

High-resolution HDX

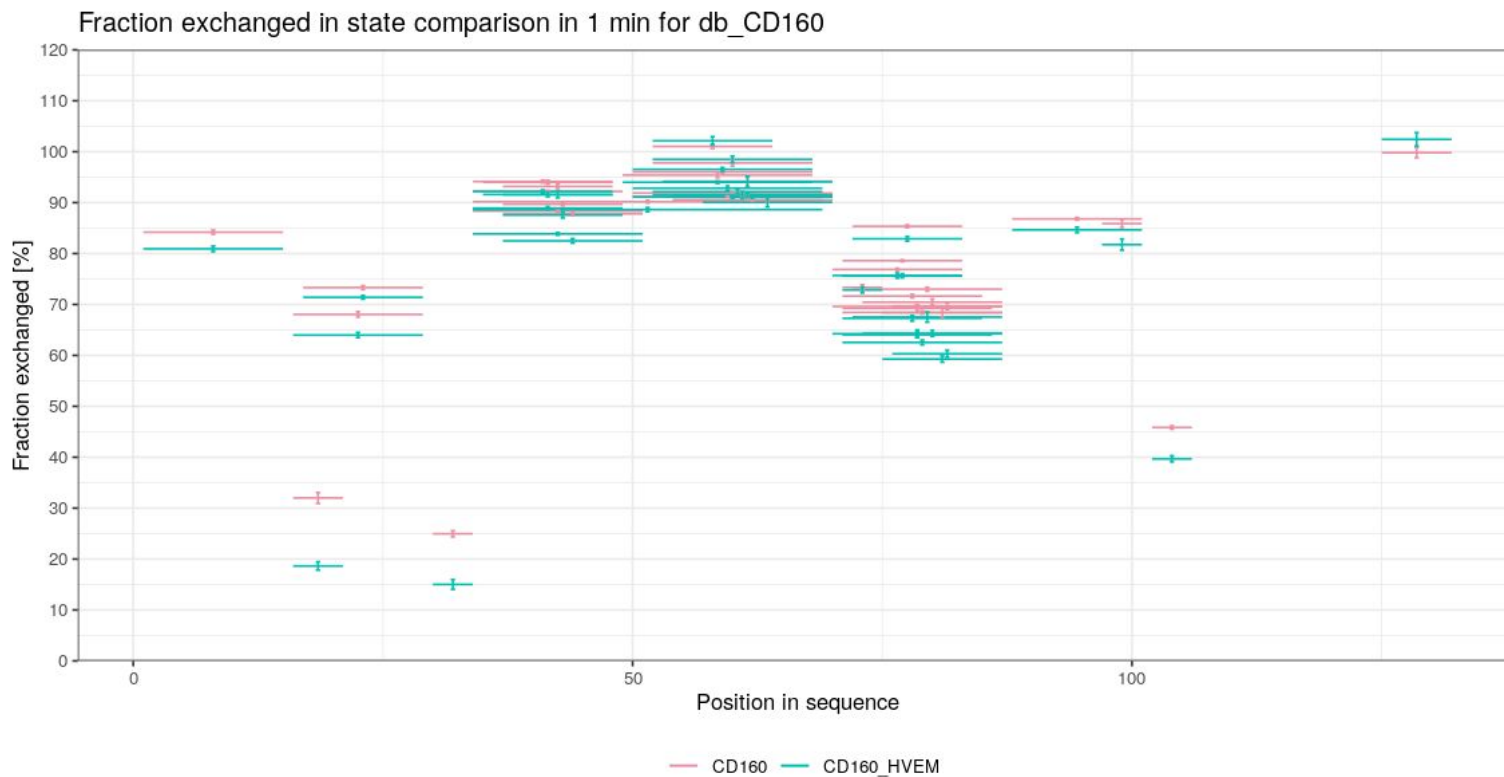
Weronika Puchała

BioGenies Seminar, 18/11/2020

HDX-MS w pigułce

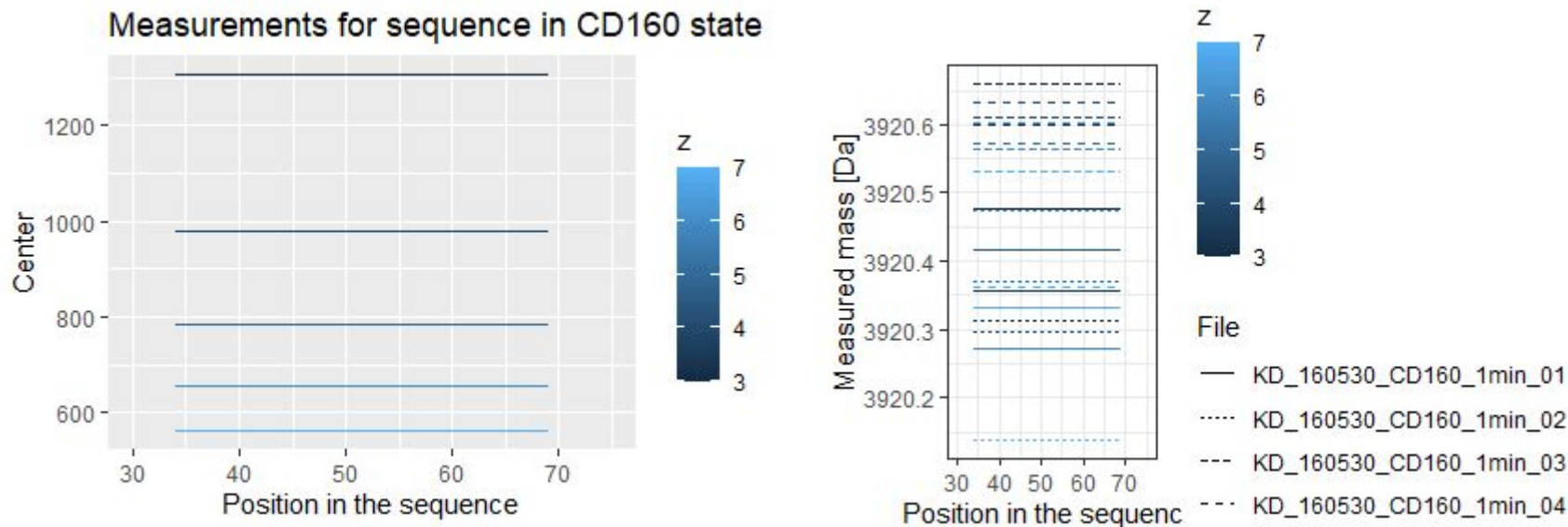
- wymiana wodoru z wiązania peptydowego na deuter z otoczenia (woda ciężka) + wymiana w drugą stronę
- mierzymy wartości dla peptydów! zestaw peptydów zależy od trawienia
