

Volcano plot

まとめ、または図の補足説明

→ 3-1 (P.34～)

日本語マニュアルの該当ページ
×対応するスライド番号

←スライド番号

1



MASCOT SERVER MIS トレーニング資料

2023.2.16 更新

- MASCOT Serverとは
- 検索方法
- 対応フォーマットとデータ変換
- パラメーターと検索方法
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS検索のオプション
- 管理プログラム・カスタマイズ



MASCOT Server とは

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

→ 1-1 (P.1～)

・質量分析装置のデータをもとに、**ペプチド配列あるいはタンパク質を同定**するソフトウェア(検索エンジン)

・入力と参照先、出力

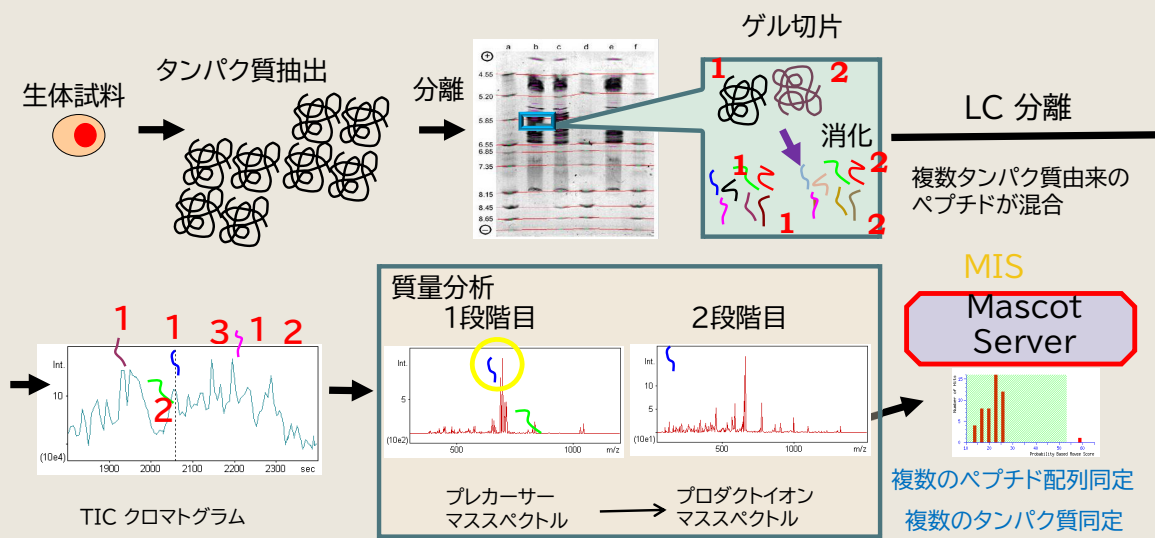
- **入力データ** : 質量分析装置から得られたスペクトル情報
- **参照先** : 配列から計算された理論スペクトル
(オプションで過去の実測スペクトルも可)
- **出力データ** : 同定ペプチド・同定タンパク質の情報

- MASCOT Serverでできる事・まとめ

- MASCOT Serverはウェブアプリケーション

2

補足説明：ショットガンプロテオミクス



- ・タンパク質は最初に分離で数十～数百 混ざった状態
- ・ペプチドになってから分離(ペプチド同定は他のペプチドの影響を受けない)

3

MASCOT Serverでできる事・まとめ①

→ 1-2 (P.1～)

定性 ペプチド断片の配列決定
タンパク質の同定
修飾の種類・位置

オプションで、ペプチド断片配列決定において以下2点も実施可能

- ・アミノ酸置換の探索
- ・クロスリンクペプチドの探索

定量 タンパク質の相対量(比)または定量指標

- * Spectral Counting/isobaric tag など限定的な手法
- * +Distiller で、isotopicな各種手法にも対応

■ MASCOT Serverとは

- MASCOT Serverでできる事・まとめ
- MASCOT Serverはウェブアプリケーション
- ネットワーク上でMASCOT Serverを指定

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

4

MASCOT Serverでできる事・補足図

→ 1-2 (P.2~)

Family	M	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	crAP	d1::sp TRY1_BOVIN	1600	28266	47	47	7	7	2.86	sp TRY1_BOVIN
2	1	SwissProt	d2::CP2CT_MOUSE	332	61419	76	76	13	13	2.00	Cytochrome P450 2C29 OS=Mus musculus OX=10090
2	2	SwissProt	d2::CP2CT_MOUSE	550	60887	27	27	8	8	0.88	Cytochrome P450 2C54 OS=Mus musculus OX=10090
2	3	SwissProt	d2::CY250_MOUSE	487	61128	27	27	10	10	1.20	Cytochrome P450 2C50 OS=Mus musculus OX=10090
2	4	SwissProt	d2::CP2F2_MOUSE	470	59267	32	32	12	12	2.11	Cytochrome P450 2F2 OS=Mus musculus OX=10090
2	5	SwissProt	d2::CP237_MOUSE	338	60590	22	22	8	8	0.89	Cytochrome P450 2C37 OS=Mus musculus OX=10090
2	6	SwissProt	d2::CP239_MOUSE	251	60856	13	13	4	4	0.37	Cytochrome P450 2C39 OS=Mus musculus OX=10090
2	7	SwissProt	d2::CP239_MOUSE	251	60856	13	13	4	4	0.37	Cytochrome P450 2C39 OS=Mus musculus OX=10090
3	3	SwissProt	d2::H571L_MOUSE	188	78552	13	13	4	4	0.28	Heat shock 70 kDa protein 1-like OS=Mus musculus OX=10090
3	1	SwissProt	d2::CYB5_MOUSE	1202	16817	42	42	5	5	3.08	Cytochrome b5 OS=Mus musculus OX=10090 GN=Cyt
4	1	SwissProt	d2::PDIA1_MOUSE	1121	64694	53	53	16	16	2.54	Protein disulfide-isomerase OS=Mus musculus OX=10090
6	1	SwissProt	d2::CP1A2_MOUSE	1053	63034	38	38	10	10	1.31	Cytochrome P450 1A2 OS=Mus musculus OX=10090
7	1	SwissProt	d2::ENPL_MOUSE	1008	103744	63	63	19	19	1.53	Endoplasmic reticulum protein OS=Mus musculus OX=10090 GN=Hsp90

同定タンパク質のリスト

(限定的だが)定量に関する情報

Protein sequence coverage: 27%

Matched peptides shown in **bold red**.

1 MDLVVFLALT LSCILLLSLN QSSSGRGLP PPTPLPIIG NFLQIDVNI
51 SQSPTNFSKA VGFVFDLVLG SKPTVILHGV EAVKEALIDR GEEFAGRGSP
101 FMAEKIIFGF GVVFNGNRRN KEMREFTMT LRNLGMSKRN IEDRVQEEAQ
151 CLVEELRATK GSPDFPFIL SCAPGVNCS IIFQNAFDYK DREVLILMDK
201 INENVKILSS FWLQVCNSFF SLIDVCFGSH HKIVNFNVL KSVILLRIKE
251 HKESLDVINF RFDIDYLIK QKQVNHIEQS EFSLENAST INDLFGATGE
301 TTSTTLRYAL LLLLYPDVT AKVQEEIDRV VGRHRSFCMQ DRSHMPYTD
351 MIHEVQRFID LLFTSLPHAV TCDIKFRYL IPKGTTVTS LSSVLHDSKE
401 FFPFEMFDFG HFLNGNGNFK KSDYFMFST GKRICAGEGL ARMELFLILT
451 TILQNEKLKS LVHPKEIDIT FVMNGFASLP PPVQLCFIL

同定ペプチド

修飾

d7783	98 - 105	577.9432	1153.8718	1153.6045	0.2673	0	35	0.0042	U R.GSFPMAEK.I
d7785	98 - 105	577.9534	1153.8922	1153.6045	0.2877	0	38	0.013	U R.GSFPMAEK.I
d8320	98 - 105	585.9059	1169.7972	1169.5994	0.1978	0	28	0.012	U R.GSFPMAEK.I + Oxidation (M)
d8321	98 - 105	585.9135	1169.8125	1169.5994	0.2131	0	25	0.035	U R.GSFPMAEK.I + Oxidation (M)
d8322	98 - 105	585.9243	1169.8341	1169.5994	0.2347	0	25	0.029	U R.GSFPMAEK.I + Oxidation (M)
d8323	98 - 105	585.9384	1169.8623	1169.5994	0.2629	0	30	0.0076	U R.GSFPMAEK.I + Oxidation (M)
d8324	98 - 105	585.9461	1169.8777	1169.5994	0.2877	0	38	0.013	U R.GSFPMAEK.I
d8325	98 - 105	585.9766	1169.9386	1169.5994	0.2877	0	38	0.013	U R.GSFPMAEK.I
d8326	98 - 105	586.0058	1169.9970	1169.5994	0.2877	0	38	0.013	U R.GSFPMAEK.I
d8327	126 - 132	521.3753	1040.7361	1040.7361	0.1001	0	22	0.004	U R.GSFPMAEK.I + Oxidation (M)
d8328	145 - 157	868.6286	1735.2427	1734.8402	0.4025	0	61	2.4e-05	U R.VQEAQCLVEELR.K
d8329	262 - 270	739.5419	1477.0693	1476.8108	0.2585	0	30	0.011	U R.DFIDYLLIK.Q
d8330	262 - 270	739.5658	1477.1170	1476.8108	0.3062	0	29	0.027	U R.DFIDYLLIK.Q
d8331	262 - 270	739.5844	1477.1542	1476.8108	0.3434	0	38	0.0026	U R.DFIDYLLIK.Q
d8332	262 - 270	739.6176	1477.2207	1476.8108	0.4100	0	63	6.2e-06	U R.DFIDYLLIK.Q
d8333	262 - 270	739.6731	1477.3317	1476.8108	0.5209	0	34	0.0055	U R.DFIDYLLIK.Q
d8334	308 - 315	618.0392	1234.0638	1233.8204	0.2335	0	43	0.0056	U R.VALLILIK.Y

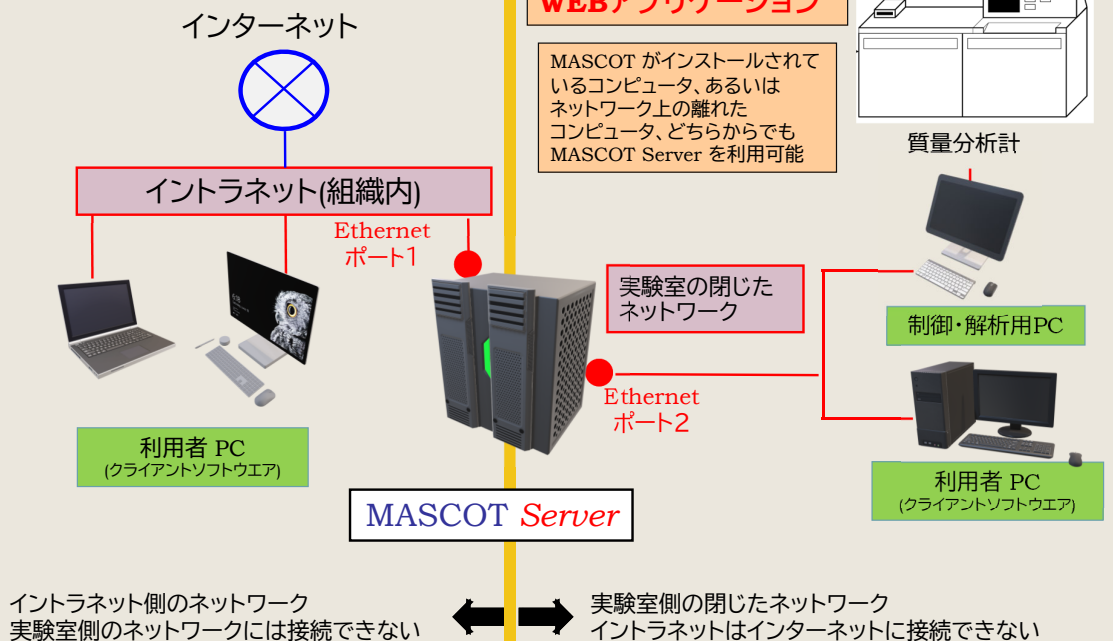
R.GSFPMAEK.I + Oxidation (M)

MASCOT はウェブアプリケーション

→ 2-1 (P.4~)

- MASCOT Serverとは
 - MASCOT Serverでできる事・まとめ
 - MASCOT Serverはウェブアプリケーション
 - ネットワーク上でMASCOT Serverを指定

- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ



ネットワーク上でMASCOT Serverを指定

→ 4-1 (P.12～)

- ・自分自身あるいは別のコンピューターからMASCOT Server を**URL**で指定

URLの記述：

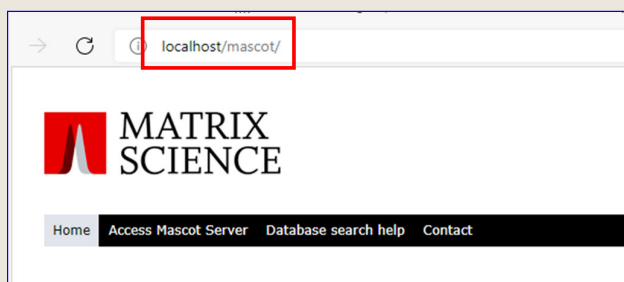
http://(computer 名)/mascot/

例：<http://mascotserver/mascot/>

http:// (IP アドレス)/mascot/

例：<http://192.168.100.222/mascot/>

- ・Serverがインストールされているコンピューターでは「**localhost**」を利用可能(下図)



- MASCOT Serverとは
 - MASCOT Serverでできる事・まとめ
 - MASCOT Serverはウェブアプリケーション
 - ネットワーク上でMASCOT Serverを指定
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

