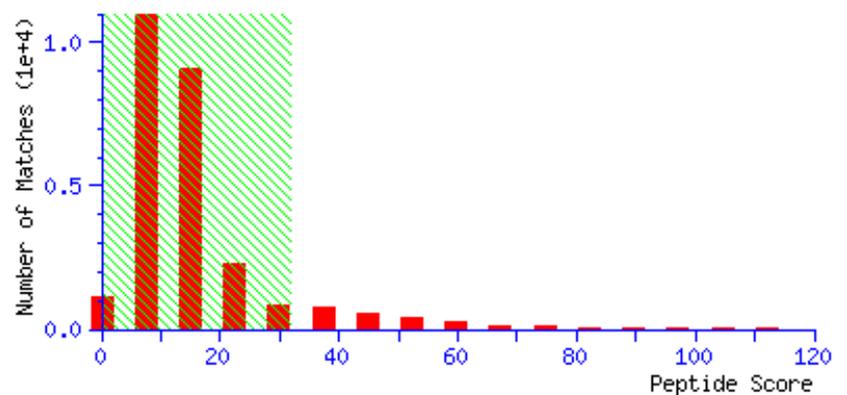


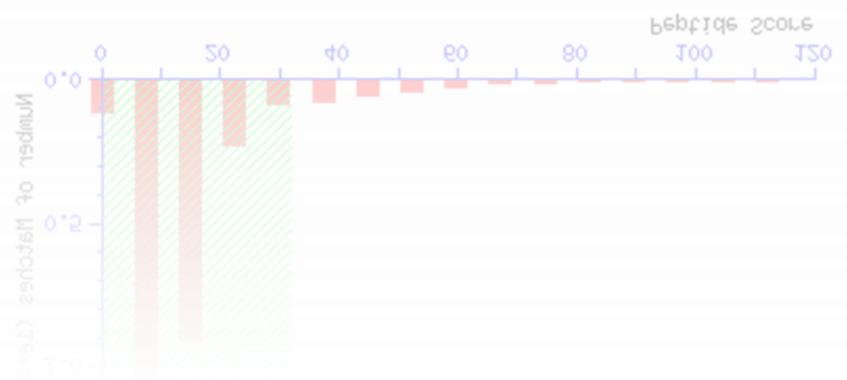
Mascot Server 2.4

取扱説明書



Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 32 indicate **identity** or **extensive homology** ($p < 0.05$).

($p < 0.02$).
ions scores > 35 indicate **identity** or **extensive homology**
probability that the observed match is a random event. Individual
peptide score distribution. Ions score is $-10\log(P)$, where P is the



◀ 目次 ▶

- ▶ 1 Mascot検索クイックスタート 3
 - 1-1 PMF検索 : Peptide Mass Fingerprint | 1-2 MIS検索 : MS/MS Ions Search

- ▶ 2 システム構成 8
 - 2-1 ソフトウェア構成 | 2-2 ネットワーク構成 | 2-3 クライアントソフトウェア
 - 2-4 フォルダ構造 | 2-5 配列データベース | 2-6 マニュアル | 2-7 Mascotの最新情報

- ▶ 3 Mascot検索の種類と検索条件 12
 - 3-1 Mascot検索の種類 | 3-2 検索条件 | 3-3 PMF検索 : Peptide Mass Fingerprint
 - 3-4 MIS検索 : MS/MS Ions Search | 3-5 SQ検索 : Sequence Query

- ▶ 4 スコアリング 18
 - 4-1 スコアと期待値 | 4-2 Identity閾値とHomology閾値 | 4-3 FDR : False Discovery Rate

- ▶ 5 検索結果 22
 - 5-1 検索結果ページの種類 | 5-2 フォーマットコントロールパネル | 5-3 タンパク質の推定
 - 5-4 Concise Protein Summary (PMF) | 5-5 Protein Summary (PMF) | 5-6 Protein View (PMF/MIS/SQ)
 - 5-7 Protein Family Summary (MIS/SQ) | 5-8 Peptide Summary (MIS/SQ)
 - 5-9 Select Summary (MIS/SQ) | 5-10 Peptide View (MIS/SQ) | 5-11 検索結果の出力 (PMF/MIS/SQ)
 - 5-12 Report Builder (MIS/SQ)

- ▶ 6 質量分析計システムとの連携 38
 - 6-1 アジレント・テクノロジー | 6-2 エービー・サイエックス | 6-3 島津製作所
 - 6-4 サーモフィッシャーサイエンティフィック | 6-5 日本ウォーターズ | 6-6 日本電子
 - 6-7 日立ハイテクノロジーズ | 6-8 ブルカー・ダルトニクス

- ▶ 7 Mascot検索の自動化 : Mascot Daemon 41
 - 7-1 インストール | 7-2 マニュアル

- ▶ 8 Mascot Serverの管理 42
 - 8-1 Welcomeトップページ | 8-2 検索ログ : Search Log
 - 8-3 配列データベース管理 : Database Status | 8-4 設定値変更 : Configuration Editor
 - 8-5 セキュリティ | 8-6 検索条件のデフォルト設定 | 8-7 エラーログ | 8-8 バックアップ

▶ 1 Mascot検索クイックスタート

1-1 PMF検索 : Peptide Mass Fingerprint

1-1-1 Mascot Serverへの接続

Webブラウザを起動し、次のアドレスにアクセスしてください。ホスト名はMascot ServerがインストールされているPCのホスト名です。

① <http://ホスト名/mascot/>

Mascot Serverに接続され、図1の「Welcome」トップページが表示されます。② [Mascot] リンクをクリックしてください。図2に示す、検索方法を選択するための「Mascot Search」ページが表示されます。

1-1-2 検索方法の選択

MSの質量スペクトルデータに対するMascot検索はPMF (Peptide Mass Fingerprint) 検索と呼んでおり、MSの質量スペクトルにマッチするタンパク質を検索・同定します。「Mascot Search」ページ(図2)の③ [Peptide Mass Fingerprint] リンクをクリックしてください。図3に示す、PMF検索条件を設定するための「MASCOT Peptide Mass Fingerprint」ページが表示されます。

1-1-3 検索条件の設定と検索の実行

「MASCOT Peptide Mass Fingerprint」ページ(図3)において、

- 検索したい配列データベース : ④ [Database(s)]
- 実験条件
 - ⑤ [Enzyme] : 消化酵素
 - ⑥ [Allow up to] : 未消化サイト数
 - ⑦ [Fixed modifications] : 化学修飾
 - ⑧ [Variable modifications] : 翻訳後修飾等
- 質量分析計の質量精度 : ⑨ [Peptide tol. ±]

を検索条件として設定してください。

[Fixed modifications]、[Variable modifications]の指定は、右側のリストボックスにあるエントリをクリックし、[<] ボタンで左側ウインドウに移してください。左側ウインドウにあるエントリを選択し、[>] ボタンを押すと右側のリストボックスに戻ります。

[Your name]、[Email]、[Search title] 入力欄への文字列入力は任意ですが、ここで入力した文字列は過去のMascot検索結果を呼び出す際に利用することができますので、入力することをお勧めします。

⑩ [Data file] の[参照] ボタンから質量データファイルを指定した後、⑪ [Start Search ...] ボタンを押してください。図3の例では検索条件として、

- ④ [Database(s)] = "SwissProt"
- ⑤ [Enzyme] = "Trypsin"