- mierzymy masę przed i po wymianie, przyrost masy to deuterium uptake
- mierzymy w paru punktach czasowych
- mierzymy w różnych stanach biologicznych!
- porównujemy zmierzoną masę z a) pomiarami kontrolnymi b) wartościami tablicowymi

Comparison plot

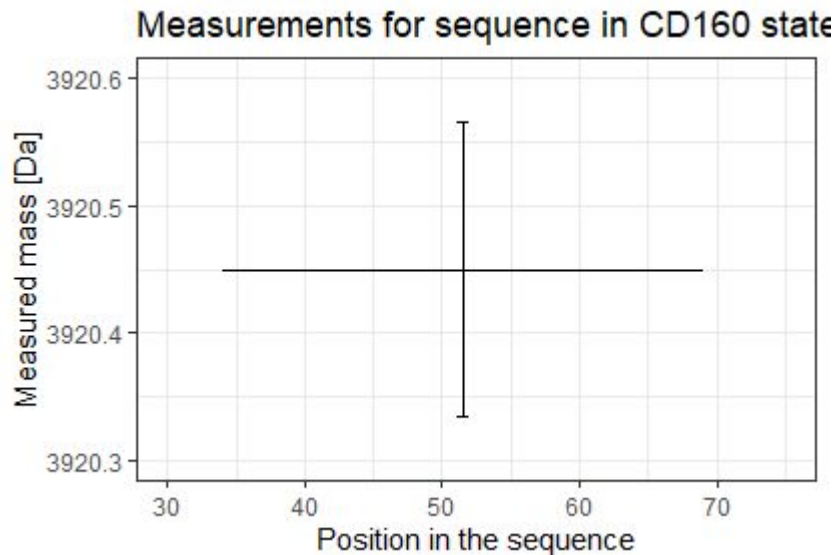
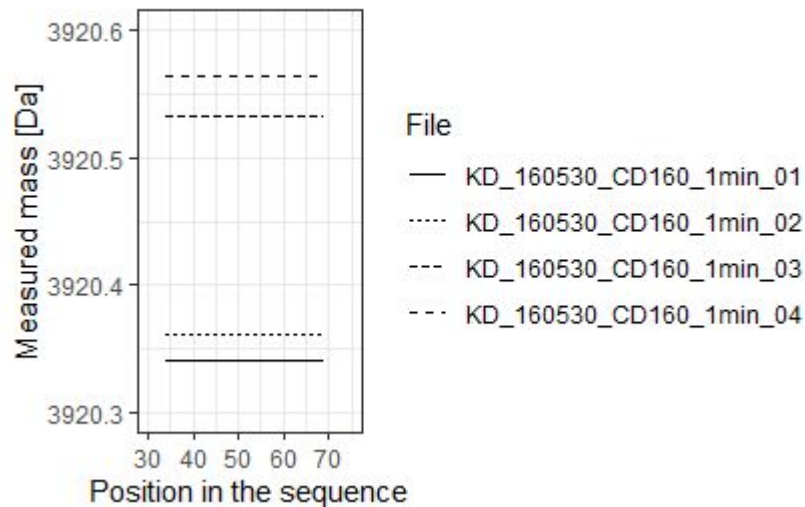


Jaka jest droga to takiego ładnego wykresu?

dane dla jednego peptydu w jednym stanie w jednym punkcie czasowym!!

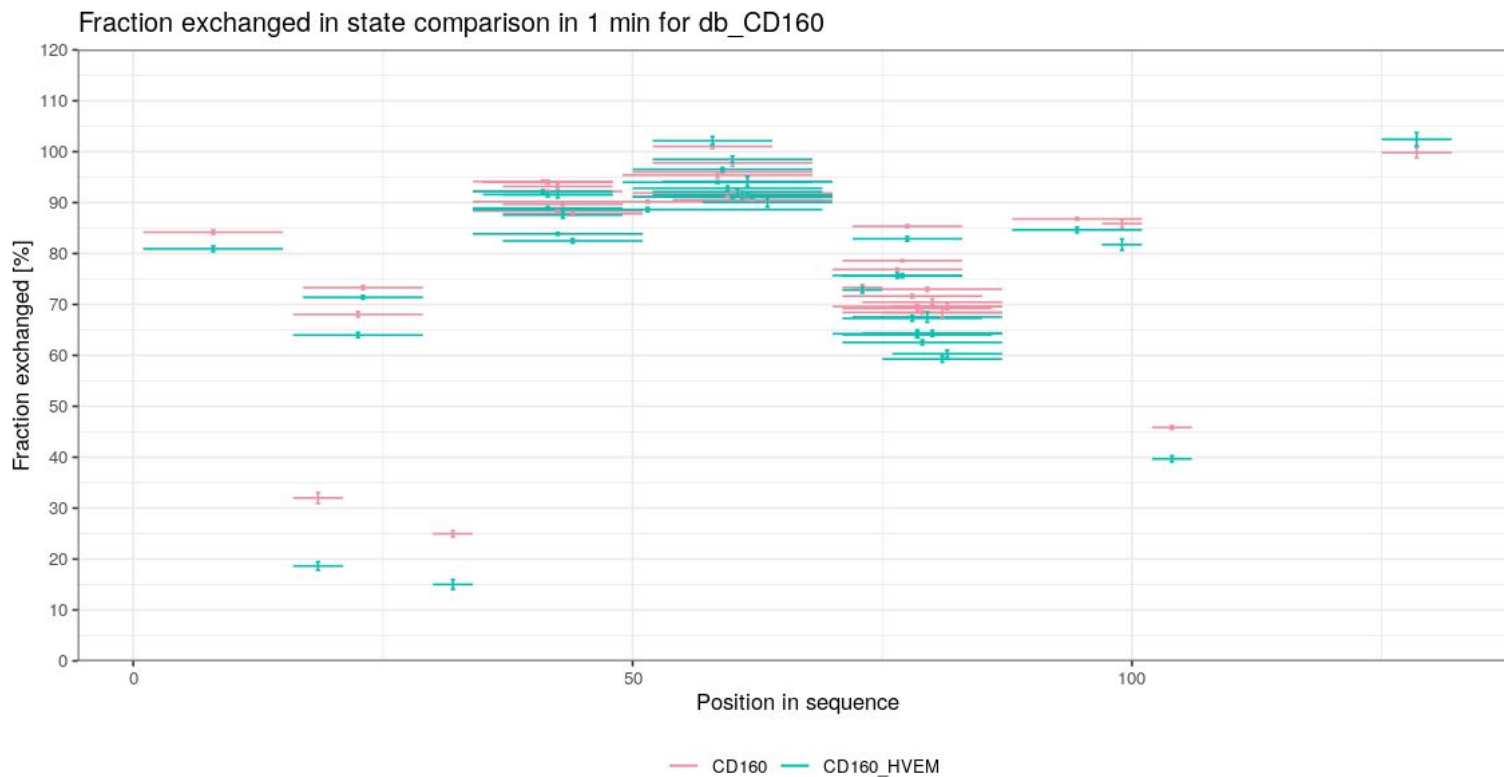


