

**Mascot Server**

日本語マニュアル



MASCOT 対応 Version : 2.8

最終更新日 : 2023/01/13





# 目次

<b>1.はじめに: MASCOT Server でできる事</b> .....	<b>1</b>
1-1. MASCOT Server とは .....	1
1-2. MASCOT Server でできる事・まとめ .....	1
1-3. 検索手法とマニュアルで対応しているページ.....	3
<b>2. Mascot Server のシステム</b> .....	<b>4</b>
2-1. MASCOT Server は WEB アプリケーションである .....	4
2-2. MASCOT Server のインストール先と各フォルダ .....	5
2-3. 3 種類の MASCOT 検索 .....	6
<b>3. MASCOT Server の入力データ</b> .....	<b>7</b>
3-1. MASCOT Server が読み込み可能なファイルフォーマット .....	7
3-1-1. PMF で対応するファイルフォーマット .....	7
3-1-2. MIS で対応するファイルフォーマット:どのような情報が抜き出されているか .....	9
3-1-3. MIS で対応するファイルフォーマット:mgf.....	9
3-1-4. MIS で対応するファイルフォーマット:mgf 以外のフォーマット.....	10
3-2 . raw データ変換プログラム.....	11
<b>4. MASCOT Server 検索方法</b> .....	<b>12</b>
4-1. 「URL」と MASCOT Server [ネットワーク上で MASCOT を指定する方法] .....	12
4-2. 変換済みファイルを WEB ブラウザで検索 .....	13
4-2-1. PMF : テキストデータを WEB ブラウザで検索.....	13
4-2-2. MIS : mgf ファイルを WEB ブラウザで検索 .....	14
4-3. Daemon を使って raw データを直接検索 .....	14
4-3-1. Daemon + ProteoWizard .....	15
4-3-2. Daemon + Distiller .....	15
4-4. 質量分析装置メーカーのソフトウェアから検索 .....	17
<b>5. 検索パラメーターとデータベース</b> .....	<b>18</b>
5-1. PMF : 検索パラメーター 一覧.....	19
5-2. Sequence Query : 検索パラメーター 一覧.....	22
5-3. MIS : 検索パラメーター 一覧.....	27
5-4. パラメーターの中でカスタマイズ可能な項目 .....	33
5-5. Database.....	34
5-6. Search form defaults .....	36
<b>6. 検索結果画面:PMF</b> .....	<b>37</b>
6-1. 表示例で使用している検索について .....	37
6-2. 表示内容の詳細 : summary 画面.....	37

6-2-1. 展開しない状態での画面概要.....	37
6-2-2. ヘッダー部分.....	39
6-2-3. Score Histogram .....	39
6-2-4. 表示内容の変更 [Format As] .....	40
6-2-5. 再検索 : Research all, search unmatched.....	41
6-2-6. 同定タンパク質の情報.....	41
6-2-7. Search parameters .....	42
<b>6-3. Protein View .....</b>	<b>42</b>
<b>6-4. 結果の Export .....</b>	<b>47</b>
<b>7. 検索結果画面:MIS.....</b>	<b>48</b>
7-1. 表示例で使用している検索について .....	48
7-2. 表示内容の詳細 : summary 画面.....	48
7-2-1. 展開しない状態での画面概要.....	48
7-2-2. ヘッダー部分.....	50
7-2-3. 再検索と結果ファイル出力: Re-search ボタン、Export ボタン.....	50
7-2-4. Search parameters .....	50
7-2-5. Score Distribution .....	51
7-2-6. Modification statistics for all protein families.....	52
7-2-7. Legend (凡例).....	52
7-2-8. 表示内容の切り替え [スコア足切りなど] .....	52
7-2-9. Sensitivity and FDR.....	53
7-2-10. 同定タンパク質とアサインペプチド.....	54
7-2-11. Report Builder タブ (タンパク質ベースの検索結果ファイル出力) .....	57
7-2-12. Unassigned タブ.....	59
<b>7-3. Protein View .....</b>	<b>59</b>
<b>7-4. Peptide View .....</b>	<b>63</b>
<b>7-5. Export 機能によるペプチドベースの検索結果ファイル出力 .....</b>	<b>68</b>
<b>8. PMF : タンパク質同定.....</b>	<b>71</b>
8-1. PMF タンパク質同定のまとめ .....	71
8-2. 入力データの調整.....	71
8-3. 配列から計算される理論ピーク .....	72
8-4. 同定タンパク質 : マッチングとスコア、同定基準値、期待値.....	73
8-5. protein inference : ユニーク/シェア ペプチド、タンパク質のグループ化 .....	74
<b>9. MIS : ペプチド同定とタンパク質同定.....</b>	<b>76</b>
9-1. MIS ペプチド同定とタンパク質同定のまとめ .....	76
9-2. 入力データの調整.....	76
9-2-1. MS2 スペクトル、ピークリストの再調整.....	76
9-2-2. ファイル内のクエリー数.....	77

9-3. ペプチド配列から行う理論値計算、ペプチドのフィルターリング .....	77
9-4. 同定ペプチド : マッチングとスコア、同定基準値、期待値、外挿的な評価 .....	78
9-5. 同定ペプチドから導き出される同定タンパク質 .....	79
9-5-1. 同定タンパク質=ユニークな同定ペプチドが1つ以上アサインされている .....	79
9-5-2. 1 Hit wonders : 同定タンパク質の Sensitivity と Specificity .....	80
9-6. protein inference : ユニーク/シェア ペプチド、タンパク質のグループ化 .....	80
<b>10. MASCOT 検索のオプション [MIS] .....</b>	<b>83</b>
<b>10-1. Spectral Library .....</b>	<b>83</b>
10-1-1. Spectral Library 概要 .....	83
10-1-2. Spectral Library 検索方法 .....	84
10-1-3. Spectral Library の検索結果 .....	85
10-1-4. Spectral Library をローカルの MASCOT Server にセットする方法 .....	86
10-1-5. Spectral Library 補足説明資料へのリンク .....	86
<b>10-2. Quantitation .....</b>	<b>87</b>
10-2-1. Quantitation 概要 .....	87
10-2-2. Quantitation 検索方法 .....	87
10-2-3. Quantitation 検索結果 .....	88
10-2-4. Quantitation 設定の作成 .....	90
10-2-5. Quantitation 補足説明へのリンク .....	90
<b>10-3. Crosslink .....</b>	<b>91</b>
10-3-1. Crosslink 検索 概要 .....	91
10-3-2. Crosslink 検索を行う方法 .....	92
10-3-3. Crosslink 検索に際し注意すべき MASCOT 設定値 .....	92
10-3-4. Crosslink 検索結果 .....	93
10-3-5. Crosslinking 設定の作成 .....	96
10-3-6. Crosslink 補足説明へのリンク .....	97
<b>10-4. Error Tolerant Search .....</b>	<b>97</b>
10-4-1. Error Tolerant Search 概要 .....	97
10-4-2. Error Tolerant Search を実行する方法 .....	98
10-4-3. Error Tolerant Search 検索結果 .....	98
<b>11. FDR と Decoy データベース .....</b>	<b>101</b>
<b>11-1. FDR はなぜ必要か? .....</b>	<b>101</b>
<b>11-2. FDR とは、Decoy データベースとは .....</b>	<b>101</b>
<b>11-3. MASCOT で FDR を算出させる方法 .....</b>	<b>102</b>
<b>11-4. FDR にまつわるトピックス .....</b>	<b>103</b>
11-4-1. Decoy データベースの利用方法、FDR 計算方法には明確な定義がありません .....	103
11-4-2. FDR 計算のカウント対象 .....	104
11-4-3. Decoy データベースでの検索結果と “1 Hit Wonders” .....	104
11-4-4. 「peptide FDR 1%」と同定タンパク質の正確さ .....	105

11-4-5. Protein FDR .....	105
11-4-6. 同定ペプチド数が少ないケースにおける FDR 調整の限界 .....	105
<b>12. MASCOT Server 管理 – データベースと検索ログ .....</b>	<b>106</b>
12-1. 現在利用可能なデータベースに関する情報 : Database Status .....	106
12-2. 検索ログ : Search log .....	111
<b>13. MASCOT Server のカスタマイズ .....</b>	<b>113</b>
13-1. カスタマイズは Configuration Editor で .....	113
13-2. Amino Acids .....	114
13-3. Modifications .....	115
13-4. Symbols .....	121
13-5. Linkers .....	122
13-6. Enzymes .....	126
13-7. Instruments .....	128
13-8. Quantitation .....	130
13-9. Crosslinking .....	138
13-10. Configuration Options .....	140
13-11. Database manager .....	142
13-12. Security .....	142

# 1.はじめに: MASCOT Server でできる事

## 1-1. MASCOT Server とは

MASCOT Server は、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**です。

入力データ側は質量分析装置から得られたスペクトル情報ですが、すべてを利用するのではなく抜粋された情報(ピークリスト)を利用しています。一方参照するデータベースについては、配列データベース(アミノ酸配列・塩基配列)から計算された理論スペクトル、あるいは配列情報と結びついている過去の実測スペクトルどちらに対しても検索を行うことができます。

ピークリストと理論値または実測スペクトルとの照合に対してマッチング度合いを示すスコアが提示され、それが別途計算された同定基準スコアを超える時、当該ペプチド/タンパク質を「同定」とみなします。すなわち、入力データであるスペクトルデータが、どのようなペプチド/タンパク質由来のスペクトルであるかを同定します。MASCOT の同定基準スコアはデフォルトでは 95%信頼度を元に算出されていますが、現在論文などでは 1% FDR(詳細は 11 章をご参照ください)を満たすよう求められていて、FDR の基準を満たすように同定基準を調整する使われ方をすることが多くなっています。

## 1-2. MASCOT Server でできる事・まとめ

### [定性]

- ・ ペプチド配列の同定
- ・ タンパク質の同定
- ・ 翻訳後修飾の同定と、候補残基が複数あるケースにおける、部位別の同定確率の提示
- ・ 設定パラメーター値以外の切断パターン、修飾、アミノ酸置換の提示
- ・ クロスリンクペプチドの同定(ペプチドの組み合わせとリンカーが付いている残基の位置)

### [定量]

- ・ タンパク質の相対量(比)または定量指標の提示
    - フラグメントピークのピーク強度を使った定量。[ MASCOT Server のみで実施可能 ]
    - Spectral Counting(emPAI) 定量指標。[ MASCOT Server のみで実施可能 ]
- \* Extracted Ion Chromatograms (XICs)のピーク強度を使った定量については、MASCOT Distiller 定量モジュールと一緒に計算を行う必要があります。

Family	M	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	CRAP	sp TRY1_BOVIN	1.00	28266	47	47	7	7	2.86	sp TRY1_BOVIN
1	1	SwissProt	#2::CP2CT_MOUSE	1.32	61419	76	76	13	13	2.00	Cytochrome P450 2C29 OS=Mus musculus OX=10090
2	2	SwissProt	#2::CP234_MOUSE	0.50	60887	27	27	8	8	0.88	Cytochrome P450 2C54 OS=Mus musculus OX=10090
2	3	SwissProt	#2::CY250_MOUSE	4.87	61128	27	27	10	10	1.20	Cytochrome P450 2C50 OS=Mus musculus OX=10090
2	4	SwissProt	#2::CP2F2_MOUSE	4.70	59267	32	32	12	12	2.11	Cytochrome P450 2F2 OS=Mus musculus OX=10090 G
2	5	SwissProt	#2::CP237_MOUSE	3.38	60590	22	22	8	8	0.89	Cytochrome P450 2C37 OS=Mus musculus OX=10090
2	6	SwissProt	#2::CP239_MOUSE	2.51	60856	13	13	4	4	0.37	Cytochrome P450 2C39 OS=Mus musculus OX=10090
2	2	SwissProt	#2::H571L_MOUSE	1.88	78552	13	13	4	4	0.28	Heat shock 70 kDa protein 1-like OS=Mus musculus OX=10090
2	1	SwissProt	#2::CYB5_MOUSE	1.20	16817	42	42	5	5	3.08	Cytochrome b5 OS=Mus musculus OX=10090 GN=Cyt
4	1	SwissProt	#2::PDIA1_MOUSE	1.121	16494	53	53	16	16	2.54	Protein disulfide-isomerase OS=Mus musculus OX=10090
6	1	SwissProt	#2::CP1A2_MOUSE	1.053	63034	38	38	10	10	1.31	Cytochrome P450 1A2 OS=Mus musculus OX=10090 C
Z	1	SwissProt	#2::ENPL_MOUSE	1.008	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 LSCILLLSLV RQSSGRGKLF PGFTLPLIIG NFLQIDVKNI
51 SQSFNFPSKA YGPFVTLVLG SKPTVILHGY EAVKEALIDR GEEFAGRGSP
101 PMAEKIKGF GVVFSNGNRW KEMRFRITMT LKVLGMSGRN IEDRVQREAR
151 CLVRELARKTK GSPCDPTFLL SCAPCNVICV IIFQNRFDYK DREFLILMDR
201 INENVRILSS FWLQVCSNFF SLIDYCPQSH HKIVNRFNYL KSYLLEKIKE
251 HRESLDVTFN RFIDYLLIK QRQVNHIEQS EFSLENLAST INDLFGAGTE
301 TTSTLRYAL LLLLKYDVT AKVQREIDRV VGRHRSFCMQ DRSDHPEYDA
351 MIHEVQRFD LLPTSLPHAV TCDIKFRKYL IPKOTTVITS LSSVLHDSKE
401 FENFEMFDPG HFLNGNGNFK KSDYFMPST GKRICAGEQL ARMEFLILIT
451 TILQNFILKS LVHFREIDIT FVMDGFASLP FFFYQQLFIPL

```

同定ペプチド

修飾

R.GSFPMAEK.I + Oxidation (M)

MASCOT Server の検索結果。同定ペプチドや同定タンパク質、修飾の情報が表示されます。

Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
#2::CO4B_HUMAN	164342	217600	3818	3818	103	103	48.75	Complement C4-B C
#2::CO4A_HUMAN	163856	217680	3814	3814	102	102	44.57	Complement C4-A C
#2::APOB_HUMAN	127385	624988	3897	3897	214	214	9.70	Apolipoprotein B-10
#2::CERU_HUMAN	59576	143199	1466	1466	50	50	14.97	Ceruloplasmin OS=H
#2::A1BG_HUMAN	58870	58330	1527	1527	19	19	11.90	Alpha-1B-glycoprote
#2::HEMO_HUMAN	44576	58934	1899	1899	30	30	144.13	Hemopexin OS=Homo
#2::CFAH_HUMAN	37520	167416	1521	1521				
#2::FHR2_HUMAN	1329	36538	82	82				Hemopexin OS=Homo sapiens OX=9606 GN=H
#2::FHR1_HUMAN	1289	43717	80	80	13	13	7.39	Complement factor H

Spectral Counting 定量指標である emPAI の表示

Proteins (545) Report Builder Unassigned (140931) 5 permalinks

Protein family members (545 proteins)

Columns (19 out of 58)

Filters: (none)

Export as CSV

Family	M	DB	Accession	Score	Mass	114/113	115/113	116/113	117/113	118/113	119/113	121/113	Matches	Match(sig)
1	1	SwissProt	#2::CO4B_HUMAN	164342	217600	1.033	1.070	1.045	1.016	1.155	1.051	1.055	3818	3818
1	2	SwissProt	#2::CO4A_HUMAN	163856	217680	1.036	1.073	1.044	1.019	1.159	1.052	1.060	3814	3814
2	1	SwissProt	#2::APOB_HUMAN	127385	624988	1.082	1.362	0.827	1.203	1.189	1.093	1.079	3897	3897
3	1	SwissProt	#2::CERU_HUMAN	59576	143199	0.884	1.080	0.711	1.047	1.283	1.027	1.027	1466	1466
4	1	SwissProt	#2::A1BG_HUMAN	58870	58330	0.949	1.201	0.994	1.124	1.181	1.041	1.085	1527	1527
5	1	SwissProt	#2::HEMO_HUMAN	44576	58934	0.970	1.161	0.940	1.086	1.334	1.053	1.103	1899	1899
6	1	SwissProt	#2::CFAH_HUMAN	37520	167416	0.962	1.130	0.872	1.103	1.306	1.096	1.132	1521	1521
6	2	SwissProt	#2::FHR2_HUMAN	1329	36538	0.858	1.123	0.687	1.132	1.406	1.065	1.071	82	82

プロダクトイオンスペクトルのピーク強度を使った定量

	Accession	Score	Mass	(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)	SD(geo)	#	(15ngR
1.1	sp P06733 ENOA_HUMA	33306	47481	1.1327	1.0902	34	
1.2	sp P13929 ENOB_HUM...	9940	47299	1.1704	1.1012	5	
2.1	sp P07900 HS90A_HU...	18698	85006	1.1109	1.1024	45	
2.2	sp P08238 HS90B_HU...	17983	83554	1.1230	1.0404	45	
2.3	sp P14625 ENPL_HUM...	7026	92696	1.1133	1.0544	29	
2.4	sp Q12931 TRAP1_HU...	2003	80345	1.1043	1.0652	9	
3.1	sp P05787 K2C8_HUM...	17641	53671	1.1180	1.1617	48	
3.2	sp P08670 VIME_HUM...	11853	53676	1.1082	1.0447	38	
3.3	sp P08729 K2C7_HUM...	8882	51411	1.1129	1.0392	31	

Distiller での定量解析結果

(Precursor XICs 定量は MASCOT Server に加え、Distiller 定量モジュールが必要です。)

### 1-3. 検索手法とマニュアルで対応しているページ

2-3 でご説明するように、MASCOT の検索は3つの方法:PMF (Peptide Mass Fingerprint),SQ (Sequence Query), MIS (Mascot Ions Search)が使用可能です。このマニュアルで各手法に関連がある章は、それぞれ以下の通りです。

**PMF** : 1,2,3,4,5,6,7,8,9,10,11,12,13

**SQ** : 1,2,3,4,5,6,7,8,9,10,11,12,13

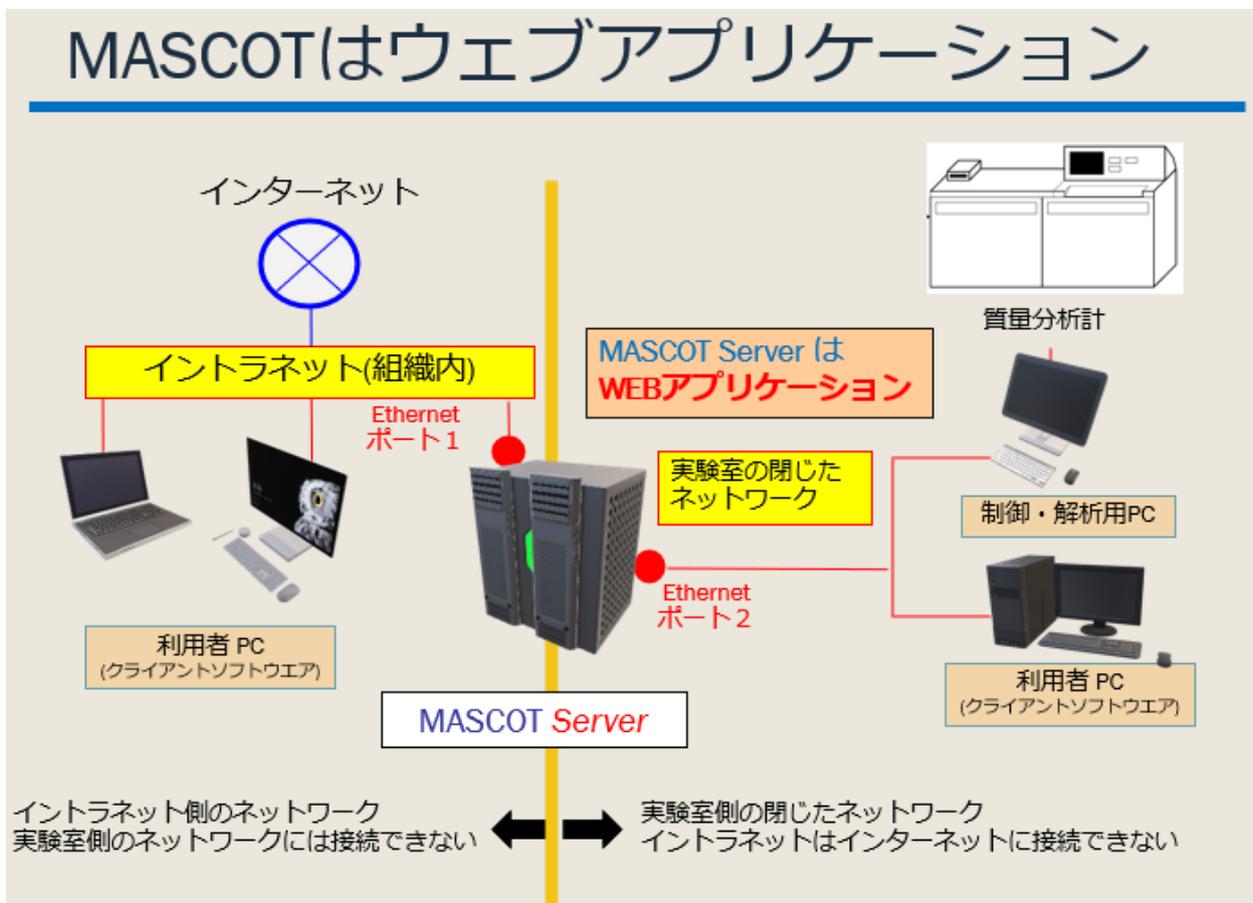
**MIS** : 1,2,3,4,5,6,7,8,9,10,11,12,13

Sequence Query については使用頻度が低い事から、本マニュアルにて Query の文法や結果の解釈について詳しく説明していません。検索時に必要な入力である Sequence Query の文法については [https://www.matrixscience.com/help/sq\\_help.html](https://www.matrixscience.com/help/sq_help.html) を、結果の解釈については MIS と概ね同じですので7章をご覧ください。

## 2. Mascot Server のシステム

### 2-1. MASCOT Server は WEB アプリケーションである

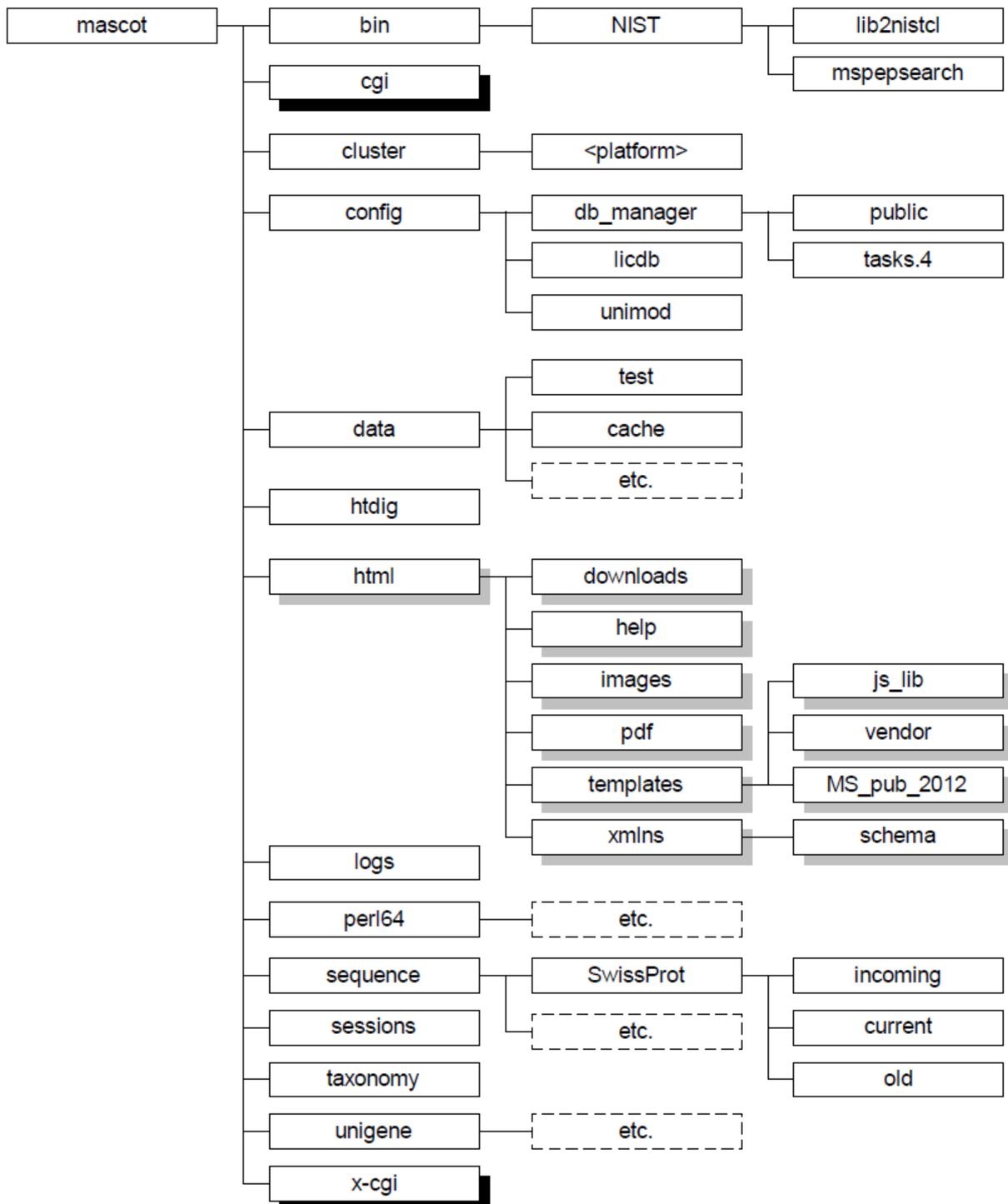
MASCOT Server は **WEB アプリケーション**で、**クライアント-サーバーシステム**です。HTTP プロトコルを介してデータのやり取りを行います。ユーザーは、MASCOT Server プログラムがインストールされているコンピューター、あるいはネットワーク上の離れたコンピューター、どちらからでも MASCOT Server を 利用する事が可能です。MASCOT Server をインストールするマシンには、WEB サーバープログラムをインストール・稼働させる必要があるほか、コンピューターが置かれているネットワーク上で HTTP プロトコルの通信が許可されている必要があります。



## 2-2. MASCOT Server のインストール先と各フォルダ

MASCOT Server に関連するファイルやプログラムはデフォルト設定で「**C:¥inetpub¥mascot**」フォルダ (Windows) あるいは **/usr/local/mascot** ディレクトリ (Linux) 以下に配置されるようインストールされます。インストール先のフォルダ/ディレクトリはインストール時に設定可能であるほか、**一部のみ全く別の場所へ配置することも可能**です。

以下、MASCOT のフォルダ/ディレクトリ の基本構成です。



各フォルダ/ディレクトリ に含まれる内容は以下の通りです(概ねフォルダ名から想定することができます)。

- bin** : CGI 以外の実行プログラム
- cgi** : CGI の実行プログラム
- cluster** : cluster システム使用時各計算ノードに配布するプログラム
- config** : 設定ファイル、ライセンスファイル
- data** : MASCOT 検索結果ファイル。結果ファイルは検索日の日付名フォルダ(yyyymmdd)の下に格納される。
- htdig** : ウェブページ検索プログラム htdig 関連のファイル
- html** : web ページドキュメントファイル
- logs** : 検索並びに各主動作のログファイル
- perl64** : perl のプライベートコピー (version 5.18)
- sequence** : 検索対象となるデータベース
  - **current** : 現在使用中のファイル
  - **incoming** : 次に使用するデータベースの一時保管場所
  - **old** : 1 つ前に使用していたデータベースの fasta ファイルを保存 (復元可能)
- sessions**: セキュリティに使用するセッション情報
- taxonomy** : 生物種情報に関するファイル
- unigene**: UniGene インデックスファイル
- x-cgi** : セキュリティ制御などに関連して使用される CGI の実行プログラム

### 2-3. 3 種類の MASCOT 検索

MASCOT Server では入力データのタイプ別に3つの検索方法を使用する事ができます。

- **PMF (Peptide Mass Fingerprint) 法** :  
MS1,ペプチドピークの組み合わせからタンパク質を同定
- **Sequence Query 法** :  
MS1 や MS2 のピーク情報に各種絞り込み条件を追加して検索  
例) 1234.2 seq(n-AC[DHK]) seq(c-HI)  
1314.7 tag(513.3,T[I]L)SP,911.5)
- **MIS (MS/MS Ions Search) 法** :  
MS1 情報と MS2 情報を組み合わせてペプチドを同定。MS1 でペプチドを絞り込み、MS2 でマッチング。同定されたペプチドをもとにタンパク質を同定(推定)

主に TOF などプレカーサーマスペクトルを測定した場合、そのデータからタンパク質を同定するためには Peptide Mass Fingerprint 法を使用します。またショットガン解析、ペプチドの MS1,MS2 情報をもとにペプチド配列やタンパク質を同定する方法としては MS/MS Ions Search 法が使用されます。

### 3. MASCOT Server の入力データ

この章では MASCOT Server で検索を行うための前処理と検索の流れを説明します。MASCOT Server で検索を行うためには、質量分析装置で測定したデータをそのまま MASCOT Server に投げるのではなく、事前にデータ処理をする必要があります。データ処理のポイントは2つあります。

- ・ノイズをカットし、ペプチド又はフラグメントの質量を反映するピークのみを集めたデータにする事 (3-1)
- ・バイナリ形式の Raw データを判読可能なテキストファイルまたは xml フォーマットに変換する事 (3-2)

すなわち **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 データを変換する事はできません。

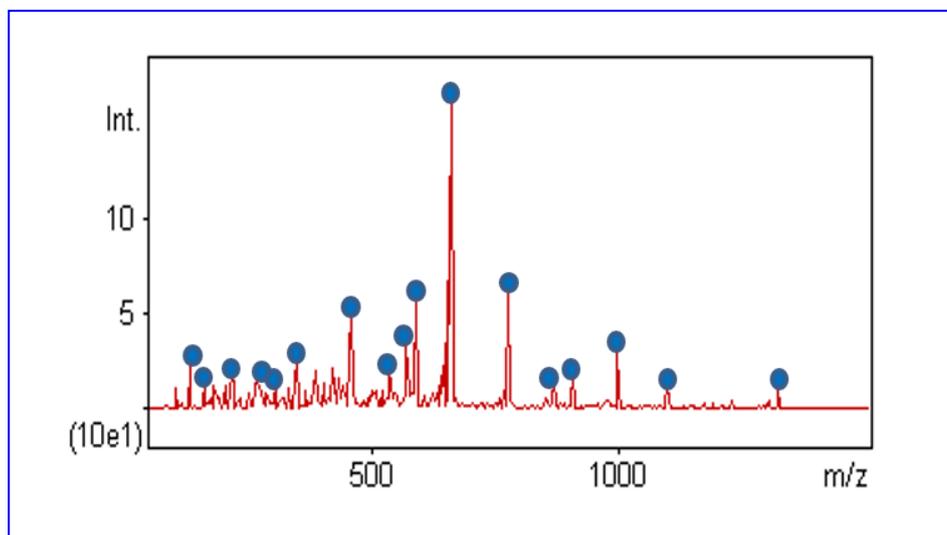
**3-1** では検索に投げる前にデータの処理が必要な理由と MASCOT Server が受け付けるファイルフォーマットについて、**3-2** では raw データから MASCOT で読み込みが可能なファイルフォーマットに変換するプログラムについてご紹介します。

#### 3-1. MASCOT Server が読み込み可能なファイルフォーマット

MASCOT Server が読み込み可能なファイルフォーマットと変換の際に行われたピーク抽出処理の内容について説明しています。

##### 3-1-1. PMF に対応するファイルフォーマット

PMF 検索の入力データはペプチドの質量が反映されたピークに関する情報です。PMF の入力データ作成の際には**ペプチド由来のピークと考えられる箇所**(下図、青色の●)を推定し、そのピークの  $m/z$  のみ、あるいは  $m/z$  と intensity 情報を利用します。ノイズに該当する個所の数値はこの段階で捨てられます。



ファイルについて、テキストファイルであれば拡張子が何であっても問題なく検索を実行できます。ファイルの中身として、下左図のように 1 行につき1つの  $m/z$  が先頭に書かれていれば検索が可能です。下右図のように後ろに **intensity** に該当する値が含まれている場合、**intensity** 情報を参考に入力データの  $m/z$  の組み合わせを再構成したいいくつかのパターンを作成し、各パターンのうちスコアが最も高いケースを真の入力データとして採用する方式となります。そのため同じ測定データでも後ろに **intensity** 情報を含まないケースと含むケースで検索結果を比べると、結果が異なるケースがあります。

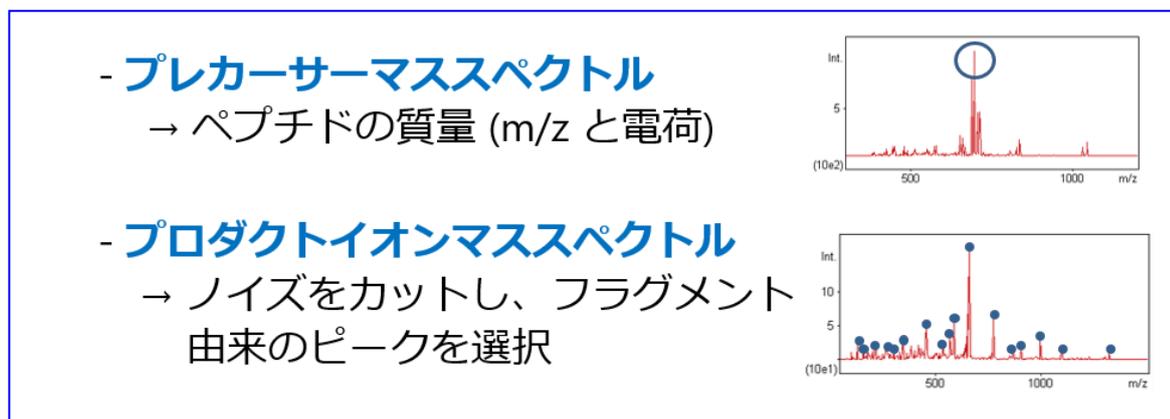
832.662	
903.342	
1186.439	
1373.681	
1403.722	
1515.444	
1727.916	
1743.951	
1759.966	
1788.721	
1804.71	
2174.812	
2190.112	
2256.871	
2273.266	
2288.489	
361.21774	4838.6552
487.26656	5281.3009
494.28934	8868.6732
505.77755	16079.047
686.36334	22677.156
723.34797	36555.65
836.78635	6731.0498
955.4842	2890.8536
1002.4743	5553.321
1020.521	2661.7174
1263.679	1759.9609
1350.7063	3770.1817
1495.685	43476.619
1533.6332	3063.8439
1675.6148	3315.1174

PMF 検索では上記のように単純なテキストファイルに加え、各装置メーカー/共通フォーマットから出力される以下のファイルの読み込みにも対応しています。

- AB SCIEX Data Explorer (.PKM)
- Bruker Analysis AutoXecute Data Report
- Bruker (.XML)
- mzData (.XML)

### 3-1-2. MIS に対応するファイルフォーマット:どのような情報が抜き出されているか

MIS 検索では入力データとしてペプチド並びにフラグメントの質量が反映されたピークの情報を集約して利用します。下図にもあるように、プレカーサーマススペクトルからはペプチドの質量を計算するために必要な  $m/z$  と電荷の情報(不明な場合は推定値)を、プロダクトイオンマススペクトルからはフラグメントを反映するピークを選び、ノイズがカットされた情報を抽出します。



### 3-1-3. MIS に対応するファイルフォーマット:mgf

**Mascot Generic Format (mgf)** は MASCOT Server の MIS 入力データとして最も使用されているフォーマットです。以下のような構成となっています。

