

Mascot Server

日本語マニュアル



MASCOT 対応 Version : 3.1

最終更新日 : 2025/3/1



目次

| | |
|---|-----------|
| 1.はじめに: MASCOT Server でできる事 | 1 |
| 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 |
| 2. Mascot Server のシステム | 4 |
| 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 |
| 3. MASCOT Server の入力データ | 7 |
| 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 |
| 4. MASCOT Server 検索方法 | 12 |
| 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 |
| 5. 検索パラメーターとデータベース | 18 |
| 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 | 32 |
| 5-5. Database | 33 |
| 5-6. Search form defaults | 35 |
| 6. 検索結果画面:PMF | 36 |
| 6-1. 表示例で使用している検索について | 36 |
| 6-2. 表示内容の詳細 : summary 画面..... | 36 |

| | |
|--|-----------|
| 6-2-1. 展開しない状態での画面概要..... | 36 |
| 6-2-2. ヘッダー部分..... | 38 |
| 6-2-3. Score Histogram | 38 |
| 6-2-4. 表示内容の変更 [Format As] | 39 |
| 6-2-5. 再検索 : Research all, search unmatched..... | 40 |
| 6-2-6. 同定タンパク質の情報..... | 40 |
| 6-2-7. Search parameters | 41 |
| 6-3. Protein View..... | 41 |
| 6-4. 結果のファイル出力 | 46 |
| 7. 検索結果画面:MIS..... | 47 |
| 7-1. 表示例で使用している検索について | 47 |
| 7-2. 表示内容の詳細 : summary 画面..... | 47 |
| 7-2-1. 展開しない状態での画面概要..... | 47 |
| 7-2-2. ヘッダー部分..... | 49 |
| 7-2-3. 再検索と結果ファイル出力: Re-search ボタン、Export ボタン..... | 49 |
| 7-2-4. Search parameters | 49 |
| 7-2-5. Score Distribution | 50 |
| 7-2-6. Modification statistics for all protein families..... | 51 |
| 7-2-7. Legend (凡例)..... | 51 |
| 7-2-8. 表示内容の切り替え [スコア足切り, refinement の再実行など] | 51 |
| 7-2-9. Sensitivity and FDR | 53 |
| 7-2-10. machine learning quality report | 53 |
| 7-2-11. 同定タンパク質とアサインペプチド..... | 57 |
| 7-2-12. Report Builder タブ (タンパク質ベースの検索結果ファイル出力) | 60 |
| 7-2-13. Unassigned タブ | 62 |
| 7-3. Protein View..... | 62 |
| 7-4. Peptide View | 66 |
| 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 | 71 |
| 8. PMF : タンパク質同定..... | 74 |
| 8-1. PMF タンパク質同定のまとめ | 74 |
| 8-2. 入力データの調整..... | 74 |
| 8-3. 配列から計算される理論ピーク値 | 75 |
| 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..... | 76 |
| 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 | 77 |
| 9. MIS : ペプチド同定とタンパク質同定..... | 79 |
| 9-1. MIS ペプチド同定とタンパク質同定のまとめ..... | 79 |
| 9-2. 入力データの準備..... | 79 |
| 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..... | 80 |

| | |
|---|------------|
| 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 | 81 |
| 9-4-1. <i>refinement</i> を実施しない場合のスコアや同定基準値 | 81 |
| 9-4-2 <i>refinement</i> を実施する場合のスコアや同定基準値 | 82 |
| 9-5. 同定ペプチドから導き出される同定タンパク質 | 82 |
| 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 | 82 |
| 9-5-2. 1 Hit wonders : 同定タンパク質の <i>Sensitivity</i> と <i>Specificity</i> | 83 |
| 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 | 83 |
| 10. MASCOT 検索のオプション [MIS] | 86 |
| 10-1. Spectral Library | 86 |
| 10-1-1. Spectral Library 概要 | 86 |
| 10-1-2. Spectral Library 検索方法 | 87 |
| 10-1-3. Spectral Library の検索結果 | 88 |
| 10-1-4. Spectral Library をローカルの MASCOT Server にセットする方法 | 89 |
| 10-1-5. Spectral Library 補足説明資料へのリンク | 89 |
| 10-2. Quantitation | 90 |
| 10-2-1. Quantitation 概要 | 90 |
| 10-2-2. Quantitation 検索方法 | 90 |
| 10-2-3. Quantitation 検索結果 | 91 |
| 10-2-4. Quantitation 設定の作成 | 93 |
| 10-2-5. Quantitation 補足説明へのリンク | 93 |
| 10-3. Crosslink | 94 |
| 10-3-1. Crosslink 検索 概要 | 94 |
| 10-3-2. Crosslink 検索を行う方法 | 95 |
| 10-3-3. Crosslink 検索に際し注意すべき MASCOT 設定値 | 95 |
| 10-3-4. Crosslink 検索結果 | 96 |
| 10-3-5. Crosslinking 設定の作成 | 99 |
| 10-3-6. Crosslink 補足説明へのリンク | 100 |
| 10-4. Error Tolerant Search | 100 |
| 10-4-1. Error Tolerant Search 概要 | 100 |
| 10-4-2. Error Tolerant Search を実行する方法 | 101 |
| 10-4-3. Error Tolerant Search 検索結果 | 101 |
| 11. 機械学習による結果の精査 (refinement) | 104 |
| 11-1. 機械学習で行う結果の精査(refinement)の概要 | 104 |
| 11-2. refinement 実施の必要要件とデータの流れ | 105 |
| 11-2-1. <i>refinement</i> 計算が実施可能な条件 | 105 |
| 11-2-2. <i>refinement</i> 計算のワークフロー | 106 |
| 11-2-3. decoy データベース | 107 |
| 11-2-4. FDR (<i>q-value</i>) | 107 |
| 11-4-5. Protein FDR | 109 |

| | |
|--|------------|
| 11-3. features | 109 |
| 11-4. 保持時間予測と MS2 スペクトル予測 | 112 |
| 11-4-1. <i>MS²Rescore</i> | 112 |
| 11-4-2. <i>DeepLC</i> | 112 |
| 11-4-3. <i>MS²PIP</i> | 115 |
| 11-5. refinement に関するレポート | 116 |
| 12. MASCOT Server 管理 – データベースと検索ログ | 117 |
| 12-1. 現在利用可能なデータベースに関する情報 : Database Status | 117 |
| 12-2. 検索ログ : Search log | 122 |
| 13. MASCOT Server のカスタマイズ | 124 |
| 13-1. カスタマイズは Configuration Editor で..... | 124 |
| 13-2. Amino Acids | 125 |
| 13-3. Modifications | 126 |
| 13-4. Symbols | 132 |
| 13-5. Linkers | 133 |
| 13-6. Enzymes | 137 |
| 13-7. Instruments | 139 |
| 13-8. Quantitation | 141 |
| 13-9. Crosslinking | 149 |
| 13-10. Configuration Options | 151 |
| 13-11. Database manager | 153 |
| 13-12. Security | 153 |

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しながら、スコア以外の要素を組み込んだ再スコアリングを行って同定ペプチド数を最大限増やす”refinement”を実施する事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 | 1400 | 28266 | 47 | 47 | 7 | 7 | 2.86 | sp TRY1_BOVIN |
| 2 | 1 | SwissProt | #2::CP2CT_MOUSE | 1332 | 61419 | 76 | 76 | 13 | 13 | 2.00 | Cytochrome P450 2C29 OS=Mus musculus OX=10090 |
| 2 | 2 | SwissProt | #2::CP234_MOUSE | 550 | 60887 | 27 | 27 | 8 | 8 | 0.88 | Cytochrome P450 2C54 OS=Mus musculus OX=10090 |
| 2 | 3 | SwissProt | #2::CY250_MOUSE | 487 | 61128 | 27 | 27 | 10 | 10 | 1.20 | Cytochrome P450 2C50 OS=Mus musculus OX=10090 |
| 2 | 4 | SwissProt | #2::CP2F2_MOUSE | 470 | 59267 | 32 | 32 | 12 | 12 | 2.11 | Cytochrome P450 2F2 OS=Mus musculus OX=10090 G |
| 2 | 5 | SwissProt | #2::CP237_MOUSE | 338 | 60590 | 22 | 22 | 8 | 8 | 0.89 | Cytochrome P450 2C37 OS=Mus musculus OX=10090 |
| 2 | 6 | SwissProt | #2::CP239_MOUSE | 251 | 60856 | 13 | 13 | 4 | 4 | 0.37 | Cytochrome P450 2C39 OS=Mus musculus OX=10090 |
| 2 | 7 | SwissProt | #2::CP238_MOUSE | 211 | 60856 | 13 | 13 | 4 | 4 | 0.37 | Cytochrome P450 2C38 OS=Mus musculus OX=10090 |
| 3 | 3 | SwissProt | #2::HS71L_MOUSE | 188 | 78552 | 13 | 13 | 4 | 4 | 0.28 | Heat shock 70 kDa protein 1-like OS=Mus musculus OX=10090 |
| 4 | 1 | SwissProt | #2::CYB5_MOUSE | 1202 | 16817 | 42 | 42 | 5 | 5 | 3.08 | Cytochrome b5 OS=Mus musculus OX=10090 GN=Cyt |
| 5 | 1 | SwissProt | #2::PDIA1_MOUSE | 1121 | 16464 | 53 | 53 | 16 | 16 | 2.54 | Protein disulfide-isomerase OS=Mus musculus OX=10090 |
| 6 | 1 | SwissProt | #2::CP1A2_MOUSE | 1053 | 63034 | 38 | 38 | 10 | 10 | 1.31 | Cytochrome P450 1A2 OS=Mus musculus OX=10090 C |
| Z | 1 | SwissProt | #2::ENPL_MOUSE | 1008 | 103744 | 63 | 63 | 19 | 19 | 1.53 | Endoplasmic OS=Mus musculus OX=10090 GN=Hsp90 |

同定タンパク質のリスト (限定定的だが) 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 SQSFTNFSKA YGPFVTLVIG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAELIKGF GVVFSNGNRW KEMRFRLTML LRLVGMGRKN IEDRVQREAR
151 CLVEELRLRRTK GSPCDPTFLL SCAPCNVICV IIFQNRFDYK DREFLLIMDR
201 INENVKRLSS FWLQVCSNFF SLIDYCPQSH HKIVNRFNYL KSYLLEKIKE
251 HRESLDVTFN RDFDIYLLIK QRQVNHIEQS EFSLENLAST INDLFGAGTE
301 TTSTLRYAL LLLLKYDVT AKVQREIDRV VGRHRSFCMQ DRSDHPEYDA
351 MIHVQRVID LLPTSLPHAV TCDIKFRKYL IFRKTTVITS LSSVLHDSKE
401 FENFEMFDPG HFLNGNGNFK KSDYMPFST GKRICAGEQL ARMEFLILLT
451 TILQNFKLKS LVHFREIDIT FVMDGFASLP FFFYQCFIPL

```

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

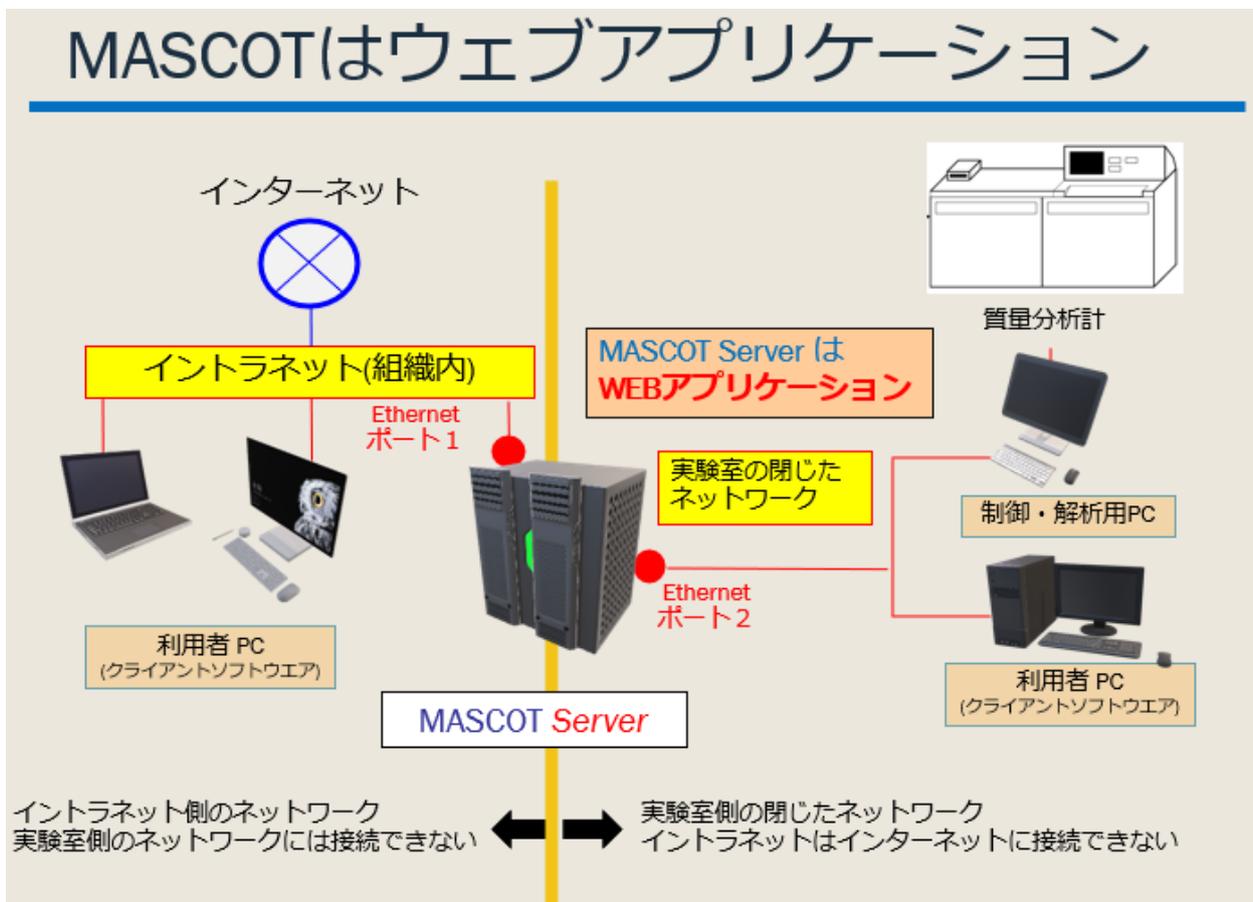
MIS : 1,2,3,4,5,6,7,8,9,10,11,12,13

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 を、結果の解釈、結果画面の見方については 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

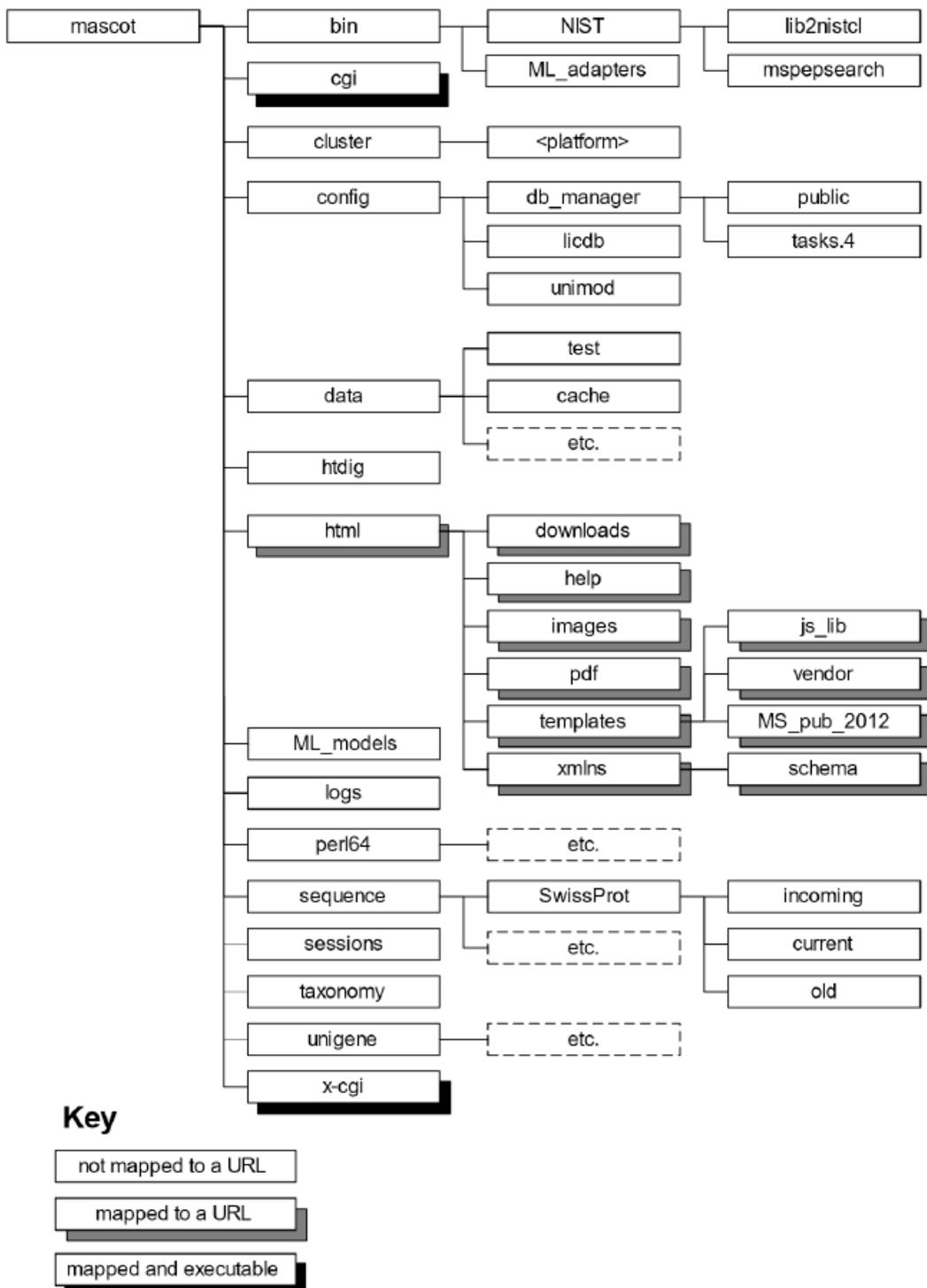
MASCOT Server は **WEB アプリケーション**で、**クライアント-サーバーシステム**を採用しています。HTTP プロトコルを介してデータのやり取りを行います。ユーザーは、MASCOT Server プログラムがインストールされているコンピューター、あるいはネットワーク上の離れたコンピューター、どちらからでも MASCOT Server を利用する事が可能です。MASCOT Server をインストールするマシンには、WEBサーバー機能を提供するプログラム(通常は Windows なら IIS, Linux なら Apache)をインストール・稼働させる必要があるほか、コン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
- ML_models** : 機械学習 refinement を実施するプログラムを格納
- 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場合、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。

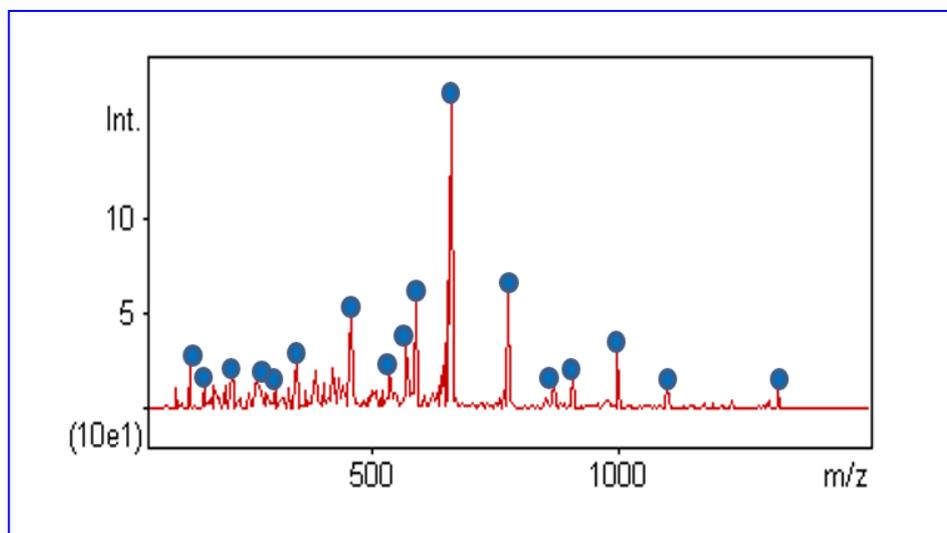
3-1 では検索に投げる前にデータの処理が必要な理由と MASCOT Server が受け付けるファイルフォーマットについて、3-2 では raw データから MASCOT で読み込みが可能なファイルフォーマットに変換するプログラムについてご紹介します。

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ファイルの拡張子に関係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 m/z の次に記載されている場合、**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検索時のピークリストが後ろに **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た場合、結果が異なるケースがあります。

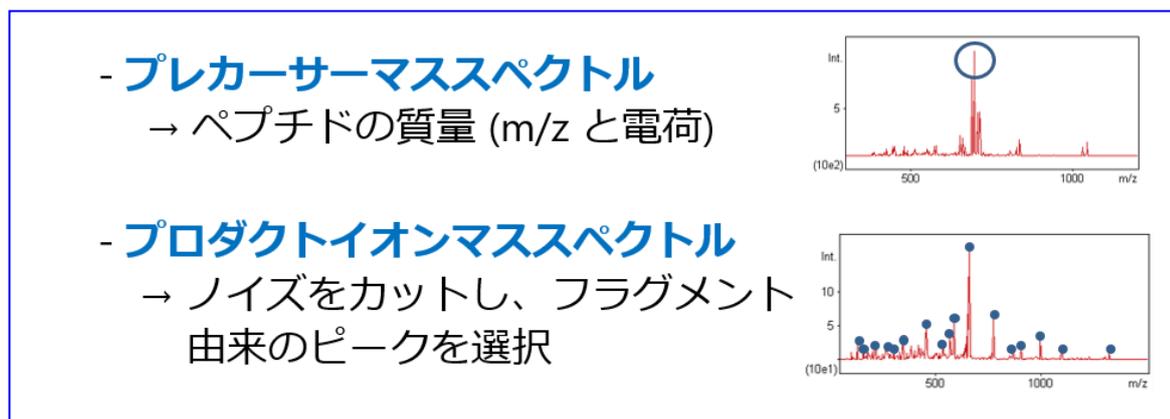
| | | |
|----------|-----------|-----------|
| 832.662 | 361.21774 | 4838.6552 |
| 903.342 | 487.26656 | 5281.3009 |
| 1186.439 | 494.28934 | 8868.6732 |
| 1373.681 | 505.77755 | 16079.047 |
| 1403.722 | 686.36334 | 22677.156 |
| 1515.444 | 723.34797 | 36555.65 |
| 1727.916 | 836.78635 | 6731.0498 |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

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
```

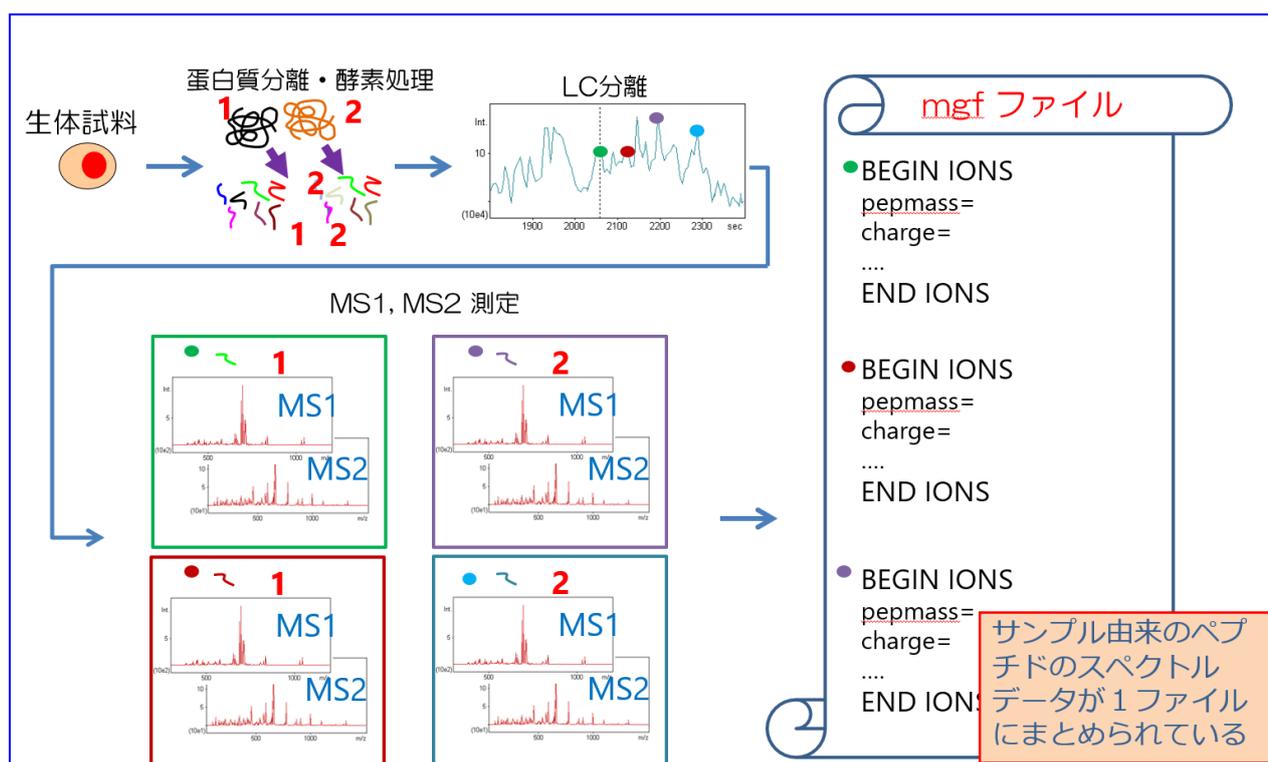
① ペプチドの質量を計算
(817.44-1.0067)*2 = 1632.866

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 の後にさら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

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)

特に、DeepLC を利用した予測保持時間情報をペプチド同定に応用する方法や、多価のフラグメントピークを検索に応用したい crosslinking、Topdown 解析などに有用です。

■ 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の概要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。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 a web browser window displaying the MASCOT Peptide Mass Fingerprint search page. The browser address bar shows the URL: localhost/mascot/search_form_select_windows.html. The page header includes the Matrix Science logo and navigation links: Home, Access Mascot Server, Database search help, and Contact. The main content area is titled "MASCOT Peptide Mass Fingerprint" and contains a search form with the following fields and options:

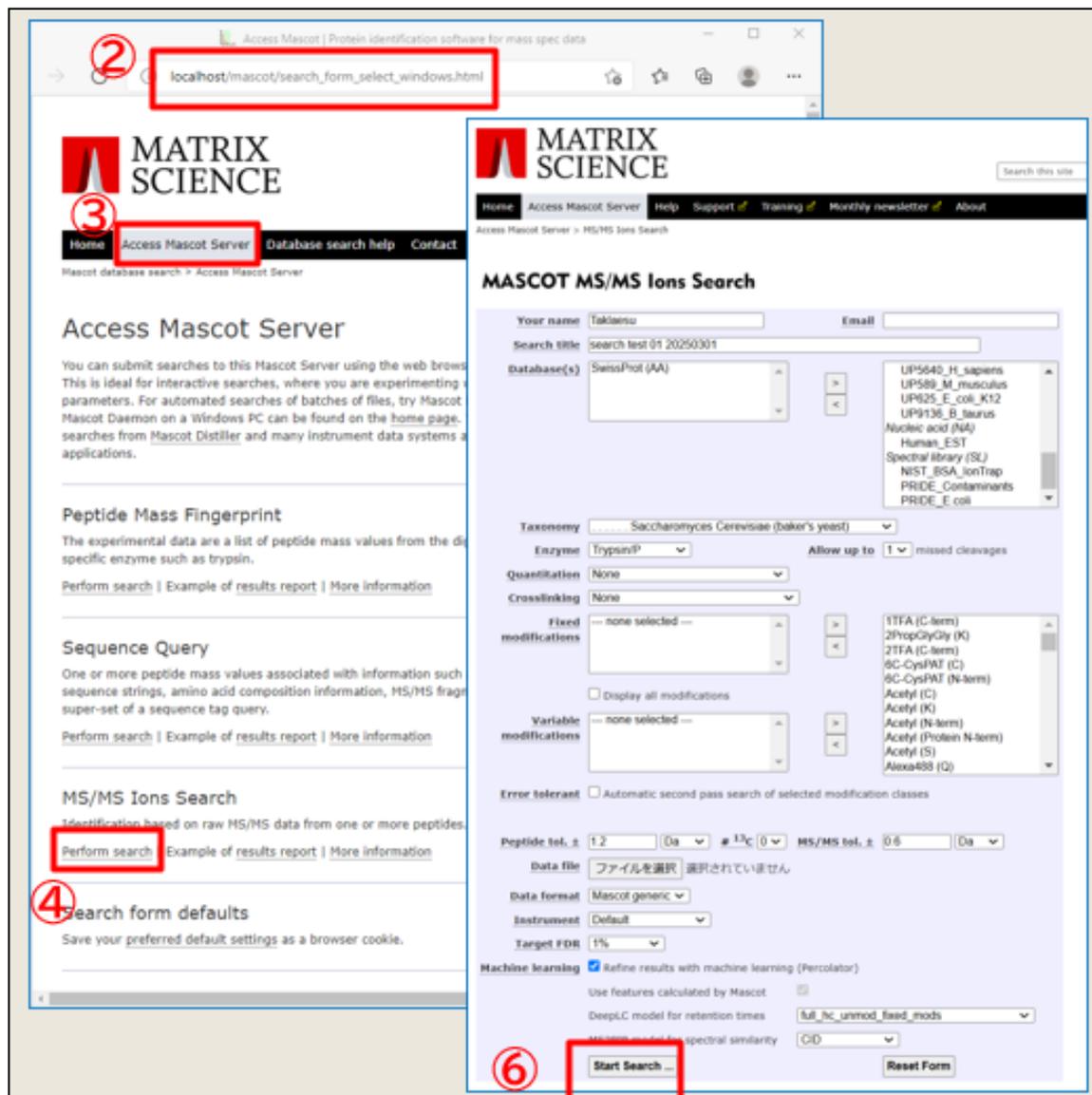
- Your name** and **Email** input fields.
- Search title** input field.
- Database(s)** dropdown menu with options: Human_EST, 2sequencesT, 7proteinsForManual, cRAP, and ssbAmong2proteins.
- Enzyme** dropdown menu set to Trypsin.
- Allow up to** dropdown menu set to 1 missed cleavages.
- Taxonomy** dropdown menu set to All entries.
- Fixed modifications** dropdown menu set to --- none selected ---.
- Variable modifications** dropdown menu set to --- none selected ---.
- Protein mass** input field (kDa).
- Peptide tol. ±** input field (1.2) and unit dropdown (Da).
- Mass values** radio buttons for MH⁺, M, and M-H⁻.
- Monoisotopic** and **Average** radio buttons.
- Data file** radio button selected, with a "ファイルの選択" button and a message "ファイルが選択されていません".
- Query** radio button.
- Data input** text area.
- Decoy** checkbox.
- Report top** dropdown menu set to AUTO.
- Start Search ...** button.
- Reset Form** button.

