

「第六回 生物系研究者のための質量分析に関する教育セミナー」報告

日時：2010年6月10日（木）15：00～16：40

場所：生物科学研究棟 3階大セミナー室（S311）

講師：マトリックスサイエンス（株） 高江洲 宏智 氏

内容： 1）MS、MS/MS スペクトルデータを基にしたタンパク質/ペプチド同定 概論

2）MASCOT 検索パラメータの説明と検索結果の見方

セミナー参加者：38名

基幹研究所：5 研究室、5 チーム、1 ユニット

脳科学総合研究センター：4 チーム、1 ユニット

今回は今までと違ってデータ解析の話だったので、常連の方や初めて又は2回目という方が参加されたことが特徴でした。

内容はデータの解析方法やデータの信頼度の話しなど、測定している研究者にとって、解析のブラックボックス的な部分が少し理解でき、重要な内容で、有意義なセミナーでした。

以上

ケミカルバイオロジー 研究基盤施設教育セミナー

第六回 生物系研究者のための 質量分析に関する教育セミナー

質量分析について、初歩から学ぶ教育セミナーを開催いたします。

日時 2010年6月10日（木）15：00～16：30

場所 生物科学研究棟 3階大セミナー室（S311）

講師 マトリックスサイエンス（株）

高江洲 宏智 氏

内容 1）MS、MS/MSスペクトルデータを基にした
タンパク質/ペプチド同定 概論

2）MASCOT検索パラメータの説明と検索結果の見方

主催 ケミカルバイオロジー研究基盤施設
バイオ解析チーム

連絡先 バイオ解析チーム
渡辺 (kowashi@riken.jp)

* 参加に当たり事前予約が必要となりますので、
連絡先(渡辺)までご連絡ください。



今後の教育セミナーの予定

第七回 2010年7月21日（水）アジレント・テクノロジー（株）

第八回 2010年8月26日（木）マトリックスサイエンス（株）

バイオ解析チームHPへの
リンクはココをクリック

著作権はRIKENにあります。
無断複製・転載を禁止します。

他の教育セミナーへの
リンクはココをクリック

MASCOT 基礎事項

- ・ MS, MS/MS のスペクトルデータを基にした
タンパク質 / ペプチド 同定 概論
- ・ MASCOT 検索パラメータの説明と
検索結果の見方

マトリックスサイエンス株式会社
高江洲 宏智

{MATRIX}
{SCIENCE}

MS, MS/MS のスペクトルデータと タンパク質 / ペプチド 同定

- ・ PMF (Peptide Mass Fingerprinting)

MS のスペクトルデータ から タンパク質 同定

- ・ MIS (Mascot Ions Search)

MS / MS のスペクトルデータ から ペプチド 同定

- ・ Protein Identification (同定)

≠ 「リストアップされたタンパク質の存在を
100% 保障する」

{MATRIX}
{SCIENCE}

* データ数は2010年6月現在

MASCOT で使用しているデータベース

- **NCBI nr**

NCBIで公開されている、nr(non-redundant)タンパク質データベース。データ数約1100万件。名前と違い網羅性高い(冗長性も高い)。配列相同性検索などにも利用。

http://www.ncbi.nlm.nih.gov/blast/blast_databases.shtml など

- **Swiss-Prot**

EBI で公開されている。データ数約52万件。論文情報などのアノテーションが豊富。

<http://au.expasy.org/sprot/>

- **IPI (human など)**

EBIで公開されている。データ数約8.7万件。網羅性高く冗長性が低いので、生物種が適合すれば最もお勧め。

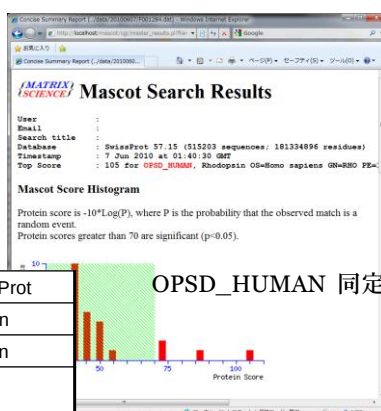
<http://www.ebi.ac.uk/IPI/Algorithm.html>

{MATRIX}
SCIENCE}

PMF 同定結果例 - OPSD_HUMAN

9 / 19

- 832.662
- 903.342
- 1186.439
- 1373.681
- 1403.722
- 1515.444
- 1617.857
- 1727.916
- 1743.951
- 1759.966
- 1788.721
- 1804.71
- 1818.963
- 2159.143
- 2174.812
- 2190.112
- 2256.871
- 2273.266
- 2288.489