```
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
```

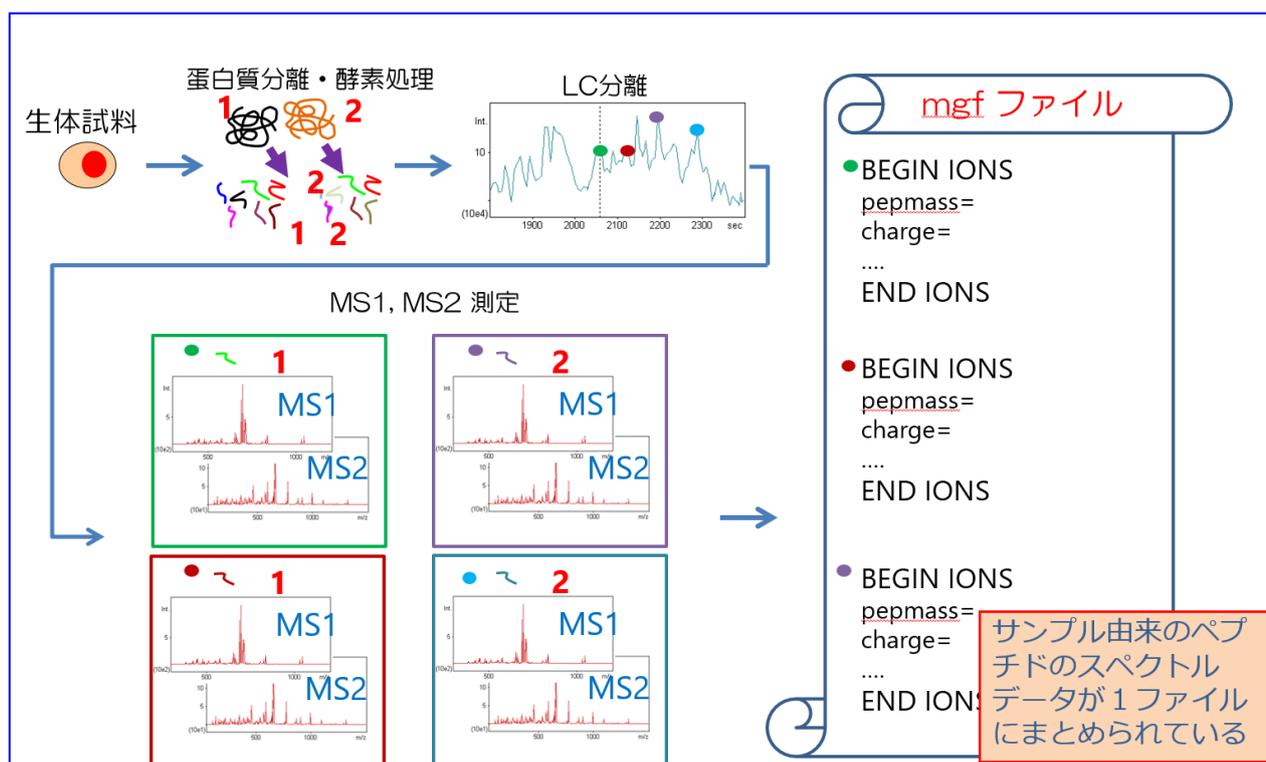
① ペプチドの質量を計算  
(817.44-1.0067)\*2 = 1632.866

② プロダクトイオンマススペクトル、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

1query 毎に **BEGIN IONS** で始まり **END IONS** で終わります。①の部分、赤い文字の行はプレカーサーマススペクトルから得られた情報に該当します。**pepmass** が  $m/z$  を、**charge** が電荷を表します。②の部分、緑の文字の行はプロダクトイオンマススペクトルから得られた情報に該当し、ノイズが

カットされフラグメントピークに該当する  $m/z$  と intensity の情報となります。なお intensity の後に電荷情報が記されている場合、MASCOT Server 側で 1 価に換算(decharge)したフラグメントピークとして扱われます。intensity の後に電荷情報が記された形でのファイル出力すべてのソフトウェアで行われるわけではなく、現在のところ MASCOT Distiller を使用する事でのみ実現可能ですのでご注意ください。

mgf ファイルには複数の query をまとめて保存する事もできます(下図)。



mgf ファイルを得る方法として、MASCOT Distiller, ProteoWizard の msconvert,そして各社装置メーカーの付属ソフトウェアから出力するオプションがあります。詳細は 3-2 で説明いたします。

### 3-1-4. MIS に対応するファイルフォーマット:mgf 以外のフォーマット

MASCOT Server の MIS 検索では mgf ファイル以外の各種フォーマット変換後ファイルの読み込みにも対応しています。対応フォーマットは以下の通りです。

- Finnigan (.ASC)
- Micromass (.PKL)
- Sequest (.DTA)
- PerSeptive (.PKS)
- Sciex API III
- Bruker (.XML)
- mzData (.XML)
- **mzML (.mzML) \* 多くの装置で出力可能なフォーマットで、mgf の次によく利用されます。**

各フォーマットの詳細は以下の WEB ページをご覧ください。

[https://www.matrixscience.com/help/data\\_file\\_help.html#FORMAT](https://www.matrixscience.com/help/data_file_help.html#FORMAT)

## 3-2 . raw データ変換プログラム

バイナリ形式の Raw データを判読可能なテキストファイルまたは xml フォーマットに変換するプログラムは、大きく分けると以下の 3 種類あります。

### ■ ProteoWizard msconvert

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です。

<https://proteowizard.sourceforge.io/>

msconvertGUI.exe というプログラムを起動してプログラム内で mgf を作成する事もできますし、MASCOT Daemon と組み合わせて利用する事もできます(4-3 で簡単な操作の紹介をしています)。

### ■ MASCOT Distiller

弊社が販売しているソフトウェアです。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です。

<https://www.matrixscience.com/distiller.html> (英語紹介ページ)

<https://www.matrixscience.co.jp/distiller.html> (日本語紹介ページ)

### ■ 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解析ソフトウェアでは 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で、それらを MASCOT 検索に利用する事ができます。mzML で使用する場合、ピークに該当するものを選び出した状態でそれを反映したファイル出力がされている事が望ましいです。

## 4. MASCOT Server 検索方法

この章では MASCOT Server で検索を行う方法についてご説明しています。

**4-1** では MASCOT Server をネットワーク上で指定する方法についてご説明します。また **4-2** では何らかの形で準備した mgf を入力データとして、WEB ブラウザを介して MASCOT 検索を実施する方法をご説明します。

**4-3** 以降では raw データから mgf への変換を意識することなく検索を実施する方法についてご紹介します。**4-3** は MASCOT Server に無料でバンドルされているソフトウェア MASCOT Daemon を使った検索方法、そして **4-4** では質量分析装置メーカーが提供しているソフトウェアを使って行う方法についてご案内いたします。

### 4-1. 「URL」と MASCOT Server [ネットワーク上で MASCOT を指定する方法]

MASCOT Server は WEB アプリケーションであり、自分自身あるいは別のコンピューターから MASCOT Server を URL で指定します。

URL の記入例としては以下のような記述となります。

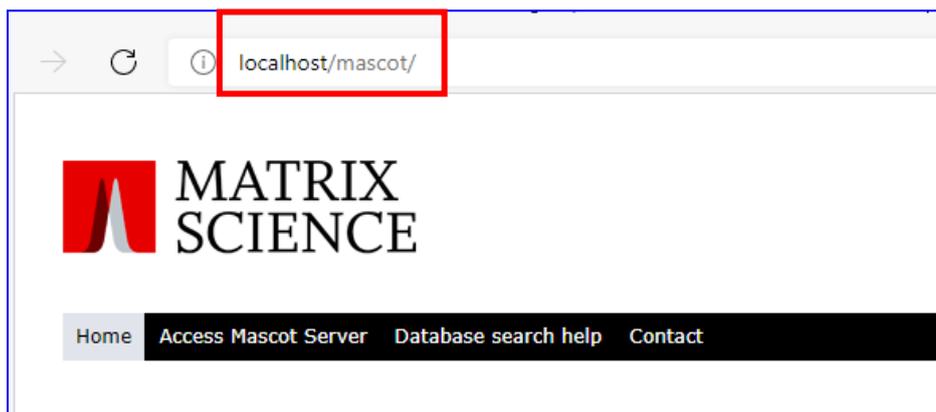
http://(computer 名)/mascot/ 例: <http://mascotserver/mascot/>

http://(IP アドレス)/mascot/ 例: <http://192.168.100.222/mascot/>

MASCOT Server のコンピューター自身から MASCOT を指定する場合、下図のように

<http://localhost/mascot/>

という URL で指定できることがほとんどです。



一方 MASCOT Server とは別のコンピューターから MASCOT Server を URL で指定する場合、コンピューター名と IP アドレスのどちらが、あるいは両方使用可能かについてはケースバイケースです。呼び出しレベルとしては IP アドレスの方がより原始的な仕組みになるため、**MASCOT Server に固定 IP アドレスが割り振られている場合は IP アドレスでの指定をする方がより可能性が高く接続できます。**一方ネットワークの仕組みとして **DHCP など動的な IP 割り当てが行われている場合はコンピューター名による指定**をする事で逐次変更される IP アドレスにも対応する事が可能です。

## 4-2. 変換済みファイルを WEB ブラウザで検索

何らかの方法でデータ変換を行った入力ファイルを手元に持っている場合、WEBブラウザを使って MASCOT Server の検索を行う事ができます。PMFの場合と MIS の場合に分けてご案内します。

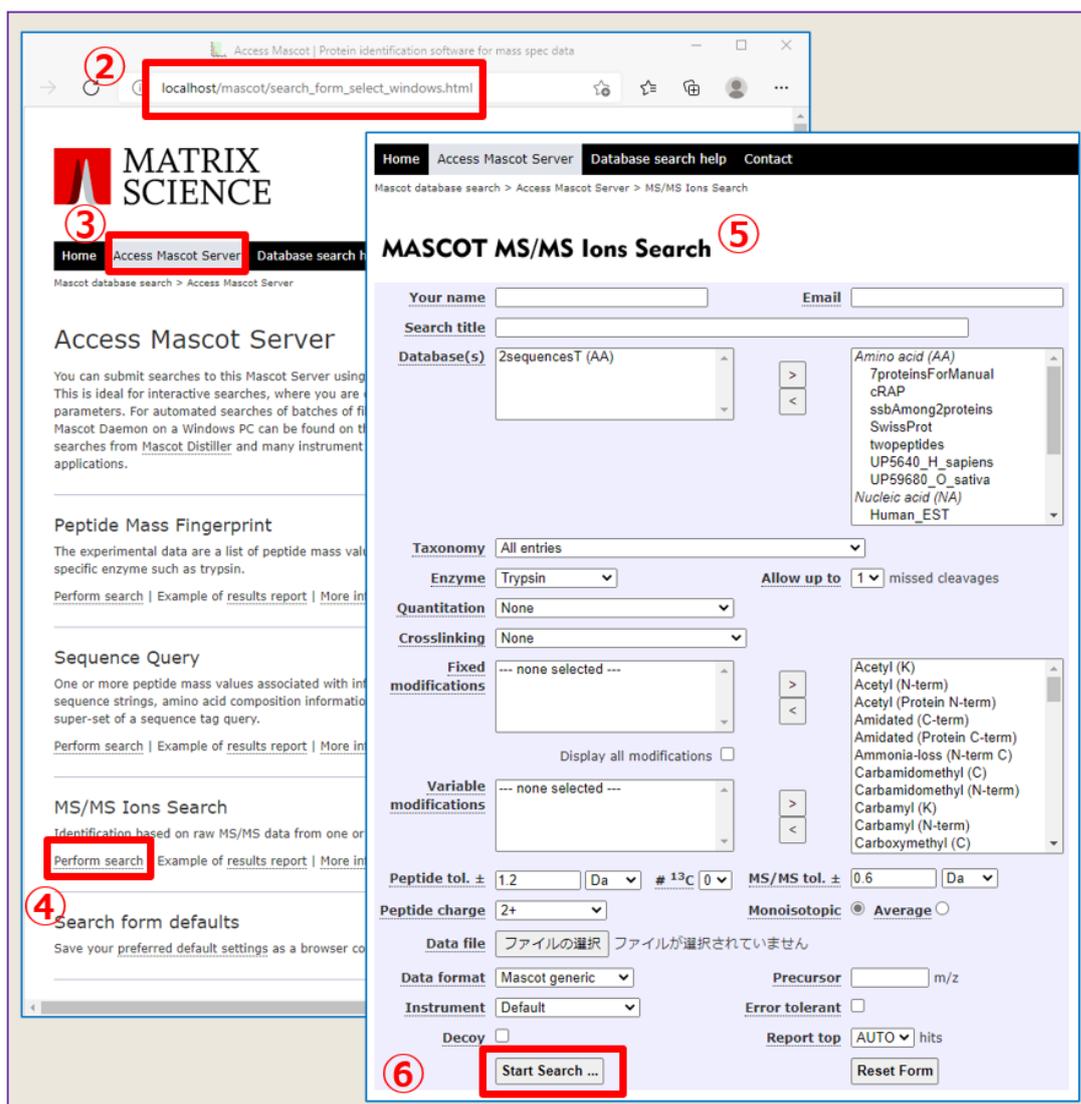
### 4-2-1. PMF : テキストデータを WEB ブラウザで検索

- ① ピークリストファイル(テキスト)を準備
- ② ブラウザを開き、MASCOT Server へアクセス
- ③ Home -> Access MASCOT Server
- ④ 「Peptide Mass Fingerprint」の「Perform search」
- ⑤ 検索パラメーター並びに入力データを指定
- ⑥ 「Start search」ボタンを押すことで検索実行

The screenshot shows the MASCOT Peptide Mass Fingerprint search interface. The browser address bar shows the URL `localhost/mascot/search_form_select_windows.html` (annotated with ②). The page title is "MASCOT Peptide Mass Fingerprint" (annotated with ⑤). The interface includes a navigation menu with "Home", "Access Mascot Server", "Database search help", and "Contact". The main content area is titled "MASCOT Peptide Mass Fingerprint" and contains several search parameters: "Your name" and "Email" (text boxes), "Search title" (text box), "Database(s)" (dropdown menu with "Human\_EST", "2sequencesT", "7proteinsForManual", "cRAP", "ssbAmong2proteins" selected), "Enzyme" (dropdown menu with "Trypsin" selected), "Allow up to" (dropdown menu with "1" selected), "missed cleavages" (checkbox), "Taxonomy" (dropdown menu with "All entries" selected), "Fixed modifications" (dropdown menu with "-- none selected --" selected), "Variable modifications" (dropdown menu with "-- none selected --" selected), "Protein mass" (text box with "kDa" unit), "Peptide tol. ±" (text box with "1.2" and "Da" unit), "Mass values" (radio buttons for "MH+", "M", "M-H"), "Monoisotopic" (radio button) and "Average" (radio button), "Data file" (radio button) with a "ファイルの選択" button and "ファイルが選択されていません" text, "Query" (radio button), and "Data input" (text area). At the bottom, there is a "Decoy" checkbox, a "Report top" dropdown menu with "AUTO" selected, and a "Start Search ..." button (annotated with ⑥). A "Reset Form" button is also present.

## 4-2-2. MIS : mgf ファイルを WEB ブラウザで検索

- ① mgf/mzML ファイルを準備
- ② ブラウザを開き、MASCOT Server へアクセス
- ③ Home -> Access MASCOT Server
- ④ 「MS/MS Ions Search」の「Perform search」
- ⑤ 検索パラメーター並びに入力データを指定
- ⑥ 「Start search」ボタンを押すことで検索実行



## 4-3. Daemon を使って raw データを直接検索

毎回 raw ファイルから mgf ファイルを準備するのが大変な場合、MASCOT Daemon 上で変換プログラムを介して raw データから自動的に変換しつつ検索する方法があります。

詳細は

[https://www.matrixscience.co.jp/supportpdf/MASCOTDaemon\\_ver28\\_manual.pdf](https://www.matrixscience.co.jp/supportpdf/MASCOTDaemon_ver28_manual.pdf)

の「3.チュートリアル」(P.11～)をご覧ください。

変換プログラムとして ProteoWizard と MASCOT Distiller を使用方法についてそれぞれご案内します。

### 4-3-1. Daemon + ProteoWizard

先程ご案内した

[https://www.matrixscience.co.jp/supportpdf/MASCOTDaemon\\_ver28\\_manual.pdf](https://www.matrixscience.co.jp/supportpdf/MASCOTDaemon_ver28_manual.pdf)

の「3.チュートリアル」(P.11～)は、ProteoWizard を使って変換して検索を行っています。詳しくはチュートリアルをご覧ください。

別資料内の内容をまとめた、簡単な手順について以下にご案内します。

- ① Daemon を起動
- ② Parameter Editor タブで検索条件を指定し、条件をファイルに保存するため「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、パラメータファイルを指定
- ④ Task Editor タブの「Data import filter」で「ProteoWizard msConvert」を選択し、「Options」の「Peak list format」で「MGF」や「mzML」を選択してから「OK」ボタン
- ⑤ Task Editor タブで「Run」ボタンを押すことで検索実行
- ⑥ Status タブに検索の進捗が表示。検索完了すると結果画面へのURLが表示されるのでクリック

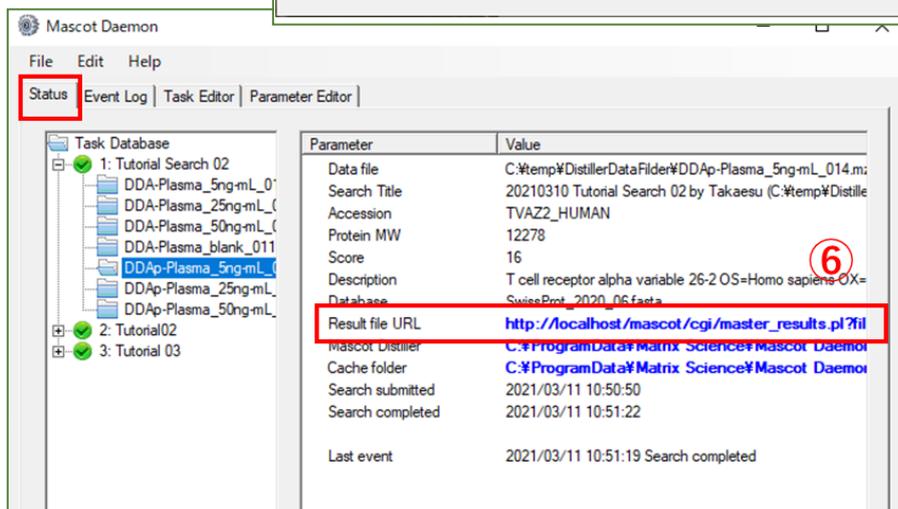
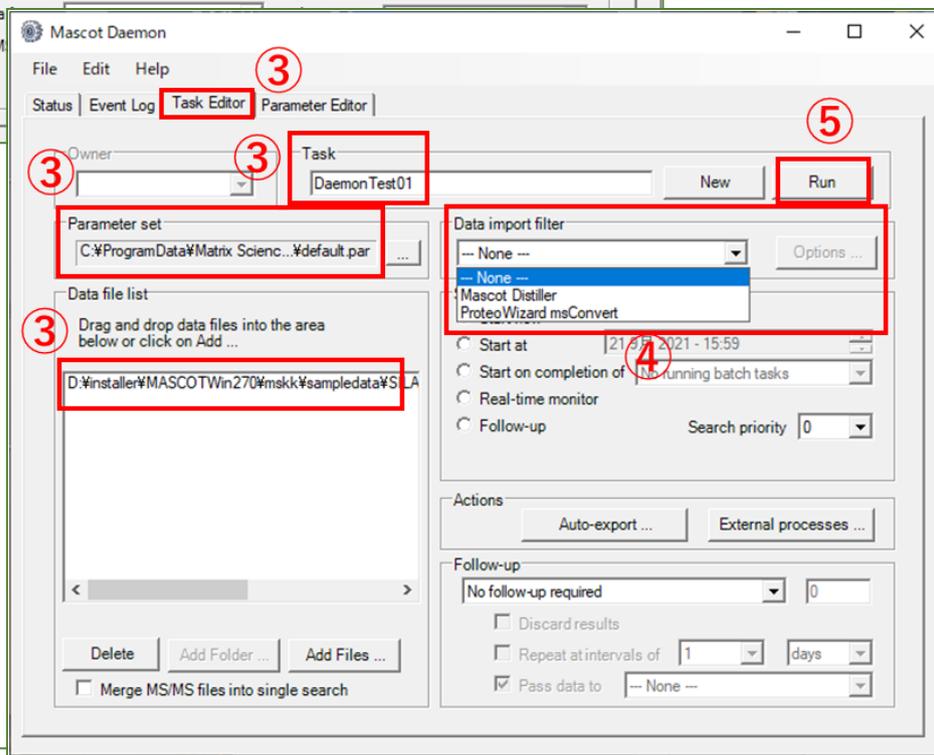
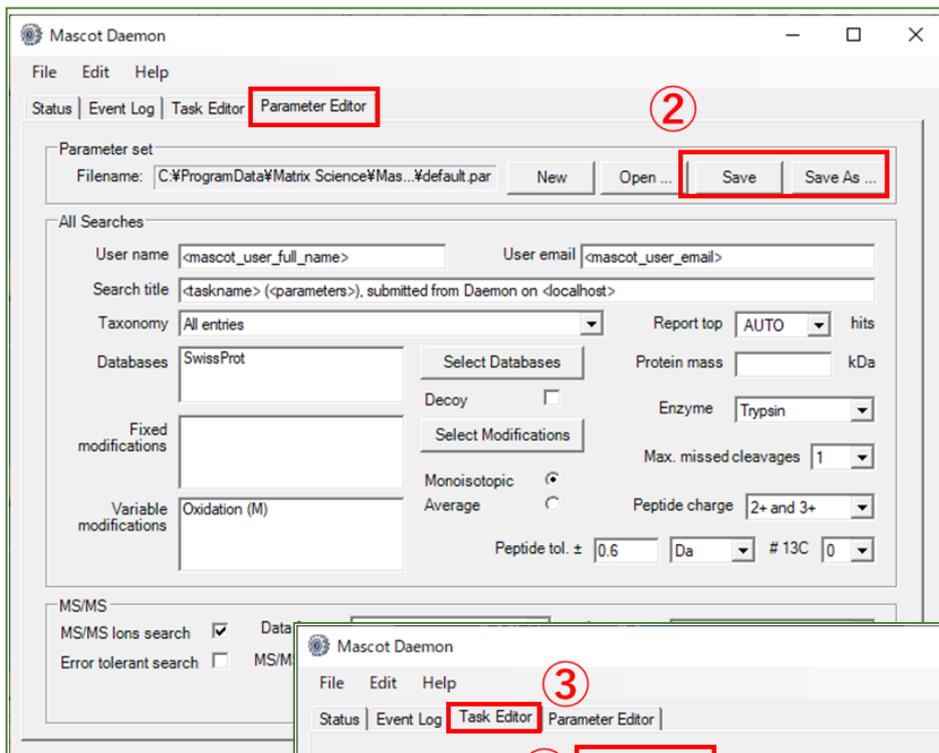
①～⑥の手順について理解の一助となる図が次頁にございますのでそちらも併せてご参照ください。

### 4-3-2. Daemon + Distiller

Mascot Distiller と同じコンピューターにインストールされた Daemon では、変換プログラムに Mascot Distiller を使用する事ができます。手順は **4-3-1** とほぼ同じですが、データ変換部分の指定箇所である④が少し異なります。

- ① Daemon を起動
- ② Parameter Editor タブで検索条件を指定し、条件をファイルに保存するため「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、パラメータファイルを指定
- ④ Task Editor タブの「Data import filter」で「Mascot Distiller」を選択し、「Options」の「Data File Format」で raw ファイルの形式を、「Mascot Distiller Processing Options」で変換の際に使用するピーク変換用のパラメータセットを選択してから「OK」ボタン
- ⑤ Task Editor タブで「Run」ボタンを押すことで検索実行
- ⑥ Status タブに検索の進捗が表示。検索完了すると結果画面へのURLが表示されるのでクリック

①～⑥の手順について理解の一助となる図が次頁にございますのでそちらも併せてご参照ください。



## 4-4. 質量分析装置メーカーのソフトウェアから検索

弊社で準備している検索の方法とは別に、質量分析装置メーカーにて販売している各種解析ソフトウェアから MASCOT を呼び出して検索を行う方法があります。その場合、MASCOT Server と装置メーカーソフトウェアは基本的に別のコンピューターにインストールされており、装置メーカーソフトウェア上で MASCOT Server を URL で指定する事になります。

接続がうまくいかない場合どのレベルで接続できていないかを確認するため、**ソフトウェアで設定を行う前に同じコンピューターの WEB ブラウザで MASCOT を URL 指定して、Home 画面を開く事ができるか必ずご確認ください事を推奨**します。

検索後、結果をどのように利用するかはソフトウェアによって異なります。検索後 MASCOT Server の検索結果画面がブラウザで表示されるケースもあれば、MASCOT Server の検索結果が装置メーカーソフトウェアに取り込まれ、そのソフトウェアの仕様で表示される事もあります。いずれのケースにおいても MASCOT Server で検索された内容は必ず MASCOT Server に残り、MASCOT の Search log などから確認する事が可能です。

## 5. 検索パラメーターとデータベース

この章では MASCOT 検索時に指定するパラメーターについて説明しています。

**5-1**～**5-3** では各検索手法におけるパラメーターについて、検索手法別に1つ1つ説明をしています。**5-4** ではパラメーターの中でカスタマイズ可能な項目について、**5-5** では検索対象のデータベースについて、**5-6** では WEB ブラウザで検索を行う際にブラウザの cookie 機能を使って予めいつも使うパラメーターの情報を保存する方法について説明しています。

**MATRIX SCIENCE**

Search this site

Home Access Mascot Server Database search help Contact Useful links

Mascot database search > Access Mascot Server

### Access Mascot Server

You can submit searches to this Mascot Server using the web browser search forms, below. This is ideal for interactive searches, where you are experimenting with the search parameters. For automated searches of batches of files, try Mascot Daemon. Links to install Mascot Daemon on a Windows PC can be found on the [home page](#). You can also submit searches from [Mascot Distiller](#) and many instrument data systems and third party software applications.

#### Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

[Perform search](#) | [Example of results report](#) | [More information](#)

#### Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

[Perform search](#) | [Example of results report](#) | [More information](#)

#### MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

[Perform search](#) | [Example of results report](#) | [More information](#)

#### Search form defaults

Save your [preferred default settings](#) as a browser cookie.

### More info

- > [Mascot overview](#)
- > [Search parameter reference](#)
- > [Data file format](#)
- > [Results report overview](#)



©2021 Matrix Science | Terms of use

## 5-1. PMF : 検索パラメーター 一覧

### MASCOT Peptide Mass Fingerprint

(1) **Your name**  (2) **Email**

(3) **Search title**

(4) **Database(s)**   
UP5640\_H\_sapiens

(5) **Enzyme**

(6) **Allow up to**  missed cleavages

(7) **Taxonomy**

(8) **Fixed modifications**

Display all modifications

(9) **Variable modifications**

(10) **Protein mass**  kDa (11) **Peptide tol. ±**

(12) **Mass values**  MH<sup>+</sup>  M<sub>r</sub>  M-H<sup>-</sup> (13) **Monoisotopic**  **Average**

(14)  **Data file**  選択されていません

**Query**

**Data input**

(15) **Decoy**

(16)  (17)

PMF 検索のパラメーターについて、以降順に説明いたします。

### (1) Your name, (2) Email, (3) Search title

次頁図(赤枠)のように、検索結果画面並びに Search log において検索の内容や検索者を確認する際に使用します。任意の文字列を入力可能です。Email については Email で検索結果を通知するなどの設定をしていない場合、メモ欄の代わりとして使用する事も可能です。

**MASCOT Search Results**

User : Monitor Test DB 0  
 E-mail :  
 Search title : MS/MS Test Search  
 MS data file : test\_search.mgf  
 Database : UP5640\_H\_sapiens 20201007 (100,100 sequences; 40,284,240 residues)  
 Timestamp : 12 Aug 2021 at 03:59:52 GMT

Re-search All Non-significant Unassigned [help] Export As XML

**MASCOT search log**

Version: 2.8.0.1 - mskk (UC2H-LHC6-Z3V8-W62R-K3M8)

Sort/filter Log File: [logs/searches.log] Start at: (-1=end, 1=start) [1] how many: [50] [6] in log: [6] after filters Data dir: [ ] GETs? [ ]

Job#	PID	dbase	User Name	Email	Title
1242	12096	SwissPro			Copy of raw 03 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default.pa submitted from
1241	5016	UP5640_H	Monitor Test DB 0		MS/MS Test Search
1240	11760	SwissPro			Copy of mgf 02 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default.pa submitted from
1239	8376	SwissPro	Monitor Test DB 0		MS/MS Test Search

#### (4) Database(s)

検索対象のデータベースを選択します。データベースには以下の 2 種類から選択可能です。

- **AA** : Amino Acid, アミノ酸配列
- **NA** : Nucleic Acid, 核酸配列

Ctrl キーを押しながらクリックする事で複数のデータベースを選択する事もできます。

#### (5) Enzyme

タンパク質の切断パターンを指定します。

#### (6) 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ますが、何度まで見逃すことを許容するかについての設定です。例えば Trypsin 設定では K,R の C 末端が切断されますが、

EGR**NR**FPFLSLSQR

という配列があった場合、missed cleavage 設定が 0 なら

EGR

NR

FPFLSLSQR

という 3 種類のペプチドのみを考慮します。missed cleavage が 1 の場合、上記に加え

EGR**NR**

NRFPFLSLSQR

の 2 種類のペプチドも考慮する対象として追加されます。

## (7) Taxonomy

**生物種**の絞り込みに関する設定です。Taxonomy 設定がされているデータベース(SwissProt など)のみ適用可能です。設定されていないデータベースに使用した場合、エラーメッセージは出ますが検索はそのまま実行する事ができます。

リストに記載されている生物種はユーザーがカスタマイズする事も可能です。

## (8) Fixed Modification, (9) Variable Modification

**修飾**に関する設定です。設定した項目について、対象アミノ酸の質量の変化を考慮します。(8)の Fixed は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します。一方 (9) の Variable は修飾がつくパターンとつかないパターンの両方を考慮します。両方を考慮する分融通が利くように見えますが、検索時間が長くなる事とスコアリングで不利になり同定しにくくなるというデメリットがあります。指定する場合は右側にあるリストから該当項目を選び、真ん中にある < を使って Fixed または Variable の内容として指定します。

またリストに初期表示される修飾は使用可能な設定のごく一部で、頻度の低い項目については初期に表示されないようになっています。それらのその他多くの設定を表示させるには、画面中央にある「**Display all modifications**」にチェックを入れる事で**右側のリストが変化**します。

## (10) Protein mass

タンパク質の質量に対する上限値です。データベースにエントリーされている配列の全長に対する質量ではなく、マッチしたペプチドのうち最も N 末端側と C 末端側でマッチした領域を対象として、その N 末端から C 末端までの配列の質量が計算対象となります。

## (11) Peptide tol.±

**ペプチドの質量**について、実測値から計算された値とペプチド配列から計算された値との**許容誤差**です。ユーザー側で装置のスペックや事前に行ったキャリブレーションの結果を基に適正値を判断します。

## (12) Mass values

PMF 検索においてクエリーの各ピークが MH<sup>+</sup>か M-H<sup>-</sup>か、あるいはイオンが負荷していない質量に換算されたものなのか(Mr)を指定します。

### (13) Monoisotopic / Average

アミノ酸の質量計算を Monoisotopic で行うか、Average で行うかを指定します。現在は monoisotopic を選択する事がほとんどです。

### (14) Data file / Query\_Data input

検索 query となるデータを指定します。ファイルで渡す場合は「Data file」で該当ファイルを選択します。一方、ファイルでなく直接データを記入して渡す場合、「Query」を指定して、「Data input」欄にデータを記入(貼り付け)します(下図)。データフォーマットについては「3-1-1.PMF に対応するファイルフォーマット」をご覧ください。

Data file ファイルを選択 選択されていません

Query

**Data input**

794.23  
836.92  
911.23  
1029.64  
1119.53  
1218.67

### (15) Decoy

同定結果の検証に必要な「FDR」を算出する事ができますが PMF ではあまり使われていません。詳細は「11.FDRとDecoyデータベース」をご覧ください。

### (16) Start Search

検索を開始します。

### (17) Reset Form

パラメーター設定をデフォルト状態に戻します。

## 5-2. Sequence Query : 検索パラメーター 一覧

**MASCOT Sequence Query**

(1) Your name (2) Email

(3) Search title

(4) Database(s) SwissProt UP5640\_H\_sapiens (5) Enzyme Trypsin (6) Allow up to 1 missed cleavages (7) Quantitation None

(8) Taxonomy All entries

(9) Fixed modifications --- none selected --- (10) Variable modifications --- none selected ---

(11) Peptide tol. ± 1.2 Da (12) # <sup>13</sup>C 0 (13) MS/MS tol. ± 0.6 Da

(14) Peptide charge Mr (15) Monoisotopic  Average C

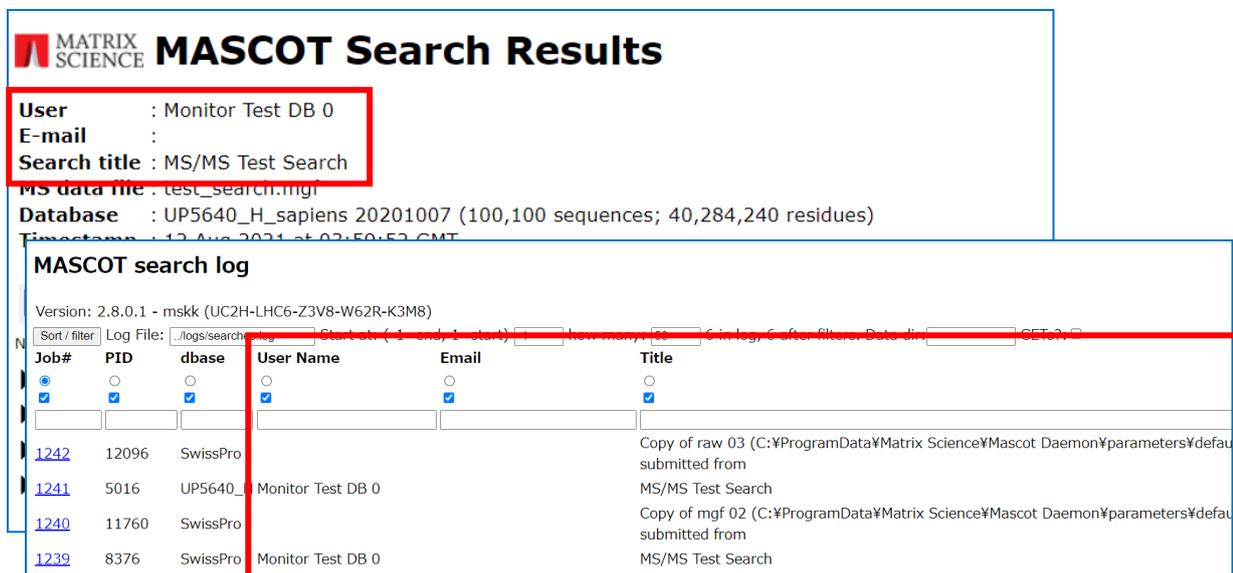
(16) Query

(17) Instrument Default (18) Decoy

(19) Start Search ... (20) Reset Form

### (1) Your name, (2) Email, (3) Search title

検索結果画面並びに Search log において検索の内容や検索者を確認する際に使用します。任意の文字列を入力可能です。Email については Email で結果を返すなどの設定をしていない場合、メモ欄の代わりに使用する事もできます。



MATRIX SCIENCE **MASCOT Search Results**

User : Monitor Test DB 0  
E-mail :  
Search title : MS/MS Test Search

MS data file : test\_search.mgf  
Database : UP5640\_H\_sapiens 20201007 (100,100 sequences; 40,284,240 residues)  
Timestamp : 12 Aug 2021 at 02:50:53 GMT

**MASCOT search log**

Version: 2.8.0.1 - mskk (UC2H-LHC6-Z3V8-W62R-K3M8)

Sort/filter Log File: [./logs/search\_log] Start at: (1 end: 1 start) [1] How many: [50] Show log: [6 after filter: Data dir: [CET37]

Job#	PID	dbase	User Name	Email	Title
1242	12096	SwissPro	Monitor Test DB 0		Copy of raw 03 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default) submitted from
1241	5016	UP5640_	Monitor Test DB 0		MS/MS Test Search
1240	11760	SwissPro	Monitor Test DB 0		Copy of mgf 02 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default) submitted from
1239	8376	SwissPro	Monitor Test DB 0		MS/MS Test Search

### (4) Database(s)

検索対象のデータベースを選択します。データベースには以下の 2 種類があります。

- **AA**: Amino Acid, アミノ酸配列
- **NA**: Nucleic Acid, 核酸配列

Ctrl キーを押しながらクリックする事で複数のデータベースを選択する事もできます。

### (5) Enzyme

タンパク質の切断パターンを指定します。

### (6) 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ますが、何度まで見逃すことを許容するかについての設定です。例えば Trypsin 設定で K,R の C 末端が切断されますが、

EG**RNR**FPFLSLSQ**R**

という配列があった場合、missed cleavage 設定が 0 なら

EG**R**

**NR**

FPFLSLSQ**R**

という 3 種類のペプチドのみを考慮します。missed cleavage が 1 の場合、上記に加え

EG**RNR**

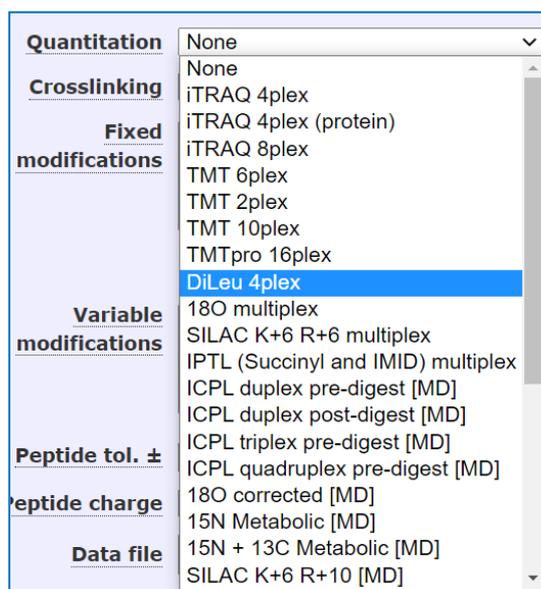
**NR**FPFLSLSQ**R**

の 2 種類のペプチドも考慮する対象として追加されます。

## (7) Quantitation

**定量計算**に関する設定項目です。MASCOT では Spectral Counting の 1 種である emPAI についてはここで指定することなく結果画面に自動的に表示されます。

**この設定項目は何の準備もなく利用できるものではありません。** 選択項目のうち、後ろに **[MD]** がついていないものについては MASCOT Server 単独でも計算が可能ですが、計算をするためにはそれに合わせた測定を予め行っておく必要があります。**[MD]** が後ろについている項目については計算のためにソフトウェア MASCOT Distiller の計算モジュールまで搭載されたものがが必要です。またラベルフリーの手法を除き、計算のためにはそれに合わせた測定をあらかじめ行っておく必要があります。計算を行うためには、様々な設定を事前に行いパッケージ化して名称を設定した項目を検索時に指定する事になります。詳細は「**10-2. Quantitation**」をご覧ください。



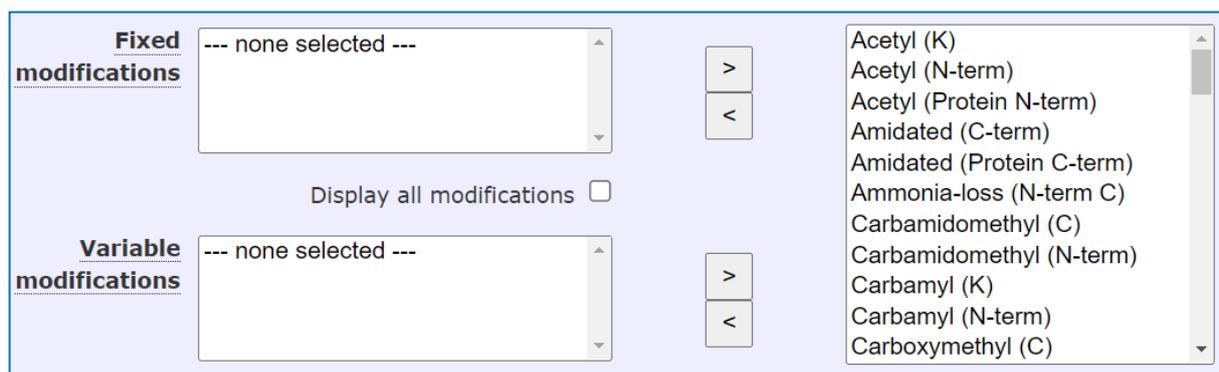
## (8) Taxonomy

**生物種**の絞り込みに関する設定です。Taxonomy 設定がされているデータベース(SwissProt など)のみ適用可能です。設定されていないデータベースに使用した場合、エラーメッセージは出ますが検索はそのまま実行する事ができます。

リストに記載されている生物種はユーザーがカスタマイズする事も可能です。

## (9) Fixed Modification, (10) Variable Modification

**修飾**に関する設定です。設定した項目について、対象アミノ酸の質量の変化を考慮します。(10)の Fixed は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します。一方 (11) の Variable は修飾がつくパターンとつかないパターンの両方を考慮します。両方を考慮する分融通が利くように見えますが、検索時間が長くなる事と同定判定で不利になり同定しにくくなるというデメリットがあります。指定する場合は右側にあるリストから該当項目を選び、真ん中にある < を使って Fixed または Variable の内容として指定します。



またリストに初期表示される修飾は使用可能な設定のごく一部で、頻度の低い項目については初期に表示されないようになっています。それらのその他多くの設定を表示させるには、画面中央にある「**Display all modifications**」にチェックを入れる事で右側のリストが変化します。

#### (11) Peptide tol.±

ペプチドの質量について、実測値から計算された値とペプチド配列から計算された値との許容誤差です。ユーザー側で装置のスペックや事前に行ったキャリブレーションの結果を基に適正値を判断します。

#### (12) #<sup>13</sup>C

質量分析装置での測定時、<sup>12</sup>C のみから構成されるペプチドではなく <sup>13</sup>C を含むペプチドを取り込んでいることがあります。その際生じる理論値とのずれを補正するためのパラメーターです。ペプチドの理論値に対して適応されます。上記(11)の Peptide to.±を TOL、実験値並びに理論値の質量をそれぞれ exp,calc と表現する場合、通常は

**TOL > |exp - calc|** の時のみをマッチとみなしますが、

この設定値を 1 とした場合は **TOL > |exp - calc - 1|** もマッチとみなします。

また設定値を 2 とした場合は上記に加え **TOL > |exp - calc - 2|** もマッチとみなします。

#### (13) MS/MS tol.±

ペプチドのフラグメントの質量について、実測値から計算された値とペプチド配列から計算された値との許容誤差です。装置のスペックや事前に行ったキャリブレーションの結果をもとに適正値を判断します。

#### (14) Peptide charge

通常は使用されないパラメーターです。クエリーデータの中に charge、すなわちペプチドの電荷に関する情報が含まれない場合、ここで指定した値が使用されます。ただしほとんどのケースにおいてペプチドの電荷に関する情報はファイルに含まれていて、その場合ここで指定した値は無視されます。ファイル内で電荷状況を示す charge 行は、仮に電荷が特定できないケースでも推定値として charge=2+,3+,4+ などと記入されていることが多いため、このパラメーターが使用されることはあまりありません。

#### (15) Monoisotopic / Average

アミノ酸の質量計算を Monoisotopic で行うか、Average で行うかを指定します。現在は monoisotopic を選択する事がほとんどです。

#### (16) Query

検索 query となる入力データを指定します。アミノ酸配列の並びや組成に関する情報をフィルターとして利用する事ができます。Sequence Query の文法については

[https://www.matrixscience.com/help/sq\\_help.html](https://www.matrixscience.com/help/sq_help.html)

をご覧ください。

## (17) Instrument

フラグメントピークと理論値をマッチする際、考慮するイオンシリーズやフラグメントピークの電荷に関する情報が定義されたセット(下図)です。設定値が変わる事で理論値とのマッチング状況が変わり、MASCOT の Ion Score が変わり、ひいては同定結果も変わってくる場合があります。

	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF TOF	ESI 4 SECT	FTMS ECD	ETD TRAP	MALDI QUAD TOF	MALDI QIT TOF	MALDI ISD
1 <sup>+</sup> fragments	X	X	X	X	X	X	X	X	X	X	X	X	X
2 <sup>+</sup> fragments if precursor 2 <sup>+</sup> or higher	X	X		X	X	X		X	X	X	X		
2 <sup>+</sup> fragments if precursor 3 <sup>+</sup> or higher													
Immonium ions			X				X	X			X	X	
a series ions	X		X				X	X				X	X
a-NH <sub>3</sub> if fragment includes RKNQ	X		X				X					X	
a-H <sub>2</sub> O if fragment includes STED			X				X					X	
b series ions	X	X	X	X	X	X	X	X			X	X	
b-NH <sub>3</sub> if fragment includes RKNQ	X	X	X	X	X	X	X	X			X	X	
b-H <sub>2</sub> O if fragment includes STED		X	X	X	X	X	X	X			X	X	
c series ions									X	X			X
x series ions													
y series ions	X	X	X	X	X	X	X	X	X	X	X	X	X
y-NH <sub>3</sub> if fragment includes RKNQ	X	X		X	X	X	X				X	X	
y-H <sub>2</sub> O if fragment includes STED		X		X	X	X	X				X	X	
z series ions								X					

## (18) Decoy

ペプチドの同定結果を検証するために必要な“FDR”を算出することができます。FDR を計算したい場合はチェックを入れてください。検索時間としては入れない場合に比べ2倍強かかります。詳細は「11. FDR と Decoy データベース」をご覧ください。

### (19) Start Search

検索を開始します。

### (20) Reset Form

パラメーター設定をデフォルト状態に戻します。

## 5-3. MIS : 検索パラメーター 一覧

### MASCOT MS/MS Ions Search

(1) **Your name**  (2) **Email**

(3) **Search title**

(4) **Database(s)**     
*Spectral library (SL)*  
PRIDE\_Contaminants  
*Amino acid (AA)*  
UP5640\_H\_sapiens

(5) **Taxonomy**

(6) **Enzyme**  (7) **Allow up to**  missed cleavages

(8) **Quantitation**

(9) **Crosslinking**

(10) **Fixed modifications**     
Acetyl (K)  
Acetyl (N-term)  
Acetyl (Protein N-term)  
Amidated (C-term)  
Amidated (Protein C-term)  
Ammonia-loss (N-term C)  
Carbamidomethyl (C)  
Carbamidomethyl (N-term)  
Carbamyl (K)  
Carbamyl (N-term)  
Carboxymethyl (C)

Display all modifications

(11) **Variable modifications**

(12) **Peptide tol.  $\pm$**    (13) **#  $^{13}\text{C}$**   (14) **MS/MS tol.  $\pm$**

(15) **Peptide charge**  (16) **Monoisotopic**  **Average**

(17) **Data file**  選択されていません

(18) **Data format**  (19) **Precursor**  m/z

(20) **Instrument**  (21) **Error tolerant**

(22) **Decoy**  (23) **Target PSM FDR**

(24)  (25)

## (1) Your name, (2) Email, (3) Search title

検索結果画面並びに Search log において検索の内容や検索者を確認する際に使用します(下図)。任意の文字列を入力可能です。Email については Email で結果を返すなどの設定をしていない場合、メモ欄の代わりに使用する事もできます。

**MASCOT Search Results**

**User** : Monitor Test DB 0  
**E-mail** :  
**Search title** : MS/MS Test Search  
**MS data file** : test\_search.mgf  
**Database** : UP5640\_H\_sapiens 20201007 (100,100 sequences; 40,284,240 residues)

**MASCOT search log**

Version: 2.8.0.1 - mskk (UC2H-LHC6-Z3V8-W62R-K3M8)

Sort/filter Log File: /logs/searches.log Start at: (-1=end, 1=start) how many: 50 6 in log, 6 after filters. Data dir: GEIS?:

Job#	PID	dbase	User Name	Email	Title
1242	12096	SwissPo			Copy of raw 03 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default.pa submitted from
1241	5016	UP5640_H	Monitor Test DB 0		MS/MS Test Search
1240	11760	SwissPo			Copy of mgf 02 (C:\ProgramData\Matrix Science\Mascot Daemon\parameters\default.pa submitted from
1239	8376	SwissPo	Monitor Test DB 0		MS/MS Test Search

## (4) Database(s)

検索対象のデータベースを選択します。データベースには以下の3種類があります。

- **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データ

Ctrl キーを押しながらクリックする事で複数のデータベースを選択する事もできます。

## (5) Taxonomy

生物種の絞り込みに関する設定です。Taxonomy 設定がされているデータベース(SwissProt など)のみ適用可能です。設定されていないデータベースに使用した場合、エラーメッセージは出ますが検索はそのまま実行する事ができます。

リストに記載されている生物種はユーザーがカスタマイズする事も可能です。

## (6) Enzyme

タンパク質の切断パターンを指定します。

## (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ますが、何度まで見逃すことを許容するかについての設定です。例えば Trypsin 設定では K,R の C 末端が切断されます。

EGR**NR**FPFLSLSQR

という配列があった場合、missed cleavage 設定が 0 なら

EGR

NR

FPFLSLSQR

という 3 種類のペプチドのみを考慮します。missed cleavage が 1 の場合、上記に加え

EGR**NR**

NRFPFLSLSQR

の 2 種類のペプチドも考慮する対象として追加されます。

## (8) Quantitation

**定量計算**に関する設定項目です。MASCOT では Spectral Counting の 1 種である emPAI についてはここで指定することなく結果画面に自動的に表示されます。**この設定項目は何の準備もなく利用できるものではありません。** 選択項目のうち、後ろに **[MD]** がついていないものについては MASCOT Server 単独でも計算が可能ですが、計算をするためにはそれに合わせた測定をあらかじめ行っておく必要があります。**[MD]** が後ろについている項目については計算のためにソフトウェア MASCOT Distiller の計算モジュールまで搭載されたものがが必要です。またラベルフリーの手法を除き、計算のためにはそれに合わせた測定をあらかじめ行っておく必要があります。

Quantitation	None
Crosslinking	None
Fixed modifications	iTRAQ 4plex iTRAQ 4plex (protein) iTRAQ 8plex TMT 6plex TMT 2plex TMT 10plex TMTpro 16plex DiLeu 4plex
Variable modifications	18O multiplex SILAC K+6 R+6 multiplex IPTL (Succinyl and IMID) multiplex ICPL duplex pre-digest [MD] ICPL duplex post-digest [MD] ICPL triplex pre-digest [MD] ICPL quadruplex pre-digest [MD]
Peptide tol. ±	18O corrected [MD]
Peptide charge	15N Metabolic [MD]
Data file	15N + 13C Metabolic [MD] SILAC K+6 R+10 [MD]

計算を行うためには、様々な設定を事前に行いパッケージ化して名称を設定した項目を検索時に指定する事になります。詳細は「**10-2.Quantitation**」をご覧ください。

## (9) Crosslinking

ペプチドがリンカーまたは共有結合で結合したデータを検索する事ができます。この設定についても **Quantitation** 同様何の準備もなく利用できるものではありません。計算を行うためには、それに合わせた測定(リンカーを添加したりSS結合を切断する還元処理を行わないなど)を予め行っておく必要があるほか、様々な設定を事前に行いパッケージ化して名称を設定した項目を検索時に指定する事になります。詳細は「**10-3.Crosslink**」をご覧ください。

[次頁に続きます]

## (10) Fixed Modification, (11) Variable Modification

修飾に関する設定です。設定した項目について、対象アミノ酸の質量の変化を考慮します。(10)の Fixed は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します。一方 (11) の Variable は修飾がつくパターンとつかないパターンの両方を考慮します。両方を考慮する分融通が利くように見えますが、検索時間が長くなる事と同定判定で不利になり同定しにくくなるというデメリットがあります。指定する場合は右側にあるリストから該当項目を選び、真ん中にある「<」を使って Fixed または Variable の内容として指定します。

リストに初期表示される修飾は使用可能な設定のごく一部で、頻度の低い項目については初期に表示されないようになっています。それらのその他多くの設定を表示させるには、画面中央にある「**Display all modifications**」にチェックを入れる事で右側のリストが変化します。

## (12) Peptide tol.±

ペプチドの質量について、実測値から計算された値とペプチド配列から計算された値との許容誤差です。ユーザー側で装置のスペックや事前に行ったキャリブレーションの結果を基に適正値を判断します。

## (13) #<sup>13</sup>C

質量分析装置での測定時、<sup>12</sup>C のみから構成されるペプチドではなく <sup>13</sup>C を含むペプチドを取り込んでいることがあります。その際生じる理論値とのずれを補正するためのパラメーターです。ペプチドの理論値に対して適応されます。上記(11)の Peptide to.±を TOL、実験値並びに理論値の質量をそれぞれ exp,calc と表現する場合、通常は

**TOL > |exp - calc|** の時のみをマッチとみなしますが、

この設定値を 1 とした場合は **TOL > |exp - calc - 1|** もマッチとみなします。

また設定値を 2 とした場合は上記に加え **TOL > |exp - calc - 2|** もマッチとみなします。

## (14) MS/MS tol.±

ペプチドのフラグメントの質量について、実測値から計算された値とペプチド配列から計算された値との許容誤差です。ユーザー側で装置のスペックや事前に行ったキャリブレーションの結果を基に適正値を判断します。

### (15) Peptide charge

通常は使用されないパラメーターです。クエリーデータの中に charge、すなわちペプチドの電荷に関する情報が含まれない場合、ここで指定した値が使用されます。ただしほとんどのケースにおいてペプチドの電荷に関する情報はファイルに含まれていて、その場合ここで指定した値は無視されます。ファイル内で電荷状況を示す charge 行は、仮に電荷が特定できないケースでも推定値として charge=2+,3+,4+ などと記入されていることが多いため、このパラメーターが使用されることはあまりありません。

### (16) Monoisotopic / Average

アミノ酸の質量計算を Monoisotopic で行うか、Average で行うかを指定します。現在は monoisotopic を選択する事がほとんどです。

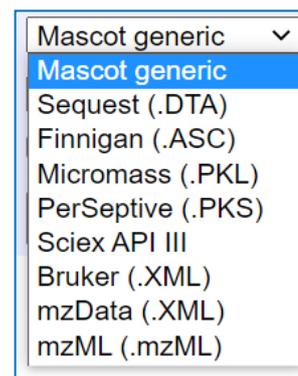
### (17) Data file

検索 query となる入力データを指定します。WEB ブラウザで検索を行う場合、raw データではなく判読可能なテキストまたは XML フォーマットに変換されたデータを指定する必要があります。

### (18) Data format

判読可能なテキストフォーマット並びに XML フォーマットのデータを読み込む際、そのフォーマットを選択します。

現在主に使用されるものとして、「Mascot generic」(mgf という拡張子、名称がよく使われます)と「mzML(.mzML)」があり、そのほかにも右図のような各種フォーマットに対応しています。詳細は「**3-1-2.**」～「**3-1-4.**」をご参照ください。



### (19) Precursor

通常は使用されないパラメーターです。クエリーデータの中に pepmass、すなわちペプチドの m/z に関する情報が含まれない場合、ここで指定した値が m/z として使用されます。ただしほとんどのケースにおいてペプチドの m/z に関する情報が含まれていて、その場合ここで指定した値は無視されます。

### (20) Instrument

フラグメントピークと理論値をマッチする際、**考慮するイオンシリーズやフラグメントピークの電荷**に関する情報が定義されたセットです(次頁図参照)。設定値が変わる事で理論値とのマッチング状況が変わり、MASCOT の Ion Score が変わり、ひいては同定結果も変わってくる場合があります。

	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF TOF	ESI 4 SECT	FTMS ECD	ETD TRAP	MALDI QUAD TOF	MALDI QIT TOF	MALDI ISD
1 <sup>+</sup> fragments	X	X	X	X	X	X	X	X	X	X	X	X	X
2 <sup>+</sup> fragments if precursor 2 <sup>+</sup> or higher	X	X		X	X	X		X	X	X	X		
2 <sup>+</sup> fragments if precursor 3 <sup>+</sup> or higher													
Immonium ions			X				X	X			X	X	
a series ions	X		X				X	X				X	X
a-NH <sub>3</sub> if fragment includes RKNQ	X		X				X					X	
a-H <sub>2</sub> O if fragment includes STED			X				X					X	
b series ions	X	X	X	X	X	X	X	X			X	X	
b-NH <sub>3</sub> if fragment includes RKNQ	X	X	X	X	X	X	X	X			X	X	
b-H <sub>2</sub> O if fragment includes STED		X	X	X	X	X	X	X			X	X	
c series ions									X	X			X
x series ions													
y series ions	X	X	X	X	X	X	X	X	X	X	X	X	X
y-NH <sub>3</sub> if fragment includes RKNQ	X	X		X	X	X	X				X	X	
y-H <sub>2</sub> O if fragment includes STED		X		X	X	X	X				X	X	
z series ions								X					

### (21) Error tolerant

1)想定外の修飾、2)非特異的切断を伴うペプチド、3)アミノ酸残基置換 の3つを検出することができる2段階検索です。詳細は「**10-4. Error Tolerant Search**」の項目をご覧ください。

### (22) Decoy

ペプチドの同定結果を検証するために必要な「**FDR**」を算出することができます。FDR を計算したい場合はチェックを入れてください。検索時間としては入れない場合に比べ2倍強かかります。詳細は「**11. FDR と Decoy データベース**」をご覧ください。

### (23) Target PSM FDR

FDRを同定基準として利用する際に設定します。カウントの対象はPSM(Peptide-Spectrum Matches)、即ち query ベースの数え上げとなります。

### (24) Start Search

検索を開始します。

### (25) Reset Form

パラメーター設定をデフォルト状態に戻します。

## 5-4. パラメーターの中でカスタマイズ可能な項目

検索パラメーターの中にはカスタマイズをして利用することができる項目があります。

**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。各設定画面については「**13. MASCOT Server のカスタマイズ**」をご覧ください。

なお設定変更が可能なのは製品版(ローカル版)MASCOT Server のみです。インターネット上で公開されている試用サーバーでは変更ができません。MASCOT Server をカスタマイズしてご利用頂きたい場合は製品版の購入をご検討ください。

## 5-5. Database

MASCOT では検索対象として最適なデータベースを自動的に選択するような仕組みはありません。ユーザーが目的に応じて適切なデータベースを自身で選択する必要があります。

MASCOT で検索可能なデータベースは大きく分けて3種類あります。

**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データ

データベースは「**Database Manager**」で既存データベースについて最新版への更新を行ったり、新規データベースの追加を行ったりする事ができます。新規データベースの追加については、MASCOT 側で予めファイル取得先やデータベース諸設定を定義されたものから選択する方法と、自身で準備した FASTA ファイルをセットする方法があります。Database manager については別紙の設定資料を準備しておりますのでそちらをご覧ください。

[https://www.matrixscience.co.jp/supportpdf/MASCOTServer\\_ver26\\_sequencedbmanage.pdf](https://www.matrixscience.co.jp/supportpdf/MASCOTServer_ver26_sequencedbmanage.pdf)

8-3 や 9-3 で後述するように、MASCOT ではデータベースにないエントリーを同定する事はできませんが、必要以上にエントリー数の多いデータベースを選択してしまうと同定基準値が高くなってしまい結果的にペプチド配列やタンパク質の同定がより難しくなってしまいます。必要十分な検索範囲のデータベースを選択する必要があります。

**使用するべきデータベースが良くわからない場合、最初の選択として「SwissProt」を選択して頂く事をお勧めします。**より網羅的な解析を希望する場合、データベース名が UP で始まり生物種名が名称内に含まれる、Uniprot 系のデータベースの使用をセットして利用する事もお勧めします。

以下、MASCOT が Predefined として準備しているデータベースのうち代表的なものについて説明をいたします。

### ■ SwissProt

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

Uniprot データベースの中の1つで、**EBI**(European Bioinformatics Institute) と **SIB** (Swiss Institute of Bioinformatics)により共同運用されたタンパク質配列のデータベース。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、2021年3月版で **565,254** 件のデータが登録されています。このエントリー数は長期に渡りあまり変動がありません。

### ■ Uniprot → MASCOT データベースでは UPN\_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 SwissProt に加え、自動かつ精査無し of データベース **TrEMBL** を併せたデータベースです。TrEMBL は SwissProt に比べ圧倒的にエントリー数が多く2021年3月版で

**219,174,961** 件のデータが登録されています。SwissProt のみのデータベース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。ただし Uniprot すべてのエントリーでは件数が多すぎるので、生物種を限定したデータベースを準備してそれに対して検索をかける事を推奨しています。

詳細な説明並びにデータベースをセットする具体的な方法については以下日本語資料「**検索対象の生物種を予め絞り込んだ UniprotKB データベースの作成手順**」をご覧ください。

[https://www.matrixscience.co.jp/supportpdf/Uniprotdb\\_20200915.pdf](https://www.matrixscience.co.jp/supportpdf/Uniprotdb_20200915.pdf)

### ■ NCBIprot (旧名称 NCBInr)

**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」です。「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。かつて網羅的な探索用のデータベースとして SwissProt の次の選択肢として弊社にて使用を推奨していました。しかし現在は冗長性膨大なエントリー数により MASCOT Server での更新やセットアップが困難になってしまった事から、**使用を推奨しておりません**。代わりに上述の Uniprot の使用をお勧めしています。

### ■ IPI

International Protein Index にてかつて公開され、一時期多くのユーザーによって利用されていたタンパク質配列のデータベースです。2011年9月に最後のリリースが行われその後更新されていないため、現在は非推奨です。

### ■ XXXXX\_EST (Human\_EST など)

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。一方でマッチング後に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(機能が不明な状態でも得られた配列情報を使って別の解析を進める、など)。

### ■ NIST\_XXXX\_YYYY (NIST\_Human\_HCD など)

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)

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。

その他のデータベースについての説明は以下のページをご参照ください。

[https://www.matrixscience.com/help/database\\_help.html](https://www.matrixscience.com/help/database_help.html)

## 5-6. Search form defaults

「Access Mascot Server」ページの一番下に「Search form defaults」という項目があります(下図)。「preferred default settings」リンクをクリックし開いた画面でパラメータをセットし保存すると、各検索方法の search form 画面を開いた際に保存内容がデフォルトの選択設定として選ばれた状態になっています。

The screenshot shows the 'Access Mascot Server' page. At the bottom, there is a section titled 'Search form defaults' with the text: 'Save your preferred default settings as a browser cookie.' The text 'preferred default settings' is highlighted with a red box. To the right, there is a 'More info' box with links: '> Mascot overview', '> Search parameter reference', '> Data file format', and '> Results report overview'. Below this is an image of a book and a CD. The page also features a navigation menu with 'Home', 'Access Mascot Server', 'Database search help', 'Contact', and 'Useful links'. A search bar is located at the top right.

The screenshot shows the 'Set Mascot search form defaults' page. The navigation menu includes 'Home', 'Mascot database search', 'Products', 'Technical support', 'Training', and 'News'. The breadcrumb trail is 'Access Mascot Server | Database search help'. The main heading is 'Set Mascot search form defaults'. The configuration options are as follows:

- Database:** A dropdown menu with options: contaminants, cRAP, SARS-CoV-2, **SwissProt** (highlighted), and UP186698\_X\_laevis.
- Taxonomy:** A dropdown menu with 'All entries' selected.
- Enzyme:** A dropdown menu with 'Trypsin' selected.
- Allow up to:** A dropdown menu with '1' selected, followed by 'missed cleavages'.
- Fixed modifications:** A list of checkboxes for: Carbamidomethyl (C), Carbamidomethyl (N-term), Carbamyl (K), Carbamyl (N-term), and Carboxymethyl (C).
- Variable modifications:** A list of checkboxes for: mTRAQ:13C(6)15N(2) (N-term), mTRAQ:13C(6)15N(2) (Y), NIPCAM (C), Oxidation (HW), and Oxidation (M).

## 6. 検索結果画面:PMF

この章では MASCOT PMF 検索における結果画面について説明します。

### 6-1. 表示例で使用している検索について

PMF の検索例として MASCOT Server 内に存在する以下の検索結果を利用します。

[公開サーバー] [https://www.matrixscience.com/cgi/master\\_results.pl?file=../data/F981122.dat](https://www.matrixscience.com/cgi/master_results.pl?file=../data/F981122.dat)

[ローカルサーバー] [http://localhost/mascot/cgi/master\\_results.pl?file=../data/F981122.dat](http://localhost/mascot/cgi/master_results.pl?file=../data/F981122.dat)

18 のピークを含んだ query(左下)で、以下のようなパラメーター設定(右下)を行った検索です。同定されるタンパク質は「**PML\_HUMAN**」となります。

814.430
958.350
1000.330
1165.390
1182.440
1191.500
1300.470
1320.400
1348.410
1355.530
1423.520
1426.570
1624.740
2265.110
2544.410
2550.300
2653.390

設定項目	設定値
<b>Database</b>	SwissProt
<b>Enzyme</b>	Trypsin/P
<b>Allow up to</b>	2
<b>Taxonomy</b>	all
<b>Peptide tol,±</b>	0.2 Da
<b>Mass Values</b>	MH <sup>+</sup>
<b>Monoisotopic/Average</b>	Monoisotopic

### 6-2. 表示内容の詳細 : summary 画面

#### 6-2-1. 展開しない状態での画面概要

URL を WEB ブラウザで指定し結果画面を開くと次頁のような画面が現れます。画面内の赤線で囲われた各パーツで記載されている内容について、以降順に説明します。

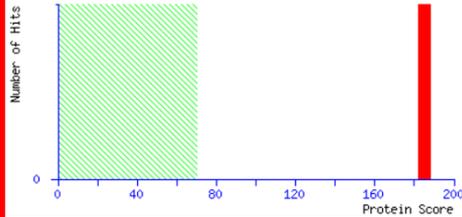
6-2-2 ヘッダー部分

User :  
 Email :  
 Search title : Peptide Mass Fingerprint Example  
 Database : SwissProt 2019\_10 (561356 sequences; 201858328 residues)  
 Timestamp : 9 Jan 2020 at 11:23:29 GMT  
 Top Score : 185 for PML\_HUMAN, Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3

Mascot Score Histogram

6-2-3 Score Histogram

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



Concise Protein Summary Report

6-2-4 表示内容の変更[Format As]

Format As  [Help](#)  
 Significance threshold p<  Max. number of hits   
 Preferred taxonomy

6-2-5 再検索

- [PML\\_HUMAN](#) Mass: 97489 Score: 185 Expect: 1.8e-013 Matches: 16  
 Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3

[RECA\\_ROSCS](#) Mass: 37935 Score: 49 Expect: 6.7 Matches: 5  
 Protein RecA OS=Roseiflexus castenholzii (strain DSM 13941 / HLO8) OX=383372 GN=recA PE=3 SV=1

[IF5A\\_PYRNV](#) Mass: 14588 Score: 47 Expect: 11 Matches: 4  
 Translation initiation factor 5A OS=Pyrobaculum neutrophilum (strain DSM 2338 / JCM 9278 / V24Sta) OX=444157 GN=if5A PE=3 SV=1

[NADD\\_CHLL2](#) Mass: 22438 Score: 44 Expect: 20 Matches: 4  
 Probable nicotinate-nucleotide adenyltransferase OS=Chlorobium limicola (strain DSM 245 / NBRC 103803 / 6330) OX=290315 GN=nadD PE=3 SV=1

[RNS10\\_HORSE](#) Mass: 23926 Score: 42 Expect: 39 Matches: 4  
 Inactive ribonuclease-like protein 10 OS=Equus caballus OX=9796 GN=RNASE10 PE=2 SV=2

[MURC\\_IDILO](#) Mass: 52994 Score: 42 Expect: 39 Matches: 5  
 UDP-N-acetylmuramate-L-alanine ligase OS=Idiomarina loihiensis (strain ATCC BAA-735 / DSM 15497 / L2-TR) OX=283942 GN=murC PE=3 SV=1

[TRGV8\\_HUMAN](#) Mass: 13327 Score: 41 Expect: 41 Matches: 3  
 T cell receptor gamma variable 8 OS=Homo sapiens OX=9606 GN=TRGV8 PE=1 SV=1

[CLPP1\\_BIFLO](#) Mass: 25812 Score: 40 Expect: 54 Matches: 4  
 ATP-dependent Clp protease proteolytic subunit 1 OS=Bifidobacterium longum (strain NCC 2705) OX=206672 GN=clpP1 PE=3 SV=1

[VGLG\\_AMPV1](#) Mass: 64475 Score: 40 Expect: 59 Matches: 6  
 Major surface glycoprotein G OS=Avian metapneumovirus (isolate Canada goose/Minnesota/15a/2001) OX=652954 GN=G PE=3 SV=1

[ISPDF\\_CAMJ8](#) Mass: 41632 Score: 40 Expect: 60 Matches: 5

6-2-6 同定タンパク質の情報

==== 中略 =====

- [LUTC\\_GEOSW](#) Mass: 26930 Score: 34 Expect: 2.4e+002 Matches: 4  
 Lactate utilization protein C OS=Geobacillus sp. (strain WCH70) OX=471223 GN=lutC PE=3 SV=1
- [RSMH\\_PARDP](#) Mass: 34298 Score: 34 Expect: 2.4e+002 Matches: 4  
 Ribosomal RNA small subunit methyltransferase H OS=Paracoccus denitrificans (strain Pd 1222) OX=318586 GN=rsmH PE=3 SV=1
- [PHEA\\_MYCBO](#) Mass: 33613 Score: 34 Expect: 2.5e+002 Matches: 4  
 Prephenate dehydratase OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) OX=233413 GN=pheA PE=1 SV=1
- [PHEA\\_MYCBP](#) Mass: 33613 Score: 34 Expect: 2.5e+002 Matches: 4  
 Prephenate dehydratase OS=Mycobacterium bovis (strain BCG / Pasteur 1173P2) OX=410289 GN=pheA PE=3 SV=1
- [PHEA\\_MYCTA](#) Mass: 33613 Score: 34 Expect: 2.5e+002 Matches: 4  
 Prephenate dehydratase OS=Mycobacterium tuberculosis (strain ATCC 25177 / H37Ra) OX=419947 GN=pheA PE=3 SV=1

Search Parameters

Type of search : Peptide Mass Fingerprint  
 Enzyme : Trypsin/P  
 Mass values : Monoisotopic  
 Protein Mass : Unrestricted  
 Peptide Mass Tolerance : ± 0.2 Da  
 Peptide Charge State : 1+  
 Max Missed Cleavages : 2  
 Number of queries : 18

6-2-7 Search parameters

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

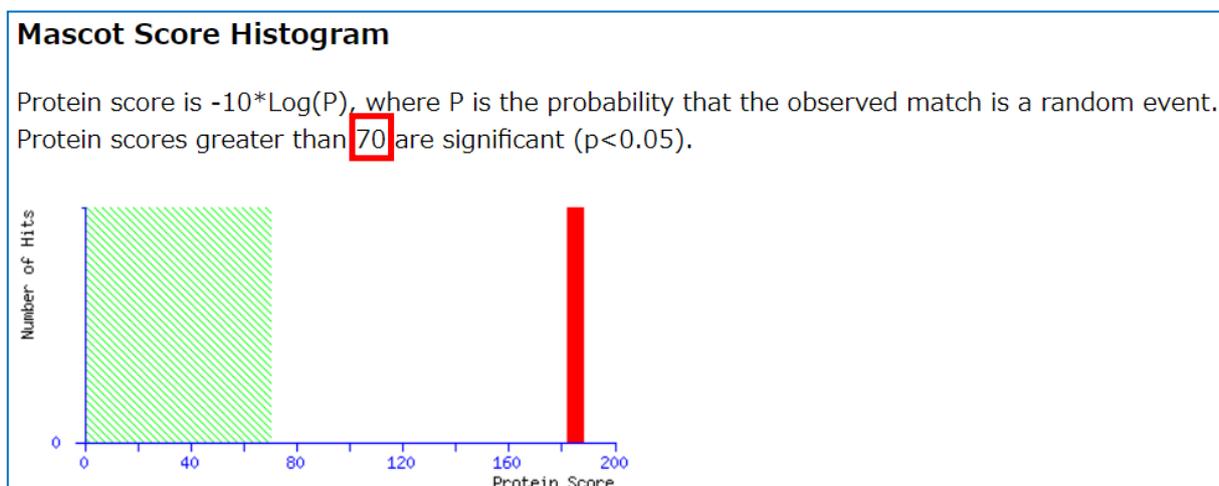
## 6-2-2. ヘッダー部分