Numbered annotations (1-6) are present on the image:

- ①: Not explicitly shown in the screenshot.
- ②: Points to the browser address bar.
- ③: Points to the "Access Mascot Server" navigation link.
- ④: Points to the "Perform search" link under the "Peptide Mass Fingerprint" section.
- ⑤: Points to the "MASCOT Peptide Mass Fingerprint" title.
- ⑥: Points to the "Start Search ..." button.

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_ver30_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_ver30_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。MGF か mzML かは、上記3で設定したパラメータファイルの file format 項目に合わせる。
- 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

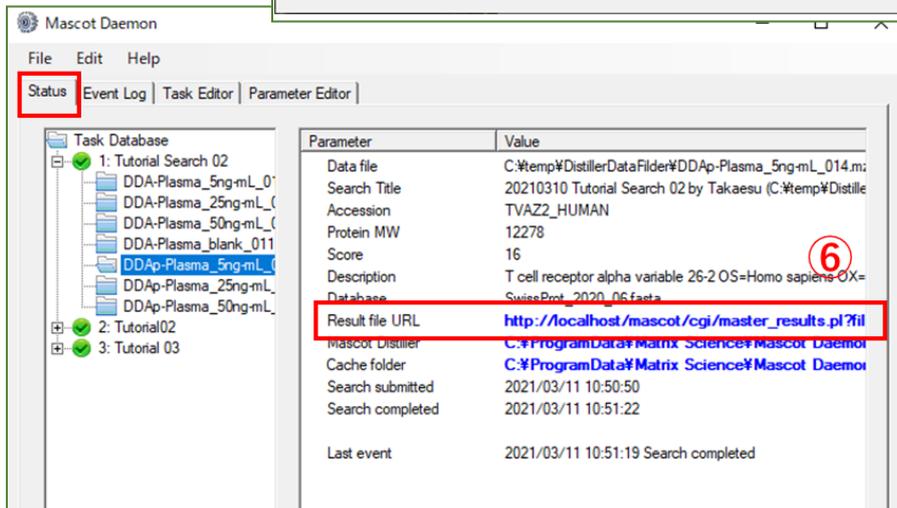
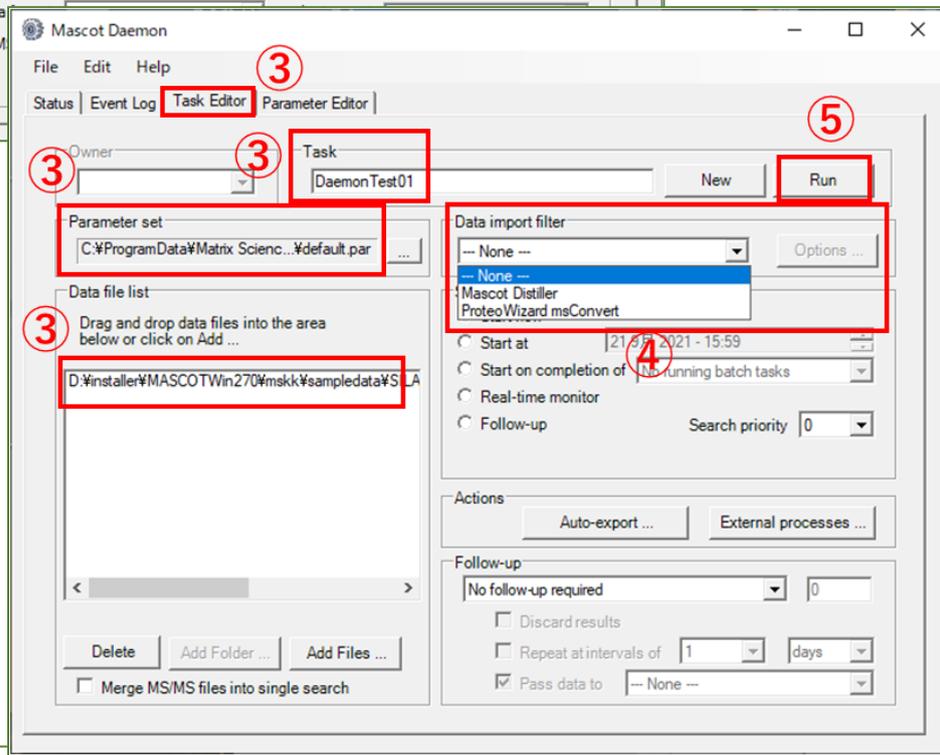
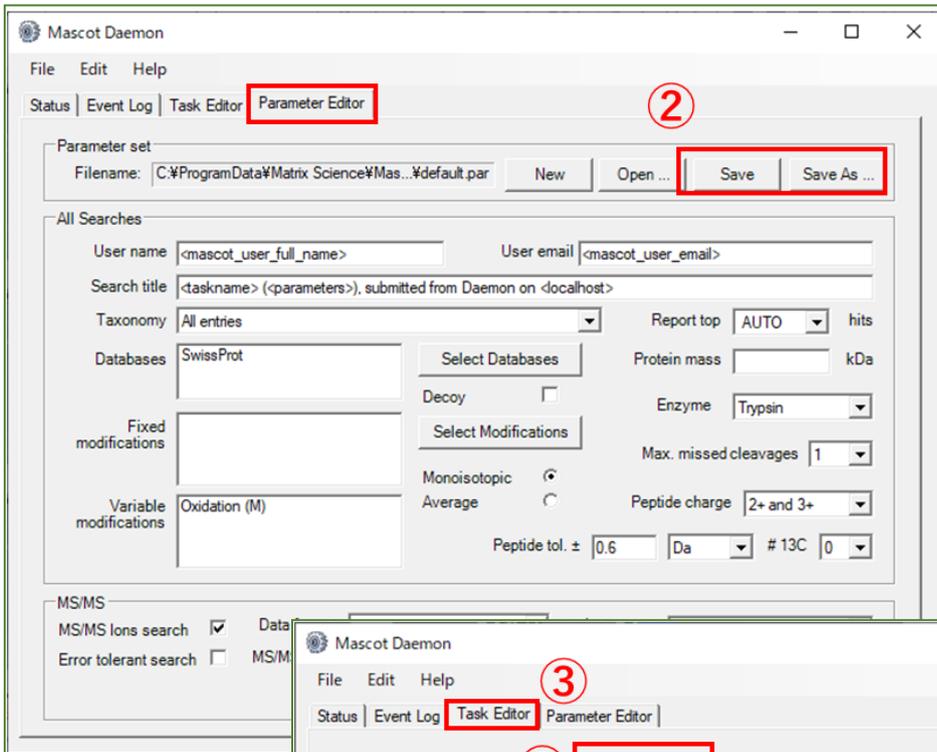
①～⑥の手順について理解の一助となる図が次頁にございますのでそちらも併せてご参照ください。

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ーセットを選択してから「OK」ボタンを押す
- 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頂く事を推奨**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 の検索結果の数値だけが装置メーカーの解析ソフトウェアに取り込まれ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)**
SARS-CoV-2
SwissProt
test2
Trembl

(5) **Enzyme**

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

(7) **Taxonomy**

(8) **Fixed modifications**
 Display all modifications

(9) **Variable modifications**
1TFA (C-term)
2PropGlyGly (K)
2TFA (C-term)
6C-CysPAT (C)
6C-CysPAT (N-term)
Acetyl (C)
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Acetyl (S)
Alexa488 (Q)

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

(12) **Mass values** MH⁺ M_r M-H⁻

(13) **Data input** Data file 選択されていません
 Query

(14) (15)

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 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 |
|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> |
| 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 N 回見逃し連結したペプチドを作成する事ができます。例えば Trypsin 設定では 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) 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、エラーメッセージが出ますが検索は問題なく実行する事ができます。

リストに記載されている生物種について、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思えますが、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。

修飾を指定する場合、右側にあるリストから該当項目を選び、真ん中にある < ボタンを押して Fixed または Variable modification 設定項目として指定します。

| | | | | |
|----------------------------|-----------------------|---|---|---------------------------|
| Fixed modifications | --- none selected --- | > | < | 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

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、MASCOT に登録されているすべての修飾がリストに表示されます。

(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。従って検索時のフィルターではなく、検索結果表示時のフィルターとして機能します。

(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值を判断します。

(12) Mass values

PMF 検索においてクエリーの各ピークが MH⁺か M-H⁻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) 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
```

(14) Start Search

検索を開始します。

(15) 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) Human_EST, contaminants, cRAP, IPI_human, Mouse

(5) Enzyme Trypsin

(6) Allow up to 1 missed cleavages

(7) Quantitation None

(8) Taxonomy All entries

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

Display all modifications

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

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

(12) Peptide charge Mr

(13) MS/MS tol. ± 0.6 Da

(14) Peptide charge Mr

(15) Query

(16) Instrument Default

(17) Start Search ... (18) Reset Form

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

次頁図にあるように、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)
Timestamp : 12 Aug 2021 at 02:50:53 GMT

MASCOT search log

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

| Job# | PID | dbase | User Name | Email | Title |
|----------------------|-------|----------|-------------------|-------|---|
| 1242 | 12096 | SwissPro | | | 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 | | | 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 N 回見逃し連結したペプチドを作成する事ができます。例えば Trypsin 設定では 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**

NRFPFLSLSQ**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。

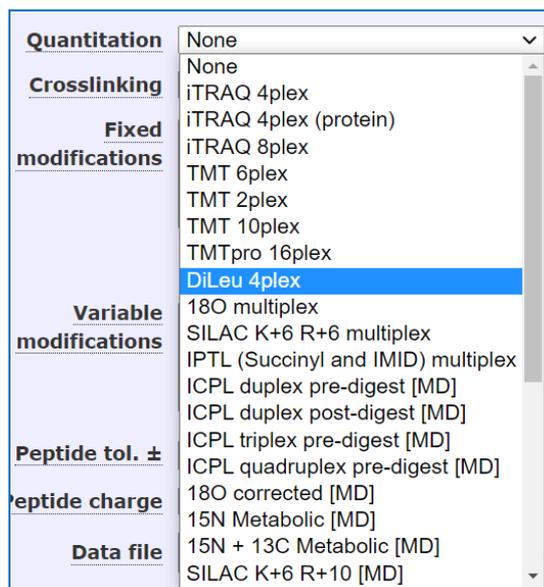
(7) Quantitation

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。

この設定は利用の際、事前に設定を準備する必要があります。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。

[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(計算モジュール搭載)が必要です。どちらの手法も、定量計算を実施するためのサンプル測定と、MASCOT での事前の設定項目の作成が必要になります。

詳細は「**10-2.Quantitation**」をご覧ください。



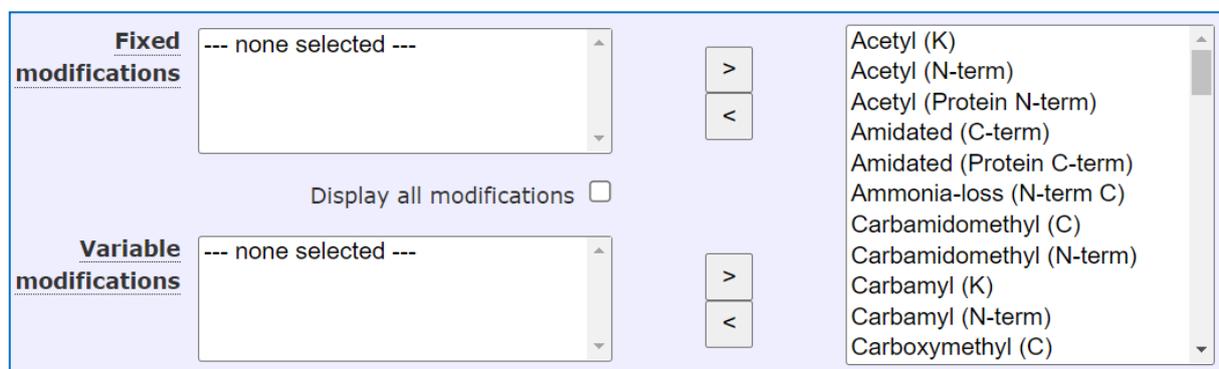
(8) Taxonomy

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、エラーメッセージが出ますが検索は問題なく実行する事ができます。

リストに記載されている生物種について、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思えますが、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。修飾を指定する場合、右側にあるリストから該当項目を選び、真ん中にある < ボタンを押して Fixed または Variable modification 設定項目として指定します。



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、MASCOT に登録されているすべての修飾がリストに表示されます。

(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) #¹³C

質量分析装置での測定時、¹²C のみから構成されるペプチドではなく ¹³C 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(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+ などの値が記入されていることが多いため、このパラメーターを使用しなければならないケースはあまりありません。

(15) 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

または

http://localhost/mascot/help/sq_help.html

をご覧ください。

(16) Instrument

フラグメントピークと理論値をマッチングする際、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(次頁図)です。MS/MS において発生するフラグメントパターンの内容に応じて選択して下さい。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 | CID+ ETD | ETchD | EAD |
|--|---------|--------------|---------------|----------|----------|-----------|---------------|------------|----------|----------|----------------|---------------|-----------|----------|-------|-----|
| <u>1</u> ⁺ | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| <u>2</u> ⁺ (pre cursor >2 ⁺) | X | X | | X | X | X | | X | X | X | X | | | X | X | X |
| <u>2</u> ⁺ (pre cursor >3 ⁺) | | | | | | | | | | | | | | | | |
| <u>imm.</u> | | | X | | | | X | X | | | X | X | | | | |
| <u>a</u> | X | | X | | | | X | X | | | | X | X | | X | X |
| <u>a*</u> | X | | X | | | | X | | | | | X | | | X | |
| <u>a0</u> | | | X | | | | X | | | | | X | | | X | |
| <u>b</u> | X | X | X | X | X | X | X | X | | | X | X | | X | X | X |
| <u>b*</u> | X | X | X | X | X | X | X | X | | | X | X | | X | X | |
| <u>b0</u> | | X | X | X | X | X | X | X | | | X | X | | X | X | |
| <u>c</u> | | | | | | | | | X | X | | | X | X | X | X |
| <u>x</u> | | | | | | | | | | | | | | | | |
| <u>y</u> | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| <u>y*</u> | X | X | | X | X | X | X | | | | X | X | | X | X | |
| <u>y0</u> | | X | | X | X | X | X | | | | X | X | | X | X | |
| <u>z</u> | | | | | | | | X | | | | | | | | |
| <u>z+1</u> | | | | | | | | | X | X | | | | X | X | X |
| <u>z+2</u> | | | | | | | | | X | X | | | X | X | X | X |
| <u>yb</u> | | | | | | | X | X | | | X | X | | | | |
| <u>ya</u> | | | | | | | X | X | | | X | X | | | | |
| <u>y must be sig.</u> | | | | | | | | | | | | | | | | |
| <u>y must be highest</u> | | | | | | | | | | | | | | | | |
| <u>d</u> | | | | | | | X | | | | | | | | | |
| <u>v</u> | | | | | | | X | | | | | | | | | |
| <u>w</u> | | | | | | | X | | | | | | | X | X | X |

(17) Start Search

検索を開始します。

(18) 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



Search this site

Home
Access Mascot Server
Help
Support
Training
Monthly newsletter
About

Access Mascot Server > MS/MS Ions Search

MASCOT MS/MS Ions Search

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

(3) **Search title**

(4) **Database(s)**

UP5640_H_sapiens
UP589_M_musculus
UP625_E_coli_K12
UP9136_B_taurus
Nucleic acid (NA)
Human_EST
Spectral library (SL)
NIST_BSA_IonTrap
PRIDE_Contaminants
PRIDE_E.coli

(5) **Taxonomy**

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

(8) **Quantitation**

(9) **Crosslinking**

(10) **Fixed modifications**

1TFA (C-term)
2PropGlyGly (K)
2TFA (C-term)
6C-CysPAT (C)
6C-CysPAT (N-term)
Acetyl (C)
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Acetyl (S)
Alexa488 (Q)

 Display all modifications

(11) **Variable modifications**

1TFA (C-term)
2PropGlyGly (K)
2TFA (C-term)
6C-CysPAT (C)
6C-CysPAT (N-term)
Acetyl (C)
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Acetyl (S)
Alexa488 (Q)

(12) **Error tolerant** Automatic second pass search of selected modification classes

(13) **Peptide tol. ±** (14) **# ¹³C** (15) **MS/MS tol. ±**

(16) **Data file**

(17) **Data format**

(18) **Instrument**

(19) **Target FDR**

Machine learning Refine results with machine learning (Percolator)

(20) Use features calculated by Mascot

(21) DeepLC model for retention times

(22) MS2PIP model for spectral similarity

(23) (24)

(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 | 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。データベースは大きく分けると以下の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。

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**

NRFPFLSLSQ**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。

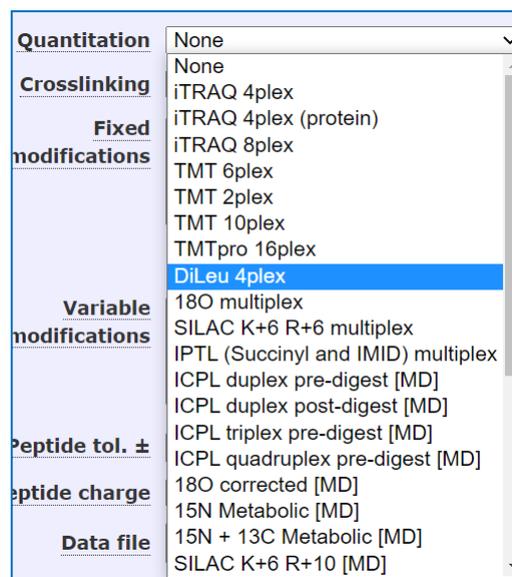
(8) Quantitation

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。

この設定は利用の際、事前に設定を準備する必要があります。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。

[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(計算モジュール搭載)が必要です。どちらの手法も、定量計算を実施するためのサンプル測定と、MASCOT での事前の設定項目の作成が必要になります。

詳細は「**10-2.Quantitation**」をご覧ください。



| | |
|------------------------|------------------------------------|
| Quantitation | None |
| Crosslinking | None |
| | iTRAQ 4plex |
| Fixed modifications | iTRAQ 4plex (protein) |
| | iTRAQ 8plex |
| | TMT 6plex |
| | TMT 2plex |
| | TMT 10plex |
| | TMTpro 16plex |
| | DilLeu 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] |
| Peptide tol. ± | ICPL quadruplex pre-digest [MD] |
| Peptide charge | 18O corrected [MD] |
| | 15N Metabolic [MD] |
| Data file | 15N + 13C Metabolic [MD] |
| | SILAC K+6 R+10 [MD] |

(9) Crosslinking

ペプチドがリンカーまたは共有結合で結合した測定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。

(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思えますが、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。修飾を指定する場合、右側にあるリストから該当項目を選び、真ん中にある < ボタンを押して Fixed または Variable modification 設定項目として指定します。

The screenshot shows the Mascot search interface for selecting modifications. On the left, there are two dropdown menus: 'Fixed modifications' and 'Variable modifications', both currently set to '--- none selected ---'. Below these is a checkbox labeled 'Display all modifications'. In the center, there are two sets of arrow buttons (> and <) for moving items between the dropdowns and the list. On the right, a scrollable list displays the following 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), and Carboxymethyl (C).

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、MASCOT に登録されているすべての修飾がリストに表示されます。

(12) 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。

(13) 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。

(14) #¹³C

質量分析装置での測定時、¹²C のみから構成されるペプチドではなく ¹³C 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。

(15) 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。

(16) 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。

(17) Data format

検索にかける query データのファイルフォーマットとして、「mascot generic (.mgf)」か、「mzML (.mzML)」かを選択します。

| | |
|-------------|------------------|
| Data format | Mascot generic ▾ |
| Instrument | Mascot generic |
| Target FDR | mzML (.mzML) |

(18) Instrument

フラグメントピークと理論値をマッチングする際、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。MS/MS において発生するフラグメントパターンの内容に応じて選択して下さい。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。なおカスタマイズにより、Machine learning に関する設定を組み込むことも可能です。

| | 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 | CID+ ETD | ETchD | EAD |
|--|---------|--------------|---------------|----------|----------|-----------|---------------|------------|----------|----------|----------------|---------------|-----------|----------|-------|-----|
| <u>1</u> ⁺ | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| <u>2</u> ⁺ (pre cursor >2 ⁺) | X | X | | X | X | X | | X | X | X | X | | | X | X | X |
| <u>2</u> ⁺ (pre cursor >3 ⁺) | | | | | | | | | | | | | | | | |
| imm. | | | X | | | | X | X | | | X | X | | | | |
| <u>a</u> | X | | X | | | | X | X | | | | X | X | | X | X |
| <u>a</u> * | X | | X | | | | X | | | | | X | | | X | |
| <u>a</u> 0 | | | X | | | | X | | | | | X | | | X | |
| <u>b</u> | X | X | X | X | X | X | X | X | | | X | X | | X | X | X |
| <u>b</u> * | X | X | X | X | X | X | X | X | | | X | X | | X | X | |
| <u>b</u> 0 | | X | X | X | X | X | X | X | | | X | X | | X | X | |
| <u>c</u> | | | | | | | | | X | X | | | X | X | X | X |
| <u>x</u> | | | | | | | | | | | | | | | | |
| <u>y</u> | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| <u>y</u> * | X | X | | X | X | X | X | | | | X | X | | X | X | |
| <u>y</u> 0 | | X | | X | X | X | X | | | | X | X | | X | X | |
| <u>z</u> | | | | | | | | X | | | | | | | | |
| <u>z</u> +1 | | | | | | | | | X | X | | | | X | X | X |
| <u>z</u> +2 | | | | | | | | | X | X | | | X | X | X | X |
| <u>y</u> <u>b</u> | | | | | | | X | X | | | X | X | | | | |
| <u>y</u> <u>a</u> | | | | | | | X | X | | | X | X | | | | |
| <u>y</u> must be sig. | | | | | | | | | | | | | | | | |
| <u>y</u> must be highest | | | | | | | | | | | | | | | | |
| <u>d</u> | | | | | | | X | | | | | | | | | |
| <u>v</u> | | | | | | | X | | | | | | | | | |
| <u>w</u> | | | | | | | X | | | | | | | X | X | X |

(19) Target FDR

同定基準となる FDR を設定します。通常は デフォルト値である 1%が採用されます。

(20) Machine learning

機械学習アルゴリズムによる結果の精査と最適化(refinement)を実施します。ショットガン解析を実施している場合、通常はチェックを入れて実施してください。

(21) DeepLC model

ペプチド配列から保持時間を予測するプログラム「DeepLC」を利用し、結果の精査と最適化(refinement)における要素の一つ(feature)として利用します。設定項目では計算に利用する数値が準備されたモデルを選択します。モデルはトレーニングデータセットの内容に基づいて名称がつけられています。

利用するためには、保持時間情報が各 query 内に書き込まれた状態で MASCOT に対して検索を行う必要があります(Mascot Distiller を利用する事で、その要件に対応させる事ができます)。

(22) MS2PIP model

ペプチド配列から MS2 スペクトルを予測するプログラム MS2PIP を利用し、結果の精査と最適化(refinement)における要素の一つ(feature)として利用します。設定項目では計算に利用する数値が準備されたモデルを選択します。モデルはトレーニングデータセットの内容に基づいて名称がつけられています。

(23) Start Search

検索を開始します。

(24) 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事ができる項目があります。

Databases

Taxonomy

Enzyme

Quantitation

Crosslinking

Modification

Instrument

Refinement 実施の各種条件など

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。

Instrument については、refinement の内容も併せて定義をすることができます。質量分析装置メーカーに付属する解析ソフトウェアから MASCOT を利用する際、instrument 項目を選ぶだけで、refinement の実施の有無や DeepLC/MS2PIP のモデルを指定するように設定することが可能です。

また上記の内容とは別に、refinement に利用するプログラムのモデルについてもカスタマイズが可能です。

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ことができます。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

MASCOT ではデータベースにないエントリーを同定することはできませんが、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ります。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、2025年3月時点で約**57万**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、自動かつ精査無しのデータベース **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2025年3月版で**252,633,200**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

■ NCBIprot (旧名称 NCBI nr)

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。

■ 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場合、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

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ー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、設定した内容がデフォルトの選択設定として選ばれた状態で画面が開きます。

The screenshot displays the Matrix Science website interface. The main content area is titled "Set Mascot search form defaults" and contains several configuration options:

- Database:** A dropdown menu with "SwissProt" selected.
- 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 modifications including Carbamidomethyl (C), Carbamidomethyl (N-term), Carbamyl (K), Carbamyl (N-term), and Carboxymethyl (C).
- Variable modifications:** A list of modifications including mTRAQ:13C(6)15N(2) (N-term), mTRAQ:13C(6)15N(2) (Y), NIPCAM (C), Oxidation (HW), and Oxidation (M).

In the left sidebar, under the "Search form defaults" section, the link "preferred default settings" is highlighted with a red box.

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。

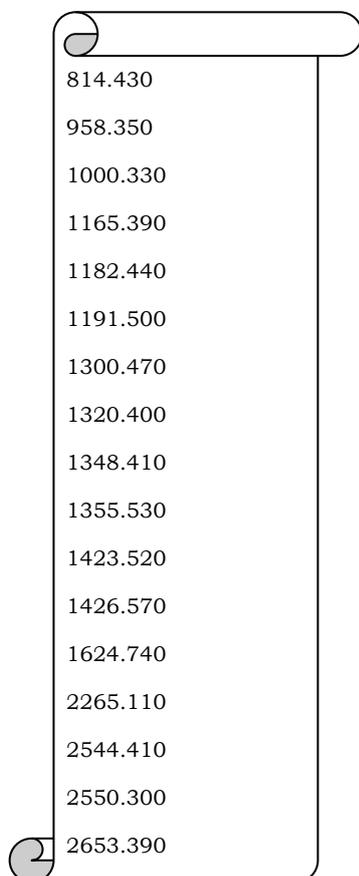
6-1. 表示例で使用している検索について

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

[ローカルサーバー] 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 |
|----------------------|-----------------|
| Database | SwissProt |
| Enzyme | Trypsin/P |
| Allow up to | 2 |
| Taxonomy | all |
| Peptide tol,± | 0.2 Da |
| Mass Values | MH ⁺ |

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。

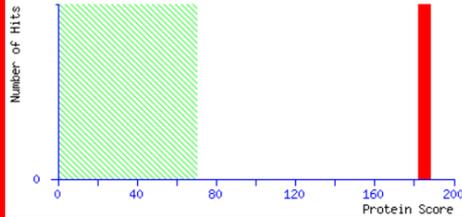
6-2-2 ヘッダー部分

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=eiF5A 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_GEOSM](#) 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/>

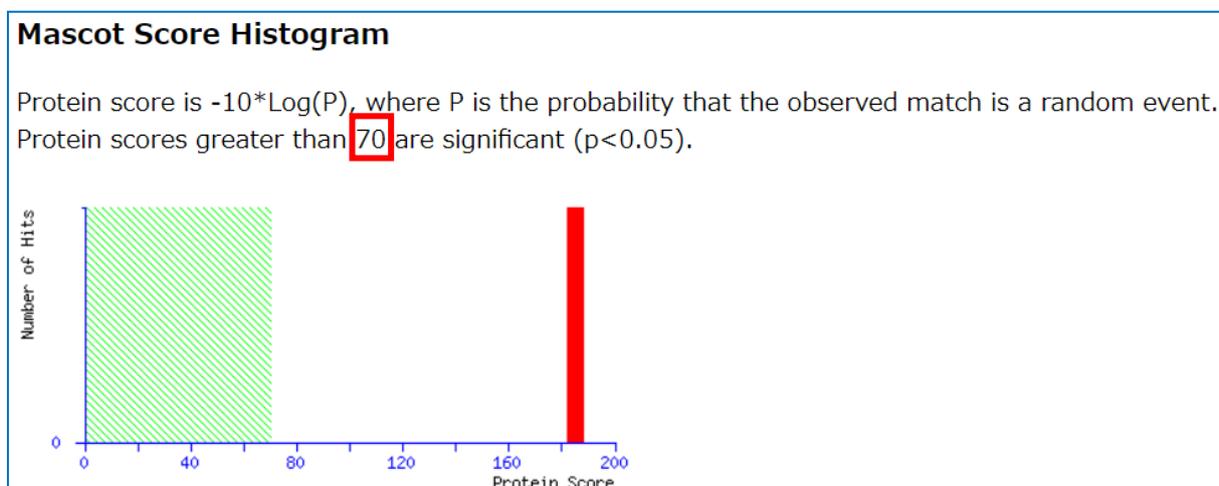
6-2-2. ヘッダー部分

```
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

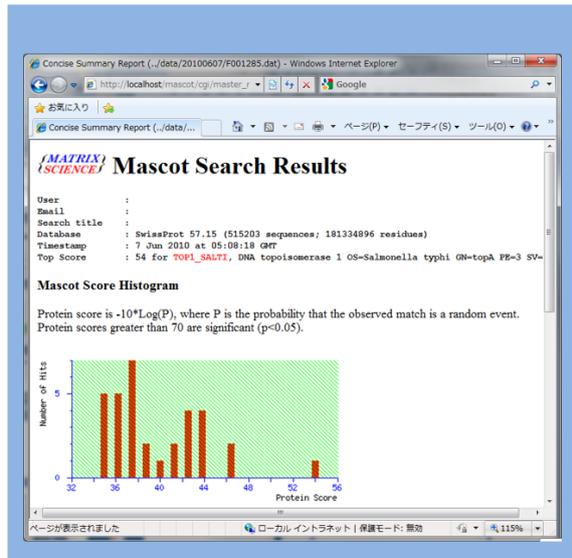


横軸が MASCOT Score,縦軸はそのスコアを持つエントリーの数です。

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。**

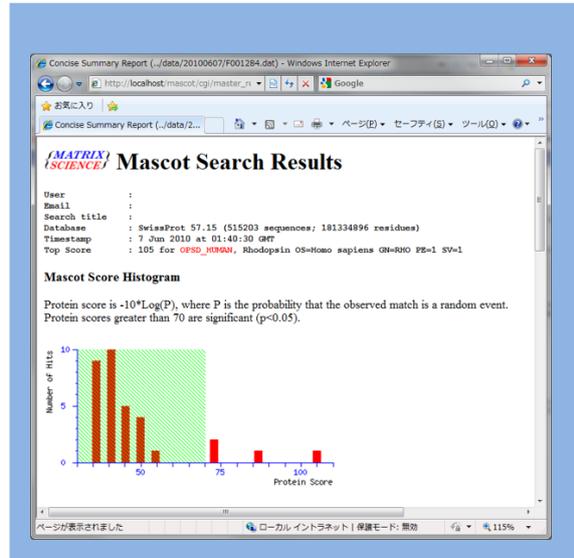
同定基準値は、グラフの緑の網掛けゾーン一番右側に対応します。**グラフの緑の網掛けより右側にタンパク質が存在するとき同定タンパク質が見つかったことを意味し、逆に最もスコアが高いタンパク質も緑の網掛けの領域を脱する事が出来なかった場合は同定タンパク質が見つからなかったことを意味します**(次頁図)。同定タンパク質が見つかった場合は検索成功で、同定内容について検証する作業に入ります。一方同定タンパク質が見つからなかった時は入力データや指定パラメーターを見直して同定できなかった理由を検証するか、新たな測定を実施して再度同定を試みる事をお勧めします。

