

## Algorytm kalibracji widm MS/MS.

Wyznaczenie parametrów kalibracji dla jonów macierzystych odbywa się na podstawie wektora  $e$ , którego elementami są błędy względne pomiaru wartości  $m/z$ , wyrażone w jednostkach ppm. Dla  $i$ -tego jonu błąd ten dany jest zależnością:

$$e_i = 10^6 \left( \frac{m_i^E - m_i}{m_i} \right),$$

gdzie  $m_i$  i  $m_i^E$  to, odpowiednio, rzeczywista i zmierzona wartość  $m/z$ . Rzeczywista wartość  $m/z$  wyznaczana jest na podstawie stopnia naładowania jonu i przypisanej mu przez system Mascot sekwencji aminokwasowej. Zakładana jest liniowa zależność błędu względnego od wartości  $m/z$ . Parametry prostej kalibracyjnej wyznaczone są metodą najmniejszych kwadratów, przy czym robione jest to w sposób zapewniający odporność na wartości odstające, wynikające z możliwych nieprawidłowych identyfikacji sekwencji dla części uwzględnionych peptydów. W tym celu wykorzystywany jest iteracyjny algorytm RANSAC (*RANdom SAmple Consensus*), który wyznacza punkty najbardziej odbiegające od optymalizowanego modelu na drodze powtarzanych w sposób iteracyjny losowych podziałów zbioru danych. Przy znanych parametrach prostej kalibracyjnej  $a$  i  $b$  możliwe jest wyznaczenie skorygowanej wartości  $m/z$  dla dowolnego jonu jako:

$$m_{CAL} = m [1 - 10^{-6}(am + b)] .$$

Kalibracja widm MS/MS wykonywana jest w taki sam sposób jak to ma miejsce dla jonów macierzystych, z tą jednak różnicą, że aby wyznaczyć wektor błędów  $e$  konieczne jest wygenerowanie teoretycznych widm fragmentacyjnych dla wszystkich uwzględnianych peptydów. Dokładniej rzecz ujmując, wyliczane są jedynie teoretyczne pozycje pików reprezentujących jednokrotnie naładowane jony z serii  $y$ , które zwykle najłatwiej zidentyfikować w eksperymentalnym widmie fragmentacyjnym. Określone na ich podstawie parametry kalibracyjne są następnie używane dla wszystkich pozostałych pików widm.

Skorygowane widma fragmentacyjne zapisywane są w formacie umożliwiającym import do Mascota. Stosowane w powtórnym przeszukaniu zakresy tolerancji wyznaczone są jako wielokrotność odchylenia standardowego estymowanego na podstawie odchylenia medianowego (*MAD – Median Absolute Deviation*):

$$MMD = K \cdot \text{median}(|E - \text{median}(E)|) / 0,6745 ,$$

gdzie  $E$  oznacza wektor błędów względnych określenia masy po wykonaniu kalibracji (uwzględniane są jedynie jony, które nie zostały uznane za skrajne przez algorytm RANSAC na etapie wyznaczania parametrów funkcji kalibracyjnej), natomiast parametr  $K$  domyślnie przyjmuje wartość 3.

Z nie do końca zrozumiałych przyczyn, dla jonów fragmentacyjnych Mascot dopuszcza jedynie użycie bezwzględnego błędu określenia masy. Dlatego też w ich przypadku konieczne jest ponowne przeliczenie tolerancji względnej (w jednostkach ppm) na bezwzględną (wyrażoną w Da), co niestety prowadzi do niepożądanego spadku dokładności w zakresie małych  $m/z$ .