```
User          :
Email         :
Search title  : Peptide Mass Fingerprint Example
Database      : SwissProt 2019_10 (561356 sequences; 201858328 residues)
Timestamp     : 9 Jan 2020 at 11:23:29 GMT
Top Score     : 185 for PML_HUMAN, Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3
```

「5-3 PMF 検索パラメーター 一覧」も併せてご覧ください。

- User** : パラメーター「**your name**」で指定した内容
- Email** : パラメーター「**Email**」で指定した内容
- Search title** : パラメーター「**Search title**」で指定した内容
- Database** : 検索対象としたデータベースとバージョン、登録件数と総残基数
- Taxonomy** : パラメーター「**Taxonomy**」で指定した生物種(指定しているときのみ)。  
生物種限定時の登録エントリー数也表示
- Timestamp** : 検索開始時間
- Top Score** : 最も高いスコアとなったタンパク質のスコア、Accession, Description

## 6-2-3. Score Histogram

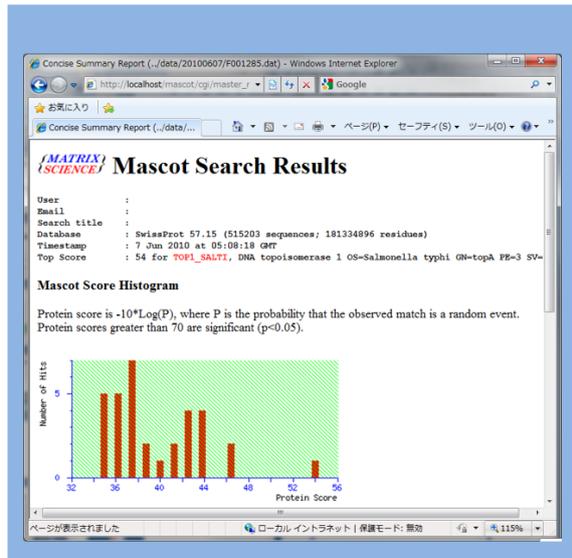


横軸が MASCOT Score,縦軸はそのスコアを持つエントリーの数です。

上記図内の赤枠で囲われた箇所(70)が同定基準値で、**同定基準値より高いスコアを持つタンパク質が MASCOT で判定された「同定タンパク質」となります。**

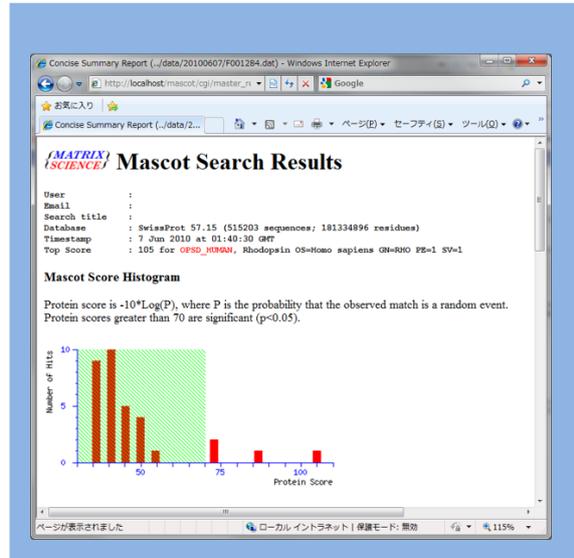
同定基準値は、グラフの緑の網掛けゾーン一番右側に対応します。**グラフの緑の網掛けより右側にタンパク質が存在するとき同定タンパク質が見つかったことを意味し、逆に最もスコアが高いタンパク質も緑の網掛けの領域を脱する事が出来なかった場合は同定タンパク質が見つからなかったことを意味します**(次頁図)。同定タンパク質が見つかった場合は検索成功で、同定内容について検証する作業に入ります。一方同定タンパク質が見つからなかった時は入力データや指定パラメーターを見直して同定できなかった理由を検証するか、新たな測定を実施して再度同定を試みる事をお勧めします。

• 失敗



→ 再検索

• 成功



→ 結果の検証

6-2-4. 表示内容の変更 [Format As]

Format As	Concise Protein Summary ▼	<a href="#">Help</a>
	Significance threshold $p <$ <input type="text" value="0.05"/>	Max. number of hits <input type="text" value="AUTO"/>
	Preferred taxonomy <input type="text" value="All entries"/>	▼

**表示 Summary 形式** : 結果画面のフォーマットを変更したり、結果ファイルの export を行います。

[Protein Summary](#) / [Concise Protein Summary](#) / [Export Search Results](#)

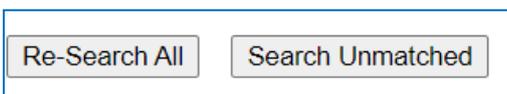
**Significance threshold** : 同定基準の **p value**(このケースでは **Expect** と同じとお考え下さい)を指定します。小さいほど同定基準が厳しい事になります。

**Max. number of hits** : 表示するタンパク質数、推奨は”AUTO”。

**Preferred taxonomy** : **優先表示させる生物種**。タンパク質の 1 つのエントリーに数種のエントリーが統合されていることがあります。検索パラメーター「taxonomy」の絞り込みはエントリーに登録されている複数生物種すべてに対応する事ができる一方、表示される生物種はエントリー時にデフォルトで定められた 1 種類のみです。結果、パラメーターで指定した生物種と明らかに異なる生物種由来と見受けられるエントリーが結果に表示されることがあります。Preferred taxonomy の設定を行う事で、優先して表示させる生物種名を指定したものに切り替える事ができます。

各種項目を変更後、”**Format As**”ボタンを押すことで設定内容に基づいて**結果画面表示が切り替わり**ます。

## 6-2-5. 再検索 : Research all, search unmatched



**再検索**を実施するボタンです。「**Re-Search All**」は入力データ全てを使った、「**Search Unmatched**」は同定基準を超えたタンパク質にマッチしたものを除いた全てのピークを使って再検索を行います。

## 6-2-6. 同定タンパク質の情報

1.	<a href="#">PML_HUMAN</a>	Mass: 97489	Score: <b>185</b>	Expect: 1.8e-013	Matches: 16
	Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3				
	<a href="#">RECA_ROSCS</a>	Mass: 37935	Score: 49	Expect: 6.7	Matches: 5
	Protein RecA OS=Roseiflexus castenholzii (strain DSM 13941 / HLO8) OX=383372 GN=recA PE=3 SV=1				
	<a href="#">IF5A_PYRNV</a>	Mass: 14588	Score: 47	Expect: 11	Matches: 4
	Translation initiation factor 5A OS=Pyrobaculum neutrophilum (strain DSM 2338 / JCM 9278 / V24Sta) OX=				
	<a href="#">NADD_CHLL2</a>	Mass: 22438	Score: 44	Expect: 20	Matches: 4
	Probable nicotinate-nucleotide adenyllyltransferase OS=Chlorobium limicola (strain DSM 245 / NBRC 10380				
	<a href="#">RNS10_HORSE</a>	Mass: 23926	Score: 42	Expect: 39	Matches: 4
	Inactive ribonuclease-like protein 10 OS=Equus caballus OX=9796 GN=RNASE10 PE=2 SV=2				
	<a href="#">MURC_IDILO</a>	Mass: 52994	Score: 42	Expect: 39	Matches: 5
	UDP-N-acetylmuramate--L-alanine ligase OS=Idiomarina loihiensis (strain ATCC BAA-735 / DSM 15497 / L2-				
	<a href="#">TRGV8_HUMAN</a>	Mass: 13327	Score: 41	Expect: 41	Matches: 3

検索した結果、スコアが高かったタンパク質について表示されます。表示内容は以下の通りです。

- Accession** : データベースの ID が表示されます。ハイパーリンクになっていてクリックするとより詳しい情報が記載されている「**Protein View**」の画面が開きます。Protein View の詳細は **6-3** をご覧ください。
- Mass** : タンパク質の質量で、データベースに登録されている配列情報から計算。
- Score** : MASCOT Score。赤字の表示は同定基準を超えている事を表します。高いほど理論値と実測値がよりよくマッチしていることを示します。
- Expect** : Score と同定基準値をもとに算出された値。ランダムマッチだった場合に検索したデータベースからどれくらいのエントリーが見つかるかを表す「期待値」です。同定基準を超えている時、値が 0.05(デフォルト設定の場合)より小さくなります。なお同定基準値は「Significance threshold p<」の値と連動します。期待値についての詳細は「**8-4.同定タンパク質:マッチングとスコア、同定基準値、期待値**」をご覧ください。
- Matches** : マッチしたピーク数。
- Description** : データベース各エントリーのヘッダー行に記載されている、タンパク質の機能に関する情報。

## 6-2-7. Search parameters

### Search Parameters

Type of search	: Peptide Mass Fingerprint
Enzyme	: Trypsin/P
Mass values	: Monoisotopic
Protein Mass	: Unrestricted
Peptide Mass Tolerance	: $\pm 0.2$ Da
Peptide Charge State	: 1+
Max Missed Cleavages	: 2
Number of queries	: 18

検索時に指定したパラメーター。詳細は「**5-1.PMF:検索パラメーター 一覧**」をご参照ください。

- Type of search** : MASCOT 3 つの検索手法のうち、どれか。
- Enzyme** : 切断パターン。
- Fixed modification** : 必ず質量置換する修飾設定 (設定時のみ。上記例図には含まれていない)。
- Variable modification** : 質量置換する/しないケースを想定する修飾設定 (設定時のみ。上記例図には含まれない)。
- Mass values** : 各種質量計算に使うアミノ酸の質量が「Monoisotopic」か「Average」か。
- Protein Mass** : タンパク質質量の上限値
- Peptide Mass Tolerance**:ペプチドの質量マッチングにおける誤差範囲
- Peptide Charge State** : ピークの電荷が MH+か Mr か M-H- か
- Max Missed Cleavages** : 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についての設定
- Number of queries** : 入力データのピーク数

## 6-3. Protein View

Summary 画面の中でタンパク質名の箇所がハイパーリンクになっています。このハイパーリンクをクリックすると、**マッチしたタンパク質についてより詳しい情報が記されている「Protein View」の画面**となります(次頁図)。

**MATRIX SCIENCE MASCOT Search Results**

**Protein View: PML\_HUMAN**

Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3

Database: SwissProt  
 Score: 185  
 Expect: 1.8e-013  
 Monoisotopic mass (M<sub>0</sub>): 97489  
 Calculated pI: 5.88  
 Taxonomy: [Homo sapiens](#)

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

**Search parameters**

Enzyme: Trypsin/P: cuts C-term side of KR.  
 Mass values searched: 18  
 Mass values matched: 16

**Protein sequence coverage: 23%**

Matched peptides shown in **bold red**.

```

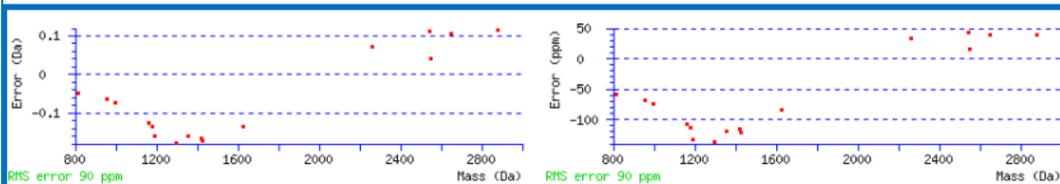
1 MEPAFARSPR PQDDPARPQE PTMPPPETPS EGRQPSPPS PTERAPASEE
51 EFQFLRCQQC QAEAKCPKLL PCLHTLCSGC LEASGMQCPPI CQAWPLGAD
101 TPALDNVFFE SLQRLSVYR QIVDAQAVCI RCKESADFWC FECEQLLCAK
151 CFEAHQWFLK HEARPLAELR NQSVREFLDG TRKTNNIFCS NPNHRTPTLT
201 SIYCRGCSKP LCCSCALLDS SHSELKCDIS AEIQQRQEEL DAMTQALQEQ
251 DSAFGAVHQA MHAAVQQLGR ARAETEELIR ERVQVVAHV RAQERELLEA
301 VDARYQRDYE EMASRLGRLD AVLQRIRTGS ALVQRMKCYA SDQEVLDMHG
351 FLRQALCRLR QEEPQSLQAA VRTDGFDEFK VRLQDLSSCI TQKDAVAVSK
401 KASPEAASTP RDPIDVDLPE EAERVKAQVQ ALGLAEAQPM AVVQSVGPAH
451 PVFVYAFSIK GPSYGEDVSN TTTAQKRKCS QTQCPRKVIK MESEEGKEAR
501 LARSSPEQPR PSTSKAVSPP HLDGPPSPRS PVIGSEVFLP NSNHVASGAG
551 EAEERVVVIS SSEDSDAENS SSRELDSSS ESSDLQLEGP STLRVLDENL
601 ADPQAEDRPL VFFDLKIDNE TQKISQLAAV NRESKFRVVI QPEAFFSIYS
651 KAVSLEVGLQ HFLSFLSSMR RPILACYKLW GPGLPNFFRA LEDINRLWEF
701 QEAIISGLAA LPLIRERVFG ASSFKLKNLA QTVLARNMSE RSAMAAVLAM
751 RDLCRILLEVS PGPQLAQHVY PFSSLQCFAS LQPLVQAVAL PRAEARLLAL
801 HNVSMFELLS AHRDRDQGGI KKYSRYSLSQ TITLPPAQPA FNLQALGTYF
851 EGLLEGPALA RAEGVSTPLA GRGLAERASQ QS
    
```

Unformatted sequence string: [882 residues](#) (for pasting into other applications).

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

Start - End	Observed	Mr (expt)	Mr (calc)	Delta M	Peptide
8 - 33	2882.5000	2881.4927	2881.3777	0.1150 2	R.SPRPQDDPARPQEP <b>PTMPPPETPSEGR</b> .Q
34 - 44	1182.4400	1181.4327	1181.5677	-0.1349 0	R.QPSPPSPPTER.A
45 - 56	1423.5200	1422.5127	1422.6779	-0.1652 0	R.APASEE <b>EFQFLR</b> .C
161 - 170	1191.5000	1190.4927	1190.6520	-0.1592 1	K. <b>HEARPLAELR</b> .N
308 - 315	1000.3300	999.3227	999.3967	-0.0740 0	R.DY <b>EMASR</b> .L
319 - 325	814.4300	813.4227	813.4708	-0.0481 0	R.LD <b>AVLQR</b> .I
359 - 372	1624.7400	1623.7327	1623.8692	-0.1365 1	R.L <b>RQEE</b> PQSLQAAVR.T
361 - 372	1355.5300	1354.5227	1354.6841	-0.1613 0	R. <b>QEE</b> PQSLQAAVR.T
361 - 382	2550.3000	2549.2927	2549.2510	0.0417 2	R. <b>QEE</b> PQSLQAAVRTD <b>GFDEFK</b> VR.L
373 - 380	958.3500	957.3427	957.4080	-0.0653 0	R.TD <b>GFDEFK</b> .V
491 - 500	1165.3900	1164.3827	1164.5081	-0.1253 1	K. <b>MESEEGKEAR</b> .L
504 - 515	1300.4700	1299.4627	1299.6419	-0.1792 1	R. <b>SSPEQ</b> PRPSTSK.A
516 - 529	1426.5700	1425.5627	1425.7365	-0.1737 0	K.AV <b>SPPHLDGPPSPR</b> .S
530 - 555	2653.3900	2652.3827	2652.2780	0.1048 0	R.SF <b>VIGSEVFLPNSNHVASGAGEAER</b> .V
574 - 594	2265.1100	2264.1027	2264.0292	0.0735 0	R.EL <b>DDSSSESSDLQLEGP</b> STLR.V
595 - 616	2544.4100	2543.4027	2543.2908	0.1120 1	R.VLDENLAD <b>PQAE</b> DRPLV <b>FFDLK</b> .I

No match to: 1320.4000, 1348.4100



ID PML\_HUMAN Reviewed: 882 AA.  
 AC P29590; E9PBR7; P29591; P29592; P29593; Q00755; Q15959; Q59FP9; Q8WUA0;  
 FC Q96841; Q9BFW2; Q9BWP7; Q9BZX6; Q9BZX7; Q9BZX8; Q9BZX9; Q9BZY0; Q9BZY2;  
 Q9BZY3

次頁以降、青色で囲われた各表示領域についてより詳しく説明します。

## Protein View: PML\_HUMAN

Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3

Database: SwissProt  
Score: 185  
Expect: 1.8e-013  
Monoisotopic mass ( $M_r$ ): 97489  
Calculated pI: 5.88  
Taxonomy: [Homo sapiens](#)

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

ページの最初にタンパク質の **Accession** と **Description** が最初に表示されます。続いて以下の情報が表示されます。

**Database** : 使用したデータベース名。  
**Score** : MASCOT Score。  
**Expect** : Score と同定基準値をもとに算出された値。同定基準を超えている時、Expect = 0.05 となる。Expect が 0.05 より小さい時同定。  
**Monoisotopic mass ( $M_r$ )**: データベースの配列から計算されたタンパク質の質量。  
**Calculated pI** : データベースの配列から計算された予測等電点。  
**Taxonomy** : 生物種。

また該当タンパク質の配列を、NCBI の BLAST(配列相同性検索プログラム)実行するためのリンクが表示されます。

## Search parameters

Enzyme: Trypsin/P: cuts C-term side of KR.  
Mass values searched: 18  
Mass values matched: 16

**Enzyme** : 切断パターン。  
**Mass values searched** : クエリーのピーク数。クエリーに強度情報が含まれている場合、クエリーセットを様々なパターンに組み合わせてマッチングスコアを検証しますが、クエリーが少ない組み合わせが採用された場合はその時のピーク数が表示されます。  
**Mass values matched** : 理論値とマッチしたピーク数。

### Protein sequence coverage: 23%

Matched peptides shown in **bold red**.

```
1  MEPAPARSPR PQQDPARPOE PTMPPPETPS EGRQPSPPSPS PTERAPASEE
51  EFQFLRCQQC QAEAKCPKLL PCLHTLCSGC LEASGMQCPI CQAPWPLGAD
101 TPALDNVFFE SLQRRLSVYR QIVDAQAVCT RCKESADFWC FECEQLLCAK
151 CFEAHQWFLK HEARPLAELR NQSVREFLDG TRKTNNIFCS NPNHRTPTLT
201 SIYCRGCSKP LCCSCALLDS SHSELKCDIS AEIQQRQEEL DAMTQALQEQ
251 DSAFGAVHAQ MHAAVGQLGR ARAETEELIR ERVRQVVAHV RAQERELLEA
301 VDARYQRDYE EMASRLGRLD AVLQRITGS ALVQRMKCYA SDQEVLDMHG
351 FLRQALCRLR QEEPQSLQAA VRTDGFDEFK VRLQDLSSCI TQGKDAAVSK
401 KASPEAASTP RDPIDVDLPE EAERVKAQVQ ALGLAEAQPM AVVQSVPGAH
451 PVPVYAFSIK GPSYGEDVSN TTTAQKRKCS QTQCPRKVIK MESEEGKEAR
501 LARSSPEQPR PSTSKAVSPP HLDGPPSPRS PVIGSEVFLP NSNHVASGAG
551 EAEERVVVIS SSEDSDAENS SSRELDSSS ESSDLQLEGP STLRVLDENL
601 ADPQAEDRPL VFFDLKIDNE TQKISQLAAV NRESKFRVVI QPEAFFSIYS
651 KAVSLEVGLQ HFLSFLSSMR RPILACYKLW GPGLPNFFRA LEDINRLWEF
701 QEASIGFLAA LPLIRERVPG ASSFKLKNLA QTYLARNMSE RSAMA AVLAM
751 RDLCLLLEVS PGPQLAQHVY PFSSLQCFAS LQPLVQAAVL PRAEARLLAL
801 HNVSEFMEELS AHRDRQGGI KKYSRYLSLQ TTTLPPAQA FNLQALGTYF
851 EGLLEGPALA RAEGVSTPLA GRGLAERASQ QS
```

続いて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、またその割合についての情報が表示されます。**Coverage** とは、全長に対するマッチペプチド残基数の割合です。マッチしたペプチド部分が赤の太字で表現されています。

Unformatted sequence string: [882 residues](#) (for pasting into other applications).

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

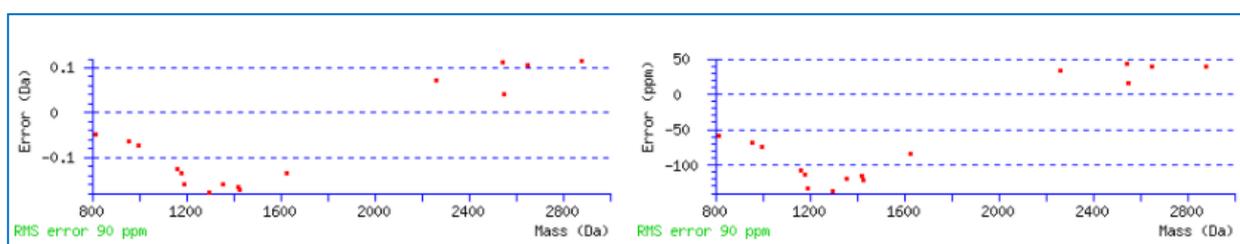
Start - End	Observed	Mr (expt)	Mr (calc)	Delta M	Peptide
8 - 33	2882.5000	2881.4927	2881.3777	0.1150 2	R.SPRPQQDPARPOEPTMPPPETPSEGR.Q
34 - 44	1182.4400	1181.4327	1181.5677	-0.1349 0	R.QPSPSPSPTER.A
45 - 56	1423.5200	1422.5127	1422.6779	-0.1652 0	R.APASEEEFQFLR.C
161 - 170	1191.5000	1190.4927	1190.6520	-0.1592 1	K.HEARPLAELR.N
308 - 315	1000.3300	999.3227	999.3967	-0.0740 0	R.DYEEEMASR.L
319 - 325	814.4300	813.4227	813.4708	-0.0481 0	R.LDAVLQR.I
359 - 372	1624.7400	1623.7327	1623.8692	-0.1365 1	R.LRQEEPQSLQAAVR.T
361 - 372	1355.5300	1354.5227	1354.6841	-0.1613 0	R.QEEPQSLQAAVR.T
361 - 382	2550.3000	2549.2927	2549.2510	0.0417 2	R.QEEPQSLQAAVVRTDGFDEFKVR.L
373 - 380	958.3500	957.3427	957.4080	-0.0653 0	R.TDGFDEFK.V
491 - 500	1165.3900	1164.3827	1164.5081	-0.1253 1	K.MESEEGKEAR.L
504 - 515	1300.4700	1299.4627	1299.6419	-0.1792 1	R.SSPEQPRPSTSK.A
516 - 529	1426.5700	1425.5627	1425.7365	-0.1737 0	K.AVSPPHLDGPPSPR.S
530 - 555	2653.3900	2652.3827	2652.2780	0.1048 0	R.SPVIGSEVFLPNSNHVASGAGEAER.V
574 - 594	2265.1100	2264.1027	2264.0292	0.0735 0	R.ELDDSSSESSDLQLEGPSTLR.V
595 - 616	2544.4100	2543.4027	2543.2908	0.1120 1	R.VLDENLADPQAEDRPLVFFDLK.I

No match to: 1320.4000, 1348.4100

さらにその下には、マッチしたペプチドがアミノ酸残基順(デフォルト設定の場合)に並んでリスト表示されています。

**Unformatted sequence string** : 残基数と共に配列をコピーしやすくなるページが開きます。他の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** : リストの並び順を指定。残基番号、質量の昇順/降順 が選択可。
- 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を選択します。
- Start-End** : タンパク質全長におけるアミノ酸残基番号。
- 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。
- Mr(calc)** : 配列から計算されたペプチドの質量。
- Delta** : Mr(expt) - Mr(calc)。
- M** : Missed cleavage。
- Peptide** : ペプチド配列。修飾も含まれる場合は併せて表示されます。



ページ下部で表示されているグラフはともに**ピークのマッチングの誤差を表すグラフ**で、左が Da、右が ppm で表現されています。ともに横軸は実験データ側のペプチド質量で、縦軸が誤差です。含まれている事が確実なタンパク質でこのグラフを確認する事で、パラメーターで指定した誤差範囲(peptide tol.)の設定値が適切であったかどうかを確認する事もできます。

```

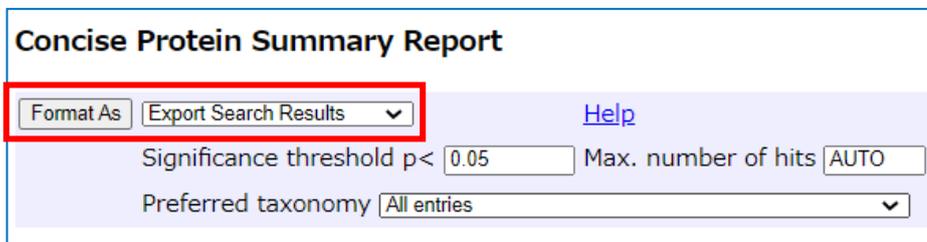
ID   PML_HUMAN                Reviewed;              882 AA.
AC   P29590; E9PBR7; P29591; P29592; P29593; Q00755; Q15959; Q59FP9; Q8WUA0;
AC   Q96S41; Q9BPW2; Q9BWP7; Q9BZX6; Q9BZX7; Q9BZX8; Q9BZX9; Q9BZY0; Q9BZY2;
AC   Q9BZY3;
DT   01-APR-1993, integrated into UniProtKB/Swiss-Prot.
DT   25-NOV-2008, sequence version 3.
DT   02-JUN-2021, entry version 248.
DE   RecName: Full=Protein PML;
DE   AltName: Full=E3 SUMO-protein ligase PML;
DE   EC=2.3.2.- {ECO:0000269|PubMed:20972456, ECO:0000269|PubMed:28250117};
DE   AltName: Full=Promyelocytic leukemia protein;
DE   AltName: Full=RING finger protein 71;
DE   AltName: Full=RING-type E3 SUMO transferase PML {ECO:0000305};
DE   AltName: Full=Tripartite motif-containing protein 19;
DE   Short=TRIM19;
GN   Name=PML; Synonyms=MYL, PP8675, RNF71, TRIM19;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM PML-3). AND DISEASE.

```

画面の最下部には**タンパク質の詳細情報**が表示されます。ただしデータベース側で情報表示に関する適切な設定がある時のみ表示されます。

## 6-4. 結果の Export

同定結果をファイル出力することができます。Format As の選択肢で「**Export Search Results**」を選択してから「**Format As**」ボタンを押します。



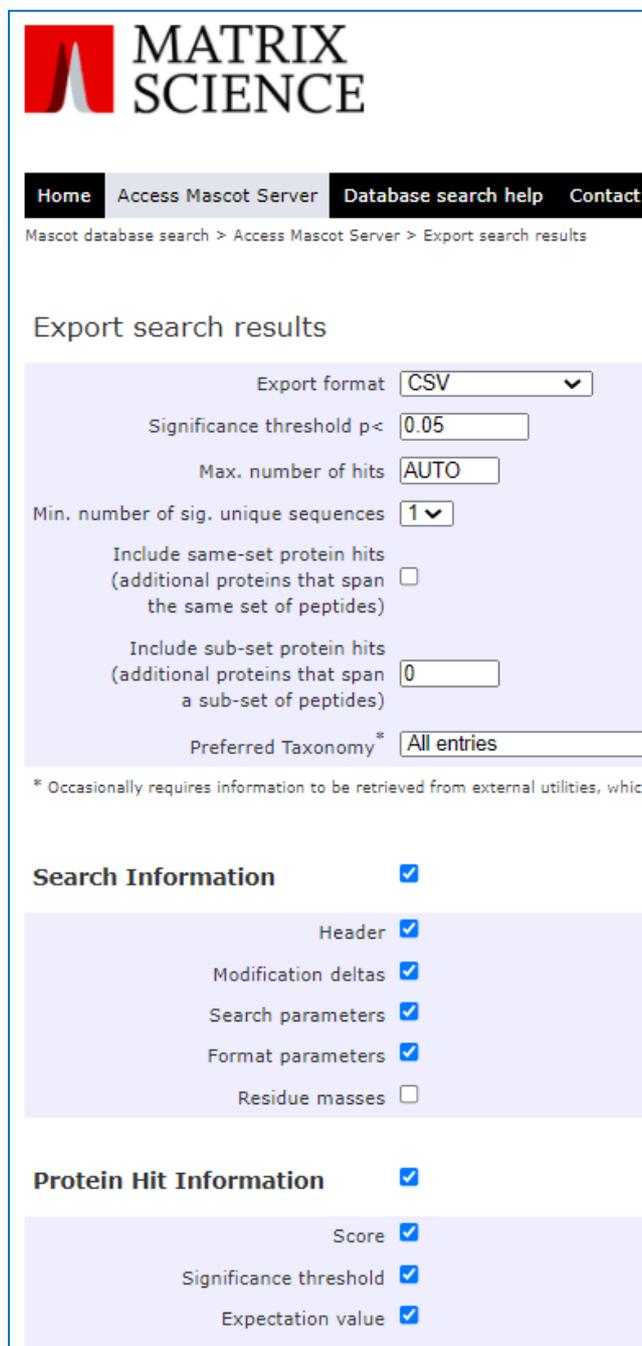
Concise Protein Summary Report

Format As **Export Search Results** [Help](#)

Significance threshold p<  Max. number of hits

Preferred taxonomy

出力のファイルフォーマットや条件、出力項目などを選択し、画面下部の「**Export Search Results**」ボタンを押すとファイル出力が実行されます。



MATRIX SCIENCE

Home Access Mascot Server Database search help Contact

Mascot database search > Access Mascot Server > Export search results

Export search results

Export format

Significance threshold p<

Max. number of hits

Min. number of sig. unique sequences

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)

Preferred Taxonomy\*

\* Occasionally requires information to be retrieved from external utilities, which

**Search Information**

Header

Modification deltas

Search parameters

Format parameters

Residue masses

**Protein Hit Information**

Score

Significance threshold

Expectation value

## 7. 検索結果画面:MIS

この章では MASCOT MIS 検索における結果画面において説明します。

### 7-1. 表示例で使用している検索について

MIS の検索例として MASCOT Server 内に存在する以下の検索結果を利用します。

[公開サーバー] [https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981139.dat](https://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](http://localhost/mascot/master_results_2.pl?file=../data/F981139.dat)

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

query 数 : 33,191

検索の結果、以下のスペクトルとタンパク質が同定されています。

同定スペクトル数 : 2,812

同定タンパク質数 : 480 (448 グループ)

### 7-2. 表示内容の詳細 : summary 画面

#### 7-2-1. 展開しない状態での画面概要

URL を WEB ブラウザで指定し結果画面を開くと、次頁のような画面が現れます。画面内の赤線で囲われた各パーツで記載されている内容について、以降順に説明します。

(次頁図は表示領域の関係で2つの画像を1つに組み合わせています。)

# MASCOT Search Results

User :  
E-mail :  
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; 201,858,328 residues)  
Taxonomy : 1: (none)  
2: Mus. (17,083 sequences)  
Timestamp : 9 Jan 2020 at 14:12:31 GMT

7-2-2 ヘッダー部分

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

7-2-3 再検索と結果ファイル出力

▼ Search parameters  
Type of search : MS/MS Ion Search  
Enzyme : Trypsin/P  
Fixed modifications : [\[help\]](#)ITRAQ4plex (K), [\[help\]](#)ITRAQ4plex (N-term), [\[help\]](#)Methylthio (C)  
Variable modifications : [\[help\]](#)Acetyl (Protein N-term), [\[help\]](#)Gln->pyro-Glu (N-term Q), [\[help\]](#)Oxidation (M)  
Mass values : Monoisotopic  
Protein mass : Unrestricted  
Peptide mass tolerance : ± 0.9 Da  
Fragment mass tolerance : ± 0.6 Da  
Max missed cleavages : 1  
Instrument type : ESI-TRAP  
Number of queries : 33,191

7-2-4 search parameters

▶ Score distribution  
▶ Modification statistics for all protein families  
▶ Legend

7-2-5 スコア分布, 7-2-6 modification 一覧,

7-2-7 凡例

## Protein Family Summary

7-2-8 表示内容の切り替え[スコア足切りなど]

Format Significance threshold p < 0.05 Max. number of families AUTO [\[help\]](#)  
Target FDR (overrides sig. threshold) (not set) 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) 7-2-9 Sensitivity and FDR

Proteins (480) [Report Builder](#) [Unassigned \(30379\)](#) [\\$ permalink](#)

## 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

Accession	Protein Name	Score	Description
1::sp TRY1_BOVIN	1600 sp TRY1_BOVIN		
▶ 2	1 2::CP2CT_MOUSE	1332	Cytochrome P450 2C29 OS=Mus musculus OX=...
	6 2::CP239_MOUSE	251	Cytochrome P450 2C39 OS=Mus musculus OX=...
	7 2::CP238_MOUSE	150	Cytochrome P450 2C38 OS=Mus musculus OX=...
	2 2::CP254_MOUSE	550	Cytochrome P450 2C54 OS=Mus musculus OX=...
	3 2::CY250_MOUSE	487	Cytochrome P450 2C50 OS=Mus musculus OX=...
	5 2::CP237_MOUSE	338	Cytochrome P450 2C37 OS=Mus musculus OX=...
	4 2::CP2F2_MOUSE	470	Cytochrome P450 2F2 OS=Mus musculus OX=1...
▶ 3	1 2::BIP_MOUSE	1302	Endoplasmic reticulum chaperone BIP OS=Mus...
	2 2::HS1_MOUSE	110	1 kDa protein OS=Mus m...
	3 2::HS2_MOUSE	110	protein 1-like OS=Mus musc...
▶ 4	2::CYB5_MOUSE	1202	Cytochrome b5 OS=Mus musculus OX=10090 G...
▶ 5	2::PDIA1_MOUSE	1121	Protein disulfide-isomerase OS=Mus musculus ...
▶ 6	2::CP1A2_MOUSE	1053	Cytochrome P450 1A2 OS=Mus musculus OX=1...
▶ 7	2::ENPL_MOUSE	1008	Endoplasmic reticulum protein OS=Mus musculus OX=10090 GN...
▶ 8	1 2::RDH7_MOUSE	995	Retinol dehydrogenase 7 OS=Mus musculus OX=...
	2 2::H17B6_MOUSE	594	17-beta-hydroxysteroid dehydrogenase type 6 ...
	3 2::RDH16_MOUSE	497	Retinol dehydrogenase 16 OS=Mus musculus O...
▶ 9	2::MGST1_MOUSE	860	Microsomal glutathione S-transferase 1 OS=Mu...
▶ 10	2::RL7A_MOUSE	768	60S ribosomal protein L7a OS=Mus musculus O...

7-2-10 同定タンパク質とアサインペプチド

ペプチド情報は展開時に表示

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

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

## 7-2-2. ヘッダー部分

```
User      :
E-mail    :
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; 201,858,328 residues)
Taxonomy   : 1: (none)
             2: Mus. (17,083 sequences)
Timestamp  : 9 Jan 2020 at 14:12:31 GMT
```

「5-3 MIS 検索パラメーター 一覧」も併せてご覧ください。

- User** : パラメーター「**your name**」で指定した内容。
- Email** : パラメーター「**Email**」で指定した内容。
- Search title** : パラメーター「**Search title**」で指定した内容。
- MS data file** : 検索で使った実験データのファイル名。
- Database(s)** : 検索対象としたデータベースとバージョン、登録件数と総残基数。
- Taxonomy** : パラメーター「**Taxonomy**」で指定した生物種(使用した時のみ)。その生物種として登録されているエントリー数も表示されます。
- Timestamp** : 検索開始時間

## 7-2-3. 再検索と結果ファイル出力: Re-search ボタン、Export ボタン



「**Re-search**」は再検索を実施するためのボタンです。再検索を行う query の種類によって、**All**(すべての query)、**Non-significant**(同定基準に満たなかった query)、**Unassigned**(同定タンパク質に帰属しない query)の3種類を選択することができます。

「**Export**」は検索結果をペプチド単位でファイル出力するためのオプションです。様々なフォーマット、同定ペプチド/タンパク質の条件、出力項目を設定してファイル出力することができます。詳細は「7-5.Export 機能によるペプチドベースの検索結果ファイル出力」でまとめましたのでそちらをご覧ください。

## 7-2-4. Search parameters