検索方法

→ 4 (P.12～)

- ・検索を実行する方法：以下3種類

変換済みファイルをWEBブラウザで検索

Daemonを使ってrawデータを直接検索

質量分析装置メーカーのソフトウェアから直接検索

- ・検索方法により準備するデータが異なる

検索方法	準備するデータ
WEBブラウザ	mgf (またはmzML)
Daemon	mgf ,mzML,rawデータ
質量分析装置メーカーのソフトウェア経由	データ準備の必要なし

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

MASCOT Serverとは

■ 検索方法

- 変換済みファイルをWEBブラウザで検索
- Daemonでrawデータを直接検索
- 質量分析装置メーカーのソフトウェアから直接検索

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

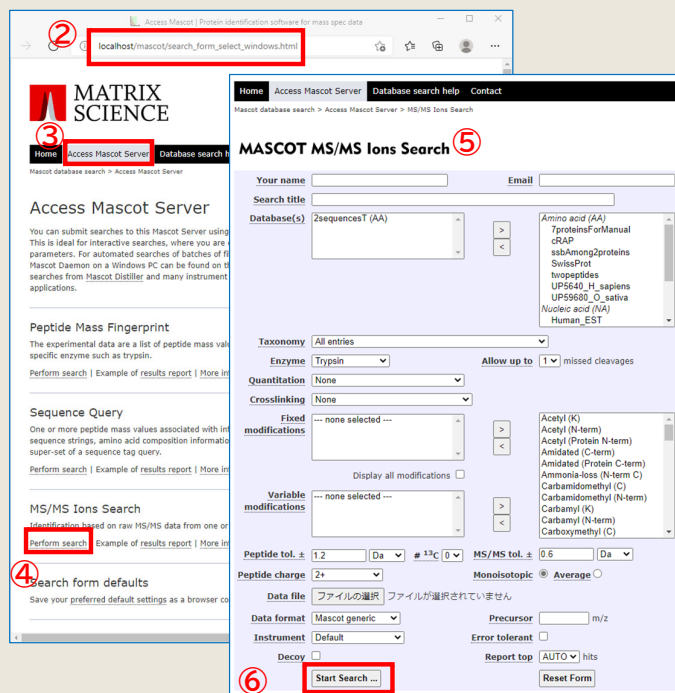
■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

変換済みファイルをWEBブラウザで検索

→ 4-2-2 (P.14~)



- ① ピークリストファイルを準備
- ② ブラウザを開きURLを指定してMASCOT Serverへアクセス
- ③ Home -> Access MASCOT Server
- ④ 「MS/MS Ions Search」の「Perform Search」
- ⑤ 検索パラメータ並びに入力データを指定
- ⑥ Start Search ボタンで検索開始

MASCOT Serverとは

■ 検索方法

- 変換済みファイルをWEBブラウザで検索
- Daemonでrawデータを直接検索
- 質量分析装置メーカーのソフトウェアから直接検索

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

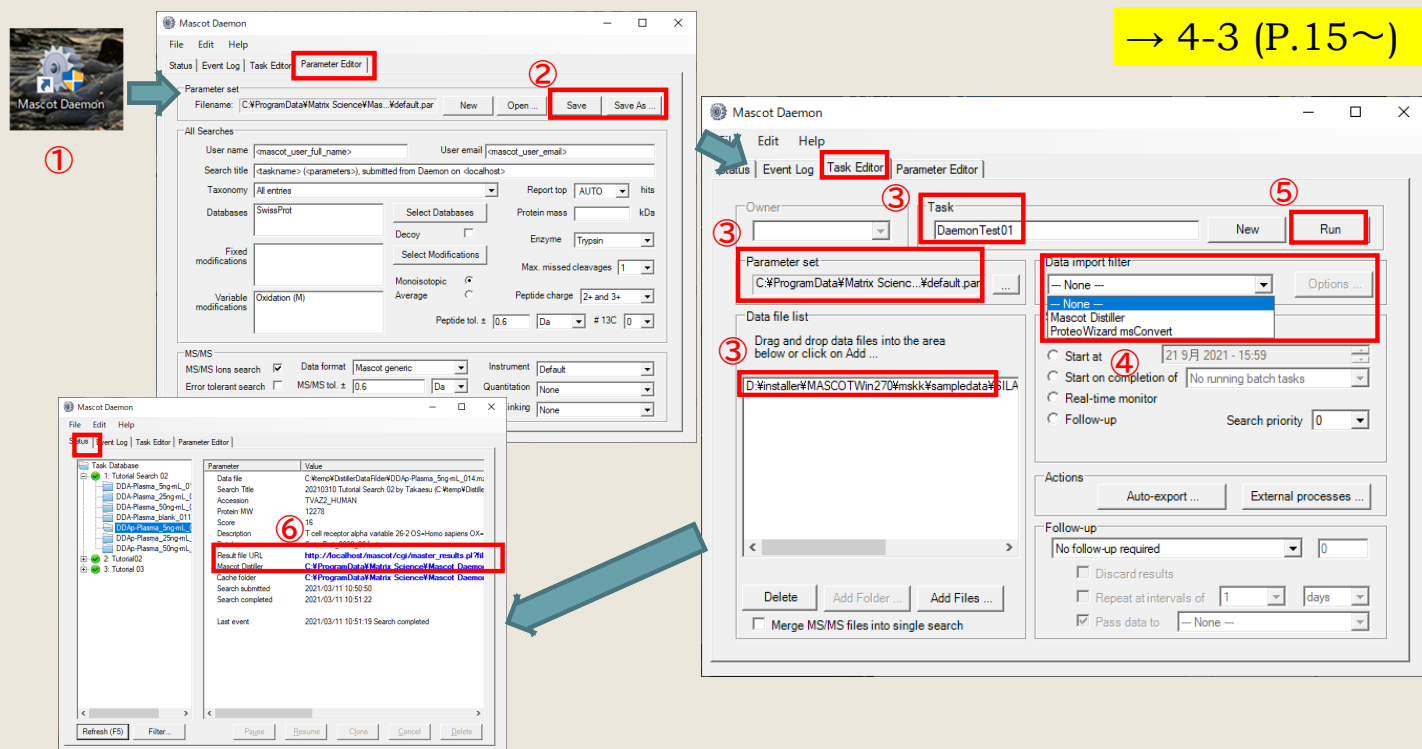
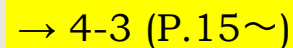
■ 管理プログラム・カスタマイズ

Daemonでrawデータを直接検索

→ 4-3 (P.14~)

- ① Daemonを起動
- ② Parameter Editorタブで検索条件を指定し、条件をファイルで保存するため「Save」または「Save As」
- ③ Task Editor タブでタスク名、検索対象のファイル、パラメーターファイルを指定
- ④ 「Data import filter」で変換プログラム [MASCOT Distiller] または [ProteoWizard msConvert]を選択、隣の「options」ボタンでファイルフォーマットなどを指定
- ⑤ Task Editorで「Run」を押すと検索実行
- ⑥ Statusタブに進捗が表示。検索完了すると結果へのURLが表示されるのでクリック

補足説明：Daemonでrawデータを直接検索_図



質量分析装置メーカーのソフトウェアから直接検索

→ 4-4 (P.17~)

- ・ 質量分析装置には、概ね解析ソフトウェアがついています。
- ・ 装置メーカーソフトウェアとMASCOT Serverは基本的に別のコンピュータにインストールされています。
- ・ 装置メーカーソフトウェア上で、MASCOT ServerをURL指定します。
- ・ 接続の前に、装置メーカーソフトウェアがインストールされているコンピュータからWEBブラウザでMASCOT Serverへアクセスできるかのテストを行ってください。
- ・ 装置メーカーソフトウェアからMASCOT Serverに行った検索はすべてMASCOT Serverにも検索ログが残ります。

ご利用方法の詳細は質量分析装置メーカーのサポートにお問い合わせください。

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

→ 3 (P.7~)

MASCOT 検索を行う前に、装置のrawデータを**判読可能で必要な情報のみ抽出したデータに変換**する必要があります。MASCOT Server 自身ではrawデータを変換することはできません。

抽出されている情報

対応する主なファイルフォーマット: mgf

Rawデータ変換プログラム

→ 3-1-2 (P.9~)

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

- 抽出されている情報

- 対応する主なファイル

フォーマット: mgf

- rawデータ変換プログラム

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

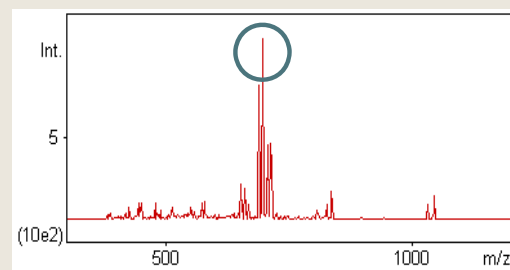
■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

・ 測定データから検索に必要な情報を抽出

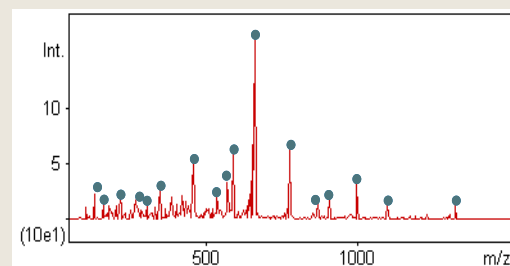
- **プレカーサーマスペクトル**

→ ペプチドの質量 (m/z と電荷)



- **プロダクトイオンマスペクトル**

→ ノイズをカットし、フラグメント由来のピークを選択



- ペプチド同定には RT 情報不要 (一部解析除く)

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
 - 抽出されている情報
 - 対応する主なファイルフォーマット : mgf
 - rawデータ変換プログラム
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

対応する主なファイルフォーマット : mgf

→ 3-1-3 (P.9~)

BEGIN IONS

TITLE=1: Scan 2 (rt=182.28)

PEPMASS=817.44 123456

CHARGE=2+

566.70161 445127.71

734.56885 253205.82

1092.2019 445679.33

.....

1426.779 1569294.9

1847.8916 3574730.8

END IONS

BEGIN IONSで始まり、END IONSで終わる

プレカーサーマスペクトル(MS1)由来の情報

① ペプチドの質量を計算
 $(817.44 - 1.0067) * 2 = 1632.866$

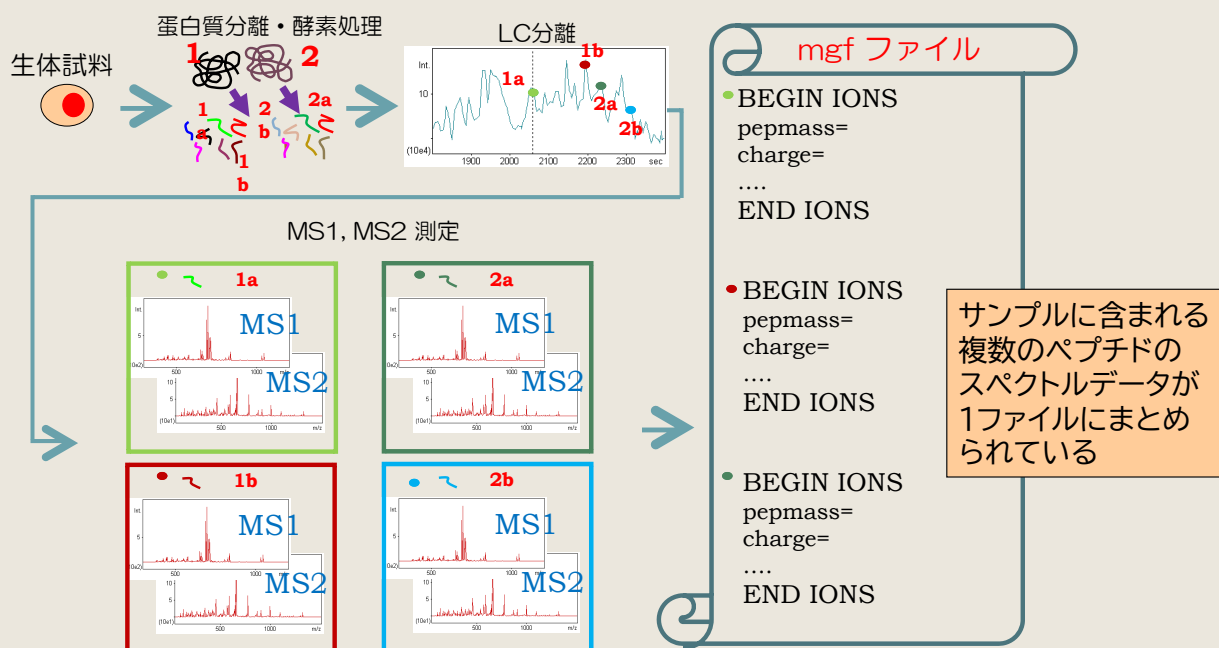
プロダクトイオンマスペクトル(MS2)由来の情報

② プロダクトイオンマスペクトル、
ノイズ等をカット。理論フラグメント質量とのマッチングに利用

* mgf以外の対応フォーマットについては、3-1-4(P.10~) 参照のこと

補足説明 : mgfは複数queryを1ファイル内に含む

→ 3-1-3 (P.10~)



rawデータ変換プログラム

→ 3-2 (P.11～)

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

- 抽出されている情報
- 対応する主なファイルフォーマット : mgf
- rawデータ変換プログラム

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

ProteoWizard msconvert

- プロテオームデータ解析向け、オープンソースでクロスプラットフォームのツールまたはライブラリ
- msconvertGUI.exeというプログラムを起動してプログラム内でmgfを作成したり、MASCOT Daemonと組み合わせて利用する事も可

MASCOT Distiller

- 弊社が販売しているソフトウェア。MASCOT Server/Daemonといった弊社取り扱いソフトウェアとの連携の良さや多価フラグメントピークデータも検索に組み込む事ができるのが特徴