図1 「Welcome」 トップページ

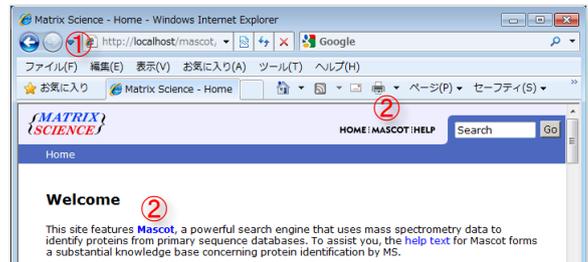
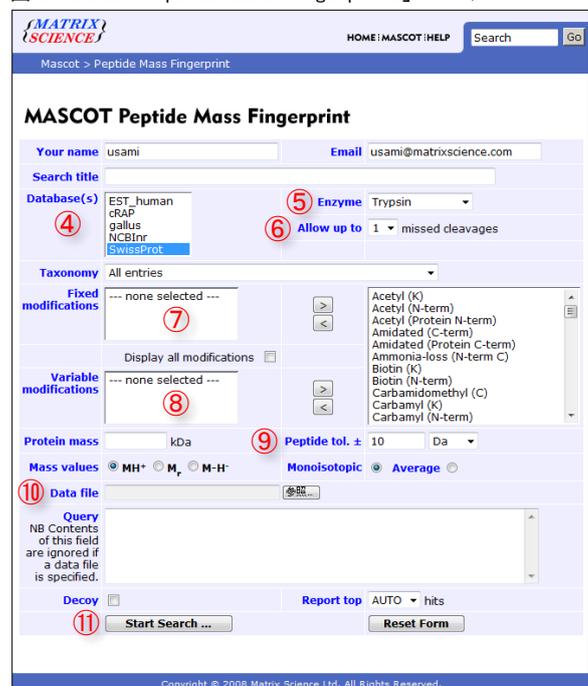


図2 「Mascot Search」 ページ



図3 「MASCOT Peptide Mass Fingerprint」 ページ



- ⑥ [Allow up to] "1" missed cleavages
- ⑦ [Fixed modifications] = "Carbamidomethyl (C)"
- ⑧ [Variable modifications] = "Oxidation (M)"
- ⑨ [Peptide tol. ±] = "0.2" "Da"
- ⑩ [Data file] = Mascot Server PC上の "C:\inetpub\mascot\mskk\sampladata\pmfSample.mgf" ファイル

を指定しています。

1-1-4 検索結果の表示

⑪ [Start Search ...] ボタンを押すと、Mascot検索の進捗状況を示すページが表示された後、検索が終了すると検索結果をまとめた「Mascot Search Results」ページが表示されます(図4)。「Mascot Search Results」ページは次のような項目で構成されています。

(1) ヘッダ情報

- ユーザ名
- 電子メールアドレス
- 検索タイトル
- MSデータファイル名
- 配列データベース名とタンパク質エントリ数
- 検索日時
- 最も高いスコアを持つタンパク質名

(2) Mascot Score Histogram

- 閾値スコア(図4の例では「70」)
- スコア分布図(検索でヒットしたタンパク質のスコア分布)

(3) Concise Protein Summary Report

PMF検索では、表示内容が異なる2種類の検索結果ページ(「Concise Protein Summary Report」と「Protein Summary Report」)を用意していますが、図4の例では「Concise Protein Summary Report」ページを表示しています。

(4) タンパク質情報

ヒットしたタンパク質に関する次の情報を表示しています。

- ヒット番号(図4では「1」)
- アクセション番号(図4では「OPSD_HUMAN」など)
- Mass(タンパク質の質量: 図4では「39437」)
- Score(プロテインスコア: 図4では「102」)
- Expect(期待値: 図4では「3.3e-005」)
- Matches(マッチしたペプチド数: 図4では「11」)
- タンパク質情報(図4では「Rhodopsin OS=Homo ...」など)
- サブセットなタンパク質(図4では「OPSD_MACFA」など)

(5) Search Parameters

この検索に用いた検索条件をページの下方に表示しています。

アクセション番号(図4では「OPSD_HUMAN」など)は「Protein View」ページへのリンクになっており、図5に示すように、そのタンパク質に関するより詳細な情報を見ることができます。

このPMF検索結果(図4)では、閾値スコア(70)より大きなプロテインスコアを持つ、OPSD_HUMAN(Rhodopsin: 人, プロテインスコア102)が同定されています。

検索結果の詳しい見方については「5 検索結果」をご覧ください。

図4 「Mascot Search Results」ページ

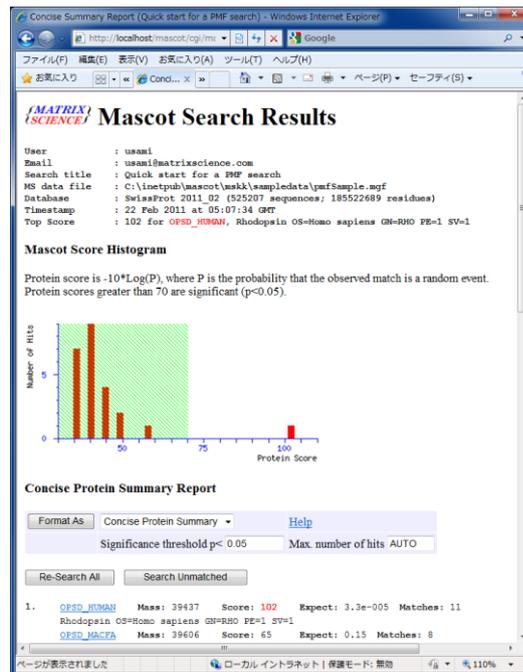
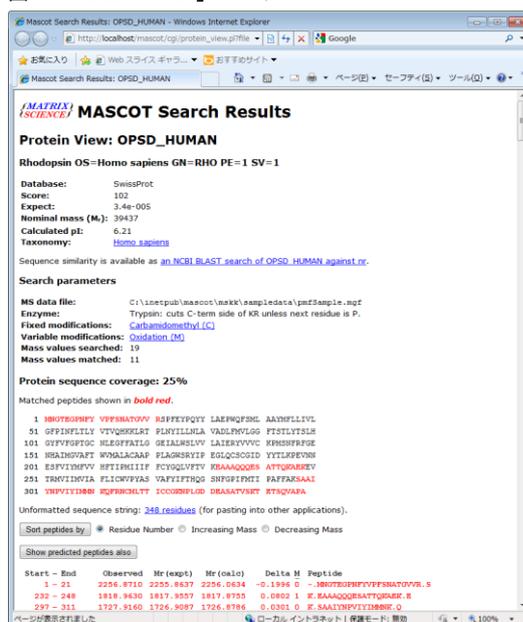


図5 「Protein View」ページ



1-2 MIS検索 : MS/MS Ions Search

1-2-1 Mascot Serverへの接続

Webブラウザを起動し、次のアドレスにアクセスしてください。**ホスト名**はMascot ServerがインストールされているPCのホスト名です。

- ① <http://ホスト名/mascot/>

Mascot Serverに接続され、図6の「Welcome」トップページが表示されます。② [Mascot]リンクをクリックしてください。図7に示す、検索モードを選択するための「Mascot Search」ページが表示されます。

1-2-2 検索モードの選択

MS/MSの質量スペクトルデータに対するMascot検索はMIS(MS/MS Ion Search) 検索と呼んでおり、MS/MSの質量スペクトルにマッチするペプチドを検索・同定するとともに、節約の原理(Principle of Parsimony)に基づいて、同定されたペプチドが帰属するタンパク質を整理し、実験サンプルに含まれるタンパク質を推定します。「Mascot Search」ページ(図7)の③ [MS/MS Ion Search]リンクをクリックしてください。MIS検索条件を設定するための「MASCOT MS/MS Ions Search」ページが表示されます(図8)。