```
▼Search parameters
Type of search      : MS/MS Ion Search
Enzyme              : Trypsin/P
Fixed modifications : iTRAQ4plex \(K\), iTRAQ4plex \(N-term\), Methy
Variable modifications : Acetyl \(Protein N-term\), Gln->pyro-Glu \(N-term\)
Mass values        : Monoisotopic
Protein mass       : Unrestricted
Peptide mass tolerance : ± 0.9 Da
Fragment mass tolerance : ± 0.6 Da
Max missed cleavages : 1
Instrument type    : ESI-TRAP
Number of queries  : 33,191
```

パラメーター内容の詳細は、「**5-3:MIS 検索パラメーター 一覧**」をご覧ください。

- Type of search** : MASCOT 3 つの検索手法のうち、どれか。
- Enzyme** : 切断パターン。
- Fixed modification** : 修飾、すべての対象アミノ酸の質量を指定内容に入れ替える。
- Variable modification** : 修飾、対象アミノ酸の質量を指定内容と入れ替えるケースと入れ替えないケースの両方を計算する。
- Mass values** : 各種質量計算に使うアミノ酸の質量が、「Monoisotopic」か「Average」か。
- Protein Mass** : タンパク質質量の上限値。
- Peptide Mass Tolerance**: ペプチドの質量マッチングにおける誤差範囲。
- Fragment Mass Tolerance** : フラグメントの質量マッチングにおける誤差範囲。
- Peptide Charge State** : ペプチドの価数 (precursor イオンレベル)。
- #<sup>13</sup>C** : ペプチドに含まれる <sup>13</sup>C を考慮した回数。
- Max Missed Cleavages** : 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についての設定。
- Instrument type** : 考慮するイオンシリーズやフラグメントピークの電荷に関する情報が定義されたセット。
- Number of queries** : 入力データの測定データ数。

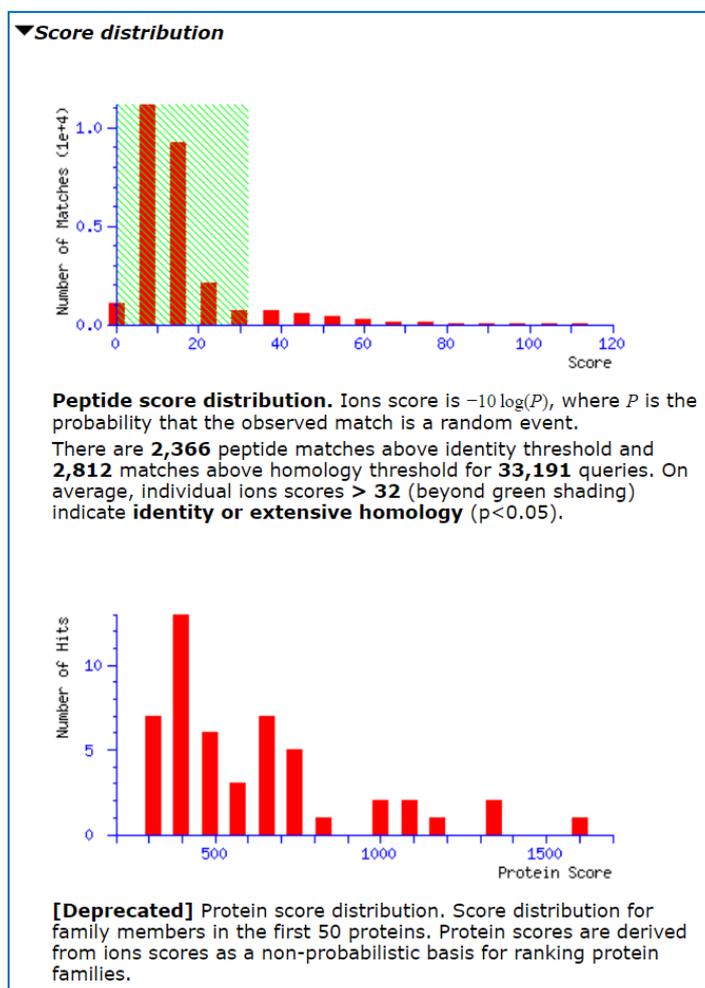
## 7-2-5. Score Distribution

### ▶ Score distribution

「**Score distribution**」の箇所をクリックすると、2つのスコアのヒストグラムが表示されます。上は query、下は同定タンパク質です。

上の query のグラフは、横軸がマッチング度合いを表す Mascot Ions スコア、縦軸が個数です。各 query により同定基準値が異なりますが、その同定基準の平均値が緑の網掛けラインの一番右側となります。

下は同定タンパク質のスコア(ペプチドのスコアをタンパク質毎に集計したスコア)分布です。タンパク質には同定基準スコアはなく、グラフには緑の網掛けは現れません。



## 7-2-6. Modification statistics for all protein families

### ▶ Modification statistics for all protein families

「**Modification statistics for all protein families**」の箇所をクリックすると、同定 query のうち修飾を含んでいた内容について、種類別の個数がカウントされリスト表示されます。

Modification	Delta	Type	Site	Total matches
iTRAQ4plex	144.102063	fixed	N-term	2746
iTRAQ4plex	144.102063	fixed	K	1732
iTRAQ4plex	144.102063	fixed	C	172
Oxidation	15.994915	variable	M	143
Gln->pyro-Glu	-17.026549	variable	N-term	35
Acetyl	42.010565	variable	Protein N-term	21

## 7-2-7. Legend(凡例)

### ▶ Legend

「Legend」の箇所をクリックすると、結果画面の見方に関する説明が表示されます。

▼ Legend

Peptide columns and rows

Dupes	...	Expect	Rank	U	1	2	Peptide	
		0.037	▶	2			GAYSLR	significant
		9	▶	1			GFPLFVEGGR	top ranking
		6.4e-05	▶	1			GSSIFGLAPGK	significant and top ranking
		1.3e-06	▶	1			SSGTSYDPVLK	peptide is found in all proteins in family member 1
		6.2e-07	▶	1			VCNVYSWIK	peptide is found in some but not all proteins in family member 2
		6.4e-05	▶	1	U		GSSIFGLAPGK	unique
▶	2	5.7e-05	▶	1			LNTLETEEWFFK	peptide has two duplicates
		0.18	▶	1			LNTLETEEWFFK	duplicate peptide

Right-facing triangle (▶) in the Dupes or Rank column indicates content that can be expanded by clicking on it. Down-facing triangle (▼) indicates the content is expanded and can be collapsed. For more details about particular columns, see [results format help](#).

Protein quantitation ratios

Score	...	114/113	115/113
CFAH_HUMAN	37559	0.962	<b>1.129</b>
FHR2_HUMAN	1330	<b>0.859</b>	1.128

When quantitation method is Reporter (e.g. iTRAQ) or Multiplex (e.g. iPTL), protein ratios are displayed when a family is expanded. Ratios in *italic* indicate that the peptide log-ratios do not appear to come from a normal distribution. **Bold** indicates that if you can assume peptide ratios are normally distributed, the protein ratio is significantly different from 1.0 (at significance level 0.05).

Note that lack of bold or italic can also mean that significance or normality testing has not been performed (for example, if protein ratio type does not support it).

## 7-2-8.表示内容の切り替え【スコア足切りなど】

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

表示するタンパク質やペプチドに関する条件を指定し、表示内容を変更します。

### Significance threshold p < :

有意性の閾値 p の設定で、デフォルトは 0.05。現バージョンでは実質、「ペプチドの期待値の閾値」と同義。"Target FDR"の値を設定するとそちらが優先され、FDR の設定を満たすような値が自動的に調整・設定されます。

### Max.number of families :

結果として表示するタンパク質(ファミリー)数の上限値。デフォルトは Auto、すなわち同定基準を超えるタンパク質をすべて表示する設定です。

### Target FDR (overrides sig.threshold) :

FDR の設定値。詳細は「11. FDR と Decoy データベース」を参照。

### FDR type :

FDR のカウント対象。PSM は query を数え上げるのに対し、Sequence は同じペプチド配列にマッチする query を1つにまとめてカウントします。

### Display non-sig. matches :

同定基準を超えるペプチドのみ表示されるのがデフォルト設定ですが、タンパク質にアサインされているペプチドすべてを結果に表示させたり、スコアで閾値を設けて表示させたりする事が可能です。

### Min. number of sig. unique sequences :

同定タンパク質の基準として、アサインされるユニークなペプチドをいくつとするかという設定です。MASCOT のデフォルトは 1 です。9-5 で後述するように、設定値を 2 以上にすることで同定タンパク質のリストの信頼性を引き上げる事ができる反面、同定タンパク質数は大幅に減ってしまいます。

### Show Percolator scores :

Percolator の計算を実行している場合のみ。Percolator のスコアと期待値が結果に表示されます。

### Dendrograms cut at :

ファミリータンパク質の類似度を表しているデンドログラムについて、スコアのカットオフを指定します。デフォルトではカットオフを実行しない 0 です。カットオフの結果によって類似するタンパク質が1つにまとめられることがあります。

### Unigene index :

(例図になし、Unigene の設定を行っている塩基配列データベースで検索した時のみ。)タンパク質を遺伝子ベースのファミリーにクラスタリングするために使用する UniGene インデックスを選択し結果画面で表示します。

### Error tolerant matches :

Error tolerant 検索を行った時のみ。検索内容に対する信頼度に基づき検索結果の表示内容を切り替える事ができます(右図)。

Error tolerant matches:	Reliable ▾
Preferred taxonomy	Reliable
	None
	All

### Preferred taxonomy :

優先表示させる生物種。タンパク質の 1 つのエントリーには複数生物種エントリーが統合されている事があります。検索パラメーター「taxonomy」の絞り込みはエントリーに登録されている複数生物種すべてに対応できる一方、表示される生物種はデフォルトで定められた1種類のみです。結果、パラメーターで指定した生物種と明らかに異なる生物種由来と見受けられるエントリーが結果に表示されることがあります。Preferred taxonomy の設定を行う事で、優先して表示させる生物種名を指定した生物種に切り替える事ができます。

## 7-2-9. Sensitivity and FDR

### ▶ Sensitivity and FDR (reversed protein sequences)

「Sensitivity and FDR」の箇所をクリックすると、protein FDR の値と peptide または PSM の FDR の値が表示されます(次頁図)。

▼ **Sensitivity and FDR (reversed protein sequences)**

	Target	Decoy	FDR
Protein family members	480	171	35.62%
PSMs <input type="text" value="above"/> homology <input type="text" value="homology"/>	2812	219	7.79%

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

FDR の値を同定基準として利用したい場合は その上に表示されている「**Format**」欄の「**Target FDR**」で数値を設定して、「**Format**」ボタンを押してください。

FDR に関する詳細は、「**11.FDR と Decoy データベース**」をご覧ください。

## 7-2-10. 同定タンパク質とアサインペプチド

同定タンパク質に関する情報がまとめられた箇所です。

The screenshot shows a web interface for protein families. At the top, there are tabs for 'Proteins (480)', 'Report Builder', and 'Unassigned (30379)'. Below this, the title is 'Protein families 1-10 (out of 448)'. There are navigation controls for 'per page' (set to 10) and page numbers (1, 2, 3, 4, 5, 6, ..., 45, Next). There are also 'Expand all' and 'Collapse all' buttons. Below the navigation is a search bar with 'Accession' and 'contains' dropdowns, a 'Find' button, and a 'Clear' button. The main content area shows a list of protein families with a dendrogram on the left. The list includes:

ID	Accession	Count	Description
1	1::sp TRY1_BOVIN	1600	sp TRY1_BOVIN
2	2::CP2CT_MOUSE	1332	Cytochrome P450 2C29 OS=Mus musculus OX=...
6	2::CP239_MOUSE	251	Cytochrome P450 2C39 OS=Mus musculus OX=...
7	2::CP238_MOUSE	150	Cytochrome P450 2C38 OS=Mus musculus OX=...
2	2::CP254_MOUSE	550	Cytochrome P450 2C54 OS=Mus musculus OX=...
3	2::CY250_MOUSE	487	Cytochrome P450 2C50 OS=Mus musculus OX=...
5	2::CP237_MOUSE	338	Cytochrome P450 2C37 OS=Mus musculus OX=...
4	2::CP2F2_MOUSE	470	Cytochrome P450 2F2 OS=Mus musculus OX=1...

上図のうち上側の部分を拡大して表示したのが以下の図です。

This is a zoomed-in view of the top part of the interface. The 'Proteins (480)' tab is highlighted with a red box. The title 'Protein families 1-10 (out of 448)' also has '448' highlighted with a red box. The navigation controls and search bar are visible below.

赤で囲われた部分のうち、数字 **480** の箇所はリストアップされた**同定タンパク質**の種類、数字 **448** の部分は**同定タンパク質ファミリー**の数です。ファミリーとはシェアペプチドを持つタンパク質をまとめたグループです。詳細は「**9-6. Protein inference**」をご覧ください。

上記情報以外に、1 ページに表示するタンパク質ファミリーの数、現在表示されているページとそれ以外のページへの移動、表示内容の全展開/全収縮、そして項目の検索欄が準備されています。

続いて各ファミリーの情報が記された箇所について説明します。

▶2 など、各タンパク質ファミリーの先頭には三角マークと rank の数字が表示されています。三角部分をクリックすると、タンパク質のマッチングに関するより詳しい情報、タンパク質にアサインされているペプチドのマッチング情報などが表示されます(下図)。

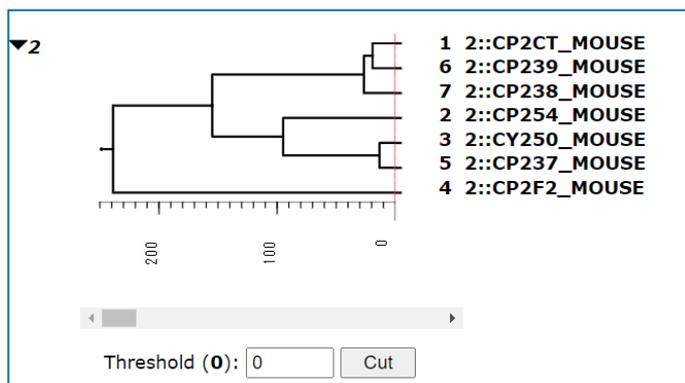
The screenshot displays a search results page with the following components:

- Search Bar:** Accession contains Find Clear
- Dendrogram:** A tree diagram showing hierarchical clustering of protein families. A red vertical line is positioned at a score of 0. The families listed are:
  - 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
- Sequence List:** A list of protein entries with their accession numbers and descriptions.
  - 1600 sp|TRY1\_BOVIN|
  - 1332 Cytochrome P450 2C29 OS=Mus musculus OX=10090 GN=Cyp2c29 PE=...
  - 251 Cytochrome P450 2C39 OS=Mus musculus OX=10090 GN=Cyp2c39 PE=...
  - 150 Cytochrome P450 2C38 OS=Mus musculus OX=10090 GN=Cyp2c38 PE=...
  - 550 Cytochrome P450 2C54 OS=Mus musculus OX=10090 GN=Cyp2c54 PE=...
  - 487 Cytochrome P450 2C50 OS=Mus musculus OX=10090 GN=Cyp2c50 PE=...
  - 338 Cytochrome P450 2C37 OS=Mus musculus OX=10090 GN=Cyp2c37 PE=...
  - 470 Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=1 S...
- Table:** A table with columns: Rank, Score, Mass, Matches, Sequences, emPAI.
 

Rank	Score	Mass	Matches	Sequences	emPAI
2.1	1332	61419	76 (76)	13 (13)	2.00
2.2	550	60887	27 (27)	8 (8)	0.88
2.3	487	61128	27 (27)	10 (10)	1.20
2.4	470	59267	32 (32)	12	1.20
2.5	338	60590	22 (22)	4	0.37
2.6	251	60856	13 (13)	4 (4)	0.37
2.7	150	61356	9 (9)	4 (4)	0.37
- Peptide Match Table:** A table with columns: Query Dupes, Observed, Mr(expt), Mr(calc), Delta M, Score, Expect, Rank, U, 1-7, Peptide.
 

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	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	U								VQEEIDR.V
4447	521.2416	1560.7029	1559.8187	0.8842	0.59	1.2e-05	▶1	U								K.NISQSFTNFSK.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.EALVDHGEEFAGR.G
4741	526.3505	1050.6864	1050.5323	0.1541	0.34	0.005	▶1	U								R.CLVEELR.K
5544	540.3247	1078.6349	1078.5385	0.0964	0.54	0.0002	▶1	U								R.ICAGEGLR.M
5605	541.3848	1080.7551	1080.6059	0.1492	0.53	0.00035	▶1	U								K.YPDVTAK.V
7790	577.9297	1153.8449	1153.6045	0.2404	0.49	0.00049	▶1	U								R.GSFPMAEK.I
8013	581.2500	1160.4854	1160.6167	-0.1313	0.43	0.0017	▶1	U								R.LCLGEPLR.M
8340	586.0058	1169.9970	1169.5994	0.3976	0.33	0.0015	▶1	U								R.GSFPMAEK.I + Oxidation (M)
10395	618.0392	1234.0638	1233.8304	0.2335	0.43	0.0056	▶1*	U								R.YAIIILLK.Y
11198	630.7186	1889.1341	1889.9427	-0.8087	0.59	7.6e-06	▶1	U								R.TSMPYTDVAVIEVQR.F
11589	636.3614	1906.0624	1905.9377	0.1247	0.32	0.0034	▶1	U								R.TSMPYTDVAVIEVQR.F + Oxidation (M)

展開された時に表示される内容について、上図の青で囲われた各パーツ単位でより詳しい説明を行います。



デンドログラムの箇所には、共通ペプチドを持つタンパク質の類似度が表現されています。より右側で結合しているタンパク質がより類似している事を表します。類似度を表すスコア(横軸)が可動バーになっていて、変更した場合の結果がインタラクティブに表示されます。(スコアを直接数字入力し Cut ボタンを押す事でも切り替わります。)

		Score	Mass	Matches	Sequences	emPAI		
<input checked="" type="checkbox"/>	2.1	<a href="#">2::CP2CT_MOUSE</a>	1332	61419	76 (76)	13 (13)	2.00	Cytochrome P450 2C29 OS=Mus m
<input checked="" type="checkbox"/>	2.2	<a href="#">2::CP254_MOUSE</a>	550	60887	27 (27)	8 (8)	0.88	Cytochrome P450 2C54 OS=Mus m
<input checked="" type="checkbox"/>	2.3	<a href="#">2::CY250_MOUSE</a>	487	61128	27 (27)	10 (10)	1.20	Cytochrome P450 2C50 OS=Mus m
<input checked="" type="checkbox"/>	2.4	<a href="#">2::CP2F2_MOUSE</a>	470	59267	32 (32)	12		Cytochrome P450 2C50 OS=Mus musculus SV=2
<input checked="" type="checkbox"/>	2.5	<a href="#">2::CP237_MOUSE</a>	338	60590	22 (22)			
<input checked="" type="checkbox"/>	2.6	<a href="#">2::CP239_MOUSE</a>	251	60856	13 (13)	4 (4)	0.37	Cytochrome P450 2C39 OS=Mus m
<input checked="" type="checkbox"/>	2.7	<a href="#">2::CP238_MOUSE</a>	150	61356	9 (9)	4 (4)	0.37	Cytochrome P450 2C38 OS=Mus m

Redisplay All None

ファミリーに属する各タンパク質の情報が表示されています。図例は rank 2 ですが、rank2 の中でスコア順にサブの番号がつけられ、それぞれ 2.1, 2.2 などと表現されています。そのほかの項目は以下の通りです。

**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 Ions 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。詳細は 9-5 をご覧ください。

**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。( )内の数字は其中で同定基準を超えているもの。現在のデフォルト設定では同定基準を超えている query のみ結果画面に表示されているため( )内外の数字が同じです。

**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。タンパク質にアサインされたペプチド配列数をもとに算出し、かつタンパク質の大きさをもとに標準化されているため結果内の異なる大きさのタンパク質間でも比較が可能。詳細は以下 URL をご覧ください。

[https://www.matrixscience.com/help/quant\\_empai\\_help.html](https://www.matrixscience.com/help/quant_empai_help.html)

続いて、アサインペプチドに関して表示しています。

▼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
<a href="#">3466</a>		503.3162	1004.6178	1004.5083	0.1095	0	31	0.015	▶1	U			■				R.MPTLEDR.T
<a href="#">3505</a>	▶4	503.8846	1005.7547	1005.6093	0.1454	0	36	0.016	▶1	U			■				R.FSVQILR.N
<a href="#">4193</a>		516.8977	1031.7808	1031.5369	0.2439	0	32	0.05	▶1		■		■				VQEEIDR.V
<a href="#">4447</a>	▶2	521.2416	1560.7029	1559.8187	0.8842	0	59	1.2e-05	▶1	U	■						K.NISQSFTNFSK.A
<a href="#">4466</a>		521.3753	1040.7361	1040.5810	0.1551	0	22	0.031	▶1		■						R.FTLMTLR.N + Ox
<a href="#">4705</a>		525.4566	1573.3479	1572.7654	0.5824	0	71	1.3e-05	▶1			■		■			K.EALVDHGEEFAGR.

**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。Query 番号が小さいデータはペプチドの質量が小さい事を示します。

- 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 欄に格納されます。Dupes で表示される数字は query がいくつまとめられているかを表し、三角印をクリックするとその query が表示されます。
- Observed** : 実験値側のペプチドの m/z
- 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
- Mr(calc)** : 配列から計算された、理論値側のペプチドの質量
- Delta** : Mr(expt) – Mr(calc)
- M** : Missed cleavage が実際に適用された数
- Score** : Mascot Ions 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
- 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、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を示します。「Rank」の箇所をクリックして展開すると、下図のようにマッチングとスコアリングを行った他のペプチドとそのスコアについて rank 順に表示されます。また上部にはその query で定められた2つの同定基準値も併せて表示されます(詳細は「9-4.同定ペプチド」をご覧ください)。

Locus: 3.645.3			
Score > 33 indicates identity			
Score > 23 indicates homology			
3	0.4100 0	63	6.2e-06 ▼ <sub>1</sub> ■ ■ <b>R.DFIDYYLIK.Q</b>
	-0.5813 0	7	2.5 2 DFPETNNILK
	0.3187 1	6	3.4 3 TPPIIHRDLK
	-0.7060 1	5	3.5 4 KETMALILK
	0.3816 0	5	4.1 5 DTLSINATNIK
	0.5306 1	5	4.1 6 DPYRDLDMHR + Oxidation (M)
	0.5273 1	4	5 7 SRGACEPDPSLR
	0.4106 0	3	6.7 8 SEVLPQEMLK + Oxidation (M)
	-0.5813 0	3	6.9 9 SLLDANFEPGK
	-0.3703 0	2	7.4 10 MFGCDVGS DWR + Oxidation (M)

- 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 ペプチドである事を意味します。
- 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。同じファミリー内で■が1つのみ表示されているペプチドがユニークペプチドという事になります。
- 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されます。

## 7-2-11. Report Builder タブ (タンパク質ベースの検索結果ファイル出力)

- 7-2-10 でご案内した同定タンパク質の表示個所について、表示領域全体がタブ構成になっています。7-2-10 でご案内したのは「Proteins」タブでしたが、それ以外に2つのタブがあります。「Report Builder」(下図)は、同定タンパク質ベースのリストを作成するのに便利なページです。

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

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 (480) Report Builder Unassigned (30379) [S\\_permalink](#)

**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	<a href="#">sp TRY1_BOVIN </a>	1600	28266	47	47	7	7	2.86	sp TRY1_BOVIN
2	1	SwissProt	<a href="#">2::CP2CT_MOUSE</a>	1332	61419	76	76	13	13	2.00	Cytochrome P450 2C29 OS=
2	2	SwissProt	<a href="#">2::CP254_MOUSE</a>	550	60887	27	27	8	8	0.88	Cytochrome P450 2C54 OS=
2	3	SwissProt	<a href="#">2::CY250_MOUSE</a>	487	61128	27	27	10	10	1.20	Cytochrome P450 2C50 OS=
2	4	SwissProt	<a href="#">2::CP2F2_MOUSE</a>	470	59267	32	32	12	12	2.11	Cytochrome P450 2F2 OS=M
2	5	SwissProt	<a href="#">2::CP237_MOUSE</a>	338	60590	22	22	8	8	0.89	Cytochrome P450 2C37 OS=
2	6	SwissProt	<a href="#">2::CP239_MOUSE</a>	251	60856	13	13	4	4	0.37	Cytochrome P450 2C39 OS=
2	7	SwissProt	<a href="#">2::CP238_MOUSE</a>	150	61356	9	9	4	4	0.37	Cytochrome P450 2C38 OS=

上図のようなリストが作成され、表示内容を CSV ファイルで出力する事ができます。さらに、表示する項目を調整する「Columns」、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「Filters」などといった機能を使用する事も出来ます(下図)。

Proteins (480) 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
- pl

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

Num. of significant matches ≥ 2 Remove

AND Sequence coverage < >

Update

## 7-2-12. Unassigned タブ

「Unassigned」タブでは、同定タンパク質にアサインされていないペプチドの一覧が表示されます。表示項目については、7-2-10 で示したタンパク質にアサインされているペプチドとほとんど同じですので、不明な項目がありましたらそちらをご参照ください。

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 1::sp|TRY1\_BOVIN| 1600 sp|TRY1\_BO



Proteins (480) Report Builder **Unassigned (30379)**

Unassigned peptides, 1-100 (out of 30379)

100 per page 1 2 3 4 5 6 ... 304 Next Sort by Decreasing score

Query Filter Clear

Peptide matches not assigned to protein families (no details means no match)

Query	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	Peptide
<a href="#">6212</a>	552.5647	1103.1149	1102.6590	0.4559	0 35	0.051	▶1	LQEQLGK
<a href="#">23411</a>	854.5979	2560.7719	2560.3563	0.4156	0 34	0.051	▶1	SIVHPSYNSNTLN
<a href="#">17701</a>	731.5877	1461.1608	1460.7932	0.3677	0 34	0.064	▶1	VLPAPMLQYGGR + Oxidation (M)
<a href="#">6704</a>	560.5627	1119.1109	1118.6790	0.4319	0 33	0.05	▶1	EELLSLK
<a href="#">6649</a>	560.0669	1118.1193	1117.7314	0.3878	0 33	0.052	▶1	IVTSLGLK
<a href="#">11432</a>	422.7268	1265.1586	1264.6511	0.5075	1 33	0.058	▶1	KEECPAVR
<a href="#">7108</a>	567.4183	1132.8220	1132.7059	0.1161	0 33	0.058	▶1	SGVSTLALK

## 7-3. Protein View

結果画面では様々なハイパーリンクがあります。タンパク質名(Accession)のハイパーリンクをクリックすると、マッチしたタンパク質についてより詳しい情報が記されている「Protein View」の画面となります(次頁図)。

次々頁以降、次頁図の青で囲われた各表示領域についてより詳しく説明します。

**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 (M<sub>r</sub>): 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:\iPRG2008\mgf\merged.mgf  
 Enzyme: Trypsin/P: cuts C-term side of KR.  
 Fixed modifications: [ITRAQ4plex \(K\)](#), [ITRAQ4plex \(N-term\)](#), [Methylthio \(C\)](#)  
 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 LSCILLLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
51 SQSFNFNSKA YGVPVTLVYG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVVELRKTK GSPCDPTFIL SCAPCNVICV IIFQNRFDYK DKEFLILMDK
201 INENVKILSS PVLQVCNSFP SLIDYCPGSH HRIVKFNFLY KSYLLERIKI
251 HRRESLDVTNP RFIDYLLIK QKQVNHIEQS EFSLENLAST INDLFGAGTE
301 TTSTTLRYAL LLLLKYPDVT AKVQEEIDRV VGRHRSPCMQ DRSHMPYTD
351 MIHEVQRFDI LLFTSLPHAV TCDIKFRKYL IPKGTTVITS LSSVLHDSKE
401 FPNPEMFDPG HFLNGNGNFK KSDYFMPFST GKRICAGEGL ARMELFLILT
451 TILQNFKLS LVHPEIDIT FVMNGFASLP PFYQLCFIPL
    
```

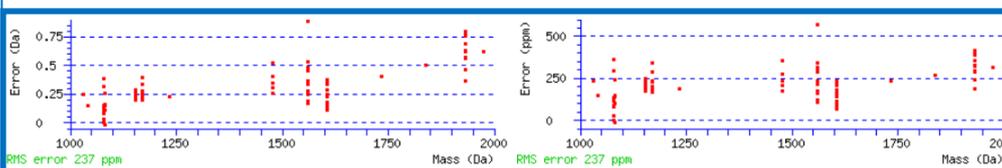
Unformatted sequence string: [490 residues](#) (for pasting into other applications).

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	
<a href="#">20249</a>	49 - 59	781.0005	1559.9865	1559.8187	0.1677	0	76	1.1e-06	1	U	K.NISQSFNFNSK.A
<a href="#">20252</a>	49 - 59	781.0128	1560.0110	1559.8187	0.1922	0	34	0.011	1	U	K.NISQSFNFNSK.A
<a href="#">20256</a>	49 - 59	781.0421	1560.0696	1559.8187	0.2509	0	76	6.3e-07	1	U	K.NISQSFNFNSK.A
<a href="#">20257</a>	49 - 59	781.0544	1560.0942	1559.8187	0.2754	0	57	4.4e-05	1	U	K.NISQSFNFNSK.A
<a href="#">20262</a>	49 - 59	781.0857	1560.1569	1559.8187	0.3381	0	71	1.9e-06	1	U	K.NISQSFNFNSK.A
<a href="#">20263</a>	49 - 59	521.0639	1560.1699	1559.8187	0.3512	0	49	0.00015	1	U	K.NISQSFNFNSK.A
<a href="#">20264</a>	49 - 59	781.1019	1560.1892	1559.8187	0.3704	0	87	8.1e-08	1	U	K.NISQSFNFNSK.A
<a href="#">20267</a>	49 - 59	781.1422	1560.2698	1559.8187	0.4511	0	90	2e-08	1	U	K.NISQSFNFNSK.A
<a href="#">20268</a>	49 - 59	781.1516	1560.2887	1559.8187	0.4700	0	86	9.7e-08	1	U	K.NISQSFNFNSK.A
<a href="#">20269</a>	49 - 59	781.1604	1560.3063	1559.8187	0.4875	0	74	8.5e-07	1	U	K.NISQSFNFNSK.A
<a href="#">20270</a>	49 - 59	521.1236	1560.3491	1559.8187	0.5304	0	54	4.8e-05	1	U	K.NISQSFNFNSK.A
<a href="#">4447</a>	49 - 59	521.2416	1560.7029	1559.8187	0.8842	0	59	1.2e-05	1	U	K.NISQSFNFNSK.A
<a href="#">21389</a>	85 - 97	536.3181	1605.9324	1605.8232	0.1091	1	43	0.0014	1	U	K.EALIDRGEFAGR.G
<a href="#">21390</a>	85 - 97	536.3278	1605.9617	1605.8232	0.1384	1	46	0.0014	1	U	K.EALIDRGEFAGR.G
<a href="#">21392</a>	85 - 97	536.3355	1605.9846	1605.8232	0.1613	1	44	0.00077	1	U	K.EALIDRGEFAGR.G
<a href="#">21395</a>	85 - 97	536.3434	1606.0083	1605.8232	0.1850	1	46	0.002	1	U	K.EALIDRGEFAGR.G

=====**中略**=====

<a href="#">5544</a>	434 - 442	540.3247	1078.6349	1078.5385	0.0964	0	54	0.0002	1	R	R.ICAGEGLAR.M
<a href="#">5548</a>	434 - 442	540.3388	1078.6630	1078.5385	0.1246	0	45	0.0016	1	R	R.ICAGEGLAR.M
<a href="#">5549</a>	434 - 442	540.3471	1078.6796	1078.5385	0.1412	0	33	0.011	1	R	R.ICAGEGLAR.M
<a href="#">5550</a>	434 - 442	540.3498	1078.6851	1078.5385	0.1466	0	39	0.0074	1	R	R.ICAGEGLAR.M
<a href="#">5566</a>	434 - 442	540.4353	1078.8560	1078.5385	0.3175	0	39	0.008	1	R	R.ICAGEGLAR.M
<a href="#">5573</a>	434 - 442	540.4708	1078.9271	1078.5385	0.3886	0	40	0.0072	1	R	R.ICAGEGLAR.M



ID CP2CT\_MOUSE Reviewed: 490 AA.  
 AC Q64458; E9QMD8; Q8WUN8;  
 DT 15-JUL-1999, integrated into UniProtKB/Swiss-Prot.  
 DT 27-JUL-2011, sequence version 2.  
 DT 02-JUN-2021, entry version 172.  
 DE RecName: Full=Cytochrome P450 2C29;  
 DE EC=1.14.14.1 (ECO:0000269|PubMed:9721182);  
 DE AltName: Full=Aldehyde oxygenase;  
 DE AltName: Full=CYP11C29;  
 DE AltName: Full=Cytochrome P-450 MUT-2;  
 DE Flags: Precursor;  
 GN Name=Cyp2c29 (ECO:0000303|PubMed:9721182, ECO:0000312|MGI:MGI:103238);  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
 OC Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae;  
 OC Murinae; Mus; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].

## Protein View: CP2CT\_MOUSE

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

Database: SwissProt  
Score: 1332  
Monoisotopic mass (M<sub>r</sub>): 61419  
Calculated pI: 8.57  
Taxonomy: [Mus musculus](#)

Sequence similarity is available as [an NCBI BLAST search of CP2CT\\_MOUSE against nr](#).

ページの最初にタンパク質の **Acession** と **Description** が表示されます(上図)。続いて以下のような情報が表示されます。

**Database** : 使用したデータベース名  
**Score** : タンパク質のスコア  
**Monoisotopic mass (Mr)** : データベースの配列から計算されたタンパク質の質量  
**Calculated pI** : データベースの配列から計算された予測等電点  
**Taxonomy** : 生物種

また該当タンパク質の配列を NCBI の BLAST(配列相同性検索プログラム)実行するためのリンクが表示されます。

### Search parameters

MS data file: D:\iPRG2008\mgf\merged.mgf  
Enzyme: Trypsin/P: cuts C-term side of KR.  
Fixed modifications: [iTRAQ4plex \(K\)](#), [iTRAQ4plex \(N-term\)](#), [Methylthio \(C\)](#)  
Variable modifications: [Acetyl \(Protein N-term\)](#), [Gln->pyro-Glu \(N-term Q\)](#), [Oxidation \(M\)](#)

**MS data file** : 入力データファイルの名称

**Enzyme** : 切断パターン

その他、検索で使用了パラメーターに関する情報が表示されます。

### Protein sequence coverage: 27%

Matched peptides shown in **bold red**.

```
1 MDLVVFLALT LSCLILLSLW 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 HKIVKNFNLYL KSYLLEKIKE
251 HKESLDVTNP RFIDYYLIK QKQVNHIEQS EFSLENLAST INDLFGAGTE
301 TTSTTLRYAL LLLLKYPDVT AKVQEEIDRV VGRHRSPCMQ DRSHMPYTDA
351 MIHEVQRFID LLPTSLPHAV TCDIKFRKYL IPKGTTVITS LSSVLHDSKE
401 FPNPEMFDPG HFLNGNGNFK KSDYFMPFST GKRICAGEGL ARMELFLILT
451 TILQNFKLKS LVHPKEIDIT PVMNGFASLP PPYQLCFIPL
```

続いて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、その割合についての情報が表示されます。**Coverage** とは、全長に対するマッチペプチド残基数の割合です。マッチしたペプチド部分が赤の太字で表現されています。

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

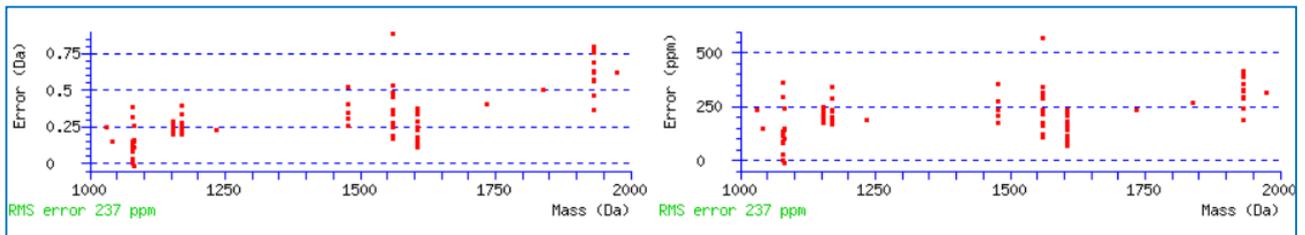
Query	Start - End	Observed	Mr(expt)	Mr(calc)	Delta	M	Score	Expect	Rank	U	Peptide
<a href="#">20249</a>	49 - 59	781.0005	1559.9865	1559.8187	0.1677	0	76	1.1e-06	1	U	K.NISQSFTNFSK.A
<a href="#">20252</a>	49 - 59	781.0128	1560.0110	1559.8187	0.1922	0	34	0.011	1	U	K.NISQSFTNFSK.A
<a href="#">20256</a>	49 - 59	781.0421	1560.0696	1559.8187	0.2509	0	76	6.3e-07	1	U	K.NISQSFTNFSK.A
<a href="#">20257</a>	49 - 59	781.0544	1560.0942	1559.8187	0.2754	0	57	4.4e-05	1	U	K.NISQSFTNFSK.A
<a href="#">20262</a>	49 - 59	781.0857	1560.1569	1559.8187	0.3381	0	71	1.9e-06	1	U	K.NISQSFTNFSK.A
<a href="#">20263</a>	49 - 59	521.0639	1560.1699	1559.8187	0.3512	0	49	0.00015	1	U	K.NISQSFTNFSK.A
<a href="#">20264</a>	49 - 59	781.1019	1560.1892	1559.8187	0.3704	0	87	8.1e-08	1	U	K.NISQSFTNFSK.A
<a href="#">20267</a>	49 - 59	781.1422	1560.2698	1559.8187	0.4511	0	90	2e-08	1	U	K.NISQSFTNFSK.A
<a href="#">20268</a>	49 - 59	781.1516	1560.2887	1559.8187	0.4700	0	86	9.7e-08	1	U	K.NISQSFTNFSK.A
<a href="#">20269</a>	49 - 59	781.1604	1560.3063	1559.8187	0.4875	0	74	8.5e-07	1	U	K.NISQSFTNFSK.A
<a href="#">20270</a>	49 - 59	521.1236	1560.3491	1559.8187	0.5304	0	54	4.8e-05	1	U	K.NISQSFTNFSK.A
<a href="#">4447</a>	49 - 59	521.2416	1560.7029	1559.8187	0.8842	0	59	1.2e-05	1	U	K.NISQSFTNFSK.A
<a href="#">21389</a>	85 - 97	536.3181	1605.9324	1605.8232	0.1091	1	43	0.0014	1	U	K.EALIDRGEFAGR.G
<a href="#">21390</a>	85 - 97	536.3278	1605.9617	1605.8232	0.1384	1	46	0.0014	1	U	K.EALIDRGEFAGR.G
<a href="#">21392</a>	85 - 97	536.3355	1605.9846	1605.8232	0.1613	1	44	0.00077	1	U	K.EALIDRGEFAGR.G

タンパク質全長の下にはマッチしたペプチドが、アミノ酸残基順(デフォルト設定の場合)に並んでリスト表示されています(上図)。表示されている情報は以下の通りです。

**Unformatted sequence string :**

残基数と共に配列をコピーしやすくなるページが開きます。他の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** : リストの並び順を指定します。残基番号、質量の昇順/降順 が選択できます。
- 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を選択します。
- Query** : Query 番号。入力データについてペプチド質量の小さい順に付けられた番号。
- Start-End** : タンパク質全長におけるアミノ酸残基番号。
- 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。
- Mr(calc)** : 配列から計算されたペプチドの質量。
- Delta** : Mr(expt) - Mr(calc)。
- M** : Missed cleavage。
- Score** : Mascot Ions score、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。
- Expect** : query 毎に計算される期待値、スコアと同定基準値から算出。
- 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を示します。
- 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 ペプチドである事を意味します。
- 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されます。



ページ下部で表示されているグラフはともに**ピークのマッチングの誤差を表すグラフ**で、左が Da、右が ppm で表現されています(上図)。ともに横軸は実験データ側のペプチド質量で、縦軸が誤差です。含まれている事が確実なタンパク質でこのグラフを確認する事で、パラメーターで指定した誤差範囲 (peptide tol.)の設定値が適切であったかどうかを確認する事もできます。

```

ID   CP2CT_MOUSE                               Reviewed;              490 AA.
AC   Q64458; E9QMD8; Q8WUN8;
DT   15-JUL-1999, integrated into UniProtKB/Swiss-Prot.
DT   27-JUL-2011, sequence version 2.
DT   02-JUN-2021, entry version 172.
DE   RecName: Full=Cytochrome P450 2C29;
DE           EC=1.14.14.1 {ECO:0000269|PubMed:9721182};
DE   AltName: Full=Aldehyde oxygenase;
DE   AltName: Full=CYP11C29;
DE   AltName: Full=Cytochrome P-450 MUT-2;
DE   Flags: Precursor;
GN   Name=Cyp2c29 {ECO:0000303|PubMed:9721182, ECO:0000312|MGI:MGI:103238};
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae;
OC   Murinae; Mus; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA].
RC   STRAIN=C57BL/6J; TISSUE=Liver;

```

画面の最下部には**タンパク質の詳細情報**が表示されます(上図)。ただしデータベース側で情報表示に関する適切な設定がある時のみ表示されます。

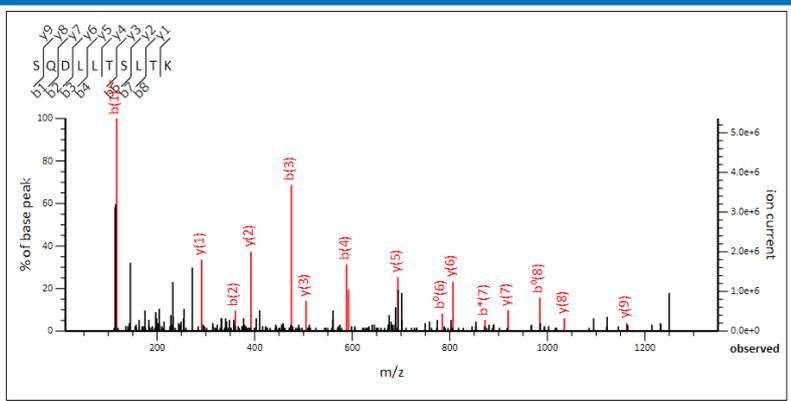
## 7-4. Peptide View

Summary 画面、あるいは Protein view 画面で Query 番号のハイパーリンクをクリックすると、**MS/MS データと理論フラグメントピークとのマッチング度合いなどを確認できる Peptide View** の画面が現れます(次頁図)。

**MATRIX SCIENCE Mascot Search Results**

**Peptide View**

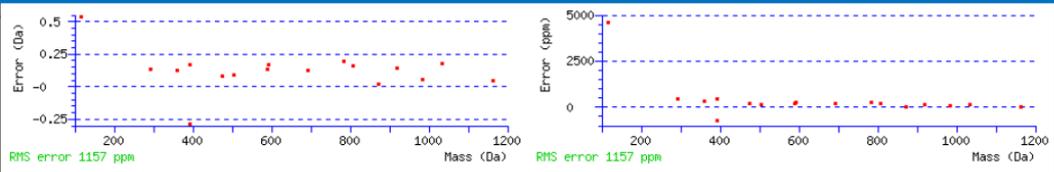
MS/MS Fragmentation of **SQDLLTSLTK**  
 Found in **CP2F2\_MOUSE** in **SwissProt**, Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=1 SV=1  
 Match to Query 15524: 1392.862848 from(697.438700,2+)  
 Title: Locus:1.559.3  
 Data file D:\¥IPRG2008¥mgf¥merged.mgf



Label all possible matches  Label matches used for scoring

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: 76 Expect: 4.2e-08  
 Matches: 17/104 fragment ions using 24 most intense peaks (help)

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	γ	γ <sup>++</sup>	γ*	γ <sup>+++</sup>	γ <sup>0</sup>	γ <sup>0++</sup>	#
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	T	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	S	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



NCBI **BLAST** search of **SQDLLTSLTK**  
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)  
 Other BLAST [web gateways](#)

**All matches to this query**

Score	Mr(calc)	Delta	Sequence
75.7	1392.8068	0.0561	<a href="#">SQDLLTSLTK</a>
13.6	1392.8544	0.0085	<a href="#">SLRSTLSTIK</a>
11.8	1393.7115	-0.8487	<a href="#">GAGMPQSDVTK</a>
10.1	1392.7340	0.1288	<a href="#">SENV DVEVSK</a>
9.6	1392.7231	0.1397	<a href="#">SSWSLSPSRSR</a>
9.5	1392.7735	0.0894	<a href="#">QLHLLLEDAVTK</a>
9.0	1392.8068	0.0561	<a href="#">SVSSAIETAIK</a>
7.6	1392.8002	0.0626	<a href="#">SVRSPLMSTK</a>
7.3	1393.7478	-0.8850	<a href="#">SNEMAINLAK</a>
5.4	1392.8002	0.0626	<a href="#">SQELRMTLK</a>

以降、前頁の青色で囲われた各表示領域について詳しく説明します。

MS/MS Fragmentation of **SQDLLTSLTK**  
Found in **CP2F2\_MOUSE** in **SwissProt**, Cytochrome P450 2F2 OS=Mus musculus OX=10090 GN=Cyp2f2 PE=1 SV=1  
Match to Query 15524: 1392.862848 from(697.438700,2+)  
Title: Locus:1.559.3  
Data file D:\iPRG2008\mgf\merged.mgf

ペプチド配列並びにこのペプチドがアサインされたタンパク質の情報が表示されます(上図)。また入力データ側の情報として、query 番号や title 行の情報、そして元のファイルのパスや名称に関する情報も表示されます。

以降の箇所を説明する前に、重要なポイントである入力データの再調整と表示ピークについて説明をします。

Label all possible matches  Label matches used for scoring

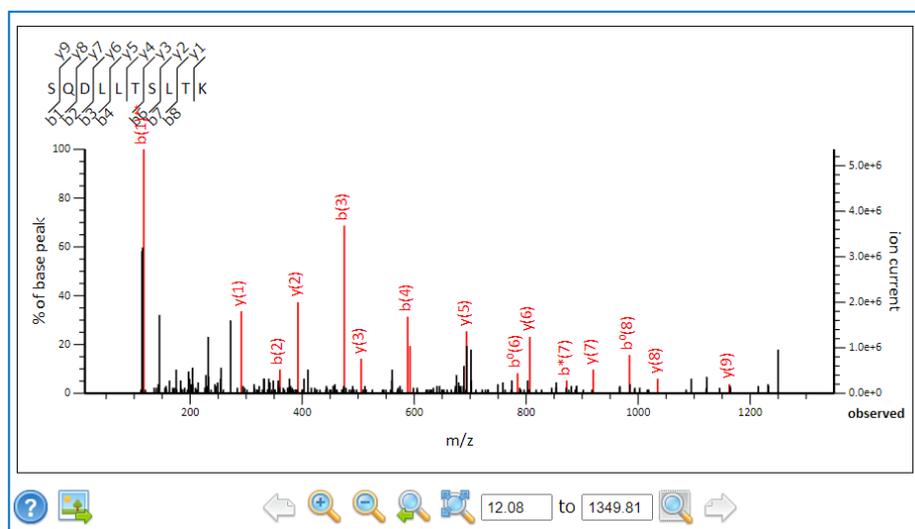
#### [Label matches used for scoring]

スコアリングに使用されたピークだけを表示してマッチング有無の表示対象とするか

#### [Label all possible matches]

スコアリングに使われなかったピークも表示対象とするか

「9-2 入力データの調整」でも説明していますが、MASCOT では入力データについてピーク強度等を基に再構成を行います。入力データ(mgf などピークリストファイル)にあるすべてのピークがスコアリングに使われるわけではありません。元のピークリストファイルを見ると存在するピークが、peptide view のマッチング表・図で理論値とマッチングしてない場合、そのピークが何らかの理由でスコアリングに使われるピークの選別から漏れたためと考えられます。マッチングしていない事に疑問を感じた場合はまず「Label all possible matches」を選択して、該当ピークが表示されるか、改めてご確認ください。



選択項目の上に表示されているのが、**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** (前頁図)です。理論値とマッチした場所が赤で表示されます。また表示領域の設定や、マッチング図のファイル出力に関するアイコンも併せて表示されています。( \*ここで使用されている(Xi Spectrum Viewer は、エジンバラ大学の Rappsilber 研究室によって開発され、Artistic License 2.0 で公開されています。また関連するいくつかのアイコンは、Farm-fresh web icons, released under the Creative Commons Attribution 3.0 License.から引用されています。) 色の設定なども変更可能です。

このグラフでは、peptide molecular ion 周辺のピークがスペクトルから除去されています。MASCOT の検索ではこのピークを使わないのでその事を反映させるためですが、別の場所に表示させた元のスペクトルと見た目が変わる事があるのでご注意ください。

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: 76 Expect: 4.2e-06  
 Matches : 17/104 fragment ions using 24 most intense peaks ([help](#))

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	γ	γ <sup>++</sup>	γ*	γ <sup>+++</sup>	γ <sup>0</sup>	γ <sup>0++</sup>	#
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	T	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	S	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

続いて理論値の一覧表をベースとしたマッチングの確認画面が現れます(上図)。表の上にはペプチドの質量や修飾、スコアと期待値の情報が表示されます。「Matches」の行については少し補足説明をします。

**Matches** : 17/104 fragment ions using 24 most intense peaks ([help](#))

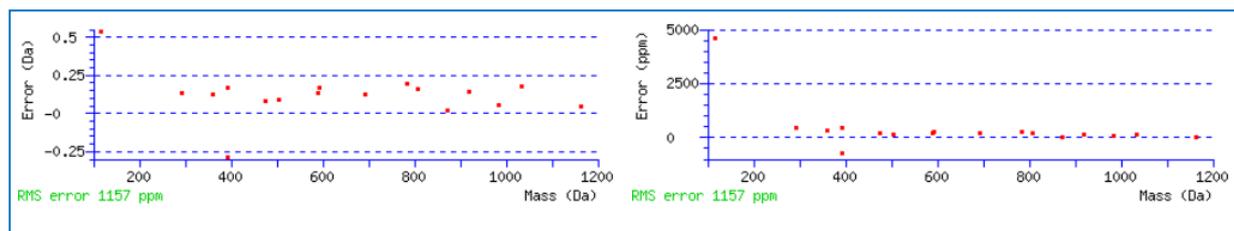
**104** は表の中の数字すべて、その中で理論値と実測値がマッチしたのが **17**(表の中の赤い文字の数すべて)であった事を意味します。**24** most intense peaks は、**入力データ**のうち 24 のピークが**入力データ**側の代表ピークとして使われていることを示しています。

表の中で使用されている文字の色・形態についても補足説明をします。4種類あります。

- 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

上記の赤色の中で区別されるポイントは、**イオンシリーズごとに見てランダムマッチを超えていると言え**

るほどの数がマッチしているかどうか)がです。イオンシリーズ単位である程度の数が見つからないと、そのイオンシリーズはスコアリングにおいて使用されない事を意味します。



マッチング表の下に表示されるグラフはともにピークのマッチングの誤差を表すグラフで、左が Da、右が ppm で表現されています(上図)。ともに横軸は実験データ側のフラグメントの質量で、縦軸が誤差です。含まれている事が確実なペプチドでこのグラフを確認する事で、誤差範囲(MS/MS tol.)の設定値が適切であったかどうかを確認する事もできます。

NCBI **BLAST** search of [SQDLLTSLTK](#)  
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)  
 Other BLAST [web gateways](#)

続いて、ヒットしたペプチド配列を NCBI の BLAST で検索するためのリンク、並びに BLAST を稼働しているその他の外部サイトへのリンクが表示されています(上図)。

Score	Mr(calc)	Delta	Sequence
75.7	1392.8068	0.0561	<a href="#">SQDLLTSLTK</a>
13.6	1392.8544	0.0085	<a href="#">SLRSTLSTIK</a>
11.8	1393.7115	-0.8487	<a href="#">GAGMPQSDVTK</a>
10.1	1392.7340	0.1288	<a href="#">SENV DVEVSK</a>
9.6	1392.7231	0.1397	<a href="#">SSWSLSPRSR</a>
9.5	1392.7735	0.0894	<a href="#">QLHLLEDAVTK</a>
9.0	1392.8068	0.0561	<a href="#">SVSSAIETAIK</a>
7.6	1392.8002	0.0626	<a href="#">SVRSPLMSTK</a>
7.3	1393.7478	-0.8850	<a href="#">SNEMAINLAK</a>
5.4	1392.8002	0.0626	<a href="#">SQELRMTLK</a>

最後に、データベース検索でマッチングとスコアリングを行ったほかの結果について、ページ内で表示されているペプチドも含めた形で rank 順に表示します(上図)。また 1 つの修飾がペプチド内で複数の箇所につく可能性がある場合、上記表の Sequence 列の右側にアミノ酸残基位置特定の確率が表示されます。例えば一方が 99%でもう一方が 1%、といったような数字が表示されます。

## 7-5. Export 機能によるペプチドベースの検索結果ファイル出力

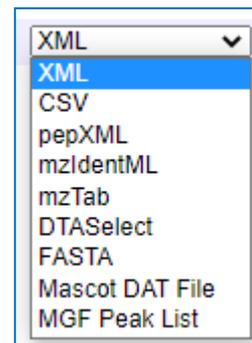
同定結果をファイル出力する事ができます。7-2-11 で紹介した Report builder では同定タンパク質に着目したファイル出力でしたが、**同定ペプチド情報をベースとしてまとめたファイル**や、その他解析結果に関連するファイルの出力をするには、7-2-3 で紹介した Export 機能を利用します。

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



対応しているファイルフォーマットは右図の通りです。

同定ペプチド情報を EXCEL で読み込んで表示可能な形式である「**CSV**」、結果ファイルの共通フォーマットとして認識されている「**mzIdentML**」や MASCOT の結果ファイルである「**Mascot DAT File**」、MASCOT の入力データフォーマットである「**MGF Peak List**」などが比較的使用される事の多い選択肢です。



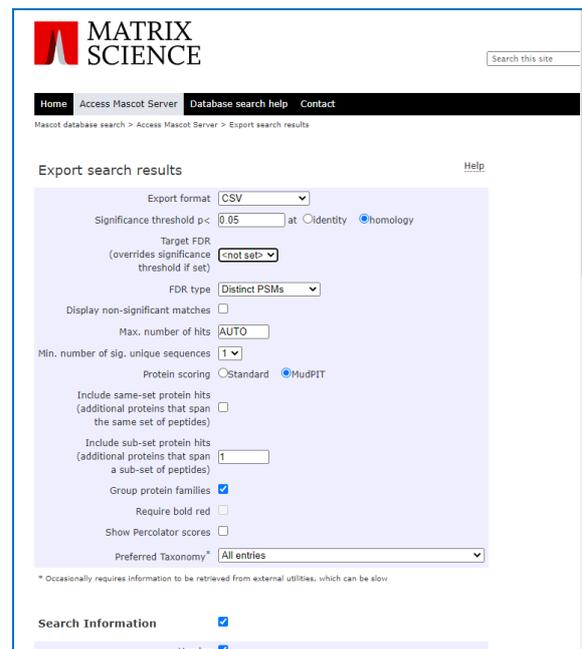
Export ボタンを押すと画面が遷移します(右図)。出力を希望するパラメーターを選択したうえで、画面最下部にある「**Export search results**」ボタンを押し、続いて現れる画面で「**Download**」ボタンを押す事で、サーバー側でのファイル作成とそのファイルのクライアントへのダウンロードが始まります。

出力されるファイルの項目のうち、XML ファイル出力に関する詳細については以下2つの URL をご参照 ください。

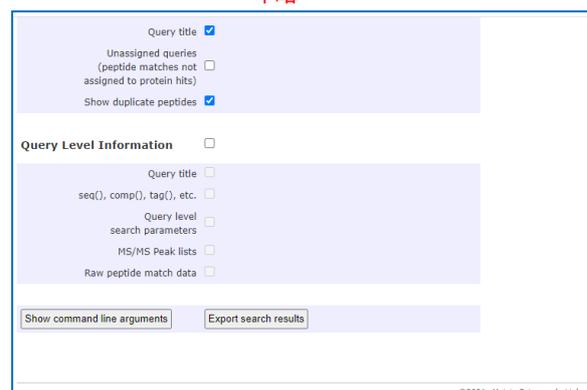
[https://www.matrixscience.com/xmlns/schema/mascot\\_search\\_results\\_2/mascot\\_search\\_results\\_2.xsd](https://www.matrixscience.com/xmlns/schema/mascot_search_results_2/mascot_search_results_2.xsd)

[https://www.matrixscience.com/xmlns/schema/mascot\\_search\\_results\\_2/index.html](https://www.matrixscience.com/xmlns/schema/mascot_search_results_2/index.html)

export で CSV として出力されたファイルの表示列について、次頁以降で説明します。



=====中略=====



<b>prot_hit_num</b>	同定タンパク質ファミリーの順位
<b>prot_family_member</b>	同定タンパク質がファミリーの中で割り振られたサブの番号
<b>prot_acc</b>	タンパク質の Accession
<b>prot_desc</b>	タンパク質の Description
<b>prot_score</b>	タンパク質のスコア(アサインペプチドのスコアをもとに算出)
<b>prot_thresh</b>	同定基準値(タンパク質、PMF の時のみ)
<b>prot_expect</b>	期待値(タンパク質、PMF のみ)
<b>prot_mass</b>	タンパク質の配列から計算された質量
<b>prot_matches</b>	タンパク質にアサインされた query 数
<b>prot_matches_sig</b>	タンパク質にアサインされた、有意基準を超える query 数
<b>prot_sequences</b>	タンパク質にアサインされたペプチド数
<b>prot_sequences_sig</b>	タンパク質にアサインされた、有意基準を超えるペプチド数
<b>prot_cover</b>	シーケンスカバレッジ
<b>prot_len</b>	タンパク質の残基長
<b>prot_pi</b>	タンパク質の配列から計算された予測等電点
<b>prot_tax_str</b>	生物種情報
<b>prot_tax_id</b>	NCBI の Taxonomy ID
<b>prot_seq</b>	タンパク質の配列
<b>prot_empai</b>	emPAI
<b>prot_acc_alpha</b>	クロスリンクペプチドの結果における、1つめのタンパク質の Accession
<b>prot_acc_beta</b>	クロスリンクペプチドの結果における、2 つめのタンパク質の Accession
<b>pep_query</b>	ペプチドの query 番号
<b>pep_rank</b>	ペプチドの rank(マッチング順位)
<b>pep_isbold</b>	1...有意基準を超えている、0...有意基準を超えていない
<b>pep_isunique</b>	ユニークペプチド(タンパク間でシェアされていない)かどうか。1...ユニーク 0...ユニークでない
<b>pep_exp_mz</b>	ペプチド・実験値側の $m/z$
<b>pep_exp_mr</b>	ペプチド・実験値側の質量
<b>pep_exp_z</b>	ペプチド・実験値側の 電荷
<b>pep_calc_mr</b>	ペプチド・理論値側の質量
<b>pep_delta</b>	ペプチドの質量誤差、実験値 - 理論値
<b>pep_start</b>	タンパク質におけるペプチド残基の位置、開始点
<b>pep_end</b>	タンパク質におけるペプチド残基の位置、終了点
<b>pep_miss</b>	該当ペプチドにおける missed cleavage の数
<b>pep_score</b>	Mascot Ions Score(query のマッチングスコア)
<b>pep_homol</b>	query の homology treshold(homology 同定基準値)
<b>pep_ident</b>	query の identity threshold(identity 同定基準値)

<b>pep_expect</b>	query の期待値
<b>pep_res_before</b>	タンパク質において該当ペプチドの1つ前に存在するアミノ酸残基
<b>pep_seq</b>	ペプチドの配列
<b>pep_res_after</b>	タンパク質において該当ペプチドの1つ後ろに存在するアミノ酸残基
<b>pep_frame</b>	翻訳のフレーム番号(塩基配列の検索時のみ)
<b>pep_var_mod</b>	Variable の修飾
<b>pep_var_mod_pos</b>	Variable の修飾がペプチド内で存在する位置を数字の文字列で表しています。「.」で挟まれたのがペプチドで数はペプチド残基長と同じ、最初の「.」の前が N 末端、最後の「.」の前が C 末端です。0 は修飾がないことを、数字は各修飾に割り当てられた修飾の番号(ファイル上部に表示)を意味します。
<b>pep_summed_mod_pos</b>	「pep_var_mod」の補足で、同時に1つのアミノ酸残基に修飾がついているケースに対応するための項目です。
<b>pep_local_mod_pos</b>	Query レベルでの「pep_var_mod_pos」と同様の情報
<b>pep_var_mod_conf</b>	複数の修飾候補領域が同一ペプチド内にある場合、各位置における同定確率
<b>pep_num_match</b>	スコアに使用されたフラグメントマッチの数
<b>pep_scan_title</b>	ピークリストの「title」行に記された情報
<b>pep_source</b>	データベースの種類 (AA, NA, XA, SL)
<b>pep_linked_sites</b>	クロスリンクの状況を表す、3 数字 x 2 で構成された文字列。 <b>例: 1:0:6,2:5:7.</b> アルファペプチド (= 1) のペプチド上の位置 0 (N-term)でリンカーの選択肢が 6 (リンカーと番号の結びつきは「Linkers table」にあります)。一方ベータペプチド (= 2)のペプチド上の位置は 5 残基目、リンカーの選択肢は 7。
<b>pep_res_before_beta</b>	クロスリンクペプチド、ベータペプチドの pep_res_before
<b>pep_seq_beta</b>	クロスリンクペプチド、ベータペプチドの pep_seq
<b>pep_res_after_beta</b>	クロスリンクペプチド、ベータペプチドの pep_res_after
<b>pep_frame_beta</b>	クロスリンクペプチド、ベータペプチドの pep_frame
<b>pep_var_mod_beta</b>	クロスリンクペプチド、ベータペプチドの pep_var_mod
<b>pep_var_mod_pos_beta</b>	クロスリンクペプチド、ベータペプチドの pep_var_mod_pos
<b>pep_summed_mod_pos_beta</b>	クロスリンクペプチド、ベータペプチドの pep_summed_mod_pos
<b>pep_local_mod_pos_beta</b>	クロスリンクペプチド、ベータペプチドの pep_local_mod_pos
<b>pep_monolink_mod_pos</b>	クロスリンクペプチドのうちモノリンクペプチドだったケースでの修飾が付いたアミノ酸残基の位置
<b>pep_monolink_mod_pos_beta</b>	クロスリンクペプチド、ベータペプチドにおける「pep_monolink_mod_pos」
<b>pep_loopleft_pos</b>	クロスリンクペプチド、ループリンクペプチドだったケースでの、ループ残基の位置
<b>pep_loopleft_pos_beta</b>	クロスリンクペプチド、ベータペプチドにおける「pep_loopleft_pos」

## 8. PMF : タンパク質同定

### 8-1. PMF タンパク質同定のまとめ

この章では PMF でのタンパク質同定において説明いたします。重要な点を列挙すると以下の通りです。

- **PMF** では1度の検索で基本的に1種類のタンパク質が同定される [8-4]
- 入力データはピーク作成時とマッチング時の2段階で選別・組み換えされる [8-2]
- スコアとは理論値と実測値のマッチング度合いを評価した数値で、高いほどよくマッチしている [8-4]
- 検索規模をもとに同定基準値が計算される。スコアが同定基準値を超えた時タンパク質を同定とみなす（デフォルト値の場合、信頼度は**95%**） [8-4]
- タンパク質間でシェアされているペプチド情報に十分注意して結果を解釈しなければならない [8-5]
- 「検索対象のデータベースの選択」も結果に大きく影響する [8-3]

以降、順に説明いたします。

### 8-2. 入力データの調整

**入力データは2段階で調整されます。**3-1-1 で説明したように、MASCOT Server ではスペクトルデータをそのまま受け付けておらず検索の前段階でペプチドのピークを抽出しノイズをカットしたものを入力データとして受け付けていて、これが**1段階目の調整**にあたります。入力データに  $m/z$  に加え強度情報も含まれている場合、MASCOT Server の検索プログラムは入力データから強度情報に基づいて**10種類のサブセット入力データ**を作成し、それぞれのサブセットに対して理論ピークとのマッチングを行って最もスコアが高くなった結果を採用します。このように MASCOT Server 側で強度を使ってデータの再構成を行うのが**2段階目の調整**です。

MASCOT Server プログラムは2段階目の処理を行う事ができますが、あくまでも**1段階目の前処理としてピーク抽出が行われていることを前提**としています。また、上記のように**サブセットの選別に強度情報**を使っていますが、**マッチングの評価そのものに関してはピーク強度の情報を利用していません**。強度の高いピークに対してマッチした場合、よりスコアが高くなる、というような判断はありません。

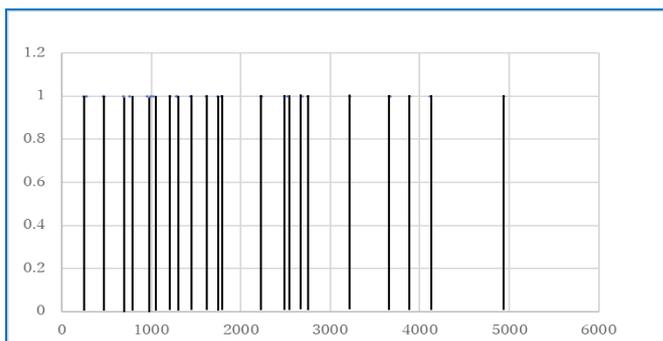
### 8-3. 配列から計算される理論ピーク

入力データとマッチングを行う参照側は、配列から計算されたタンパク質の理論的なピークリストです。検索時に指定したデータベースに登録されているタンパク質1つ1つに対して、パラメーターをもとに理論ペプチドを作成しその質量を計算します。タンパク質毎の理論ペプチドセットを作成して、入力データとのマッチングを行います。この検索方法では「答が含まれているデータベース」で検索をしなければ同定できません。データベースに含まれないタンパク質を答えとしてレポートする事はなく、同定タンパク質の検出も行われません。しかし 8-4 で説明するように、登録エントリー数が多いデータベースで検索をすると同定基準値が高くなり同定には不利です。データベースの選択は検索対象が小さすぎず大きすぎない、必要十分なエントリーが含まれるデータベースである必要があります。

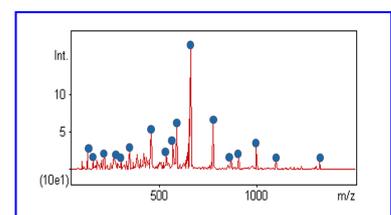
```
>sp|A0A0C4DH27|TRGV8_HUMAN T cell receptor gamma variable 8
MLLALALLLAFLLPPASQKSSNLEGRTKSVTRPTGSSAVITCDLPVENAVYTHWYLNHQQEGKAPQRL
LLYYDSYNSRVVLESGISREKYHTYASTGKSLKFILENLIERDSGVYYCATWDR
```



2 - 18	1778.1070	0	LLALALLLAFLLPPASQK
2 - 25	2521.4632	1	LLALALLLAFLLPPASQKSSNLEGR
2 - 27	2750.6058	2	LLALALLLAFLLPPASQKSSNLEGRTK
19 - 25	761.3668	0	SSNLEGR
19 - 27	990.5094	1	SSNLEGRTK
19 - 31	1433.7587	2	SSNLEGRTKSVTR
26 - 27	247.1532	0	TK
26 - 31	690.4024	1	TKSVTR
26 - 60	3886.9312	2	TKSVTRPTGSSAVITCDLPVENAVYTHWYLNHQQE
28 - 31	461.2598	0	SVTR
28 - 60	3657.7886	1	SVTRPTGSSAVITCDLPVENAVYTHWYLNHQQEGK
28 - 64	4110.0382	2	SVTRPTGSSAVITCDLPVENAVYTHWYLNHQQEGK
32 - 60	3214.5394	0	PTGSSAVITCDLPVENAVYTHWYLNHQQEGK
32 - 64	3666.7889	1	PTGSSAVITCDLPVENAVYTHWYLNHQQEGKAPQF
32 - 74	4941.3821	2	PTGSSAVITCDLPVENAVYTHWYLNHQQEGKAPQF
61 - 64	470.2601	0	APQR
61 - 74	1744.8533	1	APQRLLYYDSYNSR
61 - 83	2685.3875	2	APQRLLYYDSYNSRVVLESGISR
65 - 74	1292.6037	0	LLYYDSYNSR
65 - 83	2233.1379	1	LLYYDSYNSRVVLESGISR
65 - 85	2490.2755	2	LLYYDSYNSRVVLESGISREK
75 - 83	958.5447	0	VVLESGISR
75 - 85	1215.6823	1	VVLESGISREK
75 - 94	2224.1488	2	VVLESGISREKYHTYASTGK
84 - 85	275.1481	0	EK
84 - 94	1283.6146	1	EKYHTYASTGK
84 - 97	1611.8257	2	EKYHTYASTGKSLK
86 - 94	1026.4771	0	YHTYASTGK



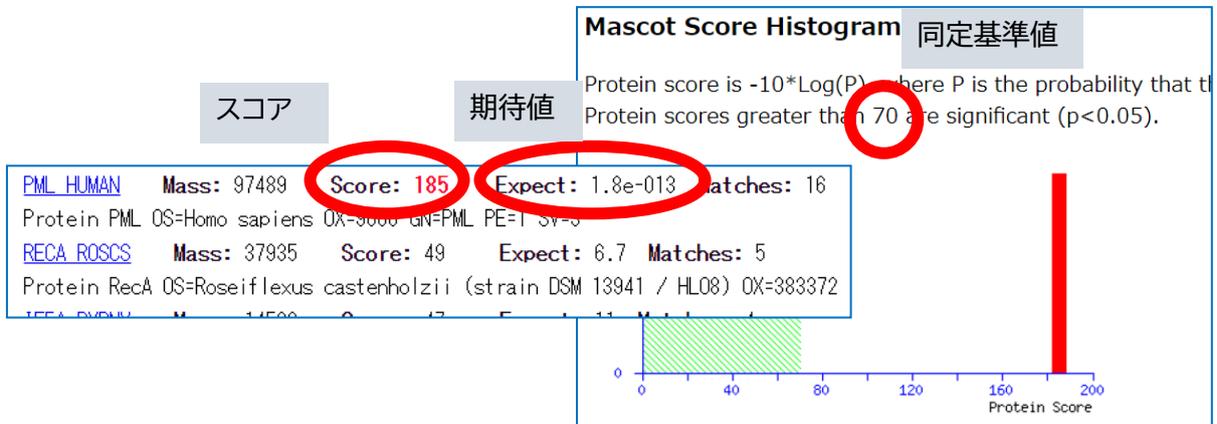
入力データ



マッピング

## 8-4. 同定タンパク質：マッチングとスコア、同定基準値、期待値

アルゴリズムは非公開ですが、基本的な考え方についてご説明いたします。



### スコア (S)

実測スペクトルと理論スペクトルとのマッチング度合いを表します。スコアが高いほど両スペクトルが良くマッチしています。MASCOT では Probability based scoring、確率論ベースのスコアリングです。実測スペクトルと理論スペクトルのマッチングがランダムな事象である確率を Pro とした場合、スコアは  $-10 \log_{10}(\text{Pro})$  と表すことができます。

### 同定基準値(St)

検索毎に計算される同定基準。同定基準は検索エンジンが算出した信頼度 95%を満たすスコアがデフォルト設定では提示され、主に検索対象となったデータベースのエントリー数を基に算出されます(データベースのエントリー数が少ない場合など一部例外のケースもあります)。スコア1位の結果でも同定基準値を超えない限りそのタンパク質が同定されたとはみなしません。

### 期待値 (Expect)

検索対象のデータベース中に、同様のランダムマッチをする事が期待されるタンパク質の数。デフォルト設定では、期待値が 0.05 より小さい時同定としています。

### スコアと同定基準値、期待値の関係

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

MASCOT の検索アルゴリズムの基となっているのは Mowse です。詳細は以下のページからご覧下さい。

<https://www.matrixscience.com/help/history.html>

[https://www.matrixscience.com/help/scoring\\_help.html](https://www.matrixscience.com/help/scoring_help.html)

## 8-5. protein inference : ユニークシェア ペプチド、タンパク質のグループ化

データベースに登録されているタンパク質は互いに類似の配列を持っていることがあり、切断ペプチド単位でも同じペプチド配列、同じ質量になる事があります。検索結果はペプチド質量ピークの組み合わせから同定の判定を行います。質量分析装置のデータだけではこれら類似配列のタンパク質を見分ける事が難しかったり、不可能であったりするケースがあります。

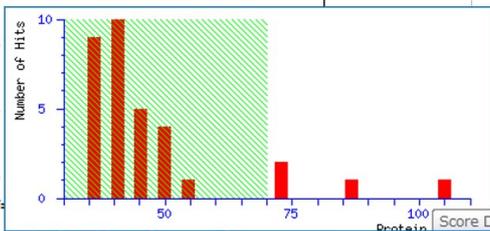
結果解釈の際 MASCOT Server がこれらのデータをどのように扱っているのか、ユーザー側が注意すべき点はどこか、についてご紹介します。

ある PMF 検索の結果について、13 のピーク(入力データ側)がデータベースに登録されている複数のタンパク質を候補として列挙しましたが、そのうち4つのタンパク質にフォーカスを当て、どのようにマッチしているかを表したのが以下の表です。「■」がマッチしていることを表し、「■」はピックアップした4つのタンパク質のうち1種類のタンパク質でしかマッチしていない事を表します。このようなペプチドはよく「ユニークペプチド」とこの分野では表現します。一方複数のタンパク質にマッチしている場合、それを「シェアペプチド」と表現します。

ピーク	OPSD_HUMAN 	OPSD_PHOVI 	OPSD_MACFA 	OPSD_CRIGR 
832.662	■	■	■	■
903.342	■	■	■	■
1186.439	■		■	■
1403.722	■		■	
1617.857	■		■	■
1727.916	■	■		
1743.951			■	
1759.966				■
1788.721				
1818.963	■	■		■
2159.143	■	■		
2174.812			■	
2256.871	■		■	■

この4つのタンパク質を含む検索結果画面が次頁図です。

1.	<b>OPSD_HUMAN</b>	Mass: 38866	Score: 105	Expect: 1.6e-05	Matches: 9
	Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1				
	<b>OPSD_PHOVI</b>	Mass: 38947	Score: 45	Expect: 16	Matches: 5
	Rhodopsin OS=Phoca vitulina GN=RHO PE=1 SV=1				
	<b>OPSD_MOUSE</b>	Mass: 39002	Score: 43	Expect: 28	Matches: 5
	Rhodopsin OS=Mus musculus GN=Rho PE=1 SV=1				
	<b>OPSD_SHEEP</b>	Mass: 38866	Score: 43	Expect: 28	Matches: 5
	Rhodopsin OS=Ovis aries GN=RHO PE=1 SV=2				
2.	<b>OPSD_MACFA</b>	Mass: 39036	Score: 88	Expect: 0.00075	Matches: 8
	Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1				
	<b>LUXS2_LACDB</b>	Mass: 17500	Score: 45	Expect: 17	Matches: 4
	S-ribosylhomocysteine lyase 2 OS=Lactobacillus delbrueckii subsp. bulgari				
3.	<b>OPSD_CRIGR</b>	Mass: 39071	Score: 73	Expect: 0.027	Matches: 7
	Rhodopsin OS=Cricetulus griseus GN=RHO PE=1 SV=1				
	<b>OPSD_CANFA</b>	Mass: 38936	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Canis familiaris GN=RHO PE=1 SV=1				
	<b>OPSD_FELCA</b>	Mass: 39023	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Felis catus GN=RHO PE=1 SV=1				
	<b>OPSD_RABIT</b>	Mass: 38968	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Orctolagus cuniculus GN=RHO PE=1 SV=1				
	<b>OPSD_SMICR</b>	Mass: 39037	Score: 56	Expect: 1.2	Matches: 6
	Rhodopsin OS=Orctolagus cuniculus GN=RHO PE=1 SV=1				



○である OPSD\_HUMAN と ○の OPSD\_PHOVI は 1 位のグループとしてまとめられています。一方 ○の OPSD\_MACFA と OPSD\_CRIGR は 2 位、3 位として別に報告されています。

前頁の表と上図の結果を見比べると、まとめられ方の理由がわかります。ポイントはユニークペプチドの有無です。OPSD\_PHOVI は OPSD\_HUMAN が持っていないペプチドが存在しません。OPSD\_PHOVI がはっきりと存在する理由がないため、OPSD\_HUMAN のグループに属して表示されます。MASCOT では、マッチしたペプチドの組み合わせが全く同じものを **same-set**, 包含関係にあるものを **sub-set** と呼びます。same-set や sub-set のタンパク質はグループとしてまとめられ、基本的にはグループの中で最もスコアが高い、あるいはデータベースに含まれている順番の関係で選ばれた代表タンパク質のみが考慮の対象となります。OPSD\_PHOVI は OPSD\_HUMAN の subset と認定されています。

一方○である OPSD\_MACFA と OPSD\_CRIGR は、表の「■」が示すように、○である OPSD\_HUMAN にはマッチしていなかったピークとマッチしています。PMF 検索ではこのような場合、異なるタンパク質がそれぞれ含まれている可能性を考慮し、異なるグループとして認識され、順位も別に表示されます。実際に、**OPSD\_HUMAN の他に OPSD\_MACFA や OPSD\_CRIGR が本当に別に含まれていたかについてはケースバイケースで、確率はあまり高くないと言えます。**「■」のピークの部分がランダムマッチである可能性は低くなく、否定できないためです。ユニークなペプチドマッチの扱いについて、MIS 検索の場合は意味合いがより重くなります。1つ1つのペプチドについてプロダクトイオンマススペクトルとのマッチングを行い確からしさを評価しているため、同定基準をクリアしているユニークなペプチドが存在は重要性が増し、それぞれのタンパク質が同定できたと考える事ができます。

このように、PMF では類似タンパク質の同定について難しい点を含んでおり、それを回避するためには MALDI の測定でポイントとなるペプチドを取り出して MS2 データを測定して MIS を行ったり、最初からショットガンなど MIS 検索を実施する必要があると言えます。

## 9. MIS : ペプチド同定とタンパク質同定

### 9-1. MIS ペプチド同定とタンパク質同定のまとめ

この章では MIS でのペプチド同定とタンパク質同定について説明いたします。重要な点を列挙すると以下の通りです。

- **query** 毎にペプチドに対する検索が行われる [9-3]
- 入力データはピーク作成時とマッチング時の2段階で選別・組み換えされる [9-2]
- スコアとは理論値と実測値のマッチング度合いを評価した数値で、高いほどよくマッチしている [9-4]
- スコアが同定基準値を超えた時ペプチド配列の同定とみなす (デフォルト設定値の信頼度は **95%**) [9-4]
- **FDR** を適用して信頼度にあたる数字を **99%**まで引き上げる事が多い [9-4, 11 章]
- 同定ペプチドが見つかったタンパク質の情報から同定タンパク質をリストアップする [9-5]
- タンパク質間でシェアされているペプチド情報に十分注意して結果を解釈する [9-6]
- 検索対象のデータベースの選択も結果に大きく影響する [9-3]

以降、順に説明します。

### 9-2. 入力データの調整

#### 9-2-1. MS2 スペクトル、ピークリストの再調整

入力データは **2 段階で調整**されます。**3-1-2, 3-1-3** でご案内したように、MIS のデータではスペクトルデータをそのまま受け付けておらず検索の前段階で**フラグメントの質量を反映したピークを抽出しノイズをカット**したものを入力データとして受け付けており、これが **1 段階目の調整**です。MASCOT Server 側では受け取ったデータに対してさらに**強度情報に基づいて入力データからサブセットデータを作成し、**

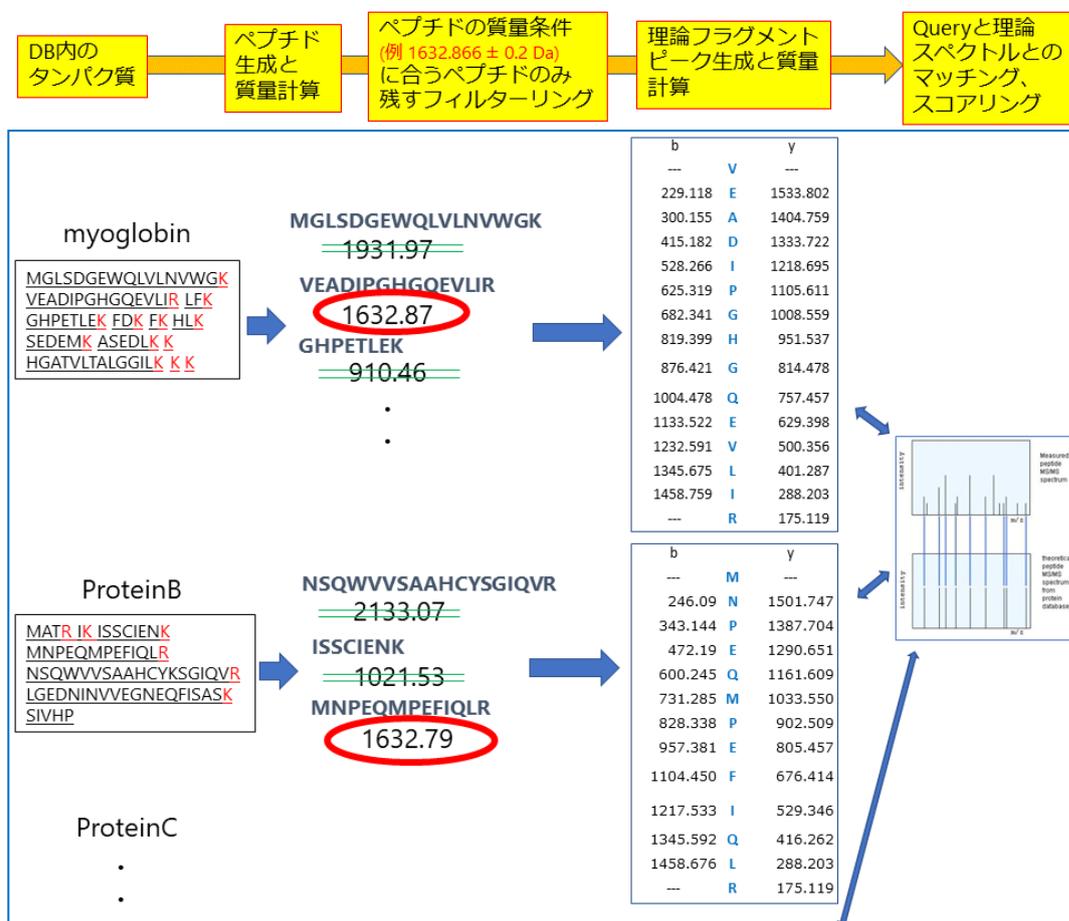
それぞれのサブセットに対して理論値とのマッチングを行って最もスコアが高くなった結果を採用しており、これが **2段階目の調整**に当たります。このように MASCOT Server プログラム側でも 2段階目の処理を行う事ができますが、あくまでも **1段階目の前処理としてピーク抽出が行われていることを前提**としています。また、上記のようにサブセットの選別に強度情報を使っていますが、**マッチングの評価そのものに関してはピーク強度の情報を利用していません**。強度の高いピークに対してマッチされた場合、よりスコアが高くなる、というような判断はありません。

## 9-2-2. ファイル内のクエリー数

PMF では入力データとして1つのスペクトルデータを受け付けていました。一方 MIS では **3-1-3** でご紹介しているように、BEGIN IONS からはじまり END IONS で終わる領域を1つの query としますが、**1ファイルの中に複数の query を含めてそれらを同時に検索する事ができます**。各 query のマッチングやスコアリングは他の query の影響を受ける事はなく、**1つずつばらばらに query を検索してもまとめて検索しても同定ペプチドの評価(スコア、同定基準値)は変わりません**。一方 **9-5** で説明しますが同定タンパク質は同定ペプチドの情報を組み合わせてレポートされるので、入力データの状況によって同定タンパク質の内容は変わってきます。

基本的には**1サンプルから由来する query はすべてまとめて検索をする事をお勧め**しています。そうする事で、**レポートされる同定タンパク質が「サンプルに含まれるタンパク質のリスト」と**なります。

## 9-3. ペプチド配列から行う理論値計算、ペプチドのフィルターリング



フィルターリングと理論値生成、入力データとのマッチングまでの流れを示したのが前頁図です。入力データとマッチングを行う参照側はペプチド配列から計算されたフラグメントの理論的なピークリストです(データベースが AA,NA の場合)。検索時に指定したデータベースに登録されているタンパク質1つ1つに対して、パラメーターをもとに理論ペプチドを作成しその質量を計算します。生成したペプチドについて質量を計算し、**query** と指定誤差範囲内でマッチしたペプチドのみを残す作業を最初に行います(フィルターリング)。残ったペプチドのみ、理論フラグメントのピークを生成し、入力データ **MS2** のピークリスト部分とマッチングを行ってその結果をスコアリングします。

この検索方法では測定データのペプチドがデータベース中に含まれている事が前提となっています。データベースに含まれないタンパク質・ペプチドからは当然理論値も作成されず、マッチング・同定タンパク質の検出も行われません。しかし 9-4 で説明するように、登録エントリー数が多いデータベースで検索をすると同定基準値が高くなり同定には不利です。**データベースの選択はエントリー数が多すぎず少なすぎない、必要十分なエントリーが含まれるデータベース**である必要があります。

## 9-4. 同定ペプチド：マッチングとスコア、同定基準値、期待値、外挿的な評価

以下のスコア・同定基準値・期待値・外挿的な評価、についてはペプチドに対してのものです。タンパク質の評価については 9-5 をご覧ください。

Locus: 3.645.3		同定基準値			
Score >	33	indicates identity			
Score >	23	indicates homology			
0.4100	0	63	6.2e-06	1	R.DFIDYYLIK.Q
-0.5813	0				DFPETNNILK
0.3187	1	スコア	期待値		TPPIIHRDLK
-0.7060	1	5	3.5 4		KETMALILK

### スコア (S)

実測スペクトルと理論スペクトルとのマッチング度合いを表します。**スコアが高いほど両スペクトルが良くマッチ**しています。MASCOT のスコアリングは確率論に基づいています。実測スペクトルと理論スペクトルのマッチングがランダムな事象である確率を Pro とした場合、スコアは  $-10\log_{10}(\text{Pro})$  と表す事ができます。

### 同定基準値(St)

query 毎に計算される同定基準です。**2種類**あり、それぞれ **Identity threshold** と **Homology threshold** という名称ですが、配列の同一性や類似性などは名称と関係がありません。両者は常に Identity threshold スコア  $\geq$  Homology threshold スコア という関係性があり、また Homology threshold は存在しない事があります。**現在の MASCOT の結果画面では基本的に Homology threshold を同定基準値とし、以下に説明する期待値の計算などにも利用**しています。

**Homology threshold** がない場合は **Identity threshold** が同定基準値となります。スコアが1位の結果でも、同定基準値を超えない限りそのペプチド配列が同定されたとはみなしません。

## 期待値 (Expect)

検索対象のデータベース中に、同様のマッチングがランダムで起こった場合に見つかるであろうペプチド数。検索エンジンのデフォルト設定では期待値が **0.05** より小さい時同定とします(FDR を同定基準として適用した場合を除く)。スコアと同定基準値、期待値には以下の関係が成り立ちます。

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

## 外挿的な評価 : FDR

当初 MASCOT をはじめとする各種検索エンジンは各々の同定基準、MASCOT でいえば有意基準を超える信頼度 95%に該当する同定基準値のみをレポートに提示していました。しかし論文に提出される検索結果について疑念を持たれるケースなどが発生する中で、**各検索エンジンの同定基準について別の観点から評価をする事が好ましい**と考えられるようになりました。また後述する同定ペプチドと同定タンパク質の関係性から、**信頼度に該当する数値は 95%でなく 99%の方が好ましいと考えられるようになりました**。これらの事を踏まえ外挿的に評価する FDR という数字を各検索結果で算出し、それが **1%**となるように同定基準を調整するようになり、現在はその方法が主流になっています(但し MASCOT で提示される同定基準はデフォルト設定の場合信頼度 95%のまま)。**各 query の同定基準値は、FDR が 1%となるような期待値になるよう調整され使用されます**。検索時、パラメーター「Decoy」にチェックを入れ、「Target PSM FDR」の数字を 1%とする事で実現可能です。**FDR については 11 章で詳しく説明します**。

## 9-5. 同定ペプチドから導き出される同定タンパク質

### 9-5-1. 同定タンパク質=ユニークな同定ペプチドが1つ以上アサインされている

スペクトルデータと理論値のマッチングを行う際、理論値の生成元はあくまでもペプチド配列です。スコアや期待値は**測定データの基となるペプチドが、マッチング対象としているペプチド配列であることを示す確からしさ**を表していることとなります。

一方 MASCOT ではタンパク質のスコアも表示されます。タンパク質のスコアは、そのタンパク質にアサインされたペプチドのスコアをもとに算出されます。タンパク質のスコアが高い事は、信頼度が高くスコアも高いペプチドが多く存在する事を意味します。**しかし現在ではタンパク質のスコアをもとに同定タンパク質であるかどうかを判定する事はしていません**。

現在 MASCOT で採用されている同定タンパク質を判定する基準は、**同定基準を超え、他のタンパク質にはアサインされていないユニークなペプチドが1つ以上アサインされているかどうか**となります(sub-set,same-set を除く)。この基準は現在広く採用されていますが、同定タンパク質の内容に厳密性を持たせようとする問題があります(9-5-2 で説明します)。

## 9-5-2. 1 Hit wonders : 同定タンパク質の Sensitivity と Specificity

ペプチド FDR 1%を同定基準とした場合、ユニークな同定ペプチドが1つのみアサインされているタンパク質を 100 集めると、**False Positive**、すなわち本来不正解の配列であるにも関わらず同定基準を超えて同定ペプチドと判定されているペプチドが1つ見つかります。**そのペプチドがアサインされているタンパク質も False Positive** という事になります。**MASCOT がレポートする同定タンパク質の信頼度のレベルもこれに準じている事になります**。このようなタンパク質の存在を指して「1 Hit wonders」問題という事がありますが、より厳しい基準を求めるケースでは、**同定タンパク質を判定する際に最低限アサインされているペプチド数を 2 に引き上げる事で問題を回避できます**。False Positive がランダムマッチであると考え、2つの False Positive ペプチドが数多くあるデータベース内のタンパク質に対して同時にアサインされることはほとんどない、というのが根拠です(ランダムマッチによるアサインペプチド数をポワソン分布で考慮)。

こういった基準については定まった厳密なルールが準備されておらず、状況に応じてユーザーが使い分ける事になります。

## 9-6. protein inference : ユニークシェア ペプチド、タンパク質のグループ化

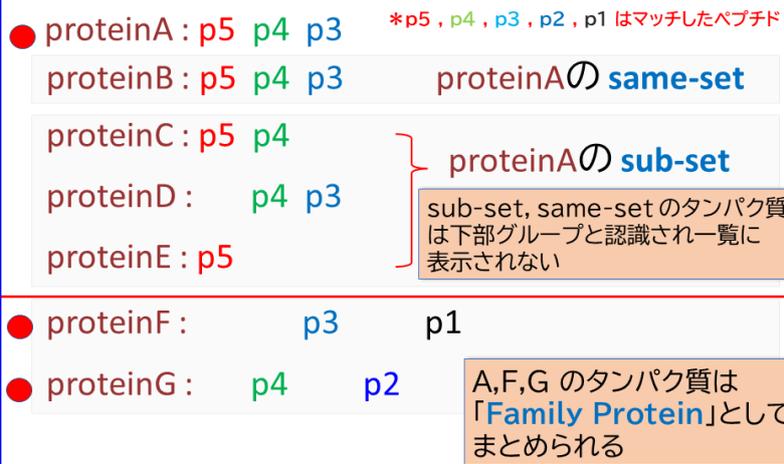
データベースに登録されているタンパク質は互いに類似の配列を持っていることがあり、測定結果において複数エントリーで完全に同一である配列部分の切断ペプチドが同定される事もしばしばあります。**質量分析装置のデータだけではこれら類似配列のタンパク質を見分ける事が難しかったり、不可能であったりするケースがあります**。結果解釈の際 MASCOT Server がこれらのデータをどのように扱っているのか、ユーザー側が注意すべき点はどこか、についてご説明します。

**同定ペプチドが複数タンパク質にシェアされている場合を「シェアペプチド」、唯一のタンパク質にて見つかった場合を「ユニークペプチド」と呼びます**。MASCOT においてシェアペプチドとユニークペプチドの状況により同定タンパク質がどのように扱われるかを説明するために、7つのタンパク質 **A~G**、5つのペプチド p1~p5 を使って同定タンパク質のグループ分けをします。5つのペプチドが各タンパク質にアサインされている状況を「■」(または「■」)で表したのが以下の表示です。

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

またこれを、タンパク質 A を中心に、シェアされているペプチドの状況によって A 以外のタンパク質をどのようにみなすのかを記したのが次頁の図です。

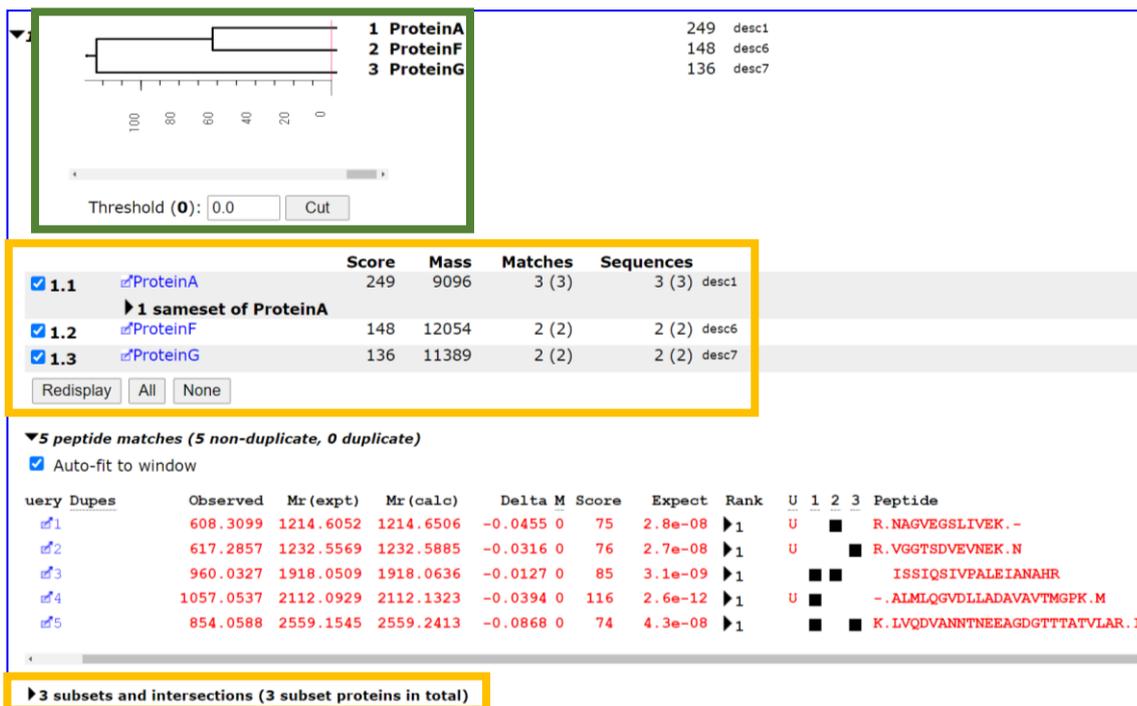
## シェアペプチド、ユニークペプチド

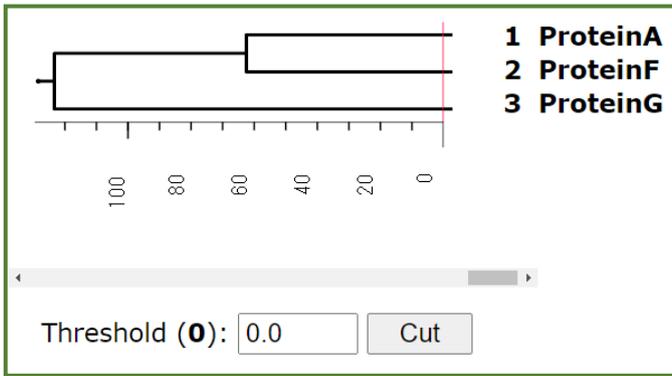


proteinB のように、proteinA と全く同じ組み合わせのペプチドが同定されているケースを **same-set protein**、proteinC,D,E のように proteinA の一部のペプチドが同定されていて包含関係的に proteinA の方が上位、かつ proteinA に含まれないペプチドがないケースについて、**sub-set protein** とそれぞれ呼んでいます。**same-set** や **sub-set** については代表タンパク質が1つのみが表示され、それ以外のタンパク質は最初の結果表示では現れません。

一方、proteinF や proteinG のように、proteinA にアサインされているペプチドと共有するペプチド(シェアペプチド)があるものの、**proteinA** にはアサインされていないユニークペプチドもある場合、**proteinA** とは異なる同定タンパク質としてリストに表示されますが、同時にシェアペプチドを含む分だけ配列が似ているという事で「**Family Protein**」として、まとめられて表示されます。またペプチドがシェアされている状況をもとに類似度を表す樹形図のようなクラスター表示も併せて表示されます。

下図は検索結果です。protein A~G まですべて含むデータベースで、ペプチド p1~p5 が同定されるようなデータで検索しましたが、結果の初期表示には proteinA,F,G の3種類のみが表示されます。





クラスター表示は本来であれば各タンパク質の配列相同性を専用アルゴリズムで計算した結果を使用するのが好ましいですが、計算の簡易化のため **MASCOT** では互いのユニークペプチドのスコアをもとに類似度を算出しています。そのため本来ファミリータンパク質に属するような配列の類似度をもつタンパク質がグループ化されなかったり、逆にシェアされた数少ない

ペプチドの存在の影響により、類似度が低いタンパク質がファミリータンパク質としてまとめられる事があります。また配列の相同性とは異なる状況で類似度表示されている事もあります。

**same-set** や **sub-set** のタンパク質はデフォルト表示では確認できませんが、結果画面内から展開表示させるとその内容を確認する事ができます。前ページのオレンジで囲った部分の三角の部分、展開ボタンをクリックすると、下図のように該当タンパク質を表示させることができます。ProteinA と same set である ProteinB, subset である proteinC, proteinD, proteinE も表示されます。

	Score	Mass	Matches	Sequences
<input checked="" type="checkbox"/> 1.1 <a href="#">ProteinA</a>	249	9096	3 (3)	3 (3) desc1
▼ 1 <b>same set of ProteinA</b>				
<input checked="" type="checkbox"/> <a href="#">ProteinB</a>	249	9855	3 (3)	3 (3) desc2
<input checked="" type="checkbox"/> 1.2 <a href="#">ProteinF</a>	148	12054	2 (2)	2 (2) desc6
<input checked="" type="checkbox"/> 1.3 <a href="#">ProteinG</a>	136	11389	2 (2)	2 (2) desc7

Redisplay All None

▼ 3 subsets and intersections (3 subset proteins in total)				
	Score	Mass	Subset of	
<input checked="" type="checkbox"/> <a href="#">ProteinC</a>	177	9288	1.1	desc3
<input checked="" type="checkbox"/> <a href="#">ProteinD</a>	146	9947	1.1	desc4
<input checked="" type="checkbox"/> <a href="#">ProteinE</a>	116	11608	1.1	desc5

CSVなどでファイル出力させる場合は、same-set のタンパク質も出力させるか選択する事ができます。また sub-set についても出力のオプションがあり、共有されるペプチドの度合いの情報を使って出力タンパク質の調整が可能です。

## 10. MASCOT 検索のオプション [MIS]

9 章までは汎用的に使用されるペプチド同定・タンパク質同定を中心とした説明でした。一方で MASCOT では通常の検索とは少し異なる、いくつかの検索オプションが存在します。この章では MIS で使用可能な 4 つの検索オプションについて説明いたします。なおこれらのオプションは PMF ではすべて適用できません。

- 10-1. Spectral Library : ピークライブラリに対する検索
- 10-2. Quantitation : 定量解析
- 10-3. crosslink : クロスリンクペプチド検索
- 10-4. Error Tolerant Search : 拡張2段階検索

### 10-1.Spectral Library

#### 10-1-1. Spectral Library 概要

従来プロテオミクス・DDA の解析では入力データの参照先データとして、配列から計算された理論値が使用されてきました。一方解析データの蓄積により、**検索に使用したピークリストデータと正解とみなされるペプチド配列の組み合わせに関する情報が蓄積し、それを検索に利用する試みが行われるようになりました。**

MASCOT でも Spectral Library(ピークリスト情報と正解のペプチド配列情報を含むデータベース)に対する検索を行う事ができます。Spectral Library 単独または FASTA データベースと組み合わせて検索が可能です。検索エンジンには NIST Mass Spectrometry Data center の MS PepSearch (<https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:mspepsearch>)を利用しています。また検索対象の Spectral Library には、インターネット上で公開されているデータベースと、ご自身の検索結果から作成したものの2種類を利用する事ができます。MS PepSearch 検索時には以下のような引数で実行しています。

\* 下記は本来一行で実行するコマンドです。

```
MSPepSearch.exe m a P /ZPPM 100 /M 0.509902 /LIB [ライブラリへのパス] /INP [MGF ファイルへのパス] /OUTTAB [出力ファイルへのパス] /HITS 10 /MinMF 0 /NumCompared /OutPrecursorMz /OutDeltaPrecursorMz /OutSpecNum
```

検索結果例として、以下の WEB ページをご覧ください。

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=./data/F981140.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=./data/F981140.dat)

ピークリストライブラリは右図のような msp というフォーマットで MASCOT Server 上に格納されています。ライブラリには、インターネット上で公開されているものを利用する方法と、ご自身の MASCOT Server で行った検索結果をライブラリ化してセットして利用する方法があります。

```
Name: SIPAYLAETLYYAMK/3
MW: 2021.0820791015626
Comment: Spec=Consensus Mods=2/0, ^, iTRAQ4plex/
Parent=674.702 Nreps=3 Naa=15 MaxRatio=1.000 P
DeltaMass=0.00 ClusterId=5258eb56-ee94-44f7-8a
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
```

## 10-1-2. Spectral Library 検索方法

MASCOT Server で Spectral Library に対して検索を行うためには、**データベースの選択時に Spectral Library を選んで検索するだけ**です(下図)。Spectral Library はデータベース一覧の中で「**Spectral Library(SL)**」と記された箇所の下にまとめて存在します。

The screenshot shows the MASCOT MS/MS Ions Search interface. It includes fields for 'Your name', 'Email', and 'Search title'. The 'Database(s)' dropdown is currently set to 'contaminants (AA)'. A list of available databases is displayed, with 'Spectral library (SL)' and 'PRIDE\_Human' highlighted with red boxes. The 'Taxonomy' dropdown is set to 'All entries'.

Spectral Library 検索では **Modification, enzyme, missed cleavages, taxonomy, instrument, charge** などのパラメーターは指定しても無視されます。この検索では単純に入力データとライブラリとのマッチングを行うため、通常検索で理論値のバリエーションを変更させたり、検索対象を調整するようなパラメーターが使用する事ができないためです。

また現段階では Decoy データベースへの検索並びに FDR 計算を行う事ができません(**将来的に対応予定**です)。Error Tolerant 検索も行う事ができません。

### 10-1-3. Spectral Library の検索結果

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981140.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981140.dat)

の結果をもとに説明いたします。必要に応じて上記 WEB ページも併せてご参照ください。

#### ■ 結果画面の表示内容

**FASTA** で検索を行った時と概ね同じです。表示内容を調整する「Format」欄で、使用できない一部の項目が表示されません。タンパク質 Accession をクリックした際に表示される protein view、ペプチドの query 番号をクリックした際に表示される peptide view の表示内容も通常の検索と同じです。タンパク質の配列については Spectral Library 内にはありませんがタンパク質の ID 情報が含まれており、かつ Spectral Library 作成時に同じ Accession 系列の配列データベースを設定しているためそこから配列情報を取得し結果画面などで表示されます。

#### ■ Score と同定基準

検索エンジンとして使用している NIST MS PepSearch のスコアは 0(一致しない)から 999(完全一致)の範囲で表されます。ライブラリ検索のみを行った場合は特にこのスコアの補正などは行われません。**同定基準スコアは 300** となっています。期待値については、スコア  $s$  の時の期待値  $E(s)$  が以下の式から計算しています。

$$E(s) = 0.05 \cdot 10^{(300-s)/100}$$

#### ■ FASTA データベースと Spectral Library を一緒に検索した時

検索例として以下の WEB ページも併せてご覧ください。

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981141.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981141.dat)

この例のように Spectral Library 検索と FASTA 検索が統合された結果の場合、同定基準を使った期待値の調整が行われます。ライブラリと FASTA のマッチの期待値について、両方でマッチしているデータを使ってその平均と分散が同じになるように調整をします。例として query1451 の調整の様子を以下の図で表示します。**ライブラリだけの検索では同定基準スコアが 300 で期待値が 0.00012** でしたが、**FASTA と共に行った検索では同定基準が 493 に、期待値が 0.0044** になっています(両者のスコアは共に 561 です)。

#### ライブラリ単独での検索結果

1451	453.2757	904.5369	904.5381	-1.37	0	561	SL	0.00012	1	U K.AKEFGILK.K
1452	302.5196	904.5370	904.5380	-1.16	0	589	SL	6.4e-05		Score > 300 indicates identity
1720	466.2325	930.4505	930.4521	-1.81	0	772	SL	9.5e-07		

#### ライブラリ+FASTA での検索結果

1451	453.2757	904.5369	904.5381	-1.37	0	561	SL	0.0044	1	U K.AKEFGILK.K
1452	302.5196	904.5370	904.5380	-1.16	0	589	SL	0.0016	1	Mascot score > 20 indicates identity Library score > 493 indicates identity
1720	466.2325	930.4505	930.4521	-1.81	0	772	SL	2.2e-06	1	

## 10-1-4. Spectral Library をローカルの MASCOT Server にセットする方法

詳細は、データベース管理マニュアル(日本語版)

[https://www.matrixscience.co.jp/supportpdf/MASCOTServer\\_ver26\\_sequencedbmanage.pdf](https://www.matrixscience.co.jp/supportpdf/MASCOTServer_ver26_sequencedbmanage.pdf)

の P.52～ をご覧ください。

インターネット上で公開されているライブラリをセットする方法は P.53～、ご自身の MASCOT Server で行った検索結果をライブラリ化してセットする方法は P.83～ をご覧ください。

## 10-1-5. Spectral Library 補足説明資料へのリンク

### ■ 弊社インターネットサイトでのヘルプページ

[https://www.matrixscience.com/help/spectral\\_library.html](https://www.matrixscience.com/help/spectral_library.html)

### ■ MSPepSearch について

- Stein, S. and Scott, D. R. (1994). Optimization and testing of mass spectral library search algorithms for compound identification, *J. Am. Soc.Spectrom.*, 5, 859-66  
[https://dx.doi.org/10.1016/1044-0305\(94\)87009-8](https://dx.doi.org/10.1016/1044-0305(94)87009-8)

### ■ PRIDE の Spectral Library について

- Griss, J., Foster, J. M., Hermjakob, H., and Vizcaíno, J. A., PRIDE Cluster: building a consensus of proteomics data, *Nature Methods* 10, 95-96 (2013)  
<https://dx.doi.org/10.1038/nmeth.2343>
- Griss, J., et al., Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets, *Nature Methods* 13, 651-656 (2016)  
<https://dx.doi.org/10.1038/nmeth.3902>

### ■ Spectral Library の公開元サイト

- PeptideAtlas / ISB  
<http://www.peptideatlas.org/speclib/>
- NIST  
<http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:cdownload>
- PRIDE / EBI  
<https://www.ebi.ac.uk/pride/cluster/#/libraries>

## 10-2. Quantitation

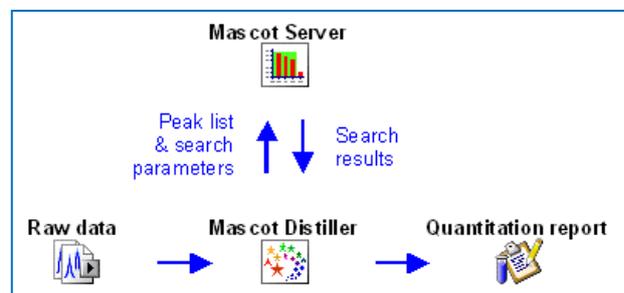
### 10-2-1. Quantitation 概要

質量分析データを用いたタンパク質の定量解析には様々なアプローチがあります。その中で**ピークリスト内の情報、または検索結果の情報だけを使う手法については MASCOT Server 単独で定量計算が可能です**。具体的には以下の手法です (手法名は MASCOT 内で使用している呼称)。

- **reporter** : MS2 スペクトル内のある領域のフラグメントピークの相対的な強度に基づいた定量
- **multiplex** : MS2 スペクトル内のイオンシリーズのフラグメントピークの相対的な強度に基づいた定量
- **emPAI** : タンパク質にアサインされたペプチドの情報に基づく定量指標。  
Spectral Counting の1種。

このうち emPAI については、MS/MS スペクトル数が 100 以上の時自動的に結果画面に表示されます。

**上記3手法以外**については raw データからプレカーサーペプチドに関する追加の情報を得る必要があり、**MASCOT Server 単独では定量計算ができません**。弊社ソフトウェアで言えば **MASCOT Distiller 定量モジュールがあると計算が可能です**(右図)。



検索結果例としては以下の WEB ページをご参照ください。

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981131.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981131.dat)

### 10-2-2. Quantitation 検索方法

検索パラメーター「**Quantitation**」で、予め設定しておいた定量手法の組み合わせを選択して検索を行うだけです(右図)。定量に使用する修飾なども Quantitation の各項目に紐づけられているため、それらをあえて Modification から指定する必要はありません。

デフォルト設定の名称については、名称の後ろに「**MD**」が付いている項目が MASCOT Distiller が必要な手法で、ついていないものが Server 単独で計算可能な手法です。

定量手法の組み合わせを自身でカスタマイズしたり新たに作成したりするのは専用の設定画面で行います。

**10-2-4,13-8** で詳しく説明いたします。

Enzyme	Trypsin
<b>Quantitation</b>	None
Crosslinking	None
Fixed modifications	ITRAQ 4plex ITRAQ 4plex (protein) ITRAQ 8plex TMT 6plex TMT 2plex DiLeu 4plex 18O multiplex SILAC K+6 R+6 multiplex IPTL (Succinyl and IMID) multiplex
Variable modifications	ICPL duplex pre-digest [MD] ICPL duplex post-digest [MD] ICPL triplex pre-digest [MD] ICPL quadruplex pre-digest [MD] 18O corrected [MD]
Peptide tol. ±	15N Metabolic [MD]
Peptide charge	15N + 13C Metabolic [MD] SILAC K+6 R+10 [MD] SILAC K+6 R+10 Arg-Pro [MD]
Data file	SILAC K+6 R+6 [MD]

### 10-2-3. Quantitation 検索結果

後ろに[MD]が付いていない Quantitation の項目を選択した状態で検索を行った際、結果画面では各ペプチドの定量値とそのペプチドがアサインされているタンパク質の定量値が表示されます(下図)。

Query	Dupes	Observed	Mr (expt)	Mr (calc)	ppm	M	Score	Expect	Rank	U	114/113	115/113	116/113	117/113	118/113	119/113	121/113	1	2	Pep
52618		517.3046	1032.5946	1032.5983	-3.52	0	22	0.026	1		0.909	1.155	0.945	-0.096	1.064	1.202	0.791	■	■	K.G
58592	1	533.2921	1064.5696	1064.5703	-0.64	0	30	0.0079	1		1.084	1.260	1.285	1.150	1.313	1.346	1.099	■	■	K.A
58697		533.7829	1065.5512	1065.5543	-2.90	0	24	0.012	1		0.973	1.420	0.857	0.963	0.734	0.820	0.850	■	■	K.A
62050	4	548.3186	1094.6226	1094.6238	-1.08	0	29	0.0047	1		0.864	1.101	0.889	0.795	1.075	0.821	0.879	■	■	K.L
70990	4	598.2969	1194.5792	1194.5792	0.033	0	17	0.024	1		0.959	1.120	0.871	0.952	1.154	0.969	0.880	■	■	R.V
79303		419.2471	1254.7195	1254.7140	4.38	0	29	0.015	1		1.006	0.954	1.220	0.793	1.234	0.837	0.896	■	■	R.E
81345	6	634.8576	1267.7006	1267.7014	-0.57	0	32	0.0014	1		1.171	1.311	1.458	0.999	1.210	0.912	0.953	■	■	R.G
81352	12	423.5749	1267.7029	1267.7014	1.18	0	35	0.00061	1		0.934	1.581	1.814	0.721	0.943	0.883	0.946	■	■	R.G
81457	27	635.3070	1268.5994	1268.5989	0.46	0	24	0.023	1		1.030	1.090	1.039	1.069	1.396	1.159	1.218	■	■	K.F
81483	24	423.8742	1268.6008	1268.5989	1.51	0	36	0.0022	1		1.015	0.846	1.267	0.816	1.345	1.500	0.890	■	■	K.F
82476		639.2738	1276.5330	1276.5339	-0.68	0	25	0.0043	1		0.856	1.140	1.163	0.848	1.046	0.802	1.121	■	■	R.C

下に拡大図①

下に拡大図②

拡大図①

es	emPAI	114/113	115/113	116/113	117/113	118/113	119/113	121/113
3)	48.75	1.033	1.070	1.045	1.016	1.155	1.051	1.055
2)	44.57	1.036	1.073	1.044	1.019	1.159	1.052	1.060

拡大図②

Expect	Rank	U	114/113	115/113	116/113	117/113	118/113	119/113	121/113	1	2	Peptide
0.026	1		0.909	1.155	0.945	-0.096	1.064	1.202	0.791	■	■	K.GQAGLQR.A
0.0079	1		1.084	1.260	1.285	1.150	1.313	1.346	1.099	■	■	K.AAANQMR.N
0.012	1		0.973	1.420	0.857	0.963	0.734	0.820	0.850	■	■	K.AAANQMR.N + D
0.0047	1		0.864	1.101	0.889	0.795	1.075	0.821	0.879	■	■	K.LTSLSDR.Y
0.024	1		0.959	1.120	0.871	0.952	1.154	0.969	0.880	■	■	R.VQQPDCR.E
0.015	1		1.006	0.954	1.220	0.793	1.234	0.837	0.896	■	■	R.EFHLHLR.L

アサインされている各ペプチドの定量値からタンパク質の定量値がどのように計算されるかについては、「Quantitation」パラメーターで選択された項目にデフォルト設定が定義されているほか、検索結果画面から再調整する事が可能です(下図、赤線で囲われた黄色の領域)。

Format	Significance threshold p<	0.05	Max. number of families	AUTO	[help]
	Target FDR (overrides sig. threshold)	(not set)	FDR type	PSM	
	Display non-sig. matches	<input type="checkbox"/>	Min. number of sig. unique sequences	1	
	Show Percolator scores	<input type="checkbox"/>	Dendrograms cut at	0	
	Preferred taxonomy	All entries			
	Protein ratio type	Median	Normalise to	None	[help]
	Min. precursor charge	1	<input type="radio"/> of all peptides		
	Min. # peptides	2	<input type="radio"/> of peptides assigned to accession(s)		
	Unique peptides only	<input type="checkbox"/>	<input type="radio"/> of peptide sequence(s)		
	Outlier removal	Automatic			
	Peptide threshold	At least homology	0.05		

パラメーターで設定できる事項については以下の通りです。

- Protein ration type** : ペプチドの ratio からタンパク質の ratio をどのように計算するか指定
- **Average** : ペプチド ratio の幾何平均
  - **Median** : ペプチド ratio のメジアン
  - **Summed intensities** : (reporter プロトコルのみ)ピーク強度の和
- Min. precursor charge** : ペプチド定量値の計算対象とするペプチド電荷の最小値
- Min. # peptides** : タンパク質定量値の計算対象とする、最低限必要なアサインペプチド数
- Unique peptides only** : ユニークペプチドのみを計算に利用するかどうか
- Outlier removal** : 外れ値を除く方法の指定、以下から選択
- **None** : 外れ値を除く処置を行わない
  - **Automatic** : データ数によって Dixon 法または Rosner 法のどちらかを選択
  - **Dixon's method** : Dixon 法
  - **Grubbs' method** : Grubbs 法
  - **Rosner's method** : Rosner 法
- Peptide threshold** : 定量計算対象とするペプチドについて、スコアや期待値、同定基準値を使ったフィルターリング
- Normalise to** : データの Normalisation 実施。サンプル間比較で、サンプルに属するペプチドの ratio を使って調整
- **None** : normalisation を実施しない
  - **Average ratio** : ペプチドの ratio について幾何平均が同じになるようにする
  - **Median ratio** : ペプチドの ratio についてメジアンが同じになるようにする
  - **Summed intensities** : (reporter プロトコルのみ) MS/MS スペクトルのレポーターイオンの強度の和が同じになるようにする

さらに上記 Normalisation について計算の対象を特定する事が可能です。各サンプルで等量含まれている事が確定している内容を選択可能です。選択肢はすべてのペプチド、配列が特定されたペプチド(複数指定可)、Accession で特定されたタンパク質(複数指定可)にアサインされているすべてのペプチド、です。

タンパク質の定量値をまとめた情報をファイル出力したい場合、**Report Builder** をご利用ください。

結果画面の Report Builder タブで「Columns」を展開し、出力内容で定量値を選択の上「**Apply**」ボタンを押します(右図)。

The screenshot shows the 'Report Builder' tab in a software interface. At the top, there are three tabs: 'Proteins (545)', 'Report Builder', and 'Unassigned (140931)'. Below the tabs, the main heading is 'Protein family members (545 proteins)'. Underneath, it says 'Columns: Standard 12 out of 58'. There are two columns of options: 'Enabled' and 'Available'. The 'Enabled' column lists various fields like Family, Member, Database, Accession, Score, Mass, and Description. The 'Available' column lists statistical fields like SD(geo), p-value, Number of peptides, Significant, Not-normal, and emPAI. At the bottom right, there is an 'Apply' button.