• 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 ▼ | Help |
| | Significance threshold $p <$ 0.05 | Max. number of hits AUTO |
| | Preferred taxonomy 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

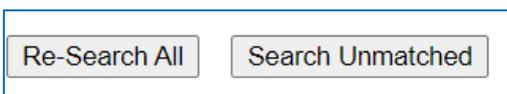
Significance threshold : 同定基準の **p value**(このケースでは **Expect** と同じとお考え下さい)を指定します。小さいほど同定基準が厳しい事になります。

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ータベース側で定められたデフォルトの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 の設定を行うと、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。

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. | 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= | | | | |
| | NADD_CHLL2 | Mass: 22438 | Score: 44 | Expect: 20 | Matches: 4 |
| | Probable nicotinate-nucleotide adenyllyltransferase OS=Chlorobium limicola (strain DSM 245 / NBRC 10380 | | | | |
| | 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- | | | | |
| | TRGV8_HUMAN | 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 $p < \dots$ 」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 | : ± 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 (M₀): 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**.

```

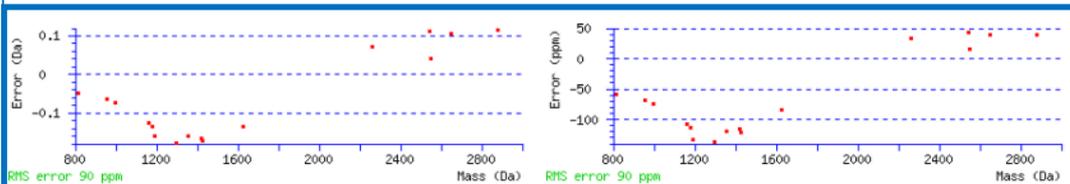
1 MEPAFARSPR PQDDPARPQE PTMPPPETPS EGRQPSPPS PTERAPASEE
51 EPQFLRCQQC QAEAKCPKLL PCLHTLCSGC LEASGMQCPPI CQAWPLGAD
101 TPALDNVFFE SLQRLSVYR QIVDAQVCT RCKESADFWC FECEQLLCAK
151 CFEAHQWFLK HEARPLAELR NQSVREFLDG TRKTNNIFCS NPNHRTPTLT
201 SIYCRGCSKP LCCSCALLDS SHSELKCDIS AEIQQRQEEEL DAMTQALQEQ
251 DSAFGAVHQA MHAAVQQLGR ARAETEELIR ERVQVVAHV RAQERELLEA
301 VDARYQRDYE EMASRLGRLD AVLQRIRTGS ALVQRMKCYA SDQEVLDMHG
351 FLRQALCRLR QEEPSLQAA 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 ADPQAE DRPL VFFDLKIDNE TQKISQLAAV NRESKFRVVI QPEAFFSIYS
651 KAVSLEVGLQ HFLSFLSSMR RPILACYKLW GPGLPNFFRA LEDINRLWEF
701 QEAIISGFLAA LPLIRERVFG ASSFKLKNLA QTVLARNMSE RSAMAAVLAM
751 RDLCRLLVVS 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.SPRPQDDPARPQEPPTMPPPETPSEGR.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.APASEEFPQFLR.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.DYEEMASR.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.LRQEEPSLQAQAVR.T |
| 361 - 372 | 1355.5300 | 1354.5227 | 1354.6841 | -0.1613 0 | R.QEEPSLQAQAVR.T |
| 361 - 382 | 2550.3000 | 2549.2927 | 2549.2510 | 0.0417 2 | R.QEEPSLQAQAVRTDGFDEFKVR.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.SSPQPRPSTSK.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.SFVIGSEVFLPNSNHVASGAGEAER.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.VLDENLADPQAE DRPLVFFDLK.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 Q96S41; 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 AVLQIRITGS 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 HNVSEMFELLS AHRDRQGGK 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

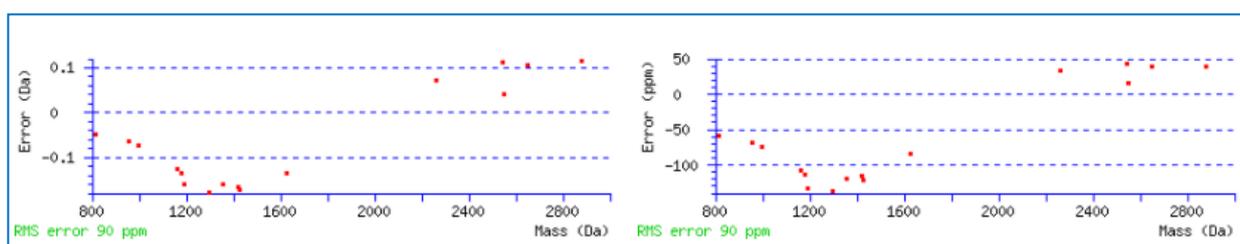
| Start - End | Observed | Mr (expt) | Mr (calc) | Delta M | Peptide |
|-------------|-----------|-----------|-----------|-----------|---------------------------------|
| 8 - 33 | 2882.5000 | 2881.4927 | 2881.3777 | 0.1150 2 | R.SPRPQQDPPARPOEPTMPPPETPSEGR.Q |
| 34 - 44 | 1182.4400 | 1181.4327 | 1181.5677 | -0.1349 0 | R.QPSPSPSPTER.A |
| 45 - 56 | 1423.5200 | 1422.5127 | 1422.6779 | -0.1652 0 | R.APASEEEEFQFLR.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。

```

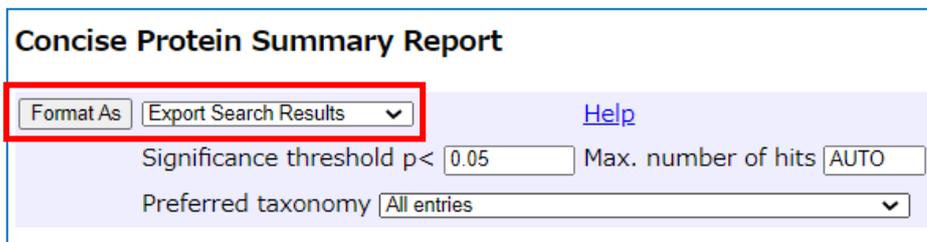
ID PML HUMAN Reviewed; 882 AA.
AC P29590; E9PBR7; P29591; P29592; P29593; Q00755; Q15959; Q59FP9; Q8WUA0;
AC Q96S41; Q9BPW2; Q9BWP7; Q9BZX6; Q9BZX7; Q9BZX8; Q9BZX9; Q9BZY0; Q9BZY2;
AC Q9BZY3;
DT 01-APR-1993, integrated into UniProtKB/Swiss-Prot.
DT 25-NOV-2008, sequence version 3.
DT 02-JUN-2021, entry 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. 結果のファイル出力

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。



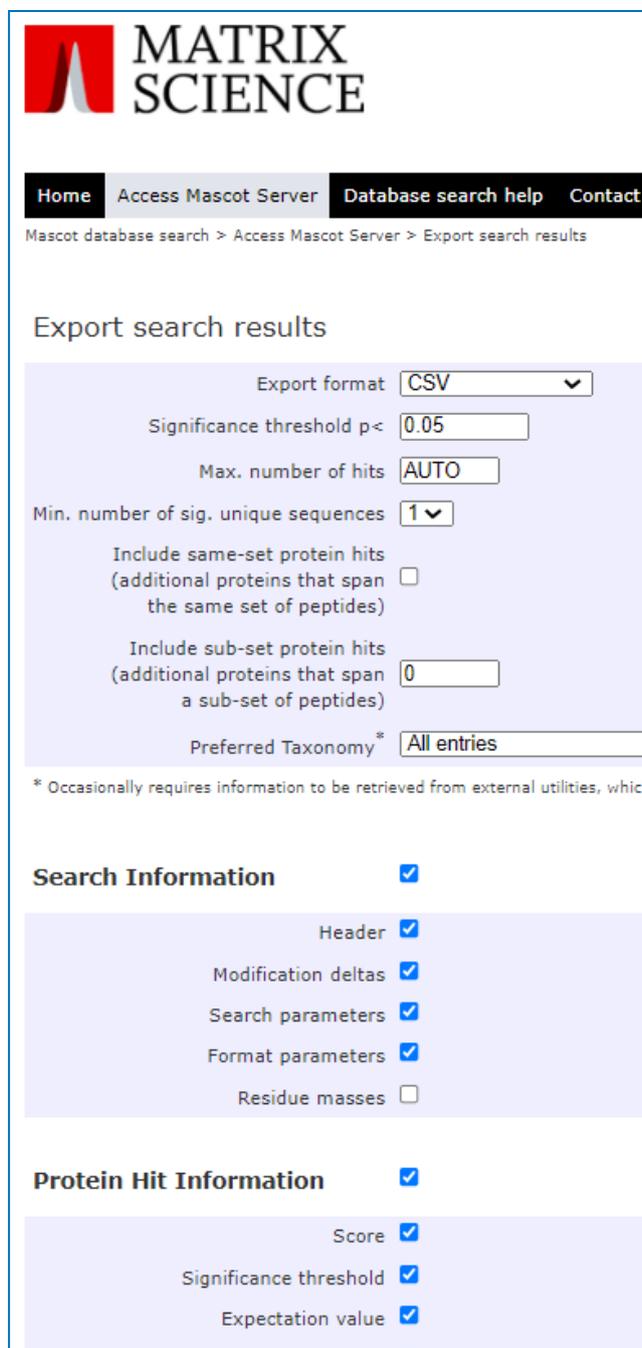
Concise Protein Summary Report

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

Significance threshold p< 0.05 Max. number of hits AUTO

Preferred taxonomy All entries

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 CSV

Significance threshold p< 0.05

Max. number of hits AUTO

Min. number of sig. unique sequences 1

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) 0

Preferred Taxonomy* All entries

* 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/F981142.msr

[ローカルサーバー] http://localhost/mascot/master_results_2.pl?file=../data/F981142.msr

| 設定項目 | 設定内容 |
|---------------------------|--------------------------|
| Enzyme | Trypsin/P |
| Fixed Modification | Carbamidomethyl (C) |
| Variable Modification | Oxidation (M) |
| Peptide tol.± | 20 ppm |
| Fragment mass tolerance ± | 0.6 Da |
| Missed cleavage | 2 |
| Instrument | Default |
| Refinement | 実施 |
| DeepLC | Full_hc_unmod_fixed_mods |
| MS ² PIP | CID |

query 数 : 8,675

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 : 3,437

同定タンパク質数 : 863 (820 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。)

MATRIX SCIENCE MASCOT Search Results

User : matrix
E-mail : support@matrixscience.com
Search title : Yeast example (CPTAC study 6)
MS data file : klc_031308p_cptac_study6_6B011.mgf
Databases : 1: contaminants 20160129 (247 sequences; 128,130 residues)
2: Sigma_UPS 20240812 (50 sequences; 11,863 residues)
3: UP2311_S_cerevisiae 20240811 (6,091 sequences; 2,950,884 residues)
Timestamp : 12 Aug 2024 at 11:09:31 GMT

7-2-2 ヘッダー部分

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

▼ Search parameters
Type of search : MS/MS Ion Search
Target FDR : 1%
Enzyme : Trypsin/P
Fixed modifications : [Carbamidomethyl \(C\)](#)
Variable modifications : [Oxidation \(M\)](#)
Mass values : Monoisotopic
Protein mass : Unrestricted
Peptide mass tolerance : ± 20 ppm
Fragment mass tolerance : ± 0.6 Da
Max missed cleavages : 2
Instrument type : Default
Number of queries : 8,675

7-2-3 再検索と結果ファイル出力

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 (results refined with [7-2-8 表示内容の切り替え\[スコア足切りなど\]](#))

Significance threshold p< 0.01247 Max. number of families AUTO [\[help\]](#)
Target FDR (overrides sig. threshold) 1% FDR type Sequence
Display non-sig. matches Min. number of sig. unique sequences 1
Dendrograms cut at 0
Preferred taxonomy All entries
Refine results using machine learning (Percolator)
- Use features calculated by Mascot [\[help\]](#)
- DeepLC model for retention times [\[help\]](#)
- MS2PIP model for spectral similarity [\[help\]](#)
full_hc_unmod_fixed_mods
CID

Apply

▼ Sensitivity and FDR (reversed protein sequences)

| | Target | Decoy | FDR |
|--------------------------|--------|-------|-------|
| Protein family members | 863 | 29 | 3.36% |
| Sequences above homology | 3437 | 34 | 0.99% |

7-2-9 Sensitivity and FDR

Details about ML model performance are available in [the machine learning quality report](#).

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

7-2-10 machine learning quality report

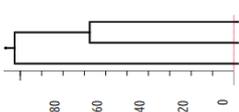
Proteins (863) [Report Builder](#) [Unassigned \(4848\)](#) [\[permalink\]](#)

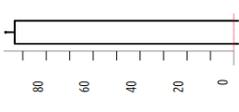
Protein families 1-10 (out of 820)

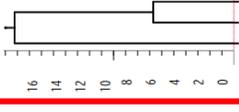
10 per page 1 2 3 4 5 6 ... 82 Next Expand all Collapse all

Accession contains Find Clear

▶ 1 3::P00549 652 Pyruvate kinase 1 OS=Saccharomyces cerevisiae (strain ATCC...

▶ 2  1 3::P10592 557 Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strai...
2 3::P10591 554 Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strai...
3 3::P16474 147 Endoplasmic reticulum chaperone BIP OS=Saccharomyces cer...

▶ 3  1 3::P07259 556 Multifunctional protein URA2 OS=Saccharomyces cerevisiae (s...
2 3::P03965 135 Carbamoyl phosphate synthase arginine-specific large chain O...

▶ 4  1 3::P00359 rogenase 3 OS=Saccharo...
2 3::P00358 453 Glyceraldehyde-3-phosphate dehydrogenase 2 OS=Saccharo...
3 3::P00360 161 Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Saccharo...

7-2-11 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)

7-2-2. ヘッダー部分

```
User : matrix
E-mail : support@matrixscience.com
Search title : Yeast example (CPTAC study 6)
MS data file : klc_031308p_cptac_study6_6B011.mgf
Databases : 1: contaminants 20160129 (247 sequences; 128,130 residues)
           2: Sigma_UPS 20240812 (50 sequences; 11,863 residues)
           3: UP2311_S_cerevisiae 20240811 (6,091 sequences; 2,950,884 residues)
Timestamp : 12 Aug 2024 at 11:09:31 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」をご覧ください。

7-2-4. Search parameters