1-2-3 検索条件の設定と検索の実行

「MASCOT MS/MS Ions Search」ページ(図8)において、

- ・ 検索したい配列データベース : ④ [Database(s)]
- ・ 実験条件
 - ⑤ [Enzyme] : 消化酵素
 - ⑥ [Allow up to] : 未消化サイト数
 - ⑦ [Fixed modifications] : 化学修飾
 - ⑧ [Variable modifications] : 翻訳後修飾等
- ・ 質量分析計の種類と質量精度
 - ⑨ [Peptide tol. ±]
 - : プリカーサイオンの質量誤差
 - ⑩ [MS/MS tol. ±]
 - : プロダクトイオンの質量誤差
 - ⑫ [Instrument] : 質量分析計の種類

を検索条件として設定してください。

[Fixed modifications]、[Variable modifications]の指定は、右側のリストボックスにあるエントリをクリックし、[<] ボタンで左側ウィンドウに移してください。左側ウィンドウのエントリを選択し、[>] ボタンを押すと右側リストボックスに戻ります。

[Your name]、[Email]、[Search title]入力欄への文字列入力任意ですが、ここで入力した文字列は過去のMascot検索結果を呼び出す際に利用することができますので、入力することをお勧めします。

⑬ [Data file]の[参照]ボタンから質量データファイルを指定した後、⑬ [Start Search ...] ボタンを押してください。図8の例では検索条件として、

- ④ [Database(s)] = "SwissProt"
- ⑤ [Enzyme] = "Lys-C/P"
- ⑥ [Allow up to] "1" missed cleavages

図6 「Welcome」 トップページ

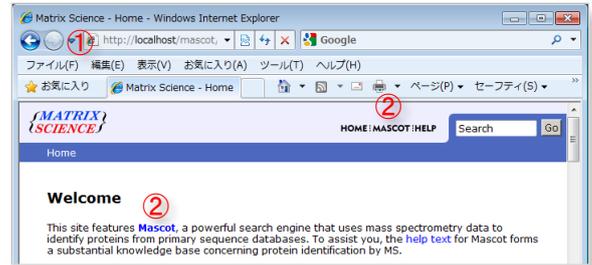


図7 「Mascot Search」 ページ

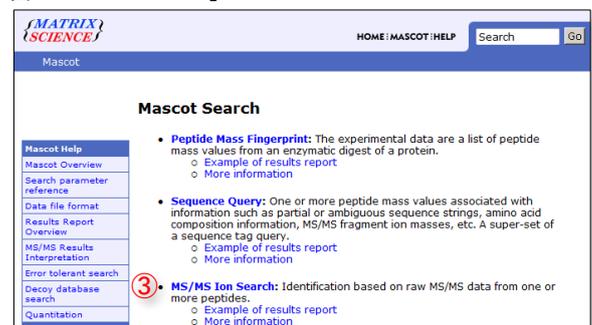


図8 「MASCOT MS/MS Ions Search」 ページ



- ⑦ [Fixed modifications] = “Carbamidomethyl (C)”
- ⑧ [Variable modifications] = “Phospho (ST)”
- ⑨ [Peptide tol. ±] = “50” “ppm”
- ⑩ [Data file] = Mascot Server PC上の “C:\inetpub\wwwroot\mascot\mskk\samp\data\error\TolSample.mgf” ファイル

を指定しています。

1-2-4 検索結果の表示

⑬ [Start Search...] ボタンを押すと、Mascot検索の進捗状況を示すページが表示された後、検索が終了すると検索結果をまとめた「Mascot Search Results」ページが表示されます(図9)。「Mascot Search Results」ページは次のような項目で構成されています。

(1) ヘッダ情報

- ユーザ名
- 電子メールアドレス
- 検索タイトル
- MSデータファイル名
- 配列データベース名とタンパク質エントリ数
- 検索日時
- 閾値スコアを超えたタンパク質名

(2) Mascot Score Histogram

- 閾値スコア(図9の例では37)
- スコア分布図(ヒットしたタンパク質のスコア分布)

(3) 検索結果ページ名

MIS検索では、表示内容が異なる3種類の検索結果ページ(「Protein Family Summary Report」、 「Peptide Summary Report」、 「Select Summary Report」)を用意していますが、図9の例では「Peptide Summary Report」を表示しています。

(4) タンパク質情報

ヒットしたタンパク質に関する次の情報を表示しています。

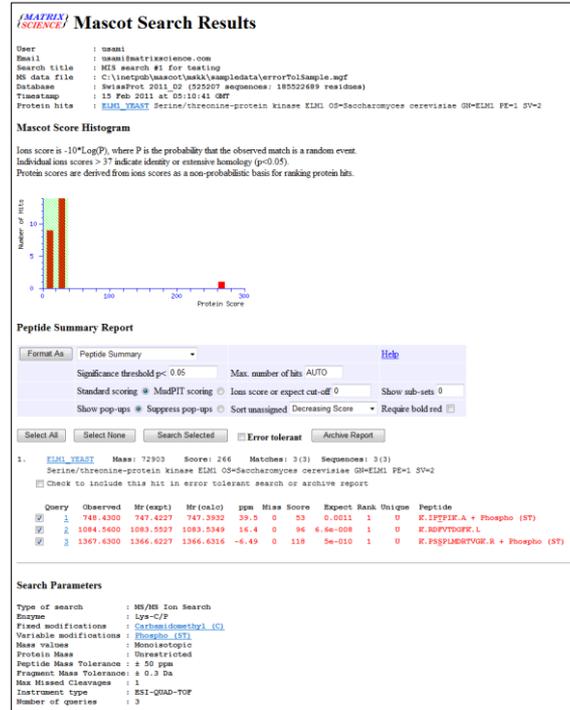
- ヒット番号(図9では「1」)
- アクセッション番号(図9では「ELM1_YEAST」)
- Mass(タンパク質の質量: 図9では「72903」)
- Score(プロテインスコア: 図9では「266」)
- Matches(マッチしたペプチド数: 図9では「3」、閾値スコア以上のイオンスコアを持つペプチド数は括弧内の「3」)
- Sequences(マッチした配列数: 図9では「3」、閾値スコア以上のイオンスコアを持つ配列数は括弧内の「3」)
- タンパク質情報(図9では「Serine/threonine-protein...」)
- タンパク質に帰属するペプチドに関する情報(Query: MS/MSの質量スペクトルデータの通し番号(クエリ番号と呼んでいます)、 Observed: プリカーサイオン質量実験値、 Mr (expt): プリカーサイオン質量実験値のMr値、 Mr (calc): プリカーサイオン質量理論値のMr値、 ppm: プリカーサイオン質量の実験値と理論値の差(ppm表示)、 Miss: 未消化サイト数、 Score: イオンスコア(ペプチドスコアとも呼んでいます)、 Expect: 期待値、 Rank: イオンスコア順位、 Unique: このペプチドがこのタンパク質のみに存在する場合は U を表示、 Peptide: ペプチドのアミノ酸配列と修飾の種類・結合サイト)

(5) Search Parameters

検索条件を表示しています。

アクセッション番号(図9では「ELM1_YEAST」)は「Protein View」のリンクになっており、図10に示すように、そのタンパク質に関するより詳細な情報を見ることができます。

図9 検索結果ページ (Peptide Summary Report)