Proteins (545) Report Builder Unassigned (140931) S\_permai

**Protein family members (545 proteins)**

▶ Columns (19 out of 58)

▶ Filters: (none)

Export as CSV

#	Family	M	DB	Accession	Score	Mass	114/113	115/113	116/113	117/113	118/113	119/113	121/113	Matches	Match(sig)
1	1	SwissProt	2::CO4B_HUMAN	164342	217600	1.033	1.070	1.045	1.016	1.155	1.051	1.055	3818	3818	
1	2	SwissProt	2::CO4A_HUMAN	163856	217680	1.036	1.073	1.044	1.019	1.159	1.052	1.060	3814	3814	
2	1	SwissProt	2::APOB_HUMAN	127385	624988	1.082	1.362	0.827	1.203	1.189	1.093	1.079	3897	3897	
3	1	SwissProt	2::CERU_HUMAN	59576	143199	0.884	1.080	0.711	1.047	1.283	1.027	1.027	1466	1466	
4	1	SwissProt	2::A1BG_HUMAN	58870	58330	0.949	1.201	0.994	1.124	1.181	1.041	1.085	1527	1527	
5	1	SwissProt	2::HEMO_HUMAN	44576	58934	0.970	1.161	0.940	1.086	1.334	1.053	1.103	1899	1899	

Apply ボタンを押すと Report Builder の表示内容が変わります。ファイル出力は表示内容と同じ順番、データが出力されるのでこの段階で好みの様式に調整してください。調整後「Export CSV」ボタンを押すとファイル出力が行われます。

### 10-2-4. Quantitation 設定の作成

Quantitation で選択する設定項目は、予め MASCOT で作成済みである設定を使う場合でも、ピークシフトなどの設定や修飾と定量計算に使用するかなどの内容についてより正確なデータ解析を行うため微調整をした方が好ましいケースがあります。また事前に準備された設定内容とは大きく異なる方法で定量を行う場合、新たに設定を作成する事もできます。

設定画面は 13-8 でご紹介している「Quantitation」(Home -> Configuration Editor ->

Family	Member	Database	Accession	Score	Mass	114/113	115/113	116/113	117/113	118/113	119/113	121/113	Num. of m	Num. of si	Num. of se	Num. of si emPAI	Description		
31																	Normalisa none		
32																	Min. precu 1		
33																	Outlier renauto		
34																	Min. numb 2		
35																	Peptide th at least homology		
36																	Unique pejno		
37																			
38	Family	Member	Database	Accession	Score	Mass	114/113	115/113	116/113	117/113	118/113	119/113	121/113	Num. of m	Num. of si	Num. of se	Num. of si emPAI	Description	
39	1	1	SwissProt	CO4B_HU	164342	217600	1.033	1.07	1.045	1.016	1.155	1.051	1.055	3818	3818	103	103	48.75	Complement C4-B (
40	1	2	SwissProt	CO4A_HU	163856	217680	1.036	1.073	1.044	1.019	1.159	1.052	1.06	3814	3814	102	102	44.57	Complement C4-A (
41	2	1	SwissProt	APOB_HU	127385	624988	1.082	1.362	0.827	1.203	1.189	1.093	1.079	3897	3897	214	214	9.7	Apolipoprotein B-10
42	3	1	SwissProt	CERU_HU	59576	143199	0.884	1.08	0.711	1.047	1.283	1.027	1.027	1466	1466	50	50	14.97	Ceruloplasmin OS=
43	4	1	SwissProt	A1BG_HU	58870	58330	0.949	1.201	0.994	1.124	1.181	1.041	1.085	1527	1527	19	19	11.9	Alpha-1B-glycoprot
44	5	1	SwissProt	HEMO_HU	44576	58934	0.97	1.161	0.94	1.086	1.334	1.053	1.103	1899	1899	30	30	144.13	Hemopexin OS=Hor
45	6	1	SwissProt	CFAH_HU	37520	167416	0.962	1.13	0.872	1.103	1.306	1.096	1.132	1521	1521	65	65	21.74	Complement factor
46	6	2	SwissProt	FHR2_HU	1329	36538	0.858	1.123	0.687	1.132	1.406	1.065	1.071	82	82	10	10	4.94	Complement factor
47	6	3	SwissProt	FHR1_HU	1289	43717	0.903	1.191	0.816	1.152	1.356	1.068	1.112	80	80	13	13	7.39	Complement factor
48	6	4	SwissProt	FHR5_HU	699	77592	0.896	1.249	0.941	1.018	1.496	1.075	1.116	35	35	7	7	0.52	Complement factor
49	7	1	SwissProt	ITIH2_HU	35418	126842	1.025	1.183	0.915	1.233	1.264	1.103	1.099	1126	1126	42	42	16.62	Inter-alpha-trypsin i
50	8	1	SwissProt	FETUA_HU	35203	45131	0.907	1.074	0.652	1.071	1.233	0.928	1.019	753	753	15	15	8.63	Alpha-2-HS-glycopr
51	9	1	SwissProt	CFAB_HU	31648	103009	0.95	1.164	0.815	1.115	1.362	1.09	1.086	843	843	43	43	21.75	Complement factor
52	9	2	contamina	718067.1	3489	102935	0.911	1.08	0.762	1.027	1.424	1.001	1.078	95	95	5	5	0.31	(Bos taurus) Compl
53	10	1	SwissProt	APOH_HU	30084	48761	0.936	1.118	0.707	1.168	1.338	0.972	1.002	751	751	21	21	27.28	Beta-2-glycoprotein

Quantitation」で行いますのでそちらをご参照ください。

### 10-2-5. Quantitation 補足説明へのリンク

■ MASCOT の Quantitation 全般に関する HELP ページ

[https://www.matrixscience.com/help/quant\\_overview\\_help.html](https://www.matrixscience.com/help/quant_overview_help.html)

■ **MASCOT の Quantitation 結果画面表示内容に関する HELP ページ**

[https://www.matrixscience.com/help/quant\\_format\\_help.html](https://www.matrixscience.com/help/quant_format_help.html)

■ **ラベルフリー定量解析のチュートリアル [日本語]**

[https://www.matrixscience.co.jp/supportpdf/MascotDistiller\\_replicatesQuantTutorial.pdf](https://www.matrixscience.co.jp/supportpdf/MascotDistiller_replicatesQuantTutorial.pdf)

■ **SILAC 解析のチュートリアル [日本語]**

[https://www.matrixscience.co.jp/supportpdf/MascotDistiller\\_Quantitative\\_quick\\_start.pdf](https://www.matrixscience.co.jp/supportpdf/MascotDistiller_Quantitative_quick_start.pdf)

■ **Reporter プロトコルの解析結果例(iTRAQ)**

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981131.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981131.dat)

■ **Multiplex プロトコルの解析結果例 (SILAC K+6, R+6 multiplex)**

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F981133.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F981133.dat)

## 10-3. Crosslink

### 10-3-1. Crosslink 検索 概要

MASCOT では質量分析のデータからタンパク質の立体構造解析やタンパク質間の相互作用解析に応用される手法、ペプチドの **crosslink** 解析にも対応しています。

MASCOT で単一のペプチド同定を行うのとは異なり、候補ペプチドが N 個存在する crosslink 解析は  $N^2$  のデータ数が解析対象となり莫大な数が対象となってしまう事から、検索対象を慎重に選定せざるを得ません。検索前に **crosslink** 解析のターゲットとするタンパク質をかなり限定するとともに、リンカーの種類や結合パターンを考慮する範囲を特定した設定を予め作成しておきパッケージ化しておいて、検索時にその設定項目を選択します。これは「Quantitation」検索と同じ方式です。

予め設定しておくべき主な内容とは以下の通りです。

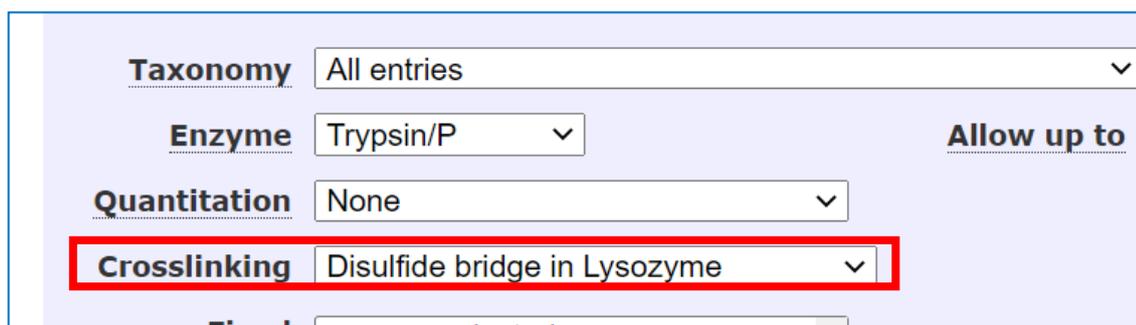
- ・ リンカーの組成、リンカーが付くアミノ酸残基
- ・ crosslink で使用するリンカー
- ・ リンカーの結合パターン（下記4種類）
- ・ リンカーが結合するタンパク質、またはデータベース

リンカーの結合パターンには以下の4種類があります。

- ・ **intralinks** : 同一タンパク質内の別ペプチド間の結合
- ・ **interlinks** : 異なる 2 種類のタンパク質のペプチド間の結合
- ・ **loolinks** : 同一ペプチド内での結合
- ・ **monolinks** : リンカーの一方にのみ結合した状態(単純な修飾)

## 10-3-2. Crosslink 検索を行う方法

MASCOT で Crosslink 検索を行いたい場合、予め設定をしておいた Crosslinking 項目を選んで検索を実行してください(下図)。



The screenshot shows a search configuration panel with several dropdown menus. The 'Crosslinking' dropdown is highlighted with a red border and contains the text 'Disulfide bridge in Lysozyme'. Other visible options include 'Taxonomy' (All entries), 'Enzyme' (Trypsin/P), and 'Quantitation' (None). The text 'Allow up to' is visible to the right of the 'Enzyme' dropdown.

Crosslinking 設定には、対象とするタンパク質(または非常に小規模のデータベース)も予め指定しておく必要があります、**実質解析の度に何かしらの項目を変更したものを使用しなければならない**と言えます。設定変更については、**10-3-5** で全体像に対する説明をしているほか、「**13-5. Linkers**」, 「**13-9. crosslinking**」の項目で設定画面についてもう少し詳しく説明しています。必要な場合はそちらも併せてご参照ください。

## 10-3-3. Crosslink 検索に際し注意すべき MASCOT 設定値

Crosslink 検索において、いくつかの MASCOT Server 設定値は2つのペプチドを合わせた数字が制限対象となります。そのため**通常の検索で問題にならないような設定値が、Crosslink 検索では不足し検索結果に悪影響を及ぼす事があります。**

関連する設定値は以下のようにまとめられます。

### ■ 両ペプチド合わせた数字が適用されるもの

- **MaxPepModArrangements** (Variable modification の組み合わせパターンについて考慮する組み合わせ数の最大値)

### ■ 両ペプチドのうち長い方の数字が適用されるもの

- **MinPepLenInSearch** (検索対象とするペプチド長さの下限)
- **MinPepLenInPepSummary** (結果画面表示対象とするペプチド長さの下限)

### ■ それぞれのペプチドに対して適用されるもの

- **missed cleavages** (検索パラメーター)
- **MaxPepNumVarMods** (1 ペプチドにおいて考慮する Variable modification(種類)数の最大値)
- **MaxPepNumModifiedSites** (1 ペプチドにおいて考慮する、Variable modification を考慮するアミノ酸残基数の最大値)

設定不可な内容として、MASCOT Server には Peptide mass の上限 16kDa という制限がありますが、これは2つのペプチドの合計値に対して適用されます。また検索パラメーターの中で、**Error tolerant**, **Decoy**, **Quantitation**, そして検索後のオプション **Percolator** については Crosslink 検索と同時に 行う事ができません。また設定値とは関係がありませんが、Crosslink 検索では各々の電荷が合算される 事が多く、MS2 フラグメントピークの電荷も多価になる事が多くなる一方、MASCOT Server では 2 価の フラグメントまでしか考慮できません。多価のフラグメントを 1 価に換算してピークリストに書き込む (弊社ではこれを **decharge** と呼んでいます)事が可能であれば、そのようなオプションをご利用ください。 弊社ソフトウェア **MASCOT Distiller** では **decharge** が可能ですので、ご興味ございましたら弊社まで ご連絡ください。

### 10-3-4. Crosslink 検索結果

以下の結果を使って説明いたします。より分かりやすい理解のためには、結果画面を開いてご参照 ください。ジスルフィド結合で、同一タンパク質内の結合を考慮した検索です。

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F002553.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F002553.dat)

結果の表示は通常の検索とほぼ同じです。Crosslink ペプチドのスコアや同定基準、期待値の算出に ついては通常検索と同様に行われます。

647.0693	2584.2480	2584.2712	-8.96	1	27	0.0028	▶1	U ■	K.VFGRCELAAAMK.R C5<-Xlink:Disulfide->C1 R.CKGTDVQAWIR.G + Oxidation (M)
904.7667	2711.2782	2711.2605	6.54	0	93	2.7e-07	▶1	U ■	R.NLCNIPCSALLSSDITASVNCAK.K + 3 Nethylmaleimide (C)
1356.6492	2711.2838	2711.2605	8.63	0	42	0.038	▶1	U ■	R.NLCNIPCSALLSSDITASVNCAK.K + 3 Nethylmaleimide (C)
1055.1664	3162.4774	3162.4602	5.42	1	114	2.1e-11	▶1	U ■	K.FESNFNTQATNRNEDGSTDYGILQINSR.W
791.6272	3162.4795	3162.4602	6.10	1	37	0.00041	▶1	U ■	K.FESNFNTQATNRNEDGSTDYGILQINSR.W
680.3362	3396.6445	3396.6244	5.91	1	73	1.3e-07	▶1	U ■	R.HGLDNYRGYSLGNWVCAAK.F C16<-Xlink:Disulfide->C1 R.CKGTDVQAWIR.G
850.1686	3396.6451	3396.6244	6.08	1	27	0.0031	▶1	U ■	R.HGLDNYRGYSLGNWVCAAK.F C16<-Xlink:Disulfide->C1 R.CKGTDVQAWIR.G
1174.2014	3519.5823	3519.5679	4.09	0	91	2.8e-09	▶1	U ■	R.NLCNIPCSALLSSDITASVNCAK.K C7<-Xlink:Disulfide->C3 R.WWCNDGR.T + 2 Nethylmaleimide (C)
880.9040	3519.5870	3519.5679	5.42	0	75	8.4e-08	▶1	U ■	R.NLCNIPCSALLSSDITASVNCAK.K C7<-Xlink:Disulfide->C3 R.WWCNDGR.T + 2 Nethylmaleimide (C)

結果のペプチド表示において、2 つのペプチドとそのペプチドがどの位置で結合しているかが示されて います。例えば **query 1496** の結果(上図の緑で囲われた部分、下図は緑で囲われた部分の拡大図)を 見ると、最初に表示されたペプチド[αペプチド]の 7 番目の C と、後に表示されたペプチド(βペプチド)の 3 番目の C がリンカー「Xlink:Disulfide」(ジスルフィド結合の定義)で結合している事を示しています。

```
R.NLCNIPCSALLSSDITASVNCAK.K C7<-Xlink:Disulfide->C3
R.WWCNDGR.T + 2 Nethylmaleimide (C)
```

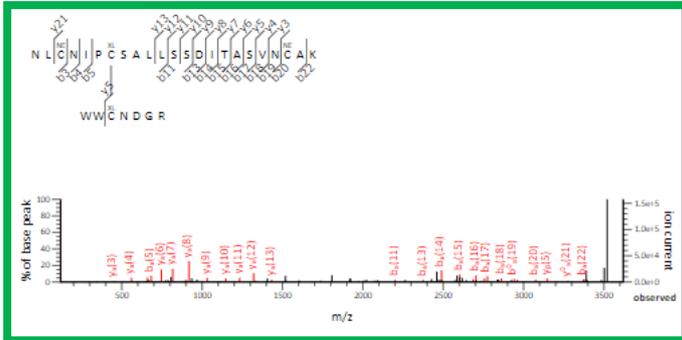
Query 番号をクリックするとマッチング内容を確認する peptide view ページを見る事ができます。 次頁図は query1496 の peptide view です。

**MASCOT SEARCH RESULTS**

**Peptide View**

MS/MS Fragmentation of **NLCNIPCSALLSSDITASVNC AK C7<-Xlink:Disulfide->C3 WWCNDGR**  
 Found in **LYSC\_CHICK** in **SwissProt**, Lysozyme C OS=Gallus gallus OX=9031 GN=LYZ PE=1 SV=1

Match to Query 1496: 3519.587056 from(880.904040,4+) intensity(11983704.0000) scans(4325) rawscans(sn4325) rtinseconds(2044.9118) index(1203)  
 Title: 1204: Scan 4325 (rt=2044.91) [C:\Users\Y\john\Downloads\Lysozyme\_2p2m\_1minute.raw]  
 Data file Lysozyme\_2p2m\_1minute.temp.mgf

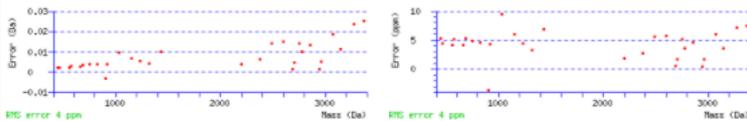


Label all possible matches  Label matches used for scoring

Monoisotopic mass of neutral linked peptide Mr (calc) : 3519.5879  
 Monoisotopic mass of neutral free alpha peptide Mr (calc) : 2586.2127  
 Monoisotopic mass of neutral free beta peptide Mr (calc) : 258.3706  
 Intact link : Alpha C7<-Xlink:Disulfide->C3 Beta  
 Variable modifications:  
 Alpha C3 : Methionine (C)  
 Alpha C21 : Methionine (C)  
 Ions Score: 75 Expect: 8.4e-08  
 Matches : 30/234 Fragment ions using 54 most intense peaks [calc]

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	115.0502	58.0287	98.0237	49.5155			N							23
2	228.1343	114.5708	211.1077	106.0575			L	3406.5323	1703.7698	3389.5058	1695.2565	3388.5217	1694.7645	22
3	<b>456.1911</b>	228.5992	439.1646	220.0859			C	3293.4482	1647.2278	3276.4217	1638.7145	<b>3275.4377</b>	1638.2225	21
4	<b>570.2341</b>	285.6207	553.2075	277.1074			N	3065.3914	1533.1993	3048.3648	1524.6861	3047.3808	1524.1940	20
5	<b>683.3181</b>	342.1627	666.2916	333.6494			I	2951.3484	1476.1779	2934.3219	1467.6646	2933.3379	1467.1726	19
6	780.3709	390.6891	763.3443	382.1758			P	2838.2644	1419.6358	2821.2378	1411.1226	2820.2538	1410.6305	18
7	1816.7352	908.8713	1799.7087	900.3580			C	2741.2116	1371.1094	2724.1851	1362.5962	2723.2011	1362.1042	17
8	1903.7673	952.3873	1886.7407	943.8740	1885.7567	943.3820	S	1704.8473	852.9273	1687.8207	844.4140	1686.8367	843.9220	16
9	1974.8044	987.9058	1957.7778	979.3925	1956.7938	978.9005	A	1617.8152	809.4113	1600.7887	800.8980	1599.8047	800.4060	15
10	2087.8884	1044.4479	2070.8619	1035.9346	2069.8779	1035.4426	L	1546.7781	773.8927	1529.7516	765.3794	1528.7676	764.8874	14
11	<b>2200.9725</b>	1100.9899	2183.9460	1092.4766	2182.9619	1091.9846	L	<b>1433.6941</b>	717.3507	1416.6675	708.8374	1415.6835	708.3454	13
12	2288.0045	1144.5059	2270.9780	1135.9926	2269.9940	1135.5006	S	<b>1320.6100</b>	660.8086	1303.5835	652.2954	1302.5994	651.8034	12
13	<b>2375.0366</b>	1188.0219	2358.0100	1179.5086	2357.0260	1179.0166	S	<b>1233.5780</b>	617.2926	1216.5514	608.7794	1215.5674	608.2873	11
14	<b>2490.0635</b>	1245.5354	2473.0369	1237.0221	2472.0529	1236.5301	D	<b>1146.5460</b>	573.7766	1129.5194	565.2633	1128.5354	564.7713	10
15	<b>2603.1476</b>	1302.0774	2586.1210	1293.5641	2585.1370	1293.0721	I	<b>1031.5190</b>	516.2631	1014.4925	507.7499	1013.5084	507.2579	9
16	<b>2704.1952</b>	1352.6013	2687.1687	1344.0880	<b>2686.1847</b>	1343.5960	T	<b>918.4349</b>	459.7211	901.4084	451.2078	<b>900.4244</b>	450.7158	8
17	<b>2775.2324</b>	1388.1198	2758.2058	1379.6065	<b>2757.2218</b>	1379.1145	A	<b>817.3873</b>	409.1973	800.3607	400.6840	799.3767	400.1920	7
18	<b>2862.2644</b>	1431.6358	2845.2378	1423.1226	2844.2538	1422.6305	S	<b>746.3502</b>	373.6787	729.3236	365.1654	728.3396	364.6734	6
19	<b>2961.3328</b>	1481.1700	2944.3062	1472.6568	<b>2943.3222</b>	1472.1648	V	<b>659.3181</b>	330.1627	642.2916	321.6494			5
20	<b>3075.3757</b>	1538.1915	3058.3492	1529.6782	3057.3652	1529.1862	N	<b>560.2497</b>	280.6285	543.2232	272.1152			4
21	3303.4326	1652.2199	3286.4060	1643.7067	3285.4220	1643.2147	C	<b>446.2068</b>	223.6070	429.1802	215.0938			3
22	<b>3374.4697</b>	1687.7385	3357.4432	1679.2252	3356.4591	1678.7332	A	218.1499	109.5786	201.1234	101.0653			2
23							K	147.1128	74.0600	130.0863	65.5468			1

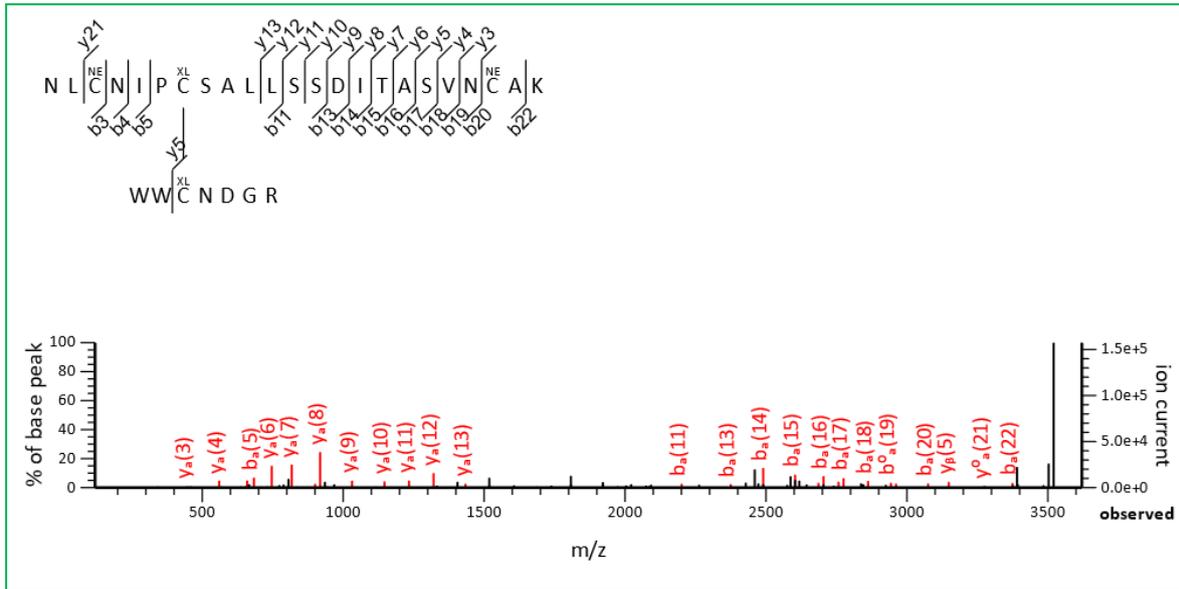
#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	187.0866	94.0469					W							7
2	373.1659	187.0866					W	3334.4959	1667.7516	3317.4693	1659.2383	3316.4853	1658.7463	6
3	3060.3722	1530.6897					C	<b>3148.4166</b>	1574.7119	3131.3900	1566.1986	3130.4060	1565.7066	5
4	3174.4151	1587.7112	3157.3886	1579.1979			N	461.2103	231.1088	444.1837	222.5955	443.1997	222.1035	4
5	3289.4421	1645.2247	3272.4155	1636.7114	3271.4315	1636.2194	D	347.1674	174.0873	330.1408	165.5740	329.1568	165.0820	3
6	3346.4635	1673.7354	3329.4370	1665.2221	3328.4530	1664.7301	G	232.1404	116.5738	215.1139	108.0606			2
7							R	175.1190	88.0631	158.0924	79.5498			1



NCBI BLAST search of **NLCNIPCSALLSSDITASVNC AK**  
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)  
 Other BLAST [web gateways](#)

上図について、緑で囲われた領域について次頁に拡大図を準備しています。

画面上部(下図)はスペクトルベースで理論値がどのようにマッチングしているのかを確認する事ができる画面で、クロスリンクされたペプチドそれぞれに $\alpha$ 、 $\beta$ という記号を割り振り、マッチングがどちらのペプチド由来のフラグメントなのかも示されています。



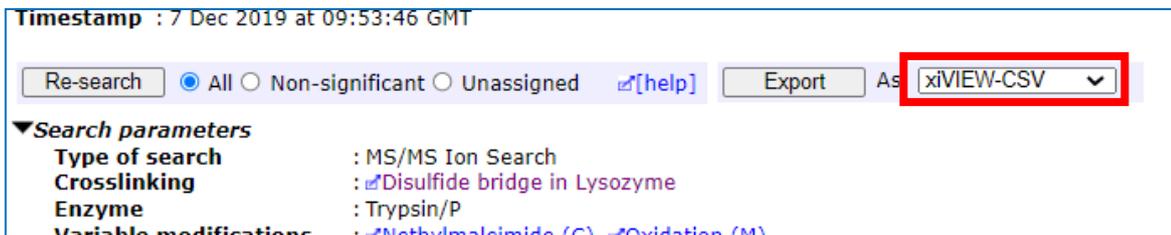
その下には理論値ベースのマッチング確認画面があり、これも両ペプチドそれぞれの理論値表が準備されています。

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	115.0502	58.0287	98.0237	49.5155			N							23
2	228.1343	114.5708	211.1077	106.0575			L	3406.5323	1703.7698	3389.5058	1695.2565	3388.5217	1694.7645	22
3	<b>456.1911</b>	228.5992	439.1646	220.0859			C	3293.4482	1647.2278	3276.4217	1638.7145	<b>3275.4377</b>	1638.2225	21
4	<b>570.2341</b>	285.6207	553.2075	277.1074			N	3065.3914	1533.1993	3048.3648	1524.6861	3047.3808	1524.1940	20
5	<b>683.3181</b>	342.1627	666.2916	333.6494			I	2951.3484	1476.1779	2934.3219	1467.6646	2933.3379	1467.1726	19
6	780.3709	390.6891	763.3443	382.1758			P	2838.2644	1419.6358	2821.2378	1411.1226	2820.2538	1410.6305	18
7	1816.7352	908.8713	1799.7087	900.3580			C	2741.2116	1371.1094	2724.1851	1362.5962	2723.2011	1362.1042	17
8	1903.7673	952.3873	1886.7407	943.8740	1885.7567	943.3820	S	1704.8473	852.9273	1687.8207	844.4140	1686.8367	843.9220	16
9	1974.8044	987.9058	1957.7778	979.3925	1956.7938	978.9005	A	1617.8152	809.4113	1600.7887	800.8980	1599.8047	800.4060	15
10	2087.8884	1044.4479	2070.8619	1035.9346	2069.8779	1035.4426	L	1546.7781	773.8927	1529.7516	765.3794	1528.7676	764.8874	14
11	<b>2200.9725</b>	1100.9899	2183.9460	1092.4766	2182.9619	1091.9846	L	<b>1433.6941</b>	717.3507	1416.6675	708.8374	1415.6835	708.3454	13
12	2288.0045	1144.5059	2270.9780	1135.9926	2269.9940	1135.5006	S	<b>1320.6100</b>	660.8086	1303.5835	652.2954	1302.5994	651.8034	12
13	<b>2375.0366</b>	1188.0219	2358.0100	1179.5086	2357.0260	1179.0166	S	<b>1233.5780</b>	617.2926	1216.5514	608.7794	1215.5674	608.2873	11
14	<b>2490.0635</b>	1245.5354	2473.0369	1237.0221	2472.0529	1236.5301	D	<b>1146.5460</b>	573.7766	1129.5194	565.2633	1128.5354	564.7713	10
15	<b>2603.1476</b>	1302.0774	2586.1210	1293.5641	2585.1370	1293.0721	I	<b>1031.5190</b>	516.2631	1014.4925	507.7499	1013.5084	507.2579	9
16	<b>2704.1952</b>	1352.6013	2687.1687	1344.0880	<b>2686.1847</b>	1343.5960	T	<b>918.4349</b>	459.7211	901.4084	451.2078	<b>900.4244</b>	450.7158	8
17	<b>2775.2324</b>	1388.1198	2758.2058	1379.6065	<b>2757.2218</b>	1379.1145	A	<b>817.3873</b>	409.1973	800.3607	400.6840	799.3767	400.1920	7
18	<b>2862.2644</b>	1431.6358	2845.2378	1423.1226	2844.2538	1422.6305	S	<b>746.3502</b>	373.6787	729.3236	365.1654	728.3396	364.6734	6
19	<b>2961.3328</b>	1481.1700	2944.3062	1472.6568	<b>2943.3222</b>	1472.1648	V	<b>659.3181</b>	330.1627	642.2916	321.6494			5
20	<b>3075.3757</b>	1538.1915	3058.3492	1529.6782	3057.3652	1529.1862	N	<b>560.2497</b>	280.6285	543.2232	272.1152			4
21	3303.4326	1652.2199	3286.4060	1643.7067	3285.4220	1643.2147	C	<b>446.2068</b>	223.6070	429.1802	215.0938			3
22	<b>3374.4697</b>	1687.7385	3357.4432	1679.2252	3356.4591	1678.7332	A	218.1499	109.5786	201.1234	101.0653			2
23							K	147.1128	74.0600	130.0863	65.5468			1

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	187.0866	94.0469					W							7
2	373.1659	187.0866					W	3334.4959	1667.7516	3317.4693	1659.2383	3316.4853	1658.7463	6
3	3060.3722	1530.6897					C	<b>3148.4166</b>	1574.7119	3131.3900	1566.1986	3130.4060	1565.7066	5
4	3174.4151	1587.7112	3157.3886	1579.1979			N	461.2103	231.1088	444.1837	222.5955	443.1997	222.1035	4
5	3289.4421	1645.2247	3272.4155	1636.7114	3271.4315	1636.2194	D	347.1674	174.0873	330.1408	165.5740	329.1568	165.0820	3
6	3346.4635	1673.7354	3329.4370	1665.2221	3328.4530	1664.7301	G	232.1404	116.5738	215.1139	108.0606			2
7							R	175.1190	88.0631	158.0924	79.5498			1

Summary の結果画面に話を戻します。crosslink 検索を行った場合、ファイル出力オプションで「xiVIEW-CSV」というファイルが出力可能となります。



このファイルは、サイト xiVIEW

[https://xiview.org/xiNET\\_website/index.php](https://xiview.org/xiNET_website/index.php)

において、結合状況を図示可能な入力データとして利用することができます。



### 10-3-5. Crosslinking 設定の作成

Crosslink の検索を行う際パラメーターとして設定項目を選択する必要がありますが、この項目については検索内容に応じて予め設定を作成しておく必要があります。設定変更は、リンカーや結合対象といった基本的な内容に加え、検索対象とするタンパク質またはデータベース名についても指定する必要がありますなど、若干複雑な構造となっています。リンカーの設定については「13-5. Linkers」を、また Crosslinking の項目自体の設定については、「13-9. crosslinking」の設定をご覧ください。

なお、crosslink 検索、特に interlink(タンパク質間の結合)は本来探索的な用途で利用したいケースも多く、対象タンパク質が 1 つや2つで済まない事もしばしばです。しかし検索対象とするタンパク質は多すぎると検索が終わらず、同定基準も高くなりあまり機能しません。目安として、**ターゲットのタンパク質組み合わせは 100 まで**、という提案をしています。これは **intralink** であれば **100 エントリー程度**、**interlink** であれば **10 エントリー(10x10=100)程度**となり、デフォルト設定もそのようになっています。

## 10-3-6. Crosslink 補足説明へのリンク

### ■ Crosslink 検索 MASCOT HELP ページ

<https://www.matrixscience.com/help/crosslink.html>

### ■ S-S 結合検索結果例

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F002553.dat](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F002553.dat)

\* 同じ結果をお手元の MASCOT Server で開けば、この結果を再検索したり少し条件を変えた検索を試すことができます。

[http://localhost/mascot/cgi/master\\_results\\_2.pl?file=../data/F002553.dat](http://localhost/mascot/cgi/master_results_2.pl?file=../data/F002553.dat)

### ■ DSS 結合検索結果例

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=../data/F002555.dat; ignoreionsscorebelow=0.05](https://www.matrixscience.com/cgi/master_results_2.pl?file=../data/F002555.dat; ignoreionsscorebelow=0.05)

\* 同じ結果をお手元の MASCOT Server で開けば、この結果を再検索したり少し条件を変えた検索を試すことができます。

[http://localhost/mascot/cgi/master\\_results\\_2.pl?file=../data/F002555.dat; ignoreionsscorebelow=0.05](http://localhost/mascot/cgi/master_results_2.pl?file=../data/F002555.dat; ignoreionsscorebelow=0.05)

## 10-4. Error Tolerant Search

### 10-4-1. Error Tolerant Search 概要

MASCOT の MIS 検索で数多くの query を同時に検索した際、どのような配列ともマッチしない query が必ずと言っていいほど多数存在します。マッチしない理由として以下のような事が考えられます。

1. データベースにないペプチド配列だった
2. 指定した誤差範囲を超えていた
3. Precursor の電荷が適切でなかった
4. 切断パターンが想定していたものと異なるペプチドだった
5. 想定外の修飾

これらの原因を検証するためにはパラメーターを変えながらの再検索が必要となりますが、そのための便利なアプローチが「Error Tolerant Search」です。Error Tolerant Search は最初、通常 conditions で検索を行います。この時同定基準を超えた query をそのまま結果として採用しつつ、同定基準を超えなかった query については以下の 3 点を考慮して 2 段階目の検索を行います。

- A. **選択した酵素の切断パターンを「半特異的」に変更**します。すなわち片方が切断ルールに則った切断でもう片方が任意の切断、となります。
- B. 修飾の追加。修飾リストに含まれるすべての項目が対象で、**追加で1つだけ修飾が付くケースを網羅的に探索**します。
- C. **アミノ酸置換**の確認。DNA の 1 塩基置換によってもたらされるアミノ酸の置換パターンを網羅的に探索します。さらにデータベースが塩基配列であった場合は挿入や欠失も考慮します。

A は前頁で挙げた5つのマッチしない理由のうちケース 4、B はケース 5、C はケース 1 に対してある程度フォローする事ができる検索方法となります。2 段階目の検索では 1 回目と 2 回目の試行で検索対象となったペプチド数をもとに同定基準値が算定され、1 回だけ行われた検索の結果に比べ 2 段階目の検索では同定基準値が厳しくなります。

### 10-4-2. Error Tolerant Search を実行する方法

検索パラメーターで「**Error tolerant**」にチェックを入れてから検索(Start Search)を行います。この時「Decoy」へのチェックと「Target PSM FDR」での数値の指定を行っている、2 段階目への query 選別における FDR 閾値の適用、並びに検索後の同定基準算出に Target PSM FDR の設定値が使用されます(下図)。Error Tolerant 検索の信頼度を高めるためにもご利用をお勧めします。

Data format	Mascot generic	Precursor		m/z
Instrument	Default	<b>Error tolerant</b>	<input checked="" type="checkbox"/>	
Decoy	<input checked="" type="checkbox"/>	Target PSM FDR	1%	
Start Search ...		Reset Form		

### 10-4-3. Error Tolerant Search 検索結果

以下のページをもとに結果画面を使った説明をいたします。可能であれば実際に WEB ページで開いてご参照ください。

[https://www.matrixscience.com/cgi/master\\_results\\_2.pl?file=./data/F981130.dat; sigthres=0.06427; target\\_fdr=1](https://www.matrixscience.com/cgi/master_results_2.pl?file=./data/F981130.dat; sigthres=0.06427; target_fdr=1)

rank2 のタンパク質ファミリー、2 の隣の三角をクリックし展開してファミリータンパク質へのペプチドマッチング状況を確認する画面を開いてください。その中の、PPB1\_HUMAN へのアサインペプチドをご確認ください(次頁図)。

[次頁に続きます]

	Score	Mass	Matches	Sequences
<input checked="" type="checkbox"/> 2.1	499	58259	31 (31)	16 (16) Alkaline phosphatase, placental type OS=Homo sapiens OX=9606 GN=ALPP PE=1 S
<input checked="" type="checkbox"/> 2.2	352	57626	25 (25)	12 (12) Alkaline phosphatase, germ cell type OS=Homo sapiens OX=9606 GN=ALPG PE=2 S
<input checked="" type="checkbox"/> 2.3	70	57119	8 (8)	7 (7) Intestinal-type alkaline phosphatase OS=Homo sapiens OX=9606 GN=ALPI PE=1 S

Redisplay All None

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

Auto-fit to window

Query	Dupes	Observed	Mr (expt)	Mr (calo)	Delta	M	Score	Expect	Rank	U	1	2	3	Peptide
41		517.1760	1032.3375	1032.5604	-0.2229	0	70	1.1e-05	1		■	■		R.GSSIFGLAPGK.A
46		532.1837	1062.3528	1062.5710	-0.2181	0	60	0.00016	1	U		■		R.GSSIFGLAPSK.A
53		545.6818	1089.3491	1089.5819	-0.2327	0	54	0.014	1		■	■		R.GSSIFGLAPGK.A + [+57.0215 at S2]
65		567.6566	1133.2987	1133.5499	-0.2511	0	45	0.007	1		■	■	■	R.GNEVISVMNR.A + Oxidation (M)
86		614.2001	1226.3856	1226.6329	-0.2473	0	28	0.039	1	U		■		K.LGPEIPLAMDR.F + Oxidation (M)
100		653.2101	1304.4057	1304.6837	-0.2780	0	87	3.6e-07	1		■	■		K.GNFQTIGLSAAAR.F
124		710.2235	1418.4324	1418.7154	-0.2829	0	92	3.5e-06	1	U		■	■	K.ANFQTIGLSAAAR.F + [+100.0160 at
133		499.1349	1494.3828	1494.6694	-0.2866	0	89	8.3e-06	1		■	■		L.DPSLMEMTEAALR.L + 2 Oxidation (M)
136		754.6864	1507.3582	1507.6691	-0.3109	0	43	0.36	1		■	■		R.NWYSADAVPASAR.Q + [+57.0215 at N
145		526.1538	1575.4396	1575.7780	-0.3384	0	81	5e-05	1		■	■		R.ALTETIMFDDAIER.A + [-48.0034 at
156		820.7283	1639.4420	1639.7763	-0.3343	0	106	6.3e-09	1		■	■		R.ALTETIMFDDAIER.A + Oxidation (M)
158		552.5010	1654.4812	1654.8315	-0.3503	0	54	0.03	1		■	■		K.HVPDSGATATAYLCGVK.G + [-33.9877
162		836.2372	1670.4598	1670.8052	-0.3454	0	88	1.1e-05	1	U		■		G.VIPAEENPAFWR.Q
165		841.2310	1680.4474	1680.8029	-0.3554	0	93	3.9e-06	1		■	■		R.ALTETIMFDDAIER.A + [+57.0215 at
170		864.2888	1726.5629	1726.9294	-0.3664	0	43	0.0036	1		■	■		K.AYTVLLYNGPGYVLK.D
175		586.4951	1756.4635	1756.8420	-0.3786	0	48	0.13	1		■	■		G.IIPVEENPDFWR
176		879.2425	1756.4705	1756.8420	-0.3715	0	83	3.7e-05	1		■	■		G.IIPVEENPDFWR
179		593.4834	1777.4285	1777.7764	-0.3478	0	46	0.22	1		■	■		K.HVPDSGATATAYLCGVK.G + [+31.9357
208		975.8100	1949.6055	1950.0245	-0.4190	0	86	7.7e-08	1		■	■		K.NLIIFLGDGMGVSTVTAAR.I + Oxidati
209		976.2340	1950.4534	1950.8555	-0.4021	1	27	0.024	1		■	■		K.DGARDVTESESGSPEYR.Q
211		656.1752	1965.5039	1964.8712	0.6327	1	68	0.0016	1		■	■		K.DGARDVTESESGSPEYR.Q + [+14.015
213		664.5518	1990.6336	1991.0510	-0.4174	0	65	0.0025	1		■	■		K.NLIIFLGDGMGVSTVTAAR.I + [+57.02
214		665.1736	1992.4991	1992.9132	-0.4141	0	54	0.043	1	U		■		R.DSTLDPSLMEMTEAALR.L + 2 [+57.02
216		1001.2027	2000.3908	2000.8058	-0.4150	0	65	3.4e-05	1	U		■		R.MGTPDPEYDDYSQGGTR.L + Oxidatio
217		667.8046	2000.3919	2000.8058	-0.4139	0	73	8.9e-07	1	U		■		R.MGTPDPEYDDYSQGGTR.L + Oxidatio
218		670.1561	2007.4466	2007.8770	-0.4304	1	75	0.00029	1		■	■		K.DGARDVTESESGSPEYR.Q + Acetyl (N-term); [+15.0109 at N-te
222		681.8205	2042.4397	2042.8164	-0.3767	0	58	9.1e-05	1	U		■		R.MGTPDPEYDDYSQGGTR.L + Acetyl (N-term); Oxidation (M)
245		766.2128	2295.6165	2296.1084	-0.4919	0	55	3.3e-05	1	U		■		R.QQSAVPLDEETHAGEDVAVFAR.G
252		784.5440	2350.6103	2351.1030	-0.4927	0	68	0.0022	1	U		■		R.QQSAVPLDEETHAGEDVAVFAR.G + [-17.0265 at N-term]
253		790.2186	2367.6341	2368.1295	-0.4954	0	93	1e-08	1	U		■		R.QQSAVPLDEETHAGEDVAVFAR.G
260		809.2208	2424.6406	2425.1510	-0.5104	0	66	0.0037	1	U		■		R.QQSAVPLDEETHAGEDVAVFAR.G + [+57.0215 at N-term]
274		914.9160	2741.7263	2741.2306	0.4956	0	44	0.58	1		■	■	■	R.QEGCQDIATQLISNMDIDVILGGGR.K + Oxidation (M); [+79.9568 at C4]
275		920.5878	2758.7415	2759.3218	-0.5804	0	126	3.7e-09	1		■	■	■	R.QEGCQDIATQLISNMDIDVILGGGR.K + [+57.0215 at M15]

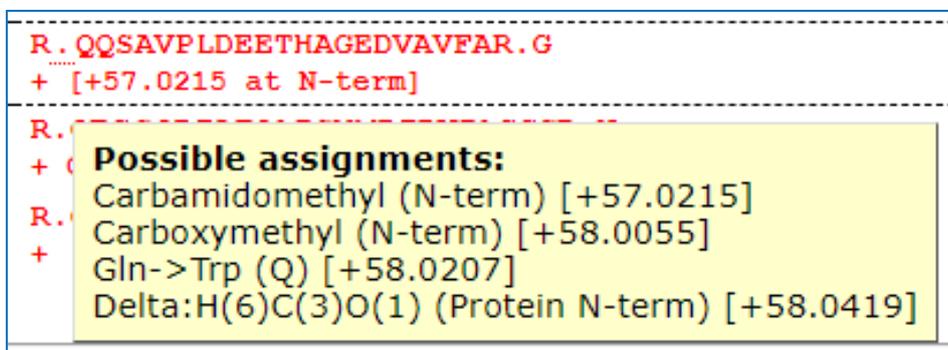
結果には Expect の値も表示されています。2 段階目の検索結果は 1 段階目で同定済みの結果に比べ、同定基準が高めになっています。

以下、Error Tolerant 検索で見つかった結果にフォーカスしていきます。上図の赤線で囲われた部分を拡大したのが下図です。

query 133 は N 末端側がトリプシンのルールに従わないペプチドを検出しています。

■ ■ L.DPSLMEMTEAALR.L + 2 Oxidation (M)

Query 260 では N 末端側に+57.0215 修飾の可能性が指摘されており、[+57.0215 at N-term]の部分にカーソルを合わせると、その質量変動を引き起こす修飾やアミノ酸の置換が一覧で表示されます(下図)。質量分析装置のデータでは単に質量の変動のみがわかるので、この中のどれであるかについては特定する事はできません。状況から考えてありえないような選択肢も含まれることがあります。MASCOT では単に質量の変化から候補を提供しているためその内容について判別する事ができません。



R. QQSAVPLDEETHAGEDVAVFAR.G  
+ [+57.0215 at N-term]

R. **Possible assignments:**  
+ Carbamidomethyl (N-term) [+57.0215]  
R. Carboxymethyl (N-term) [+58.0055]  
+ Gln->Trp (Q) [+58.0207]  
Delta:H(6)C(3)O(1) (Protein N-term) [+58.0419]

Error Tolerant Search は便利な検索ではあり、同定基準値に基づいた期待値も表示されてはいますが、結果の採用については慎重な判断が必要です。このページの結果に対する議論については以下 HELP ページをご覧ください。現段階ではまだ「有力な正解候補となるペプチドを探すためのツール」程度に考えておくのがよいかもしれません。

[https://www.matrixscience.com/help/error\\_tolerant\\_help.html#RESULTS](https://www.matrixscience.com/help/error_tolerant_help.html#RESULTS)

## 11. FDR と Decoy データベース

### 11-1. FDR はなぜ必要か？

プロテオームの解析において、検索エンジンは確率論を基にしたアルゴリズムで同定ペプチドを検出しますが、MASCOT では同定基準を統計的有意性を元に判断しており、信頼度としては 95%が解析当初から採用されていました。

解析が広まるにつれ、同定ペプチド並びに同定タンパク質が、想定した有意水準とかけ離れているのではないかと思われるケースが散見されるようになり、**検索エンジンが提唱する同定基準とは別の観点から同定結果を検証**する事が求められるようになりました。また、好ましい信頼度の数字についても議論されるようになりました。その流れの中で広く広まった評価方法が Decoy データベースを使った **FDR** による方法であり、**FDR の基準値として 1%が採用される事が非常に多くなりました。**

### 11-2. FDR とは、Decoy データベースとは

FDR(False Discovery Rate)という用語を単独でインターネット検索するとその多くが多重検定に関する事項として結果が出ますが、MASCOT で使用されている(、並びにプロテオミクスの定性で使われる)FDR は具体的な計算方法という面で大きく異なります。

検査・検定における Sensitivity, Specificity についてその各種性能を論じる際よく以下のような表が使われますが、MASCOT で使われている **FDR は検査(判定)の Positive 全体における False Positive の割合**を示します。以下の表における文字を使うなら、**FDR とは  $B / (A+B)$**  です。

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

MASCOT(、並びにプロテオミクスの定性で使われる)における FDR のポイントは False Positive, すなわち B の見積り方にあります。FDR の計算において分母にあたる **A+B は MASCOT が同定とみなしたペプチドすべての数**にあたります。検索結果の同定数そのものであり容易に使用できます。**しかし B については通常の検索結果からはその数がいくつであるかを判断する事ができません。**もし何かしらの判断により False Positive の結果とわかるようならそれらは誤った同定結果として取り除かれるべきですが、当然そのようにする事ができません。

B の算出に一役買っているのが「Decoy」データベースです。正解が一つも含まれていないはずのデータベース、「Decoy」(撒き餌、疑似餌)データベースを準備して、そのデータベースに対して同じ検索条件・同じ入力データを検索した際同定基準スコアを超えてきたデータ数を、最初の検索の False Positive B と概ね同じくらいの数字であろうと仮定して、FDR 計算における分子 B の代わりに利用しています。

すなわち

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

と計算しています。

ペプチドでカウントした際の FDR 1%とは、「同定基準スコアを超えたペプチドの中で、正解とは異なる配列が含まれている割合が 1%」とみなす事ができます。

なお Decoy データベースは「正解の配列を含まないデータベース」という事になりますが、通常の検索とデータベースの性質(残基長分布、エントリー数、アミノ酸出現頻度)を揃えるため、通常の検索で使用したデータベースの各エントリーの配列を逆向きまたはランダム配列に変更したものを利用します。

### 11-3. MASCOT で FDR を算出させる方法

検索パラメーターにある「Decoy」のチェックを入れて検索を実行する事で FDR が計算可能となります(下図)。また「Target PSM FDR」の数値を設定する事で、FDR の設定値を目安として同定基準値が調整された状態で検索結果が表示されます。現在 FDR の設定値として 1%が採用されることがほとんどです。

The screenshot shows the MASCOT search configuration interface. The 'Decoy' checkbox is checked and highlighted with a red box. The 'Target PSM FDR' dropdown menu is also highlighted with a red box and set to '1%'. Other visible settings include: Peptide tol. ± 1.2 Da, # 13C 0, MS/MS tol. ± 0.6 Da, Peptide charge 2+, Monoisotopic (selected), Average (unselected), Data file (File selection), Data format Mascot generic, Instrument Default, Precursor (m/z), and Error tolerant (unchecked). Buttons for 'Start Search ...' and 'Reset Form' are visible at the bottom.

