

An effective algorithm for PTM identifying 2-PTMs from tandem mass spectra

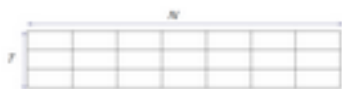
Soohan Ahn¹, Hyunwoo Kim¹, Eunok Paek¹, Heejin Park¹
¹Hanyang University, KOREA

1. Overview

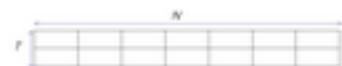
- o MODa 알고리즘은 MS/MS 데이터에서 Dynamic programming을 기반으로 단백질 데이터베이스에서 PTM을 포함한 펩타이드를 동정하는 알고리즘이다.
- o MODa에는 PTM을 한 개 이하만 허용하는 OneMOD, 그리고 복수의 개수를 허용하는 MultiMOD 두 가지 버전이 있다.
- o 이 논문에서는 두 개 이하의 PTM을 허용하는 펩타이드를 더욱 효율적으로 동정하고자 TwoMOD 버전을 고안하였다.
- o 기존 MultiMOD의 시간 복잡도는 $O(C_{mul} T^2 N(L + R))$ 이지만, TwoMOD에서는 $O(C_{two} N(L + R))$ 로 향상시켰다.
- o 실제 실험 결과, 기존의 MultiMOD에 비하여 약 35%의 시간이 단축되었으며, 두 개 이하의 PTM을 갖는 펩타이드도 성공적으로 동정하였다.

2. MODa

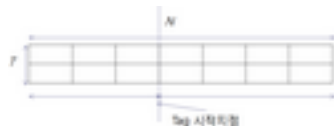
- o Unrestrictive Search.
- o 기존의 MS-Alignment에 비하여 약 40배의 속도 향상.
- o Dynamic programming 기반의 알고리즘 사용.
Dynamic Table에서 행의 이동이 PTM을 의미한다.
- o Version:
 - o **MultiMOD**
 - T : 후보군 펩타이드 당 태그의 수. (3 이상)
 - N : 후보군 펩타이드의 길이.



- o **OneMOD**
 - T : 후보군 펩타이드 당 태그의 수. (2로 고정.)
 - N : 후보군 펩타이드의 길이.



- o **TwoMOD**
 - Tag를 기준으로 후보군 펩타이드를 두 부분으로 분할.
 - 각 분할에 대하여 OneMOD를 수행.

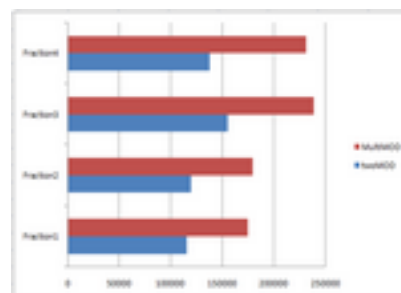


- o 시간 복잡도

OneMOD	$O(C_{one} N \max(L, R))$
TwoMOD	$O(C_{two} N(L + R))$
MultiMOD	$O(C_{mul} T^2 N(L + R))$
C_k : 후보군 펩타이드의 수	
T : 후보군 펩타이드당 태그의 수 + 2.	
N : 후보군 펩타이드의 길이.	
L, R : 각각 후보군 펩타이드 중 왼쪽, 오른쪽의 flanking mass에 해당하는 아미노산 길이.	

3. Result

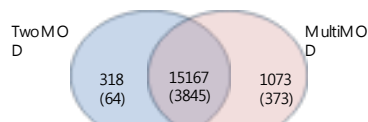
- o 실험 데이터
 - 단백질 DB: Uniprot human database 2013-5 version.
 - 스펙트럼: Human Plasma data (67,648 MS/MS Spectra)
- o 67,648개의 스펙트럼을 4개로 분할하여 MODa의 TwoMOD와 MultiMOD를 수행하여 비교.
- 모든 분할에서 약 35~40%의 속도 향상.



- MultiMOD와 유사한 비율로 2개의 PTM을 갖는 펩타이드를 검색.

	MultiMOD	TwoMOD
0 Modifications	12,909	12,557
1 Modifications	3,186	2,761
2 Modifications	142	167
> 2 Modifications	3	0

- o PSM / Unique peptide 비교

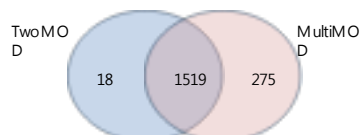


- o 포함된 PTM 개수별로 Unique peptide 비교.

-2 PTMs



-1 PTMs



-0 PTMs