Database	SwissProt
Taxonomy	Human
Enzyme	Trypsin
Missed Cleavage	1
Fixed mod	なし
Var mod	なし
Protein mass	なし
Peptide tol	±0.4 Da

OPSD_HUMAN 同定

- 832.662
- 903.342
- 1186.439
- 1373.681
- 1403.722
- 1515.444
- 1617.857
- 1727.916
- 1743.951
- 1759.966
- 1788.721
- 1804.71
- 1818.963
- 2159.143
- 2174.812
- 2190.112
- 2256.871
- 2273.266
- 2288.489

{MATRIX}
SCIENCE}

Probability based Scoring - MASCOT

Random match or not ?

9 / 19

• 832.662
• 903.342
• 1186.439
• 1373.681
• 1403.722
• 1515.444
• 1617.857
• 1727.916
• 1743.951
• 1759.966
• 1788.721
• 1804.71
• 1818.963
• 2159.143
• 2174.812
• 2190.112
• 2256.871
• 2273.266
• 2288.489

• 832.662
• 903.342
• 1186.439
• 1373.681
• 1403.722
• 1515.444
• 1617.857
• 1727.916
• 1743.951
• 1759.966
• 1788.721
• 1804.71
• 1818.963
• 2159.143
• 2174.812
• 2190.112
• 2256.871
• 2273.266
• 2288.489

各タンパク質のScore

P ... ランダムマッチだったとした場合の確率
(1/10,000 、 1/1,000,000 など)

$$\text{Score} = -10 \log P$$

例)

P が100万分の1、 10^{-6} なら

Score は 60

P が1億分の1、 10^{-8} なら

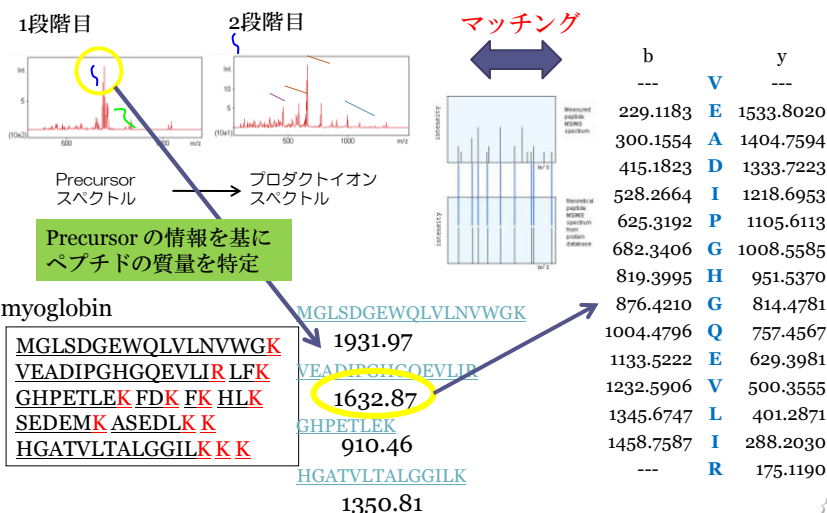
Score は 80

同定基準値

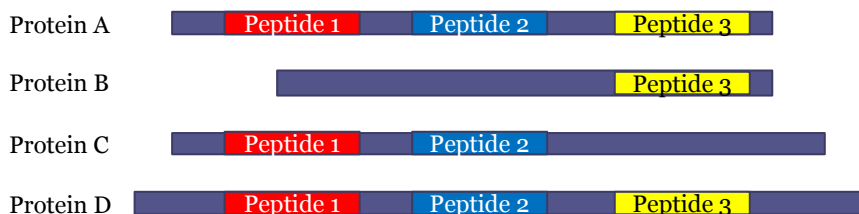
データベースのサイズ（総エントリー数）を
基に、期待値が0.05となるようなPを算出し、
その時のスコアを同定基準値とする