Decoy のオプションを使用すると、使用しない時に比べ2倍強の時間がかかります。

**Protein Family Summary**

Format

Significance threshold  $p <$   Max. number of families  [\[help\]](#)

Target FDR (overrides sig. threshold)  FDR type

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

Show Percolator scores  Dendrograms cut at

Preferred taxonomy

**▼ Sensitivity and FDR (reversed protein sequences)**

	Target	Decoy	FDR
Protein family members	302	11	3.64%
<input type="text" value="PSMs"/> above <input type="text" value="homology"/>	1821	18	0.99%

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

FDRの数値については、上図にあるように結果画面の Format 欄、「**Target FDR**」「**FDR type**」で調整する事ができます(FDR type については「**11-4-1.Decoy データベースの利用方法、FDR 計算方法には明確な定義がありません**」もご覧ください)。また「**Sensitivity and FDR**」の箇所をクリックすると、**Target** (通常のデータベースの事)と **Decoy** それぞれでどれだけの数が同定基準を超えていたのかについて確認する事ができます。

## 11-4. FDR にまつわるトピックス

### 11-4-1. Decoy データベースの利用方法、FDR 計算方法には明確な定義がありません

Decoy データベースを準備する方法や Decoy のカウント対象、並びに FDR を計算する際にどのようなデータを数え上げるのかについての**厳密な定義はなされていません**。そもそも数え上げでなく別の方法で類似の概念を持つ計算式により算出されている事もあるなど、**研究者により多少異なるのが現状です**。従ってその内容には注意が必要であると共に、ご自身でデータを示す際にも説明の記述に注意する必要があります。論文や論文投稿の際データをアップロードするのに利用されている repository site などでデータ処理の項目を見てどのように処理されたかを確認する必要があります。

主要な論文やコンソーシアムでは一定の枠組みでのガイドラインがありますのでご紹介します。

- HUPO

<https://www.hupo.org/HPP-Data-Interpretation-Guidelines>

- MCP Guidelines

<https://www.mcponline.org/guidelines>

FDR の適用についてどのようなパターンがあるか調べる意味で、あるいはご自身で記載する際の参考のため、PRIDE, jPOST などの repository site の投稿データを実際に眺めてその記述内容をご確認頂く事をお勧めいたします。次頁図は PRIDE というサイトで FDR に関する表記の例です。peptide FDR

だけでなく protein FDR も適用している事、FDR の設定値が 1%である事、Decoy データベースにはランダムでなく逆向き(reverse)配列を適用している事などがわかります。

**Description**  
Staphylococcus aureus causes invasive infections and easily acquires antibiotic resistances. Even antibiotic susceptible S. aureus can survive antibiotic therapy and persist, requiring prolonged treatment and surgical interventions. These so-called persisters display an arrested-growth phenotype, tolerate high antibiotic concentrations and are associated with chronic and recurrent infections. To c...

[Read more](#)

**Sample Processing Protocol**  
S. aureus cells (2-5\*108 per sample) were lysed in 50 µL of lysis buffer (1% sodium deoxycholate (SDC), 10 mM TCEP, 100 mM Tris, pH=8.5) using thirty cycles of sonication (30 sec on, 30 sec off per cycle) on a Bioruptor (Dianode). Following sonication, proteins in the bacterial lysate were reduced by TCEP at 95°C for 10 min. Proteins were then alkylated using 15 mM chloroacetamide at 37°C for 30 m...

[Read more](#)

**Data Processing Protocol**  
The acquired raw files were imported into the Progenesis QI software (v2.0, Nonlinear Dynamics Limited), which was used to extract peptide precursor ion intensities across all samples applying the default parameters. The generated mgf-files were searched using MASCOT against a decoy database containing normal and reverse sequences of the Staphylococcus aureus proteome (UniProt, NCTC 8325, release date: 09.12.2016) and commonly observed contaminants generated using the SequenceReverser tool from the MaxQuant software (Version 1.0.13.13). The following search criteria were used: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 3 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) and protein N-terminal acetylation were applied as variable modifications; mass tolerance of 10 ppm (precursor) and 0.02 Da (fragments). **The database search results were filtered using the ion score to set the false discovery rate (FDR) to 1% on the peptide and protein level, respectively, based on the number of reverse protein sequence hits in the datasets.** Quantitative analysis results from label-free quantification were processed using the SafeQuant R package v.2.3.4 (<https://github.com/eahrne/SafeQuant/>) (32). Data processing included computation of total peak/reporter areas across all LC-MS runs, and summation of peak areas per protein and LC MS/MS run. Only isoform specific peptide ion signals were considered for quantification. Data were then normalized using the variance stabilization normalization provided by the VSN R package v3.52.0 (33). The normalized protein expression values were used for statistical testing of differentially abundant proteins between conditions. Here, empirical Bayes moderated t-tests were applied, as implemented in the R limma package v3.40.2 (34). The resulting per protein and condition comparison p-values were adjusted for multiple testing using the Benjamini Hochberg method. The assessment of

**Organism part**  
Unknown

**Diseases**  
Unknown

**Modification**  
[monohydroxylated residue](#)  
[acetylated residue](#)  
[iodoacetamide derivatized residue](#)

**Instrument**  
Q Exactive HF

**Software**  
Unknown

**Experiment Type**  
Unknown

**Quantification**  
Unknown

**Dataset reuses**  
Not available

**Similar Studies**

- [Bacterial persistence - Bacterial persistence is an active  \$\sigma\$ S stress response to metabolic flux limitation](#)  
2015-03-24
- [Complement C5a impairs phagosomal maturation in the neutrophil through phosphoproteomic remodelling.](#)  
2020-01-14
- [Phosphoproteomic analysis of](#)

## 11-4-2. FDR 計算のカウント対象

MASCOT のデフォルト設定では、PSM、即ち query 数をベースとして数え上げた FDR を使用していますが、これについては Format の「**FDR type**」というパラメーターで「Sequence」に変更する事ができます。「Sequence」にすると、同じペプチドにマッチしている query は1つにまとめて数え上げます。

## 11-4-3. Decoy データベースでの検索結果と “1 Hit Wonders”

結果画面にある「**Sensitivity and FDR**」を展開した際一番下に表示される「**the decoy report**」ハイパーリンクをクリックすると、Decoy データベースに対して行った検索結果画面を確認する事ができます。**9-5-2** でランダムマッチなペプチドが1つのタンパク質に2つ以上アサインされることはほとんど

ないという説明をしましたが、Decoy の結果を見ると実際に同定基準を超えるペプチドが2つ以上アサインされるタンパク質はほとんどない事がわかります。

#### 11-4-4. 「peptide FDR 1%」と同定タンパク質の正確さ

peptide FDR 1%というのは、同定ペプチドを 100 集めた際、1つのペプチドが不正解である可能性がある事を意味しています。同定ペプチドのリストの正確さについても同様の評価が可能です。

同定タンパク質については、前述のようにアサインペプチドが2つ以上である場合、その2つとも **False Positive** ペプチドであることはほとんどありません。一方アサインペプチドが1つしかない場合、そのタンパク質同定の確からしさはペプチドと同じです。すなわち peptide FDR1%なら、「アサインペプチドが1つである同定タンパク質を 100 集めた場合、1つ間違えている同定タンパク質である可能性がある」と言えます。

現状では同定タンパク質数を確保する目的もあり、このような正確性での解析レベルがほとんどの研究者に受け入れられています。すなわち 9-5-1 で説明したように「同定タンパク質とは、ユニークな1つの同定ペプチドがアサインされているタンパク質」という基準となっています。

#### 11-4-5. Protein FDR

結果画面の「**Sensitivity and FDR**」を展開した時に表示されている「**Protein Family members**」の行で表示されている FDR は、通常良く言われるところの「**protein FDR**」になります。MASCOT では peptide (PSM または Sequence) と異なり、**protein FDR で足切りをすることができません**。近年 protein FDR も条件として求められる事がありますが、MASCOT で protein FDR の数字も何かの条件を満たすようにするには、**PSM または sequence FDR を上げたり下げたりしながらこちらの表示をご自身で確認して頂く必要があります**。ほとんどのケースでは protein FDR と peptide FDR に同じ設定値(1%)を適用した場合、protein FDR の方が条件が厳しく実質 protein FDR を適用する事で両条件を達成します。

Protein FDR を調整する際、**タンパク質にアサインされるペプチド数の最低限必要な数を 1 から 2 に変更する事で対応する事もできます**。この場合 protein FDR の値は 0 になる事がほとんどです。

#### 11-4-6. 同定ペプチド数が少ないケースにおける FDR 調整の限界

FDR が 1%という数字は、100 の同定ペプチドがあると 1 つ False Positive ペプチドがあるレベルである事を示します。FDR 計算という観点から見ると、**Decoy データベースでの検索で 1 つが同定基準を超えた場合、通常データベースでの検索結果で最低 100 のペプチドが同定基準を超えないと FDR1%以上になってしまう事になります**。すなわち最低 100 のペプチドを同定できないようなデータでは FDR 1%以下にするのはとても難しく、このようなケースにおいてもどうにか FDR を無理やり適用させたい場合、Decoy での検索において同定基準を超えるペプチドが 0 になるよう、Decoy ペプチドの最高スコアより少し上に基準値を置くしかありません。Decoy の結果のランダム性に大きく引っ張られ、FDR そのものの信用度が揺らぐケースです。

## 12. MASCOT Server 管理 – データベースと検索ログ

この章では MASCOT Server の管理に利用するプログラムについてご紹介します。

12-1 では MASCOT Server で使用しているデータベースや各種ログを確認する事ができる **Database Status** について、12-2 では検索のログである **Search log** についてご紹介しています。

### 12-1. 現在利用可能なデータベースに関する情報 : Database Status

Database status では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(下図)。

The screenshot shows the 'MASCOT search status page' in a web browser. The page displays the following information:

- Version: 2.8.0.1 - mskk (UC2H-LHC6-Z3V8-W62R-K3M8) [Licence Info](#)
- 12 logical, 1 physical Intel processors (hyper-threading enabled, 6 core). CPUs: 0 1 2 3 4 5 6 7 8 9 10 11 available, using: 0 1 2 3 4 5 6 7 8 9 10 11 . [0 searches running]
- Active databases: 3. Inactive databases: 0. Max databases: 256.
- Navigation links: [Search log](#), [monitor log](#), [error log](#), [Error message descriptions](#), [Do not auto refresh this page](#)
- Database details (3 entries):
  - PRIDE Contaminants**: Family = C:/inetpub/mascot/sequence/PRIDE\_Contaminants/current/PRIDE\_Contaminants; Filename = PRIDE\_Contaminants\_20160906.msp; Pathname = C:/inetpub/mascot/sequence/PRIDE\_Contaminants/current/PRIDE\_Contaminants; Status = In use; State Time = Fri Aug 13 02:30:49 # searches = 0; Mem mapped = NO; Request to mem map = NO; Request unmap = NO; Mem locked = NO; Number of threads = 1; Current = YES; Type = Spectral library.
  - 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; 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.
  - UP5640 H sapiens**: Family = C:/inetpub/mascot/sequence/UP5640\_H\_sapiens/current/UP5640\_H\_sapiens\_\*.f; Filename = UP5640\_H\_sapiens\_20201007.fasta; Pathname = C:/inetpub/mascot/sequence/UP5640\_H\_sapiens/current/UP5640\_H\_sapiens\_20201007.fasta; Status = In use; 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.

図内の表記について説明いたします。\$MASCOT は MASCOT のインストールフォルダでを意味します。

- ① : 使用している MASCOT のバージョンと MASCOT のライセンス
- ② : MASCOT Server が搭載されているコンピューターの CPU の中で、MASCOT Server がどのコアを動かしているか、並びに現在稼働中の検索件数

③ : 各種ログへの移動

- Search log** : 検索ログ (Home にある「Search log」と同じ)
- monitor log** : MASCOT Server の動作ログ。\$MASCOT¥logs¥monitor.log の内容
- error log** : MASCOT Server のエラーログ。\$MASCOT¥logs¥errorlog.txt の内容
- Error message descriptions** : MASCOT のエラー番号に関する情報
- Do not auto refresh this page** : database status 画面を自動・定期的に更新するかしないかの切り替え

④ : 各データベースの状況。例として以下に SwissProt の情報を表示した部分の拡大図を示します。

```
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
```

各項目が示す内容は以下の通りです。

- Name** : MASCOT Server で登録されているデータベースの名称
- Family** : データベース側で登録されている、ファイルの path 並びにファイル名称ルールの情報
- Filename** : 現在認識されている fasta ファイルの名称
- Pathname** : 現在認識されている fasta とファイルが置かれている path
- Status** : データベースの現在の状況を表しています。データベースは fasta ファイルが更新された場合、その fasta ファイルから MASCOT で使用する複数のファイルが作成され、すべてのファイルが作成された後に最初の検索テストが実施されます。また設定内容によってはデータベースファイルがメモリ上にマッピングされ、最後に使用可能となります。データベース構築から使用可能になるまでの一連の状況を知らせるため、Status 項目で以下のように表示されます。

**Creating compressed files**

**Running 1st test**

**First test just run OK**

**Trying to memory map files**

**Just enabled memory mapping**

**In Use**

またどこかの段階でエラーになった場合などは「Halted:」などのように問題を含む状況であることを表すメッセージを表示します。

- 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
- 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** : データベースの種類が Amino Acid か Nucleic Acid か、あるいは Spectral Library か

表示のいくつかの箇所はハイパーリンクになっていて、クリックする事でさらに詳しい情報を確認する事ができます。以下、各ハイパーリンク先の画面について説明します。

・**データベースの名称(Name の項目)** : 該当データベースで行われている検索の進捗状況、または過去に行った検索の内容。実行中の検索がある場合、”**Current jobs**”にその内容が表示されます。”Job”の検索番号がハイパーリンクになっており、クリックするとさらに詳細の情報を確認できます(次頁)。

Mascot database status - SwissProt							
<b>Current jobs</b>							
<a href="#">Job</a>	<a href="#">PID</a>	<a href="#">Start time</a>	<a href="#">Dur.</a>	<a href="#">Status</a>	<a href="#">User</a>	<a href="#">UserID</a>	<a href="#">Title</a>
<a href="#">1243</a>	14420	Wed Sep 1 16:34:14	14	Searching....		0	Copy of mgf 02 (C:\ProgramData\Matrix Science\Ma
<b>Completed jobs</b>							
<a href="#">Job</a>	<a href="#">PID</a>	<a href="#">Start time</a>	<a href="#">Dur.</a>	<a href="#">Status</a>	<a href="#">User</a>	<a href="#">UserID</a>	<a href="#">Title</a>

検索が完了すると、”Completed jobs”に移行します。

Mascot database status - SwissProt							
<b>Current jobs</b>							
<a href="#">Job</a>	<a href="#">PID</a>	<a href="#">Start time</a>	<a href="#">Dur.</a>	<a href="#">Status</a>	<a href="#">User</a>	<a href="#">UserID</a>	<a href="#">Title</a>
<b>Completed jobs</b>							
<a href="#">Job</a>	<a href="#">PID</a>	<a href="#">Start time</a>	<a href="#">Dur.</a>	<a href="#">Status</a>	<a href="#">User</a>	<a href="#">UserID</a>	<a href="#">Title</a>
<a href="#">1243</a>	14420	Wed Sep 1 16:34:14	810	User read res		0	Copy of mgf 02 (C:\ProgramData\Matr

•[job 番号] :

“Percent complete” で検索の進捗状況を確認する事ができます。

### Mascot Job status - Job 1243

```

Database       : SwissProt
Job Number    : 1243
Process ID    : 14420
Task ID       : 163048165201
User Name     :
User ID       : 0
User email    :
Search title  : Copy of mgf 02 (C:\ProgramData\Matrix Science\Mascot Daemon)
Percent complete : 6%
Intermed file : /data/20210901/F001243.dat
Start time    : Wed Sep 1 16:34:14 2021
End time      :
Searching... time : 0
Upload time   : 0
Query prep time : 0
Whole process time: 0
Job status    : Searching....
Priority       : current value 0. Change this by -5 -1 +1 +5
IP address    : 192.168.1.19
Type of srch  : MIS
Enzyme?       : Yes
CPU utilisation : 0%
Job requests? : No requests. Kill / Pause / Resume

```

Database Status のハイパーリンクに話を戻します。

```

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

```

•**Unidentified taxonomy** :

データベースのエントリーの中で taxonomy 情報との紐づけできなかったものをピックアップして表示します。データベースファイル自身の不具合で生じるケースもあります。

•**Statistics** :

データベースの登録情報に関する各種数値。画面上部(左下図)ではデータベースの登録件数や残基数、生物種エントリー毎の登録件数などが表示されます。画面下部には各アミノ酸残基数やエントリーの残基長分布が表示されます(右下図)。

```

Time files compressed : Thu Aug 12 12:49:30 2021
Time files compressed (int) : 1628740170
Time / date of fasta file : Mon May 24 19:20:09 2021
Time of fasta files (int) : 1621851609
Number of residues : 203519613
Number of sequences : 564638
Number with invalid residues: 0
Number of sequences too long: 0
Length of longest sequence : 35213
Maximum Accession Length : 11
Version of Mascot : 2.8.0.1
Version of this file : 5
Type of fasta file : AA
Parse rule for accession : >..|[^\]|*|¥|([^\ ])*¥
Seqs with invalid taxon tree: 0
Num sequences for taxonomy : All entries=564638
Num sequences for taxonomy : Archaea (Archaeobacteria)=19638
Num sequences for taxonomy : Eukaryota (eucaryotes)=193019
Num sequences for taxonomy : Alveolata (alveolates)=1141
Num sequences for taxonomy : Plasmodium falciparum (malaria par
Num sequences for taxonomy : Other Alveolata=793
Num sequences for taxonomy : Metazoa (Animals)=107932
Num sequences for taxonomy : Caenorhabditis elegans=4226
Num sequences for taxonomy : Drosophila (fruit flies)=5949

```

Residue	Frequency
A	16807181
B	276
C	2817673
D	11117166
E	13691007
F	7870527
G	14406614
H	4633761
I	12043802
J	0
K	11820774
L	19648473
M	4914035
N	8264502
O	29
P	9644044
Q	8003286
R	11259991
S	13512057
T	10903032
U	329
V	13968080
W	2240442
X	8280
Y	5944003
Z	249

Length	Count
2	2
3	5
4	22
5	41
6	35
7	116

### ・Recompress file :

現在構築済みのデータベースについて、再度データベースの構築を開始します。Taxonomy の設定などを変更しそれを反映させる時などに利用します。

### ・retry :

Status が Halted になった際表示される retry をクリックすると、データベースの再構築または1st テスト検索の再実施を行います。エラーが生じた際、エラーへの対処を行ったのち再度データベース構築や検索を試みる際に利用します。

Database Status で**同じデータベースが2つ表示されることがあります。データベース入れ替え中、または入れ替え済み**であることを示しています。

入れ替え途中の過程では、現在使用可能な(status = in use)データベースと、構築中である新たなデータベースの進捗状況が2つとも表示されます(下図)。

```
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

Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
Filename = SwissProt_2021_03.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2021_03.fasta
Status = Creating compressed files 1% complete
State Time = Thu Sep 2 10:18:18 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = NO Type = Amino acid
```

また入れ替え済みとなった後、現在使用可能な「status = in use」のデータベースと、かつて使用していた現在「status = not in use」となったデータベースについても引き続き表示されます(下図)。

```
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 = Not in use Statistics
State Time = Thu Sep 2 10:28:34 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = NO Type = Amino acid

Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
Filename = SwissProt_2021_03.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2021_03.fasta
Status = In use Statistics Unidentified taxonomy Recompress file
State Time = Thu Sep 2 10:28:34 # 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
```

もし2つ表示されている状況が気になるようであれば、MASCOT Service またはコンピューター自身を再起動する事で、最新の1つのみが表示されるようになります。

## 12-2. 検索ログ : Search log

MASCOT Server で行った検索はすべて検索結果を格納したファイルが保存され、検索のログも残ります。Home → 「**Search log**」をクリックするか、Database Status で「Search log」のハイパーリンクをクリックする事で、Server で行った検索のログを開く事ができます。

If this is your first visit, [start here](#). If you include results from Mascot in a publication, please cite either [www.matrixscience.com](http://www.matrixscience.com) or Electrophoresis, **20(18)** 3511-67 (1999) (abstract).

### Security

Mascot incorporates a role based security model, which allows the system administrator to control which functionality is available to individual users. If security has been enabled, you may be prompted to log in if you attempt to access a page or perform a task that requires authorisation. If security is not enabled, which is the default, the following links will simply confirm this.

- \* [Log in](#)
- \* [Log out](#)
- \* [Change password](#)
- \* [Edit settings](#)
- \* [Current session information](#)

Whether or not Mascot security is enabled, in order to comply with the licensing conditions, access to this Mascot server must be restricted to authorised users. We strongly recommend that your server is protected by a secure firewall.

### Mascot Utilities

Your system administrator may have restricted access to these utilities

#### Database Status

View the status of all the sequence databases. Links to database statistics, search status, log files, etc.

[Search Log](#)

A tabular view of the search log. Can be filtered to find specific search results.

下図のような Search log 画面が表示されます。

MASCOT search log

Version: 2.8.0 - mskk (YRNB-5YZ8-GFBC-T9W9-CYNQ)

Sort/filter | Log File: ../logs/searches.log | Start at: (-1=end, 1=start) | -1 | how many: 50 | 273 in log, 273 after filters. Data dir: | GETs?:

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\mascotsearchtest2021\803
1509	14116	UP5640_H			be		Tue Aug 3 23:49:12 2021	415	User read res	0	MIS	Yes	0		C:\temp\mascotsearchtest2021\803
1508	2452	UP5640_H			be		Tue Aug 3 23:38:02 2021	414	User read res	0	MIS	Yes	0		C:\temp\mascotsearchtest2021\803
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		Lysozyme_2p2m_1minute.temp.mgf
1505	716	SwissPro			IT		Thu Jul 8 02:01:25 2021	17	User read res	0	MIS	Yes	0		C:\temp\WTRAQ8plex6data\fromlog
1504	18348	SwissPro			IT		Thu Jul 8 01:57:29 2021	17	User read res	0	MIS	Yes	0		C:\temp\WTRAQ8plex6data\fromlog

以下、①～③各パートに記載されている情報や利用可能な機能について、詳しく説明します。

①

Version: 2.8.0 - mskk (YRNB-5YZ8-GFBC-T9W9-CYNQ)

Sort/filter | Log File: ../logs/searches.log | Start at: (-1=end, 1=start) | -1 | how many: 50 | 273 in log, 273 after filters. Data dir: | GETs?:

バージョンとライセンスの情報が先頭行に記載されています。

- Sort/filter** : この行で指定した情報に基づいて検索ログが表示されます。条件を変更して新たに表示させる場合、このボタンを押します。
- Log File** : MASCOT の検索ログのファイルの path
- Start at** : 検索ログを、検索番号の降順(-1)で表示するか昇順(1)で表示するか
- how many** : 表示するログ件数
- Data dir** : デフォルト設定以外の data フォルダへのパスを使用するとき指定
- Gets?** : getseq 情報を表示するか

②

Job#	PID	dbase	User Name	Email	Ti	In
<input type="radio"/>						
<input checked="" type="checkbox"/>						
<input type="text"/>						
<a href="#">1511</a>	5436	UP2195_D	Monitor Test DB 0		MS	<a href="#">...</a>
<a href="#">1510</a>	12840	UP5640_H			be	<a href="#">...</a>
<a href="#">1509</a>	14116	UP5640_H			be	<a href="#">...</a>

**Job#** : 検索番号。デフォルトでは 1234 番から実行順に割り振られています

**PID** : MASCOT Server で使用しているプロセス ID

**dbase** : 使用したデータベース

**User Name** : 検索ユーザー(検索時に指定)

**Email** : メールアドレス。ただし local 版ではメールアドレスでなく短いテキスト情報でもよい

**Title(Ti とデフォルト表示)** : 検索のタイトル (検索時に指定)

**Intermediate File (In とデフォルト表示)** : 結果ファイルの置かれている場所

③

start time	Durati	Status	Prio	Type	Enzyme IP	User ID	Peak list data file
<input type="radio"/>							
<input checked="" type="checkbox"/>							
<input type="text"/>							
Fri Aug 6 11:57:49 2021	1	No email setu	0	MIS	Yes	0	test_search.mgf
Wed Aug 4 00:00:25 2021	410	User read res	0	MIS	Yes	0	C:%temp%\mascotsearchtest20
Tue Aug 3 23:49:12 2021	415	User read res	0	MIS	Yes	0	C:%temp%\mascotsearchtest20
Tue Aug 3 23:38:02 2021	414	User read res	0	MIS	Yes	0	C:%temp%\mascotsearchtest20

**start time** : 検索開始時間

**Duration** : 検索時間

**Status** : 検索結果の状態

**Priority** : 検索時に指定した、検索の優先順位

**Type** : PMF / SQ / MIS

**Enzyme** : 特異性を持つ設定で行ったか(Yes)、特異性を持たない設定 None で実行したか (No)

**IP address** : 検索をかけたコンピューターの IP アドレス

**User ID** : MASCOT のセキュリティ設定で指定しているユーザーID

**Peak list data file** : 入力データファイルのパス並びに名称

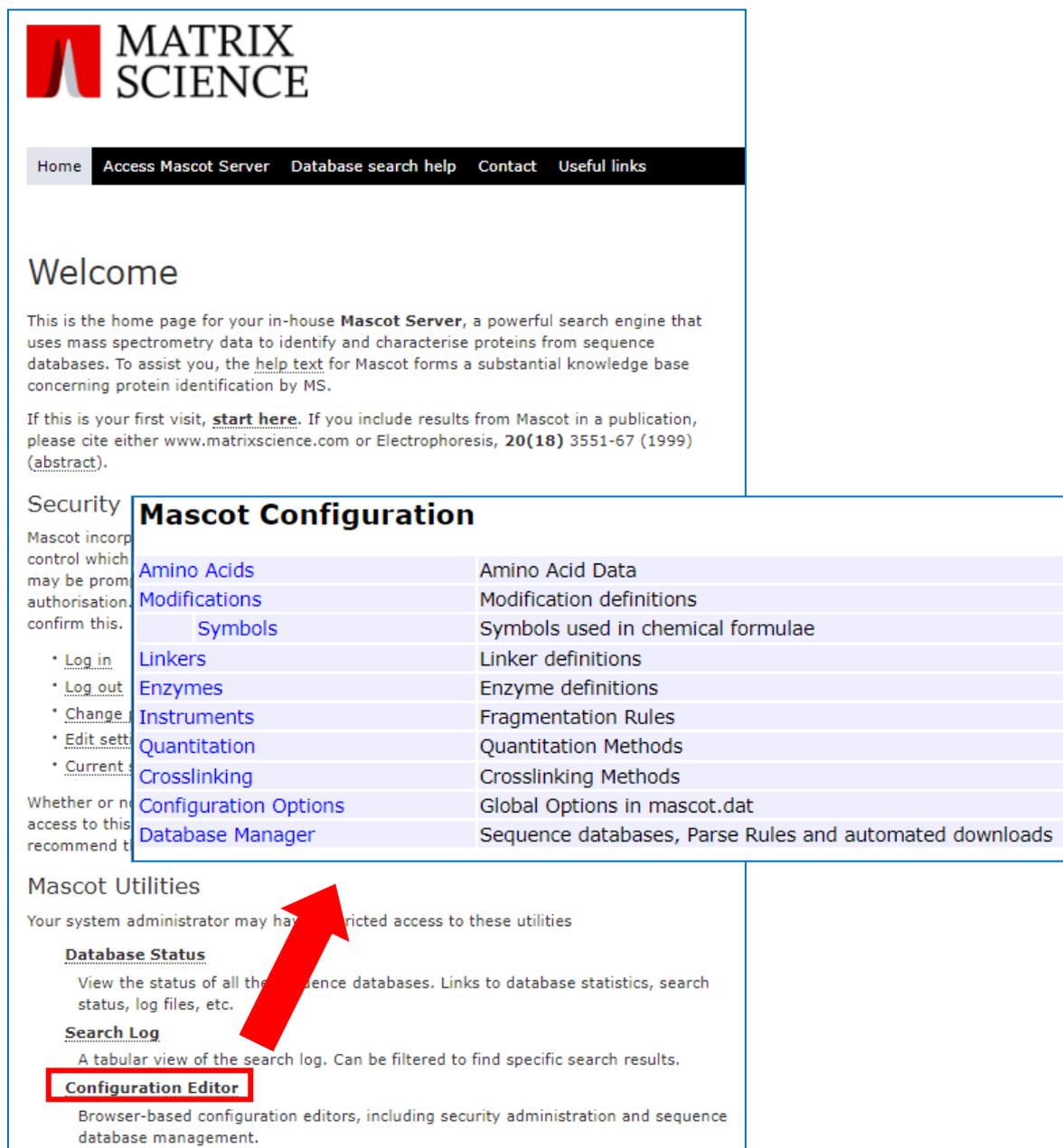
各項目の一番上にある入力欄はフィルターとして利用できます。例えば SwissProt で検索した結果のみを表示させたい場合、“dbase”項目に“SwissProt”と入力して“sort/filter”ボタンを押すことでフィルターリングが実現します。

またその上にあるチェックボックスは項目の表示を広げてすべて読めるようにするかどうか、さらにその上にあるラジオボタン(1項目のみ選択可能)は選択項目で並び替えを行う事を意味します。並び替えについては通常”Job#”が選択されていて、検索を行った順番で並べられています。

## 13. MASCOT Server のカスタマイズ

### 13-1. カスタマイズは Configuration Editor で

MASCOT Server の各種設定変更を行うのが「**Configuration Editor**」です。Home 画面にある「Configuration Editor」をクリックすると、設定内容一覧が現れます。各設定画面について、以降より詳しく説明いたします



**MATRIX SCIENCE**

Home Access Mascot Server Database search help Contact Useful links

## Welcome

This is the home page for your in-house **Mascot Server**, a powerful search engine that uses mass spectrometry data to identify and characterise proteins from sequence databases. To assist you, the [help text](#) for Mascot forms a substantial knowledge base concerning protein identification by MS.

If this is your first visit, [start here](#). If you include results from Mascot in a publication, please cite either [www.matrixscience.com](http://www.matrixscience.com) or Electrophoresis, **20(18)** 3551-67 (1999) ([abstract](#)).

### Security

Mascot incorporates security control which may be promoted or restricted by authorisation. Please contact your administrator to confirm this.

- [Log in](#)
- [Log out](#)
- [Change password](#)
- [Edit settings](#)
- [Current session](#)

Whether or not you have access to this page, we recommend that you contact your administrator.

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

### Mascot Utilities

Your system administrator may have restricted access to these utilities

- [Database Status](#)  
View the status of all the sequence databases. Links to database statistics, search status, log files, etc.
- [Search Log](#)  
A tabular view of the search log. Can be filtered to find specific search results.
- [Configuration Editor](#)  
Browser-based configuration editors, including security administration and sequence database management.

## 13-2. Amino Acids

アミノ酸の質量に関する設定を行う画面です。Home -> Configuration Editor -> **Amino Acids** で開く事ができます。

最初に定義一覧が現れます(下図)。基本的には定義されている内容をそのままご利用頂く事になります。

Mascot Configuration: Amino Acids					
Amino Acids					
1	3	Fullname	Monoisotopic (Da)	Average (Da)	Composition
A	Ala	Alanine	71.037114	71.0779	H(5) C(3) N O
R	Arg	Arginine	156.101111	156.1857	H(12) C(6) N(4) O
N	Asn	Asparagine	114.042927	114.1026	H(6) C(4) N(2) O(2)
D	Asp	Aspartic acid	115.026943	115.0874	H(5) C(4) N O(3)
C	Cys	Cysteine	103.009185	103.1429	H(5) C(3) N O S
E	Glu	Glutamic acid	129.042593	129.1140	H(7) C(5) N O(3)
Q	Gln	Glutamine	128.058578	128.1292	H(8) C(5) N(2) O(2)
G	Gly	Glycine	57.021464	57.0513	H(3) C(2) N O
H	His	Histidine	137.058912	137.1393	H(7) C(6) N(3) O
I	Ile	Isoleucine	113.084064	113.1576	H(11) C(6) N O
L	Leu	Leucine	113.084064	113.1576	H(11) C(6) N O
K	Lys	Lysine	128.094963	128.1723	H(12) C(6) N(2) O
M	Met	Methionine	131.040485	131.1961	H(9) C(5) N O S
F	Phe	Phenylalanine	147.068414	147.1739	H(9) C(9) N O
P	Pro	Proline	97.052764	97.1152	H(7) C(5) N O
S	Ser	Serine	87.032028	87.0773	H(5) C(3) N O(2)
T	Thr	Threonine	101.047679	101.1039	H(7) C(4) N O(2)
W	Trp	Tryptophan	186.079313	186.2099	H(10) C(11) N(2) O
Y	Tyr	Tyrosine	163.063329	163.1733	H(9) C(9) N O(2)
V	Val	Valine	99.068414	99.1311	H(9) C(5) N O
N-term	N-term	N-term	1.007825	1.0079	H
C-term	C-term	C-term	17.002740	17.0073	H O
U	Sec	Selenocysteine	150.953633	150.0379	H(5) C(3) N O Se <a href="#">Edit</a>
J	Xle	Leu or Ile	113.084064	113.1576	H(11) C(6) N O <a href="#">Edit</a>
O	Pyr	Pyrrrolysine	237.147727	237.2982	H(19) C(12) N(3) O(2) <a href="#">Edit</a>
B	Bbb	Asn or Asp	114.534940	114.5950	ambiguity code
X	Xxx	Any residue	111.000000	111.0000	ambiguity code
Z	Zzz	Glu or Gln	128.550590	128.6216	ambiguity code

Main menu

**U, J, O** の 3 文字についてはユーザーがカスタマイズする事ができます。カスタマイズを希望する場合、カスタマイズを希望する文字の行にある「**Edit**」ボタンをクリックします。

B	Bbb	Asn or Asp	114.534940	114.5950	ambiguity code
X	Xxx	Any residue	111.000000	111.0000	ambiguity code
Z	Zzz	Glu or Gln	128.550590	128.6216	ambiguity code

---

Letter **U**      Short name       Full name

Composition     

Monoisotopic      **150.953633**

Average            **150.0379**

---

このカスタマイズは、**20** 種類のアミノ酸以外のアミノ酸をターゲットとしたい時や、あるアミノ酸の特定修飾について配列に組み込んでおきたい場合などに、配列データベースのカスタマイズと一緒に使用するなどといった使い方があります。

## 13-3. Modifications

修飾、アミノ酸の質量の変化に関する設定を行う画面です。Home -> Configuration Editor -> **Modifications** で開く事ができます。

最初に、設定済みの定義一覧が現れます(下図)。既にある設定をそのまま微調整する場合は項目名のリンクを、現存設定を残しコピーしてそちらで書き換える場合は「**Copy**」のリンクをクリックします。新規の設定を作成するには画面下部の「**Add new modification**」ボタンを押します。

**Mascot Configuration: Modifications**

Displaying 1499/1499

**Visibility:**

 Short list  
 Long list  
 Mixed  
 Not listed

**Error tolerant:**

 Yes  
 No  
 Mixed

**Classifications:** clear

-  
 Post-translational  
 Co-translational  
 Pre-translational  
 Chemical derivative  
 Artefact

**Source:**

 Unimod  
 Edited Unimod  
 Local

**Apply to selected: 0**

Modifications								
<input type="checkbox"/> Title ↑	Monoisotopic	Average	Composition	Source	Visibility	Err Tol		
<input type="checkbox"/> 15N-oxobutanoic	-18.023584	-18.0239	H(-3) 15N(-1)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 2-dimethylsuccinyl	144.042259	144.1253	H(8) C(6) O(4)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 2-monomethylsuccinyl	130.026609	130.0987	H(6) C(5) O(4)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 2-nitrobenzyl	135.032028	135.1201	H(5) C(7) N O(2)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 2-succinyl	116.010959	116.0722	H(4) C(4) O(4)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 2HPG	282.052824	282.2476	H(10) C(16) O(5)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 3-deoxyglucosone	144.042259	144.1253	H(8) C(6) O(4)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 3-hydroxybenzyl-phosphate	186.008196	186.1018	H(7) C(7) O(4) P	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 3-phosphoglyceryl	167.982375	168.0420	H(5) C(3) O(6) P	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 3sulfo	183.983029	184.1693	H(4) C(7) O(4) S	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 4-ONE	154.099380	154.2063	H(14) C(9) O(2)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 4-ONE+Delta:H(-2)O(-1)	136.088815	136.1910	H(12) C(9) O	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> 4AcAllylGal	372.142033	372.3671	H(24) C(17) O(9)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> a-type-ion	-46.005479	-46.0254	H(-2) C(-1) O(-2)	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> AccQTag	170.048013	170.1674	H(6) C(10) N(2) O	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> Acetyl	42.010565	42.0367	H(2) C(2) O	Unimod	mixed	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> Acetyl:13C(2)	44.017274	44.0220	H(2) 13C(2) O	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> Acetyl:2H(3)	45.029395	45.0552	H(-1) 2H(3) C(2) O	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> Acetyldeoxyhypusine	97.089149	97.1582	H(11) C(6) N	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>
<input type="checkbox"/> Acetylhypusine	113.084064	113.1576	H(11) C(6) N O	Unimod	long	yes	<a href="#">Copy</a>	<a href="#">Print</a>

Page 1/75 Go to page  <<"/>  Page size

■ 最初の一覧表で表示されている項目(列)は以下の通りです。

- Title** : 修飾の名称。詳しくは <https://www.unimod.org/names.html>
- Monoisotopic** : Monoisotopic 質量
- Average** : Average 質量
- Composition** : 化学式、修飾が付く残基における質量の増減
- Source** : unimod 由来の項目か、unimod 内容を一部変更したか、自身で作成した項目か
- Visibility** : modification リストのデフォルト表示(short)か、”show all mod”オプションを使用した時に初めて表示される項目(long)か、修飾が付くアミノ酸残基によってそれらの設定が異なり混ざっている状態 (mixed)か
- Err Tol** : Error Tolerant 検索 (10-4)の時に考慮される修飾かどうか

115

表示されている項目のうち青い太字で表されている Title, Monoisotopic, Average については、項目名をクリックする事で降順/昇順 に並べ替える事ができます。目的とする項目を探す際に、Title(名称)のアルファベット順にするか、Monoisotopic 質量の順に並べて計算値から探すことで見つけやすいです。

■ 前頁図一覧下に表示されているボタンは以下の通りです。

**Page N/M Go to page , page size N**

: 表示しているページの変更や 1 ページで表示される項目数の変更

**Add new modification** : 新しい修飾を作成

**Main menu** : Configuration Editor の一覧表示に戻ります

**Check Unimod** : 修飾設定は Unimod というサイトでまとめられています。「check Unimod」ボタンは Unimod にある修飾設定ファイルの内容を確認し、**手元のファイルより新しかった場合にそれを取得して設定をアップデートします**。なお Unimod ファイルをアップデートしても自身で作成した修飾はなくなりません。

■ 前頁図左フレームは、表示内容をフィルターリングしたり、項目を複数選んで設定をまとめて変更する際に利用する欄です。

以下は表示内容をフィルターリングする項目です。

**Visibility** : modification リストのデフォルト表示(short)か、”show all mod”オプションを使用した時に初めて表示される項目(long)か、修飾が付く残基によってそれらの設定が異なり混ざっている(mixed)か、あるいはリストには記入されておらず修飾以外の特定の設定内のみ有効な内容か(Not listed)。

**Error tolerant** : Error tolerant 検索で考慮される修飾かどうか。「**Yes**」:使用される設定、「**No**」:使用されない設定、「**Mixed**」:残基により使用される/されない 設定が混ざっている

**Classifications** : 修飾の分類。検索などで使用するタグのようなものですが、検索そのものには「AA substitution」以外どれを選んでいても影響を与えません。

**Source** : Unimod 登録内容か、Unimod 登録内容をユーザーが一部変更したか、完全にユーザーが作成したものか

また以下は設定をまとめて変更するためのオプションです。

**Apply to selected** : 設定をまとめて変更します。以下の選択肢があります。

- **Include in short list** :

右側の表で選択(行先頭のチェック欄)された内容を、すべて修飾の short list (パラメーター選択時にデフォルトで表示される修飾リスト)に含める

- **Include in long list :**  
右側の表で選択(行先頭のチェック欄)された内容を、すべて修飾の long list (パラメーター選択時、“show all mod”オプションを指定する事で初めて表示される項目)に含める
- **Include in error tolerant / Exclude from error tolerant:**  
右側の表で選択(行先頭のチェック欄)された内容を、Error tolerant 検索で考慮する修飾に含む、あるいは考慮対象から外す
- **Delete :** 選択した修飾をリストから除く

以降、さらに修飾の各設定項目について、その設定画面の説明をいたします。例として項目「Phospho」(リン酸化)の設定を利用いたします。

### View Modification :Phospho

**Name**

Title Phospho  
Fullname Phosphorylation

Delta
Specificity
Ignore Masses
Misc
References

**Delta**

Monoisotopic	79.966331
Average	79.9799
Composition	H O(3) P

OK
Make editable

「Phospho」リンクをクリックするとまず **Title** , **Fullname** 欄に名称並びに詳細情報が記された画面が現れます(上図)。この段階では閲覧のみ可能で編集ができません。画面右下にある「**Make editable**」をクリックすると編集が可能となります(下図)。

Delta
Specificity
Ignore Masses
Misc
References

**Delta**

Monoisotopic	79.966331
Average	79.9799
Composition	H O(3) P

Symbols
13C
▼
1
▼
Add

Save changes
Cancel
Update
Revert to Unimod
Show differences

「**Delta**」タブでは、修飾が付いた場合の質量の増減分を「**Composition**」に記載します。設定の際すぐ下にある「**Symbols**」で対象の元素記号(または化合物を表す記号)とその数を指定して「**Add**」を押す事で「**Composition**」に反映されます。減少分はマイナスの数字がありそれを指定する事で対応可能です。また入力ミスなどの場合、Composition 部分の数字を直接書き換えたり、同じ記号で数字を変えて再び Add 操作を行う事でも対応可能です。

Delta	<b>Specificity</b>	Ignore Masses	Misc	References			
Specificity							
Specificity	Site	E	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	R	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	K	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	H	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	C	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	D	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	Y	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	T	Position	Anywhere	Copy	Delete	Show Details
Specificity	Site	S	Position	Anywhere	Copy	Delete	Show Details
New Specificity Definition		Show All Details					

「**Specificity**」タブでは修飾が付くアミノ酸残基、あるいは N 末端/C 末端の位置 について設定します (上図)。設定項目がある場合、「**Show Details**」ボタンを押すことでより詳しい設定内容が確認できます (下図)。

Specificity	Site	D	Position	Anywhere	Copy	Delete	Show Details	
Specificity	Site	Y	Position	Anywhere	Copy	Delete	Show Details	
Specificity	Site	T	Position	Anywhere	Copy	Delete	Show Details	
Specificity	Site	S	Position	Anywhere	Copy	Delete	Hide Details	
Classification	Post-translational		Group	1				
Visibility	In short list	<input checked="" type="checkbox"/>	In long list	<input checked="" type="checkbox"/>	Use in error tolerant search	<input checked="" type="checkbox"/>		
Notes	<input type="text"/>							
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide				Delete			
Composition	<input type="text"/>				Symbols	13C	1	Add
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide				Delete			
Monoisotopic: <b>97.976896</b> Average: <b>97.9952</b>								
Composition	H(3) O(4) P				Symbols	13C	1	Add
New Neutral Loss								
New Specificity Definition		Hide All Details		Show All Details				
Save changes			Cancel		Update		Revert to Unimod	Show differences

表示項目は以下の通りです。

**Site** : 修飾が付くアミノ酸またはペプチドの位置(N 末端/C 末端)。

**Position** : Site のアミノ酸についての位置の絞り込み。

**Classification** : 修飾の分類。但しフィルターリングの際使用するタグのような意味合いがあるものの、検索そのものには「AA substitution」以外影響を与えません。

- Group** : 異なる specificity で同じ group 番号が割り振られている場合、phospho(ST)などのようにアミノ酸が1つにまとめられます。一方 phospho の場合 ST と Y は異なる group 番号が割り振られており、修飾リストでもそれらが分けられて表示されます。
- Visibility** : modification リストでの表示のされ方。
- **short list** : modification の選択項目として最初から表示。
  - **long list** : "show all modifications" のオプションをオンにしないと表示されない。
- Notes** : 修飾についてのメモ。

また Neutral loss、すなわち precursor の段階では損失しておらずプロダクトイオンマススペクトル測定の段階で損失する官能基などの塊やそれに対応する質量減少に関する設定もあります。

**Neutral loss** : 基本的に「**Scoring**」を指定します。Neutral Loss に派生して生じるピークについて、その存在を特に重視したり、逆に影響を抑える場合などに別のオプションを利用します。

- **Scoring** : NeutralLoss により生じるフラグメント側のピークが、マッチング・スコアリングの対象となります(通常はこちらを選択します)。
- **Satellite** : Neutral loss で生じたピークをピークリストから外します。マッチング対象の理論値は Neutral loss をしていないものとなります。Neutral loss のピークがノイズと認識されない分、マッチングスコアが上昇する事が期待される時に選択します。
- **Peptide** : プレカーサーから Neutral loss 分が消失して生じたピークをピークリストから除きます。マッチングしない無駄なデータが減る分、マッチングスコアが上昇する事が期待される時に選択します。
- **Required Peptide** :  
上記「Peptide」の性質を持ち、かつプレカーサーから Neutral loss 分が消失して生じたピークが存在しない場合、Neutral loss に関連する考慮を全くしなくなります。 : 参考 : <https://www.matrixscience.com/pdf/2005WKSHP2.pdf>