```
▼Search parameters
Type of search : MS/MS Ion Search
Target FDR : 1%
Enzyme : Trypsin/P
Fixed modifications : ☑Carbamidomethyl (C)
Variable modifications : ☑Oxidation (M)
Mass values : Monoisotopic
Protein mass : Unrestricted
Peptide mass tolerance : ± 20 ppm
Fragment mass tolerance : ± 0.6 Da
Max missed cleavages : 2
Instrument type : Default
Number of queries : 8,675
```

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)。
- #¹³C** : ペプチドに含まれる ¹³C を考慮した回数(例図には表示なし)。
- 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。

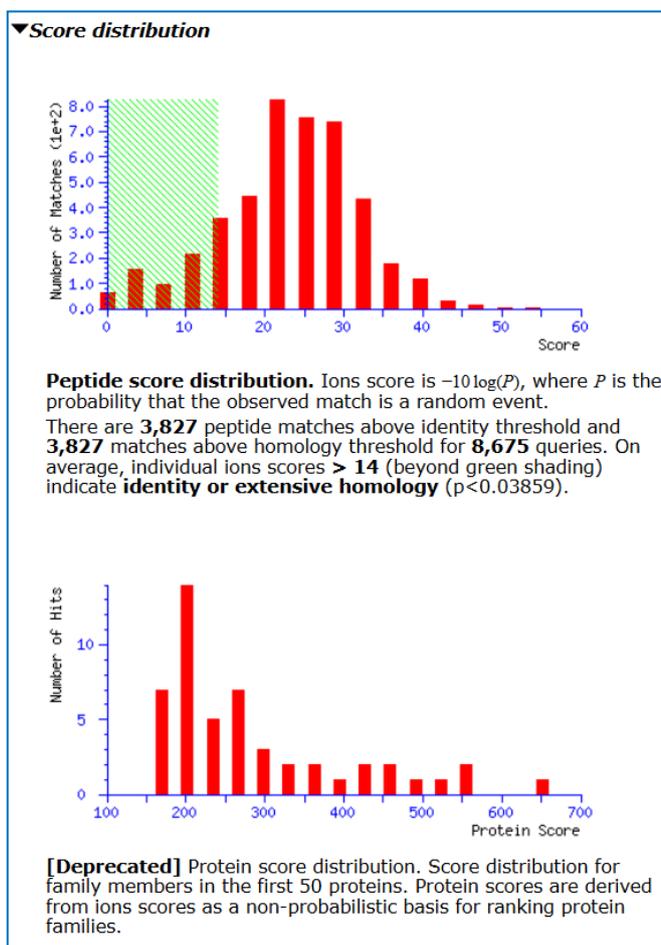
7-2-5. Score Distribution

▶ 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。ペプチドの同定基準値に該当するスコア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 statistics for all protein families

| Modification | Delta | Type | Site | Total matches |
|-----------------|-----------|----------|------|---------------|
| Carbamidomethyl | 57.021464 | fixed | C | 327 |
| Oxidation | 15.994915 | variable | M | 58 |

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 | | | GAYSLSLR | significant |
| | | 9 | ▶ | 1 | | | GFFLFVEGGR | 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 | | | VCNYSVSWIK | 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 | 1.129 |
| FHR2_HUMAN | 1330 | | 0.859 | 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, refinement の再実行など]

Protein Family Summary (results refined with machine learning)

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

Preferred taxonomy Dendrograms cut at

Refine results using machine learning (Percolator) [\[help\]](#)

- Use features calculated by Mascot [\[help\]](#)

- DeepLC model for retention times [\[help\]](#)

- MS2PIP model for spectral similarity [\[help\]](#)

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 の設定です。現バージョンでは「ペプチドの期待値で見た同定基準値」が示されます。**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。

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。設定値を 2 以上にすると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ります。

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。

Refine results using machine learning (Percolator) :

Refinement 実施。実施する事で同定ペプチド数が増えます。

Use features calculated by Mascot :

Percolator の core features を計算・利用するかどうか。現バージョンでは refinement を実施した場合必ずこの項目にチェックが入ります。

DeepLC model for retention times:

ペプチド配列データから保持時間予測の計算の実施と、その結果を Percolator で利用するかどうか。選択肢は計算時に使うパラメーターセットで、どのようなトレーニングデータセットを使ったかに基づいた名称がつけられています。

model に関する説明は、以下の help ページをご覧ください。

https://www.matrixscience.com/help/ms2rescore_help.html#DEEPLCMODELS

MS2PIP model for spectral similarity :

ペプチド配列データから MS2 スペクトル予測の計算の実施と、その結果を Percolator に引き渡すかどうか。選択肢は計算時に使うパラメーターセットで、どのようなトレーニングデータセットを使ったかに基づいた名称がつけられています。

model に関する説明は、以下の help ページをご覧ください。

https://www.matrixscience.com/help/ms2rescore_help.html#MS2PIPMODELS

7-2-9. Sensitivity and FDR

protein FDR の値と peptide(MASCOT では Sequences と表記)または PSM の FDR の値が表示されます(下図)。

| ▼Sensitivity and FDR (reversed protein sequences) | | | |
|--|--------|-------|-------|
| | Target | Decoy | FDR |
| Protein family members | 863 | 29 | 3.36% |
| <input type="text" value="Sequences"/> above <input type="text" value="homology"/> | 3437 | 34 | 0.99% |

7-2-10. machine learning quality report

機械学習アルゴリズムによる同定結果の refinement による確からしさや効果、再スコアリングに影響を及ぼした要素について確認する事ができます。関連したグラフを多数表示させる事ができます。

Details about ML model performance are available in [the machine learning quality report](#).

ハイパーリンクをクリックするとレポートが表示されます。レポートは「**Overview**」、「**Target-Decoy evaluation**」、「**Rescoring features**」の3つのタブで構成されています。

「**Overview**」タブでは、refinement 適用前後で結果がどのように変わったかに関する情報がまとめられています。

General statistics では、refinement 適用前後での同定スペクトル数、または Peptide 数の変化がまとめられています。

Score comparison では、各 query に関して、refinement 適用前後のスコアに基づいたプロットです。target(青)と decoy (赤)両方の結果が同時に含まれています。

False discovery rate comparison では、FDR の区切りを変化させた場合、同定スペクトル数がどのように推移するかまとめられています。

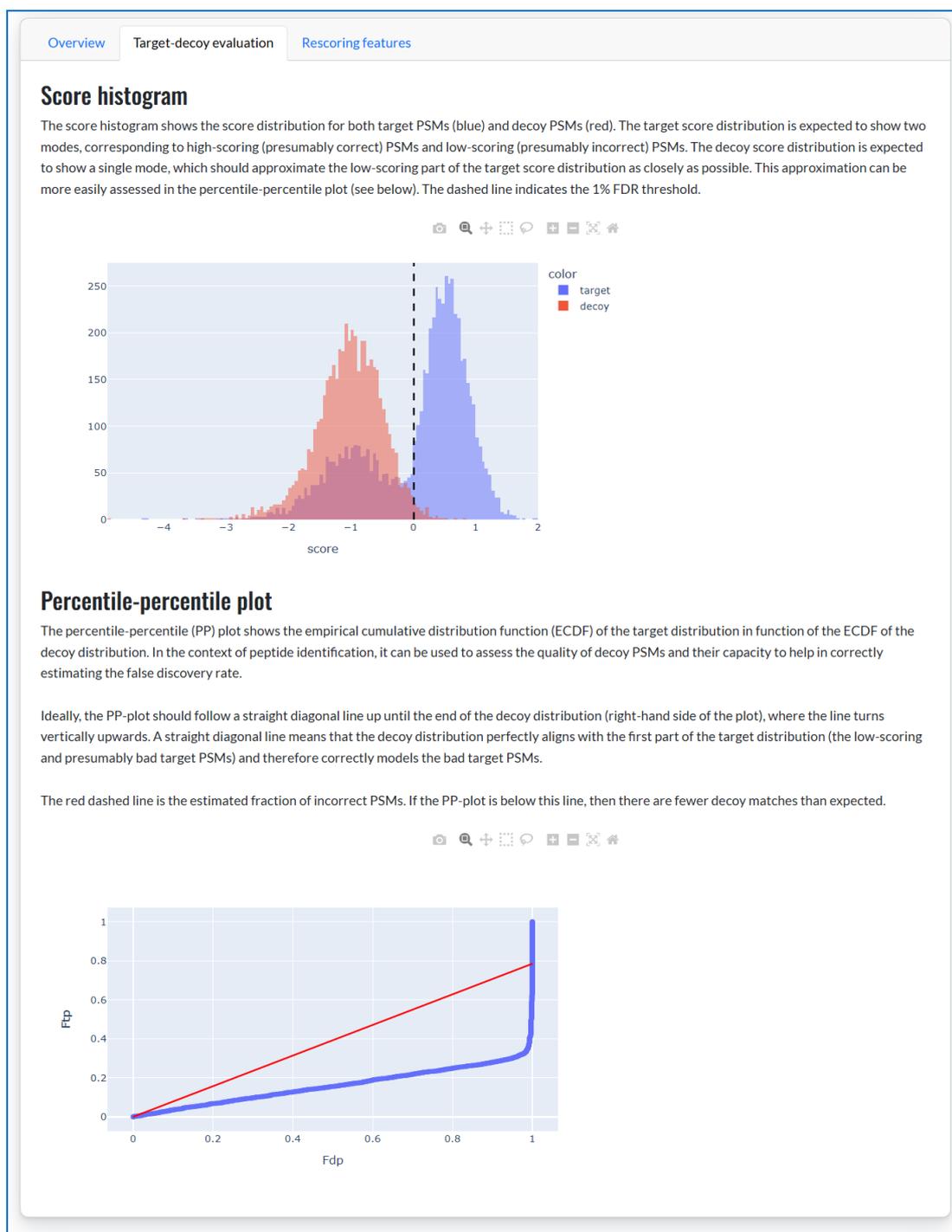
Identification overlap では、refinement 適用前後により query の結果が変わったもの、変わらないものがそれぞれどれくらいあるかを示します。



「**Target-decoy evaluation**」タブでは、target データベースと decoy データベースにおける結果の違いについて検証するためのグラフが提供されます。

Score histogram では target と decoy のスコア分布、特に両者の分布がどれくらい分離されているかを確認する事ができます。同定基準値が点線で示されています。

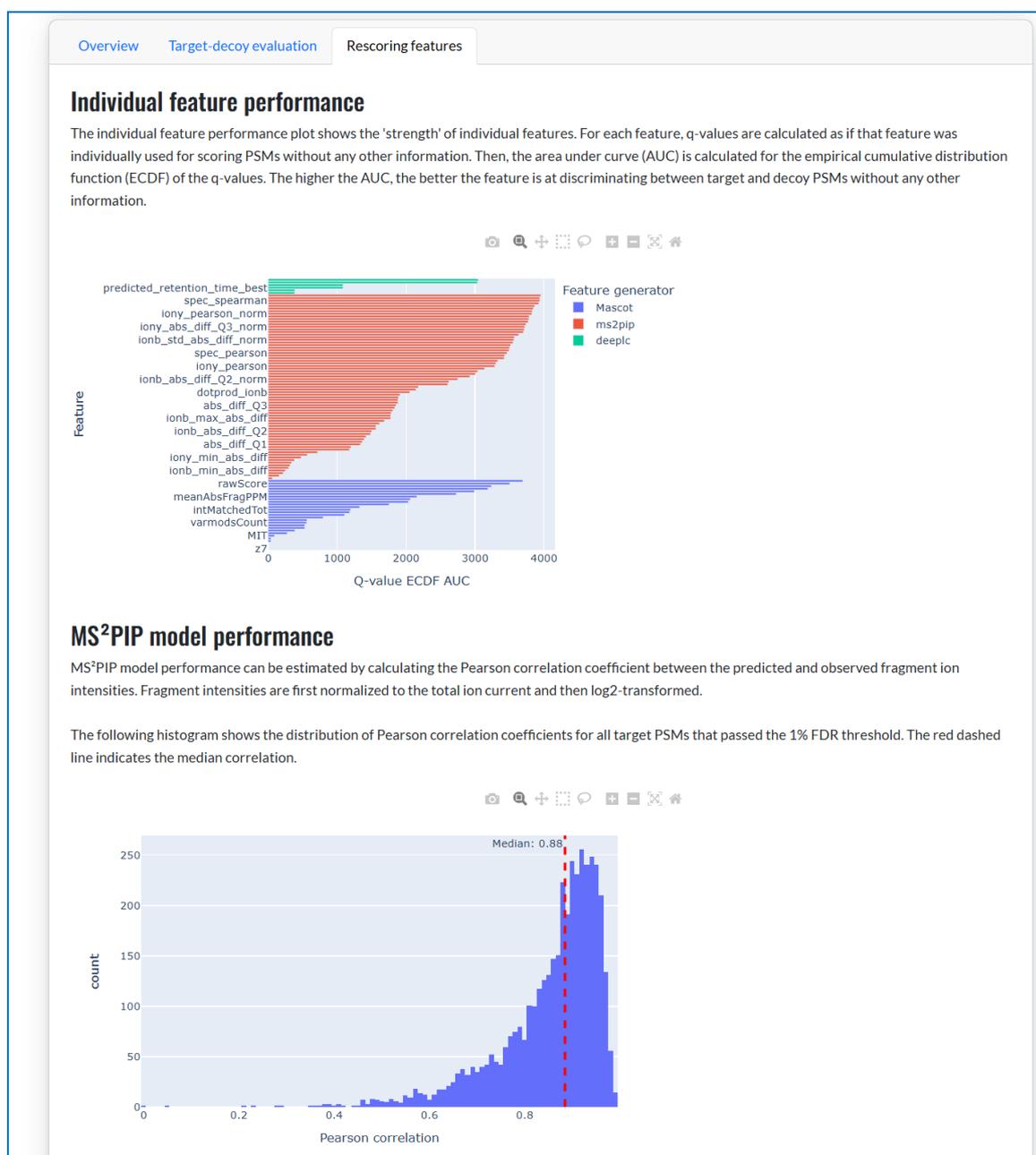
Percentile-percentile plot (PP plot)では Decoy データベースと Target データベースそれぞれに対して行った検索のスコア分布について累積度を比較しています。



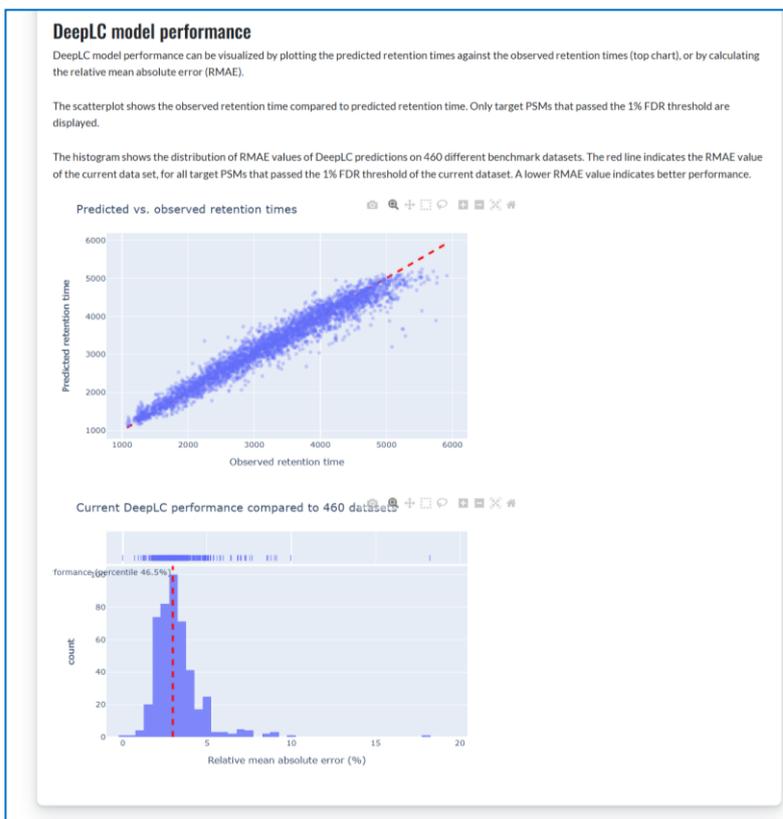
「**Rescoring features**」タブでは、Percolator が再評価の際に使用した features に関するグラフが表示されます。

Individual feature performance では、再評価の際 Percolator の判断において features がどれくらい影響を及ぼしたのかを示しています。

MS²PIP model performance では、実測値と予測値ピアソン相関のヒストグラムが表示されます。配列から予測されたスペクトルと実測データとの一致度は、feature の1つとして refinement 計算に利用されます。



DeepLC model では、保持時間の
 実測値と予測値について、散布図
 並びにピアソン相関のヒストグラムが
 表示されます。予測+キャリブレーション
 された保持時間と実測データとの
 一致度は、feature の1つとして
 refinement 計算に利用されます。



7-2-11. 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。

Proteins (863) Report Builder Unassigned (4848) [\\$. permalink](#)

Protein families 1-10 (out of 820)

10 per page 1 2 3 4 5 6 ... 82 Next Expand all Collapse all

Accession contains Find Clear

| | | | |
|----|-------------|-----|---|
| ▶1 | 3::P00549 | 652 | Pyruvate kinase 1 OS=Saccharomyces cerevisiae (strain ATCC... |
| ▶2 | 1 3::P10592 | 557 | Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strai... |
| | 2 3::P10591 | 554 | Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strai... |
| | 3 3::P16474 | 147 | Endoplasmic reticulum chaperone BIP OS=Saccharomyces cer... |

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。

Proteins (863) Report Builder Unassigned (4848)

Protein families 1-10 (out of 820)

赤で囲われた部分のうち、数字 **863** 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 **820** 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、同定タンパク質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 shows a web interface for protein analysis. At the top, there are buttons for 'Proteins (863)', 'Report Builder', and 'Unassigned (4848)'. Below this, the section 'Protein families 1-10 (out of 820)' is displayed. A dropdown menu shows '10' per page, and there are navigation buttons for '1', '2', '3', '4', '5', '6', '82', and 'Next'. There are also 'Expand all' and 'Collapse all' buttons. A search bar with 'Accession' and 'contains' dropdowns is present, along with 'Find' and 'Clear' buttons.

A dendrogram is shown, with a red box highlighting the first two clusters. Cluster 1 is labeled '3::P00549' and cluster 2 is labeled '1 3::P10592', '2 3::P10591', and '3 3::P16474'. To the right of the dendrogram, a list of protein entries is shown:

- 652 Pyruvate kinase 1 OS=Saccharomyces cerevisiae (strain ATCC 2045...
- 557 Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ATC...
- 554 Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strain ATC...
- 147 Endoplasmic reticulum chaperone BiP OS=Saccharomyces cerevisiae

Below the dendrogram, there is a 'Threshold (0): 0.0' field with a 'Cut' button. A table of peptide matches is shown below:

| | Score | Mass | Matches | Sequences | emPAI | |
|-----|-------|-------|---------|-----------|-------|---|
| 2.1 | 557 | 69599 | 41 (41) | 26 (26) | 4.15 | Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ATCC 20450... |
| 2.2 | 554 | 69786 | 41 (41) | 28 (28) | 4.44 | Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strain ATCC 20450... |
| 2.3 | 147 | 74479 | 10 (10) | 9 (9) | 0.67 | Endoplasmic reticulum chaperone BiP OS=Saccharomyces cerevisiae (strain ATCC 20450... |

Below the table, there are 'Redisplay', 'All', and 'None' buttons. A section titled '59 peptide matches (49 non-duplicate, 10 duplicate)' is shown, with a checked 'Auto-fit to window' option. A table of peptide matches is displayed:

| Query Dupes | Observed | Mx (expt) | Mx (calc) | ppm | M | Score | Expect | Rank | U | 1 | 2 | 3 | Peptide |
|-------------|----------|-----------|-----------|-------|---|-------|--------|------|---|---|---|---|-----------------------------|
| 364 | 380.7134 | 759.4122 | 759.4127 | -0.54 | 0 | 23 | 0.0047 | ▶1 | | ■ | ■ | ■ | R.NSTIPTK.K |
| 472 | 389.1940 | 776.3735 | 776.3738 | -0.44 | 0 | 23 | 0.0046 | ▶1 | | ■ | ■ | ■ | K.MVAEAEK.F |
| 890 | 417.2091 | 832.4037 | 832.4039 | -0.19 | 0 | 15 | 0.033 | ▶1 | | ■ | ■ | ■ | K.DLSTNQR.A |
| 1012 | 426.2002 | 850.3857 | 850.3742 | 13.6 | 0 | 15 | 0.034 | ▶3 | U | ■ | ■ | ■ | R.MVEEAEK.F + Oxidation (M) |
| 1238 | 439.7519 | 877.4892 | 877.4909 | -1.94 | 0 | 23 | 0.0055 | ▶1 | U | ■ | ■ | ■ | K.VAYPITSK.L |
| 1300 | 444.7366 | 887.4585 | 887.4600 | -1.64 | 0 | 21 | 0.0083 | ▶1 | | ■ | ■ | ■ | R.STLDPVEK.V |
| 2222 | 490.7442 | 979.4739 | 979.4757 | -1.79 | 0 | 25 | 0.0034 | ▶1 | U | ■ | ■ | ■ | R.ALSSQMSTR.I |
| 2422 | 334.8570 | 1001.5490 | 1001.5505 | -1.51 | 1 | 19 | 0.012 | ▶1 | | ■ | ■ | ■ | K.TKDNLLCK.F |

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。

This screenshot shows a detailed view of the protein analysis interface. It includes the same top navigation and search elements as the previous screenshot. The 'Protein families 1-10 (out of 820)' section is expanded to show a dendrogram and a list of protein entries. A red box highlights the first two clusters, similar to the previous screenshot. The list of protein entries is the same as in the previous screenshot. A 'permalink' button is visible in the top right corner.

デンドログラムが表示されている箇所では、同定タンパク質のうち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 | 3::P10592 | 557 | 69599 | 41 (41) | 26 (26) | 4.15 Heat shock protein SSA2 OS=Saccharo |
| <input checked="" type="checkbox"/> | 2.2 | 3::P10591 | 554 | 69786 | 41 (41) | 28 (28) | 4.44 Heat shock protein SSA1 OS=Saccharo |
| <input checked="" type="checkbox"/> | 2.3 | 3::P16474 | 147 | 74479 | 10 (10) | 9 (9) | 0.67 Endoplasmic reticulum chaperone BiP O |

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。

Mass :

タンパク質の質量。データベースに登録されている配列情報から計算。

Matches :

タンパク質にアサインされた query 数。()内の数字はその中で同定基準を超えているもの。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

続いて、アサインペプチドに関して表示しています。

▼59 peptide matches (49 non-duplicate, 10 duplicate)

Auto-fit to window

| Query Dupes | Observed | Mr (expt) | Mr (calc) | ppm | M | Score | Expect | Rank | U | 1 | 2 | 3 | Peptide |
|----------------------|----------|-----------|-----------|-------|---|-------|--------|------|---|---|---|---|-----------------------------|
| 364 | 380.7134 | 759.4122 | 759.4127 | -0.54 | 0 | 23 | 0.0047 | ▶1 | ■ | ■ | ■ | | R.NSTIPTK.K |
| 472 | 389.1940 | 776.3735 | 776.3738 | -0.44 | 0 | 23 | 0.0046 | ▶1 | ■ | ■ | ■ | | K.MVAEAEK.F |
| 890 | 417.2091 | 832.4037 | 832.4039 | -0.19 | 0 | 15 | 0.033 | ▶1 | ■ | ■ | ■ | | K.DLSTNQR.A |
| 1012 | 426.2002 | 850.3857 | 850.3742 | 13.6 | 0 | 15 | 0.034 | ▶3 | U | | | | R.MVEEAEK.F + Oxidation (M) |
| 1238 | 439.7519 | 877.4892 | 877.4909 | -1.94 | 0 | 23 | 0.0055 | ▶1 | U | | | | K.VAYPIISK.L |
| 1300 | 444.7366 | 887.4585 | 887.4600 | -1.64 | 0 | 21 | 0.0083 | ▶1 | ■ | ■ | ■ | | R.STLDPVEK.V |
| 2222 | 490.7442 | 979.4739 | 979.4757 | -1.79 | 0 | 25 | 0.0034 | ▶1 | U | | | | R.ALSSQMSTR.I |
| 2422 | 334.8570 | 1001.5490 | 1001.5505 | -1.51 | 1 | 19 | 0.012 | ▶1 | ■ | ■ | ■ | | K.TKDNLLGK.F |

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** : 実験データ側のペプチドの m/z
- Mr(expt)** : m/z と電荷から計算された、実験データ側のペプチドの質量
- Mr(calc)** : データベースの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** : 理論ペプチド作成時に Missed cleavage 設定が実際に適用された回数
- 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 ▼ ₁ ■ ■ R.DFIDYYLIK.Q |
| | -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 MFGCDVGSQWR + 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 ペプチドであることを意味します。
- 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。

7-2-12. 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 (863) **Report Builder** Unassigned (4848)

Protein families 1-10 (out of 820)

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 (863) Report Builder Unassigned (4848)

Protein family members (863 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 | UP2311_S_cerevisiae | ☞3::P00549 | 652 | 54909 | 54 | 54 | 34 | 34 | 13.67 | Pyruvate kinase 1 OS=Saccharomyces cerevisiae |
| 2 | 1 | UP2311_S_cerevisiae | ☞3::P10592 | 557 | 69599 | 41 | 41 | 26 | 26 | 4.15 | Heat shock protein SSA2 OS=Saccharomyces ce |
| 2 | 2 | UP2311_S_cerevisiae | ☞3::P10591 | 554 | 69786 | 41 | 41 | 28 | 28 | 4.44 | Heat shock protein SSA1 OS=Saccharomyces ce |
| 2 | 3 | UP2311_S_cerevisiae | ☞3::P16474 | 147 | 74479 | 10 | 10 | 9 | 9 | 0.67 | Endoplasmic reticulum chaperone BiP OS=Sacch |
| 3 | 1 | UP2311_S_cerevisiae | ☞3::P07259 | 556 | 246198 | 40 | 40 | 39 | 39 | 0.96 | Multifunctional protein URA2 OS=Saccharomyces |
| 3 | 2 | UP2311_S_cerevisiae | ☞3::P03965 | 135 | 124465 | 11 | 11 | 11 | 11 | 0.45 | Carbamoyl phosphate synthase arginine-specific |
| 4 | 1 | UP2311_S_cerevisiae | ☞3::P00359 | 508 | 35838 | 35 | 35 | 20 | 20 | 12.15 | Glyceraldehyde-3-phosphate dehydrogenase 3 O |
| 4 | 2 | UP2311_S_cerevisiae | ☞3::P00358 | 453 | 35938 | 33 | 33 | 18 | 18 | 9.33 | Glyceraldehyde-3-phosphate dehydrogenase 2 O |
| 4 | 3 | UP2311_S_cerevisiae | ☞3::P00360 | 161 | 35842 | 12 | 12 | 10 | 10 | 2.23 | Glyceraldehyde-3-phosphate dehydrogenase 1 O |
| 5 | 1 | UP2311_S_cerevisiae | ☞3::P16521 | 477 | 116727 | 35 | 35 | 24 | 24 | 1.66 | Elongation factor 3A OS=Saccharomyces cerevis |
| 6 | 1 | UP2311_S_cerevisiae | ☞3::P19097 | 463 | 207388 | 37 | 37 | 34 | 34 | 1.00 | Fatty acid synthase subunit alpha OS=Saccharo |

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

↑ ↓ Apply

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

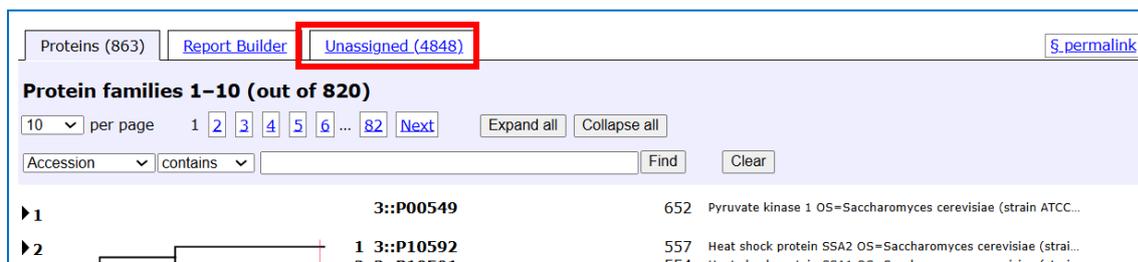
Num. of significant matches ≥ 2 Remove

AND Sequence coverage < >

Update

7-2-13. 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、項目の内容についてはそちらをご参照ください。



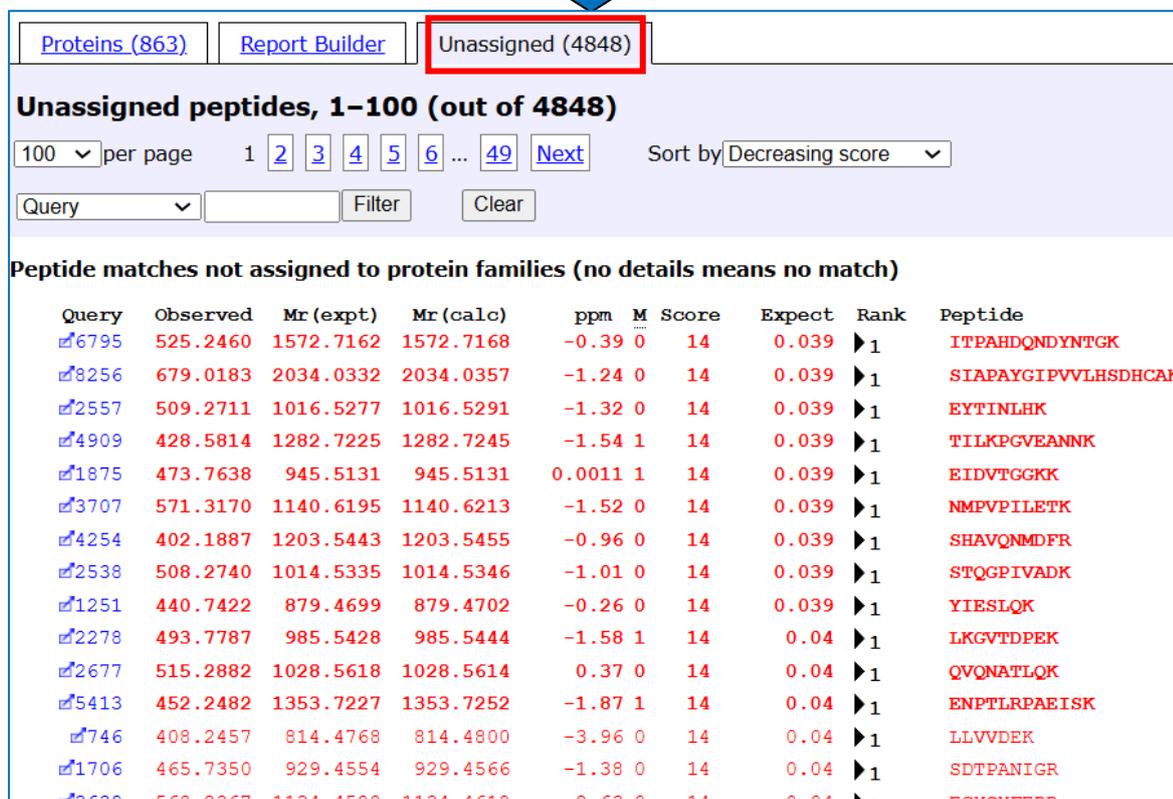
Proteins (863) Report Builder **Unassigned (4848)** [\\$ permalink](#)

Protein families 1-10 (out of 820)

10 per page 1 2 3 4 5 6 ... 82 Next Expand all Collapse all

Accession contains Find Clear

| | | | |
|----|-------------|-----|---|
| ▶1 | 3::P00549 | 652 | Pyruvate kinase 1 OS=Saccharomyces cerevisiae (strain ATCC... |
| ▶2 | 1 3::P10592 | 557 | Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strai... |
| | 2 3::P10591 | 554 | Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strai... |



Proteins (863) Report Builder **Unassigned (4848)**

Unassigned peptides, 1-100 (out of 4848)

100 per page 1 2 3 4 5 6 ... 49 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) | ppm | M | Score | Expect | Rank | Peptide |
|----------------------|----------|-----------|-----------|--------|---|-------|--------|------|---------------------|
| 6795 | 525.2460 | 1572.7162 | 1572.7168 | -0.39 | 0 | 14 | 0.039 | ▶1 | ITPAHQNDYNTGK |
| 8256 | 679.0183 | 2034.0332 | 2034.0357 | -1.24 | 0 | 14 | 0.039 | ▶1 | SIAPAYGIPVVLHSDHCAK |
| 2557 | 509.2711 | 1016.5277 | 1016.5291 | -1.32 | 0 | 14 | 0.039 | ▶1 | EYTLNLHK |
| 4909 | 428.5814 | 1282.7225 | 1282.7245 | -1.54 | 1 | 14 | 0.039 | ▶1 | TILKPGVEANNK |
| 1875 | 473.7638 | 945.5131 | 945.5131 | 0.0011 | 1 | 14 | 0.039 | ▶1 | EIDVTGGKK |
| 3707 | 571.3170 | 1140.6195 | 1140.6213 | -1.52 | 0 | 14 | 0.039 | ▶1 | NMPVPILETK |
| 4254 | 402.1887 | 1203.5443 | 1203.5455 | -0.96 | 0 | 14 | 0.039 | ▶1 | SHAVQNMDFR |
| 2538 | 508.2740 | 1014.5335 | 1014.5346 | -1.01 | 0 | 14 | 0.039 | ▶1 | STQGPVADK |
| 1251 | 440.7422 | 879.4699 | 879.4702 | -0.26 | 0 | 14 | 0.039 | ▶1 | YIESLQK |
| 2278 | 493.7787 | 985.5428 | 985.5444 | -1.58 | 1 | 14 | 0.04 | ▶1 | LKGVTDPEK |
| 2677 | 515.2882 | 1028.5618 | 1028.5614 | 0.37 | 0 | 14 | 0.04 | ▶1 | QVQNTLQK |
| 5413 | 452.2482 | 1353.7227 | 1353.7252 | -1.87 | 1 | 14 | 0.04 | ▶1 | ENPTLRPAEISK |
| 746 | 408.2457 | 814.4768 | 814.4800 | -3.96 | 0 | 14 | 0.04 | ▶1 | LLVVEK |
| 1706 | 465.7350 | 929.4554 | 929.4566 | -1.38 | 0 | 14 | 0.04 | ▶1 | SDTPANIGR |
| 3639 | 568.2367 | 1134.4588 | 1134.4618 | 2.63 | 0 | 14 | 0.04 | ▶1 | RYGVYEDR |

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。

Protein View: P10592

Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SSA2 PE=1 SV=3

Detailed information about this protein hit is shown below. [\(help\)](#)

Database: UP2311_S_cerevisiae
Score: 557
Monoisotopic mass (M_r): 69599
Calculated pI: 4.95

Sequence similarity is available as [an NCBI BLAST search of P10592 against nr](#).

Search parameters

MS data file: k1c_031308p_cptac_study6_6B011.mgf
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: [Carbamidomethyl \(C\)](#)
Variable modifications: [Oxidation \(M\)](#)

Protein sequence coverage: 44%

Matched peptides shown in **bold red**.

```

1 MSKAVGIDLG TTYSCVAHFS NDRVDIILAND QGNRTTPSFV GFTDTERLIG
51 DAAKNQAAMN PANTVFDAKR LIGRNFNDE VQGMKHFFP KLIDVDGKRPQ
101 IQVEFKGETK NFTPEQISSM VLGKMKETA SYLGAKVNDA VVTVPAYFND
151 SQRQATKDAG TIAGLNVLR INEPTAAATA YGLDKKKEE HVLIFDLGGG
201 TFDVSLLSIE DGIFEVKATA GDTHLGGEDF DNRLVNHFIQ EFKRKNKDL
251 STNQRALRR RTACERAKRT LSSSAQTSVE IDSLFEGIDF YTSITRARFE
301 ELCADLFRST LDPVEKVLRD ARLDKSGQVDE IVLVGGSTRI PKVQKLVTDY
351 FNGKEPNRSI NPDEAVAYGA AVQAAITLGD ESSKTQDLLL LDVAPLSLGI
401 ETAGGVMTKL IPRNSTIPTK KSEVFSYAD NQPGVLIQVF EGERAKTKDN
451 NLLGKPELSG IPPAPRGVPO IEVTFDSDN GILNVSAREK GTGKSNKITI
501 TNDKGRLSKE DIEKVVAAAE KPKEDDEKES QRIASKNQL SIAYSLKNTI
551 SEAGDKLEQA DKDAVTKGAE ETIAWLDSNT TATKEEFDDQ LKELQEVANP
601 IMSKLYQAGG APEGAAPGGF PGGAPPAPPA EGPTVEEVD
    
```

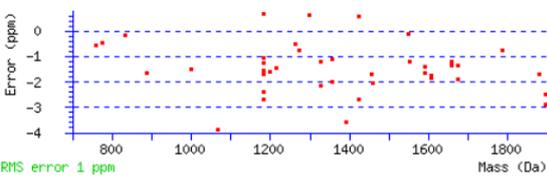
Unformatted sequence string: [639 residues](#) (for pasting into other applications).

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

| Query | Start - End | Observed | Mr (expt) | Mr (calc) | ppm | M | Score | Expect | Rank | U | Peptide |
|----------------------|-------------|----------|-----------|-----------|--------|---|-------|---------|------|---|-------------------------------------|
| 4341 | 24 - 34 | 607.8089 | 1213.6033 | 1213.6051 | -1.46 | 0 | 39 | 0.00013 | 1 | | R.VDIIANDQGNR.T |
| 6135 | 35 - 47 | 729.3478 | 1456.6810 | 1456.6835 | -1.71 | 0 | 29 | 0.0013 | 1 | U | R.TTPSFVGTDTTER.L |
| 6879 | 55 - 69 | 796.3790 | 1590.7434 | 1590.7460 | -1.64 | 0 | 32 | 0.00068 | 1 | U | K.NQAAMNPANTVFDAK.R |
| 6880 | 55 - 69 | 531.2552 | 1590.7438 | 1590.7460 | -1.39 | 0 | 19 | 0.013 | 1 | U | K.NQAAMNPANTVFDAK.R |
| 6962 | 55 - 69 | 804.3763 | 1606.7380 | 1606.7409 | -1.83 | 0 | 37 | 0.00019 | 1 | U | K.NQAAMNPANTVFDAK.R + Oxidation (M) |
| 6963 | 55 - 69 | 804.3763 | 1606.7381 | 1606.7409 | -1.77 | 0 | 20 | 0.0093 | 2 | U | K.NQAAMNPANTVFDAK.R + Oxidation (M) |
| 5694 | 75 - 86 | 697.3038 | 1392.5930 | 1392.5980 | -3.58 | 0 | 25 | 0.0029 | 1 | U | R.NFNDPEVQGMK.H |
| 6686 | 111 - 124 | 775.8977 | 1549.7809 | 1549.7810 | -0.099 | 0 | 22 | 0.006 | 1 | | K.NFTPEQISSMVLGK.M |
| 5251 | 125 - 136 | 664.3303 | 1326.6461 | 1326.6489 | -2.13 | 1 | 30 | 0.0011 | 1 | | K.MKETAESYLGAK.V |

=====中略=====

| | | | | | | | | | | | |
|----------------------|-----------|----------|-----------|-----------|-------|---|----|---------|---|---|----------------------|
| 4751 | 537 - 547 | 422.5625 | 1264.6656 | 1264.6663 | -0.52 | 0 | 26 | 0.0025 | 1 | | K.NQLESIAYSLK.N |
| 7936 | 568 - 584 | 626.9852 | 1877.9339 | 1877.9370 | -1.69 | 1 | 30 | 0.00095 | 1 | U | K.KAETIAWLDSNTTATK.E |
| 5444 | 593 - 604 | 679.8515 | 1357.6884 | 1357.6911 | -1.99 | 0 | 20 | 0.0092 | 2 | U | K.ELQEVANPIMSK.L |
| 5445 | 593 - 604 | 679.8521 | 1357.6896 | 1357.6911 | -1.10 | 0 | 23 | 0.0053 | 1 | U | K.ELQEVANPIMSK.L |



```

ID HSP72_YEAST Reviewed; 639 AA.
AC P10592; D6VXY0;
DT 01-JUL-1989, integrated into UniProtKB/Swiss-Prot.
DT 23-JAN-2007, sequence version 3.
DT 05-FEB-2025, entry version 242.
DE RecName: Full=Heat shock protein SSA2;
GN Name=SSA2; OrderedLocusNames=YLL024C; ORFNames=L0931;
OS Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast).
OC Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=559292;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=ATCC 204508 / S288c;
    
```

Protein View: P10592

Heat shock protein SSA2 OS=*Saccharomyces cerevisiae* (strain ATCC 204508 / S288c) OX=559292 GN=SSA2 PE=1 SV=3

Detailed information about this protein hit is shown below. ([help](#))

Database: UP2311_S_cerevisiae

Score: 557

Monoisotopic mass (M_r): 69599

Calculated pI: 4.95

Sequence similarity is available as [an NCBI BLAST search of P10592 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 (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 (上図では表示されていません)

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: klc_031308p_cptac_study6_6B011.mgf

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

Fixed modifications: [Carbamidomethyl \(C\)](#)

Variable modifications: [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: 44%

Matched peptides shown in **bold red**.

```
1 MSKAVGIDLG TTYSCVAHFS NDRVDIIAND QGNRTTPSFV GFTDTERLIG
51 DAAKNQAAMN PANTVFDAKR LIGRNFNDPE VQGMKHFPF KLIDVDGKPKQ
101 IQVEFKGETK NFTPEQISSM VLGKMKETAE SYLGAKVNDA VVTVPAYFND
151 SQRQATKDAG TIAGLNVLRI INEPTAAAIA YGLDKKGKEE HVLIFDLGGG
201 TFDVSLLSIE DGIFEVKATA GDTHLGGEDF DNRLVNHFIQ EFKRKNKKDL
251 STNQRALRRL RTACERAKRT LSSSAQTSVE IDSLFEGIDF YTSITRARFE
301 ELCADLFRST LDPVEKVLRD AKLDRSQVDE IVLVGGSTRI PKVQKLVTDY
351 FNGKEPNRSI NPDEAVAYGA AVQAAILTGD ESSKTQDLLL LDVAPLSLGI
401 ETAGGVMTKL IPRNSTIPTK KSEVPSTYAD NQPGVLIQVF EGERAKTKDN
451 NLLGKFELSG IPPAPRGVPO IEVTFDVDSN GILNVSAVEK GTGKSNKITI
501 TNDKGRLSKE DIERMVAEAE KFKEEDEKES QRIASRNQLE SIAYSSLKNTI
551 SEAGDKLEQA DKDAVTKAE ETIAWLDSNT TATKEEFDDQ LKELQEVANP
601 IMSKLYQAGG APEGAAPGGF PGGAPPAPEA EGPTVEEVD
```

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: [639 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) | ppm | M | Score | Expect | Rank | U | Peptide |
|----------------------|-------------|----------|-----------|-----------|--------|---|-------|---------|----------|---|------------------------------------|
| 4341 | 24 - 34 | 607.8089 | 1213.6033 | 1213.6051 | -1.46 | 0 | 39 | 0.00013 | <u>1</u> | | R.VDIIANDQGNR.T |
| 6135 | 35 - 47 | 729.3478 | 1456.6810 | 1456.6835 | -1.71 | 0 | 29 | 0.0013 | <u>1</u> | | R.TTPSFVGFDTTER.L |
| 6879 | 55 - 69 | 796.3790 | 1590.7434 | 1590.7460 | -1.64 | 0 | 32 | 0.00068 | <u>1</u> | | K.NQAAMNPANTVFDK.R |
| 6880 | 55 - 69 | 531.2552 | 1590.7438 | 1590.7460 | -1.39 | 0 | 19 | 0.013 | <u>1</u> | | K.NQAAMNPANTVFDK.R |
| 6962 | 55 - 69 | 804.3763 | 1606.7380 | 1606.7409 | -1.83 | 0 | 37 | 0.00019 | <u>1</u> | | K.NQAAMNPANTVFDK.R + Oxidation (M) |
| 6963 | 55 - 69 | 804.3763 | 1606.7381 | 1606.7409 | -1.77 | 0 | 20 | 0.0093 | <u>2</u> | | K.NQAAMNPANTVFDK.R + Oxidation (M) |
| 5694 | 75 - 86 | 697.3038 | 1392.5930 | 1392.5980 | -3.58 | 0 | 25 | 0.0029 | <u>1</u> | | R.NFNDPEVQGMK.H |
| 6686 | 111 - 124 | 775.8977 | 1549.7809 | 1549.7810 | -0.099 | 0 | 22 | 0.006 | <u>1</u> | | K.NFTPEQISSMVLGK.M |
| 5251 | 125 - 136 | 664.3303 | 1326.6461 | 1326.6489 | -2.13 | 1 | 30 | 0.0011 | <u>1</u> | | K.MKETAESYLGAK.V |
| 5253 | 125 - 136 | 443.2230 | 1326.6473 | 1326.6489 | -1.20 | 1 | 27 | 0.0019 | <u>1</u> | | K.MKETAESYLGAK.V |
| 3052 | 127 - 136 | 534.7619 | 1067.5093 | 1067.5135 | -3.88 | 0 | 26 | 0.0027 | <u>1</u> | | K.ETAESYLGAK.V |

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 : 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。

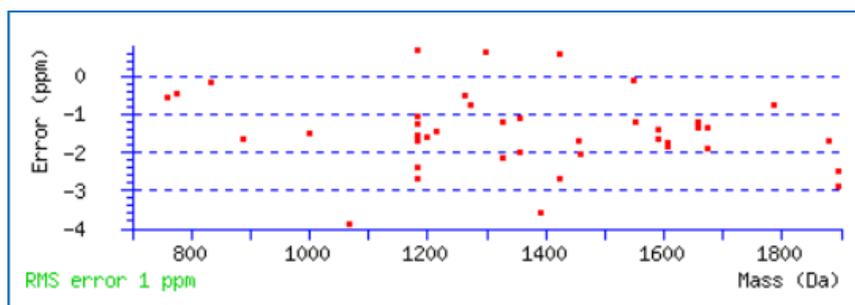
Score : Mascot Ions score、実験値の MS2 ピークと理論値とのマッチング度合いを表します。

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**です(上図)。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ができます。

