

R E C E N Z J A

Praca doktorska pt:
Analizy QSPR wielkich bibliotek związków chemicznych na przykładzie bazy
danych Abamachem

Autor:
Mgr Urszula KUCIA
Promotor
Prof. dr hab. inż. Jarosław POLAŃSKI

Pierwsze wrażenia

Pierwsza lektura bardzo obszernej pracy doktorskiej może wprowadzić czytelnika z zakresu technologii chemicznej w pewne zakłopotanie. Po pierwsze – kierunek badań *in silico* nie jest wciąż powszechny w technologii. Po drugie – główna problematyka pracy, którą jest badanie efektów ekonomicznych w chemii jest w tymże piśmiennictwie spotykane bardzo rzadko. W czytelniku odzywa się wątpliwość, czy w ogóle ilościowe uzależnienie ceny związku chemicznego od jego struktury chemicznej jest możliwe. Tytuł dysertacji nie zdradza, że to właśnie analiza ekonomiczna będzie przedmiotem wnikliwych poszukiwań w pracy doktorskiej, przechodząc od metodologii QSAR do QSPR i dalej do QSER, czyli właściwości w kategoriach ekonomicznych. Ostatecznie Doktorantka stawia sobie jako cel pracy opracowanie metod analizy wielkich zbiorów danych chemicznych (t zw. big data). Tutaj jest to biblioteka Abamachem zawierająca ponad 2,2 mln związków organicznych.

Kolejny krok to potrzeba umiejscowienia tej pracy w jakiejś płaszczyźnie umożliwiającej porównanie do standardowych prac doktorskich w tej konwencji. Już proste zestawienie w spisie literatury wskazuje, że jest to oryginalna praca w obszarze projektowania leków, od lat z dużym powodzeniem prowadzonym przez prof. J. Polańskiego, na 118 cytowanych pozycji 18 jest z udziałem Profesora, wśród których znajduje się ważny dla tej pracy rozdział w książce pt „Advances in QSAR Modeling Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences” wydanej przez Springer’a w 2017 r a cytowany pod poz. [34]. Należy podkreślić, że podane pozycje [97,98,109] są również dorobkiem Autorki dysertacji. Nie

ulega żadnym wątpliwości, że ta praca będąca ukoronowaniem wieloletnich badań w Zakładzie Chemii Organicznej Instytutu Chemii U.Śl. jest cennym wkładem w rozwój tej problematyki. W szczególności warto odnotowania jest w pracy doktorskiej opracowanie efektywnych metod zarządzania wielkimi zbiorami (big data) opisującymi numeryczne reprezentacje związków chemicznych.

Recenzowana praca doktorska formalnie dotyczy związków organicznych. Jasne jest jednak, że stanowią one substytut leków, dla których dostępność danych jest znacznie mniejsza. Doktorantka omawia „regułę pięciu” Christopher’a Lipińskiego, która pozwala zrozumieć jak szcążkowe jest wciąż nasze zrozumienie efektów aktywności biologicznej związków chemicznych.

Zgodnie z regułą istnieje prawdopodobieństwo, że dany związek chemiczny zostanie lekiem, jeżeli spełnia przynajmniej dwa spośród wymienionych następujących kryteriów (posiada nie więcej niż 5 donorów wiązań wodorowych, posiada nie więcej niż 10 akceptorów wiązania wodorowego, masa cząsteczkowa jest poniżej 500 Da, logP jest mniejszy niż 5).

Należy zdawać sobie sprawę z faktu, że efekty ekonomiczne są jeszcze bardziej ulotne ale ta praca przekonuje, że o losach i szansach nowych innowacyjnych rozwiązań decydują właśnie kategorie ekonomiczne. Bardzo ważnym elementem pracy jest poszukiwanie deskryptora, który może być skorelowany z efektami ekonomicznymi. Jako najprostszy sposób opisu złożoności cząsteczki praca przyjmuje MW, w zasadzie rozszerzając metodę wydajności ligandu (LE – ligand efficiency). Dalej Doktorantka bada przydatność opisu złożoności syntezy związku chemicznego opracowaną przez prof. Johanna Gasteigera (np. SAS1) co udokumentowano w tej pracy oraz publikacji własnej [97].

Można również w pierwszym czytaniu odczuwać dyskomfort napotyając na pojęcia: ekonomia atomowa, cena molowa itd.

Doktorantka wprowadza jednak czytelnika w świat bardzo licznej grupy pojęć w tym zakresie. To bardzo ciekawy fragment pracy, który pozwala połączyć wiele odległych od siebie pojęć które zlewają się w komplementarną całość w głównym nurcie łączącym efekty chemiczne i ekonomiczne.

