czytelnik może dopiero uzyskać w materiałach uzupełniających w Tabeli A2 na stronie 130, i to samodzielnie licząc etykiety przypisane kolejnym obiektom. Tu kolejne pytanie, w jaki sposób dzielono populację na zbiór treningowy i testowy. Warto powiedzieć, że te czynniki mają fundamentalne znaczenie dla wyników (por. np. Polanski, et al. Modeling robust QSAR. J.Chem.Inf.Mod. 46.6 (2006): 2310-2318 lub Golbraikh, A., & Tropsha, A. (2002). Beware of q²!. Journal of molecular graphics and modelling, 20(4), 269-276). Poza operacjami obejmującymi deskryptory Doktorant ekstrahuje z literatury dostępne wyniki modelowanej właściwości HCO₂. Ponieważ zestaw tak otrzymanych danych zawierał wartości pomiarów przeprowadzonych w szerokim zakresie temperatur, w celu wykluczenia wpływu tego parametru na planowane modele matematyczne, dane sortuje i filtruje w taki sposób, aby otrzymać jak największy zbiór wartości uzyskanych w identycznych warunkach laboratoryjnych. Otrzymuje zestawy 55 pomiarów w temperaturze 298 K oraz 39 pomiarów w temperaturze 303 K. (Tabela 3.1).

Konstrukcja rozprawy jest typowa dla pracy doktorskiej. Składa się z trzech części: (i) literaturowej, zatytułowanej Wprowadzenie, gdzie Autor omawia problemy chemii cieczy jonowych oraz modelowania QSPR, rozpuszczalności gazów, zielonych rozpuszczalników, ditlenku węgla jako gaz cieplarniany, oraz wolnej objętości w cieczach jonowych, jako

czynników związanych z właściwościami modelowanego układu CO₂-ciecz jonowa. Dalej omawia (ii) cele i tezy pracy oraz (iii) metody badawcze, (iv) wyniki, (v) wnioski, (vi) podsumowanie, (vii) bibliografię. Pracę uzupełnia spis dorobku naukowego, spis ilustracji i tabel oraz materiały uzupełniające. W części bibliograficznej cytuje 249 pozycji literaturowych.

Kilka uwag krytycznych do dyskusji, które nasunęły mi się podczas lektury:

1. Formalnie metoda stosowana przez Doktoranta jest interpretowalna, chociaż nie jest to interpretacja bezpośrednia, taka, jaką w metodach QSAR (QSPR) omawia Polischuk, Polishchuk, P. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. J. Chem. Inf. Model. 2017, 57 (11), 2618–2639. Ponieważ rozumiemy coraz lepiej jak ważne są problemy interpretowalności nawet w metodach AI (Explainable Artificial Intelligence, XAI), ciekaw jestem jak broniłby się Autor, kiedy postawiłbym tezę, że stosowane metody stoją nisko w rankingu XAI. Czy prezentowane metody należą do AI?
2. Jak zinterpretować opisane znaczenie *indeksu lekopodobieństwa*?
3. Chciałbym aby Doktorant przedstawił precyzyjnie problemy podziału obiektów na zbiór testowy i treningowy. Jakie formalizmy tu stosujemy, dlaczego jest to ważne, i jakie czynniki będą decydować, że efekt ten będzie mniej lub bardziej istotny dla wyników modelowania.
4. Chciałbym się dowiedzieć, czy prowadzone modelowania miały aspekt praktyczny, tzn. czy planowane są syntezy nowych cieczy jonowych w oparciu o uzyskane modele.

Liczę na atrakcyjną dyskusję w czasie obrony.

Pracę oceniam wysoko. To interesujące studium które mieści się w nowych trendach chemii materiałowej. Uzyskane wyniki stanowią istotny wkład w rozwój modelowania właściwości cieczy jonowych. Atutem pracy jest dobry język opisu. O wysokiej wartości pracy świadczy także dorobek Autora. To cztery publikacje (jedna w przygotowaniu). W trzech Doktorant jest pierwszym autorem. O zainteresowaniu świadczy duża liczba cytowań. Praca Eichenlaub, J., Rakowska, P. W., & Kloskowski, A. (2022). User-assisted methodology targeted for building structure interpretable QSPR models for boosting CO₂ capture with ionic liquids. Journal of Molecular Liquids, 350, 118511 ma już 24 cytowania (GS).

Konkludując, wyniki uzyskane w rozprawie doktorskiej stanowią istotny wkład w rozwój modelowania właściwości cieczy jonowych. Po analizie przedstawionych treści rozprawy potwierdzam ścisły związek pomiędzy przedstawionym celem pracy, a wnioskami końcowymi. Stwierdzam, że rozprawa doktorska mgr. inż. Joachima Eichenlauba spełnia wszystkie warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020 r. poz. 85 z późniejszymi zmianami). Wnioskuję zatem o dopuszczenie mgr. inż. Joachima Eichenlauba do dalszych etapów przewodu doktorskiego.