質量分析装置付属のソフトウェア

- ほとんどの質量分析装置付属のソフトウェアではmgfファイル(ピークリストのテキストファイル)やmzMLフォーマットに出力が可能

検索パラメーター

→ 5 (P.18～)

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

- 参照先の配列データベースの各エントリーについて、理論ピークリストを作成する際、並びにマッチングの条件指定に検索パラメーターの情報を利用
- MASCOT Server local版では各種カスタマイズが可能
- 「データベースにないタンパク質・ペプチド」は見つからない
- データベースの種類と特徴

検索パラメーター一覧

カスタマイズ可能な項目

データベース

検索パラメーター 一覧

→ 5-3 (P.27~)

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

- 検索パラメーター 一覧
- カスタマイズ可能な項目
- データベース

結果画面

ペプチド同定

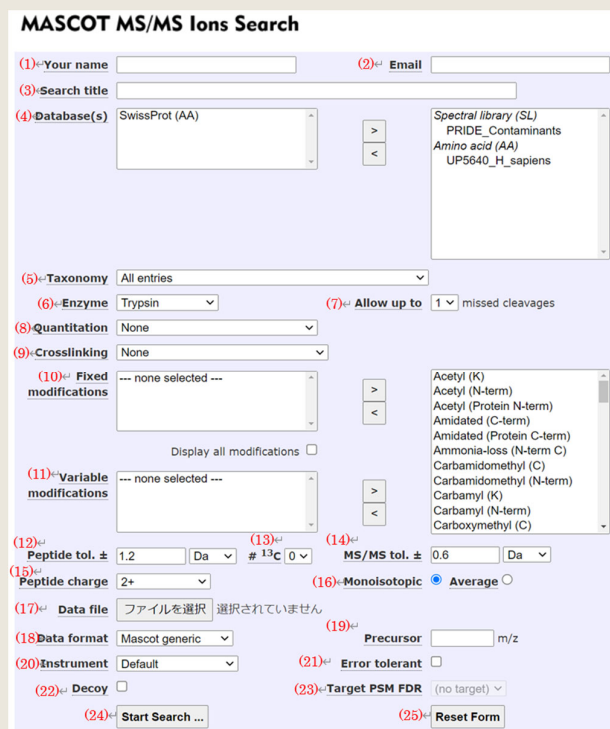
同定ペプチドの検証

タンパク質同定

MIS 検索のオプション

管理プログラム・カスタマイズ

19



補足説明：検索パラメーター各項目について

No	項目名	説明	No	項目名	説明	No	項目名	説明
1	Your name	ユーザー名	10	Fixed modifications	修飾、対象のすべてのアミノ酸について指定した内容に質量が変更	19	Precursor	[通常使用されない]ピークリストにprecursor のm/z情報がない場合に割り当てる値
2	Email	電子メール。但しlocal版ではメモとして使用可能	11	Variable modifications	修飾、修飾がつくパターンとつかないパターンの両方を考慮	20	Instrument	理論フラグメント作成を行うイオンシリーズの定義セット
3	Search title	検索タイトル	12	Peptide tol. ±	ペプチドの実測値と理論値の誤差範囲	21	Error tolerant	拡張2段階検索の実施
4	Database	検索対象のデータベース	13	#13C	13Cを含むペプチドの補正	22	Decoy	Decoyへの計算とFDR算出の実施
5	Taxonomy	生物種絞り込み	14	MS/MS tol. ±	フラグメントの実測値と理論値の誤差範囲	23	Target PSM FDR	同定基準としてのPSM FDRの設定値
6	Enzyme	タンパク質の切断パターン	15	Peptide charge	[通常は使用されない]ピークリストに電荷情報がない場合に割り当てる電荷の情報	24	Start Search	検索開始
7	Allow up to	Enzyme設定について、切断箇所と認定された箇所を見逃し連結したペプチドを作成する事ができるが、何度まで見逃すことを許容するか	16	Monoisotopic/Average	アミノ酸の質量計算をMonoisotopicで行うか、Averageで行うか	25	Reset Form	パラメータ設定を初期地に戻す
8	Quantitation	タンパク質の定量計算	17	Data file	検索queryとなる入力データ			
9	Crosslinking	リンカー又は共有結合で結合したペプチドの検索	18	Data format	入力データのフォーマット			

好ましい
デフォルト設定に
ついて
→ 口頭で説明

20

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
 - 検索パラメーター一覧
 - カスタマイズ可能な項目
 - データベース
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

Databases Taxonomy Enzyme Quantitation Crosslinking Modification Instrument

ほとんどの設定は**Configuration Editor**
(Home -> configuration Editor)で行います。各設定画面は
「**13. MASCOT Serverのカスタマイズ(P.113～)**」をご覧ください。

- MASCOT Serverとは
- 検索方法
- ネットワーク・検索方法
- 検索パラメーター
 - 検索パラメーター一覧
 - カスタマイズ可能な項目
 - データベース
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

- ・大きく分けて3種類のデータベースを使用可能
 - **AA** : **Amino Acid** : アミノ酸配列
 - **NA** : **Nucleic Acid** : 核酸配列
 - **SL** : **Spectral Library** : ピークリスト/過去に測定したペプチドのフラグメントピーク
 - ・使用しているデータベースの更新や新規追加は「**Database Manager**」で
 - ・「データベースにないタンパク質・ペプチド」は見つからない
 - ・ユーザーが適切なデータベースを選択しなければならない
- * Uniprotで生物種別に分けられたデータベース
- ・1st チョイスは「**SwissProt**」。→だめなら「UP_NNNN_BBBBBB」
 - ・MASCOTで準備しているPredefined データベース がある(次頁)

補足説明 : Predefined Database について 1

■ SwissProt : AA

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

Uniprot データベースの中の1つ、タンパク質配列を集めたデータベース。各エントリーに対して、機能・ドメイン構造・修飾・バリエーション・論文情報・他データベースへのリンク、など、精査されたアノテーション情報が手動で付与されています。配列の冗長性はできるだけ無いように調整されていて、2023年2月時点で **568,744** 件のデータが登録。**最初に選ぶデータベースとして最適**です。

■ Uniprot proteomes : AA → MASCOTデータベースでは UP_N_B と表記 (Nは番号、Bは生物種)

<https://www.uniprot.org/>

手動アノテーションされた上記 Uniprotに加え、自動かつ精査無しのデータベース**TrEMBL** の内容も併せたデータベース。TrEMBLの方がSwissProtに比べ圧倒的にエントリー数が多く、2023年2月版で **219,174,961** 件のデータが登録。SwissProtのみに比べ配列のカバー範囲が広く、**SwissProtでマッチしなかった場合の次の選択肢に最適**です。

■ NCBIprot (旧名称 NCBItr) : AA

NCBI(National Center for Biotechnology Information)で公開されているタンパク質データベース「nr」。「nr」とは「non-redundant」の略ですが、実際にはほぼ同じ配列のデータが数多く登録されており、それが膨大なエントリー数の要因となっています。現在は使用をお勧めしていません。

補足説明 : Predefined Database について 2

■ XXXXX EST (Human ESTなど) : NA

ESTとはExpressed Sequence Tag の事で、cDNAライブラリの部分配列にあたります。mRNAレベルでの塩基配列情報が含まれているデータベースで、タンパク質データベースでは見つからないようなペプチド配列などが検出される事が期待されます。一方でマッチしても次のステップに進みにくいため、目的をもって利用される事をお勧めします。

■ NIST XXXX YYYY (NIST Human HCD など) : SL

ピークリスト(過去に測定したペプチドのフラグメントピークの情報)データベースで、**NIST**(National Institute of Standards and Technology)にて公開しているものです。

■ PRIDE XXXXX (PRIDE Human など) : SL

ピークリスト(過去に測定したペプチドのフラグメントピークの情報)データベースで、data repository siteの**PRIDE**(Proteomics IDentification database)にて公開しているものです。

検索前のまとめ：具体的な検索方法の確認

URL http :// /mascot/(index.html)


検索方法	データの種類	rawデータ変換方法	その他
<ul style="list-style-type: none"> □WEBブラウザ □Daemon □メーカーのソフトウェア () 	<ul style="list-style-type: none"> □mgf □raw □その他 (mzMLなど) □ 気にしなくて良い (ソフトから直接検索) 	<ul style="list-style-type: none"> □最初からmgf □ProteoWizard □Distiller □その他 () 	<ul style="list-style-type: none"> • [注意事項など]

WEBブラウザ → スライド 9
 Daemon → スライド 10
 メーカーのソフトウェア → スライド 12

25

表示例の検索について

→ 7-1 (P.48～)



Mascot Server

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
 - 検索query数
 - 同定ペプチド数
 - 同定タンパク質(グループ)数
 - 同定タンパク質の具体的な内容
 - タンパク質にアサインされているペプチド
 - スペクトルとペプチド理論値のマッチング
 - 結果ファイルの出力
- ペプチド同定
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

[入力データ]
query数 :
 33,191


[同定数]
同定スペクトル数 :
 2,812

同定タンパク質数 :
 480 (448グループ)

設定項目	設定内容
Enzyme	Trypsin/P
Fixed Modification	iTRAQ4plex(K), iTRAQ 4plex(N-term), Methylthio(C)
Variable Modification	Acetyl(Protein N-term), Gln->pyro-Glu(N-term Q), Oxidation(M)
Peptide tol.±	0.9 Da
Fragment mass tolerance ±	0.6 Da
Missed cleavage	1
Instrument	ESI-TRAP

[公開サーバー]
http://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981139.dat
[ローカルサーバー]
http://localhost/mascot/master_results_2.pl?file=../data/F981139.dat

26



Mascot Server

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

ペプチド同定

同定ペプチドの検証

タンパク質同定

MIS 検索のオプション

管理プログラム・カスタマイズ

27

結果画面で確認したい事

[入力データ]

query数：
33,191

[同定数]

同定スペクトル数
2,812
同定タンパク質数
480 (448グループ)

1. 検索query数

2. 同定ペプチド数


3. 同定タンパク質(グループ)数

4. 同定タンパク質の具体的な内容

5. タンパク質にアサインされているペプチド[修飾など]

6. スペクトルとペプチド理論値のマッチング状況

7. 結果ファイルの出力



Mascot Server

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

ペプチド同定

同定ペプチドの検証

タンパク質同定

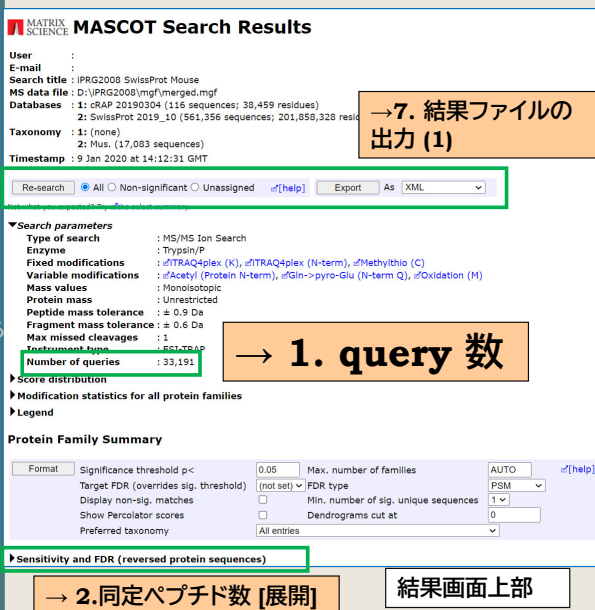
MIS 検索のオプション

管理プログラム・カスタマイズ

28

検索query数,各項目の場所

→ 7-2 (P.48～)



MASCOT Search Results

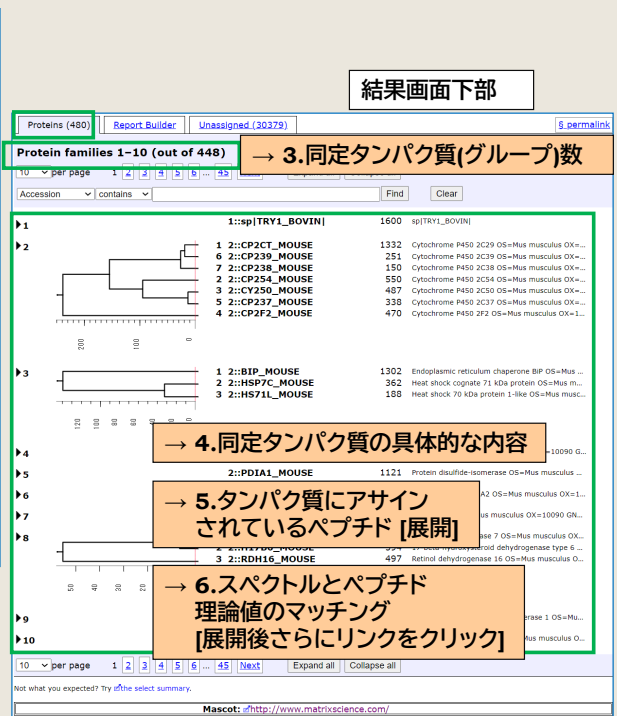
User: [blank]
E-mail: [blank]
Search title: IPRG2008 SwissProt Mouse
MS data file: D:\IPRG2008\mgf\merged.mgf
Databases: 1: cRAP 20190304 (116 sequences; 38,459 residues)
2: SwissProt 2019_10 (561,356 sequences; 201,658,328 residues)
Taxonomy: 1: (none)
2: Mus, (17,083 sequences)
Timestamp: 9 Jan 2020 at 14:12:31 GMT

→ 7. 結果ファイルの出力 (1)

→ 1. query 数

→ 2. 同定ペプチド数 [展開]

結果画面上部



Protein families 1-10 (out of 448)

→ 3. 同定タンパク質(グループ)数

→ 4. 同定タンパク質の具体的な内容

→ 5. タンパク質にアサインされているペプチド [展開]

→ 6. スペクトルとペプチド理論値のマッチング [展開後さらにリンクをクリック]

結果画面下部



Mascot Server

MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

29

同定ペプチド数

→ 7-2-9 (P.53~)

Timestamp : 9 Jan 2020 at 14:12:31 GMT

Re-search ☒ All ☐ Non-significant ☐ Unassigned

Not what you expected? Try [the select summary](#).

Search parameters
Score distribution
Modification statistics for all protein families
Legend

Protein Family Summary

Format Significance threshold p< Target FDR (overrides sig. threshold) Display non-sig. matches Show Percolator scores Preferred taxonomy

Sensitivity and FDR (reversed protein sequences)

	Target	Decoy	FDR
Protein family members	302	11	3.64%
PSMs	1821	18	0.99%

Decoy results are available in [the decoy report](#).

Proteins (302) [Report Builder](#) [Unassigned \(31370\)](#)

Protein families 1-10 (out of 280)

10 per page 1 2 3 4 5 ... 28 [Next](#) [Expand all](#) [Collapse all](#)

Accession contains Find Clear

1::sp|TRY1_BOVIN| 1153 sp|TRY1_BOVIN|

“Sensitivity and FDR”をクリックして展開すると、同定ペプチド数とFDRが表示される

Target : 通常データベース
Decoy : Decoyデータベース(後述)

Sequence : 同一ペプチドにマッチするスペクトルを1つにまとめてカウント

PSMs : 同一ペプチドにマッチするスペクトルを1つにまとめないでカウント



Mascot Server

MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

30

同定タンパク質(グループ)数

→ 7-2-10 (P.54~)

Legend

Protein Family Summary

Format Significance threshold p< 0.05 Max. number of families AUTO [\[help\]](#)
Target FDR (overrides sig. threshold) 1% FDR type PSM
Display non-sig. matches Min. number of sig. unique sequences 1
Show Percolator scores Dendrograms cut at 0
Preferred taxonomy All entries

Sensitivity

Proteins (480) [Report Builder](#) [Unassigned \(31370\)](#)

Protein families 1-10 (out of 448)

10 per page 1 2 3 4 5 ... 45 [Next](#) [Expand all](#) [Collapse all](#)

Accession contains Find Clear

1::sp|TRY1_BOVIN| 1600 sp|TRY1_BOVIN|

1 2::CP2CT_MOUSE
6 2::CP239_MOUSE
7 2::CP238_MOUSE
2 2::CP254_MOUSE
3 2::CY250_MOUSE
5 2::CP237_MOUSE
4 2::CP2F2_MOUSE

1332 Cytochrome P450 2C29 OS=Mus musculus
251 Cytochrome P450 2C39 OS=Mus musculus
150 Cytochrome P450 2C38 OS=Mus musculus
550 Cytochrome P450 2C54 OS=Mus musculus
487 Cytochrome P450 2C50 OS=Mus musculus
338 Cytochrome P450 2C37 OS=Mus musculus
470 Cytochrome P450 2F2 OS=Mus musculus

Proteins の後ろのある括弧内の数字が同定タンパク質数

“out of N”の数字Nが同定タンパク質のグループ数
*「グループ」の意味については後述

Mascot Server
 Mascot Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

■ ペプチド同定

■ 同定ペプチドの検証

■ タンパク質同定

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

31

同定タンパク質の具体的な内容

→ 7-2-10 (P.54～)

Proteins (480)

Report Builder

Unassigned (30379)

Protein families 1~10 (out of 448)

10 per page

1 2 3 4 5 6 ... 45 Next

Expand all

Collapse all

Accession

contains

Find

Clear

1 2 3 4 5 6 7 8 9 10

1::sp|TRY1_BOVIN|

1600 sp|TRY1_BOVIN|

1332 Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

251 Cytochrome P450 2C39 OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

150 Cytochrome P450 2C38 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=100

550 Cytochrome P450 2C54 OS=Mus musculus OX=10090 GN=Cyp2c54 PE=100

487 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100

338 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=100

470 Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=100

2::CP2CT_MOUSE

6 2::CP239_MOUSE

7 2::CP238_MOUSE

2 2::CP254_MOUSE

3 2::CY250_MOUSE

5 2::CP237_MOUSE

4 2::CP2F2_MOUSE

2::BIP_MOUSE

2 2::HSP70_MOUSE

3 2::HSP71_MOUSE

2::CYB5_MOUSE

1202 Cytochrome b5 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

1121 Protein disulfide-isomerase OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

1053 Cytochrome P450 1A2 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=100

1008 Cytochrome P450 2C54 OS=Mus musculus OX=10090 GN=Cyp2c54 PE=100

1009 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100

1010 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=100

1011 Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=100

2::MGST1_MOUSE

860 Microsomal glutathione S-transferase 1 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

768 60S ribosomal protein L7a OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

2::RL7A_MOUSE

768 60S ribosomal protein L7a OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

同定タンパク質を10件ずつ表示。
表示方法は設定で変更可能

同定タンパク質のうち、
シェアペプチドを持つタンパク
質をグループ化
*グループ化については後述

各ランクの▲マークをクリックするとより詳細な情報
を表示 [→次スライド以降]

タンパク質スコアが高い順に結果表示。
順位付けにはグループ内の最高スコアを採用

Summary画面：同定タンパク質とアサインペプチド ①

→ 7-2-10 (P.54～)

Accession

contains

Find

Clear

1 2 3 4 5 6 7 8 9 10

1::sp|TRY1_BOVIN|

1600 sp|TRY1_BOVIN|

1332 Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

251 Cytochrome P450 2C39 OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

150 Cytochrome P450 2C38 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=100

550 Cytochrome P450 2C54 OS=Mus musculus OX=10090 GN=Cyp2c54 PE=100

487 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100

338 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=100

470 Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=100

2::CP2CT_MOUSE

6 2::CP239_MOUSE

7 2::CP238_MOUSE

2 2::CP254_MOUSE

3 2::CY250_MOUSE

5 2::CP237_MOUSE

4 2::CP2F2_MOUSE

2::BIP_MOUSE

2 2::HSP70_MOUSE

3 2::HSP71_MOUSE

2::CYB5_MOUSE

1202 Cytochrome b5 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

1121 Protein disulfide-isomerase OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

1053 Cytochrome P450 1A2 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=100

1008 Cytochrome P450 2C54 OS=Mus musculus OX=10090 GN=Cyp2c54 PE=100

1009 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100

1010 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=100

1011 Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=100

2::MGST1_MOUSE

860 Microsomal glutathione S-transferase 1 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100

768 60S ribosomal protein L7a OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

2::RL7A_MOUSE

768 60S ribosomal protein L7a OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100

Threshold (0): 0

Cut

2.1 2::CP2CT_MOUSE

1332 61419 76 (76) 13 (13) 2.00 Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=100 SV=2

2.2 2::CP239_MOUSE

550 60887 27 (27) 8 (8) 0.88 Cytochrome P450 2C39 OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100 SV=1

2.3 2::CY250_MOUSE

487 61128 27 (27) 10 (10) 1.20 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100 SV=2

2.4 2::CP2F2_MOUSE

470 59267 32 (32) 12 (12) 0.37 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=100 SV=2

2.5 2::CP237_MOUSE

338 60590 22 (22) 4 (4) 0.37 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=100 SV=2

2.6 2::CP239_MOUSE

251 60856 13 (13) 4 (4) 0.37 Cytochrome P450 2C39 OS=Mus musculus OX=10090 GN=Cyp2c39 PE=100 SV=2

2.7 2::CP238_MOUSE

150 61356 9 (9) 4 (4) 0.37 Cytochrome P450 2C38 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=100 SV=2

Redisplay

All

None

▼134 peptide matches (43 non-duplicate, 91 duplicate)

Auto-fit to window

Query Dups

Observed

Mr (expt)

Mr (calc)

Delta M

Score

Expect

Rank

U

1

2

3

4

5

6

7

Peptide

3466

503.3162

1004.6178

1004.5083

0.1095

0.31

0.015

1

U

R.MPTLEDK.T

3505

503.8846

1005.7547

1005.6093

0.1454

0.36

0.016

1

U

R.FSVQILK.N

4193

516.8977

1031.7808

1031.5369

0.2439

0.32

0.05

1

U

VQEEIDK.V

4447

521.2416

1560.7029

1559.8187

0.8842

0.59

1.2e-05

1

U

K.NISQSTFNK.A

4466

521.3753

1040.7361

1040.5810

0.1551

0.22

0.031

1

U

R.FTLMTLR.N + oxidation (M)

4705

525.4566

1573.3479

1572.7654

0.5824

0.71

1.3e-05

1

U

K.EALVDHGEFAGK.G

4741

526.3505

1050.6864

1050.5323

0.1541

0.34

0.005

1

U

R.CLVEELK.K

5544

540.3247

1078.6349

1078.5385

0.0964

0.54

0.0002

1

U

R.ICAGEGLK.M

同定タンパク質の情報
(詳しい情報は後述)

同定ペプチドの情報
(詳しい情報は後述)

Summary画面：同定タンパク質とアサインペプチド ②

→ 7-2-10 (P.56～)

		Score	Mass	Matches	Sequences	emPAI	
<input checked="" type="checkbox"/> 2.1	#2::CP2CT_MOUSE	1332	61419	76 (76)	13 (13)	2.00	Cytochrome P450 2C29 OS=Mus m
<input checked="" type="checkbox"/> 2.2	#2::CP254_MOUSE	550	60887	27 (27)	8 (8)	0.88	Cytochrome P450 2C54 OS=Mus m
<input checked="" type="checkbox"/> 2.3	#2::CY250_MOUSE	487	61128	27 (27)	10 (10)	1.20	Cytochrome P450 2C50 OS=Mus m
<input checked="" type="checkbox"/> 2.4	#2::CP2F2_MOUSE	470	59267	32 (32)	12		Cytochrome P450 2C50 OS=Mus musculus SV=2
<input checked="" type="checkbox"/> 2.5	#2::CP237_MOUSE	338	60590	22 (22)			
<input checked="" type="checkbox"/> 2.6	#2::CP239_MOUSE					0.37	Cytochrome P450 2C39 OS=Mus m
<input checked="" type="checkbox"/> 2.7	#2::CP238_MOUSE					0.37	Cytochrome P450 2C38 OS=Mus m

Redisplay All None

タンパク質のAccession (ID)部分を
クリックすると、より詳細な情報が表示
されるProtein View 画面となる

表示項目	説明
Score	タンパク質のスコア。タンパク質にアサインされているペプチドのIon Scoreをもとに算出され、大きいほど信頼度の高いペプチドが多いことを示しますが、現在のMASCOTではProtein Scoreをもとに同定タンパク質を判定する事はしていません。
Mass	タンパク質の質量。データベースに登録されている配列情報から計算。
Matches	タンパク質にアサインされたquery数。()内の数字は、その中で同定基準を超えているもの。
Sequences	queryのうち同じペプチド配列にマッチしている内容を1つにまとめてカウントしたもの。()についてはMatchesと同じく同定基準を超えているもののみをカウント。
emPAI	Spectral Countingの1つであるemPAI。値が大きいほど量が多いという判断の基準となります。

Protein View 画面

→ 7-3 (P.59～)

タンパク質について、マッチしたペプチドの情報などを確認できます

右図内の緑で囲われた部分について、以降でさらに詳細な説明があります

MATRIX SCIENCE

MASCOT Search Results

Protein View: CP2CT_MOUSE

Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=1 SV=2

Database: SwissProt

Score: 1332

Monoisotopic mass (M₀): 61419

Calculated pI: 8.57

Taxonomy: Mus musculus

Sequence similarity is available as an [NCBI BLAST search of CP2CT_MOUSE against nr](#).

Search parameters

MS data file: D:\ipr2008\mgf\merged.mgf

Enzyme: Trypsin/P; cuts C-term side of KR.

Fixed modifications: ITRAQ4plex (K), ITRAQ4plex (N-term), Methylation (C)

Variable modifications: Acetyl (Protein N-term), Gln->pyro-Glu (N-term Q), Oxidation (M)

Protein sequence coverage: 27%

Matched peptides shown in **bold red**.