また、クエリ番号(MS/MSの質量スペクトルデータの通し番号、図9では1、2、3の番号)は「Peptide View」ページのリンクになっており、図11(クエリ番号3)に関する「Peptide View」ページ)に示すように、プロダクトイオンスペクトルのピークに対応するイオンシリーズの帰属状況などを一覧することができます。

図9のMIS検索結果では、閾値イオンスコアより大きいイオンスコアを持つ3つのペプチドが同定され、これらのペプチドは ELM1_YEAST に唯一帰属することが示されています。従いまして、実験サンプルに含まれるタンパク質は ELM1_YEAST であると推定することができます。

図11はクエリ番号3のMS/MSスペクトルに対して、最も高いイオンスコア(117)でマッチしたペプチドPSSPLMDRTV GKの検索結果を示しています。スペクトル図の下にあるマッチング状況説明文にあるように、N末端側から3番目のセリン(S3)がリン酸化されていますが、フラグメンテーションの際にこのリン酸基の脱離(ニュートラルロス)が起こっているために、プロダクトイオン y(10) と y(11) の質量は98Da 小さくなり、スペクトル図のプロダクトイオンピークに付けられたラベルでも確認することができますが、y(10)-98 および y(11)-98 のピークにマッチしていることがわかります。

クエリ番号3のMS/MSスペクトルにマッチするペプチドは多数存在しますが、「Peptide View」ページの下方にはマッチしたペプチドのうちのイオンスコア上位10件のペプチドリストが図12のように表示されます。「Score」項はマッチしたペプチドのイオンスコアを示していますが、一般的に、有意にマッチした「ランク1位」のペプチドと偶然にマッチした「ランク2位」以降のペプチドのイオンスコアには大きな差が認められます。「Sequence」項の配列はリンクになっており、クリックすると「Peptide View」ページはこの配列に対応する内容に書き換えられます。

図12において、ランク1位、2位、4位のペプチドのイオンスコアはそれぞれ117.2、102.4、30.2で、同じアミノ酸配列を持っており、リン酸化されたアミノ酸の位置(S3またはS2、T9)が異なります。「Site Analysis」項にはこれらのスコアの差から計算された、リン酸が存在する位置の確率が表示されており、この例では、S3がリン酸化されたランク1位のペプチドが85%の確率で正しいことを示しています。

検索結果の詳しい見方については「5 検索結果」をご覧ください。

図10 「Protein View」ページ

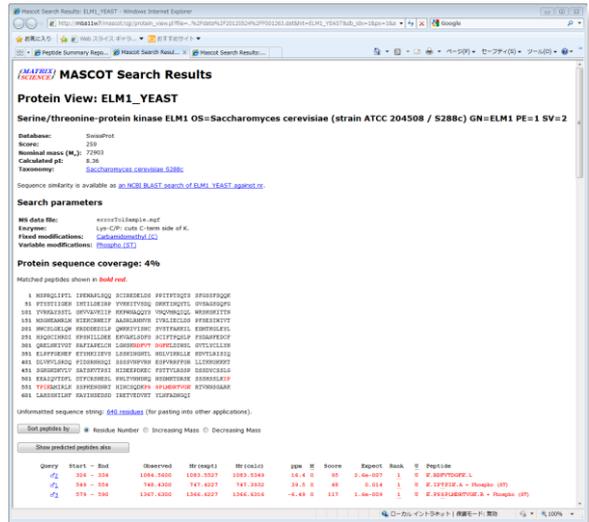


図11 「Peptide View」ページ

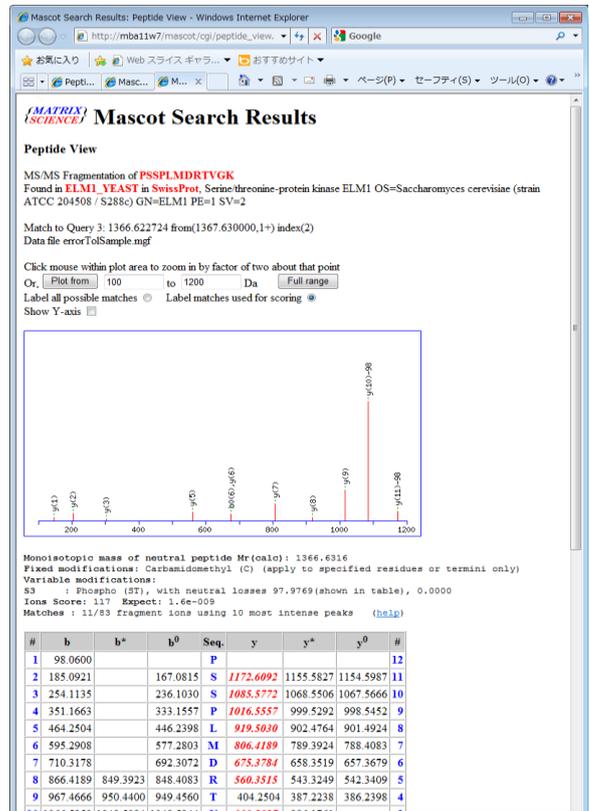


図12 イオンスコア上位 10 件のペプチド

Score	Mr(calc)	Delta	Sequence	Site Analysis
117.2	1366.6316	-0.0089	PSSPLMDRTV GK	Phospho S3 84.59%
102.4	1366.6316	-0.0089	PSSPLMDRTV GK	Phospho S2 15.41%
31.0	1366.5959	0.0268	PSSGGGGLFSFFK	
30.2	1366.6316	-0.0089	PSSPLMDRTV GK	Phospho T9 0.00%
23.2	1366.6616	-0.0389	SDVDYIDSLV GK	
22.6	1366.5959	0.0268	PSSGGGGLFSFFK	
20.4	1365.6639	0.9589	FAETPMLPGFK	
20.1	1366.6605	-0.0378	LASRLEEER GK	
19.4	1365.5925	1.0302	PALDTNDAARDK	
19.3	1366.5588	0.0639	PSHCEGVSISSK	

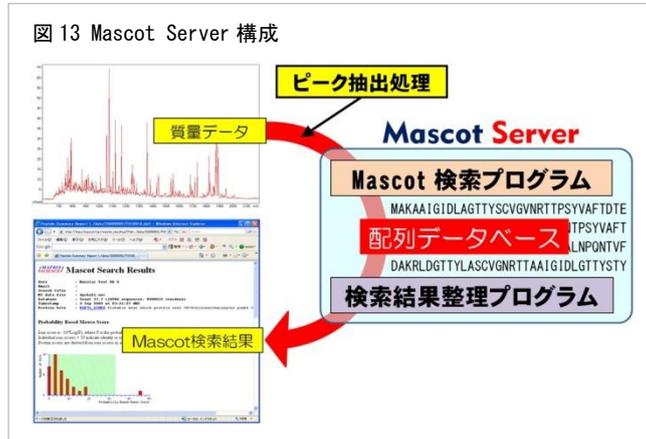
▶ 2 システム構成

2-1 ソフトウェア構成

Mascot Serverは、質量分析計から得られた質量スペクトルに一致するタンパク質あるいはペプチドを配列データベースから検索し、実験サンプルに含まれるタンパク質を同定するソフトウェアです。Mascot Serverは図13に示すように、3つの部分から構成されています。

