Algorytm predykcji czasów retencji peptydów.

Metoda predykcji czasów elucji oparta na jest modelu zjawiska retencji, którego parametry poddawane są optymalizacji przy użyciu algorytmu ewolucyjnego. Podstawą modelu jest zestaw dwudziestu współczynników retencji *Rc*, reprezentujących niezależną od pozycji w sekwencji hydrofobowość poszczególnych reszt aminokwasowych. Wyznaczona przez ich sumowanie hydrofobowość całkowita jest jednak następnie modyfikowana w sposób zależny od sekwencji, ze szczególnym uwzględnieniem jej N-końcowego fragmentu.

Wpływ dodatniego ładunku N-końca na obserwowany czas retencji uwzględniany jest w prezentowanym modelu przez wprowadzenie addytywnego czynnika korekcyjnego H_{Nt}, o wartości zależnej od N-końcowych reszt aminokwasowych. Założone zostało, że dodatni ładunek grupy aminowej powoduje osłabienie naturalnej hydrofobowości lub hydrofilowości znajdujących się w jego sąsiedztwie reszt aminokwasowych. Jego wpływ rozciąga się na reszty zajmujące w sekwencji peptydu pozycje od 1 do L_{Nt} i maleje wykładniczo wraz z oddalaniem się od N-końca. Ostateczna postać poprawki dana jest wzorem:

$$H_{Nt} = \sum_{i=1}^{L_{Nt}} -(Rc_{(i)} - \bar{Rc})\,\alpha_{Nt} e^{-\beta_{Nt}(i-1)} \quad ,$$

gdzie $Rc_{(i)}$ jest współczynnikiem retencji reszty aminokwasowej zajmującej *i*-tą pozycję w sekwencji, a $\bar{Rc}$ jest średnią współczynników retencji wszystkich reszt aminokwasowych. Maksymalny zasięg L_{Nt} zależny jest od decydującego o szybkości spadku wartości funkcji wykładniczej współczynnika B_{Nt}: uwzględniane są pozycje, dla których wartość tej funkcji jest większa od 0,05.

Podobne, choć mające mniejszy wpływ na czas retencji zjawisko można zaobserwować dla występujących wewnątrz sekwencji peptydu zasadowych reszt aminokwasowych (argininy – R, lizyny – K, histydyny – H), których łańcuchy boczne niosą dodatnie ładunki. Charakter wpływu, jego interpretacja i sposób uwzględnienia go w modelu są w tym przypadku analogiczne jak dla N-końca, z tą tylko różnicą, że poprawka działa w sposób dwustronny. Wystąpienie na *k*-tej pozycji w sekwencji którejś z zasadowych reszt aminokwasowych skutkować będzie poprawką w postaci:

$$H_X = \sum_{i=1}^{L_X} -(Rc_{(k-i)} - \bar{Rc})\,\alpha_X e^{-\beta_X i} + \sum_{i=1}^{L_X} -(Rc_{(k+i)} - \bar{Rc})\,\alpha_X e^{-\beta_X i} \quad ,$$

gdzie *X* należy do zbioru {R, K, H}.

Dodatkowym elementem jest przedziałami liniowy multiplikatywny czynnik korekcyjny K_L, związany z długością L^P sekwencji peptydu. Jego zasadnicza postać jest zgodna z modelem Krokhina:

$$K_L = \begin{cases} 1 - a_1(10 - L^P) & \text{dla } L^P < 10 \\ 1 & \text{dla } L^P \in \langle 10;20 \rangle \\ 1 - a_2(L^P - 20) & \text{dla } L^P > 20 \end{cases},$$

z tą jednak różnicą, że nachylenia prostych a_1 i a_2 nie są wartościami z góry ustalonymi, lecz poddawane są optymalizacji.