CD drogi do ładnego wykresu

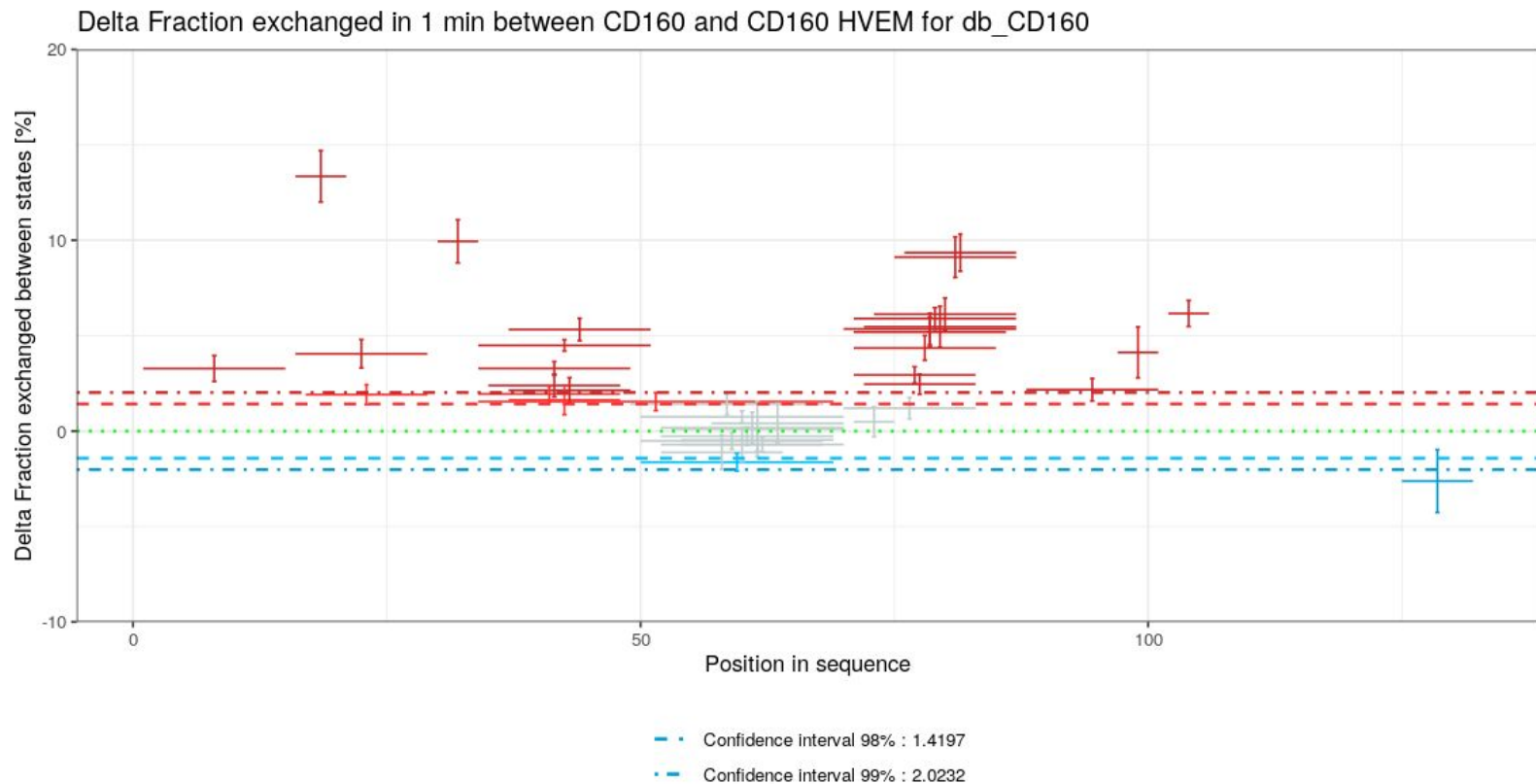


to jest tylko wartość masy - bez porównania z minimalną/maksymalną kontrolą

Comparison plot