**Composition** : 「**Delta**」で指定した修飾がついている状態を基本としてそこからの増減を指定します。隣にある「Symbols」と数字、「Add」ボタンと連動していますまた直接の記入にも対応しています。

Neutral loss するパターンとしないパターンの両方を考慮したい場合、例の図のように片側は空欄(しない場合)の設定を作成しておく必要があります。

以降タブがありますが、記入が必須ではなく重要度はあまり高くない設定項目です。

「Ignore Masses」タブ(下図)は解析に付随して現れる特定のピークについて、そのピークを入力データから取り除きマッチングの対象から外す際に利用します。

「Misc」タブ(下図)は登録エントリーに関してこれまでの設定項目以外のメモ、「その他」となります。

「Reference」タブは、この Modification 情報を設定するにあたり参考にした情報について記入します。

## 13-4. Symbols

元素や分子の部分構造の質量に割り当てた記号(Symbol)とその質量を確認する事ができる画面です(下図)。Home -> Configuration Editor -> **Symbols** で開く事ができます。設定変更はできず確認のみとなります。この画面内で定義されている Symbol は、modification, linker などのページ内で利用する事が可能です。また各記号に対して実際に割り当てられた質量や部分構造を確認する際にも利用します。

Mascot Configuration: Symbols				
Symbols				
Symbol ↑	Name	Monoisotopic	Average	Composition
13C	Carbon 13	13.003355	13.0034	13C
15N	Nitrogen 15	15.000109	15.0001	15N
18O	Oxygen 18	17.999160	17.9992	18O
2H	Deuterium	2.014102	2.0141	2H
Ac	Acetate	42.010565	42.0367	C(2) H(2) O
Ag	Silver	106.905092	107.8682	Ag
Al	Aluminium	26.981539	26.9815	Al
As	Arsenic	74.921594	74.9216	As
Au	Gold	196.966543	196.9666	Au
B	Boron	11.009306	10.811	B
Br	Bromine	78.918336	79.904	Br
C	Carbon	12	12.0107	C
Ca	Calcium	39.962591	40.078	Ca
Cd	Cadmium	113.903357	112.411	Cd
Cl	Chlorine	34.968853	35.453	Cl
Co	Cobalt	58.933198	58.9332	Co
Cr	Chromium	51.940510	51.9961	Cr
Cu	Copper	62.929599	63.546	Cu
dHex	Deoxy-hexose	146.057909	146.1412	C(6) H(10) O(4)
F	Fluorine	18.998403	18.9984	F
Fe	Iron	55.934939	55.845	Fe
H	Hydrogen	1.007825	1.0079	H
Hep	Heptose	192.063388	192.1666	C(7) H(12) O(6)
Hex	Hexose	162.052824	162.1406	H(10) C(6) O(5)
HexA	Hexuronic acid	176.032088	176.1241	C(6) H(8) O(6)

[次頁に続きます]

## 13-5. Linkers

クロスリンクの設定に関連する Linkers に関する設定を行う画面です。Home -> Configuration Editor -> **Linkers** で開く事ができます。

最初に、設定済みの定義一覧が現れます(下図)。既にある設定をそのまま微調整する場合は項目名のリンクを、現存する設定を残しコピーしてそちらで書き換える場合は「Copy」のリンクをクリックします。新規の設定を作成するには画面下部の「Add new linker」ボタンを押します。

「Linkers」の設定内容は、**13-3** の modification と概ね同じで、modification と一部連動しています。この画面の表示内容で不明な点がある場合は **13-3** 前半をご覧ください。

### Mascot Configuration: Linkers

Displaying 16/16

**Visibility:**  
 Short list  
 Long list  
 Mixed  
 Not listed

**Error tolerant:**  
 Yes  
 No  
 Mixed

**Classifications:** [clear](#)  
Cross-link  
CID cleavable cross-link  
Photo cleavable cross-link  
Other cleavable cross-link

**Source:**  
 Unimod  
 Edited Unimod  
 Local

Title ↑	Monoisotopic	Average	Composition	Source	Visibility	Err Tol
<a href="#">Xlink:BS2G</a>	96.021129	96.0841	H(4) C(5) O(2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:BuUrBu</a>	196.084792	196.2032	H(12) C(9) N(2) O(3)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:Disulfide</a>	-2.015650	-2.0159	H(-2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:Dityrosine</a>	-2.015650	-2.0159	H(-2)	local	long	no <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DMP</a>	122.084398	122.1677	H(10) C(7) N(2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DSS</a>	138.068080	138.1638	H(10) C(8) O(2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DSSO</a>	158.003765	158.1750	H(6) C(6) O(3) S	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DST</a>	113.995309	114.0563	H(2) C(4) O(4)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DTBP</a>	172.012890	172.2711	H(8) C(6) N(2) S(2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:DTSSP</a>	173.980921	174.2406	H(6) C(6) O(2) S(2)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:EDC</a>	-18.010565	-18.0153	H(-2) O(-1)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:EGS</a>	226.047738	226.1828	H(10) C(10) O(6)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:SDA</a>	82.041865	82.1005	H(6) C(5) O	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:SMCC</a>	219.089543	219.2365	H(13) C(12) N O(3)	Unimod	long	yes <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:test1</a>	185.058912	185.1821	C(10) H(7) N(3) O	local	long	no <a href="#">Copy</a> <a href="#">Print</a>
<a href="#">Xlink:test2</a>	159.068414	159.1846	C(10) H(9) N O	local	long	no <a href="#">Copy</a> <a href="#">Print</a>

Page 1/1 Go to page [Xlink:B to Xlink:t](#) << >> Page size [20](#)

[Add new linker](#) [Main menu](#) [Check Unimod](#)

以降、各項目の設定画面について、「Xlink:DSS」の設定画面を使って説明します。

Xlink:DSS のリンクをクリックするとまず Title ,Fullname 欄に名称並びに詳細情報が記された画面が現れます。この段階では閲覧のみ可能で編集ができません。画面右下にある「**Make editable**」をクリックすると編集が可能となります。

### View Linker :Xlink:DSS

**Name**  
Title Xlink:DSS  
Fullname disuccinimidyl suberate (DSS)

[Delta](#) [Specificity](#) [Ignore Masses](#) [Misc](#) [References](#)

**Delta**  
Monoisotopic **138.068080**  
Average **138.1638**  
Composition H(10) C(8) O(2)

[OK](#) [Make editable](#)

「**Delta**」タブ(上図)では、リンカーが付く際の質量の増減分を”Composition”に記載します。増減については結合対象となる 2 つのアミノ酸配列を併せた数字で考えます。設定の際、すぐ下にある「Symbols」で対象の元素記号(または化合物を表す記号)とその数を指定して「Add」ボタンを押す事で”Composition”に反映されます。減少分は数字にマイナスのものがありませんのでそれを指定する事でも対応可能です。また入力ミスなどの場合は Composition 部分の数字を書き換えたり、同じ記号で数字を変えながら再び Add 操作を行う事でも対応可能です。

「**Specificity**」タブ(上図)ではリンカーが付くアミノ酸残基、あるいは N 末端/C 末端の位置 について設定します。設定項目がある場合、「Show Details」ボタンを押すことでより詳しい設定内容が確認できます。(次頁図)

[次頁に続く]

Delta				Specificity	Ignore Masses	Misc	References
Specificity	Site	N-term	Position	Protein N-term	Copy	Delete	Show Details
Specificity	Site	K	Position	Anywhere	Copy	Delete	Hide Details
Classification		Cross-link	Group	1			
Visibility	In short list	<input type="checkbox"/>	In long list	<input checked="" type="checkbox"/>	Use in error tolerant search	no	
Notes	<input type="text"/>						
Neutral loss	Code	I	Pairs With	<input type="text"/>	Delete		
Composition		<input type="text"/>	Symbols	13C	1	Add	
Description	Intact cross-link						
Neutral loss	Code	A	Pairs With	<input type="text"/>	Delete		
Monoisotopic: -17.026549 Average: -17.0305							
Composition		H(-3) N(-1)	Symbols	13C	1	Add	
Description	Ammonia quenched monolink						
Neutral loss	Code	W	Pairs With	<input type="text"/>	Delete		
Monoisotopic: -18.010565 Average: -18.0153							
Composition		H(-2) O(-1)	Symbols	13C	1	Add	
Description	Water quenched monolink						
Neutral loss	Code	T	Pairs With	<input type="text"/>	Delete		
Monoisotopic: -121.073893 Average: -121.1350							
Composition		C(-4) H(-11) N(-1) O(-3)	Symbols	13C	1	Add	
Description	Tris quenched monolink						
New Neutral Loss							
New Specificity Definition							
Hide All Details							
Show All Details							

- Site** : 修飾が付くアミノ酸またはペプチドの位置(N 末端/C 末端)。
- Position** : Site のアミノ酸についての位置特定。
- Classification** : リンカーの分類ですが、表示の絞り込みなどで使用するタグのようなもので検索そのものには影響を与えません。
- Group** : 異なる specificity で同じ group 番号が割り振られている場合、Link(ST)などのようにアミノ酸が1つにまとめられます。
- Visibility** : modification リストでの表示のされ方
- **short list** : modification の選択項目として最初から表示
  - **long list** : "show all modifications"のオプションをオンにしないと表示されない

また Neutral loss に関する設定もあります。必ずしも Loss だけでなく、質量が増加する場合もこちらで設定をします。この部分は crosslinking の monolink 設定と連動しており、modification にはない独自の設定項目があるので注意が必要です。

**Code** : Crosslinking の Monolink 設定で、NeutralLoss のパターンを指定する際に選択する記号として表示されます。記号は任意で、MASCOT Server 側で特に準備されたものはありません。ユーザーが連想しやすいものをご利用ください。Linker がペア

ペプチドを構成せず構造の一部が変化するパターンは1つでない事も多いですが、その複数のパターンを予め作成しておくことができます。

**Pairs With** : monolink パターンの中で特に対になっている組み合わせがあれば、相手側の Code を記入します。

**Composition** : 「Delta」で指定した修飾がついている状態を基本としてそこからの増減を指定します。隣にある「Symbols」と数字、「Add」ボタンによる入力に対応しています。

**Description** : Code で指定したパターンの内容について後で確認するときを使うメモ欄です。

以下、重要度があまり高くない設定欄です。不明な場合は空欄でも問題ないことがほとんどです。

「Ignore Masses」タブ(下図)は解析に付随して現れる特定のピークについて、そのピークを入力データから取り除きマッチングの対象から外す際に利用します。

Delta Specificity **Ignore Masses** Misc References

**Ignore Masses**

Ignore Mass 1

Composition

Symbols 13C 1

[Revert to Unimod](#) [Show differences](#)

The chemical composition of the modification as a delta between the modified and unmodified residue or terminus. For example, if the modification removes an H and adds a CH3 group, the Composition would be shown as H(2) C. The formula is displayed and entered as 'atoms', optionally followed by a number in parentheses. The number may be negative and, if there is no number, 1 is assumed. Hence, H(2) C is the same as H(2) C(1)

「Misc」タブは登録エントリーに関してこれまでの設定項目以外のメモ、「その他」となります。

Delta Specificity Ignore Masses **Misc** References

**Misc**

Creation Date 2017-08-17 14:12:08

Last modified 2018-09-13 11:40:08

User unimod

Notes

Alternative name 1 bis[sulfosuccinimidyl] suberate (BS3)

[Revert to Unimod](#) [Show differences](#)

「Reference」タブは、この linkers 情報を設定するにあたり参考にした情報について記入します。

## 13-6. Enzymes

タンパク質からペプチドに切断するパターンに関する設定を行う画面です。Home -> Configuration Editor -> **Enzymes** で開く事ができます。

最初に設定済みの定義一覧が現れます(下図)。既にある設定をそのまま微調整する場合は「**Edit**」のリンクをクリックします。既存のものとは異なる新規の設定を作成するには画面下部の「**Add new enzyme**」ボタンを押します。

Mascot Configuration: Enzymes							
Enzymes							
Title	Sense	Cleave at	Restrict	Independent	Semispecific		
Trypsin	C-Term	KR	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Trypsin/P	C-Term	KR		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Arg-C	C-Term	R	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Asp-N	N-Term	BD		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Asp-N_ambic	N-Term	DE		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Chymotrypsin	C-Term	FLWY	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
CNBr	C-Term	M		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
CNBr+Trypsin	C-Term	M		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
	C-Term	KR	p	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Formic_acid	N-Term	D		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
	C-Term	D					
Lys-C	C-Term	K	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Lys-C/P	C-Term	K		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
LysC+AspN	N-Term	BD		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
	C-Term	K	p	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
Lys-N	N-Term	K		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
PepsinA	C-Term	FL		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
semiTrypsin	C-Term	KR	P	no	yes	<a href="#">Edit</a>	<a href="#">Delete</a>
TrypChymo	C-Term	FKLRWY	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
	N-Term	J					
	C-Term	KR	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
TrypsinMSIPI	C-Term	J					
	C-Term	J					
TrypsinMSIPI/P	N-Term	J		no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
	C-Term	JKR					
V8-DE	C-Term	BDEZ	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
V8-E	C-Term	EZ	P	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
NoCleave	C-Term	J	ABCDEFGHIJKLMNQRSTUUVWXYZ	no	no	<a href="#">Edit</a>	<a href="#">Delete</a>
None							

以降、Trypsin を例に画面について説明いたします。

### Edit Enzyme: Trypsin

**General**

Title

Independent

Semispecific

**Components**

#	Sense	Cleave At	Restrict	Delete
1	C-Term	KR	P	<input type="checkbox"/>

**Test**

Protein :

MSEELSQKPSAQSLSLREGRNRPFLSLSQREGRFFPSLSLSEDRGRKFSFLSMFSLM  
PLLEVIKIIIASVASVIFVGFACVTLGSAALVSTPVFIIFSPVLVPATIAIVLATGFTAG  
GSFGATALGLIMWLVKRRMGVKPKDNPPAGLPPNSGAGAGGAQSLIKKSKAKSKGGL

#	Start	End	Peptide
1	1	1	M
2	1	18	MSEELSQKPS SAQSLSLR
3	2	18	SEELSQKPS SAQSLSLR
4	19	21	EGR
5	22	23	NR
6	24	32	FPFLSLSQR
7	33	35	EGR
8	36	45	FFPSLSLSEER
9	46	48	DGR
10	49	49	K
11	50	67	FSFLSMFSL MPLLEVIK
12	68	140	IIIASVASVI FVGFACVTLA GSAALVST PVFIIFSPVL VPATIAIVVL ATGFTAGGSF GATALGLIMW LVK
13	141	141	R
14	142	142	R
15	143	148	MGVKPK
16	149	172	DNPPAGLPP NSGAGAGGAQ SLIK
17	173	173	K
18	174	175	SK
19	176	177	AK
20	178	179	SK
21	180	183	GGLK
22	184	187	AWCK
23	188	188	K
24	189	191	MLK
25	192	193	SK
26	194	197	FGGK
27	198	198	K
28	199	200	GK

**Title** : 設定名

**Independent** : 2 つ以上の Components が設定に含まれている時、2 つの設定をばらばらに実施してその後混入されたケースを仮定して切断されたパターンを考慮する(チェック)か、それとも同時に切断パターンが有効になるケースを考慮する(チェック無し)か

**Semispecific** : N または C 末端の片側が Components で定義した内容、もう片側が任意の場所で切断する切断方法を考慮するかどうか

**Components** : 切断パターンに関する設定。以下項目を設定します。

- **Sense** : 切断箇所が N 末端か C 末端か
- **Cleave At** : どのアミノ酸残基を認識して切断するか
- **Restrict** : 切断残基の前後(N-term 切断なら前、C-term 切断なら後ろ)に指定残基がある場合切断しない

**Test Enzyme ボタン** : 設定内容によってタンパク質配列が実際にどのように切断されるかのテスト。

切断パターンの Componentns は2つ設定する事もできます。下図は「CNBr+Trypsin」設定の例です。

Title	CNBr+Trypsin			
Independent	<input type="checkbox"/>			
Semispecific	<input type="checkbox"/>			
Components				
#	Sense	Cleave At	Restrict	Delete
1	C-Term	M		<input type="checkbox"/>
2	C-Term	KR	P	<input type="checkbox"/>
Add				Delete

## 13-7. Instruments

プロダクトイオンスペクトルのフラグメントピーク理論値作成を考慮するイオンシリーズに関する設定を行う画面です。Home -> Configuration Editor -> **Instruments** で開く事ができます。

最初に、設定済みの定義一覧が現れます。既にある設定をそのまま微調整する場合は項目名の下にある「**Edit**」リンクをクリックします。既存のものとは異なる新規の設定を作成するには画面下部の「**New Instrument**」ボタンを押します。

Mascot Configuration: Instruments															
Instruments															
Ion series	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF TOF	ESI 4SECTOR	FTMS ECD	ETD TRAP	MALDI QUAD TOF	MALDI QIT TOF	MALDI ISD	CID+ETD	ETHcD
1+	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2+ (precursor>2+)	X	X		X	X	X		X	X	X	X			X	X
2+ (precursor>3+)															
immonium			X				X	X			X	X			
a	X		X				X	X				X	X		X
a*	X		X				X					X			X
a0			X				X					X			X
b	X	X	X	X	X	X	X	X			X	X		X	X
b*	X	X	X	X	X	X	X	X			X	X		X	X
b0		X	X	X	X	X	X	X			X	X		X	X
c									X	X				X	X
x															
y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
y*	X	X		X	X	X	X				X	X		X	X
y0		X		X	X	X	X				X	X		X	X
z								X							
yb							X	X			X	X			
ya							X	X			X	X			
y must be significant															
y must be highest score															
z+1									X	X				X	X
d							X								
v							X								
w							X							X	X
z+2									X	X			X	X	X
Min internal mass															
Max internal mass	700	700	700	700	700	700	700	700	700	700	700	700		700	
	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete
	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit

New Instrument Main menu

MS2 データとのマッチングを考慮するイオンシリーズや電荷についてチェックを入れ、設定作成後に「Save changes」ボタンを押すことで設定が保存されます。

Instruments										
Ion series	New	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF	ESI 4SECTOR	FTM EC
1+	<input checked="" type="checkbox"/>	X	X	X	X	X	X	X	X	X
2+ (precursor>2+)	<input checked="" type="checkbox"/>	X	X		X	X	X		X	X
2+ (precursor>3+)	<input type="checkbox"/>									
immonium	<input type="checkbox"/>			X				X	X	
a	<input type="checkbox"/>	X		X				X	X	
a*	<input type="checkbox"/>	X		X				X		
a0	<input type="checkbox"/>			X				X		
b	<input checked="" type="checkbox"/>	X	X	X	X	X	X	X	X	
b*	<input type="checkbox"/>	X	X	X	X	X	X	X	X	
b0	<input type="checkbox"/>		X	X	X	X	X	X	X	
c	<input type="checkbox"/>									X
x	<input type="checkbox"/>									
y	<input checked="" type="checkbox"/>	X	X	X	X	X	X	X	X	X
y*	<input type="checkbox"/>	X	X		X	X	X	X		
y0	<input type="checkbox"/>		X		X	X	X	X		
z	<input type="checkbox"/>								X	
yb	<input type="checkbox"/>							X	X	
ya	<input type="checkbox"/>							X	X	
y must be significant	<input type="checkbox"/>									
y must be highest score	<input type="checkbox"/>									
z+1	<input type="checkbox"/>									X
d	<input type="checkbox"/>							X		
v	<input type="checkbox"/>							X		
w	<input type="checkbox"/>							X		
z+2	<input type="checkbox"/>									X
Min internal mass	<input type="text" value="0"/>									
Max internal mass	<input type="text" value="0"/>	700	700	700	700	700	700	700	700	700
Instrument name:	<input type="text" value="Customized"/>									
										<input type="button" value="Save changes"/>

Mascot Configuration: Instruments																
Instruments																
Ion series	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF	ESI 4SECTOR	FTMS ECD	ETD TRAP	MALDI QUAD TOF	MALDI QIT TOF	MALDI ISD	CID+ETD	ETHC	Customized
1+	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2+ (precursor>2+)	X	X		X	X	X		X	X	X	X			X	X	X
2+ (precursor>3+)																
immonium			X				X	X			X	X				
a	X		X				X	X				X	X		X	
a*	X		X				X					X			X	
a0			X				X					X			X	
b	X	X	X	X	X	X	X	X			X	X		X	X	X
b*	X	X	X	X	X	X	X	X			X	X		X	X	X
b0		X	X	X	X	X	X	X			X	X		X	X	X
c									X	X				X	X	X
x																
y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
y*	X	X		X	X	X	X				X	X		X	X	X
y0		X		X	X	X	X				X	X		X	X	
z								X								
yb							X	X			X	X				
ya							X	X			X	X				
y must be significant																
y must be highest																

## 13-8. Quantitation

タンパク質の定量解析に関する設定を行う画面です。Home -> Configuration Editor -> **Quantitation** で開く事ができます。

最初に設定済みの定義一覧が現れます(下図)。既にある設定をそのまま微調整する場合は項目名のリンクを、現存設定を残しコピーしてそちらで書き換える場合は「**Copy**」のリンクをクリックします。既存のものとは異なる新規の設定を作成するには画面下部の「**New quantitation method**」ボタンを押します。

以降、「TMT 6 plex」の設定を使って各画面の説明をいたします。

### Mascot Configuration: Quantitation Methods

Quantitation Methods				
Name	Protocol			
None	null			
<a href="#">ITRAQ 4plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ITRAQ 4plex (protein)</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ITRAQ 8plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">TMT 6plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">TMT 2plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">TMT 10plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">TMTpro 16plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">DiLeu 4plex</a>	reporter	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">18O multiplex</a>	multiplex	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+6 R+6 multiplex</a>	multiplex	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">IPTL (Succinyl and IMID) multiplex</a>	multiplex	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICPL duplex pre-digest [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICPL duplex post-digest [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICPL triplex pre-digest [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICPL quadruplex pre-digest [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">18O corrected [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">15N Metabolic [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">15N + 13C Metabolic [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+6 R+10 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+6 R+10 Arg-Pro [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+6 R+6 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC R+6 R+10 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+8 R+10 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">SILAC K+4 K+8 R+6 R+10 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICAT ABI Cleavable [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">ICAT D8 [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">Dimethylation [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">NBS Shimadzu [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">Acetylation [MD]</a>	precursor	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">Label-free [MD]</a>	replicate	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">Average [MD]</a>	average	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>

[New quantitation method](#) [Main menu](#)

Proteome Sciences 6-plex Tandem Mass Tags(R)

画面上部には「Name」と「Description」があり、設定項目の識別情報として利用します。

「Method」タブではタンパク質の定量値を計算する条件やペプチド定量値からタンパク質の定量値を計算する方法に関連する設定項目が集められています。

### Edit Quantitation Method:TMT 6plex

**Name**

Name 
Description

Method
Protocol
Component
Report Ratio
Integration
Quality
Outliers
Normalisation
XML

**Method**

Property	Value	Action
Constrain Search	<input type="checkbox"/>	
Protein Ratio Type	median ▾	
Protein score Type	Define	
Report Detail	<input checked="" type="checkbox"/>	
Show subsets	Define	
Require bold red	<input checked="" type="checkbox"/>	Clear
Minimum Peptides	<input type="text" value="2"/>	
Significance Threshold	Define	
comp qualifier		Add Composition
seq qualifier		Add Sequence
Modification groups	TMT fixed	Delete Add Modification Group

Save changes
Cancel

If any modification group specifies exclusive mode, then apply this constraint during the search so that only matches that can be used for quantitation will be returned.

- Constrain Search** : 「Quality」タブの「Exclusion」項目で設定された修飾について、検索結果から除いてレポートするかどうかを指定します。
- Protein Ratio Type** : アサインペプチドの定量情報からタンパク質の定量値を計算する方法。  
**median, average, weighted**(intensity による重みづけをした average) から選びます。結果画面で変更可能。
- Protein score Type** : タンパク質のスコア算出方法でデフォルト設定からの変更を希望する時に指定。**standard** または **mudpit** から選択します。
- Report Detail** : ペプチドレベルでの定量値を結果画面に表示するかどうか。
- Show subsets** : subset のタンパク質を結果画面に表示するかの設定でデフォルトから変更希望の際に指定。0(全く表示させない)から 1(すべて表示させる)の間の数値。
- Require bold red** : 定量計算対象のペプチドを bold red (同定基準を超えランクが1位) に限定するかどうか
- Minimum Peptides** : タンパク質にアサインされるペプチド数の最低値。結果画面で変更可能。
- Significance Threshold**: ペプチドの同定基準をデフォルト設定から変更したい場合に指定。

- comp qualifier** : ペプチドに含まれるアミノ酸残基数による限定。文法は Sequence Query 法の記述と同じです。例) **\*[C]→C を必ず1つ含む**、など
- seq qualifier** : ペプチドに含まれるアミノ酸配列による限定。文法は Sequence Query 法の記述と同じです。  
例) **\*-TSL →N 末端または C 末端から TSL という並びの配列**、など
- Modification groups** : 定量計算を行う際必ず指定する修飾の組を設定します。図例では「**TMT fixed**」というグループがあり、グループ内で TMT6plex (N-term) と TMT6plex (K) が定義されています。グループの設定画面は以下の通りです。

Property	Value	Action
Name	TMT fixed	
Mode	fixed	
Required	<input type="checkbox"/>	
Modification	TMT6plex (N-term)	Delete
	TMT6plex (K)	Delete
Unmodified		Add unmodified
Local definitions		Add local definition

OK Cancel

- Name** : 修飾グループの名称。
- Mode** : **fixed** または **variable**。
- Required** : 定量計算を行う上で、この修飾が必ず付いていることが求められるかどうか
- Modification** : 修飾設定。
- Unmodified** : 検索対象から除外するアミノ酸残基や N/C 末端。
- Local definitions** : Quantitation 内部だけで使用する修飾を指定可。

「**Protocol**」タブでは、実施する定量解析が MASCOT で定義するどのパターンの定量解析方法かを指定するとともに、その手法独自の関連設定項目が表示されます(下図)。表示例は **reporter** の設定項目ですが、その他の設定値は選択されたプロトコルによって大きく異なります(実際に変更してみると関連する設定項目が切り替わります)。項目名にカーソルを合わせると、下の欄に説明文が現れるので詳しくはそちらをご参照ください。

Property	Value
Protocol	reporter
Reporter Tolerance	Define
Reporter Tolerance Unit	Define

Save changes Cancel

**Protocol** : MASCOT で定義される定量解析プロトコルの選択、以下の項目があります。

- **reporter** : MS2、特定領域に表れる各ピークの強度を定量計算に利用。
- **precursor** : タグなどの影響で質量が異なる同一ペプチドのペアを 1 データ内から探し、同定ペプチドの MS1(XICs)情報を利用して定量計算に利用。Mascot Distiller が必要
- **multiplex** : MS2、同定ペプチドのフラグメントピークの強度を定量計算に利用。
- **replicate** : ラベルフリー定量、複数のデータにまたがった解析。同定ペプチドの MS1(XICs)情報を利用して定量計算に利用。Mascot Distiller が必要。
- **average** : ラベルフリー定量、複数のデータにまたがった解析。各タンパク質の定量値について強度が強い top N(N は整数)のペプチドデータを使い計算するとともに reference との比較で定量値を算出

Method	Protocol	Component	Report Ratio	Integra
<b>Protocol</b>				
Property	Value			
Protocol	average ▾			
Num peptides	3			
Selection	unique_sequence ▾			
Reference Accession				
Reference Database				
Reference Amount	1.0			

右図例は Average プロトコルの場合です。上記と表示項目が異なります。

「**Component**」タブでは、定量計算を行うために定義した各データシリーズ(Component)について、名称や質量などの設定を行います(下図)。こちらも Protocol により設定項目が変化します。以下 reporter の例については各項目について簡単に説明しますが、各プロトコル、各項目の詳細は実際の画面でカーソルを各項目にあわせた際、画面下部の説明欄に表示される内容をご覧ください。

Method	Protocol	Component	Report Ratio	Integration	Quality	Outliers	Normalisation	XML
<b>Component</b>								
Components:	126 ▾		New		Copy		Delete	
Property	Value							Action
Component	126							
M/Z	Monoisotopic : 126.127726		Average : 126.2188					
Corrections	Type:	AB certificate ▾	Shift:	-2				
	Element:	▾		0.0	Delete			
	Type:	AB certificate ▾	Shift:	-1				
	Element:	▾		0.1	Delete			
	Type:	AB certificate ▾	Shift:	1				
	Element:	▾		8.5	Delete			
	Type:	AB certificate ▾	Shift:	2				
	Element:	▾		0.5	Delete		Add correction	

**Components** : 定義済みの「Component」一覧(右図)

**Component** : 選択している Component の名称

**M/Z** : reporter のピークの位置

**Corrections** : 各 component で生じるピークシフトについて、シフトする数値をその割合の設定。

Components:	126 ▾
<b>Property</b>	126
Component	128
M/Z	129
Corrections	130
	131
Element:	

「**Report Ratio**」タブでは結果で表示する比について定義します。**分母や分子に使用される Components を指定**しますが、単純な設定だけでなく、要素同士を足したりする設定も可能です。

Property	Value	Action
Report Ratio	127/126	Delete Report Ratio
Numerator	127 ▼ Coefficient : 1.0	Add numerator
Denominator	126 ▼ Coefficient : 1.0	Add denominator
Report Ratio	128/126	Delete Report Ratio
Numerator	128 ▼ Coefficient : 1.0	Add numerator
Denominator	126 ▼ Coefficient : 1.0	Add denominator
Report Ratio	129/126	Delete Report Ratio
Numerator	129 ▼ Coefficient : 1.0	Add numerator
Denominator	126 ▼ Coefficient : 1.0	Add denominator
Report Ratio	130/126	Delete Report Ratio
Numerator	130 ▼ Coefficient : 1.0	Add numerator
Denominator	126 ▼ Coefficient : 1.0	Add denominator
Report Ratio	131/126	Delete Report Ratio
Numerator	131 ▼ Coefficient : 1.0	Add numerator
Denominator	126 ▼ Coefficient : 1.0	Add denominator

New Report Ratio

Save changes Cancel

■ 複雑な設定については以下 URL をご覧ください。

[https://www.matrixscience.com/help/quant\\_overview\\_help.html#CONCEPTS](https://www.matrixscience.com/help/quant_overview_help.html#CONCEPTS)

■ ラベルフリー定量、replicate プロトコルでの Component や Report Ratio の設定例については以下資料の P.16～ をご覧ください。

[https://www.matrixscience.co.jp/supportpdf/MascotDistiller\\_replicatesQuantTutorial.pdf](https://www.matrixscience.co.jp/supportpdf/MascotDistiller_replicatesQuantTutorial.pdf)

「**Integration**」タブは Precursor の intensity について、同定ペプチドのピーク強度を足し合わせていく際のルールについての設定で、MS1 関連の定量手法で設定が必要な項目です。ペプチドが検出されたポイントを中心に Retention time を前後しながらターゲットとなるペプチドピークの intensity を定量値として足し合わせていく際のルールです。reporter プロトコルではデフォルト設定のように、基本的に「none」設定でご利用頂く事になります(右図)。

Property	Value
Integration Method	none ▼
Integration Source	survey ▼
Simple ratio	<input type="checkbox"/>
Allow Elution Shift	<input type="checkbox"/>
Elution Time Delta	Define
Elution Profile Correlation Threshold	Define
All Charge States	<input type="checkbox"/>
All Charge States Threshold	
Matched Rho	
XIC Threshold	

よってここでは TMT でなく「Label-free[MD]」の設定を使って Integration の各項目を説明します(下図)。

Method	Protocol	Component	Report Ratio	Integration	Quality	Out
<b>Integration</b>						
<b>Property</b>	<b>Value</b>					
Integration Method	simpsons ▾					
Integration Source	survey ▾					
Simple ratio	<input type="checkbox"/>					
Allow Elution Shift	<input checked="" type="checkbox"/>					
Elution Time Delta	500.0	Unit :	seconds ▾	Clear		
Elution Profile Correlation Threshold	999	Clear				
All Charge States	<input type="checkbox"/>					
All Charge States Threshold	0.20					
Matched Rho	0.8					
XIC Threshold	0.1					
XIC Max Width	250					
XIC Smoothing	3					

**Integration Method** : データの統合(積分)方法。

- **none** : 結合しない。
- **simpsons** : シンプソンズ則に基づいた積分。
- **trapezium**: 台形公式に基づいた積分。

**Integration Source** : 何のデータを積分するかについての設定。

- **survey** : survey scan の Precursor ピーク面積。
- **zoom** : zoom scan の Precursor ピーク面積。
- **header** : ヘッダー情報に記載されている XIC の値。
- **fragments**: MS/MS のフラグメントピーク面積の合計。
- **mrm** : MRM, multiple reaction monitoring

**Simple ratio** : チェックが入っている時は合計値がそのまま計算に利用されます。  
チェックされていない場合、近似曲線の勾配を比率とします。

**Allow Elution Shift** : 溶出時間のタイムシフトを容認するか  
(データ間の RT アライメントを実施するかどうか)。

**Elution Time Delta** : アライメントの許容時間。

**Elution Profile Correlation Threshold** :

XIC ピークの各スキャンの成分強度から近似直線を書かせた際、直線と各スキャンとの標準誤差をスキャン選定の閾値として利用。

**All Charge States** : 他の電荷で同定されたペプチドのピークを定量計算で合算するかどうか

**All Charge States Threshold** :

上記「All Charge States」にチェックが入っている場合、最も強度が強い電荷に対してここで設定した割合以上の強度ももつ電荷のピークを合算対象として定量計算に利用。

- Matched Rho** : 同位体クラスターピークの実測値と理論値の相関係数がこの設定値以上の時定量計算に利用。
- XIC Threshold** : XIC 領域最大値に対しての割合。XIC の強度について、計算対象領域は XIC 強度がこの設定値以上の割合の値であることが求められます。
- XIC Max Width** : XIC 計算対象領域の拡張を考慮するスキャン数の上限。
- XIC Smoothing** : スムージングのパラメーターで、設定値を  $n$  とした時  $2n+1$  個の Savitzky-Golay コンボリューション整数のセットによってスムージングを実施 (数字が大きいほど細くなるが計算時間がかかる)。

**Matched Rho** については以下資料も補足資料としてご参照ください。

[https://www.matrixscience.co.jp/supportpdf/MascotDistiller\\_replicatesQuantTutorial.pdf](https://www.matrixscience.co.jp/supportpdf/MascotDistiller_replicatesQuantTutorial.pdf)

**Matched Rho** は P.42～の「Correlation coefficient」に該当します。

「**Quality**」タブは定量計算の対象とするペプチドに関する選定条件です(下図)。

Property	Value	Acti
Minimum precursor charge	1	
Isolated precursor	<input type="checkbox"/>	
Isolated Precursor Threshold	0.5	
Minimum a(1)	0.0	
Peptide Threshold Type	at least homology ▾	
Peptide Threshold Value		
Unique Pepseq	<input type="checkbox"/>	
Exclusion		Add Exclusion

Save changes Cancel

**Minimum precursor charge :**

計算対象とするペプチドの電荷の下限。

**Isolated precursor :**

precursor ピーク付近の全ピーク面積に対して、precursor のクラスターピーク的面積が占める割合を XIC 計算対象とするペプチドの選定の判断に利用するかどうか。Distiller 上の結果画面で「**Fraction**」として表示される値と同じです。

[https://www.matrixscience.co.jp/supportpdf/MascotDistiller\\_replicatesQuantTutorial.pdf](https://www.matrixscience.co.jp/supportpdf/MascotDistiller_replicatesQuantTutorial.pdf)

P.40～ に、図と共に Fraction に関する説明がございます。

**Isolated Precursor Threshold :**

上記「Isolated precursor」で使用する閾値。

**Minimum a(1) :**

最大ピーク強度に対して a(1)ピークの強度として求める割合。

### Peptide Threshold Type :

定量計算対象とするペプチドについて、スコアや期待値、同定基準値を使ったフィルターリング。

### Peptide Threshold Value :

上記「Peptide Threshold Type」でスコアや期待値を選択した際、閾値に使われる数字

**Unique Pepseq** : ユニークペプチドのみを結果に表示。

**Exclusion** : ここで指定した修飾(グループ)がついているペプチドを定量計算から外します。

「**Outliers**」タブはペプチドの定量値からタンパク質の定量値を計算する際、「外れ値」の扱いをどうするかについての設定です(下図)。

### Outlier method : 外れ値を除く方法

- **none** : 外れ値を除く処置を行わない
- **dixons** : Dixon 法
- **auto** : データ数によって Dixon 法または Rosner 法のどちらかを選択
- **grubbs** : Grubbs 法
- **rosners** : Rosner 法

Property	Value
Outlier method	auto
	none
	dixons
	auto
	grubbs
	rosners

「**Normalisation**」タブは数値の normalisation に関する設定です(下図)。最初から異なる事を前提としたデータの解析の場合"none"に、normalisation に使用可能なデータがある場合は以下ご案内に基づきご利用ください。

Property	Value	Action
Normalisation method	none	
Normalisation Peptides		Add Peptide
Normalisation Proteins		Add Protein

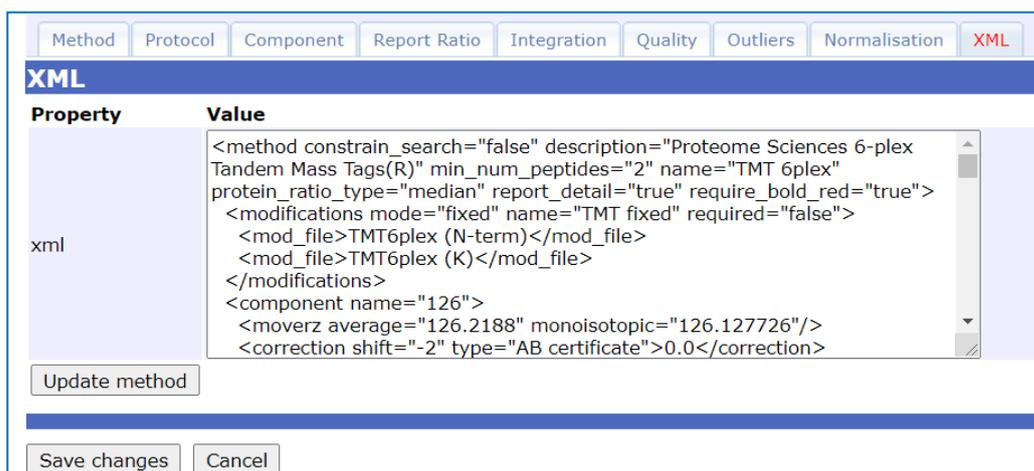
### Normalisation method :

- **none** : normalisation を実施しない。
- **sum** : (reporter プロトコルのみ) MS/MS スペクトルのレポーターイオンの強度の和がサンプル間で同じになる前提とする。
- **average** : ペプチドの ratio についてサンプル間で幾何平均が同じになる前提とする。
- **median** : ペプチドの ratio についてサンプル間で中央値が同じになる前提とする。

**Normalisation Peptides** : Normalisation に使用するペプチド(配列)を指定。

**Normalisation Proteins** : Normalisation に使用するタンパク質(Accession)を指定。

「XML」タブは設定ファイル内で設定内容が記述されている、XML フォーマットを実際に表示して確認するとともに、場合によっては直接編集するための設定欄です。記述変更後「Update method」ボタンを押すことで記述内容が適用されます



## 13-9. Crosslinking

クロスリンクペプチド検索に関する設定を行う画面です。Home -> Configuration Editor -> **Crosslinking** で開く事ができます。

最初に設定済みの定義一覧が現れます。既にある設定をそのまま微調整する場合は項目名のリンクを、現存設定を残しコピーしてそちらで書き換える場合は「**Copy**」のリンクをクリックします。既存のものとは異なる新規の設定を作成するには画面下部の「**New quantitation method**」ボタンを押します。

The screenshot shows a dialog box titled 'Mascot Configuration: Crosslinking Methods'. It has a header 'Crosslinking Methods' and a table with columns for Name, Strategy, Copy, Delete, and Print. Below the table are buttons for 'New crosslinking method' and 'Main menu'.

Name	Strategy	Copy	Delete	Print
None	None			
<a href="#">Disulfide bridge in Lysozyme</a>	Brute-force	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">EDC MND1_ARATH+HOP2_ARATH</a>	Brute-force	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>
<a href="#">HSA Xlink:DSS</a>	Brute-force	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">Print</a>

以降、各項目の設定画面について、「HSA Xlink:DSS」の設定を使って説明をいたします。

画面上部には「Name」と「Description」があり、設定の識別に利用します。

「Method」タブで GUI での設定を行う事ができます(下図)。

- Strategy** : クロスリンクのペアペプチドを探す方法。Ver.2.8 では実施の選択肢が「Brute-force」(総当たり)の 1 択です。
- InterLink** : 異なるタンパク質のペプチド間の結合を考慮するか。
- IntraLink** : 同一種類のタンパク質の異なるペプチド間の結合を考慮するか。
- LoopLink** : 同一種類のタンパク質・同一ペプチド内での結合を考慮するか。
- Linker** : リンカーの種類。リンカー設定より選択。
- Monolink** : 一方のペプチドのみにリンカーが付き実質修飾のようなになるケース。リンカーとして設定していた状態と構造が異なるため、予め「**linkers(13-5)**」内で決めておいた質量変化パターンを Neutral Loss として定義した Code を選択します。Code が何を示しているかは、Linkers の設定を確認する必要があります。
- DoesNotPairWith** : 同時に考慮しない Linker の組み合わせ。
- Accessions**
- Database name** : クロスリンク検索を考慮するタンパク質について、対象のデータベースを特定したい場合にデータベース名を記入して使用します。**Accession** の記載がない場合はデータベースエントリーを網羅的に探索しますが基本的に非推奨です。どうしても利用したい場合はエントリー数が非常に少ない(interlink なら 10 以下、intralink なら 100 以下)、ペプチド組み合わせを考慮するタンパク質の対象が 100 以下となるような選び方ができるデータベースを準備してください。

**Accession** : クロスリンク検索を考慮するタンパク質の Accession を指定

**Filters** : crosslink 検索対象とする候補ペプチドの条件

- **MinPrecursorMr** : 候補ペプチドの質量の最低値
- **MinLen** : 候補ペプチドの残基長の最低値
- **MinCharge** : 候補ペプチドの電荷の最低値

**Settings** : そのほかの設定。現在は以下一つのみ

- **MaxProteins** : 考慮するタンパク質数の最大数。サーバーへの負荷を制御したい場合などに利用

「XML」タブは設定ファイル内で設定項目の内容が記述されている、XML フォーマットを実際に表示して確認し、場合によっては直接編集するための設定欄です(下図)。編集後「**Update method**」ボタンを押すことで編集内容が適用されます。

Property	Value
xml	<pre>&lt;mxm:method description="" name="HSA Xlink:DSS" strategy="Brute-force"&gt;   &lt;mxm:linkers&gt;     &lt;mxm:linker ModFileName="Xlink:DSS (K)"&gt;       &lt;mxm:monolink&gt;A&lt;/mxm:monolink&gt;       &lt;mxm:monolink&gt;W&lt;/mxm:monolink&gt;     &lt;/mxm:linker&gt;     &lt;mxm:linker ModFileName="Xlink:DSS (Protein N-term)"&gt;       &lt;mxm:monolink&gt;A&lt;/mxm:monolink&gt;       &lt;mxm:monolink&gt;W&lt;/mxm:monolink&gt;     &lt;/mxm:linker&gt;   &lt;/mxm:linkers&gt; &lt;/mxm:method&gt;</pre>

## 13-10. Configuration Options

MASCOT Server のオプションに関する設定を行う画面です。Home -> Configuration Editor -> **Configuration Options** で開く事ができます(次頁図)。

パラメーターには空欄の場合 MASCOT Server が予め定めているデフォルト値を適用する場合と、設定値がない場合の 2 パターンがあります。また Configuration Options にデフォルトで表示されていないオプションも存在し、その場合は MASCOT Server が定めるデフォルト値が適用されていますが、その項目名と設定値をこの画面で明示する事で設定値を変更する事もできます。

設定画面の一番下にある「**Add New Options**」を押すと、一覧表示にはなかったオプションについて設定項目名と設定値を入力する欄が現れます。

すべての設定変更が完了した時点で「**Apply**」ボタンを押すことで変更内容が適用されます。

### Configuration Editor: Edit Mascot Options

Detailed descriptions of individual options can be found in Chapter 6 of the [Mascot Setup and Installation manual](#).

To drop an option, clear the value field.

No changes are written to mascot.dat until you choose APPLY.

ProxyType	Auto
proxy_server host	
proxy_server port	
proxy_username	
proxy_password	
UseHTTPProxyForFTP	<input type="radio"/> yes <input checked="" type="radio"/> no <input type="radio"/> clear
UseHTTPProxyForHTTPS	<input type="radio"/> yes <input checked="" type="radio"/> no <input type="radio"/> clear
<input type="button" value="Test Proxy Settings"/>	
SaveLastQueryAsc	<input type="radio"/> yes <input checked="" type="radio"/> no <input type="radio"/> clear
SaveEveryLastQueryAsc	<input type="radio"/> yes <input checked="" type="radio"/> no <input type="radio"/> clear
LastQueryAscFile	../logs/lastquery.asc
InterFileBasePath	C:/inetpub/mascot/data

=====中略=====

DechargeFragmentPeaks	10
PercolatorTargetRankScoreThreshold	20
PercolatorTargetRankRelativeThreshold	0.2
ExecAfterSearch_1	waitfor:0;logging:2;percolator:1, Creating percolator input,../bin/ms-createpip.exe -i %resultfilepath -o %percolator_pip
ExecAfterSearch_2	waitfor:1;logging:2;percolator:1, Percolating, ../bin/percolator.exe \$PercolatorExeFlags
<input type="button" value="Add New Option"/> <input type="button" value="APPLY"/>	

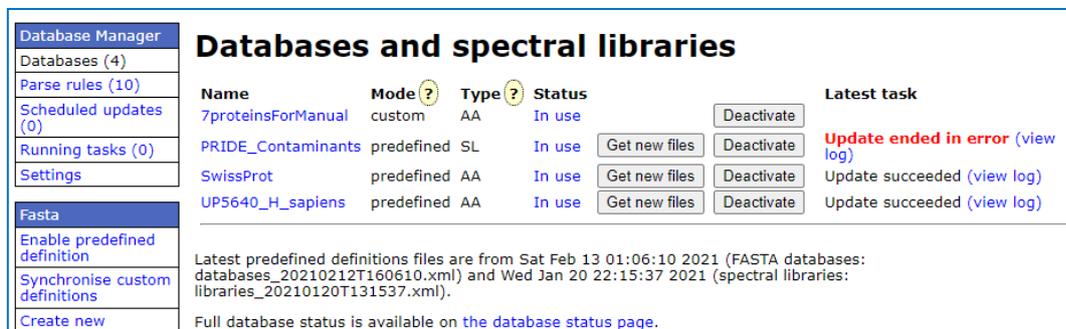
Configuration Options で初期表示されている/いない に関係なく、MASCOT Server で使用可能なオプションがあります。詳細は別紙の日本語資料資料

URL： (資料未完成、完成後公開いたします)

または、**MASCOT**の **Setup & Installation manual** (Home 画面にあるリンクで開く事ができる PDF ファイル、英語)の **6 章、「Options」**の項目をご覧ください。

## 13-11. Database manager

MASCOT で使用しているデータベースに関する設定を行う画面です。Home -> Configuration Editor -> **Database Manager** で開く事ができます。



**Database Manager**

Databases (4)

Parse rules (10)

Scheduled updates (0)

Running tasks (0)

Settings

**Fasta**

Enable predefined definition

Synchronise custom definitions

Create new

### Databases and spectral libraries

Name	Mode ?	Type ?	Status	Latest task
7proteinsForManual	custom	AA	In use	Deactivate
PRIDE_Contaminants	predefined	SL	In use	Get new files Deactivate Update ended in error (view log)
SwissProt	predefined	AA	In use	Get new files Deactivate Update succeeded (view log)
UP5640_H_sapiens	predefined	AA	In use	Get new files Deactivate Update succeeded (view log)

Latest predefined definitions files are from Sat Feb 13 01:06:10 2021 (FASTA databases: databases\_20210212T160610.xml) and Wed Jan 20 22:15:37 2021 (spectral libraries: libraries\_20210120T131537.xml).

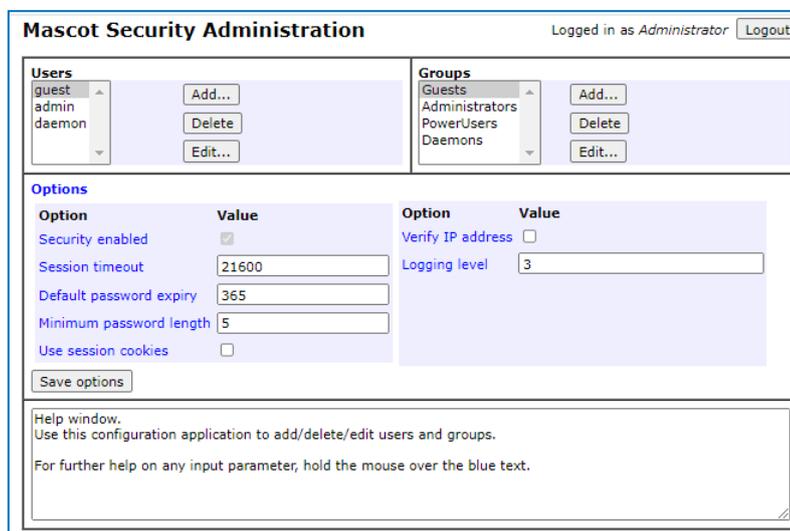
Full database status is available on [the database status page](#).

Database manager については別紙の設定資料を準備しておりますのでそちらをご覧ください。

[https://www.matrixscience.co.jp/supportpdf/MASCOTServer\\_ver26\\_sequencedbmanage.pdf](https://www.matrixscience.co.jp/supportpdf/MASCOTServer_ver26_sequencedbmanage.pdf)

## 13-12. Security

セキュリティ機能に関する設定を行う画面です。MASCOT Server で Security 機能を ON にしている時のみ、Home -> Configuration Editor -> **Security** で開く事ができます。



**Mascot Security Administration** Logged in as Administrator Logout

**Users**

guest admin daemon

Add... Delete Edit...

**Groups**

Guests Administrators PowerUsers Daemons

Add... Delete Edit...

**Options**

Option	Value	Option	Value
Security enabled	<input checked="" type="checkbox"/>	Verify IP address	<input type="checkbox"/>
Session timeout	21600	Logging level	3
Default password expiry	365		
Minimum password length	5		
Use session cookies	<input type="checkbox"/>		

Save options

Help window.  
Use this configuration application to add/delete/edit users and groups.  
For further help on any input parameter, hold the mouse over the blue text.

セキュリティ機能については別紙の設定資料を準備しておりますのでそちらをご覧ください。

<https://www.matrixscience.co.jp/supportpdf/Security.pdf>



## 技術サポート

ご質問等ありましたら弊社技術サポートにご連絡ください。

電子メール : [support-jp@matrixscience.com](mailto:support-jp@matrixscience.com)

電 話 : 03-5807-7897