Punkt centralny pracy

Pierwszym problemem analizy danych było żmudne przekształcanie zbioru Abamachem do postaci, którą można poddać dalszym operacjom *in silico*. Dalej, opracowanie efektywnych metod badania rozbudowanych relacji w zakresie statystyki zaawansowanej na poziomie molekularnym. Związki pomiędzy wybranymi deskryptorami molekularnymi zostały przeanalizowane w konwencji znanych ujęć w literaturze ale punkt centralny to genialna w swoim

efekcie ale bardzo prosta transformacja polegająca na odwróceniu liniowości skali MW jako wielkości odwrotnie proporcjonalnej, co w najprostszym zapisie oznacza:

$$\frac{\text{deskryptor}}{MW} \text{ vs. } MW$$

Tutaj deskryptor oznacza szeroko rozumianą aktywność ale też cenę, a klasyczna masa cząsteczkowa jako $1/MW$ przybiera wymiar [mol/g] czyli ilości materii do masy.

Elementem dopełniającym i oryginalnym dlatego rodzaju rozważań jest binowana masa molowa. W ten sposób utworzono deskryptory MBM [\$ /mol], WBM [\$ /g] oraz MW bin [g/mol]. W ten bardzo przejrzysty sposób pojęcie omawianej tutaj kategorii „cena” przestaje być odczuwana jako jednostka wyalienowana.

W ramach omawianej relacji masa cząsteczkowa jako miara złożoności cząsteczki (p. 10.2) i dalej (p. 11) analizuje się efekty hiperboliczne, przypadające na zakres MW bin = 238-274 (wcześniejszy rys. 31a). To zagadnienie jest omawiane bardzo szczegółowo w wersji graficznej i przy pomocy stałej Avogadry o symbolu N_A . Jest to również opublikowane w pracach własnych [97,98,109]. W moim przekonaniu jest to związane z faktem znanym w analizie wymiarowej przy ustalaniu bezwymiarowych kryteriów, gdy zmiennym przypisuje się najprostszy wymiar i jednocześnie koherentny, tak aby nie powstawały przeliczniki wymiarowe. Najlepszy dowód w pracy, że jest to słuszne to rys. 31b oraz 38e, gdy zmienne odnoszą się do tej samej jednostki jednego mola. Uważam, że bardzo dobrą ilustracją jest rys. 32a,b.

Ciekawy jestem czy z tym się Doktorantka zgadza.

Drugi niezmiernie ważny element to wydajność ligandu LE. Wydajność ligandu jest szeroko stosowanym parametrem projektowym w odkrywaniu leków. Wykorzystując ujęcie źródłowe tegorocznej pracy [85] (Peter W. Kenny, J. Cheminform. 11:8(2019)1-18) jest to analog:

$$\Delta g = - \frac{\Delta G^\emptyset}{HAC} \propto LE$$

który można porównać z wyrażeniem

$$\Delta G^\emptyset = + RT \ln(K_d/C^\emptyset)$$

gdzie: $\Delta G^\emptyset$ – standardowa energia swobodna Gibbsa wiązania liganda (lub inhibitorów) [J/mol], HAC - liczba atomów ciężkich, K_d - stała równowagi

dysocjacji kompleksu (asocjacja $\leftrightarrow$ dysocjacja), $C^\emptyset$ – równowagowe stężenie protein [mol], który został przekształcony przez zastąpienie ΔG aktywnością biologiczną wprost proporcjonalną do wyrażenia pIC_{50} dla $T = 300\text{ K} = \text{const}$ (wzór 2.0).

Cytowany w pracy doktorskiej prof. Gerhard Klebe ale w innej pracy (Nature Reviews, **14**(2015)95-110) wykorzystując wyniki z pracy Olssona *et al* (Olsson TSG, Williams MA, Pitt WR, Ladbury JE, J. Mol. Biol., **384**(2008)1002-1017) potwierdza, że w ostatnich latach rośnie zainteresowanie stosowaniem funkcji termodynamicznych dla profilowania interakcji ligand-receptor. Celem jest wybór optymalnych ligandów, które mogą stać się atrakcyjnymi lekami o pożądanych właściwościach fizykochemicznych. W tym celu wykorzystuje się układ współrzędnych EEC, czyli $\Delta H = T\Delta S + \Delta G$, dla stałych izolinii $\Delta G < 0$, wskazując że korzystniejsze są modyfikacje wywołane czynnikami entropowymi.

Tak więc ilustracja przedstawiona na rys. 42 b wraz ze wzrostem HAC sugeruje, że w granicznym przypadku energia swobodna (wielkość molowa) asymptotycznie zmierza do zera, ale w dosyć szerokim pasmie ujemnych wartości, co może mieć związek z ograniczeniem entropii konformacyjnej, o czym mowa na str. 54. Zgodnie z wcześniejszymi danymi przedstawionymi na rys. 41 a, b uwaga wydaje się być trafna i wszystkie elementy zilustrowane na rys. 41,42 są ze sobą spójne. Skala osi rzędnych na rys. 42 b jest ujemna, co może w pierwszej chwili może przesłaniać komplementarność wyводу i konieczna jest myślowa inwersja układu.

A więc kolejne pytanie, czy bifurkacje w skali MW bin dla dużych mas cząsteczkowych nie są wynikiem zmienności deskryptorów z powodów zróżnicowania termodynamicznego?