Woods Plot



Przedziały ufności - Houde

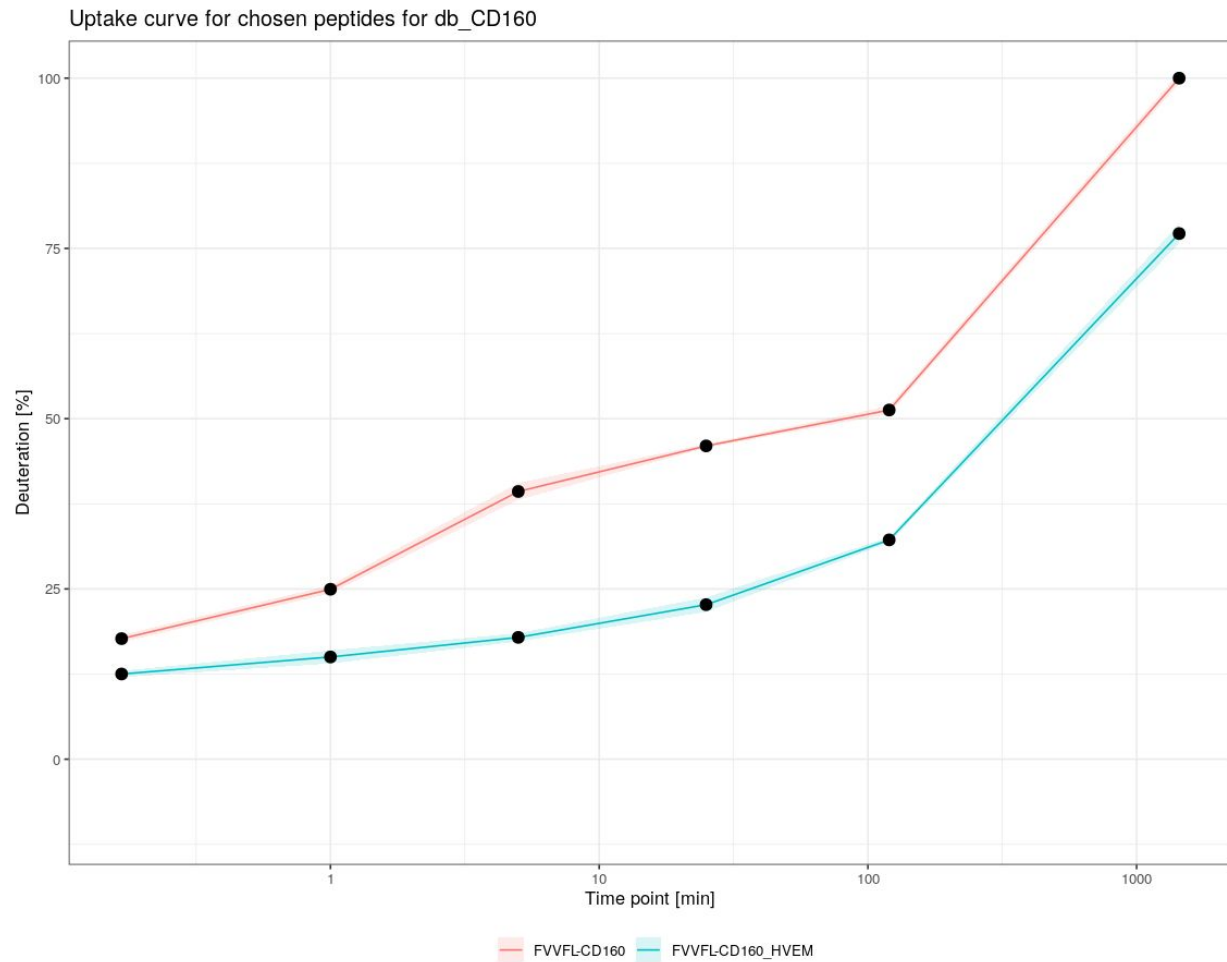
- robimy średnią z wszystkich wartości (czyli: różnic między dwoma stanami dla peptydów)