1 MDLVVFLALT LSCILLISIN RQSSGRKLP PGPTPLPIIG NFLQIDVKNI
51 SQSFTNFSKA YGVFTLLYLQ SEPTVILNGY EAVFEALIDR GEEFAGRGSF
101 PMAEKIINGF GVVFSGNKNR KEMRPTLMT LANLNGMERN IEDRVQREAG
151 CLVEKLKRTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLIMDK
201 INENVLKSS PWLQVNSFP SLIDYCPGSH HKIVKNFNLY KSYLLEKIK
251 HRESLDVNSP RQFDIDYLIK QGVNHIQS EFSLENLAST INDLFAGTE
301 TTSTTLRYAL LLLKYDVT AKVQREIDRV VGRHRSFCMQ DRSHMPYTD
351 MIHEVQRFDI LLTSLPHAV TCDIKFRYL IFKGTTVITS LSSVLHDSKE
401 FPHFVDFDPG HFLNGNGNFR KSDYMPFST GKRCAGEGL ARNELFLILT
451 TILQNFELKS LVHREIDIT FVNGFASLP PFVQLCFIPL

Unformatted sequence string: 490 residues (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass

Show ☒ matched peptides only ☐ predicted peptides also

Query	Start	End	Observed	Mr (expt)	Mr (calc)	Delta	M	Score	Expect	Rank	U	Peptide
#20242	49	59	781.0005	1559.9865	1559.8187	0.1677	0	76	1.1e-06	1	U	K.NISQSFNFSK.A
#20242	49	59	781.0128	1560.0110	1559.8187	0.1922	0	34	0.011	1	U	K.NISQSFNFSK.A
#20246	49	59	781.0421	1560.0696	1559.8187	0.2509	0	76	6.3e-07	1	U	K.NISQSFNFSK.A
#20247	49	59	781.0544	1560.0942	1559.8187	0.2754	0	57	4.4e-05	1	U	K.NISQSFNFSK.A
#20242	49	59	781.0857	1560.1569	1559.8187	0.3381	0	71	1.9e-06	1	U	K.NISQSFNFSK.A
#20243	49	59	521.0639	1560.1699	1559.8187	0.3512	0	49	0.00015	1	U	K.NISQSFNFSK.A
#20244	49	59	781.1019	1560.1892	1559.8187	0.3704	0	87	8.1e-08	1	U	K.NISQSFNFSK.A
#20247	49	59	781.1422	1560.2698	1559.8187	0.4511	0	90	2e-08	1	U	K.NISQSFNFSK.A
#20248	49	59	781.1516	1560.2887	1559.8187	0.4700	0	86	9.7e-08	1	U	K.NISQSFNFSK.A
#20249	49	59	781.1604	1560.3063	1559.8187	0.4875	0	74	8.5e-07	1	U	K.NISQSFNFSK.A
#20242	49	59	521.1230	1560.3494	1559.8187	0.5304	0	34	4.0e-05	1	U	K.NISQSFNFSK.A
#4447	49	59	521.2416	1560.7029	1559.8187	0.8842	0	59	1.2e-05	1	U	K.NISQSFNFSK.A
#1312	85	97	536.3181	1605.9324	1605.8232	0.1091	1	43	0.0014	1	U	K.EALIDGEEFAGR.G
#1312	85	97	536.3278	1605.9617	1605.8232	0.1384	1	46	0.0014	1	U	K.EALIDGEEFAGR.G
#1312	85	97	536.3355	1605.9846	1605.8232	0.1613	1	44	0.00077	1	U	K.EALIDGEEFAGR.G
#1312	85	97	536.3434	1606.0083	1605.8232	0.1850	1	46	0.002	1	U	K.EALIDGEEFAGR.G

34

Protein View: sequence coverage

→ 7-3 (P.61~)

Protein sequence coverage: 27%

Matched peptides shown in **bold red**.

```

1 MDLVVFLALT LSCILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
51 SQSFTNFSKA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
201 INENVKILSS PWLQVCNSFP SLIDYCPGSH HKIVKNFNYL KSYLLEKIKE
251 HKESLDVTNP RDFIDYYLIK QKQVNHIEQS EFSLENLAST INDLFGAGTE
301 TTSTTLRYAL LLLLKYPDVT AKVQEEIDRV VGRHRSPCMQ DRSHMPYTDA
351 MIHEVQRFID LLPTSLPHAV TCDIKFRKYL IPKGTTVITS LSSVLHDSKE
401 FPNPEMFDPG HFLNGNGNEFK KSDYFMPFST GKRICAGEGL ARMELFLILT
451 TILQNFKLKS LVHPKEIDIT PVMNGFASLP PPYQLCFIPL
    
```

タンパク質全長に対してマッチしたペプチドがどの部位にあたるのか、並びにその割合についての情報を表示

Protein sequence coverage : 全長に対するマッチペプチド残基数の割合

35

Protein View : matched peptides

→ 7-3 (P.62~)

Unformatted sequence string: [490 residues](#) (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass
Show ☒ matched peptides only ☐ predicted peptides also

Query	Start - End	Observed	Mr (expt)	Mr (calc)	Delta	M	Score	Expect	Rank	U	Peptide
20249	49 - 59	781.0005	1559.9865	1559.8187	0.1677	0	76	1.1e-06	1	U	K.NISQSFTNFSK.A
20252	49 - 59	781.0128	1560.0110	1559.8187	0.1922	0	34	0.011	1	U	K.NISQSFTNFSK.A
20256	49 - 59	781.0421	1560.0696	1559.8187	0.2509	0	76	6.3e-07	1	U	K.NISQSFTNFSK.A
20257	49 - 59	781.0544	1560.0942	1559.8187	0.2754	0	57	4.4e-05	1	U	K.NISQSFTNFSK.A
20262	49 - 59	781.0857	1560.1569	1559.8187	0.3381	0	71	1.9e-06	1	U	K.NISQSFTNFSK.A
20263	49 - 59	521.0639	1560.1699	1559.8187	0.3512	0	49	0.00015	1	U	K.NISQSFTNFSK.A
20264	49 - 59	781.1019	1560.1892	1559.8187	0.3704	0	87	8.1e-08	1	U	K.NISQSFTNFSK.A
20267	49 - 59	781.1422	1560.2698	1559.8187	0.4511	0	90	2e-08	1	U	K.NISQSFTNFSK.A
20268	49 - 59	781.1516	1560.2887	1559.8187	0.4700	0	86	9.7e-08	1	U	K.NISQSFTNFSK.A

マッチしたペプチドについて、アミノ酸残基順 (デフォルト設定) に並べたリストを表示

表示項目	説明
Unformatted sequence string	ハイパーリンクをクリックすると配列をコピーしやすくなるページが開きます。他アプリケーションで配列を使用したい場合に便利です。
Sort by	リストの並び順。残基番号、質量の昇順/降順。
Show	理論値と実測値がマッチしたペプチドのみをリストに表示させるか、マッチしなかった理論ピークも表示させるか
Query	Query番号。MASCOTでは入力データについて、ペプチドの質量に換算した際小さい順に番号が割り振られます。
Start-End	アミノ酸残基番号。
Observed	ピークリストファイルの m/z 。
Mr(expt)	ピークリストの値から計算されたペプチドの質量。

表示項目	説明
Mr(calc)	配列から計算されたペプチドの質量。
Delta	Mr(expt) - Mr(calc)。
M	実際に適用された Missed cleavage。
Score	Mascot Ion score、MS2ピーク実験値と理論値とのマッチング度合いを表します。
Expect	同定基準値とScoreから計算された期待値。
Rank	データベース中の候補ペプチドとマッチングを行った際、ペプチド配列とのマッチングが全体の中での何位であったか。
U	Uマークがついている場合、他のエントリーと共有されていないUnique ペプチドであることを意味します。
Peptide	ペプチド配列。修飾も含まれる場合は併せて表示。

36

Mascot Server

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

検索query数

同定ペプチド数

同定タンパク質(グループ)数

同定タンパク質の具体的な内容

タンパク質にアサインされているペプチド

スペクトルとペプチド理論値のマッチング

結果ファイルの出力

ペプチド同定

同定ペプチドの検証

タンパク質同定

MIS 検索のオプション

管理プログラム・カスタマイズ

37

タンパク質にアサインされているペプチド

→ 7-2-10 (P.56~)

▼134 peptide matches (43 non-duplicate, 91 duplicate)

☒ Auto-fit to window

Query	Dupes	Observed	Mr(expt)	Mr(calc)	Delta	M	Score	Expect	Rank	U	1	2	3	4	5	6	7	Peptide
3466		503.3162	1004.6178	1004.5083	0.1095	0	31	0.015	1	U								R.MPTLEDR.T
3505	4	503.8846	1005.7547	1005.6093	0.1454	0	36	0.016	1	U								R.FSVQILR.N
4193		516.8977	1031.7808	1031.5369	0.2439	0	32	0.05	1									VQEEIDR.V
4447										U								K.NISQSFNFASK.A
4466																		R.FTLMTLR.N + Ox
4705																		K.EALVDHGEEFAGR.

Query番号部分をクリックすると、スペクトルとペプチド理論値のマッチングを確認できる PeptideView 画面となる

項目名	説明	項目名	説明
Query	Query番号。MASCOTでは入力データについて、ペプチドの質量に換算した際小さい順に番号が割り振られます。	Score	Mascot Ion scoreで、実験値のMS2ピークと理論値とのマッチング度合いを表します。
Dupes	同じペプチド配列でかつ修飾や電荷も同じ結果にマッチしたqueryは、スコアが最も高いもののみ表示されそれ以外はDupes欄に格納。数字はqueryがいくつまとめられているかを表します。	Expect	同定基準値とScoreから計算された期待値。 デフォルト設定の場合0.05以下なら同定。
Observed	実験値側のペプチドの m/z。	Rank	データベース中の候補ペプチドとマッチングを行った際、表示されているペプチド配列とのマッチングが全体の中での何位であったか。
Mr(expt)	m/zと電荷から計算された、実験値側のペプチドの質量。	U	Uマークがついている場合、他のエントリーと共有されていないUnique ペプチドであることを意味します。
Mr(calc)	配列から計算された、理論値側のペプチドの質量。	数字	数字はファミリーに属するタンパク質と連動しています。■印があるタンパク質に該当ペプチドがアサインされていることを示します。
Delta	Mr(expt) - Mr(calc)。	Peptide	ヒットしたペプチド配列。修飾や前後のアミノ酸残基なども併せて表示。
M	Missed cleavageが実際に適用された数。		

Mascot Server

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

検索query数

同定ペプチド数

同定タンパク質(グループ)数

同定タンパク質の具体的な内容

タンパク質にアサインされているペプチド

スペクトルとペプチド理論値のマッチング

結果ファイルの出力

ペプチド同定

同定ペプチドの検証

タンパク質同定

MIS 検索のオプション

管理プログラム・カスタマイズ

38

スペクトルとペプチド理論値のマッチング

→ 7-4 (P.63~)

ペプチドについて、スペクトルとのマッチング情報などを確認できます [Peptide view 画面]

右図内の緑で囲われた部分について、以降でさらに詳細な説明があります

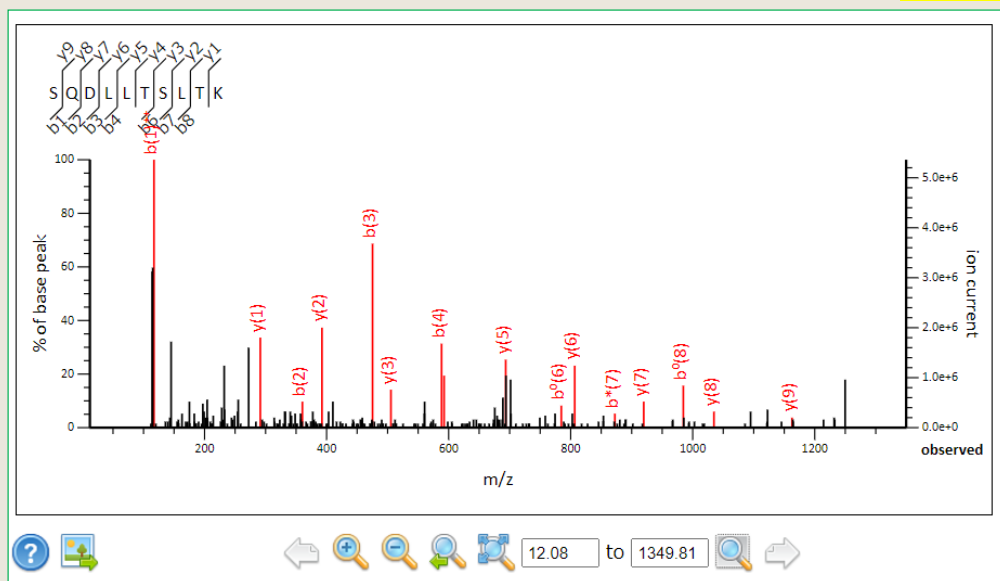
Mascot Search Results

Peptide View

MS/MS Fragmentation of **EQQLTSLTK**
Found in **CP2F2_MOUSE** in **SwissProt**, Cytochrome P450 2F2 O5=Mus musculus OX=10090 GN=Cyp2f2 PE=1 SV=1
Match to Query 15524: 1392.862848 from(697.438700,2+)
Title: Locust 1.559.3
Data file D:\K0602000\img\merged.mgf

Peptide View : スペクトルベースのマッチング確認

→ 7-4 (P.65~)



入力データのスペクトルベースで見たマッチングの確認画面です。
理論値とマッチした場所が赤で表示されます。

39

Peptide View: 理論値ベースのマッチング確認

→ 7-4 (P.66~)

Matches : 17/104 fragment ions using 24 most intense peaks (help)

Monoisotopic mass of neutral peptide Mr(calc): 1392.8068
Fixed modifications: iTRAQ4plex (K), iTRAQ4plex (N-term), Methylthio (C) (apply to specified residues or termini only)
Ions Score: 30 Expected: 4.2e-08
Matches : 17/104 fragment ions using 24 most intense peaks (help)

#	b	b ⁺⁺	b [*]	b ⁺⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	232.1414	116.5743			214.1308	107.5690	S							10
2	360.1999	180.6036	343.1734	172.0903	342.1894	171.5983	Q	1162.6800	581.8436	1145.6534	573.3303	1144.6694	572.8383	9
3	475.2269	238.1171	458.2003	229.6038	457.2163	229.1118	D	1034.6214	517.8143	1017.5948	509.3011	1016.6108	508.8090	8
4	588.3110	294.6591	571.2844	286.1458	570.3004	285.6538	L	919.5944	460.3009	902.5679	451.7876	901.5839	451.2956	7
5	701.3950	351.2011	684.3685	342.6879	683.3845	342.1959	L	806.5104	403.7588	789.4838	395.2456	788.4998	394.7535	6
6	802.4427	401.7250	785.4161	393.2117	784.4321	392.7197	S	693.4263	347.2168	676.3998	338.7035	675.4158	338.2115	5
7	889.4747	445.2410	872.4482	436.7277	871.4642	436.2357	T	592.3786	296.6930	575.3521	288.1797	574.3681	287.6877	4
8	1002.5588	501.7830	985.5322	493.2698	984.5482	492.7777	L	505.3466	253.1769	488.3201	244.6637	487.3360	244.1717	3
9	1103.6065	552.3069	1086.5799	543.7936	1085.5959	543.3016	T	392.2625	196.6349	375.2360	188.1216	374.2520	187.6296	2
10							K	291.2149	146.1111	274.1883	137.5978			1