Można więc potwierdzić, że wykorzystane w pracy doktorskiej LE czy LLE są analogami ceny w [\$/g].

Ocena pracy doktorskiej

Praca jest bardzo dobrze zorganizowana, część literaturowa zaczyna się od wybranych zagadnień projektowania leków *in silico* i w dalszej części szczegółowo omawia problematykę deskryptorów molekularnych. W nawiązaniu do grupy rozważań QSAR rozważa indywidualne efekty fizykochemiczne poprzez analizę stosownych równań. W moim przekonaniu przyjęta została konwencja dr Karola Kamela z tym, że dodano szereg nowych elementów w tym zakresie, a także prawo Erooma i wiek leku.

Część inicjującą „Badania własne” nie wiadomo dlaczego zaopatrzone w nową numerację zaczynającą się od 1, jest kompozycją prac literaturowych własnych i obcych. W zasadzie jest to do zaakceptowania, chociaż trudno znaleźć wyraźną

granicę bo obserwujemy bardzo przejrzyste przejście do zagadnień podstawowych, dalej do zagadnień warsztatowych (binowanie, y-randomizacja, walidacja krzyżowa, omówienie big data Abamachem). Wykorzystane techniki świadczą o bardzo dojrzałym i świadomym operowaniu wymienionymi procedurami.

Punkty 6 do 11 stanowią zwarty i nie przeładowany raport w zakresie merytorycznym, zakończony „Podsumowaniem i wnioskami” które są zarówno bardzo wyważone, interesujące a nawet o zabarwieniu heurystycznym (np. zasada azotowa). Słowa na str. 97 w brzmieniu o niepotrzebnie „przesadnie racjonalnym” podejściu, znakomicie wplatają się w badany sens oryginalnej i nawet unikatowej problematyki. Wniosek o płatności za indywidua chemii organicznej za ilość materii jest z pewnością nowym spojrzeniem na relacje QSPR (właściwości), QSAR (aktywność) względem struktury tych związków.

Bardzo dobrze, że w dalszej części pracy nazwanej część eksperymentalna, przedstawiono szczegóły działań warsztatowo-technicznych, przez co tekst podstawowy zyskał na zwartości i klarowności, w sumie jednak bardzo żmudnej w jej analizowaniu, przede wszystkim z powodu operowania nie tylko ogromną ilością danych (w ilości ponad 2,2 mln indywiduów) ale poprzez częste stosowanie skrótów.

Uwagi techniczne

1. W wykazie skrótów nie podano LE, LLE czy PLS,PCA.
2. Rys.7 –literówka (bilogia, w publikacji [109] poprawnie).
3. Str.48⁶ – tekst niejasny.
4. Str.49/50 – tekst nie jest zrozumiały.
5. Str.53, rys. 23a powtarza się na str.90 jako rys. 42a.
6. Począwszy od str.57, w pracy stosuje się pojęcia współczynnik korelacji (r), determinacji (r^2), R (str.71) i R^{bin} , ten ostatni nie jest wyjaśniony w sposób jednoznaczny, podobnie jak przedstawiono na str.129 przy omawianiu cross validation.
7. Rys.31 powtarza się jako rys. 38 d,e.
8. Str.85₅ – ciężar pozostał jako kategoria fizyczna, bowiem MW jest tu stosunkiem masy do ilości materii.
9. Str.88₂ – wzór (2.3) opis symboli do równania nie jest jasny.
- 10.Str.89₁ – literówka.
- 11.Str.90 – podpisy pod rys.42 nie są zgodne z symbolami osi.
- 12.Str.116 [65], literówka.
- 13.Str.116 [67], niepełne cytowanie, bowiem dane zmieniają się w czasie, obecnie dostępny jest 2010 rok.

Usterki nie mają wpływu na jakość pracy, są mało istotne i nie oczekuję ustosunkowania się do nich.

Wnioski końcowe

Praca jest absolutnie nowatorska i nieoczekiwanie wskazuje na ogromnie silne związki w obszarach QSAR, QSPR z łącznikiem ekonomicznym, jakim jest QSER. Zawiera mnogość przeanalizowanych deskryptorów molekularnych w odniesieniu do ogromnie dużej populacji indywidualnych związków organicznych, wykorzystując czasochłonne sondowania i transformowanie danych pod określone cele. Stawiam ją w rzędzie prac na zaskakująco wysokim poziomie.

Jak na dysertację zawierającą ponad 150 stron nie znajduje się ani dłużyzn ani przewartościowań, bowiem jest bardzo dobrze zorganizowana i w najważniejszej części opublikowana w renomowanych czasopismach.

Praca doktorska wypełnia znamiona ustawy z dnia 21.04.2017 o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. z dn.28.04.2017, poz.859).

Stawiam wniosek aby dopuścić Panią **mgr Urszulę KUCIA** do dalszych etapów przewodu doktorskiego.

W przypadku pozytywnego rozstrzygnięcia, proponuję wyróżnienie dysertacji za bardzo innowacyjny charakter i zaskakująco logiczne i przejrzyste powiązanie domeny chemicznej z ekonomiczną.