$$D(\Delta M_{i,t}) = S_{\text{ref}}(M_{i,t}) - S_{\text{exp}}(M_{i,t})$$

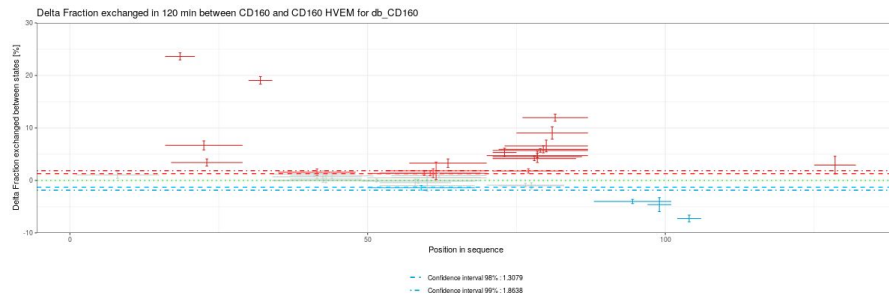
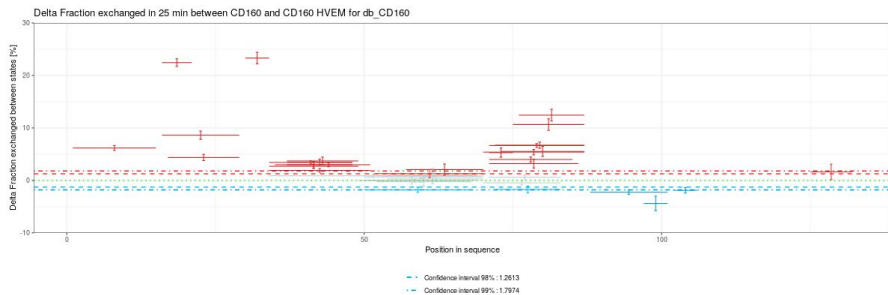
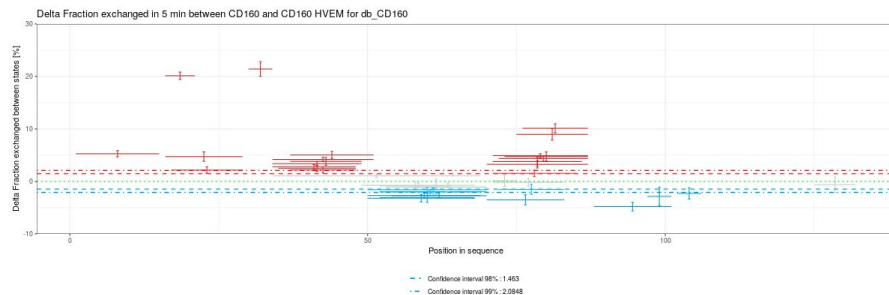
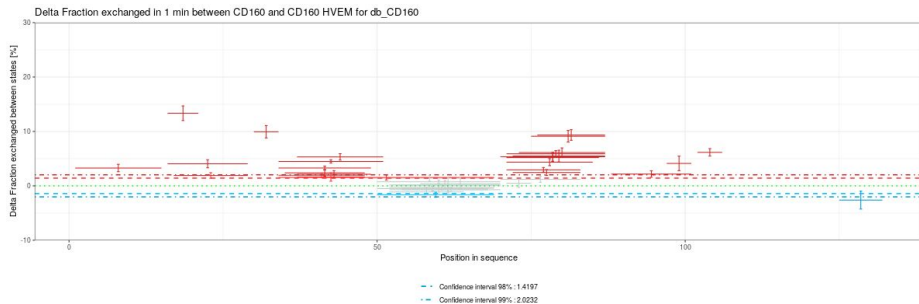
-
- liczymy standard error of the mean
- SEM mnożymy razy wartość t-Studenta dla 2 stopni swobody (czemu dwa stopnie? bo to różnica między dwoma stanami?) -> wynik
- przedział symetryczny względem 0
- pamiętamy, że te dane wcześniej były wielokrotnie agregowane!

“Using this value for SEM and multiplying it by the appropriate Student’s t table value for the 98% confidence limit (or interval) for two degrees of freedom gave an estimated 98% confidence limit² of about ± 0.5 Da (rounded to one decimal position) for any mean value for $D(\Delta M_{i,t})$ calculated from three separate H/DX-MS experiments. “

Kinetyka



Widać zmianę w czasie



Pytania:

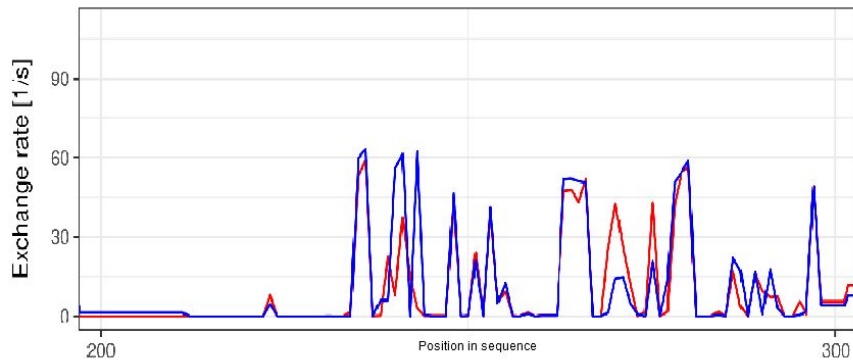
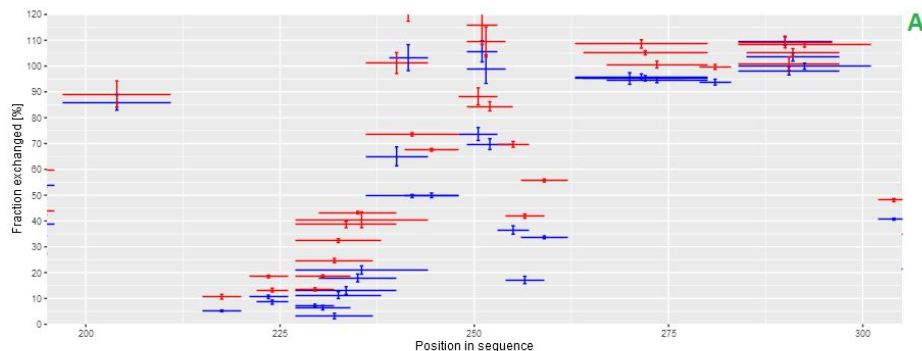
- jak można przedstawić te dane, żeby były możliwie informatywne ale czytelne? i atrakcyjne?