104 : 理論値すべて
17 : マッチしたピーク
24 : 入力データのうち
24が代表ピークと
して選ばれている

赤・太字斜体 : イオンシリーズ単位で有意と判定され、スコアリングに利用
赤・太字 : イオンシリーズ単位ではぎりぎり有意と判定も、スコアリングには不使用
赤・太字でない : マッチはしたがイオンシリーズ単位ではランダムマッチの域を超えないという判断で、スコアリングには不使用
黒 : マッチしていない

40

Mascot Server

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

- 検索query数
- 同定ペプチド数
- 同定タンパク質(グループ)数
- 同定タンパク質の具体的な内容
- タンパク質にアサインされているペプチド
- スペクトルとペプチド理論値のマッチング
- 結果ファイルの出力

ペプチド同定

同定ペプチドの検証

タンパク質同定

MIS 検索のオプション

管理プログラム・カスタマイズ

41

結果ファイルの出力 ①

→ 7-5 (P.68~)

ペプチドベースのファイル出力

Re-search
☒ All
☐ Non-significant
☐ Unassigned
[\[help\]](#)
Export
As
XML

Exportの選択肢で希望するファイルフォーマットを選択しボタンを押します。

XML

XML
CSV
pepXML
mzIdentML
mzTab
DTASelect
FASTA
Mascot DAT File
MGF Peak List

Export search results

Export format: CSV
Significance threshold p<: 0.05
at: ☐ Identity ☒ Homology
Target FDR (overrides significance threshold if set): <not set>
FDR type: Distinct PSMs
Display non-significant matches: ☐
Max. number of hits: AUTO
Min. number of sig. unique sequences: 1
Protein scoring: ☐ Standard ☒ MudPIT
Include same-set protein hits (additional proteins that span the same set of peptides): ☐
Include sub-set protein hits (additional proteins that span a sub-set of peptides): 1
Group protein families: ☒
Require bold red: ☐
Show Percolator scores: ☐
Preferred Taxonomy: All entries

画面最下部にある「**Export search results**」ボタンを押す
↓
続いて現れる画面で「**Download**」ボタンを押す
↓
サーバー側でのファイル作成とそのファイルのクライアントへのダウンロード開始

結果のファイル出力 ② Report Builder

→ 7-2-11 (P.57~)

同定タンパク質ベースのファイル出力

Proteins (490)
Report Builder
Unassigned (30379)

Protein families 1~10 (out of 448)

10 per page
1 2 3 4 5 ... 45 Next
Expand all Collapse all

Report Builder タブをクリックすると、表示が下図のように切り替わります

Proteins (490)
Report Builder
Unassigned (30379)

Protein family members (480 proteins)

Columns: Standard (12 out of 16)

Filters: (none)

Export as CSV

Family	M	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	cRAP	#1::sp TRY1_BOVIN	1600	28266	47	47	7	7	2.86	sp TRY1_BOVIN
2	1	SwissProt	#2::CP2CT_MOUSE	1332	61419	76	76	13	13	2.00	Cytochrome P450 2C29 OS=
2	2	SwissProt	#2::CP254_MOUSE	550	60807	27	27	8	8	0.88	Cytochrome P450 2C54 OS=
2	3	SwissProt	#2::CY250_MOUSE	487	61128	27	27	10	10	1.20	Cytochrome P450 2C50 OS=
2	4	SwissProt	#2::CP2F2_MOUSE	470	59267	32	32	12	12	2.11	Cytochrome P450 2F2 OS=
2	5	SwissProt	#2::CP237_MOUSE	338	60590	22	22	8	8	0.89	Cytochrome P450 2C37 OS=
2	6	SwissProt	#2::CP239_MOUSE	251	60856	13	13	4	4	0.37	Cytochrome P450 2C39 OS=
2	7	SwissProt	#2::CP238_MOUSE	150	61356	9	9	4	4	0.37	Cytochrome P450 2C38 OS=

リストが作成され、表示内容をCSVファイルで出力する事ができます(「**Export as CSV**」ボタン)。

「**Columns**」 : 表示する項目を調整する

「**Filters**」 : 表示するタンパク質に対して、スコアやアサインペプチド数などで絞り込み条件を与える

などといった機能を使用する事も出来ます。

Proteins (490)
Report Builder
Unassigned (30379)

Protein family members (271 proteins)

Columns: Standard (12 out of 16)

Arrangement: <custom> Load Make default

Enabled
Family Member Database Accession Score Mass Num. of matches Num. of significant matches Num. of sequences Num. of significant sequences emPAI Description

Available
Protein family members Num. of unique sequences Num. of significant unique sequences Sequence coverage pi

Filters: "Num. of significant matches" >= 2

Num. of significant matches >= 2 Remove
AND Sequence coverage < >
Update

User :
 E-mail :
 Search title : (PRO2008 SwissProt Mouse
 MS data file : D:\PRG2008\mgf\merged.mgf
 Databases : 1: CRAP 20190304 (116 sequences; 38,459 res
 2: SwissProt 2019_10 (561,356 sequences; 20
 Taxonomy : 1: (none)
 2: Mus. (17,083 sequences)
 Timestamp : 9 Jan 2020 at 14:12:31 GMT

Re-search ☒ All ☐ Non-significant ☐ Unassigned [\[?help\]](#) Export As XML

Search parameters
 Type of search : MS/MS Ion Search
 Enzyme : Trypsin/P
 Fixed modifications : dTRAQ4plex (K), dTRAQ4plex (N-term), dMethylthio (C)
 Variable modifications : dAcetyl (Protein N-term), dGlu -> pyro-Glu (N-term Q), dOxidation (M)
 Mass values : Monoisotopic
 Protein mass : Unrestricted
 Peptide mass tolerance : ± 0.9 Da
 Fragment mass tolerance : ± 0.6 Da
 Max missed cleavages : 1
 Test method type : SGL TRAP
 Number of queries : 33,191

Score distribution
 Modification statistics for all protein families
 Legend

Protein Family Summary

Format	Significance threshold p<	0.05	Max. number of families	AUTO
Target FDR (overrides sig. threshold)	(not set)	FDR type	PSM	
Display non-sig. matches	<input checked="" type="checkbox"/>	Min. number of sig. unique sequences	1	
Show Recalculator scores	<input type="checkbox"/>	Dendrograms cut at	0	
Preferred taxonomy	<input type="checkbox"/>	All entries		

Sensitivity and FDR (reversed protein sequences)

→ 2. 同定ペプチド数 [展開]

結果画面上部

→ 7. 結果ファイルの出力 (1)

Proteins (480) Report Builder Unassigned (30379) [\[?normalink\]](#)

Protein families 1-10 (out of 448) → 3. 同定タンパク質(グループ)数

Accession contains Find Clear

1:1sp|TRY1_BOVIN| 1500 sp|TRY1_BOVIN|

1 2:CP2CT_MOUSE 1332
 2 2:CP230_MOUSE 251
 3 2:CP238_MOUSE 150
 4 2:CP254_MOUSE 550
 5 2:CP250_MOUSE 467
 6 2:CP237_MOUSE 338
 7 2:CP252_MOUSE 470

→ 4. 同定タンパク質の具体的な内容

1 2:1BP_MOUSE 1302 Endoplasmic ret
 2 2:HS17L_MOUSE 362 Heart shock 70 kD
 2:CVB5_MOUSE 1202 Cytochrome b5 c
 2:PDIA1_MOUSE 1121 Protein disulfide
 1053 Cytchrome P45
 1008 Endoplasmic N
 995 Retinol dehydrog
 497 Retinol dehydrog

→ 5. タンパク質にアサインされているペプチド [展開]

→ 6. スペクトルとペプチド理論値のマッチング [展開後さらにリンクをクリック]

10 per page 1 2 3 4 5 6 ... 45 Next Expand all Collapse all

Not what you expected? Try [dFile select summary](#).

Mascot: <http://www.matrixscience.com/>

結果画面下部

Rank展開図

Accession contains Find Clear

1:1sp|TRY1_BOVIN| 1500 sp|TRY1_BOVIN|

1 2:CP2CT_MOUSE 1332 Cytchrome P450 2C29 O8-Mus musculus OX=10090 GN=Cyp2C29
 2 2:CP230_MOUSE 251 Cytchrome P450 2C39 O8-Mus musculus OX=10090 GN=Cyp2C30
 3 2:CP238_MOUSE 150 Cytchrome P450 2C38 O8-Mus musculus OX=10090 GN=Cyp2C38
 4 2:CP254_MOUSE 550 Cytchrome P450 2C54 O8-Mus musculus OX=10090 GN=Cyp2C54
 5 2:CP250_MOUSE 467 Cytchrome P450 2C50 O8-Mus musculus OX=10090 GN=Cyp2C50
 6 2:CP237_MOUSE 338 Cytchrome P450 2C37 O8-Mus musculus OX=10090 GN=Cyp2C37
 7 2:CP252_MOUSE 470 Cytchrome P450 2F2 O8-Mus musculus OX=10090 GN=Cyp2F2

Threshold (0) Cut

	Score	Mass	Matches	Sequences	empP2
2:1-CP2CT_MOUSE	1332	61419	76 (76)	13 (13)	2.00
2:2-CP230_MOUSE	550	60837	27 (27)	8 (8)	0.88
2:3-CP238_MOUSE	487	61128	27 (27)	10 (10)	1.20
2:4-CP254_MOUSE	470	59267	32 (32)	12	Cytchrome P450 2C50 O8-Mus musculus OX=10090 GN=Cyp2C50
2:5-CP250_MOUSE	338	60590	22 (22)	5 (5)	0.37
2:6-CP237_MOUSE	251	60556	13 (13)	4 (4)	0.37
2:7-CP252_MOUSE	150	61356	9 (9)	4 (4)	0.37

Redisplay All None

134 peptide matches (43 non-duplicate, 91 duplicate)

Auto-fit to window

Query Index	Observed	He (exp)	He (calc)	Delta H	Score	Expect	Rank	U	1	2	3	4	5	6	Peptide
450	3162	1001	6178	1044	5083	0.1068	0	0.018	1	U					H. VERGIZIN
503	8846	1005	7547	1005	4093	0.1454	0	0.016	1	U					H. VERGIZIN
516	8977	1001	7808	1001	5269	0.2439	0	0.05	1	U					VERGIZIN
521	2416	1561	7059	1559	9187	0.8862	0	0.1	1	U					H. VERGIZIN

結果画面：まとめ2

Matrix Science Mascot Search Results

Protein View

Protein View

Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=1 SV=2

Database: SwissProt
 Score: 1332
 Monoisotopic mass (M₀): 61419
 Calculated pI: 8.57
 Taxonomy: [Mus musculus](#)

Sequence similarity is available as an [NCBI BLAST search of CP2CT_MOUSE against nr](#).

Search parameters

MS data file: D:\IPRG2005\mgf\merged.mgf

Enzyme: Trypsin/P: cuts C-term side of K

Fixed modifications: ITRAQ4plex (K), ITRAQ4plex (N-term), Methylation (C)

Variable modifications: Acetyl (Protein N-term), Sln->pyro-Glu (N-term Q), Oxidation (M)

Protein sequence coverage: 27%

Matched peptides shown in **bold red**.

```

1 MDGVFLVLT ILCILLISG RQSGRGKSL PPTPLIPIG NPLQIDVNI
51 KQVTFVLSK IGVPTFIST EPTFVLSTQ EKVVALIND GRPFAQNSF
101 PMAKIKLRFQ QVTVENQSN ITRAC4PMTL IQRGGRIND IEDRWQKQ
151 CVLKEIKER GSKCDPTFI SCAPCVNCS LQFRRFQVQ DREFLIMKG
201 IENEVKLSL PHLQVGFHF SLIDVCGRQ KRVFNRYFL KYSLIKELKE
251 KHESLDVNP RQFDITLYLK QKQVNRKQK EPELNLAST INDLGAGTE
301 TSTTLALVA LLLKLYQVF AKVQEKIDVR VGRNRSPCQ DRSHMPTDA
351 MIEHVQKID LPLSLERAV TQDIEKRYL IFRGTVITS LSSVLDSKE
401 FPNFEMDQ HFLNGGNFF KNDYTMQFS QKIKACRGL ARNELFLIT
451 TILQNFKLS LVRKHEIDIT FQNRGASPL PPTQLCTPL
    
```

Unformatted sequence string: [490 residues](#) (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass

Show ☒ matched peptides only ☐ predicted peptides also

Query	Start - End	Observed	Mr(expt)	Mr(calcd)	Delta	M Score	Expect	Rank	U	Peptide
#12624	49 - 59	781.0005	1559.9865	1559.8187	0.1677	0	1.14e-06	1	U	K.N.KSQSTFNFSK.A
#12625	49 - 59	781.0128	1560.0110	1559.8187	0.1922	0	0.011	1	U	K.N.KSQSTFNFSK.A
#12626	49 - 59	781.0421	1560.0696	1559.8187	0.2509	0	6.33e-07	1	U	K.N.KSQSTFNFSK.A
#12627	49 - 59	781.0544	1560.0942	1559.8187	0.2754	0	7.44e-05	1	U	K.N.KSQSTFNFSK.A
#12628	49 - 59	781.0807	1560.1569	1559.8187	0.3811	0	1.94e-06	1	U	K.N.KSQSTFNFSK.A
#12629	49 - 59	521.0639	1560.1699	1559.8187	0.3512	0	4.90e-05	1	U	K.N.KSQSTFNFSK.A
#12630	49 - 59	781.1019	1560.1892	1559.8187	0.3704	0	8.14e-08	1	U	K.N.KSQSTFNFSK.A
#12631	49 - 59	781.1422	1560.2698	1559.8187	0.4511	0	9.0e-08	1	U	K.N.KSQSTFNFSK.A
#12632	49 - 59	781.1516	1560.2887	1559.8187	0.4700	0	9.76e-08	1	U	K.N.KSQSTFNFSK.A
#12633	49 - 59	781.1604	1560.3063	1559.8187	0.4875	0	7.85e-07	1	U	K.N.KSQSTFNFSK.A
#12634	49 - 59	521.1216	1560.3491	1559.8187	0.5304	0	4.48e-05	1	U	K.N.KSQSTFNFSK.A
#14647	49 - 59	521.2416	1560.7029	1559.8187	0.8842	0	5.91e-05	1	U	K.N.KSQSTFNFSK.A
#11389	85 - 97	536.3181	1605.9324	1605.8232	0.1091	1	4.3	0.0014	1	U K.EALIDGKEEFAFG.G
#11390	85 - 97	536.3278	1605.9617	1605.8232	0.1384	1	4.6	0.0014	1	U K.EALIDGKEEFAFG.G
#11392	85 - 97	536.3356	1605.9846	1605.8232	0.1613	1	4.4	0.00077	1	U K.EALIDGKEEFAFG.G
#11393	85 - 97	536.3434	1606.0083	1605.8232	0.1850	1	4.6	0.002	1	U K.EALIDGKEEFAFG.G