```

ID   HSP72_YEAST                Reviewed;          639 AA.
AC   P10592; D6VXY0;
DT   01-JUL-1989, integrated into UniProtKB/Swiss-Prot.
DT   23-JAN-2007, sequence version 3.
DT   05-FEB-2025, entry version 242.
DE   RecName: Full=Heat shock protein SSA2;
GN   Name=SSA2; OrderedLocusNames=YLL024C; ORFNames=L0931;
OS   Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast).
OC   Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=559292;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC   STRAIN=ATCC 204508 / S288c;
RX   PubMed=2644626; DOI=10.1093/nar/17.2.805;
RA   Slater M.R., Craig E.A.;
RT   "The SSA1 and SSA2 genes of the yeast Saccharomyces cerevisiae.";

```

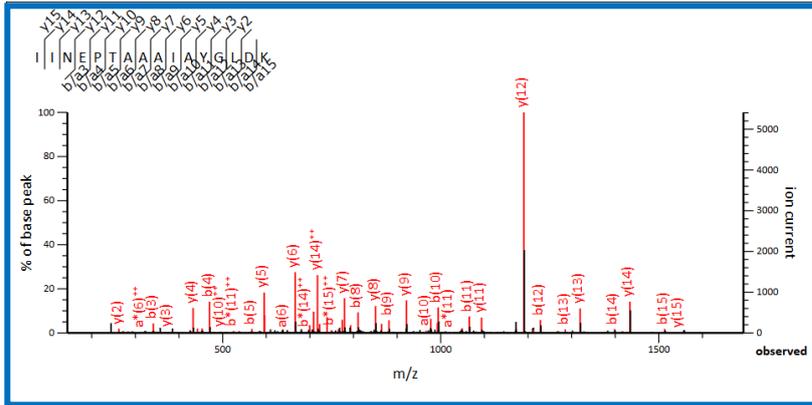
画面の最下部には**タンパク質の詳細情報**が表示されます(上図)。ただしデータベースに対して、情報表示に関する適切な設定が MASCOT 側で行われている時のみ表示されます。

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)。

Peptide View

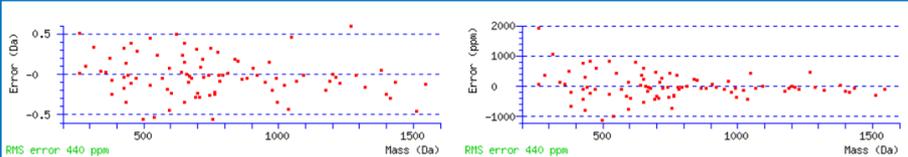
MS/MS Fragmentation of **IINEPTAAAIYGLDK**
 Found in **P10592** in **UP2311_S_cerevisiae**, Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292
 GN=SSA2 PE=1 SV=3
 Match to Query 7200: 1658.885668 from(830.450110,2+) intensity(29102.746) scans(9314) rawscans(sn9314) rtinseconds(4381.2834)
 index(6805)
 Title: 6806: Scan 9314 (rt=4381.28) [C:\Downloads\klc_031308p_cptac_study6_6B011.RAW]
 Data file klc_031308p_cptac_study6_6B011.mgf



144.19 to 1693.9

Monoisotopic mass of neutral peptide Mr(calc): 1658.8879
 Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
 Ions Score: 45 Expect: 3.2e-05
 Peak matches: 31/172 fragment ions using 42 most intense peaks
 Annotated fragments: 87/172 (help)

| # | a | a ⁺⁺ | a [*] | a ⁺⁺⁺ | b | b ⁺⁺ | b [*] | b ⁺⁺⁺ | Seq. | y | y ⁺⁺ | y [*] | y ⁺⁺⁺ | # |
|----|-----------|-----------------|----------------|------------------|-----------|-----------------|----------------|------------------|------|-----------|-----------------|----------------|------------------|----|
| 1 | 86.0964 | 43.5519 | | | 114.0913 | 57.5493 | | | I | | | | | 16 |
| 2 | 199.1805 | 100.0939 | | | 227.1754 | 114.0913 | | | I | 1546.8112 | 773.9092 | 1529.7846 | 765.3959 | 15 |
| 3 | 313.2234 | 157.1153 | 296.1969 | 148.6021 | 341.2183 | 171.1128 | 324.1918 | 162.5995 | N | 1433.7271 | 717.3672 | 1416.7005 | 708.8539 | 14 |
| 4 | 442.2660 | 221.6366 | 425.2395 | 213.1234 | 470.2609 | 235.6341 | 453.2344 | 227.1208 | E | 1319.6842 | 660.3457 | 1302.6576 | 651.8324 | 13 |
| 5 | 539.3188 | 270.1630 | 522.2922 | 261.6498 | 567.3137 | 284.1605 | 550.2871 | 275.6472 | P | 1190.6416 | 595.8244 | 1173.6150 | 587.3111 | 12 |
| 6 | 640.3665 | 320.6869 | 623.3399 | 312.1736 | 668.3614 | 334.6843 | 651.3348 | 326.1710 | T | 1093.5888 | 547.2980 | 1076.5623 | 538.7848 | 11 |
| 7 | 711.4036 | 356.2054 | 694.3770 | 347.6921 | 739.3985 | 370.2029 | 722.3719 | 361.6896 | A | 992.5411 | 496.7742 | 975.5146 | 488.2609 | 10 |
| 8 | 782.4407 | 391.7240 | 765.4141 | 383.2107 | 810.4356 | 405.7214 | 793.4090 | 397.2082 | A | 921.5040 | 461.2556 | 904.4775 | 452.7424 | 9 |
| 9 | 853.4778 | 427.2425 | 836.4512 | 418.7293 | 881.4727 | 441.2400 | 864.4462 | 432.7267 | A | 850.4669 | 425.7371 | 833.4403 | 417.2238 | 8 |
| 10 | 966.5619 | 483.7846 | 949.5353 | 475.2713 | 994.5568 | 497.7820 | 977.5302 | 489.2687 | I | 779.4298 | 390.2185 | 762.4032 | 381.7053 | 7 |
| 11 | 1037.5990 | 519.3031 | 1020.5724 | 510.7898 | 1065.5939 | 533.3006 | 1048.5673 | 524.7873 | A | 666.3457 | 333.6765 | 649.3192 | 325.1632 | 6 |
| 12 | 1200.6623 | 600.8348 | 1183.6358 | 592.3215 | 1228.6572 | 614.8322 | 1211.6307 | 606.3190 | Y | 595.3086 | 298.1579 | 578.2821 | 289.6447 | 5 |
| 13 | 1257.6838 | 629.3455 | 1240.6572 | 620.8322 | 1285.6787 | 643.3430 | 1268.6521 | 634.8297 | G | 432.2453 | 216.6263 | 415.2187 | 208.1130 | 4 |
| 14 | 1370.7678 | 685.8876 | 1353.7413 | 677.3743 | 1398.7627 | 699.8850 | 1381.7362 | 691.3717 | L | 375.2238 | 188.1155 | 358.1973 | 179.6023 | 3 |
| 15 | 1485.7948 | 743.4010 | 1468.7682 | 734.8877 | 1513.7897 | 757.3985 | 1496.7631 | 748.8852 | D | 262.1397 | 131.5735 | 245.1132 | 123.0602 | 2 |
| 16 | | | | | | | | | K | 147.1128 | 74.0600 | 130.0863 | 65.5468 | 1 |



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

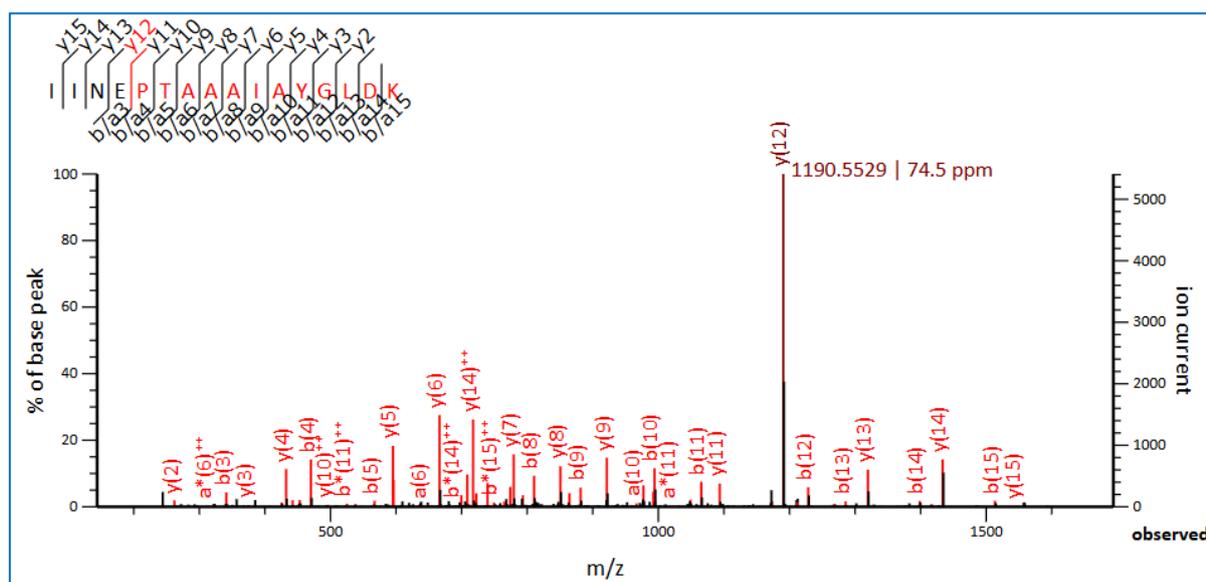
All matches to this query

| Score | Mr(calc) | Delta | Sequence |
|-------|-----------|---------|-----------------|
| 45.0 | 1658.8879 | -0.0022 | IINEPTAAAIYGLDK |

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 **IINEPTAAAIAYGLDK**
Found in **P10592** in **UP2311_S_cerevisiae**, Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SSA2 PE=1 SV=3
Match to Query 7200: 1658.885668 from(830.450110,2+) intensity(29102.746) scans(9314) rawscans(sn9314) rtinseconds(4381.2834) index(6805)
Title: 6806: Scan 9314 (rt=4381.28) [C:\Downloads\klc_031308p_cptac_study6_6B011.RAW]
Data file klc_031308p_cptac_study6_6B011.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。



続いて、**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** (上図)が表示されます。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。色の設定なども変更可能です。なお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。

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、別の場所で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): 1658.8879
 Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
 Ions Score: 45 Expect: 3.2e-05
 Peak matches: 31/172 fragment ions using 42 most intense peaks
 Annotated fragments: 87/172 ([help](#))

| # | a | a++ | a* | a*** | b | b++ | b* | b*** | Seq. | y | y++ | y* | y*** | # |
|----|-----------|----------|-----------|----------|-----------|----------|-----------|----------|------|-----------|----------|-----------|----------|----|
| 1 | 86.0964 | 43.5519 | | | 114.0913 | 57.5493 | | | I | | | | | 16 |
| 2 | 199.1805 | 100.0939 | | | 227.1754 | 114.0913 | | | I | 1546.8112 | 773.9092 | 1529.7846 | 765.3959 | 15 |
| 3 | 313.2234 | 157.1153 | 296.1969 | 148.6021 | 341.2183 | 171.1128 | 324.1918 | 162.5995 | N | 1433.7271 | 717.3672 | 1416.7005 | 708.8539 | 14 |
| 4 | 442.2660 | 221.6366 | 425.2395 | 213.1234 | 470.2609 | 235.6341 | 453.2344 | 227.1208 | E | 1319.6842 | 660.3457 | 1302.6576 | 651.8324 | 13 |
| 5 | 539.3188 | 270.1630 | 522.2922 | 261.6498 | 567.3137 | 284.1605 | 550.2871 | 275.6472 | P | 1190.6416 | 595.8244 | 1173.6150 | 587.3111 | 12 |
| 6 | 640.3665 | 320.6869 | 623.3399 | 312.1736 | 668.3614 | 334.6843 | 651.3348 | 326.1710 | T | 1093.5888 | 547.2980 | 1076.5623 | 538.7848 | 11 |
| 7 | 711.4036 | 356.2054 | 694.3770 | 347.6921 | 739.3985 | 370.2029 | 722.3719 | 361.6896 | A | 992.5411 | 496.7742 | 975.5146 | 488.2609 | 10 |
| 8 | 782.4407 | 391.7240 | 765.4141 | 383.2107 | 810.4356 | 405.7214 | 793.4090 | 397.2082 | A | 921.5040 | 461.2556 | 904.4775 | 452.7424 | 9 |
| 9 | 853.4778 | 427.2425 | 836.4512 | 418.7293 | 881.4727 | 441.2400 | 864.4462 | 432.7267 | A | 850.4669 | 425.7371 | 833.4403 | 417.2238 | 8 |
| 10 | 966.5619 | 483.7846 | 949.5353 | 475.2713 | 994.5568 | 497.7820 | 977.5302 | 489.2687 | I | 779.4298 | 390.2185 | 762.4032 | 381.7053 | 7 |
| 11 | 1037.5990 | 519.3031 | 1020.5724 | 510.7898 | 1065.5939 | 533.3006 | 1048.5673 | 524.7873 | A | 666.3457 | 333.6765 | 649.3192 | 325.1632 | 6 |
| 12 | 1200.6623 | 600.8348 | 1183.6358 | 592.3215 | 1228.6572 | 614.8322 | 1211.6307 | 606.3190 | Y | 595.3086 | 298.1579 | 578.2821 | 289.6447 | 5 |
| 13 | 1257.6838 | 629.3455 | 1240.6572 | 620.8322 | 1285.6787 | 643.3430 | 1268.6521 | 634.8297 | G | 432.2453 | 216.6263 | 415.2187 | 208.1130 | 4 |
| 14 | 1370.7678 | 685.8876 | 1353.7413 | 677.3743 | 1398.7627 | 699.8850 | 1381.7362 | 691.3717 | L | 375.2238 | 188.1155 | 358.1973 | 179.6023 | 3 |
| 15 | 1485.7948 | 743.4010 | 1468.7682 | 734.8877 | 1513.7897 | 757.3985 | 1496.7631 | 748.8852 | D | 262.1397 | 131.5735 | 245.1132 | 123.0602 | 2 |
| 16 | | | | | | | | | K | 147.1128 | 74.0600 | 130.0863 | 65.5468 | 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。

Peak matches: 31/172 fragment ions using 42 most intense peaks
 Annotated fragments: 87/172 ([help](#))

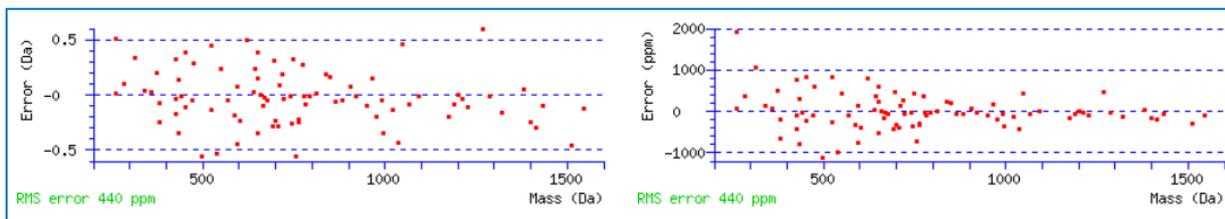
MASCOT では入力データについてピーク強度等を基に再構成を行います。入力データ(mgf などピークリストファイル)にあるすべてのピークがマッチング・スコアリングに使われるわけではありません。

172 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 **31**(表の中の赤い文字の数すべて)であった事を意味します。**42** most intense peaks は、**入力データ**のうち 42 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。”Annotated fragments”の分母である 172 は先ほど同様、理論フラグメントピーク全てで、**87** はスコアリングに使用されなかった実験データ側のピークをすべてカウントした場合の数値です。

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。5 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、スコアリングに大きく貢献
- 赤・太字** : イオンシリーズ単位ではぎりぎり有意と判定、スコアリングへの影響は小
- 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
- 青** : マッチングのピーク選定の際に選ばれなかった m/z でスコアリングにも使用されていないが、数値としては理論値と一致
- 黒** : マッチしていない、mgf 側にもピークがない

赤色の中で区別されるポイントは、**イオンシリーズ単位**で見てマッチした数が多いかどうか**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です。



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 [IINEPTAAAIAYGLDK](#)
 (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)。

All matches to this query

| Score | Mr(calc) | Delta | Sequence |
|-------|-----------|---------|----------------------------------|
| 45.0 | 1658.8879 | -0.0022 | IINEPTAAAIAYGLDK |

最後に、該当 query において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、rank 順に表示します(上図)。今回の例は refinement を実施したこともあり top1 のみ表示されています。

| Score | Mr(calc) | Delta | Sequence | Site Analysis |
|-------|-----------|--------|----------------------------------|-------------------|
| 83.4 | 1846.7179 | 0.1889 | DIGSESTEDQAMEDIK | Phospho S4 84.56% |
| 75.8 | 1846.7179 | 0.1889 | DIGSESTEDQAMEDIK | Phospho S6 14.73% |
| 62.7 | 1846.7179 | 0.1889 | DIGSESTEDQAMEDIK | Phospho T7 0.72% |
| 26.9 | 1846.7808 | 0.1261 | KLNSNPENYCESELK | |
| 22.8 | 1846.7729 | 0.1339 | KMEDSVGCLETAEEVK | |

Refinement を実施しない場合などで、該当 query とデータベース内に存在する他のペプチド配列との間におけるマッチング結果について、それぞれの配列とスコアが表示されます。また複数の候補が 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 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。上記例では 4 番目の S にリン酸基が付く確率が 84.56%、6 番目の S に付く確率が 14.73%である事を示しています。

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

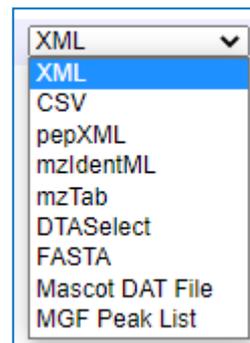
同定結果をファイル出力する事ができます。7-2-11 で紹介した Report builder では同定タンパク質に着目したファイル出力でしたが、**同定ペプチド情報をベースとしてまとめたファイル**や、その他解析結果に関連するファイルの出力をするには、7-2-3 で紹介した Export 機能を利用します。

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。

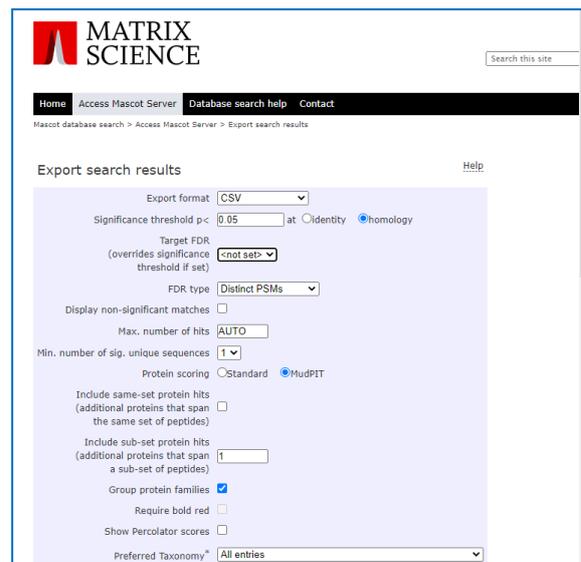


Export ボタンを押すと画面が遷移します(右図)。出力を希望するパラメーターを選択したうえで、画面最下部にある「**Export search results**」ボタンを押し、続いて現れる画面で「**Download**」ボタンを押す事で、サーバー側でのファイル作成とそのファイルのクライアントへのダウンロードが始まります。

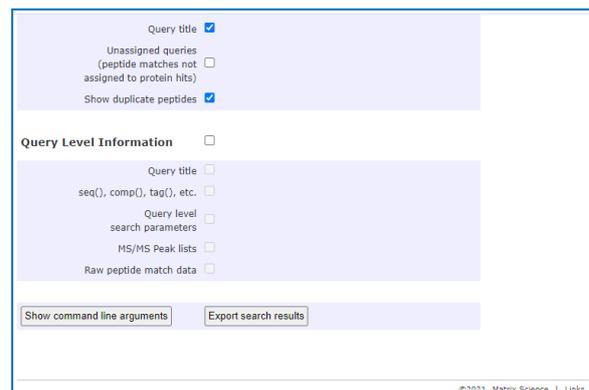
出力されるファイルの項目のうち、XML ファイル出力に関する詳細については以下2つの URL をご参照 ください。

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/index.html



===== 中略 =====



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。

| | |
|---------------------------|--|
| prot_hit_num | 同定タンパク質ファミリーの順位 |
| prot_family_member | 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 |
| prot_acc | タンパク質の Accession |
| prot_desc | タンパク質の Description |
| prot_score | 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) |
| prot_thresh | 同定基準値(タンパク質、PMF の時のみ) |
| prot_expect | 期待値(タンパク質、PMF のみ) |
| prot_mass | タンパク質の配列から計算された質量 |
| prot_matches | タンパク質にアサインされた query 数 |
| prot_matches_sig | タンパク質にアサインされた、有意基準を超える query 数 |
| prot_sequences | タンパク質にアサインされたペプチド数 |
| prot_sequences_sig | 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 |
| prot_cover | シーケンスカバレッジ |
| prot_len | タンパク質の残基長 |
| prot_pi | タンパク質の配列から計算された予測等電点 |
| prot_tax_str | 生物種情報 |
| prot_tax_id | NCBI の Taxonomy ID |
| prot_seq | タンパク質の配列 |
| prot_empai | emPAI |
| prot_acc_alpha | 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 |
| prot_acc_beta | 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 |
| pep_query | ペプチドの query 番号 |
| pep_rank | ペプチドの rank(マッチング順位) |
| pep_isbold | 1...有意基準を超えている、0...有意基準を超えていない |
| pep_isunique | 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 |
| pep_exp_mz | ペプチド・実験値側の m/z |
| pep_exp_mr | ペプチド・実験値側の質量 |
| pep_exp_z | ペプチド・実験値側の 電荷 |
| pep_calc_mr | ペプチド・理論値側の質量 |
| pep_delta | ペプチドの質量誤差、実験値 - 理論値 |
| pep_start | タンパク質におけるペプチド残基の位置、開始点 |
| pep_end | タンパク質におけるペプチド残基の位置、終了点 |
| pep_miss | 該当ペプチドにおける missed cleavage の数 |
| pep_score | Mascot Ions Score(query のマッチングスコア) |
| pep_homol | query の homology treshold(homology 同定基準値) |

| | |
|----------------------------------|--|
| pep_ident | query の identity threshold(identity 同定基準値) |
| pep_expect | query の期待値 |
| pep_res_before | 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 |
| pep_seq | ペプチドの配列 |
| pep_res_after | 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 |
| pep_frame | 翻訳のフレーム番号(塩基配列の検索時のみ) |
| pep_var_mod | Variable の修飾 |
| pep_var_mod_pos | 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。 |
| pep_summed_mod_pos | 「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。 |
| pep_local_mod_pos | Query レベルでの「pep_var_mod_pos」と同様の情報 |
| pep_var_mod_conf | 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 |
| pep_num_match | スコアに使用されたフラグメントマッチの数 |
| pep_scan_title | ピークリストの「title」行に記された情報 |
| pep_source | データベースの種類 (AA, NA, XA, SL) |
| pep_linked_sites | 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。 例: 1:0:6,2:5:7. 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。 |
| pep_res_before_beta | クロスリンクペプチド、ベータペプチドの pep_res_before |
| pep_seq_beta | クロスリンクペプチド、ベータペプチドの pep_seq |
| pep_res_after_beta | クロスリンクペプチド、ベータペプチドの pep_res_after |
| pep_frame_beta | クロスリンクペプチド、ベータペプチドの pep_frame |
| pep_var_mod_beta | クロスリンクペプチド、ベータペプチドの pep_var_mod |
| pep_var_mod_pos_beta | クロスリンクペプチド、ベータペプチドの pep_var_mod_pos |
| pep_summed_mod_pos_beta | クロスリンクペプチド、ベータペプチドの pep_summed_mod_pos |
| pep_local_mod_pos_beta | クロスリンクペプチド、ベータペプチドの pep_local_mod_pos |
| pep_monolink_mod_pos | 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 |
| pep_monolink_mod_pos_beta | 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」 |
| pep_loopl原因_pos | 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 |
| pep_loopl原因_pos_beta | クロスリンクペプチド、ベータペプチドにおける「pep_loopl原因_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]**
- 入力データセット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]**
- 検索対象のエントリー数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 **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する必要がある **[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種類の入力データのサブセット**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 側で強度情報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**ピーク抽出という1段階目の前処理が行われていることを前提**としています。また、上述のように**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するとよりスコアが高くなる、というような判断はありません。

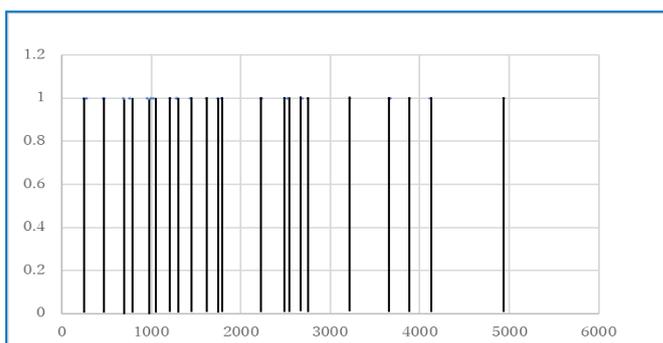
8-3. 配列から計算される理論ピーク値

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
```

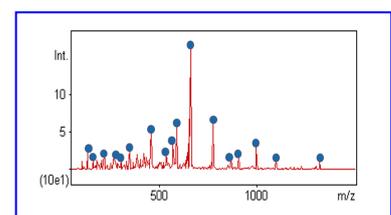


| | | | |
|---------|-----------|---|-------------------------------------|
| 2 - 18 | 1778.1070 | 0 | LLALALLLAFLLPPASQK |
| 2 - 25 | 2521.4632 | 1 | LLALALLLAFLLPPASQKSSNLEGR |
| 2 - 27 | 2750.6058 | 2 | LLALALLLAFLLPPASQKSSNLEGRTK |
| 19 - 25 | 761.3668 | 0 | SSNLEGR |
| 19 - 27 | 990.5094 | 1 | SSNLEGRTK |
| 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 |



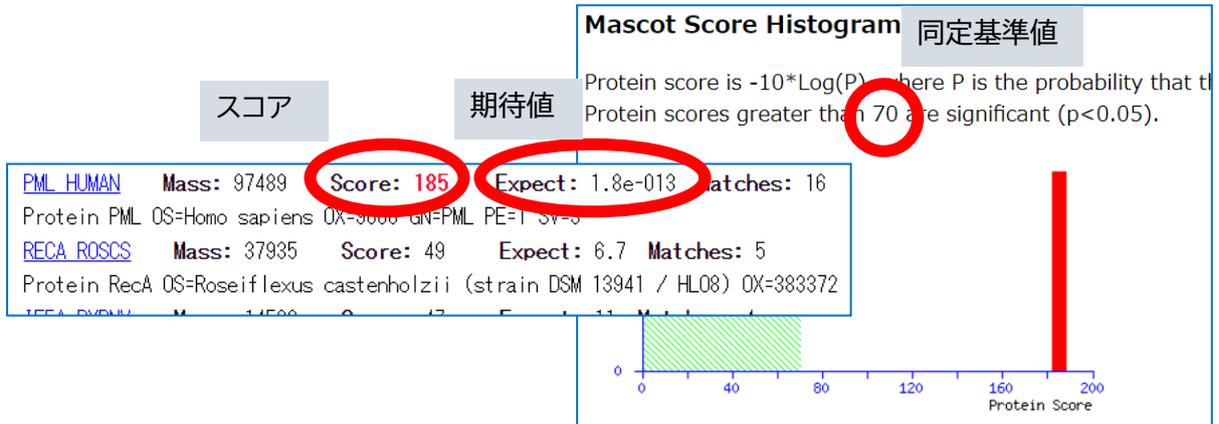
入力データ

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る事を表しています。MASCOT 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%を満たすスコアです。PMF の場合、同定基準値は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非常に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

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事があり、切断ペプチド単位でみても同じペプチド配列、同じ質量の組み合わせをもつタンパク質が存在する事があります。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。

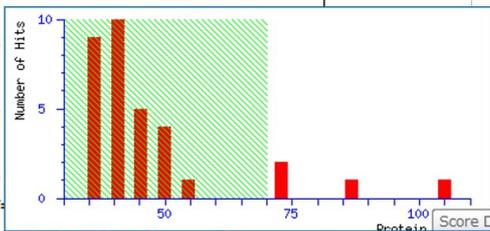
結果解釈の際 MASCOT Server がこれらのデータをどのように扱っているのか、ユーザー側が注意すべき点はどこか、についてご紹介します。

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とマッチしたケースを想定します。そのうち4つのタンパク質にフォーカスを当て、どのようにペプチドピークが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. | OPSD_HUMAN | Mass: 38866 | Score: 105 | Expect: 1.6e-05 | Matches: 9 |
| | Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1 | | | | |
| | OPSD_PHOVI | Mass: 38947 | Score: 45 | Expect: 16 | Matches: 5 |
| | Rhodopsin OS=Phoca vitulina GN=RHO PE=1 SV=1 | | | | |
| | OPSD_MOUSE | Mass: 39002 | Score: 43 | Expect: 28 | Matches: 5 |
| | Rhodopsin OS=Mus musculus GN=Rho PE=1 SV=1 | | | | |
| | OPSD_SHEEP | Mass: 38866 | Score: 43 | Expect: 28 | Matches: 5 |
| | Rhodopsin OS=Ovis aries GN=RHO PE=1 SV=2 | | | | |
| 2. | OPSD_MACFA | Mass: 39036 | Score: 88 | Expect: 0.00075 | Matches: 8 |
| | Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1 | | | | |
| | LUXS2_LACDB | Mass: 17500 | Score: 45 | Expect: 17 | Matches: 4 |
| | S-ribosylhomocysteine lyase 2 OS=Lactobacillus delbrueckii subsp. bulgari | | | | |
| 3. | OPSD_CRIGR | Mass: 39071 | Score: 73 | Expect: 0.027 | Matches: 7 |
| | Rhodopsin OS=Cricetulus griseus GN=RHO PE=1 SV=1 | | | | |
| | OPSD_CANFA | Mass: 38936 | Score: 56 | Expect | |
| | Rhodopsin OS=Canis familiaris GN=RHO PE=1 SV=1 | | | | |
| | OPSD_FELCA | Mass: 39023 | Score: 56 | Expect | |
| | Rhodopsin OS=Felis catus GN=RHO PE=1 SV=1 | | | | |
| | OPSD_RABIT | Mass: 38968 | Score: 56 | Expect | |
| | Rhodopsin OS=Oryctolagus cuniculus GN=RHO PE=1 SV=1 | | | | |
| | OPSD_SMICR | Mass: 39037 | Score: 56 | Expect: 1.2 | Matches: 6 |



○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ユニークペプチドの有無である事がわかります。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○のタンパク質 OPSD_MACFA と OPSD_CRIGR は、表の「■」が示すように、○のタンパク質 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、順位も別にして表示されます。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は MIS 検索の場合、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さを評価しているため意味合いがより重くなります。同定基準を超えているユニークなペプチドの存在は、それぞれのタンパク質が同定できたと考えるための根拠となります。

このように、PMF では類似タンパク質の同定(区別)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 の測定で両者を隔てるカギとなるペプチドを取り出して MS2 データを測定し 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ります。