- (1) Mascot検索プログラム
- (2) 配列データベース管理プログラム
- (3) 検索結果整理プログラム

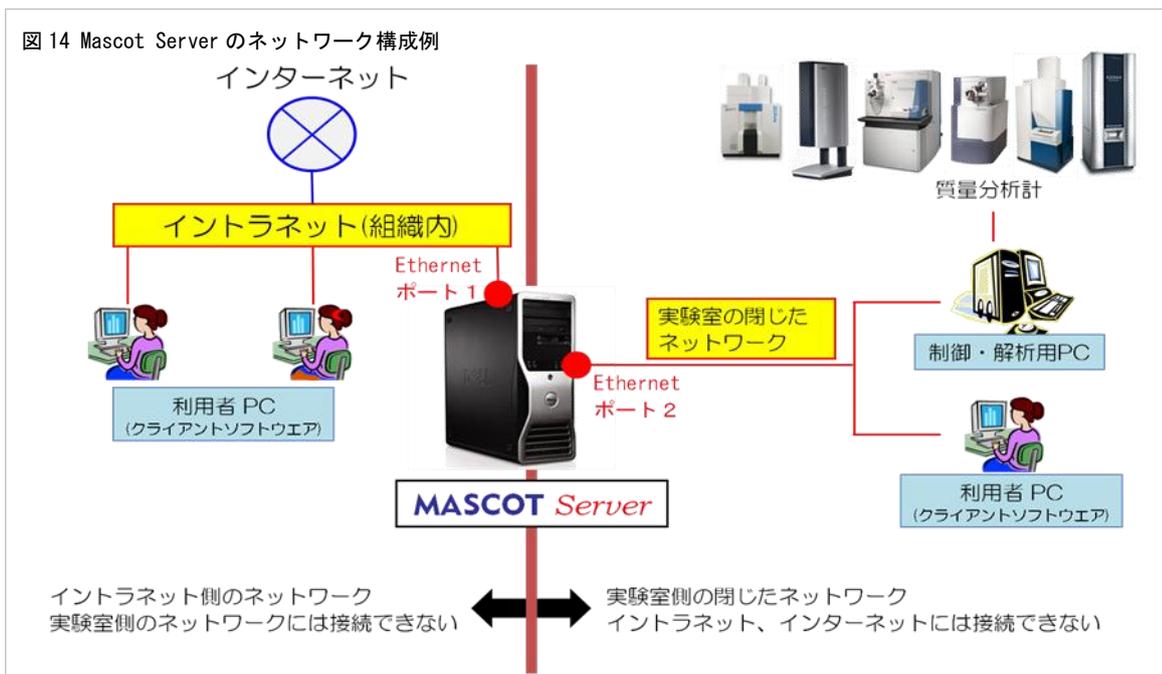
質量スペクトルは実験サンプルに含まれるタンパク質由来しますので、求めるタンパク質が配列データベースに存在するとすれば、「配列データベースを利用して、質量スペクトルをタンパク質IDに変換するソフトウェア」と表現することもできます。



2-2 ネットワーク構成

Mascot ServerはWebサーバとともに動作します。Mascot ServerはPCにインストールされたWebサーバ(IISやApache)上に構築されたWebサイト(Mascotサイト)として構成され、IE(インターネット・エクスプローラ)のようなWebブラウザや質量分析計に付随する解析ソフトウェアなどのクライアントソフトウェアからHTTPプロトコルを使ってアクセスし、利用します。従いまして、ネットワーク上のクライアントソフトウェアを介して複数のユーザが同時にMascot検索を実行することができます。

図14は、それぞれ独立した2つのネットワーク(「イントラネット(組織内)」および「実験室の閉じたネットワーク」)内に存在するクライアントソフトウェアからMascot Serverにアクセスできるようにしたネットワーク構成例です。



「イントラネット(組織内)」内の利用者PCからはMascot Server PCの「Ethernetポート1」を介してMascot Serverにアクセスすることはできますが、「実験室の閉じたネットワーク」内のPCにはアクセスすることはできません。逆に、「実験室の閉じたネットワーク」内の利用者PCや解析用PCからはMascot Server PCの「Ethernetポート2」を介してMascot Serverにアクセスする

ことはできますが、「イントラネット(組織内)」内のPCやインターネットにはアクセスすることができません。Mascot Serverはイントラネットを経由してインターネットにアクセスし、配列データベースの更新を行うことができます。

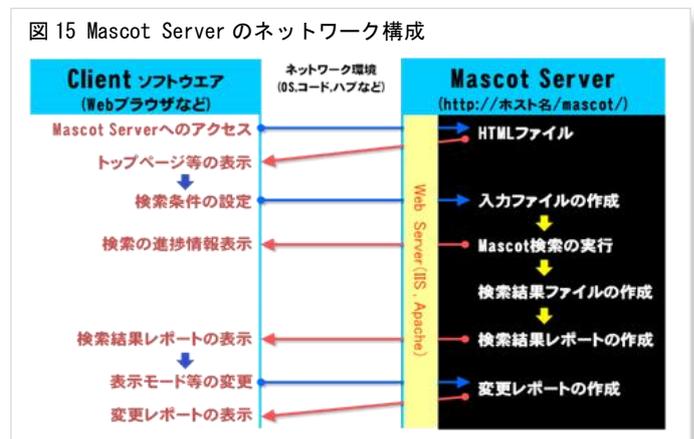
ネットワークの構造・環境はお客様により異なりますので、Mascot Serverをネットワークに配置して運用する際は、お客様の情報システム・ネットワーク管理者にご相談ください。

2-3 クライアントソフトウェア

Mascot Serverに対してクライアントとして動作するソフトウェアには次のようなものがあります。

- (1) Webブラウザ (IE, Safari, Firefox, Chromeなど)
- (2) Mascot Daemon, Mascot Distillerなど弊社製品
- (3) Analyst QS, BioTools, Launchpad, Bioworks, Proteome Discoverer, MassLynxなどの質量分析計ベンダーの解析ソフトウェア

Mascot 検索の際、クライアントソフトウェア (Client側) と Mascot Server (Server側) は、ネットワークを介して図15に示すようなデータのやり取りを行っています。クライアントソフトウェアは Mascot Server とネットワークで接続されていれば、どこに設置されていてもかまいません。



2-4 フォルダ構造

Mascot Serverは `C:\inetpub\mascot` フォルダにインストールされており、図16に示すようなフォルダ構造を持っています。cgi、x-cgi、html フォルダは次の URL にマップされています。

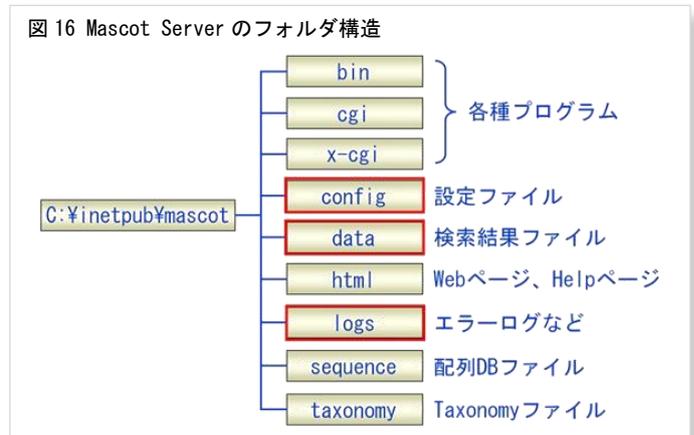
```

cgi → http://ホスト名/mascot/cgi
x-cgi → http://ホスト名/mascot/x-cgi
html → http://ホスト名/mascot/

```

Mascot Serverが必要とする様々な設定ファイルは `config` フォルダに、検索結果ファイルは `data` フォルダに、検索ログやエラーログなどのログ情報ファイルは `logs` フォルダに、配列データベースファイルは `sequence` フォルダに格納されています。