ペプチド同定とタンパク質同定の流れ

→ 9-1 (P.76~)

ペプチド同定

ペプチド同定
結果の検証

タンパク質の
同定[推定]

本資料でのトピック

- ・ ペプチド配列の同定方法
- ・ スコア、同定基準値、期待値

- ・ 統計的な検証
- ・ FDR,Decoy

- ・ 同定タンパク質の基本的なルール
- ・ 類似タンパク質の取り扱い

ペプチド配列決定の方法

ペプチド配列を決定する方法

■ 参照無し

- ・ *denovo* sequencing



■ 参照あり

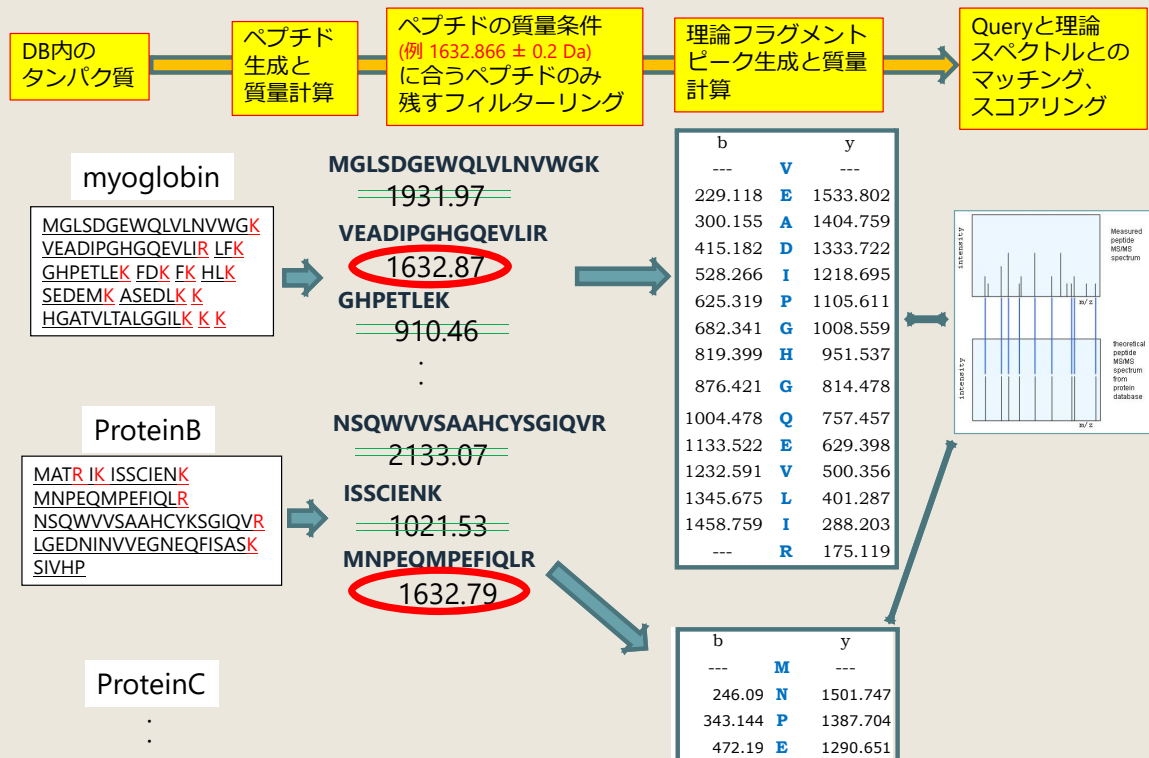
- ・ ピークリスト
- ・ 配列データから計算された理論ピーク



- ・ 主にMASCOTで利用されている方法
- ・ データベースの配列から理論フラグメントピーク（強度情報なし）のパターンを作成

ペプチド配列決定の方法：理論値計算、ペプチドのフィルターリング

→ 9-3 (P.77~)



47

スコア、同定基準値、期待値

→ 9-4 (P.78~)

Locus: 3.645.3

Score > 33 indicates identity
Score > 23 indicates homology

Score	Ident	Expect	Protein
0.4100	63	6.2e-06	R.DFIDYYLIK.Q
-0.5813	0	7	2.5 2
0.3187	1	6	3.4 3
-0.7060	1	5	3.5 4

DFPETNNILK
TPPIIHRDLK
KETMALILK

Score(S) マッチング内容を元に評価。
-10logP (P:ランダムマッチだった場合の確率)

同定基準値(St) 検索毎に計算される基準値で、
検索対象となったペプチド数を元に算出

期待値(Expect) データベース中にランダムで同じヒットが見つかる「期待値」。
(MASCOT デフォルト設定の場合) 0.05以下で同定。

$$\text{Expect} = 0.05 \cdot 10^{-(S-St)/10}$$

Mascot Server

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
 - ペプチド同定とタンパク質同定の流れ
 - ペプチド配列決定の方法
 - スコア、同定基準値、期待値
- 同定ペプチドの検証
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

48

ペプチド同定結果の検証

→ 11 (P.101~)

- 検索エンジンが提示した同定ペプチド/タンパク質が、
 - 本当に含まれているの？
 - どれくらい正しいか別の観点からの評価はないの？
 という疑問に答えるため実施

- 検索結果を統計的に検証する事が求められる

- FDR, Percolatorなどのq-value : 1%
[MASCOT基準は5%に相当]

ペプチド同定のValidation:FDR

→ 11-2 (P.101~)

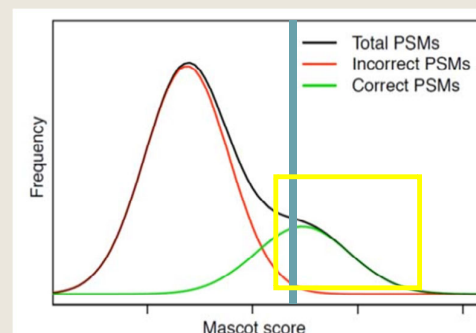
- Correctのスコア分布の山と Incorrect (またはランダム)のスコア分布の山を想定する

- 検索結果はCorrectとIncorrectを足し合わせた「Total」として現れる

- あるスコア(右図青線)で区切った際、FDRは以下のように表現できる

$$\text{FDR} = \frac{\text{Incorrect PSMs}}{\text{Total PSMs}}$$

- Incorrect PSMsのスコア分布を、Decoyデータベース(次スライド)に対する検索で得る



ペプチド同定のValidation:Decoyデータベース

■ Normalの各エントリーの配列を random / reverse に変更

■ データベースの性質 がNormalと同じ

- エントリー数
- 残基長分布
- 出現アミノ酸数

■ 検索におけるquery、パラメーターもNormalと同じ条件

■ 検索の条件をできるだけそろえつつ、参照先のDecoyには正解が含まれない

→ 結果をIncorrectのスコア分布として利用可能

別の観点から見たFDR

→ 11-2 (P.101~)

		正解配列と予測配列の一致	
		+	-
検索エンジンの判定	+	True Positive A	False Positive B
	-	False Negative C	True Negative D

$$\text{FDR} = \frac{B}{A+B}$$

- ・ A+B → 同定基準を超えたペプチド全て
- ・ **B**は不明。そこで...

Bを、「Decoyデータベースで検索した時同定基準を上回る結果数」で見積る

$$\text{FDR} = \frac{\text{Decoyデータベースへの検索で同定基準を超えたデータ数}}{\text{通常データベースへの検索で同定基準を超えたデータ数}}$$

ペプチドのスコア、同定基準、FDR

Normal,Decoy 検索結果(ペプチドと期待値)を、結果の良い順に混ぜて並べる

Sequence	E-value	N or D	N累計	D累計	FDR[D/N] (%)
NAGVEGSLIVEK	0.000034	N	1	0	0
VEEVIVTK	0.000038	N	2	0	0
:	:	:	:	:	:
TLNDELEIIEGMK	0.00062	N	118	0	0
MATRIK	0.00071	D	118	1	0.847
ISSIQSIVPALEIANAHR	0.00075	N	119	1	0.840
:	:	:	:	:	:
VGLQVVAVK	0.030	N	512	5	0.977
TAKAESK	0.035	D	512	6	1.171
LSDGVAVLK	0.037	N	513	6	1.170
SCAFFLDK	0.039	D	513	7	1.365

FDRが1%を超える直前を同定基準値とする

MASCOTでFDRを算出させる方法 → 11-3 (P.102~)

Mascot Server

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- ペプチド同定
- 同定ペプチドの検証
 - FDR
 - Decoyデータベース
 - FDR計算方法例
 - MASCOTでFDRを算出させる方法
 - FDRにまつわるトピックス
- タンパク質同定
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

Peptide tol. ± 1.2 Da # 13C 0 MS/MS tol. ± 0.6 Da

Peptide charge 2+ Monoisotopic Average

Data file ファイルを選択 選択されていません

Data format Mascot generic

Instrument Default

Decoy ☒

Target PSM FDR 1%

Start Search ... Reset Form

検索時：パラメーター「Decoy」にチェックを入れるだけ

Protein Family Summary

Format

Significance threshold p < 0.05 Max. number of families AUTO

Target FDR (overrides sig. threshold) 1% FDR type PSM

Display non-sig. matches Min. number of sig. unique sequences 1

Show Percolator scores Dendrograms cut at 0

Preferred taxonomy All entries

▼Sensitivity and FDR (reversed protein sequences)

	Target	Decoy	FDR
Protein family members	302	11	3.64%
PSMs above homology	1821	18	0.99%

Decoy results are available

検索後：

- ・「**Sensitivity and FDR**」でFDRに関わるまとめを確認可能
- ・表の上の行が「**protein FDR**」、下の行が「**PSM/peptide FDR**」に相当
- ・「**Format**」でFDRの閾値並びにカウント対象(PSM,peptide)を変更可能

タンパク質同定:同定タンパク質の基準

→ 9-5-1 (P.79~)

■ MASCOTの同定基準：

ユニークな同定ペプチドが1つ以上アサイン

■ Protein Score は同定基準に使っていない

■ Protein FDR 1% が条件に加わる事もある

■ Protein inference：「類似タンパク質」の取り扱い

タンパク質同定:同定タンパク質のグループ化

→ 9-6 (P.80~)

例：7タンパク質のペプチドシェア状況

ペプチド		タンパク質						
	ペプチド配列	A	B	C	D	E	F	G
p5	LVQDVANNTNEEAGDGTTTATVLAR	■	■	■		■		
p4	ALMLQGVDLLADAVAVTMGPK	■	■	■	■			■
p3	ISSIQSIVPALEIANahr	■	■		■		■	
p2	VGGTSDVEVNEK							■
p1	NAGVEGSLIVEK						■	

補足説明：7タンパク質のペプチドシェア状況と結果画面 1 → 9-6 (P.81～)

*p5, p4, p3, p2, p1 はマッチしたペプチド

proteinA : p5 p4 p3

proteinB : p5 p4 p3

proteinAの **same-set**

proteinC : p5 p4

proteinD : p4 p3

proteinE : p5

proteinAの **sub-set**

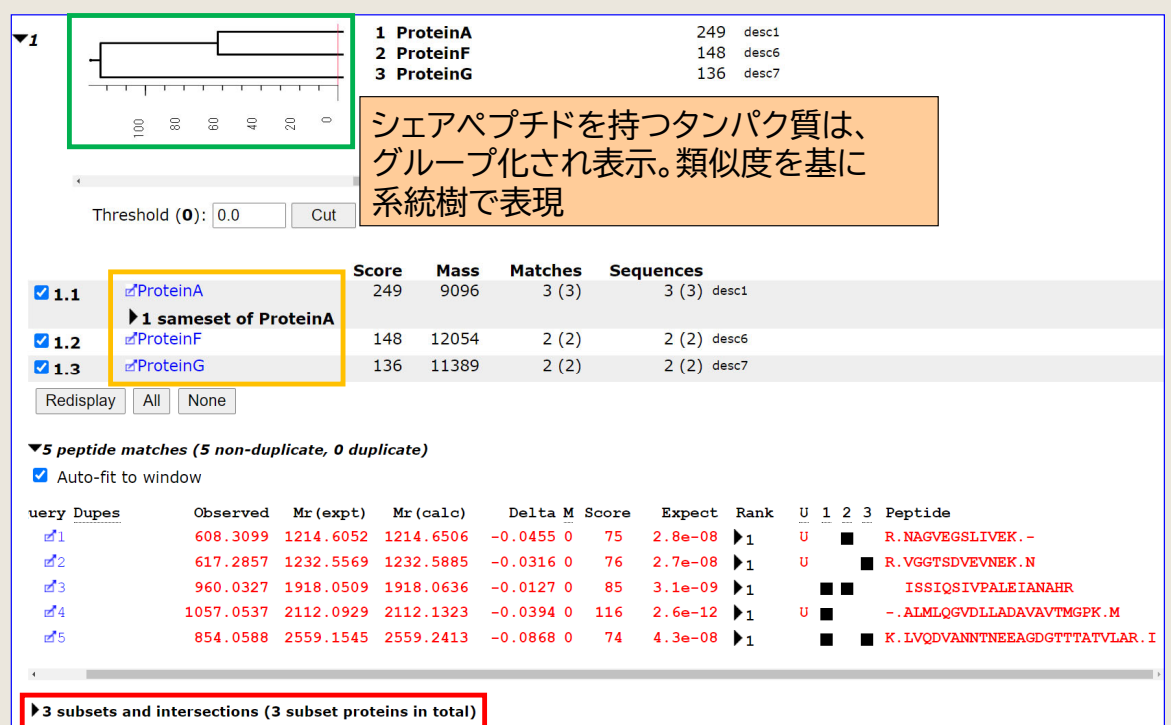
sub-set, same-set のタンパク質は
下部グループと認識され一覧に
表示されない

proteinF : p3 p1

proteinG : p4 p2

A,F,G のタンパク質は「**Family Protein**」としてまとめられる

補足説明：7タンパク質のペプチドシェア状況と結果画面 2 → 9-6 (P.81～)