Po wyznaczeniu wartości wszystkich czynników korekcyjnych, całkowita hydrofobowość H peptydu wyznaczana jest z zależności:

$$H = K_L \left(\sum_{i=1}^{20} N_i Rc_i + H_{Nt} + H_R + H_L + H_K \right) ,$$

gdzie N_i jest liczbą wystąpień reszty aminokwasowej o współczynniku retencji równym Rc_i. Przy znanej hydrofobowości H i danych parametrach gradientowej zmiany stężenia acetonitrylu w fazie ruchomej (nachyleniu A_g i opóźnieniu t_g), przewidywany czas zejścia peptydu z kolumny chromatograficznej może być określony na podstawie zależności:

$$t_r = A_g H + t_g ,$$

Wartości parametrów opisanego powyżej modelu wyznaczane są na podstawie zbioru sekwencji N_t peptydów i wektora $\boldsymbol{t}^S$ zawierającego średnie czasy sekwencjonowania tych peptydów, zaobserwowane w związanych z eksperymentem przebiegach LC-MS/MS. Optymalizacja parametrów realizowana jest przy użyciu algorytmu ewolucyjnego. Genotyp każdego z osobników populacji jest rzeczywistoliczbowym wektorem w postaci:

$$[Rc_1, \ldots, Rc_{20}, \alpha_{Nt}, \beta_{Nt}, \alpha_R, \beta_R, \alpha_K, \beta_K, \alpha_H, \beta_H, a_1, a_2] .$$

Kolejne pozycje wektora reprezentują optymalizowane parametry modelu, którymi są: współczynniki retencji dla każdej z reszt aminokwasowych (Rc_i, dla i = 1, .., 20), współczynniki eksponencjalnych poprawek dla N-końca ($\alpha_{Nt}, \beta_{Nt},$) i zasadowych reszt aminokwasowych ($\alpha_R, \beta_R, \alpha_L, \beta_L, \alpha_H, \beta_H$) oraz nachylenia prostych modelujących wpływ długości peptydu na hydrofobowość (a_1, a_2). Fenotyp osobnika ma postać wektora $\boldsymbol{t}^P$, którego elementami są przewidywane czasy retencji dla wszystkich peptydów ze zbioru uczącego, wyznaczone na podstawie wartości parametrów modelu tworzących jego genotyp. Wartością przystosowania osobnika jest współczynnik korelacji liniowej pomiędzy wektorami $\boldsymbol{t}^S$ i $\boldsymbol{t}^P$, czyli pomiędzy rzeczywistymi a przewidywanymi czasami retencji peptydów:

$$r(\boldsymbol{t}^S, \boldsymbol{t}^P) = \frac{\sum_{i=1}^{N_t} (t_i^S - \bar{t}^S)(t_i^P - \bar{t}^P)}{\sqrt{\sum_{i=1}^{N_t} (t_i^S - \bar{t}^S)^2} \sqrt{\sum_{i=1}^{N_t} (t_i^P - \bar{t}^P)^2}} ,$$

gdzie :

$$\bar{t}^S = \frac{1}{N_t}\sum_{i=1}^{N_t} t_i^S \quad ; \quad \bar{t}^P = \frac{1}{N_t}\sum_{i=1}^{N_t} t_i^P \quad .$$

Użyty algorytm ewolucyjny charakteryzuje się populacją o stałej liczebności. Stosowany jest schemat sukcesji elitarnej, gdyż, jak wykazały testy, pozwala to przyspieszyć osiągnięcie zbieżności. Aby jednak uniknąć łatwego osiadania w maksimach lokalnych funkcji celu, elita jest niewielka (5% wielkości populacji) i krótkożyciowa (maksymalny czas życia wynosi 3 pokolenia). Prawdopodobieństwo reprodukcji osobników jest liniowo zależne od ich rangi, ustalanej przez posortowanie całej populacji według nierosnących wartości przystosowania (w taki jednak sposób, aby osobniki o jednakowej wartości przystosowania otrzymały takie same rangi). Osobniki potomne powstają poprzez krzyżowanie równomierne i mutację o rozkładzie Cauchy'ego. Osobniki niemieszczące się w ograniczeniach przestrzeni poszukiwań są z pewnym prawdopodobieństwem naprawiane przez lustrzane odbicie wobec ograniczeń. Kryterium zatrzymania algorytmu jest nieosiągnięcie przez pewną zdefiniowaną liczbę pokoleń poprawy w stosunku do najlepszego z dotychczasowych osobników.