

Algorytm wyszukiwania widm mas jonów peptydowych.

Ze względu na zdecydowanie większą rozdzielczość i powtarzalność pomiarów spektrometru mas w porównaniu z systemem HPLC, wyszukanie peptydu w praktyce sprowadza się do wyznaczenia pozycji widma reprezentującego go jonu wzdłuż osi czasu retencji. Przez właściwy czas retencji rozumiany jest taki, dla którego funkcja opisująca obwiednię izotopową peptydu $f^P(m)$ wykazuje najlepsze dopasowanie do danych pomiarowych.

Algorytm wyszukiwania składa się z dwóch etapów: w pierwszej kolejności identyfikowane są wszystkie możliwe położenia widma peptydu wzdłuż osi czasu retencji, z których następnie wybierane jest to, które charakteryzuje się najlepszym dopasowaniem. Pierwszy etap oparty jest na wykorzystaniu pewnej funkcji $c^P(t)$, która w idealnym przypadku przyjmuje wartości niezerowe jedynie dla poprawnego czasu retencji widma poszukiwanego peptydu. W proponowanym algorytmie używana jest w tym celu funkcja będąca średnią geometryczną chromatogramów $c_i^P(t)$ dla wartości m/z odpowiadających N_C najwyższym pikom obwiedni izotopowej peptydu. Dodatkowo uwzględniany jest fakt, że położenie poszukiwanego widma wzdłuż osi czasu retencji nie jest dowolne i nie powinno znacząco odbiegać od teoretycznego czasu zejścia z kolumny chromatograficznej określanego przez parametr t_r^P modelu. Dlatego wartości chromatogramu są modyfikowane przez funkcję kary w postaci gaussoidy o maksimum dla czasu t_r^P i szerokości wynikającej z wartości parametru σ_C . Ostateczna postać funkcji $c_i^P(t)$ dana jest zależnością:

$$c^P(t) = \exp\left(-\frac{(t-t_r^P)^2}{2\sigma_C^2}\right) \sqrt[N_C]{\prod_{i=1}^{N_C} c_i^P(t)} .$$

W każdym skanie wartość chromatogramu cząstkowego $c_i^P(t)$ wyznaczana jest jako maksymalna wysokość widma w przedziale wartości m/z $\langle m_i^P - \Delta m_C; m_i^P + \Delta m_C \rangle$, gdzie m_i^P jest położeniem i -tego, pod względem teoretycznej wysokości, pików obwiedni, a bezwzględna tolerancja Δm_C określana jest na podstawie wyrażonej w jednostkach ppm względnej tolerancji δm_C , będącej parametrem przeszukiwania:

$$\Delta m_C = 10^{-6} \delta m_C m_i^P .$$

Oznacza to, że wraz ze wzrostem wartości m/z zwiększana jest tolerancja wyznaczenia chromatogramu, co jest zgodne ze spadkiem dokładności wzdłuż osi m/z . Dodatkowym wymogiem jest występowanie maksimum lokalnego widma w rozpatrywanym przedziale wartości m/z .

Po wyznaczeniu ostatecznej postaci całkowitego chromatogramu poddawany jest on wygładzeniu przy użyciu filtru Savitzkiego-Golaya, po którym następuje detekcja pików. Wykryte w chromatogramie piki uznawane są za potencjalne miejsca występowania poszukiwanego widma

jonu peptydowego i w obszarze każdego z nich podejmowana jest próba dopasowania teoretycznego modelu do danych eksperymentalnych.

Dopasowanie funkcji opisującej kształt obwiedni izotopowej następuje w skanie, w którym występowało maksimum pików chromatogramu. Uwzględniane są przy tym pochodzące z tego skanu wartości widma w_j odpowiadające punktom m_j ($j = \{1, \dots, N_w\}$) znajdującym się w zakresie m/z teoretycznej obwiedni izotopowej. Wyznaczenie modelującej dane eksperymentalne funkcji $f^E(m/z)$ odbywa się metodą najmniejszych kwadratów, przy użyciu iteracyjnego algorytmu Levenberga–Marquardta. Minimalizacji poddawane jest wyrażenie:

$$S(\boldsymbol{\beta}) = \sum_{j=1}^{N_w} [w_j - f^E(m_j | \boldsymbol{\beta})]^2 ,$$

gdzie $\boldsymbol{\beta}$ jest wektorem parametrów opisujących położenia, wysokości i szerokości połówek dla K uwzględnianych pików obwiedni. Początkowe wartości parametrów określone są przez model teoretyczny widma jonu. W wypadku wykrycia słabego dopasowania, proces może być iteracyjnie powtarzany ze zmodyfikowanymi początkowymi wartościami parametrów. Jako miara jakości dopasowania do danych eksperymentalnych stosowany jest współczynnik determinacji (*coefficient of determination*):

$$R_w^2 = 1 - \frac{\sum_{j=1}^{N_w} (w_j - f_j^E)^2}{\sum_{j=1}^{N_w} (w_j - \bar{w})^2} ,$$

gdzie f_j^E są wartościami funkcji $f^E(m/z)$ w punktach m_j , a \bar{w} jest średnią arytmetyczną wartości widma. Sama wartość współczynnika R_w^2 nie jest jednak wystarczającym wskaźnikiem jakości dopasowania: z jednej strony nie jest ona czuła na odstępstwa od początkowych oszacowań położenia i wysokości pików, a z drugiej może być degradowana przez występujące w rozpatrywanym obszarze widma piki nie należące do poszukiwanej obwiedni. Dlatego też wyznaczana jest również miara dopasowania do idealnej obwiedni izotopowej:

$$R_p^2 = 1 - \frac{\sum_{j=1}^{N_w} (f_j^{SP} - f_j^E)^2}{\sum_{j=1}^{N_w} (f_j^{SP} - \bar{f}^P)^2} ,$$

gdzie f_j^{SP} są wartościami funkcji $f^{SP}(m/z)$, która powstaje przez takie przeskalowanie i przesunięcie teoretycznej obwiedni $f^P(m/z)$, aby jej pik monoizotopowy miał położenie i wysokość zgodne z wartościami określonymi przez funkcję $f^E(m/z)$. Wartość R_p^2 jest podstawą wyboru najlepszego spośród wszystkich rozpatrywanych pików chromatogramu.

Dopasowanie danego przez funkcję $g^P(t)$ kształtu przekroju widma w kierunku osi czasu retencji odbywa się na podstawie chromatogramu wykonanego dla najwyższego pików obwiedni i

przebiega w sposób analogiczny do opisanego dla kierunku m/z .

Należy zwrócić uwagę, że stanowiące podstawę algorytmu wyszukiwania modele widm tworzone są w oparciu o skład peptydów oraz parametry czasowe przebiegów LC-MS/MS i tym samym nie mogą z góry uwzględniać błędów systematycznych związanych z pomiarami widm LC-MS. Dlatego też lepsze rezultaty wyszukiwania można osiągnąć dzięki dostosowaniu zarówno samych modeli, jak i parametrów wyszukiwania do aktualnie analizowanego widma. W tym celu opisany powyżej proces wyszukiwania powtarzany jest dwukrotnie, przy czym pierwsze powtórzenie służy do wyznaczenia parametrów korygujących ewentualne różnice w kalibracji spektrometru mas czy gradiencie chromatograficznym, które następnie są uwzględniane w drugim powtórzeniu. W pierwszym powtórzeniu stosowane są szerokie przedziały tolerancji dla wartości m/z i czasów retencji, którym jednak towarzyszą ostre wymagania dotyczące jakości dopasowania do teoretycznej obwiedni. Dla znalezionych widm określone są względne odstępstwa wobec teoretycznego położenia w kierunku osi m/z i czasu retencji. Na podstawie wartości błędów wyznaczane są parametry krzywych kalibracyjnych korygujących teoretyczne położenia w modelach oraz nowe, zawężone wartości parametrów δm_c i σ_c związanych z zakresami tolerancji chromatogramów cząstkowych i funkcji kary. Stosowana przy ich wyznaczeniu metoda jest taka sama jak ta zaprezentowana w podczas rekaliibracji widm fragmentacyjnych.