

Algorytmy i modele statystyczne do przewidywania schematu fragmentacji cząsteczek podczas dysocjacji wywołanej kolizją

Spektrometria mas jest metodą eksperymentalną stosowaną w chemii do pomiaru masy cząsteczek. Jest to bardzo przydatne narzędzie o bardzo szerokim zastosowaniu w wielu dziedzinach, od rolnictwa, przez farmację i medycynę po kryminalistykę. Pomaga w identyfikacji i charakteryzowaniu nieznanymi związków (białek lub mniejszych, takich jak metabolity), opracowywaniu i projektowaniu nowych związków, takich jak leki. Tandemowa spektrometria mas jest jedną z typowych konfiguracji eksperymentów spektrometrycznych, która składa się z dwóch spektrometrów połączonych jeden po drugim. W pierwszym spektrometrze zjonizowane cząsteczki są rozdzielane na podstawie ich stosunku masy do ładunku (m/z), a następnie wykrywane. W drugim jony o wybranych wartościach m/z są fragmentowane, a ich fragmenty są wykrywane. Popularną techniką fragmentacji jest dysocjacja wywołana kolizją (CID). Polega na przyspieszaniu cząsteczek do wyższej energii kinetycznej, a następnie zderzaniu ich z niskoenergetycznymi, obojętnymi atomami gazów szlachetnych, takimi jak hel. Badanie schematów fragmentacji jest interesującym tematem badawczym, ponieważ pomaga w lepszym zrozumieniu struktur cząsteczek. Chociaż fragmentacja CID jest dobrze scharakteryzowana dla białek, wciąż w przypadku metabolitów i innych małych cząsteczek stanowi wyzwanie badawcze. Jest to spowodowane ich skomplikowaną, nieliniową strukturą, w przeciwieństwie do regularnej, liniowej struktury białek. Jednym z ważnych problemów spektrometrii mas, który nadal wymaga dalszych badań, jest automatyczne wyjaśnienie mechanizmu fragmentacji dla danej cząsteczki. Rozwiązanie tego problemu wymaga zastosowania zaawansowanego algorytmicznego i statystycznego podejścia ze względu na wspomnianą wcześniej złożoność.

Głównym celem tego projektu jest program komputerowy, który przewiduje schemat fragmentacji dla danej struktury chemicznej prekursora i jego widma. Z jednej strony, program powinien zaproponować najbardziej realny schemat fragmentacji, z drugiej strony powinien wyjaśniać dane widmo jak najlepiej. Narzędzie będzie mogło być używane w codziennej pracy laboratoryjnej dzięki łatwo zrozumiałym komunikatom i prostemu interfejsowi.

Proponujemy kombinatoryczne rozwiązanie tego problemu, które polega na symulacji in-silico wykonalnych reakcji fragmentacji i wyborze tych, które optymalnie wyjaśniają widmo masowe. Cała przestrzeń wszystkich możliwych reakcji jest zbyt duża, aby mogła być symulowana przez typowy komputer. Używamy technik algorytmicznych, takich jak kolejki priorytetowe i funkcje oceny heurystycznej, aby symulować najbardziej prawdopodobne reakcje. Algorytmy są inspirowane przez te używane w sztucznej inteligencji. Optymalny mechanizm reakcji jest następnie wybierany z symulowanych reakcji, aby wyjaśnić jak najwięcej widma w najprostszy możliwy sposób.

Naszym celem jest przeanalizowanie dość szerokiego zakresu możliwych fragmentacji przy użyciu kilku prostych zasad. Z tego powodu opisujemy reakcję jako sekwencję prostych wzorców reakcji: podziały wiązań, przegrupowanie wiązań wielokrotnych, czy połączenia fragmentów. Nawet jeśli połączenia nie występują w rzeczywistych procesach fragmentacji, są one wykorzystywane do symulacji złożonych reakcji, które obejmują kilka atomów z różnych części cząsteczki, takich jak przegrupowanie McLafferty.

Wstępna implementacja prototypu narzędzia programowego pokazuje, że nasze podejście daje bardziej wykonalną fragmentację w porównaniu z podejściem znanym z innych popularnych narzędzi. Włączenie zestawu podstawowych reakcji pozwala na lepsze wyjaśnienie widma i lepiej odzwierciedla rzeczywiste procesy fragmentacji. Nadal jednak istnieje szeroka przestrzeń do opracowywania algorytmów i narzędzi, które proponują dobre wyjaśnienie widma z rozsądną złożonością obliczeniową.