najlepiej: nowy sposób, którym zrobimy furrorę

- jaką alternatywną statystykę można zaproponować? ta jest budząca wątpliwości ale: jest prosta, stosowana i działa!

Główny zamysł:
HIGH
RESOLUTION

niestety nie
single residue



Co się dzieje w głębi

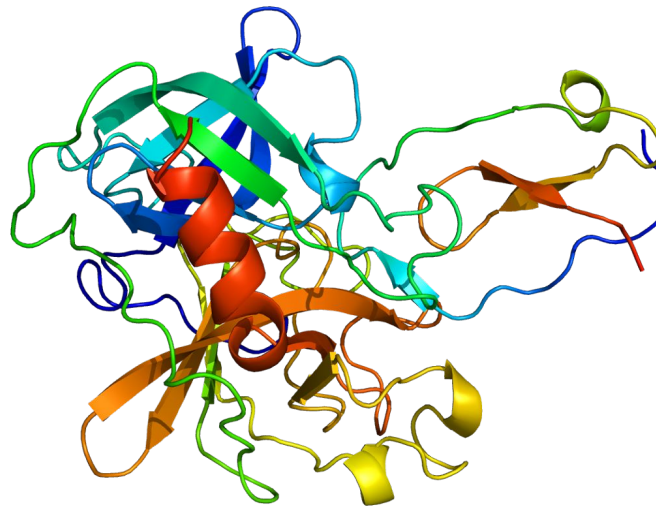
wymianę opisują stałe:

- k_{int} - policzone z tablic, Bai et al
- k_{exp} - eksperymentalne
- k_{open} - stała otwarcia, nieznana
- k_{close} - stała zamknięcia, nieznana

Stosunek tych zmiennych -> protection factor (definiowany później)

Co ogranicza wymianę?

- warunki zewnętrzne (temp, pH) - to jest uwzględnione w stałej wymianie k
- własności fizykochemiczne amidu i jego sąsiadów
- dostępność do deuteru czyli
- struktura białka
- stan biologiczny



Gessner et al 1

- bazują na sub-fragmentach x (części wspólne nakładających się peptydów)
- model liniowy
 - z danych exp, fitują wartość dla subfragmentu (minus proliny plus błąd) - szukamy x

$$D_{F_{n,t}}^{\text{exp}} = \sum_{j=1}^{N_n} x_{j,t} - p + E_{F_{n,t}}^{\text{lin}}$$

- robią to minimalizując błąd (najmniejsze kwadraty) po wszystkich fragmentach i punktach czasowych

$$GE^{\text{lin}} = \sum_{t=1}^T \sum_{n=1}^M \left(E_{F_{n,t}}^{\text{lin}} \right)^2$$

mamy komplet wartości x

Gessner et al 2

- model nieliniowy
 - cel: zebrać k_{exp} (pamiętamy: zależne od rzeczy typu pH, temp, seq, structure itd)
 - DU na reszcie i w czasie t , zakładając wymianę pierwszego rzędu:

$$D_{i,t} = (1 - e^{-k_{\text{ex},i} t})$$

- nie piszą tego wprost, ale to D to jest to wcześniejsze DU na subfragmencie, na reszcie
- back-exchange
 - możemy sobie policzyć (Bai) stałe wymiany w drugą stronę:

$$D_i^{\text{bk}} = e^{-k_{\text{bk},i} t_{\text{lag}}}$$

Gessner et al 3

- optymalizacja
 - uwzględniamy back-exchange

$$D_{i,t}^{\text{calc}} = (1 - e^{-k_{ex,i} t})(e^{-k_{bk,i} t_{\text{lag}}})$$

- rozpisujemy na amidy (już nie na subfragmenty!!)