`C:\inetpub\mascot` フォルダを定期的にバックアップするのが理想的ですが、バックアップ先の記憶容量等の問題ですべてのフォルダのバックアップが難しい場合は、少なくとも `config`、`data`、`logs` フォルダをバックアップするようにしてください。



2-5 配列データベース

FASTAフォーマットのアミノ酸配列データベースおよび塩基配列データベースをセットアップし、Mascot検索に利用することができます。FASTAフォーマットに関しては次のページをご覧ください。

<http://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml>

配列データベースの管理に関しては次のページをご覧ください。

http://ホスト名/help/seq_db_setup.html (または、http://www.matrixscience.com/help/seq_db_setup.html)
http://www.matrixscience.jp/pdf/jap_database_setup.pdf

2-6 マニュアル

図1の「Welcome」トップページからアクセスすることができます。

(1) 日本語マニュアル

「Welcome」トップページ下方にある[日本語マニュアル]リンクをクリックするか、次の URL にアクセスしてください。

http://www.matrixscience.jp/jap_manual_index.html

(2) 英文マニュアル

「Welcome」トップページ下方にある[Setup & Installation Manual]リンクをクリックしてください。

(3) 英文ヘルプ

「Welcome」トップページ右上にある[HELP]リンクをクリックするか、次の URL にアクセスしてください。

http://ホスト名/help_index.html

http://www.matrixscience.com/help_index.html

表示された「Help Topic Index」ページは次のような項目で構成されています。

MASCOT - General : ASMSなどで発表したプレゼンテーション資料などをご覧いただくことができます。

- [Mascot search overview](#)
- [A History of Mascot and Mowse](#)
- [ASMS 2001 ~ 2012 User Meeting presentations](#)
- [Mascot FAQ's](#)

Using Mascot : Mascot Serverの使い方をまとめてあります。

- [Search parameter reference](#)
- [Data file format](#)
- [Scoring algorithm](#)
- [Result Report Overview](#)
- [Summary Reports for PMF](#)
- [Summary Reports for MS/MS](#)
- [MS/MS Results Interpretation](#)
- [Sharing result reports](#)
- [Error tolerant search](#)
- [Exporting result reports](#)
- [Decoy Databases](#)
- [Top-down Searches](#)
- [Percolator](#)
- [Quantitation](#)

Mascot System Administration : 配列データベースのセットアップ方法をまとめてあります。

- [Sequence database setup](#)
- [PC Hardware for Mascot Server](#)

Instrument Specific Tips : 質量分析計のデータ処理システムとMascot Serverの連携方法をまとめてあります。

- [AB SCIEX Analyst](#)
- [AB SCIEX Data Explorer](#)
- [AB SCIEX 4000 / 5000 Series \(TOF/TOF\)](#)
- [Micromass Masslynx](#)
- [Thermo Finnigan Xcalibur](#)

Protein Identification : タンパク質同定の原理などをまとめてあります。

- [Peptide Mass Fingerprint](#)
- [Sequence Query](#)
- [MS/MS Ion Search](#)

Protein chemistry : タンパク質同定に関連するタンパク質科学に関してまとめてあります。

- [Post translational modifications](#)
- [Enzymes](#)
- [Autolysis](#)
- [Sequence databases](#)
- [Nucleic acid translation](#)
- [Amino acid reference](#)
- [Contaminants](#)
- [BLAST & FastA](#)

Mass spectrometry : 質量分析計に関わるトピックスを紹介しています。

- [MS/MS fragmentation](#)
- [Accuracy & resolution](#)

Miscellaneous : その他の関連事項についてまとめてあります。

- [Mascot brochure \(PDF 2.7 Mb\)](#)
- [PC Hardware for Mascot](#)
- [Privacy FAQ](#)
- [Web Browser Compatibility](#)

2-7 Mascotの最新情報

弊社のホームページ (<http://www.matrixscience.com/>) にアクセスしてください。

(1) [\[WHAT'S NEW\]](#)リンク

バージョンアップ情報やプレスリリース等に関する最新情報を掲載しています。

(2) [\[SUPPORT\]](#)リンク

製品別の最新技術情報を入手することができます。また、製品試用版をダウンロードすることができます。

▶ 3 Mascot検索の種類と検索条件

3-1 Mascot検索の種類

Mascot Serverは入力として与える質量データの種類に対応して3つの検索方法をサポートしています。

(1) PMF検索 : Peptide Mass Fingerprint : ペプチドマスフィンガープリント法

MSの質量スペクトル(ペプチドイオン質量のセット)にマッチするタンパク質を配列データベースから検索する方法です。

(2) SQ検索 : Sequence Query : シーケンスクエリ法

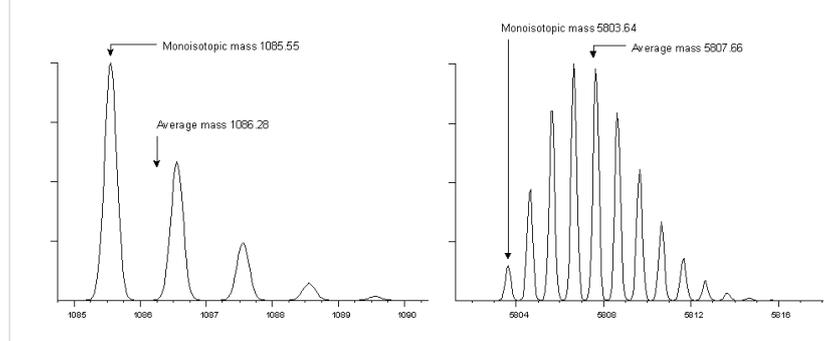
プリカーサイオン質量を構成する部分要素(アミノ酸配列、構成アミノ酸、プロダクトイオン質量など)にマッチするペプチドを配列データベースから検索する方法です。

(3) MIS検索 : MS/MS Ions Search : MS/MSイオンサーチ法

MS/MSの質量スペクトル(プリカーサイオン質量とプロダクトイオン質量・強度のセット)にマッチするペプチドを配列データベースから検索する方法です。

生データとしての質量スペクトルデータには同位体ピークが含まれています。たとえば、 m/z の値がそれぞれ1085と5803の1価のプロトン化ペプチドは図17に示すような同位体分布を持っています。電荷は1価(MH⁺)ですので、同位体ピークは1 Daの幅で分布します。 m/z の値が大きくなるほど観測される同位体ピークの数は増え、同位体分布の一番左にあるモノアイソトピック質量ピークの強度は他の同位体ピークのそれに比べて相対的に小さくなります(図17の右図)。

図 17 $m/z=1085$ と 5803 のペプチドの同位体分布



Mascot Serverはモノアイソトピックなピーク質量に対して検索を行います。モノアイソトピックなピーク質量のセットをピークリストと呼んでおり、これをMascot検索用の入力データとして使います。ピークリストは各質量分析装置に付属する解析用のソフトウェアや弊社のMascot Distillerを利用して作成しますが、通常はピークリストファイルとして保存し、Mascot検索の際にピークリストファイルを指定します。

SQ検索およびMIS検索では、プロダクトイオンの電荷は1価または2価(たとえば、**b** および **b⁺** や **y** および **y⁺** など)を仮定して検索します。プロダクトイオンの電荷が3価以上の場合、MS/MSの質量スペクトルデータを deconvolve (逆畳み込み)して1価の状態に変換した質量データを作成する必要があります。この変換作業には弊社製品のMascot Distillerを利用することができます。Mascot Distillerについては次のページをご覧ください。