{MATRIX}
{SCIENCE}

MS/MS のスペクトルデータと データベース検索

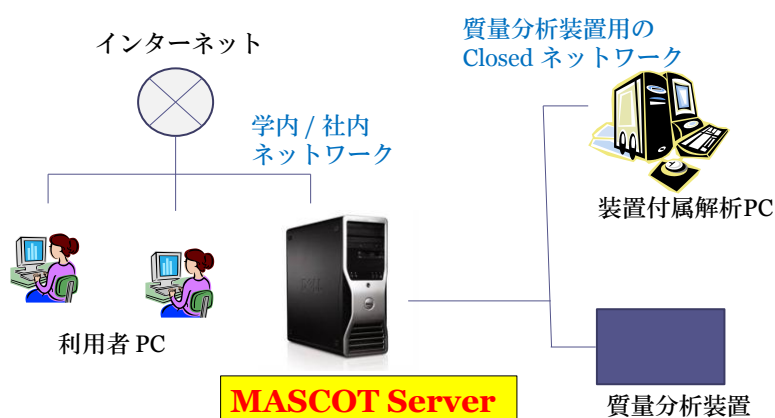


MIS はペプチドを同定するが、タンパク質を一意に絞れないケースがある



- Peptide 1 ~ 3 が同定された場合、Protein A,D が存在する可能性が最も高いが、Protein B や Protein C については、あるともないとも言い切れない
- Protein A とペプチドのヒットの仕方が同じである Protein D については、区別がつけられない

MASCOT Server システム (local版)



MASCOT Server ... **WEBアプリケーションソフトウェア**。
通常、弊社から PC とセットでご購入頂いております。

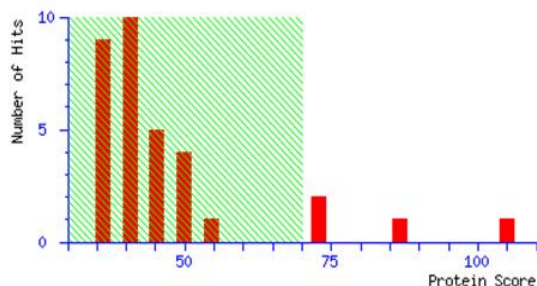
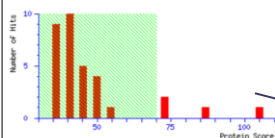
PMF 検索結果の見方 - グラフ

(MATRIX) Mascot Search Results

User : H.Takasu
Email : takasu@matrixscience.com
Search title : Search Test
Database : SwissProt 57.15 (515203 sequences; 181334896 residues)
Timestamp : 7 Jun 2010 at 05:14:26 GMT
Top Score : 105 for **OSD_HUMAN**, Rhodopsin OS=Homo sapiens GI=181334896

Mascot Score Histogram

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed Protein scores greater than 70 are significant ($p < 0.05$).



Concise Protein Summary Report

Format As: Concise Protein Summary Help
Significance threshold $p < 0.05$ Max. number of hits AUTO

Re-Search All Search Unmatched

1. **OSD_HUMAN** Mass: 38846 Score: 105 Expect: 1.6e-005 Matches: 9
Rhodopsin OS=Homo sapiens GI=181334896 PE=1 SV=1
OSD_PPC1 Mass: 38947 Score: 45 Expect: 1.6 Matches: 5
Rhodopsin OS=Homo sapiens GI=181334896 PE=1 SV=1
OSD_MOUSE Mass: 39052 Score: 43 Expect: 28 Matches: 5
Rhodopsin OS=Mus musculus GI=181334896 PE=1 SV=1
OSD_SHEEP Mass: 38846 Score: 43 Expect: 28 Matches: 5

横軸：MASCOT Score

縦軸：タンパク質数

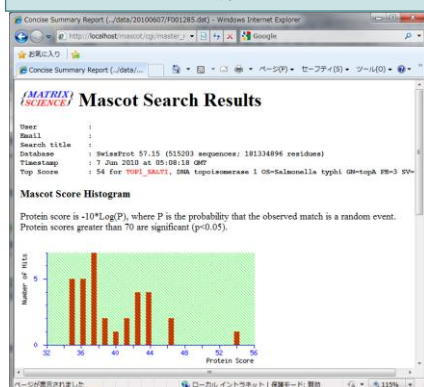
MASCOT スコア分布。緑の網掛けの一番右側（このケースでは70）が同定基準値

(MATRIX) SCIENCE

検索後、同定できたかは一目でわかる

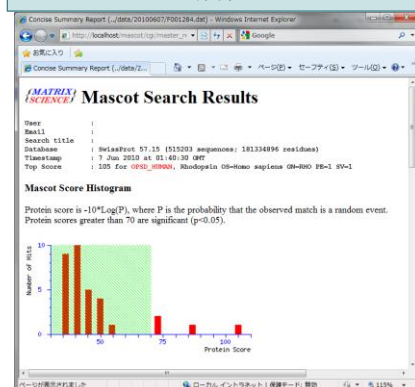
→ スコア分布のグラフで、緑の網掛けから出ているタンパク質があるかどうか

失敗



→ 再検索

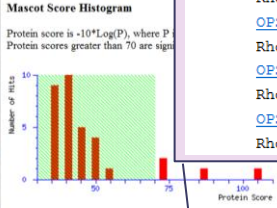
成功



→ 結果の検証

(MATRIX) SCIENCE

PMF 検索結果の見方 - タンパク質



Mascot Score Histogram
Protein score is $-10 \cdot \log(P)$, where P is the probability of a random match.
Protein scores greater than 70 are significant.