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。

- 各 **query** の検索は独立して行われる。 [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 [9-4]
- 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]
- 結果の信頼度が **99%**以上となるように同定基準が調整される [9-4, 11 章]
- 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タンパク質を同定タンパク質とみなす [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]

以降、順に説明します。

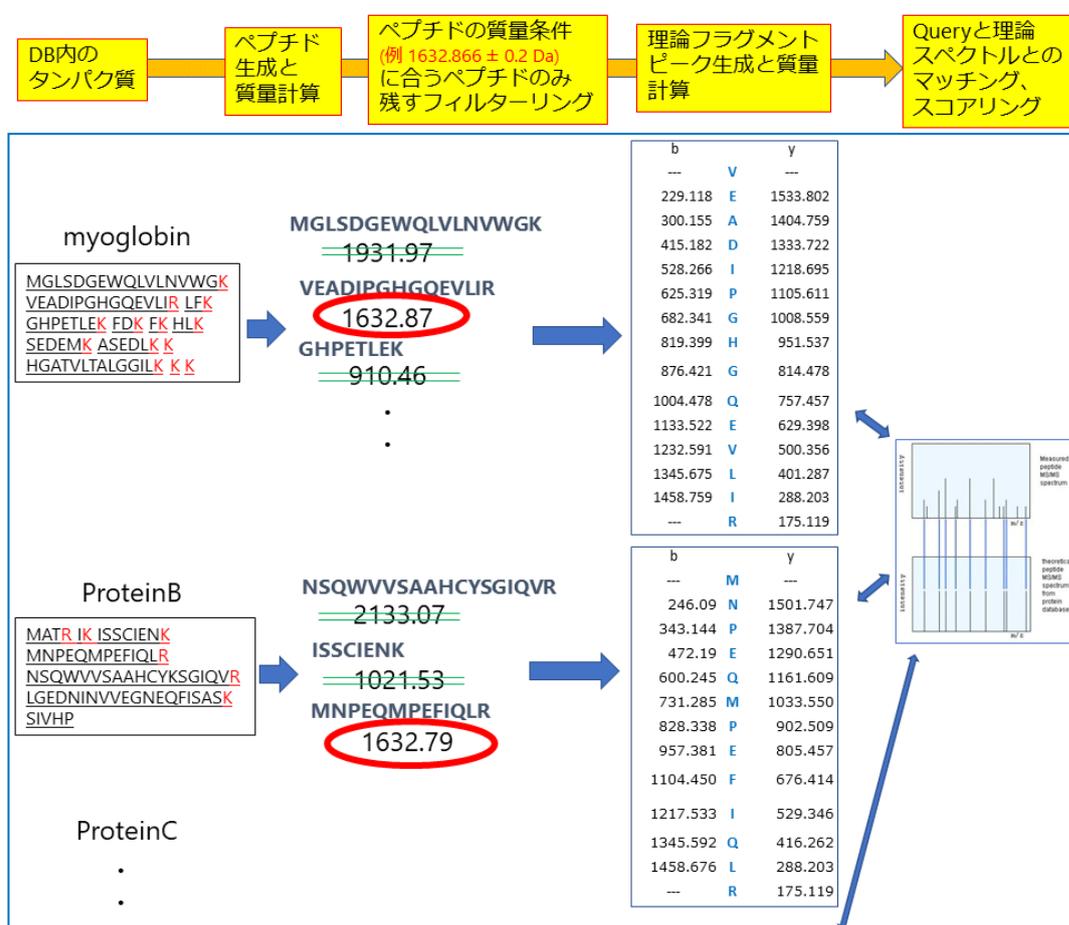
9-2. 入力データの準備

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。**

入力データの各 query データは、2段階で選別・組み分けされます。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 段階目の処理を行っていますが、これはあくまでも **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、**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するとよりスコアが高くなる、というような判断はありません。

9-3. 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(データベースの配列タイプが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。

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

9-4-1. refinement を実施しない場合のスコアや同定基準値

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、かつ検索時に refinement を実施しない場合です。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 |
| -0.5813 | 0 | | |
| 0.3187 | 1 | スコア | 期待値 |
| -0.7060 | 1 | 5 | 3.5 4 |

R.DFIDYYLIK.Q
DFPETNNILK
TPPIHRDLK
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)。**同定基準値は、FDR が 1%となるよう調整されます**。検索時、パラメーター「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**。

9-4-2 refinement を実施する場合のスコアや同定基準値

Refinement については、11 章にて詳しく説明します。

ここでは、Refinement を実施した場合、結果にて表示されるスコア、期待値、同定基準値がそれぞれ何を意味するのかについて説明します。

Refinement 実施により、Percolator プログラムが稼働します。Percolator は MASCOT に対して p-value, q-value, PEP(posterior error probability)を返します。Refinement を実施した場合、MASCOT の結果で示されるスコアや期待値は、PEP を基にしたスコアの表現に置き換わります。

スコア

$-10\log_{10}(\text{PEP})$

期待値

PEP

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。ペプチドスコアや期待値は**ペプチドの測定データが、マッチング対象であるペプチド配列である事の確認からしさを表していること**になります。

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 MASCOT 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)。この基準は現在各所で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した場合、False Positive が混ざってしまう可能性があるという問題があります (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** です。**MASCOT がレポートする同定タンパク質の信頼度のレベルがまさにこのレベル**ということになります。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、というのがその根拠です(ポワソン分布)。ただし必要な同定ペプチド数2にすると、同定タンパク質数は大きく低下します。求める信頼度に応じてユーザーが基準を使い分ける必要があります。

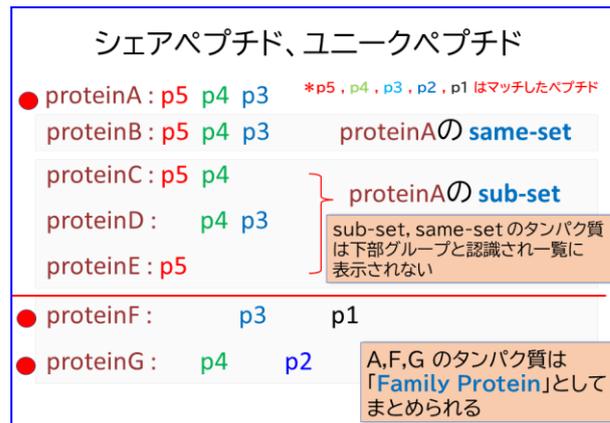
9-6. protein inference : ユニークシェア ペプチド、タンパク質のグループ化

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アサインされている場合を「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がアサインされている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 |
| p5 | LVQDVANNTNEEAGDGTTTATVLR | ■ | ■ | ■ | | ■ | | |
| p4 | ALMLQGVDLLADAVAVTMGPK | ■ | ■ | ■ | ■ | | | ■ |
| p3 | ISSIQSIVPALEIANHR | ■ | ■ | | ■ | | ■ | |
| p2 | VGGTSDVEVNEK | | | | | | | ■ |
| p1 | NAGVEGSLIVEK | | | | | | ■ | |

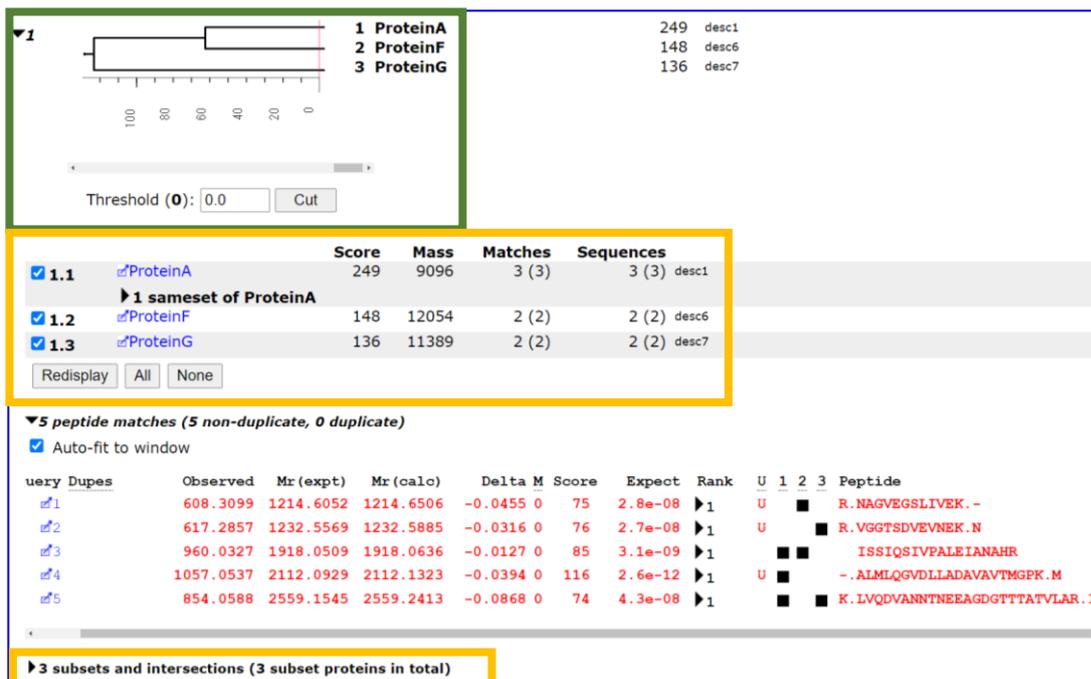
タンパク質 A を中心に、シェアされているペプチドの状況によって A 以外のタンパク質をどのようにみなすのかを記したのが次頁の図です。



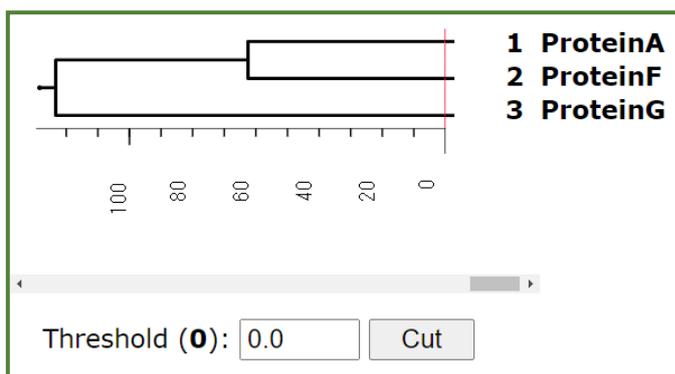
proteinB のように、proteinA と全く同じ組み合わせのペプチドが同定されているケースを **same-set protein** と呼んでいます。また proteinC, D, E のように proteinA の一部のペプチドが同定されていて包含関係的に proteinA の方が上位で proteinA に含まれないペプチドが存在しない場合、**sub-set protein** と呼んでいます。**same-set** や **sub-set** については代表タンパク質が1つのみが表示され、それ以外のタンパク質は結果表示画面に現れません。

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。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それらは「**Family Protein**」としてまとめられます。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。



前ページ下部の図について、以下に詳しく説明します。



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。

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。前頁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 ProteinA | 249 | 9096 | 3 (3) | 3 (3) desc1 |
| | ▼ 1 same set of ProteinA | | | | |
| | ProteinB | 249 | 9855 | 3 (3) | 3 (3) desc2 |
| <input checked="" type="checkbox"/> | 1.2 ProteinF | 148 | 12054 | 2 (2) | 2 (2) desc6 |
| <input checked="" type="checkbox"/> | 1.3 ProteinG | 136 | 11389 | 2 (2) | 2 (2) desc7 |

Redisplay All None

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

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使い方が異なる、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 ではすべて実施できません。

- 10-1. Spectral Library : ピークリスト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

初期のプロテオミクス・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

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. The 'Database(s)' dropdown is set to 'contaminants (AA)'. A list of databases is shown, 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発生パターンを変更させたり、検索対象を調整するようなパラメーターが使用すること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。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

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 $E(s)$ を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

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

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

■ 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

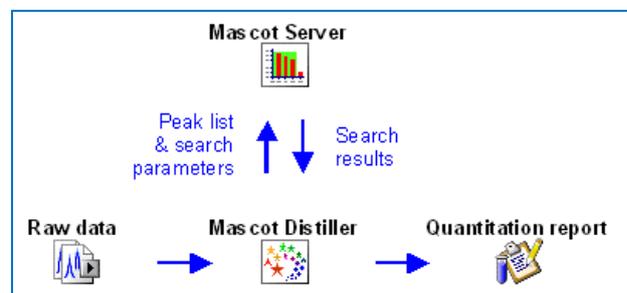
10-2-1. Quantitation 概要

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**ピークリストファイルに出力された情報、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ます。**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れば計算が可能です**(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

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 |
| Quantitation | 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)。

| 1.1 | 2::CO4B_HUMAN | Score | Mass | Matches | Sequences | emPAI | 114/113 | 115/113 | 116/113 | 117/113 | 118/113 | 119/113 | 121/113 | |
|-----|---------------|--------|--------|-------------|-----------|-------|---------|---------|---------|---------|---------|---------|---------|--------------|
| 1.1 | 2::CO4B_HUMAN | 164342 | 217600 | 3818 (3818) | 103 (103) | 48.75 | 1.033 | 1.070 | 1.045 | 1.016 | 1.155 | 1.051 | 1.055 | Complement C |
| 1.2 | 2::CO4A_HUMAN | 163856 | 217680 | 3814 (3814) | 102 (102) | 44.57 | 1.036 | 1.073 | 1.044 | 1.019 | 1.159 | 1.052 | 1.060 | Complement C |

下に拡大図①

下に拡大図②

| 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 | 0.0026 | 1 | | 0.909 | 1.155 | 0.945 | -0.096 | 1.064 | 1.202 | 0.791 | ■ | ■ | K.G |
| 58592 | 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 | 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 | 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 | 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 | 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 | 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 | 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 |

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 | Cc |
| 2) | 44.57 | 1.036 | 1.073 | 1.044 | 1.019 | 1.159 | 1.052 | 1.060 | Cc |

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 の中央値
 - **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 について中央値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)'. The main title is 'Protein family members (545 proteins)'. Below this, there is a dropdown menu for 'Columns: Standard' with '12 out of 58' items. An 'Arrangement' dropdown is set to '<custom>' with 'Load' and 'Make default' buttons. The interface is divided into two columns: 'Enabled' and 'Available'. The 'Enabled' column lists various fields like Family, Member, Database, Accession, Score, Mass, and Description. The 'Available' column lists statistical fields such as 'SD(geo) (117/113)', 'p-value (117/113)', 'Number of peptides (118/113)', 'Significant (118/113) (p-value < 0.05)', and 'Not-normal (118/113)'. There are left and right arrow buttons between the columns. At the bottom right, there is an 'Apply' button.

Proteins (545) Report Builder Unassigned (140931) S_permln

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を利用する場合、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です。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 -> Quantitation)で行いますのでそちらをご参照ください。

| 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_HUI | 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 C |
| 40 | 1 | 2 | SwissProt | CO4A_HUI | 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 C |
| 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_HUI | 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_HUI | 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_HL | 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_HUI | 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_HUI | 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_HUI | 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_HUI | 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_HUM | 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_HL | 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-glycop |
| 51 | 9 | 1 | SwissProt | CFAB_HUI | 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 |

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

■ **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

■ **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

■ **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

■ **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

■ **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

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。

予め設定しておく必要がある設定内容は以下の通りです。

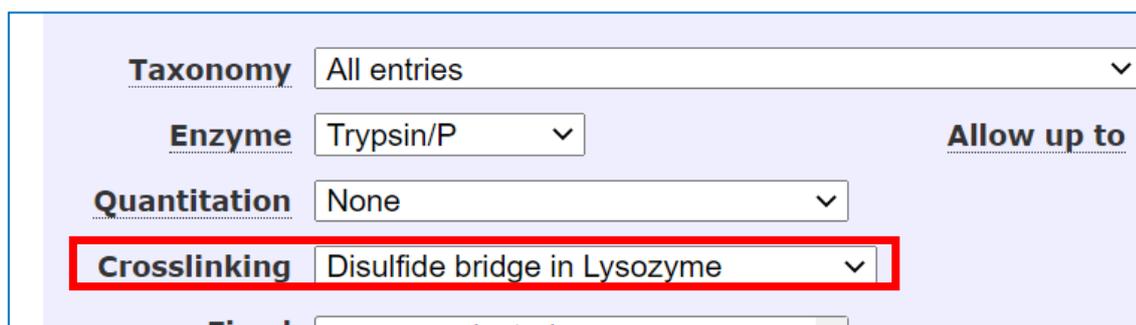
- ・ リンカーの組成、リンカーが付くアミノ酸残基
- ・ crosslink で使用するリンカー
- ・ リンカーの結合パターン（下記4種類）
- ・ リンカーが結合するタンパク質、またはデータベース

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 window with several dropdown menus. The 'Crosslinking' dropdown is highlighted with a red border and contains the text 'Disulfide bridge in Lysozyme'. Other visible dropdowns 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、2つのペプチドを組み合わせた数字が MASCOT Server 設定値の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。Crosslink 検索を実施する場合は必要に応じてこれらの検索の設定値を変更して(数値を増やして)ください。

■ 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ペプチドの質量の上限 (16kDa) という制限がありますのでご注意ください。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 検索と同時に進行することができません。また設定値ではありませんが検索に影響を及ぼしやすいのがペプチドの電荷です。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によるピークリスト作成機能をご利用ください。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