<http://www.matrixscience.com/distiller.html>

3-2 検索条件

表1に検索条件項目とその内容および各検索項目に対応する検索方法(✓で示しています)をまとめました。

検索条件は、配列データベース(あらかじめMascot Serverにセットアップすることが必要です)、実験条件、質量分析計の質量特性、質量データなどです。ほとんどの検索条件は自分で決められるものではありませんので、「事実」を設定し、Mascot 検索を実行してください。検索条件に関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/search_field_help.html

http://www.matrixscience.com/help/search_field_help.html

表1 検索条件項目とその内容および対応する検索方法

検索条件項目	内容	PMF	SQ	MIS
Your name	名前を入力してください。	✓	✓	✓
Email	Emailアドレスを入力してください。	✓	✓	✓
Search title	検索タイトル(サンプルに関わる情報など)を入力してください。後日、この検索結果にアクセスしやすくするためにできるだけ詳しい内容を入力することをお勧めします。	✓	✓	✓
Database(s)	検索したい配列データベースを選択してください。[Ctrl]キーを押しながらクリックすることにより、複数の配列データベースを選択することができます。	✓	✓	✓
Enzyme	タンパク質を消化するために使用した消化酵素を選択してください。	✓	✓	✓
Allow up to # missed cleavages	タンパク質の消化状態を選択してください。たとえば 2 を選択すると、(未消化サイト数 2, 1, 0) のペプチドに対して検索します。	✓	✓	✓
Quantitation	サンプルに含まれるタンパク質の量を解析するために使用した実験手法名を選択してください。最後に[MD]がついているエントリは Mascot Distiller (別途ライセンスが必要です)と連動して動作します。		✓	✓
Taxonomy	検索対象としたい生物種を選択してください。NCBIInrなどの総合配列データベースに対して有効です。	✓	✓	✓
Fixed modifications	修飾を選択してください。指定されたすべての(Fixされた)アミノ酸に対する修飾を考慮して検索を行います。システインの誘導体化のような意図された化学修飾に対応します。	✓	✓	✓
Variable modifications	修飾を選択してください。指定されたアミノ酸に対する修飾を考慮する場合としない場合の2通りの検索を行います。実験サンプルの前処理の段階で発生する酸化などの偶発的な修飾や翻訳後修飾に対応します。	✓	✓	✓
Protein mass	検索範囲ウィンドウを指定してください。たとえば、“30”を指定した場合は30KDaの範囲にある連続したアミノ酸配列を対象として検索を行います。	✓		
Peptide tol. ±	質量分析計の特性値としてのペプチド質量の誤差(Da, mmu, %, ppm)の単位で指定してください。たとえば、1.0 Daを指定した場合は(ペプチド質量-1.0)Daから(ペプチド質量+1.0)Daの質量範囲にマッチするペプチドを検索します。	✓	✓	✓
# ¹³ C	¹³ C同位体ピークを検出してMS/MSのスキャンを実行した可能性がある場合に選択してください。[Peptide tol. +/-]で指定した範囲に加え、マイナス側の1または2Da先にその範囲を拡張して検索します。		✓	✓

検索条件項目	内容	PMF	SQ	MIS
MS/MS tol. ±	質量分析計の特性値としてのプロダクトイオン質量の誤差を (Da, mmu) の単位で指定してください。		✓	✓
Peptide Charge	通常、質量データファイルにはプリカーサイオンの電荷が記載されていますので、この項目は無視してください。		✓	✓
Mass values	ペプチドの電荷状態 (MH+, Mr, M-H) を指定してください。	✓		
Monoisotopic または Average	質量データが Monoisotopic か Average かを指定してください。	✓	✓	✓
Data file	質量データファイルを[ファイルの選択]ダイアログを通じて指定してください。	✓		✓
Query	[Data file]で質量データファイルを指定しない場合は、質量データを入力してください。	✓	✓	
Data format	ピークリストファイルの書式を選択してください。			✓
Precursor	通常、質量データファイルにはプリカーサイオン質量の値が記載されていますので、この項目は無視してください。			✓
Instrument	プリカーサイオンの開裂様式(発生するプロダクトイオンの種類)を指定するために、測定に用いた質量分析計の種類を選択してください。		✓	✓
Error tolerant	チェックボックスをチェックしてください。自動的にError tolerant検索を実行します。			✓
Decoy	チェックボックスをチェックしてください。Decoyデータベースの同時検索を実行します。	✓	✓	✓
Report top	検索結果ページに表示させたいタンパク質数を選択してください。 [AUTO]を選択した場合は、有意なスコアを持つタンパク質が検索結果ページに表示されます。	✓	✓	✓
Start Search	Mascot検索が実行されます。	✓	✓	✓
Reset Form	設定した内容をリセットします。	✓	✓	✓

3-3 PMF検索 : Peptide Mass Fingerprint

PMF検索では、MSの質量スペクトルデータから抽出したモノアイソトピックなピーク質量各々に関して、検索条件に一致するペプチド(アミノ酸配列)を配列データベースから検索し、マッチしたペプチドとそれが帰属するタンパク質の関係を集計することにより、MSの質量スペクトルデータが意味する(実験サンプルに含まれる)タンパク質を同定します。

WebブラウザからPMF検索条件設定ページ(図18:「Welcome」トップページ→[Mascot]リンク→[Peptide Mass Fingerprint]リンク)にアクセスし、検索条件を入力または選択して設定してください。

表1の「PMF」の項にチェック(✓)のある検索条件項目を利用することができます。ほとんどの検索条件は実験条件(ゲル内消化の際に使用した消化酵素、メルカプト基の再結合を防ぐために使用した還元試薬など)や質量分析計の特性値(質量精度)をそのまま使用しますので、あらかじめそれらの情報を準備してください。

質量データは[Query]ウインドウにモノアイソトピックなピーク質量を一行にひとつずつリスト形式で入力することもできますが、それらがピークリストファイルとしてまとめられている場合は[Data file]の[参照]ボタンから直接読み込んでください。

PMF検索には SwissProt のような重複度の少ない配列データベースが適しています。EST やゲノムの配列データベースに対する検索は意味がありません。

検索条件の設定および質量データの指定が終わりましたら[Start Search...]ボタンを押してください。PMF検索が実行され、検索が終了すると検索結果ページが表示されます。

検索に要する時間は、検索に使用した配列データベースのサイズ(配列データベースを構成するタンパク質エントリの総数)、[missed cleavages]、[Variable modifications]、[Peptide tol +/-]の設定値、ピーク質量データ数などにより異なります。これらの値が大きくなるほど検索空間が広がるために、検索時間は長くなります。

複数のタンパク質を含む試料から得られたMSの質量スペクトルデータの品質が良い場合は複数のタンパク質を同時に同定することも可能です。次の検索例をご覧ください。

http://www.matrixscience.com/cgi/master_results.pl?file=./data/F981138.dat

図 18 PMF 検索条件設定ページ

3-4 MIS検索 : MS/MS Ions Search

MIS検索では、次の2段階の検索が進行します。

- (1) プリカーサイオン質量にマッチするペプチドを配列データベースから検索します(この時点でペプチドのアミノ酸配列と帰属するタンパク質は判明していません)。
- (2) (1)で検索された全てのペプチドに関して、検索条件として指定したイオンシリーズに対応するプロダクトイオン表を作成し、プロダクトイオンピークとのマッチングを行うことにより、最もマッチング状況が良いペプチドを検索し、MS/MSの質量スペクトルが意味するペプチドを同定します。

質量分析計の種類により生成するプロダクトイオンの種類(図19に示すようなイオンシリーズ)は異なりますので、検索条件として質量分析計の種類を選択し、検索対象となるイオンシリーズを指定します。

