

8. Übung zur Vorlesung “Algorithmische Massenspektrometrie”

Wintersemester 2020/2021

Sebastian Böcker, Kai Dührkop

Ausgabe: 07. Januar 2021, Abgabe: 14. Januar 2021, vor der Übung

1. **Aufbau einer Köderdatenbank für Peptid-Spektren:** Aus einer Proteindatenbank wird eine Spektrendatenbank berechnet, in dem die einzelnen Proteine in-silico tryptisch verdaut (siehe erste Übungsserie) und für jedes Peptid entsprechend seiner Ionenserien ein Spektrum generiert wird. Als nächstes soll eine Köderdatenbank auf ähnliche Weise berechnet werden.
 - (a) Damit Zufallstreffer in der Köderdatenbank genauso wahrscheinlich sind wie in der echten Datenbank, sollte die Köderdatenbank die gleiche Anzahl an Einträgen wie die echte Datenbank haben. Begründen Sie, warum bei der *inverted protein* Methode dies nicht immer garantiert werden kann. Zeigen Sie ein Beispiel.
 - (b) Gegeben sei eine Datenbank mit folgenden beiden Proteinen:

DDCDKPKMN
MNMQKWAWAKFIFIRVW

Geben Sie die zugehörigen Köderdatenbanken an, die Sie jeweils nach Anwendung der *inverted proteins*, *inverted peptides* und *pseudo-inverted peptides* Methode erhalten.

(5 Punkte)

2. **FDR und q-Values:** Nach einer Datenbanksuche von 27 Sequenzen in einer Datenbank (die auch die Einträge der Köderdatenbank enthält) wurde für jede Sequenz der beste zu erreichende Score und die Datenbank, aus der dieser gewonnen wurde, ausgegeben. Siehe Tabelle 1.
 - (a) Für welchen Score erhalten Sie eine FDR von 20% ?
 - (b) Geben Sie den q-Value für die Sequenz 20 mit Score 84 an.
 - (c) Wie ändert sich der q-Value für Sequenz 20, wenn die ersten 5 Sequenzen (mit Score größer 115) aus der Tabelle entfernt werden.

(4 Punkte)

#	Score	Datenbank
17	128	echt
25	125	echt
6	120	echt
10	119	echt
24	117	echt
1	115	echt
23	114	echt
16	112	echt
5	106	echt
7	102	Köder
18	100	echt
15	99	echt
22	97	echt
27	94	echt
3	90	echt
26	88	Köder
2	87	echt
20	84	echt
11	83	echt
21	81	echt
12	78	echt
13	72	Köder
14	71	echt
4	68	echt
19	65	Köder
8	62	Köder
9	61	echt

Tabelle 1: