

Nauka komputera reguł reaktywności i strategii w automatycznym planowaniu retrosyntetycznym

Sara Szymkuć

Promotor: Prof. Bartosz A. Grzybowski

Nauka komputera projektowania ścieżek syntezy związków organicznych stanowi jedno z najstarszych wyzwań chemii obliczeniowej. Pierwszy program podejmujący się rozwiązania tego problemu został opracowany już w latach 60. XX wieku. Wiele innych programów było rozwijanych na przełomie lat 80. i 90. lecz żaden z nich nie sprostał wymaganiom chemików organicznych, co przyczyniło się do utraty zainteresowania tą dziedziną nauki w latach 2000. Było to niezwykle niefortunne, gdyż w międzyczasie komputery opanowały inne umiejętności analityczne, dotychczas uważane wyłącznie za domenę ludzkiego intelektu oraz kreatywności np. symboliczne rozwiązywanie złożonych równań różniczkowych (Mathematica) czy opanowanie gier strategicznych (szachy, Go) na poziomie przewyższającym ludzkich mistrzów. W mojej pracy doktorskiej, czerpałam inspirację (oraz nadzieję) z tych dokonań, przez kilka lat rozwijając platformę obliczeniową *Chematica*, która planuje ścieżki syntezy związków organicznych.

Pierwszym krokiem w nauce komputera chemii było przyjęcie właściwego formatu danych dla reakcji chemicznych oraz cząsteczek w postaci zrozumiałej dla maszyny. Alfanumeryczna notacja SMILES/SMARTS została wybrana ze względu na szybkość wykonywania operacji pojedynczych reakcji chemicznych oraz możliwość inkorporacji szczegółowych informacji na temat stereochemii.

Kolejnym etapem była nauka komputera reakcji organicznych. Dla każdej klasy transformacji, szczegółowo badałam jej mechanizm i dokładnie wyznaczałam rdzeń na który składały się motywy strukturalne wraz z dopuszczalnymi podstawnikami. Każda reakcja zawiera także informację kontekstową opisującą niekompatybilne bądź wymagające protekcji grupy funkcyjne znajdujące się poza rdzeniem oraz typowe warunki dla danej transformacji. Sformalizowana przeze mnie struktura bazy danych oraz reguły reakcji stanowią podwalinę wiedzy chemicznej programu *Chematica* na którą obecnie składa się 60 tysięcy reakcji, z których osobiście zakodowałam około 15 tysięcy.

Nadrzędnym celem mojej pracy była nauka komputera samodzielnego projektowania ścieżek syntezy. Kluczową rolę odgrywało zdefiniowanie funkcji oceny pozwalającej algorytmowi wyszukiwania na rozpoznanie czy porusza się we właściwym kierunku oraz jak daleko znajduje się od substratów. Zaproponowałam dualną funkcję, oceniającą zarówno skomplikowanie cząsteczkowe jak i szansę realizacji poszczególnych reakcji.

Niestety, nawet prawidłowe, pojedyncze reakcje nie gwarantują jeszcze utworzenia rozsądnej ścieżki syntezy. Mając świadomość tego problemu, skoncentrowałam się na identyfikacji najbardziej obiecujących sekwencji reakcji oraz eliminacji jałowych kombinacji.

Niezależnie od nowatorskości idei kryjącej się za programem retrosyntetycznym, jego użyteczność powinna zostać zweryfikowana. Początkowo przeprowadziłam „walidację na papierze”, w której program niezależnie odtworzył opublikowane ścieżki syntezy dla wybranych związków organicznych. Po tym teście przyszedł czas na weryfikację proponowanych przez program syntez w laboratorium. Osiem ścieżek syntetycznych zaprojektowanych przez *Chematicę* zostało wykonanych przez chemików z firmy Sigma-Aldrich, Uniwersytetu Northwestern oraz ICHO PAN. Wszystkie zakończyły się sukcesem dokumentując pierwszą znaną walidację eksperymentalną programu retrosyntetycznego.

Następnie moje zainteresowania naukowe skierowały się w stronę chemii systemów. W tym duchu ostatnia część rozprawy doktorskiej opisuje komputerowe odkrywanie cykli reakcji chemicznych, m.in. wzorujących się na cyklach biologicznych i potencjalnie użytecznych jako metoda recyklingu katalizatorów czy autoamplifikacji użytecznych związków.