$$D_{F,n,t}^{\text{exp}} = \sum_{i=m}^l (1 - e^{-k_{ex,i} t})(e^{-k_{bk,i} t_{\text{lag}}}) + E_{F,n,t}^{\text{nonlin}}$$

- minimalizujemy błąd tak jak poprzednio, żeby znać k_{exp}
- z naszych k_{exp} robimy wolną energię Gibbsa -> wynik -> wykres
- jesteśmy zadowoleni

Skinner et al 1

- wprowadzają protection factor jako $PF = k_{\text{int}} / k_{\text{exp}}$ (EX2)
- DU dla peptydu:

$$D_j(t_k, \{P_i\}) = \frac{1}{n_j} \sum_{i=m_j+1}^{m_j+n_j-1} \left(1 - e^{-\frac{k_i^{\text{int}}}{P_i} t_k} \right),$$

- obcięta pierwsza reszta w ramach korekcji back-exchange
- szukamy zestawów PF, tak, żeby pasowało:

$$D_j^{\text{pred}}(t_k) = D_j^{\text{exp}}(t_k) + \varepsilon_{j,k},$$

- funkcja kosztu:

$$\begin{aligned} C(t_k, \{P_i\}) &= \sum_j \sum_k w_{jk} [D_j^{\text{pred}}(t_k, \{P_i\}) - D_j^{\text{exp}}(t_k)]^2 \\ &= \sum_j \sum_k (\varepsilon_{j,k})^2, \end{aligned}$$

waga nie wiadomo skąd.

Skinner et al 2

- losowo generujemy 10^4 zestawów takich PF (żeby pasowało) - random search
- zwycięża ten, który ma najmniejszy koszt i jest warunkiem początkowym optymalizacji - najmniejsze kwadraty
- koniec tej części
- wygenerowane zestawy clusterujemy - nie do końca wiadomo po co, bo autorzy tego nie wykorzystują.

Kan et al 1

- matlab, szeroko cytowany jako przykład wysokorozdzielczego
- wymiana odwrotna: deuter -> wodór, ale nie szkodzi
- schemat po peptydach:
 - początkowy strzał / poprawienie wartości
 - uwzględniają back exchange (z danych kontrolnych, przygotowane raz)
 - “trust-region-reflective optimization method and other functions provided in that environment”
 - liczą widmo (symulują), porównują z eksperymentalnym
 - liczą błąd
 - koniec: albo błąd wystarczająco mały/stabilny, albo max iteracji
- wynik: wartości DU per reszta
- fit do eksponenty -> stałe i PF

Smit et al 1

- szuka PF ale zakłada inną formę z PF:

$$k_{obs} = \frac{k_{int}}{1 + PF}$$

- minimalizuje funkcję kosztu (TensorFlow)

$$\min_g \frac{1}{N_p N_t} \sum_{\mu\nu} \left[D_{\mu\nu} - \sum_{\rho} X_{\mu\rho} \left(1 - \exp \left\{ \frac{-k_{int,\rho} t_{\nu}}{1 + 10^{g_{\rho}}} \right\} \right) \right]^2 + \frac{\lambda}{N_r} \sum_{\rho=1}^{N_r-1} |\Delta g_{\rho}|$$

- $g = \log_{10}(PF)$
- k_{int} liczone sposobem Skinnera
- X to macierz reszt w peptydzie
- ostatnie to człon regulujący (Zhang et al) - lambda jest tunowana

Co te podejścia mają wspólnego?

- definiują warunki, dla których jest opis
- tworzą opis wymiany
- korygują back-exchange (w opisie lub założeniach)
- poprawiają swój model, aż spełni jakiś warunek (np minimalizacja błędu)

Poza tym, prezentowane podejścia są bardzo różne - łącznie z tym, że mają różne produkty!

Pytania:

- Jak mówić o back-exchange?
 - co kiedy wymiana jest w te i we wte?
 - mamy w czasie malejący przyrost deuteru a potem wzrost. jak to interpretować? czy wymienili się z powrotem te same? inne? ??
- Jak korygować back-exchange?
 - jak uwzględnić wpływ back exchange na wymianę?
 - jak interpretować różnice w wynikach?

Pytania:

- Jaka agregacja?
 - z powtórzeń - to oczywiste
 - z punktów czasowych - to nieoczywiste
 - z sub-fragmentów - to oczywiste
 - z peptydów - to nieoczywiste
 - z całego białka - to bardzo nieoczywiste
- Jak zrobić lepszy model?