Concise Protein Summary Report

Format As: Concise Protein Summary | Help
Significance threshold p < 0.05 | Max. number of hits AUTO
Re-Search All | Search Unmatched

Rank	Protein ID	Mass	Score	Expect	Matches
1.	OPSD_HUMAN	38866	105	1.6e-005	9
	Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1				
	OPSD_PHOVI	38947	45	16	5
	Rhodopsin OS=Phoca vitulina GN=RHO PE=1 SV=1				
	OPSD_MOUSE	39002	43	28	5
	Rhodopsin OS=Mus musculus GN=Rho PE=1 SV=1				
	OPSD_SHEEP	38866	43	28	5
	Rhodopsin OS=Ovis aries GN=RHO PE=1 SV=2				

- タンパク質のID, 質量, Score, 期待値, マッチしたペプチド数, Description
- タンパク質のIDはハイパーリンクになっている。クリックすると、より詳細な検証ができるWEBページへ (Protein View)。

同定基準値を超える = 試料内に必ず存在する？

→ **No**

[理由]

- 検索エンジンの間違い
- Taxonomy の違い
- データベースが不十分 (特に“重複”が問題)

[対処法]

- 目視による結果検証
- Taxonomy の検討
- MS/MS 検索

1.	OPSD_HUMAN	Mass: 38866	Score: 105	Expect: 1.6e-005	Matches: 9
	Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1				
	OPSD_PHOVI	Mass: 38947	Score: 45	Expect: 16	Matches: 5
	Rhodopsin OS=Phoca vitulina GN=RHO PE=1 SV=1				
	OPSD_MOUSE	Mass: 39002	Score: 43	Expect: 28	Matches: 5
	Rhodopsin OS=Mus musculus GN=Rho PE=1 SV=1				
	OPSD_SHEEP	Mass: 38866	Score: 43	Expect: 28	Matches: 5
	Rhodopsin OS=Ovis aries GN=RHO PE=1 SV=2				
2.	OPSD_MACFA	Mass: 39036	Score: 88	Expect: 0.00074	Matches: 8
	Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1				
	LUXS2_LACDB	Mass: 17500	Score: 45	Expect: 17	Matches: 4
	S-ribosylhomocysteine lyase 2 OS=Lactobacillus delbrueckii subsp. bulga				
3.	OPSD_CRIGR	Mass: 39071	Score: 73	Expect: 0.026	Matches: 7
	Rhodopsin OS=Cricetulus griseus GN=RHO PE=1 SV=1				
	OPSD_CANFA	Mass: 38936	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Canis familiaris GN=RHO PE=1 SV=1				
	OPSD_FELCA	Mass: 39023	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Felis catus GN=RHO PE=1 SV=1				
	OPSD_RABIT	Mass: 38968	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Orctolagus cuniculus GN=RHO PE=1 SV=1				
	OPSD_SMICR	Mass: 39037	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Sminthopsis crassicaudata GN=RHO PE=1 SV=2				
4.	OPSD_LOXAF	Mass: 39023	Score: 72	Expect: 0.033	Matches: 7
	Rhodopsin OS=Loxodonta africana GN=RHO PE=1 SV=1				
	OPSD_FIG	Mass: 39001	Score: 72	Expect: 0.033	Matches: 7
	Rhodopsin OS=Sus scrofa GN=RHO PE=1 SV=1				

MIS 検索結果の見方

ヘッダー。検索日, 検索対象データベース、検索日、基準スコアを超えるタンパク質に関する情報

Mas

```

User: takesu
Email: takesu@matrixscience.com
Search title: ThreeMSdata.mgf
MS data file: SwissProt 2010_06 (517100 sequences; 182146551 residues)
Database: 7 Jun 2010 at 06:24:59 GMT
Protein hits:
  FAD1_YEAS  Phenylacrylic acid decarboxylase OS=Saccharomyces cerevisiae GN=FAD1 PE=1 SV=2
  CEN5_KLAL  COP9 signalosome complex subunit 5 OS=Kluyveromyces fragilis GN=CEN5 PE=1 SV=1
  RPL12_RHOC  60S ribosomal protein L12 OS=Encephalitozoon cuniculi GN=RPL12 PE=1 SV=1
  RRP5_BOVIN  Protein RRP5 homolog OS=Bos taurus GN=RRP5 PE=2 SV=1
  LRC42_MOUSE  Leucine-rich repeat-containing protein 42 OS=Mus musculus GN=LRC42 PE=2 SV=1
  RRS3A3_RHOC  Ribosome import ATP-binding protein RRS3 OS=Encephalitozoon cuniculi GN=RRS3A3 PE=1 SV=1
  
