

Wyznaczanie dokładnej obwiedni izotopowej peptydu.

Wyznaczanie postaci funkcji $f^p(m/z)$, odpowiedzialnej za opis kształtu obwiedni izotopowej, odbywa się przy użyciu algorytmu, opartego na splocie w dziedzinie masy. W swej podstawowej wersji generuje on rozkład pików sumarycznych, powstających przez połączenie wszystkich pików pochodzących od odmian izotopowych o jednakowej całkowitej liczbie nukleonów. Tym samym jest on powiązany z algorytmem Kubinyi oraz z jedną z odmian fourierowskiego algorytmu Rockwooda. Istotną różnicą wobec tych algorytmów jest jednak fakt, że wyznaczone są dokładne masy pików sumarycznych, rozumiane jako średnie mas tworzących je odmian izotopowych, ważone przez prawdopodobieństwa ich występowania. Efektem działania rozszerzonej wersji algorytmu mogą być dokładne masy, prawdopodobieństwa wystąpień oraz, co jest unikalną cechą, składy izotopowe wszystkich odmian tworzących piki sumaryczne.

Wersja podstawowa algorytmu

Prawdopodobieństwa występowania poszczególnych pików sumarycznych cząstki reprezentowane są przez elementy wektora \mathbf{p} . Wyznaczanie prawdopodobieństw odbywa się w sposób iteracyjny, będący odpowiednikiem budowania cząstki z grup atomów kolejnych pierwiastków, aż do momentu uzyskania docelowego składu chemicznego. Każdemu łączeniu odpowiada splot aktualnego wektora prawdopodobieństw cząstki \mathbf{p} z wektorem prawdopodobieństw przyłączanej grupy atomów aktualnie rozpatrywanego pierwiastka \mathbf{p}_E . Atomy tego samego pierwiastka dodawane są do cząstki w grupach o licznosciach będących potęgami liczby 2, przy czym wektor prawdopodobieństw grupy atomów o licznosci 2^{n+1} ($n = 0, 1, \dots$) jest wyznaczany na podstawie wektorów prawdopodobieństw grup atomów o licznosci 2^n . Rozwiązanie to, wzorowane na algorytmie Kubinyi, pozwala ograniczyć liczbę koniecznych mnożeń. Dalsze zwiększenie wydajności można osiągnąć przez zastosowanie tablic z obliczonymi wcześniej wartościami wektorów prawdopodobieństw grup atomów.

Na początku działania algorytmu wektor \mathbf{p} zawiera jeden element o wartości równej 1, podczas gdy wektor \mathbf{p}_E zawiera prawdopodobieństwa wystąpień izotopów pierwszego z rozpatrywanych pierwiastków. Proces dodawania atomów tego pierwiastka sterowany jest poprzez binarną reprezentację ich liczby. Pojawienie się w niej wartości 1 na n -tej pozycji oznacza konieczność przyłączenia grupy 2^n atomów. Wiąże się to z wyznaczeniem nowego wektora prawdopodobieństw cząstki \mathbf{p}' , którego k -ty element dany będzie wzorem:

$$\mathbf{p}'[k] = \sum_i \mathbf{p}[i] \mathbf{p}_E[k-i]$$

Wektor prawdopodobieństw pierwiastka \mathbf{p}_E uaktualniany jest w każdej iteracji poprzez splot

z samym sobą. Po uwzględnieniu wszystkich atomów pierwszego pierwiastka proces jest kontynuowany dla kolejnych pierwiastków, aż do osiągnięcia docelowego składu cząstki.

Aby umożliwić wyznaczenie mas pików sumarycznych, reprezentacja rozkładu izotopowego wzbogacona jest o wektor masy \mathbf{M} . Jego uaktualnianie na podstawie wektorów mas \mathbf{M}_E grup przyłączanych atomów odbywa się jednocześnie z liczeniem splotu prawdopodobieństw, przy czym k -ty element nowego wektora \mathbf{M}' dany jest zależnością:

$$\mathbf{M}'[k] = \sum_i \mathbf{p}[i] \mathbf{p}_E[k-i] (\mathbf{M}[i] + \mathbf{M}_E[k-i])$$

Po zakończeniu obliczeń każdy element wektora mas \mathbf{M} dzielony jest przez odpowiadający mu element wektora prawdopodobieństw \mathbf{p} . W rezultacie masy pików stają się średnimi ważonymi mas wszystkich odmian izotopowych, których prawdopodobieństwa składały się na dany pik sumaryczny, przy czym wagami są prawdopodobieństwa wystąpienia poszczególnych odmian. Po każdym kroku splatania rozmiar nowych wektorów prawdopodobieństw \mathbf{p} i mas \mathbf{M} będzie równy powiększonej o jeden sumie rozmiarów wektorów prawdopodobieństw łączonych fragmentów. W celu zwiększenia szybkości działania, po każdym splocie może być wykonywane przycinanie. W odróżnieniu od algorytmów wielomianowych, dotyczy ono jednak jedynie pików sumarycznych znajdujących się na krańcach rozkładu, dzięki czemu stosowana może być bardzo niska wartość progu.

Po określeniu mas i prawdopodobieństw występowania odmian izotopowych peptydu, możliwe jest wyznaczenie pełnej postaci obwiedni izotopowej odpowiadającego mu jonu molekularnego. W pierwszej kolejności konieczne jest zamienienie mas każdej z odmian izotopowych na wartości m/z wynikające ze stopnia naładowania jonu, a następnie wykonanie splotu z funkcją modelującą kształt pików. Ze względu na rodzaj spektrometru używanego trakcie opracowania prezentowanej metody, udział w całej obwiedni k -tego pików sumarycznego o wartości m/z równej m_k i prawdopodobieństwie wystąpienia p_k opisywany jest funkcją Gaussa:

$$f_k(m/z) = p_k \exp\left(-\frac{(m/z - m_k)^2}{2\sigma_k^2}\right),$$

przy czym wartość parametru σ_k , decydującego o szerokości pików, związana jest z położeniem jego maksimum m_k i rozdzielczością spektrometru RP następującą zależnością:

$$\sigma_k = \frac{m_k}{RP \sqrt{2 \ln(2)}}.$$

Całkowita obwiednia $f^P(m/z)$ jest sumą funkcji kształtów pierwszych K pików sumarycznych, których prawdopodobieństwa wystąpienia są większe od zadanego progu.

Wersja rozszerzona algorytmu

Rozszerzona wersja algorytmu umożliwia również wyznaczenie składu (a tym samym – dokładnej masy i prawdopodobieństwa wystąpienia) wszystkich odmian izotopowych tworzących piki sumaryczne.

Każda odmiana izotopowa jest reprezentowana w postaci wektora składu, którego elementy K_{ij} oznaczają liczby atomów i -tego pierwiastka znajdujących się w j -tym stanie izotopowym:

$$[K_{11}, K_{12}, \dots, K_{1I_1}; K_{21}, K_{22}, \dots, K_{2I_2}; \dots] \quad (8.1)$$

W obecnej implementacji wektor ten uwzględnia izotopy stabilne wodoru (^1H , ^2H), węgla (^{12}C , ^{13}C), azotu (^{14}N , ^{15}N), tlenu (^{16}O , ^{17}O , ^{18}O) i siarki (^{32}S , ^{33}S , ^{34}S , ^{35}S), czyli pierwiastków wchodzących w skład aminokwasów tworzących białka oraz dodatkowo często występujący w modyfikacjach potranslacyjnych fosfor (^{31}P). Tak więc najlżejsza odmiana izotopowa cząstki o składzie chemicznym $\text{C}_{254}\text{H}_{377}\text{N}_{65}\text{O}_{75}\text{S}_6$ reprezentowana jest 14-elementowym wektorem o postaci $[254, 0; 377, 0; 65, 0; 75, 0, 0; 0; 6, 0, 0, 0]$.

Dysponując składem chemicznym odmiany izotopowej, jej masę M_k oraz prawdopodobieństwo p_k wystąpienia można wyznaczyć z zależności:

$$M_k = K_{11}M_{11} + \dots + K_{1I_1}M_{1I_1} + K_{21}M_{21} + \dots + K_{1I_2}m_{1I_2} + \dots \quad (8.2)$$

$$p_k = \left(\frac{N_1!}{K_{11}! \dots K_{1I_1}!} p_{11}^{K_{11}} \dots p_{1I_1}^{K_{1I_1}} \right) \left(\frac{N_2!}{K_{21}! \dots K_{2I_2}!} p_{21}^{K_{21}} \dots p_{2I_2}^{K_{2I_2}} \right) \dots \quad (8.3)$$

W celu wyznaczenia dokładnego rozkładu izotopowego cząstki, jej odmiany grupowane są, zgodnie z całkowitą liczbą nukleonów, w listy powiązane z odpowiednimi pikami sumarycznymi. Lista odmian izotopowych k -tego piku sumarycznego jest aktualizowana w trakcie splotu podczas wyznaczania wartości prawdopodobieństwa jego występowania (por. wzór 5.15 w rozdziale 5.3.2), przy czym operacja mnożenia zastępowana jest tutaj utworzeniem nowej odmiany izotopowej na podstawie dwóch istniejących, poprzez zsumowanie ich wektorów składu, a operacji sumowania odpowiada dodanie do listy piku nowej odmiany izotopowej. Przetwarzaniu w następnej iteracji poddawane są jedynie listy związane z pikami sumarycznymi, które nie zostały wyeliminowane w kroku przycinania. Zastosowanie progu prawdopodobieństwa pików sumarycznych jest w tym przypadku konieczne, gdyż bez niego algorytm staje się odpowiednikiem algorytmu wielomianowego. Należy jednak zwrócić uwagę, że przycinanie dotyczy pików sumarycznych, a nie pików poszczególnych odmian izotopowych. W efekcie usuwane są wszystkie odmiany izotopowe związane z nieobserwowalnymi pikami sumarycznymi o znikomym prawdopodobieństwie, podczas gdy listy odmian izotopowych najbardziej nas interesujących głównych pików pozostają kompletne.