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 | 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.FESNFNTQATNRNEDGSDTYGILQINSR.W |
| 791.6272 | 3162.4795 | 3162.4602 | 6.10 | 1 | - | 37 | 0.00041 | ▶1 | U | ■ | K.FESNFNTQATNRNEDGSDTYGILQINSR.W |
| 680.3362 | 3396.6445 | 3396.6244 | 5.91 | 1 | 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 | 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 | 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 | 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)
.....

```

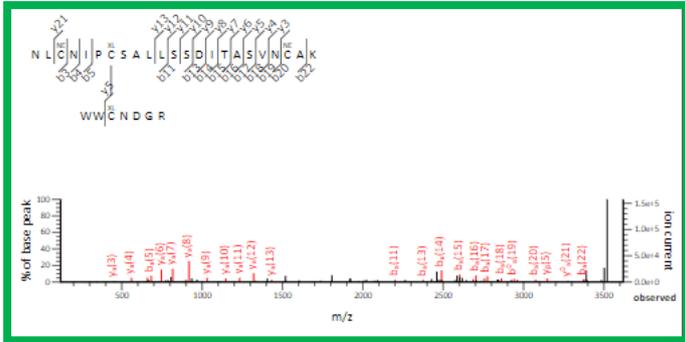
Query 番号をクリックするとマッチング内容を確認する peptide view ページを見る事ができます。次頁図は query1496 の peptide view です。

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\johnk\Downloads\Lysozyme_2p2m_1minute.raw]
 Data file Lysozyme_2p2m_1minute.temp.mgf

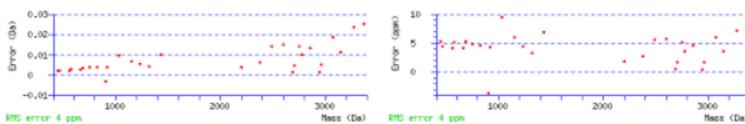


Label all possible matches Label matches used for scoring

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
 Intraot link : Alpha C7<-Xlink:Disulfide->C3 Beta
 Variable modifications:
 Alpha C3 : Methionine (O)
 Alpha C21 : Methionine (O)
 Ion Score: 75 Expect: 8.4e-08
 Matches : 30/234 Fragment ions using 54 most intense peaks [calc]

| # | b | b ⁺⁺ | b ⁺ | b ⁺⁺⁺ | b ⁰ | b ⁰⁺⁺ | Seq. | y | y ⁺⁺ | y ⁺ | y ⁺⁺⁺ | y ⁰ | y ⁰⁺⁺ | # |
|----|------------------|-----------------|----------------|------------------|------------------|------------------|------|------------------|-----------------|----------------|------------------|------------------|------------------|----|
| 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 | 456.1911 | 228.5992 | 439.1646 | 220.0859 | | | C | 3293.4482 | 1647.2278 | 3276.4217 | 1638.7145 | 3275.4377 | 1638.2225 | 21 |
| 4 | 570.2341 | 285.6207 | 553.2075 | 277.1074 | | | N | 3065.3914 | 1533.1993 | 3048.3648 | 1524.6861 | 3047.3808 | 1524.1940 | 20 |
| 5 | 683.3181 | 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 | 2200.9725 | 1100.9899 | 2183.9460 | 1092.4766 | 2182.9619 | 1091.9846 | L | 1433.6941 | 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 | 1320.6100 | 660.8086 | 1303.5835 | 652.2954 | 1302.5994 | 651.8034 | 12 |
| 13 | 2375.0366 | 1188.0219 | 2358.0100 | 1179.5086 | 2357.0260 | 1179.0166 | S | 1233.5780 | 617.2926 | 1216.5514 | 608.7794 | 1215.5674 | 608.2873 | 11 |
| 14 | 2490.0635 | 1245.5354 | 2473.0369 | 1237.0221 | 2472.0529 | 1236.5301 | D | 1146.5460 | 573.7766 | 1129.5194 | 565.2633 | 1128.5354 | 564.7713 | 10 |
| 15 | 2603.1476 | 1302.0774 | 2586.1210 | 1293.5641 | 2585.1370 | 1293.0721 | I | 1031.5190 | 516.2631 | 1014.4925 | 507.7499 | 1013.5084 | 507.2579 | 9 |
| 16 | 2704.1952 | 1352.6013 | 2687.1687 | 1344.0880 | 2686.1847 | 1343.5960 | T | 918.4349 | 459.7211 | 901.4084 | 451.2078 | 900.4244 | 450.7158 | 8 |
| 17 | 2775.2324 | 1388.1198 | 2758.2058 | 1379.6065 | 2757.2218 | 1379.1145 | A | 817.3873 | 409.1973 | 800.3607 | 400.6840 | 799.3767 | 400.1920 | 7 |
| 18 | 2862.2644 | 1431.6358 | 2845.2378 | 1423.1226 | 2844.2538 | 1422.6305 | S | 746.3502 | 373.6787 | 729.3236 | 365.1654 | 728.3396 | 364.6734 | 6 |
| 19 | 2961.3328 | 1481.1700 | 2944.3062 | 1472.6568 | 2943.3222 | 1472.1648 | V | 659.3181 | 330.1627 | 642.2916 | 321.6494 | | | 5 |
| 20 | 3075.3757 | 1538.1915 | 3058.3492 | 1529.6782 | 3057.3652 | 1529.1862 | N | 560.2497 | 280.6285 | 543.2232 | 272.1152 | | | 4 |
| 21 | 3303.4326 | 1652.2199 | 3286.4060 | 1643.7067 | 3285.4220 | 1643.2147 | C | 446.2068 | 223.6070 | 429.1802 | 215.0938 | | | 3 |
| 22 | 3374.4697 | 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 |

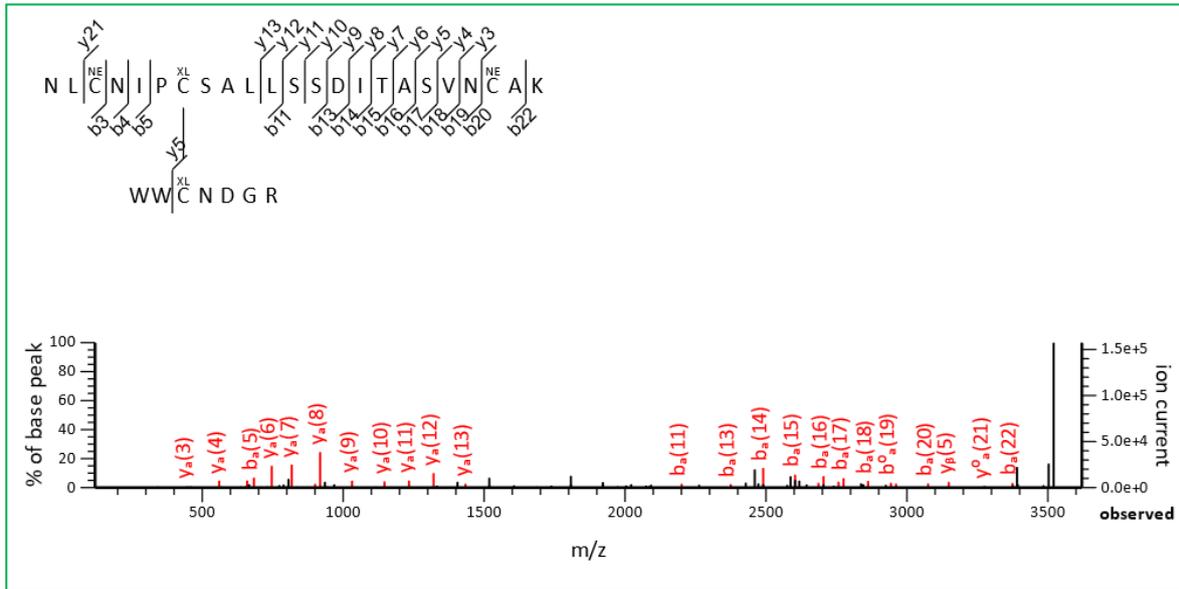
| # | b | b ⁺⁺ | b ⁺ | b ⁺⁺⁺ | b ⁰ | b ⁰⁺⁺ | Seq. | y | y ⁺⁺ | y ⁺ | y ⁺⁺⁺ | y ⁰ | y ⁰⁺⁺ | # |
|---|-----------|-----------------|----------------|------------------|----------------|------------------|------|------------------|-----------------|----------------|------------------|----------------|------------------|---|
| 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 | 3148.4166 | 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](#)

上図の緑で囲われた領域について、次頁以降にて詳しく説明します。

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 α 、 β 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。

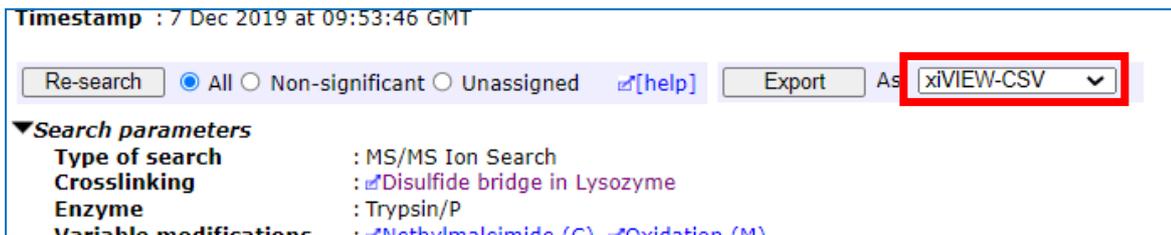


その下には理論値ベースのマッチング確認画面があり、これも両ペプチドそれぞれの理論値表が準備されています。

| # | b | b ⁺⁺ | b [*] | b ⁺⁺⁺ | b ⁰ | b ⁰⁺⁺ | Seq. | y | y ⁺⁺ | y [*] | y ⁺⁺⁺ | y ⁰ | y ⁰⁺⁺ | # |
|----|------------------|-----------------|----------------|------------------|------------------|------------------|------|------------------|-----------------|----------------|------------------|------------------|------------------|----|
| 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 | 456.1911 | 228.5992 | 439.1646 | 220.0859 | | | C | 3293.4482 | 1647.2278 | 3276.4217 | 1638.7145 | 3275.4377 | 1638.2225 | 21 |
| 4 | 570.2341 | 285.6207 | 553.2075 | 277.1074 | | | N | 3065.3914 | 1533.1993 | 3048.3648 | 1524.6861 | 3047.3808 | 1524.1940 | 20 |
| 5 | 683.3181 | 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 | 2200.9725 | 1100.9899 | 2183.9460 | 1092.4766 | 2182.9619 | 1091.9846 | L | 1433.6941 | 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 | 1320.6100 | 660.8086 | 1303.5835 | 652.2954 | 1302.5994 | 651.8034 | 12 |
| 13 | 2375.0366 | 1188.0219 | 2358.0100 | 1179.5086 | 2357.0260 | 1179.0166 | S | 1233.5780 | 617.2926 | 1216.5514 | 608.7794 | 1215.5674 | 608.2873 | 11 |
| 14 | 2490.0635 | 1245.5354 | 2473.0369 | 1237.0221 | 2472.0529 | 1236.5301 | D | 1146.5460 | 573.7766 | 1129.5194 | 565.2633 | 1128.5354 | 564.7713 | 10 |
| 15 | 2603.1476 | 1302.0774 | 2586.1210 | 1293.5641 | 2585.1370 | 1293.0721 | I | 1031.5190 | 516.2631 | 1014.4925 | 507.7499 | 1013.5084 | 507.2579 | 9 |
| 16 | 2704.1952 | 1352.6013 | 2687.1687 | 1344.0880 | 2686.1847 | 1343.5960 | T | 918.4349 | 459.7211 | 901.4084 | 451.2078 | 900.4244 | 450.7158 | 8 |
| 17 | 2775.2324 | 1388.1198 | 2758.2058 | 1379.6065 | 2757.2218 | 1379.1145 | A | 817.3873 | 409.1973 | 800.3607 | 400.6840 | 799.3767 | 400.1920 | 7 |
| 18 | 2862.2644 | 1431.6358 | 2845.2378 | 1423.1226 | 2844.2538 | 1422.6305 | S | 746.3502 | 373.6787 | 729.3236 | 365.1654 | 728.3396 | 364.6734 | 6 |
| 19 | 2961.3328 | 1481.1700 | 2944.3062 | 1472.6568 | 2943.3222 | 1472.1648 | V | 659.3181 | 330.1627 | 642.2916 | 321.6494 | | | 5 |
| 20 | 3075.3757 | 1538.1915 | 3058.3492 | 1529.6782 | 3057.3652 | 1529.1862 | N | 560.2497 | 280.6285 | 543.2232 | 272.1152 | | | 4 |
| 21 | 3303.4326 | 1652.2199 | 3286.4060 | 1643.7067 | 3285.4220 | 1643.2147 | C | 446.2068 | 223.6070 | 429.1802 | 215.0938 | | | 3 |
| 22 | 3374.4697 | 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 ⁺⁺ | b [*] | b ⁺⁺⁺ | b ⁰ | b ⁰⁺⁺ | Seq. | y | y ⁺⁺ | y [*] | y ⁺⁺⁺ | y ⁰ | y ⁰⁺⁺ | # |
|---|-----------|-----------------|----------------|------------------|----------------|------------------|------|------------------|-----------------|----------------|------------------|----------------|------------------|---|
| 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 | 3148.4166 | 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 |

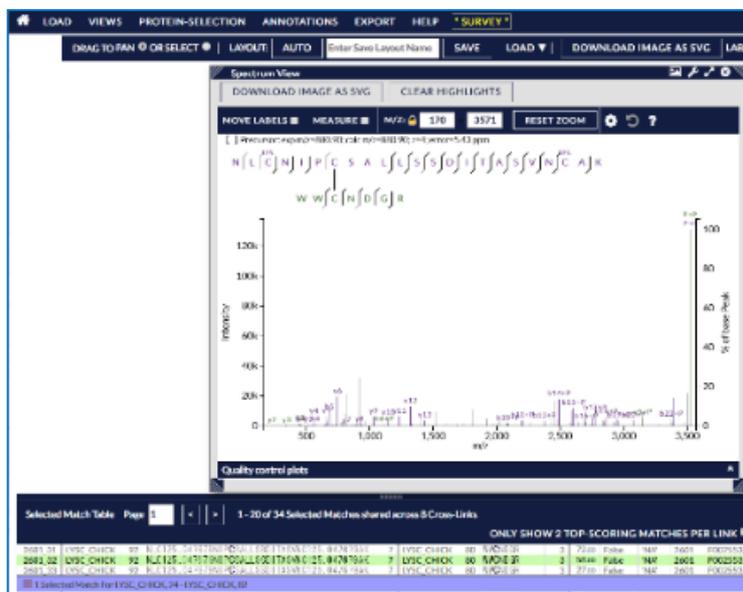
Summary の結果画面に話を戻します。crosslink 検索を行った場合、ファイル出力オプションで「xiVIEW-CSV」というファイルが出力可能となります。



このファイルは、サイト xiVIEW

https://xiview.org/xiNET_website/index.php

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。私たちは目安として、**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

* 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

■ DSS 結合検索結果例

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

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、通常の方法で検索を行います。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

まず検索パラメーターで「**Error tolerant**」にチェックを入れます。すると、考慮する修飾について機能別にグループ化されたリストが表示されます。検索対象となる修飾グループを選択し、「>」ボタンを使って右側へ移動させます。考慮する修飾グループを選択し終わったら、画面下の Start Search ボタンを押して検索を実行します。

The screenshot shows the search interface with the following details:

- Variable modifications:** --- none selected ---
- Modification classes (right):** Acetyl (N-term), Acetyl (Protein N-term), Acetyl (S), Alexa488 (Q)
- Error tolerant:** Automatic second pass search of selected modification classes
- Modification list (left):** Multiple (20), N-linked glycosylation (210), Non-standard residue (12), O-linked glycosylation (290), Other (77), Other glycosylation (24)
- Modification class list (right):** Chemical derivative (691)
- Peptide tol.:** 1.2 Da
- MS/MS tol.:** 0.6 Da

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 hold=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)。

| | Score | Mass | Matches | Sequences | | |
|---|------------|------|---------|-----------|---------|---|
| <input checked="" type="checkbox"/> 2.1 | PPBI_HUMAN | 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 | PPBN_HUMAN | 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 | PPBI_HUMAN | 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 |
| 126 | | 726.1806 | 1450.3465 | 1450.6477 | -0.3011 | 0 | 69 | 7e-06 | 1 | | | | | R.NWYSADVDPASAR.Q |
| 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.NWYSADVDPASAR.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] |

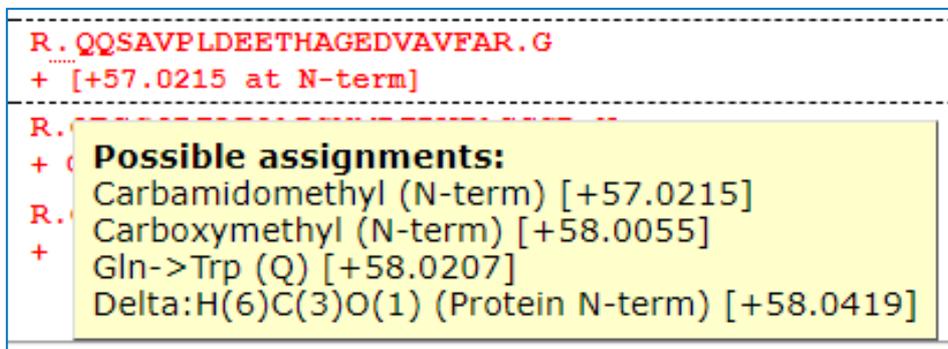
結果には Expect の値も表示されています。2 段階目の検索結果は 1 段階目で同定済みの結果に比べ、同定基準が高めになっています。

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。このような修飾を排除したい場合、検索実行時の修飾グループの選択についてご検討ください。



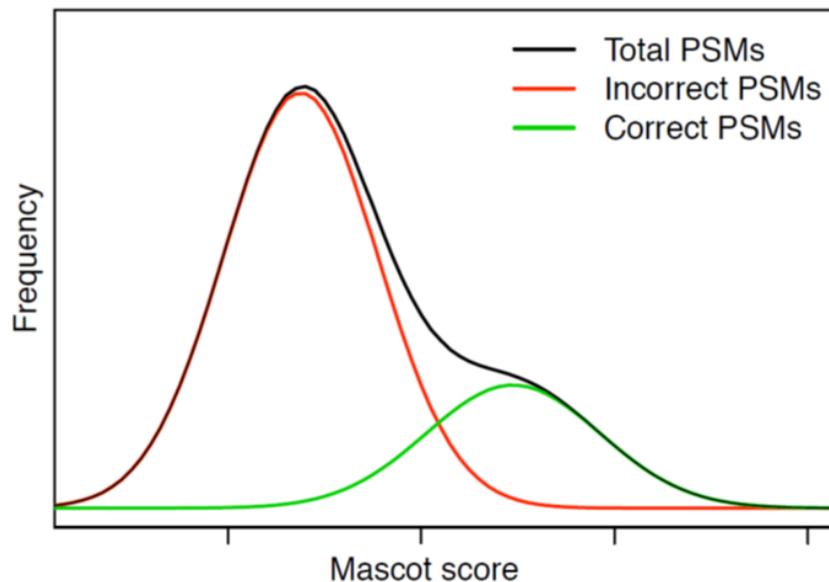
Error Tolerant Search は便利な検索であり、同定基準値に基づいた期待値も表示されていますが、結果の採用については慎重な判断が必要です。このページの結果に対する議論については以下 HELP ページをご覧ください。

https://www.matrixscience.com/help/error_tolerant_help.html#RESULTS

11. 機械学習による結果の精査 (refinement)

11-1. 機械学習で行う結果の精査(refinement)の概要

Mascot Server には **Percolator**、**MS²Rescore (+DeepLC,MS²PIP)**が搭載されています。Percolator は半教師付き機械学習アルゴリズムであり、正しいスペクトル識別と間違っただスペクトル識別の識別を改善するために利用します。タンパク質配列データベースに対して検索を行った場合、結果は常に正しいペプチドと間違っただペプチドが混在します。ピークは様々な条件により出たりでなかったりすることもあり、MASCOT のスコアリングだけで正解と不正解のペプチド情報を分離する事を目指すのは難しいと言えます。下図のように、不正解と正解のケースにおけるスコア分布において、重なり合っている領域があります。この領域をできるだけ小さくして正解と不正解の分離を容易にすることが重要です。



Percolator を適用するメリットは、**Mascot** スコア以外の要素をうまく活用し、**正解と不正解の区別を進め、より多くのペプチドを拾い上げる事ができる**という事です。正解のデータと不正解のデータでは、スコア以外の要素(features)が系統的に異なることがしばしばあります。

refinement 実施の際、学習のスタート地点として検索結果を正解データと不正解データに分けます。正解のデータは通常データベース(Target データベース)の高スコアのマッチングを、不正解のデータは Decoy データベースへの検索結果を利用します。Decoy データベースに関する詳細は「**11-2-3. Decoy データベース**」の項目で説明しています。Mascot スコア以外の各要素(features)の情報を活用して、正解と不正解のデータをより良く区分するのに最適な結果を生み出す方法を検討します(この過程について、本資料ではしばしば「再スコアリング」と呼んでいます)。

refinement を実施する際、**ペプチド配列情報から計算される予測保持時間情報や予測 MS² スペクトル情報を利用する事ができます**。これらを実施するためのプログラムが **DeepLC,MS²PIP** で、その**2つの**

プログラムと **Percolator** の間をつなぐのが **MS²Rescore** です。これらプログラムを利用する際、より適切な結果を生み出すためのパラメーターがセットされている「モデル」を指定する必要があります。**モデルはそのパラメーターを算出するために使用したデータセット群を基に名前が付けられています。**モデルの選択は、検索前にパラメーターとして指定をするか、検索後に結果画面の中にある表示コントロールの画面から各種項目を選択して下さい。

11-2. refinement 実施の必要要件とデータの流れ

11-2-1. refinement 計算が実施可能な条件

Mascot Server では Refinement を実施するために以下の条件を定めています。

- MS/MS 検索である (PMF ではない)。
- Decoy データベースへの結果結果が含まれている (ver.3.0 以降は自動的に実行されます)。
- **750 以上のクエリー**
- データベースに登録されている**エントリー数が 100 以上**である
- Error tolerant 検索ではない

上述の条件のうち、クエリー数とデータベースエントリー数については、MASCOT サーバーの設定を調整することにより、変更する事ができます(ただし弊社としては基本的にデフォルト設定値での利用をお勧めしています)。以下の項目です。

PercolatorMinQueries :750

refinement 実施を許可する最小クエリー数。

PercolatorMinSequences : 100

refinement 実施を許可するデータベースエントリーの最小数。

また Percolator 計算には基本的に rank1 のペプチド配列が利用されますが、それ以外の rank のペプチドを利用するかどうかについて、以下のパラメーターにて調整が可能です。

PercolatorTargetRankScoreThreshold :

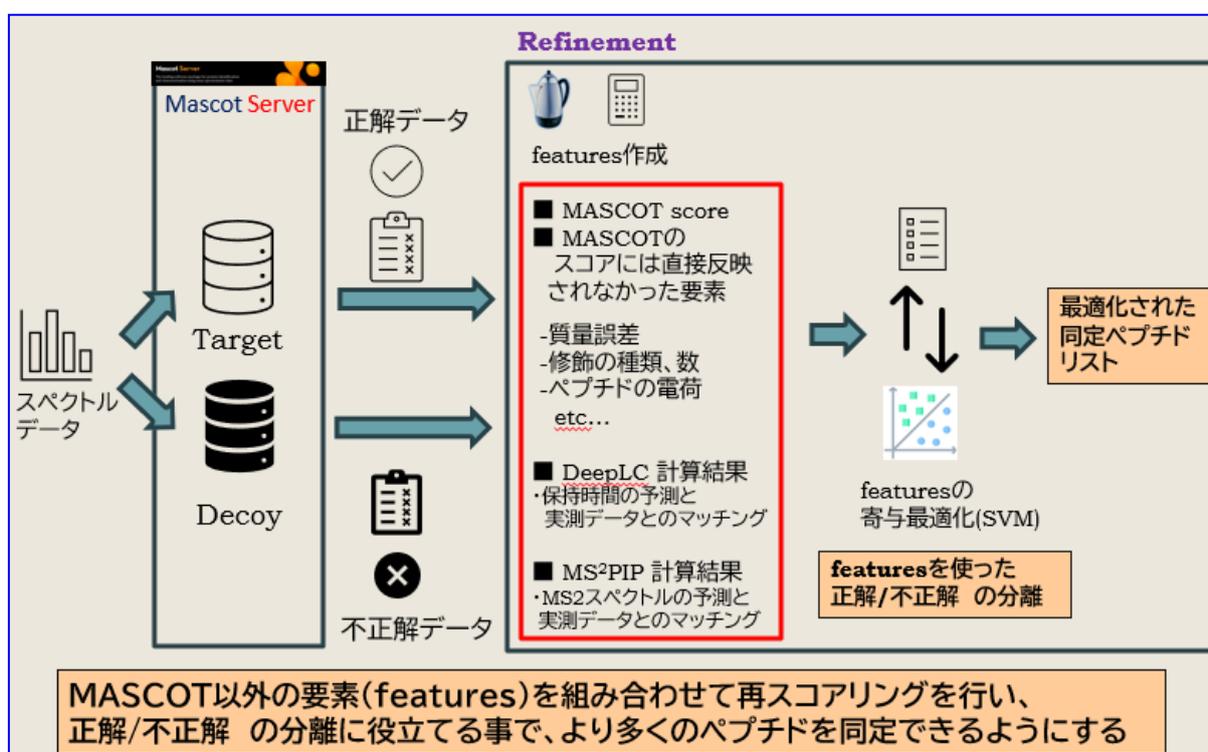
スコアがこの値より小さい場合、ランク 1 以下のマッチは使用されません。

PercolatorTargetRankRelativeThreshold :

スコアの差をランク 1 のスコアで割った値がこの値より大きい場合、ランク 1 以下のマッチは使用されません(デフォルト 0.2)

11-2-2. refinement 計算のワークフロー

Refinement 実施の計算の流れを示した以下の図です。



検索時、Target データベースとは別に Decoy データベースに対しても検索を行います。Decoy データベースの詳細については、「11-2-3. Decoy データベース」をご覧ください。

検索結果を受け取った後、refinement 計算が本格的に実施されます。まず、各 query/ペプチド配列について、Mascot score 以外の様々な要素(features)を計算します。さらに必要に応じて、予測保持時間を計算するプログラム DeepLC や予測 MS² スペクトルを計算するプログラム MS²PIP を利用した理論値と、実測データとの一致度などを feature として追加します。features を使い、refinement 計算を実施します。features については「11-3. features」をご覧ください。

Percolator では Target データベースのうちスコアが高いものを正解データとしてタグをつけ、Decoy データベースの結果については不正解データとしてタグをつけます。features 情報の寄与の割合を調整しながら、正解データと不正解データが一番きれいに分離されるような値を繰り返し探求します。これらの過程を経て最適と判断されたパラメーターを利用して認定されたペプチドを同定ペプチドとします。

[関連する論文]

Kall, L., et al., Semi-supervised learning for peptide identification from shotgun proteomics datasets, Nature Methods 4 923-925 (2007)

Kall, L., et al., Posterior error probabilities and false discovery rates: Two sides of the same coin, Journal of Proteome Research 7 40-44 (2008)

Kall, L., et al., Assigning significance to peptides identified by tandem mass spectrometry using decoy databases, Journal of Proteome Research 7 29-34 (2008)

Kall, L., et al., Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry, Bioinformatics 24 I42-I48 (2008)

Brosch, M., et al., Accurate and Sensitive Peptide Identification with Mascot Percolator, Journal of Proteome Research 8 3176-3181 (2009)

Spivak, M., et al., Improvements to the Percolator Algorithm for Peptide Identification from Shotgun Proteomics Data Sets, Journal of Proteome Research 8 3737-3745 (2009)

11-2-3. decoy データベース

Decoy データベースは、Target データベース(検索対象となるデータベース、通常のデータベース)の各エントリーの配列を ランダムまたは逆向きに変更する事で作成されます。

Decoy データベースは、「データベースの性質」が Target データベースと同じである事が重要なポイントです。「データベースの性質」とは以下の内容の事を示します。

- データベースに登録されているエントリー数
- 登録エントリーのアミノ酸残基長分布
- データベースに含まれる各アミノ酸の出現頻度

検索データも同じ、検索パラメーターも同じ、そしてデータベースの性質も同じであることが重要です。配列だけが通常と異なり正解配列と同一のものを含む可能性が極端に小さい Decoy データベースを利用することで、その検索結果を Target データベースとの対比として利用するうえで支障がなくなります。必要な条件を備えているため、デコイマッチがターゲットデータベースの不正確なマッチをモデル化したものであると考える事ができます。

11-2-4. FDR (q-value)

Percolator は MASCOT に PEP, q-value の値を返します。PEP は個々のペプチドにおける誤同定の確率を表します。一方 q-value は該当ペプチドより高いスコアを見たときに、FDR の値がどうなっているかを示す数値です。FDR という用語は q-value とほぼ同じ内容を示しますが、特に「同定基準値」として見た場合に使用されています。

MASCOTは refinement を実施している場合、スコア、期待値を以下のように置き換えて結果表示します。

スコア : $-10\log_{10}(\text{PEP})$

期待値 : PEP

同定基準は検索パラメーターの時に指定する”Target FDR”の値となり、通常は 1%です。

Target,Decoy 両データベースで検索した際のスコアを良い順に並べ、Normal の結果の累積数を分母に、Decoy の検索結果を分子にした割合(%)が q-value または FDR です(下図)。

Refinement : ペプチドのスコア、同定基準、q-value (またはFDR)

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

| Sequence | Score (PEPを反映) | Normal or Decoy | N累計 | D累計 | q-value(FDR) [D/N] (%) |
|-------------------|----------------|-----------------|-----|-----|------------------------|
| NAGVEGSLIVEK | 52 | N | 1 | 0 | 0 |
| VEEVIVTK | 51 | N | 2 | 0 | 0 |
| TLNDELEIIEGMK | 34 | N | 118 | 0 | 0 |
| MATRIK | 33 | D | 118 | 1 | 0.847 |
| ISSIQSIVPALEIANHR | 31 | N | 119 | 1 | 0.840 |
| ⋮ | | | | | |
| VGLQVVAVK | 19 | N | 512 | 5 | 0.977 |
| TAKAESK | 18 | D | 512 | 6 | 1.171 |
| LSDGVAVLK | 17 | N | 513 | 6 | 1.170 |
| SCAFFLDK | 16 | D | 513 | 7 | 1.365 |

スコア : $-10\log_{10}(\text{PEP})$
期待値 : PEP

FDR 1%を同定基準値として、q-valueが 1%より小さいペプチドは同定ペプチドとなる

どのような基準を適用すると良いか？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ると、データがどのように処理されたかを確認する事ができます。

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。

- HUPO

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

- MCP Guidelines

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

FDR の適用など同定基準についてどのようなパターンがあるか調べるため、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。下図は 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 Proteomics 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/eahme/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](#)

11-3. features

Percolator の計算には、Mascot Score 以外の様々な要素(features)が利用されます。ここではその features の項目について説明します。

| Feature name | Description |
|------------------|---|
| retentionTime | 保持時間 |
| dM | ペプチド質量の誤差、[理論値] - [実測値] (Da) |
| mScore | Mascot スコア(always on) |
| IgDScore | 次ランクとの Mascot スコア差 |
| mrCalc | ペプチド質量の理論値 |
| charge | 電荷 |
| dMppm | ペプチド質量の誤差、[理論値] - [実測値] (ppm) |
| absDM | ペプチド質量の誤差、[理論値] - [実測値] (Da)、絶対値 |
| absDMppm | ペプチド質量の誤差、[理論値] - [実測値] (ppm)、絶対値 |
| isoDM | ペプチド質量の誤差、[理論値] - [実測値] (Da)、同位体ピークの可能性を考慮 |
| isoDMppm | ペプチド質量の誤差、[理論値] - [実測値] (ppm)、同位体ピークの可能性を考慮 |
| isoDmz | ペプチド m/z の誤差、[理論値] - [実測値] (Da)、絶対値 (注:質量ではなく) |
| isoSysDM | isoDM の要素に加え全ペプチドの誤差を基に全体をオフセット調整 |
| isoSysDMppm | isoDMppm の要素に加え全ペプチドの誤差を基に全体をオフセット調整 |
| isoSysDmz | isoDmz の要素に加え全ペプチドの誤差を基に全体をオフセット調整 |
| mc | ペプチド切断設定(Missed Cleavage)の適用数。No enzyme の時は 0 |
| varmods | 修飾設定の適用数を修飾設定が適用されうるアミノ酸残基の数で割った値。 |
| varcount | 修飾設定の種類数 |
| varmodsCount | 修飾の種類数 |
| modifiable | 修飾設定が適用されうるアミノ酸残基の数 |
| modified | 修飾設定が適用されうるアミノ酸残基あるいはペプチド末端の数 |
| totInt | 入力データで 100Da 毎に 20 ピーク選んだ際の、ピーク強度の総和の log。 |
| intMatchedTot | マッチしたピーク全体のピーク強度総和の log |
| relIntMatchedTot | 入力データで 100Da 毎に 20 ピーク選んだ際の、ピーク強度の総和を、マッチしたピーク全体のピーク強度総和 で割った割合 |
| fragDeltaMed | フラグメントピークの誤差の中央値(Da) |
| fragDeltaIqr | フラグメントピークマッチの誤差、四分位範囲(Da) |

| | |
|-------------------------|---|
| fragDeltaMedPPM | フラグメントピークの誤差の中央値(ppm) |
| fragDeltaIqrPPM | フラグメントピークマッチの誤差、四分位範囲(ppm) |
| fragDeltaPolyFit | m/z(質量対電荷比)とフラグメントイオンの誤差(delta)を2次多項式(quadratic polynomial)でフィッティングし、その結果の決定係数 R ² (データポイント 100) |
| longest | 最も連続してフラグメントピークがマッチした回数。イオンシリーズ別に表示。 |
| fracIonsMatched | フラグメントマッチ率。イオンシリーズ別に表示、Neutral Loss など派生も加味。 fracIonsMatchedB1, fracIonsMatchedB1deriv, fracIonsMatchedB2, ..) |
| matchedIntensity | イオン強度、イオンシリーズ別に表示 |
| qmatch | マッチング条件にて該当ペプチドとペプチド質量がマッチしたペプチド数 |
| MIT | Mascot 同定基準値、identity threshold |
| MHT | Mascot 同定基準値、homology threshold |
| peptideLength | ペプチドの長さ |
| z1 | 電荷が 1 の場合、1 |
| z2 | 電荷が 2 または 3 の場合、1 |
| z4 | 電荷が 4,5 または 6 の場合、1 |
| z7 | 電荷が 7 以上の場合、1 |
| 12C | モノアイソトピックピークのみから構成されている時、1 |
| mc0 | ペプチド切断設定(Missed Cleavage)の適用数が 0 の時、1 |
| mc1 | ペプチド切断設定(Missed Cleavage)の適用数が 0 または 1 の時、1 |
| mc2 | ペプチド切断設定(Missed Cleavage)の適用数が 2 以上の時、1 |
| RMS | マッチしたフラグメントピークの RMS (m/z) |
| RMSppm | マッチしたフラグメントピークの RMS (ppm) |
| meanAbsFragDa | マッチしたフラグメントピークの誤差の平均値、絶対値 (Da) |
| meanAbsFragPPM | マッチしたフラグメントピークの誤差の平均値、絶対値 (ppm) |
| rawscore | 二項分布の考えを用いた確率スコア、主なイオンシリーズにおけるマッチング情報を利用。確率 p は $p = 2 * ITOL * n / 100$ (ITOL:MS2 誤差、n:100Da の中で選ばれたピークの数)で計算される |
| peptide | ペプチド配列。修飾情報は数字と合わせて表現。e.g) X.DAKAAM1AGRLM1IR.X |
| proteins | タンパク質の Accession (引数に加えるときは必ず最後の項目として指定すること) |

11-4. 保持時間予測と MS2 スペクトル予測

前述の、MASCOT Server 側が計算する features を”**core features**”と呼んでいます。それ以外にも features を利用することができます。別プログラムを使ってペプチド配列から予測保持時間と予測 MS2 スペクトルを計算し、実データとの一致度などの要素を Percolator の features として利用することで、同定ペプチド数を増やすことができます。ここでは、それらの予測プログラム並びに予測プログラムと Percolator との間をつなぐプログラムについて説明します。

11-4-1. MS²Rescore

MS²Rescore は、Gent 大学で開発されたプログラムです。ペプチド同定の再スコアリングを AI が支援するモジュール式のプラットフォームです。**MS²Rescore** は、保持時間を予測する **DeepLC** と スペクトルの **MS2** スペクトルを予測する **MS² PIP** を利用するための、共通の **Python** インターフェースを提供します。

MS²Rescore は開発者の許可を得て Mascot Server に含まれています。MS²Rescore 並びに関連するプログラム等には様々なオープンソースライセンスがあり、その詳細は Mascot Installation & Setup マニュアルに記載されています。MS²Rescore は Mascot Server Windows 版/Linux 版の両方のシステムでシームレスに動作します。

計算のために GPU を設置する必要はありません。またインターネット接続をしてどこかのサーバーに計算を実行させる必要はなく、すべてローカル環境で計算を行います。

[関連する論文]

Buur et al.: MS²Rescore 3.0 is a modular, flexible, and user-friendly platform to boost peptide identifications, as showcased with MS Amanda 3.0. J Prot Res (2024)

Declercq et al.: MS²Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates. Molecular & Cellular Proteomics (2021)

Silva et al.: Accurate peptide fragmentation predictions allow data driven approaches to replace and improve upon proteomics search engine scoring functions. Bioinformatics (2019)

11-4-2. DeepLC

DeepLC は Gent 大学で開発されたプログラムです。Deep Learning の手法を用いてペプチドの保持時間予測を行う事ができます。DeepLC はペプチド配列と修飾情報を受け取り、元素組成を計算して保持時間を予測します。そのため学習ステップで見られなかった可変修飾に対しても正確な予測ができるとい

うのが DeepLC の大きな強みです。

DeepLC を利用するためには、実測データの保持時間情報がピークリストファイルの各 query に含まれている必要があります。例えば入力フォーマットが MGF の場合、各クエリーデータに「**RTINSECONDS**」行がパラメーターとして含まれている必要があります。SCANTITLE の文字列の中に保持時間情報が埋め込まれているだけでは不十分です。入力が mzML の場合においても、保持時間は SCANTITLE ではととしてではなく、正しくタグ付けされたメタデータ(CV term)として含まれている必要があります。条件を満たすための最も簡単な検索方法は、**ピーク抽出処理を Mascot Distiller にて実行すること**です。

MASCOT であらかじめ準備している DeepLC のモデルは以下の通りです。出典情報やトレーニングデータセット数、記述内容の詳細などについては以下 URL をご覧ください。

https://www.matrixscience.com/help/ms2rescore_help.html#DEEPLCMODELS

| モデル名 | 一般名称 | カラム | 勾配 | ペプチドの特性 |
|--|--------------------|-------|--------|--|
| full_hc_PXD005573_mcp (recommended) | DIA HF | RP | 2h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term) |
| full_hc_ATLANTIS_SILICA _fixed_mods | ATLANTIS SILICA | HILIC | 90min | Tryptic, Carbamidomethyl (C) fixed |
| full_hc_LUNA_HILIC_fixed _mods | LUNA HILIC | HILIC | 90min | Tryptic, Carbamidomethyl (C) fixed |
| full_hc_LUNA_SILICA_fixe d_mods | LUNA SILICA | HILIC | 90min | Tryptic, Carbamidomethyl (C) fixed |
| full_hc_PXD008783_media n_calibrate | | | | Semi-tryptic, metabolic (14N/15N), open modification |
| full_hc_SCX_fixed_mods | SCX | SCX | 90min | Tryptic, Carbamidomethyl (C) fixed |
| full_hc_Xbridge_fixed_mo ds | Xbridge | HILIC | 90min | Tryptic, Carbamidomethyl (C) fixed |
| full_hc_arabidopsis_psms_ aligned | Arabidop sis | RP | 2h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term) |
| full_hc_dia_fixed_mods | SWATH library | RP | 135min | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_hela_hf_psms_alig ned | HeLa hf | RP | 1h | Tryptic, TMT, phosphopeptide enriched |

| モデル名 | 一般名称 | カラム | 勾配 | ペプチドの特性 |
|---|--------------------|-----|--------|---|
| full_hc_hela_lumos_1h_psms_aligned | HeLa Lumos 1h | RP | 1h | Tryptic, SILAC, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_hela_lumos_2h_psms_aligned | HeLa Lumos 2h | RP | 2h | Tryptic, SILAC, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_mod_fixed_mods | HeLa DeepRT | RP | 4h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term), Phospho (STY) |
| full_hc_pancreas_psms_aligned | Pancreas | RP | 110min | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_plasma_lumos_1h_psms_aligned | Plasma lumos 1h | RP | 1h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_plasma_lumos_2h_psms_aligned | Plasma lumos 2h | RP | 2h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M) |
| full_hc_prosit_ptm_2020 | ProteomeTools PTM | RP | 50min | Tryptic, 21 PTMs |
| full_hc_unmod_fixed_mods | Yeast DeepRT | RP | 4h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term) |
| full_hc_yeast_120min_psms_aligned | Yeast 2h | RP | 2h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term) |
| full_hc_yeast_60min_psms_aligned | Yeast 1h | RP | 1h | Tryptic, Carbamidomethyl (C) fixed, Oxidation (M), Acetyl (Protein N-term) |

■ 使用における制限

- ・ DeepLC が対応可能なペプチド配列の最大の長さは 60 残基です。
- ・ 溶出時間が大幅に短いデータへの適用は困難です。DeepLC はある程度キャリブレーションで対応させることができます。しかし溶出の全体の時間が短かったり、溶出時間が非常に短いグラジエントの保持時間を予測しようとした場合、キャリブレーションがうまくいかないことがあります。例えば 1 時間モデルを使用した場合 30~90 分のグラジエントではうまくいきますが、2 分や 5 分のデータに対しては予測が失敗することがあります。
- ・ ペプチドがトレーニングセットと大きく異なる場合、DeepLC の予測精度は低下します。例えばセレンを含むペプチドは、どのモデルもセレンを含まないため DeepLC で良い予測を行うことはできません。亜鉛が修飾として付与している場合などもトレーニングセットには含まれていないため、うまくいきません。

- DeepLC の現在のバージョンは、13C と 15N 以外の同位体をサポートしていません。例えば、2 つのペプチドが重水素 1 つだけ異なる場合、実際の RT が異なっているとしても、DeepLC は同じリテンションタイムを予測します。
- 現在のところ、クロスリンク検索、スペクトルライブラリ検索(データベースタイプが SL の検索)、Error Tolerant 検索には適用させることができません。

[関連する論文]

Bouwmeester et al.: DeepLC can predict retention times for peptides that carry as-yet unseen modifications. Nature Methods 18, 1363–1369 (2021).

11-4-3. MS²PIP

MS²PIP は、Gent 大学で開発されたプログラムです。複数のフラグメンテーションメソッド、装置、ラベリング技術に対応した高速で正確なペプチドフラグメンテーションスペクトル予測を行うことができます。MS²PIP は、**ペプチド配列、可変修飾、電荷状態を入力とし、ピーク強度を含む MS/MS フラグメンテーションスペクトルを予測** します。MS²PIP はトリプシンペプチドと非トリプシンペプチドのどちらにも対応しています。様々な修飾付与のスペクトル予測にも対応しています。

現在のところ、クロスリンク検索、スペクトルライブラリ検索(データベースタイプが SL の検索)、Error Tolerant 検索には適用させることができません。