```

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 41 indicate identity or extensive homology (p<0.05). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 41 indicate identity or extensive homology (p<0.05). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Peptide Summary Report

Format As: Peptide Summary

Significance threshold p: 0.05

Standard scoring: * Mascot scoring * Mascot scoring

Show pop-ups * Suppress pop-ups

Select All Select None Search Selected Free

スコア分布。ただし Protein Score なので注意

(MATRIX)
SCIENCE

MIS 検索結果の見方

1. [OPSD_HUMAN](#) Mass: 38866 Score: 487 Matches: 3 (3) Sequences: 3 (3)
Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1

☐ Check to include this hit in error tolerant search

Query	Observed	Mr (expt)	Mr (calc)	Delta Miss	Score	Expect	Rank	Unique	Peptide
1	832.4500	831.4427	831.4450	-0.0023	1	71	0.00011	1	U K.AEKEVTR.M
2	1403.6649	1402.6576	1402.6576	0.0000	0	163	3.1e-14	1	U K.NFLGDDDEASATVSK.T
3	2257.0700	2256.0627	2256.0634	-0.0006	0	253	1.7e-23	1	U -.MNGTEGPNFYVFPFSNATGVVR.S

Proteins matching the same set of peptides:

[OPSD_MACFA](#) Mass: 39036 Score: 487 Matches: 3 (3) Sequences: 3 (3)
Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1

Query以下、MS/MS のスペクトルデータを基に同定されたペプチドの情報

Query : MS/MS の番号。ペプチドの質量換算で小さい順に 1 から番号がついている

Observed : Precursor の m/z

Mr (expt): Precursor の m/z と charge を基に算出されたペプチドの質量

Mr (calc): データベース側のペプチド質量理論値

Score : スコア。ペプチドの確からしさを表す。

Peptide : ペプチドのアミノ酸配列

(MATRIX)
SCIENCE

MIS 検索結果の見方

1. **OPSD_HUMAN** Mass: 38866 **Score: 487** Matches: 3(3) Sequences: 3(3)
 Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1
☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1	832.4500	831.4427	831.4450	-0.0023	1	71	0.00011	1	U	K.AEKQVTR.M
2	1403.6649	1402.6576	1402.6576	0.0000	0	163	3.1e-14	1	U	K.NPLGDDASATVSK.T
3	2257.0700	2256.0627	2256.0634	-0.0006	0	253	1.7e-23	1	U	- .MNGTEGPNFYVFFSNATGVVR.S

Proteins matching the same set of peptides:
OPSD_MACFA Mass: 39036 Score: 487 Matches: 3(3) Sequences: 3(3)
 Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1

Protein の Score

各ペプチドの Score を基に
算出されている。

Standard Scoring の場合

各ペプチドの Score の和

MUDPIT Scoring の場合

信頼性の高いペプチドのみを使用、同
定基準値を超えた分だけスコアを加算

Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Unique Peptide

1	832.4500	831.4427	831.4450	-0.0023	1	71	0.00011	1	U	K.AEKQVTR.M
2	1403.6649	1402.6576	1402.6576	0.0000	0	163	3.1e-14	1	U	K.NPLGDDASATVSK.T
3	2257.0700	2256.0627	2256.0634	-0.0006	0	253	1.7e-23	1	U	- .MNGTEGPNFYVFFSNATGVVR.S

Proteins matching the same set of peptides:
OPSD_MACFA Mass: 39036 Score: 487 Matches: 3(3) Sequences: 3(3)
 Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1

2. **FAH1_YEAST** Score: 487
 Phenylacrylic acid
☐ Check to include this hit in error tolerant search

Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Unique Peptide

1	832.4500	831.4427	831.4450	-0.0023	1	71	0.00011	1	U	K.AEKQVTR.M
---	----------	----------	----------	---------	---	----	---------	---	---	-------------

タンパク質の ID 部分をクリック

→ **Protein View**

Query 番号部分をクリック

→ **Peptide View**