		Score	Mass	Matches	Sequences
<input checked="" type="checkbox"/> 1.1	ProteinA	249	9096		
	▼ 1 same set of ProteinA				
	ProteinB	249	9855		
<input checked="" type="checkbox"/> 1.2	ProteinF	148	12054	2 (2)	2 (2) desc6
<input checked="" type="checkbox"/> 1.3	ProteinG	136	11389	2 (2)	2 (2) desc7

Redisplay All None

ユニークペプチドを持つ Protein A,F,G が表示

Protein BはProtein A の「same set」として初期段階で表示されないが、クリックし展開すると表示

▼ 3 subsets and intersections (3 subset proteins in total)					
		Score	Mass	Subset of	
	ProteinC	177	9288	1.1	desc3
	ProteinD	146	9947	1.1	desc4
	ProteinE	116	11608	1.1	desc5

Protein C,D,E はProtein A(上記図では1.1と表現) の「subset」として初期段階で表示されないが、クリックし展開すると表示

ペプチド同定とタンパク質同定：まとめ

ペプチド同定

ペプチド同定
結果の検証

タンパク質の
同定[推定]

- 3種類の中で最も主流な方法：
 - 参照有
 - 配列から作成した理論ピーク
- スコアが同定基準値を上回る時、同定 (デフォルトは95%信頼度)

- 検索エンジンが提示する基準値を別角度から評価
- Decoyデータベースを使用し、False Positive マッチ数を推定
- FDR 1%となる同定基準値を採用

- 「ユニークなペプチド1つ」で同定
- 類似タンパク質の取り扱いに注意が必要

MIS検索のオプション

→ 10 (P.83~)

- MASCOT Serverとは
- 検索方法
- 対応フォーマット・データ変換
- 検索パラメーター
- 結果画面
- 同定タンパク質・同定ペプチド
- MIS 検索のオプション
- 管理プログラム・カスタマイズ

これまでご紹介した汎用的な検索とは異なるいくつかの検索オプションについてご紹介

Spectral Library [ピークストライブラリ検索]

Quantitation [定量解析]

Crosslink [クロスリンクペプチド検索]

Error Tolerant Search [拡張2段階検索]

補助検索:スペクトルライブラリ検索

→ 10-1 (P.83~)

■ スペクトルライブラリ検索の検索対象:

- × 配列から計算された理論ピーク
- 実測スペクトル (ピークリスト)

■ スペクトルライブラリは以下2種類が可能

- ・ 自身のMASCOT Serverで行った検索結果のうち、フィルターリング条件に適合したqueryのピークリスト
- ・ 各種サイトで公開されているピークリストデータベース (PRIDE, NIST)

■ 配列データベース検索と組み合わせて利用する事が可能

■ 検索エンジン : MS PepSearch [NIST]

```
Name: SIPAYLAETLYYAMK/3
MW: 2021.0820791015626
Comment: Spec=Consensus Mods=2/0,^,iTRAQ4plex
Parent=674.702 Nreps=3 Naa=15 MaxRatio=1.000
DeltaMass=0.00 ClusterId=5258eb56-ee94-44f7-8
Protein=sp|ANXA5_HUMAN|
Num peaks: 29
114.111 757.8
115.108 810.71
116.111 850.41
117.115 818.65
136.076 1030.9
204.146 73.61
213.088 32.62
232.142 282.79
291.216 257.88
299.142 91.74
317.232 79.55
327.174 28.27
332.161 61.64
345.226 312.34
346.22 72.4
422.257 104.24
429.089 26.44
444.417 1426.25
485.322 22.68
493.293 53.71
512.318 6.05
513.317 320.91
514.32 83.07
648.386 66.0
649.386 24.8
656.352 1.97
676.381 122.41
677.383 36.02
```


MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

同定タンパク質・同定ペプチド

MIS 検索のオプション

- ピークライブラリ
- 定量解析
- クロスリンクペプチド検索
- Error Tolerant Search (拡張2段階検索)

管理プログラム・カスタマイズ

補助検索:Error Tolerant 検索 [拡張2段階検索]

→ 10-4 (P.97~)

	Score	Mass	Matches	Sequences	
2.1	499	58259	31 (31)	16 (16)	Alkaline phosphatase, placental type O
2.2	352	57626	25 (25)	12 (12)	Alkaline phosphatase, germ cell type OS=Homo sapiens OX=9606 GN=ALPG PE=2 SV=
2.3	70	57119	8 (8)	7 (7)	Intestinal-type alkaline phosphatase OS=Homo sapiens OX=9606 GN=ALPI PE=1 SV=2

Redisplay All None

▼ 34 peptide matches (34 non-duplicate, 0 duplicate)

☒ Auto-fit to window

	Expect	Rank	U	1	2	3	Peptide
	1.1e-05	1					R.GSSIFGLAPGK.A
	0.00016	1	U				R.GSSIFGLAPSK.A
53	0.014	1					R.GSSIFGLAPGK.A + [+57.0215 at S2]
65	0.007	1					R.GNEVISVMNR.A + Oxidation (M)
86	0.039	1	U				K.LGFEIPLAMDR.F + Oxidation (M)
100	3.6e-07	1					K.GNFQITGLSAAAR.F
	3.5e-06	1	U				K.ANFQITGLSAAAR.F + [+100.0160 at N-term]
	7e-06	1					R.NWYSADVFPASAR.Q
133	8.3e-06	1					L.DPSLMEMTEAALR.L + 2 Oxidation (M)
136	0.36	1					R.NWYSADVFPASAR.Q + [+57.0215 at N-term]
145	5e-05	1					R.ALTETIMFDDAIAER.A + [-48.0034 at M7]
156	6.3e-08	1					R.ALTETIMFDDAIAER.A + Oxidation (M)

質量変化(修飾・アミノ酸置換)

非特異的切断

[検索対象タンパク質]

1度目の検索で同定基準を超えたタンパク質のみ

[再検索対象query]

1度目の検索で同定基準に満たなかったqueryすべて

以下の3点について、最初の検索より対象を広げて自動的に2回目の検索を行う

- 任意の箇所で切断されるペプチド断片
- 想定外の修飾 (追加で1種まで)
- アミノ酸の置換 (塩基配列の場合、挿入・欠失にも対応)

MASCOT Serverとは

検索方法

対応フォーマット・データ変換

検索パラメーター

結果画面

同定タンパク質・同定ペプチド

MIS 検索のオプション

管理プログラム・カスタマイズ

管理プログラム・カスタマイズ

→ 12-1 (P.106~)

- ・MASCOTで使用しているデータベース、検索結果などを管理する事ができる画面があります
- ・MASCOTで使用している各種項目をカスタマイズする事ができます

Database Status

Search log

Configuration Editor

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ 同定タンパク質・同定ペプチド

■ MIS 検索のオプション

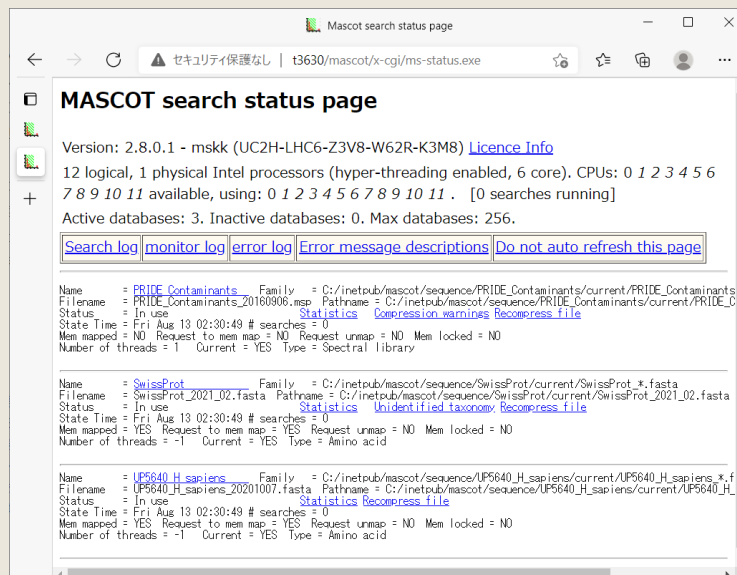
■ 管理プログラム・カスタマイズ

- Database Status
- Search log
- Configuration Editor

69

・Home → Database Status で開く画面

・各種動作のログに関する情報へのリンクや、MASCOT Server上で現在使用可能なデータベースに関する情報について表示されます



補足説明：各データベースの表示内容

```
Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
Filename = SwissProt_2021_02.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2021_02.fasta
Status = In use Statistics Unidentified taxonomy Recompress file
State Time = Fri Aug 13 02:30:49 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES Type = Amino acid
```

項目名	説明
Name	データベースの名称
Family	MASCOT Server側で管理している、ファイルのパス並びにファイルの名称ルールの情報。
Filename	現在認識されているfastaファイルの名称。 データベースのバージョンもわかります。
Pathname	現在認識されているfastaとファイルが置かれているパス。
Status	データベースの現在の状況。
State Time	現在示されている Status を認識した日時。
Mem mapped	データベースがメモリ上にマップされた状態であるか。

項目名	説明
Request to mem map	MASCOT Serverの設定でデータベースをメモリにマッピングする事を試みる設定であるか。
Request unmap	MASCOTのプログラムがメモリ上へのマッピングを解除する命令を下した状態であるか。
Mem locked	データベースのメモリ上へのマッピングを固定(lock)する状態になっているかどうか。
Number of thresholds	検索に使用可能なコア数の設定。通常は、最適設定を自動適用する設定である事を表す「-1」と表示。
Current	現在データベース関連のファイルが正しく認識されているか。
Type	配列の種類が Amino AcidかNucleic Acidか。

70

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ 同定タンパク質・同定ペプチド

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

- Database Status
- Search log
- Configuration Editor

Search log

→ 12-2 (P.111~)

・Home → Search log で開く画面

・MASCOT Serverで行ったすべての検索のログを開き、結果画面を開く事ができます

・項目別に並び替えやフィルターリング表示させることも可能

MASCOT search log

Version: 2.8.0 - mskk (YRNB-5VZ8-GFBC-T9W9-CYNQ)

Job#	PID	dbase	User Name	Email	Ti	In	start time	Durati	Status	Prio	Type	Enzyme	IP	User ID	Peak list data file
1511	5436	UP2195_D	Monitor Test DB 0		MS		Fri Aug 6 11:57:49 2021	1	No email setu	0	MIS	Yes		0	test_search.mgf
1510	12840	UP5640_H			be		Wed Aug 4 00:00:25 2021	410	User read res	0	MIS	Yes		0	C:\temp\mascotsearchtest202
1509	14116	UP5640_H			be		Tue Aug 3 23:49:12 2021	415	User read res	0	MIS	Yes		0	C:\temp\mascotsearchtest202
1508	2452	UP5640_H			be		Tue Aug 3 23:38:02 2021	414	User read res	0	MIS	Yes		0	C:\temp\mascotsearchtest202
1507	6932	SwissPro	Monitor Test DB 0		MS		Tue Aug 3 17:05:40 2021	3	No email setu	0	MIS	Yes		0	test_search.mgf
1506	9944	SwissPro	takaesu		Ly		Wed Jul 14 10:35:59 2021	23	User read res	0	MIS	Yes	19	0	Lysozyme_2p2m_1minute.ter
1505	716	SwissPro			IT		Thu Jul 8 02:01:25 2021	17	User read res	0	MIS	Yes		0	C:\temp\ITRAQ8plex6data\fro
1504	18348	SwissPro			IT		Thu Jul 8 01:57:29 2021	17	User read res	0	MIS	Yes		0	C:\temp\ITRAQ8plex6data\fro

■ MASCOT Serverとは

■ 検索方法

■ 対応フォーマット・データ変換

■ 検索パラメーター

■ 結果画面

■ 同定タンパク質・同定ペプチド

■ MIS 検索のオプション

■ 管理プログラム・カスタマイズ

- Database Status
- Search log
- Configuration Editor

Configuration Editor

→ 13 (P.113~)

・Home → Configuration Editor で開く画面

・MASCOT Serverで使用する各種設定変更・カスタマイズ

Mascot Configuration

Amino Acids	Amino Acid Data
Modifications	Modification definitions
Symbols	Symbols used in chemical formulae
Linkers	Linker definitions
Enzymes	Enzyme definitions
Instruments	Fragmentation Rules
Quantitation	Quantitation Methods
Crosslinking	Crosslinking Methods
Configuration Options	Global Options in mascot.dat
Database Manager	Sequence databases, Parse Rules and automated downloads