MASCOT に最初から搭載されているモデルは以下の通りです。モデルのバージョン、出典情報、トレーニングデータセット数などを確認したい場合は以下 URL をご参照ください。

https://www.matrixscience.com/help/ms2rescore_help.html#MS2PIPMODELS

| モデル | フラグメンテーション | MS2 検出器 | ペプチド特性 |
|-------------------|------------|-----------------|------------------------------|
| CID | CID | Linear ion trap | Tryptic |
| CID-TMT | CID | Linear ion trap | Tryptic, TMT-labeled |
| CIDch2 | CID | Linear ion trap | Tryptic, 1+ and 2+ fragments |
| HCD2019 | HCD | Orbitrap | Tryptic |
| HCD2021 | HCD | Orbitrap | Tryptic and chymotryptic |
| HCDch2 | HCD | Orbitrap | Tryptic, 1+ and 2+ fragments |
| Immuno-HCD | HCD | Orbitrap | Immunopeptides |
| TMT | HCD | Orbitrap | Tryptic, TMT-labeled |

| モデル | フラグメンテーション | MS2 検出器 | ペプチド特性 |
|---------------------|------------|---|--|
| TTOF5600 | CID | Quadrupole time-of-flight | Tryptic |
| iTRAQ | HCD | Orbitrap | Tryptic digest, iTRAQ-labeled |
| iTRAQphospho | HCD | Orbitrap | Tryptic, iTRAQ-labeled, enriched for phosphorylation |
| timsTOF2023 | CID | Ion mobility quadrupole time-of-flight | Tryptic and elastase, immuno class 1 |
| timsTOF2024 | CID | Ion mobility quadrupole time-of-flight | Tryptic and elastase, immuno class 1 & 2 |

[関連する論文]

Declercq et al.: Updated MS2PIP web server supports cutting-edge proteomics applications. Nucleic Acids Research (2023)

Gabriels et al.: Updated MS2PIP web server delivers fast and accurate MS2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. Nucleic Acids Research (2019)

Degroeve et al.: MS2PIP prediction server: compute and visualize MS2 peak intensity predictions for CID and HCD fragmentation. Nucleic Acids Research, 43(W1), W326–W330. (2015)

Degroeve, S., & Martens, L.: MS2PIP: a tool for MS/MS peak intensity prediction. Bioinformatics 29(24), 3199–203. (2013)

11-5. refinement に関するレポート

機械学習を適用した際、以下の疑問に答えるレポートを出力することができます。

- 機械学習の適用により、どれくらい結果が改善したか
- データに対して正しいモデルを選択する事ができたか？
- 保持時間予測(**DeepLC**)またはスペクトル予測(**MS²PIP**)は正確であったか？
- **feature** のうち貢献度が高かったものはどれか？

これらレポートの見方については、「**7-2-10. machine learning quality report**」にまとめていますのでそちらをご覧ください。

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 title is 'MASCOT search status page'. The browser address bar shows 't3630/mascot/x-cgi/ms-status.exe'. The page content is as follows:

Version: 3.1.0 - MSKK (CLW7-X3C6-FBUB-VZXN-TNJZ) [Licence Info](#)

24 logical, 1 physical Intel processors (hyper-threading enabled, 12 core). CPUs: 0 1 2 3 4 5 6 18 19 20 21 22 23 available, using: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Active databases: 21. Inactive databases: 8. Max databases: 256.

[Search log](#) [monitor log](#) [error log](#) [Error message descriptions](#) [Do not auto refresh this page](#)

④

| | | | | | |
|-------------------|---|------------------------------------|----------------------------|--------------------------------------|---|
| Name | = | 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_20160906.msp |
| Status | = | In use | Statistics | Compression warnings | Recompress file |
| 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 | | | |

| | | | | | |
|-------------------|---|---------------------------|----------------------------|---------------------------------------|--|
| 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 | = | UP5640_H_sapiens | Family | = | C:/inetpub/mascot/sequence/UP5640_H_sapiens/current/UP5640_H_sapiens_*.fasta |
| Filename | = | UP5640_H_sapiens_20201007.fasta | Pathname | = | C:/inetpub/mascot/sequence/UP5640_H_sapiens/current/UP5640_H_sapiens_20201007.fasta |
| Status | = | In use | Statistics | 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 | | | |

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 | | | | | | | |
|------------------------------------|---------------------|----------------------------|----------------------|------------------------|----------------------|------------------------|-----------------------------------|
| Current jobs | | | | | | | |
| Job | PID | Start time | Dur. | Status | User | UserID | Title |
| 1243 | 14420 | Wed Sep 1 16:34:14 | 14 | Searching.... | | 0 | Copy of mgf 02 (C:\ProgramData\Ma |
| Completed jobs | | | | | | | |
| Job | PID | Start time | Dur. | Status | User | UserID | Title |

検索が完了すると、”Completed jobs”に移行します。

| Mascot database status - SwissProt | | | | | | | |
|------------------------------------|---------------------|----------------------------|----------------------|------------------------|----------------------|------------------------|-----------------------------------|
| Current jobs | | | | | | | |
| Job | PID | Start time | Dur. | Status | User | UserID | Title |
| | | | | | | | |
| Completed jobs | | | | | | | |
| Job | PID | Start time | Dur. | Status | User | UserID | Title |
| 1243 | 14420 | Wed Sep 1 16:34:14 | 810 | User read res | | 0 | Copy of mgf 02 (C:\ProgramData\Ma |

•[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

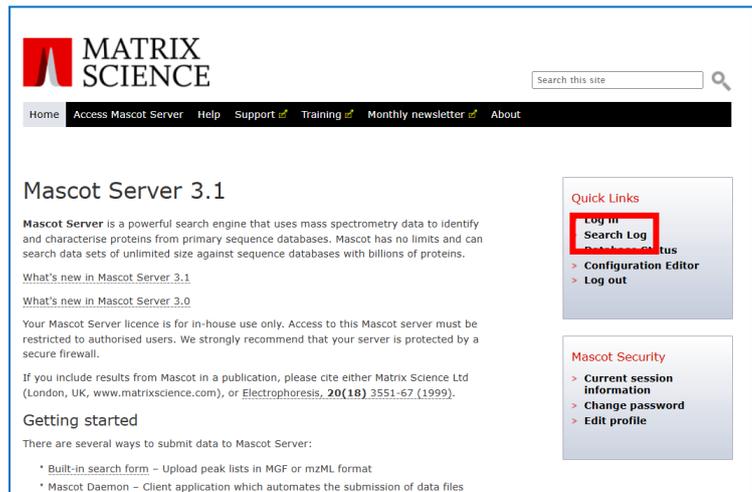
Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
Filename = SwissProt_2021_03.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2021_03.fasta
Status = In use Statistics Unidentified taxonomy Recompress file
State Time = Thu Sep 2 10:28:34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



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"/> |
| 1511 | 5436 | UP2195_D | Monitor Test DB 0 | | MS | ... |
| 1510 | 12840 | UP5640_H | | | be | ... |
| 1509 | 14116 | UP5640_H | | | be | ... |

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

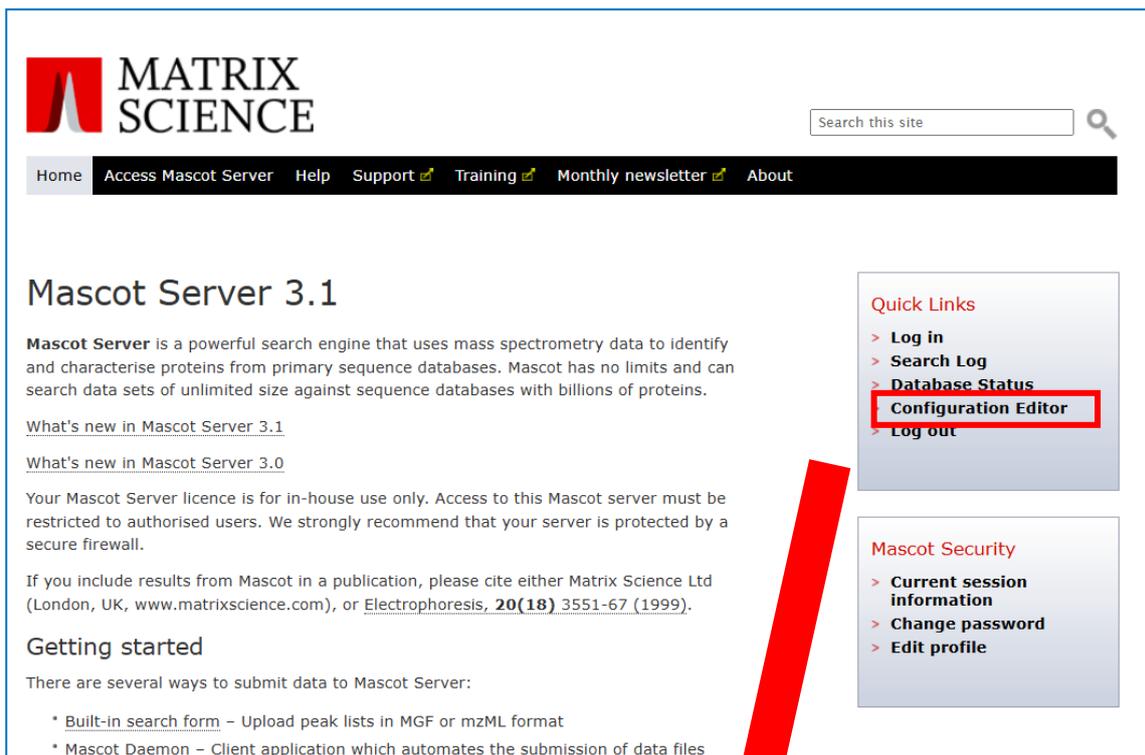
各項目の一番上にある入力欄はフィルターとして利用できます。例えば SwissProt で検索した結果のみを表示させたい場合、“dbase”項目に“SwissProt”と入力して“sort/filter”ボタンを押すことでフィルターリングが実現します。

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

Search this site

Home Access Mascot Server Help Support Training Monthly newsletter About

Mascot Server 3.1

Mascot Server is a powerful search engine that uses mass spectrometry data to identify and characterise proteins from primary sequence databases. Mascot has no limits and can search data sets of unlimited size against sequence databases with billions of proteins.

[What's new in Mascot Server 3.1](#)

[What's new in Mascot Server 3.0](#)

Your Mascot Server licence is for in-house use only. Access to this Mascot server must be restricted to authorised users. We strongly recommend that your server is protected by a secure firewall.

If you include results from Mascot in a publication, please cite either Matrix Science Ltd (London, UK, www.matrixscience.com), or *Electrophoresis*, **20(18)** 3551-67 (1999).

Getting started

There are several ways to submit data to Mascot Server:

- * [Built-in search form](#) - Upload peak lists in MGF or mzML format
- * [Mascot Daemon](#) - Client application which automates the submission of data files

Quick Links

- > [Log in](#)
- > [Search Log](#)
- > [Database Status](#)
- > **Configuration Editor**
- > [Log out](#)

Mascot Security

- > [Current session information](#)
- > [Change password](#)
- > [Edit profile](#)

Mascot Configuration

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

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 Edit |
| J | Xle | Leu or Ile | 113.084064 | 113.1576 | H(11) C(6) N O Edit |
| O | Pyr | Pyrrrolisine | 237.147727 | 237.2982 | H(19) C(12) N(3) O(2) Edit |
| 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 | | | | | | | | |
|--|--------------|----------|--------------------|--------|------------|---------|----------------------|-----------------------|
| Title | Monoisotopic | Average | Composition | Source | Visibility | Err Tol | | |
| <input type="checkbox"/> 15N-oxobutanoic | -18.023584 | -18.0239 | H(-3) 15N(-1) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 2-dimethylsuccinyl | 144.042259 | 144.1253 | H(8) C(6) O(4) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 2-monomethylsuccinyl | 130.026609 | 130.0987 | H(6) C(5) O(4) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 2-nitrobenzyl | 135.032028 | 135.1201 | H(5) C(7) N O(2) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 2-succinyl | 116.010959 | 116.0722 | H(4) C(4) O(4) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 2HPG | 282.052824 | 282.2476 | H(10) C(16) O(5) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 3-deoxyglucosone | 144.042259 | 144.1253 | H(8) C(6) O(4) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 3-hydroxybenzyl-phosphate | 186.008196 | 186.1018 | H(7) C(7) O(4) P | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 3-phosphoglyceryl | 167.982375 | 168.0420 | H(5) C(3) O(6) P | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 3sulfo | 183.983029 | 184.1693 | H(4) C(7) O(4) S | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 4-ONE | 154.099380 | 154.2063 | H(14) C(9) O(2) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 4-ONE+Delta:H(-2)O(-1) | 136.088815 | 136.1910 | H(12) C(9) O | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> 4AcAllylGal | 372.142033 | 372.3671 | H(24) C(17) O(9) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> a-type-ion | -46.005479 | -46.0254 | H(-2) C(-1) O(-2) | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> AccQTag | 170.048013 | 170.1674 | H(6) C(10) N(2) O | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> Acetyl | 42.010565 | 42.0367 | H(2) C(2) O | Unimod | mixed | yes | Copy | Print |
| <input type="checkbox"/> Acetyl:13C(2) | 44.017274 | 44.0220 | H(2) 13C(2) O | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> Acetyl:2H(3) | 45.029395 | 45.0552 | H(-1) 2H(3) C(2) O | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> Acetyldeoxyhypusine | 97.089149 | 97.1582 | H(11) C(6) N | Unimod | long | yes | Copy | Print |
| <input type="checkbox"/> Acetylhypusine | 113.084064 | 113.1576 | H(11) C(6) N O | Unimod | long | yes | Copy | Print |

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

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 | | | | | | | Specificity | | | Ignore Masses | | | Misc | | | References | | |
|-------------|------|---|----------|----------|------|--------|--------------|--|--|---------------|--|--|------|--|--|------------|--|--|
| 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: 97.976896 Average: 97.9952 | | | | | | | | | | | | | | |
| 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。Error Tolerant 検索の際、ここで分類したグループ単位で再検索時に考慮する修飾を指定することができる

- 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。

The screenshot shows the 'Ignore Masses' tab selected in a menu bar. Below the menu bar, there is a 'New Ignore Mass' button. At the bottom of the tab, there are buttons for 'Save changes', 'Cancel', and 'Update', along with links for 'Revert to Unimod' and 'Show differences'.

「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。

The screenshot shows the 'Misc' tab selected in a menu bar. The main content area displays metadata for a registration entry: 'Creation Date' (2002-08-19 19:17:11), 'Last modified' (2018-08-13 13:42:59), and 'User' (unimod). Below this is a 'Notes' section with a text area containing the text: 'Neutral loss of phosphate is typically observed from Y/H/D/E/K/C, rather than the preferential loss of phosphoric acid from S/T.' There is also a 'New Alternative Name' button. At the bottom, there are buttons for 'Save changes', 'Cancel', and 'Update', along with links for 'Revert to Unimod' and 'Show differences'.

「Reference」タブは、この link 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。

The screenshot shows the 'References' tab selected in a menu bar. The main content area displays a list of three references. Each reference has a 'Delete' button. Reference 1 has 'Source' set to 'RESID' and 'Text' set to 'AA0036'. Reference 2 has 'Source' set to 'Misc. URL' and 'Text' set to 'IonSource'. Reference 3 has 'Source' set to 'RESID' and 'Text' set to 'AA0037'. Each reference also has a 'URL' field and a 'Goto' link.

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 |
|----------------------------------|--------------|----------|----------------------|--------|------------|--|
| Xlink:BS2G | 96.021129 | 96.0841 | H(4) C(5) O(2) | Unimod | long | yes Copy Print |
| Xlink:BuUrBu | 196.084792 | 196.2032 | H(12) C(9) N(2) O(3) | Unimod | long | yes Copy Print |
| Xlink:Disulfide | -2.015650 | -2.0159 | H(-2) | Unimod | long | yes Copy Print |
| Xlink:Dityrosine | -2.015650 | -2.0159 | H(-2) | local | long | no Copy Print |
| Xlink:DMP | 122.084398 | 122.1677 | H(10) C(7) N(2) | Unimod | long | yes Copy Print |
| Xlink:DSS | 138.068080 | 138.1638 | H(10) C(8) O(2) | Unimod | long | yes Copy Print |
| Xlink:DSSO | 158.003765 | 158.1750 | H(6) C(6) O(3) S | Unimod | long | yes Copy Print |
| Xlink:DST | 113.995309 | 114.0563 | H(2) C(4) O(4) | Unimod | long | yes Copy Print |
| Xlink:DTBP | 172.012890 | 172.2711 | H(8) C(6) N(2) S(2) | Unimod | long | yes Copy Print |
| Xlink:DTSSP | 173.980921 | 174.2406 | H(6) C(6) O(2) S(2) | Unimod | long | yes Copy Print |
| Xlink:EDC | -18.010565 | -18.0153 | H(-2) O(-1) | Unimod | long | yes Copy Print |
| Xlink:EGS | 226.047738 | 226.1828 | H(10) C(10) O(6) | Unimod | long | yes Copy Print |
| Xlink:SDA | 82.041865 | 82.1005 | H(6) C(5) O | Unimod | long | yes Copy Print |
| Xlink:SMCC | 219.089543 | 219.2365 | H(13) C(12) N O(3) | Unimod | long | yes Copy Print |
| Xlink:test1 | 185.058912 | 185.1821 | C(10) H(7) N(3) O | local | long | no Copy Print |
| Xlink:test2 | 159.068414 | 159.1846 | C(10) H(9) N O | local | long | no Copy Print |

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

[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す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 | | | | | | | | | | | |
| Neutral loss | Code | I | Pairs With | | | Delete | | | | | |
| Composition | | | Symbols | 13C | 1 | Add | | | | | |
| Description | Intact cross-link | | | | | | | | | | |
| Neutral loss | Code | A | Pairs With | | | 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 | | | 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 | | | 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 link 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 | Edit | Delete |
| Trypsin/P | C-Term | KR | | no | no | Edit | Delete |
| Arg-C | C-Term | R | P | no | no | Edit | Delete |
| Asp-N | N-Term | BD | | no | no | Edit | Delete |
| Asp-N_ambic | N-Term | DE | | no | no | Edit | Delete |
| Chymotrypsin | C-Term | FLWY | P | no | no | Edit | Delete |
| CNBr | C-Term | M | | no | no | Edit | Delete |
| CNBr+Trypsin | C-Term | M | | no | no | Edit | Delete |
| | C-Term | KR | p | no | no | Edit | Delete |
| Formic_acid | N-Term | D | | no | no | Edit | Delete |
| | C-Term | D | | | | | |
| Lys-C | C-Term | K | P | no | no | Edit | Delete |
| Lys-C/P | C-Term | K | | no | no | Edit | Delete |
| LysC+AspN | N-Term | BD | | no | no | Edit | Delete |
| | C-Term | K | p | no | no | Edit | Delete |
| Lys-N | N-Term | K | | no | no | Edit | Delete |
| PepsinA | C-Term | FL | | no | no | Edit | Delete |
| semiTrypsin | C-Term | KR | P | no | yes | Edit | Delete |
| TrypChymo | C-Term | FKLRWY | P | no | no | Edit | Delete |
| | N-Term | J | | | | | |
| TrypsinMSIPI | C-Term | KR | P | no | no | Edit | Delete |
| | C-Term | J | | | | | |
| TrypsinMSIPI/P | N-Term | J | | no | no | Edit | Delete |
| | C-Term | JKR | | | | | |
| V8-DE | C-Term | BDEZ | P | no | no | Edit | Delete |
| V8-E | C-Term | EZ | P | no | no | Edit | Delete |
| NoCleave | C-Term | J | ABCDEFGHIJKLMNQRSTUUVWXYZ | no | no | Edit | Delete |
| 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
PLLEVIKIIIASVASVIFVGFACVTLGSAALVSTPVFIIFSPVLVPATIATVVLATGFTAG
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 | IIIASVASVIFVGFACVTLGSAALVSTPVFIIFSPVLVPATIATVVL 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 Components 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。

| 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 | EAD |
|---|-----------|--------------|---------------|----------|----------|-----------|---------------|-------------|----------|----------|----------------|---------------|-----------|---------|--------|--------|
| 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 | X | X | X |
| 2+ (precursor>3+) | | | | | | | | | | | | | | | | |
| Immonium | | | X | | | | X | X | | | X | X | | | | |
| a | X | | X | | | | X | X | | | X | X | X | | X | X |
| a* | X | | X | | | | X | X | | | X | X | | | X | X |
| a0 | | | X | | | | X | X | | | 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 |
| 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 | X |
| y0 | | X | | X | 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 | X |
| d | | | | | | | X | | | | | | | | | |
| v | | | | | | | X | | | | | | | | | |
| w | | | | | | | X | | | | | | | X | X | X |
| z+2 | | | | | | | | | X | X | | | X | X | X | X |
| Min internal mass | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | 700 | | |
| Max internal mass | | | | | | | | | | | | | | | | |
| Refine results with machine learning | | | | | | | | | | | | | | | | |
| DeepLC model for retention times | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) |
| MS2PIP model for spectral similarity | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) | (none) |
| | Edit | 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 | Edit |
| New Instrument | Main menu | | | | | | | | | | | | | | | |
| y or y++ must be highest scoring series | | | | | | | | | | | | | | | | |

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(上図)。

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(次ページ上図)。

設定には refinement に関する設定を組み込ませることができます。Proteome Discoverer との連携時など、Instrument 設定項目はあるのに refinement に関する設定項目がない場合などにこの設定のご利用をご検討ください。

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 | | | |
| ITRAQ 4plex | reporter | Copy | Delete | Print |
| ITRAQ 4plex (protein) | reporter | Copy | Delete | Print |
| ITRAQ 8plex | reporter | Copy | Delete | Print |
| TMT 6plex | reporter | Copy | Delete | Print |
| TMT 2plex | reporter | Copy | Delete | Print |
| TMT 10plex | reporter | Copy | Delete | Print |
| TMTpro 16plex | reporter | Copy | Delete | Print |
| DiLeu 4plex | reporter | Copy | Delete | Print |
| 18O multiplex | multiplex | Copy | Delete | Print |
| SILAC K+6 R+6 multiplex | multiplex | Copy | Delete | Print |
| IPTL (Succinyl and IMID) multiplex | multiplex | Copy | Delete | Print |
| ICPL duplex pre-digest [MD] | precursor | Copy | Delete | Print |
| ICPL duplex post-digest [MD] | precursor | Copy | Delete | Print |
| ICPL triplex pre-digest [MD] | precursor | Copy | Delete | Print |
| ICPL quadruplex pre-digest [MD] | precursor | Copy | Delete | Print |
| 18O corrected [MD] | precursor | Copy | Delete | Print |
| 15N Metabolic [MD] | precursor | Copy | Delete | Print |
| 15N + 13C Metabolic [MD] | precursor | Copy | Delete | Print |
| SILAC K+6 R+10 [MD] | precursor | Copy | Delete | Print |
| SILAC K+6 R+10 Arg-Pro [MD] | precursor | Copy | Delete | Print |
| SILAC K+6 R+6 [MD] | precursor | Copy | Delete | Print |
| SILAC R+6 R+10 [MD] | precursor | Copy | Delete | Print |
| SILAC K+8 R+10 [MD] | precursor | Copy | Delete | Print |
| SILAC K+4 K+8 R+6 R+10 [MD] | precursor | Copy | Delete | Print |
| ICAT ABI Cleavable [MD] | precursor | Copy | Delete | Print |
| ICAT D8 [MD] | precursor | Copy | Delete | Print |
| Dimethylation [MD] | precursor | Copy | Delete | Print |
| NBS Shimadzu [MD] | precursor | Copy | Delete | Print |
| Acetylation [MD] | precursor | Copy | Delete | Print |
| Label-free [MD] | replicate | Copy | Delete | Print |
| Average [MD] | average | Copy | Delete | Print |

[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 style="width: 50px;" 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

| 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。

| 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。

「**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

■ 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

「**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 |
|---------------------------------------|-------------------------------------|-----------|--------------|-------------|---------|-----|
| Integration | | | | | | |
| Property | Value | | | | | |
| 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

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 | Action |
|------------------------------|--------------------------|---------------|
| 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 |

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の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

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 |
| method of detecting outliers | none=No outlier removal |

「**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 |

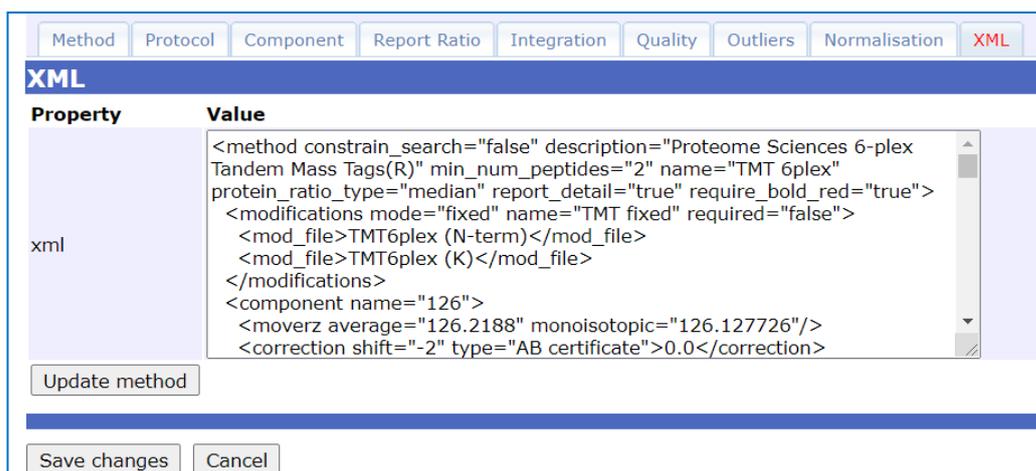
Normalisation method :

- **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。

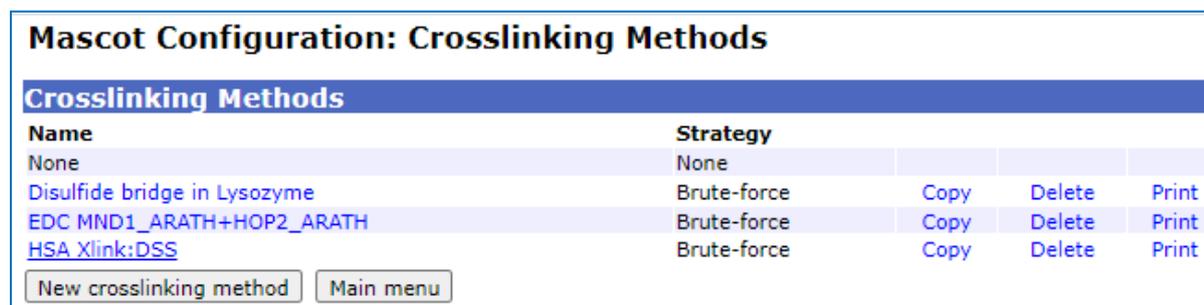
「XML」タブは設定ファイル内で設定内容が記述されている、XML フォーマットを実際に表示して確認するとともに、場合によっては直接編集するための設定欄です。記述変更後「Update method」ボタンを押すことで記述内容が適用されます



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。



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)。

| Property | Value | Action |
|------------|---|--------|
| Strategy | Brute-force | |
| InterLink | <input checked="" type="checkbox"/> | |
| IntraLink | <input checked="" type="checkbox"/> | |
| LoopLink | <input checked="" type="checkbox"/> | |
| Linkers | Linker : Xlink:DSS (K) Monolink : I, A, W, T DoesNotPairWith : Xlink:DSS (K), Xlink:DSS (Protein N-term) Delete | |
| | Linker : Xlink:DSS (Protein N-term) Monolink : I, A, W, T DoesNotPairWith : Xlink:DSS (K), Xlink:DSS (Protein N-term) Delete | |
| Accessions | Database name : Accession : ALBU_HUMAN Delete Add parameter | |
| Filters | Name : MinLen Value : 2 Delete Add parameter | |
| Settings | Add parameter | |

- 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><mxm:method description="" name="HSA Xlink:DSS" strategy="Brute-force"> <mxm:linkers> <mxm:linker ModFileName="Xlink:DSS (K)"> <mxm:monolink>A</mxm:monolink> <mxm:monolink>W</mxm:monolink> </mxm:linker> <mxm:linker ModFileName="Xlink:DSS (Protein N-term)"> <mxm:monolink>A</mxm:monolink> <mxm:monolink>W</mxm:monolink> </mxm:linker> </mxm:linkers> </mxm:method></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 checked="" type="radio"/> yes <input type="radio"/> no <input type="radio"/> clear |
| SaveEveryLastQueryAsc | <input checked="" type="radio"/> yes <input type="radio"/> no <input type="radio"/> clear |
| LastQueryAscFile | ../logs/lastquery.asc |
| InterFileBasePath | C:/inetpub/mascot/data |

=====中略=====

| | |
|--|-----------------------------|
| MaxVarMods | 15 |
| SearchSubmitAcceptedFileTypes | Mascot generic,mzML (.mzML) |
| SearchSubmitDefaultPeptideCharge | 2+, 3+ and 4+ |
| SearchSubmitOutputFormat | msr |
| AlwaysCreateDat28ResultsFile | 1 |
| ClientResultFileMimeRefining | 1 |
| <input type="button" value="Add New Option"/> <input type="button" value="APPLY"/> | |

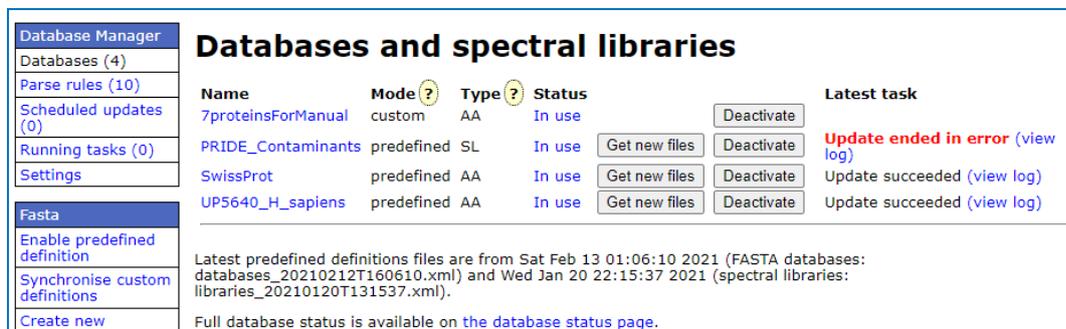
Configuration Options で初期表示されている/いない に関係なく、MASCOT Server で使用可能なオプションがあります。詳細は別紙の日本語資料資料

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).

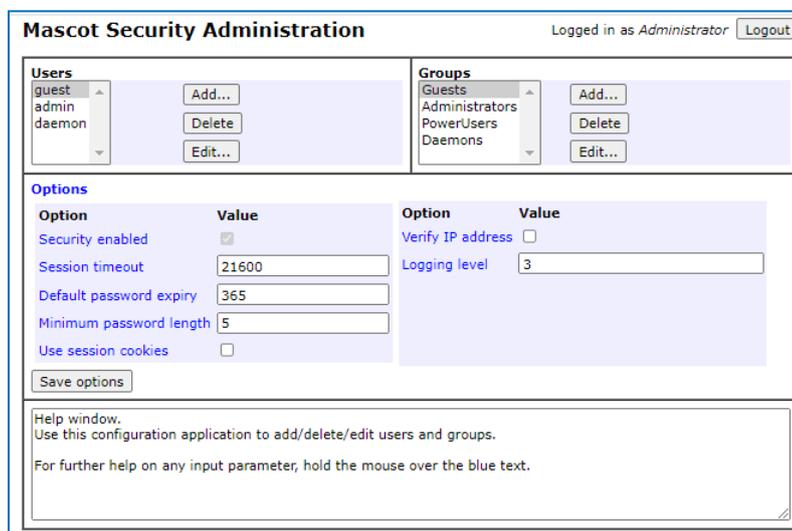
Full database status is available on [the database status page](#).

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

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

電 話 : 03-5807-7895