同定されたペプチドと、それが帰属するタンパク質の帰属関係を整理することにより、実験サンプルに含まれているタンパク質を推定します。

WebブラウザからMIS検索条件設定ページ(図20:「Welcome」トップページ→[Mascot]リンク→[MS/MS Ion Search]リンク)にアクセスし、検索条件を入力または選択して設定してください。表1に検索条件項目とその内容をまとめました。「MIS」の項にチェック(✓)のある検索条件項目を利用することができます。検索条件のほとんどは実験条件(ゲル内消化の際に使用した消化酵素、メルカプト基の再結合を防ぐために使用した還元試薬など)や質量分析計の特性値(質量精度)をそのまま使用しますので、あらかじめそれらの情報を準備してください。

質量データファイルは[Data file]の[参照]ボタンから直接読み込んでください。指定した質量データファイルの書式が「Mascot generic (拡張子は“MGF”)」ではない場合は[Data format]から対応する書式を選択してください。

検索するイオンシリーズを指定するために、質量分析計の種類を[Instrument]から選択してください。

検索条件の設定および質量データの指定が完了したら[Start Search...]ボタンを押してください。MIS検索が実行され、検索が終了すると検索結果ページが表示されます。

検索に要する時間は、検索に使用した配列データベースのサイズ(配列データベースを構成するタンパク質エントリの総数)、[missed cleavages]、[Variable modifications]、[Peptide tol +/-]の設定値、MS/MSスペクトルデータ数などにより異なります。これらの値が大きくなるほど検索空間が広がるために、検索時間は長くなります。

図 19 プロダクトイオンとイオンシリーズ

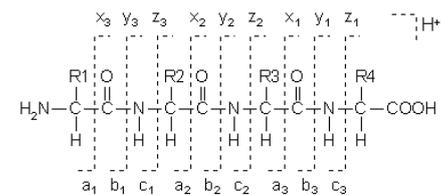


図 20 MIS 検索条件設定ページ

3-5 SQ検索 : Sequence Query

SQ検索では、プリカーサイオン質量およびプリカーサイオン質量の構成要素(アミノ酸配列、構成アミノ酸、プロダクトイオン質量、配列タグ(連続したアミノ酸配列とその両端におけるプロダクトイオン質量のセット))にマッチするペプチドを配列データベースから検索し、タンパク質を同定します。

PMF検索およびMIS検索では質量データを指定しましたが、SQ検索では次の書式で検索クエリを指定します。

M seq(...) comp(...) ions(...) tag(...), etag(...)

Mはプリカーサイオン質量です。seq(アミノ酸配列)、comp(構成アミノ酸)、ions(プロダクトイオン質量)、tag(配列タグ)、etag(Error Tolerant 配列タグ)はオプションとして指定することができ、0個以上いくつでもかまいません。複数の標準配列タグ「tag(...)」と1つ以上の Error Tolerant 配列タグ「etag(...)」が混在する場合は全て Error tolerant 配列タグ「etag(...)」として処理されます。

ions(...), tag(...), etag(...) に対しては確率的なスコアリングアルゴリズムが適応されます。一方、seq(...) と comp(...) はフィルタとして機能し、一致するペプチドが存在しない場合、このクエリは破棄されます。なお、seq(...) は指定したアミノ酸配列に一致するペプチドを検索しますが、Blast検索とは異なります。

たとえば、図21に示す検索条件と、[Query]入力欄に、

1489.430 tag(650.213,GWSV,1079.335)

を入力し、[Start Search ...]ボタンを押してMascot検索を実行してください。次のような検索結果が得られます。

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1	1489.4300	1489.4300	1489.7348	-0.3048	0	64	4e-007	1	U	K.LQGIVSWGSGCAQK.N

「K.LQGIVSWGSGCAQK.N」にスコア64(期待値は4E10-007)でマッチしますので、この配列である可能性が高いことがわかります。また、このアミノ酸配列は SwissProt の中では「TRY1_BOVIN」にのみ存在するユニークな配列ですので、結果として「TRY1_BOVIN」が同定されたことがわかります。

検索クエリの書式などに関する詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/sq_help.html

http://www.matrixscience.com/help/sq_help.html

図 21 SQ 検索条件設定ページ

▶ 4 スコアリング

4-1 スコアと期待値

Mascot Serverで使われているスコアリングの中身は公開されていませんので、検索結果ページに記載されているスコアなどの数値に対して検算することはできませんが、次の2つの資料から、Mascot Serverのスコアリングの特徴を理解することができます。

http://www.matrixscience.com/pdf/asms_tutorial_2005.pdf

http://www.matrixscience.com/help/scoring_help.html

Mascot Serverでは、質量データのペプチドに対するマッチ(MIS検索の場合はプロダクトイオンに対するマッチ)は確率事象(random event)として取り扱います。指定された検索条件の下で、質量データがペプチド(MIS検索の場合はプロダクトイオン)にマッチした時の確率 P は先験的(a priori)に決まり、配列データベースの種類やサイズ(エントリ数)には依存しません。確率 P の値は非常に小さいため、次の式を使ってスコアに変換し、検索結果ページに表示しています。

$$\text{スコア} = -10 \times \log_{10}(P)$$

たとえば、4個の異なる質量データが、それぞれ4個の異なるペプチドにマッチし、それら全てがひとつのタンパク質に帰属するときの確率Pは絶対的に決まり、そのときの確率がたとえば $P=0.000001 (=1 \times 10^{-6})$ だった場合、上の式からスコアは 60 になり、この値がプロテインスコアとして検索結果ページに表示されます。質量データがプロダクトイオンにマッチした時のスコアも同様に計算され、イオンスコア(またはペプチドスコア)として検索結果ページに表示されます。スコアは与えられた質量と質量誤差などの質量に関係する検索条件によって絶対的に決まり、配列データベースの種類やタンパク質の数(エントリ数)には依存しません。

検索によってヒットしたタンパク質が有意かどうか(“Significant match” OR “Random match”?)は、検索に使用した配列データベースのタンパク質エントリ数 N から決まる閾値スコアを使って判定しています。

$$\text{閾値スコア} = -10 \times \log_{10}(1/N * 0.05)$$

たとえば、タンパク質エントリ数が5,000件と500,000件の2種類の配列データベースを使った場合の閾値スコアはそれぞれ 50 と 70 になりますので、上記のスコア 60 のタンパク質の例では、タンパク質エントリ数が5,000件の配列データベースに対して検索した場合は閾値スコア 50 を超えていますので有意なヒットになりますが、タンパク質エントリ数が500,000件の配列データベースの場合は閾値スコアが 70 ですので、それよりも小さいため、有意なヒットではないという判定になります。すなわち、配列データベースのエントリ数によって閾値スコアは変化します。

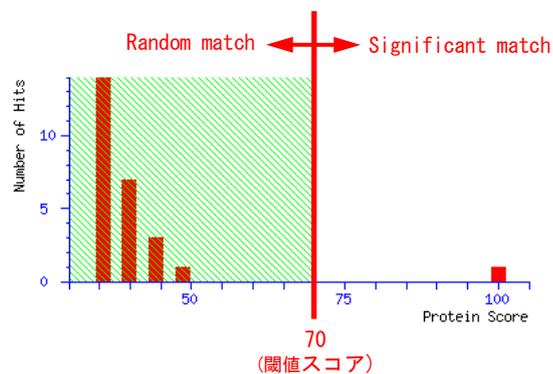
検索結果ページには、図22に示すような「Mascot Score Histogram」が表示されます。横軸はタンパク質のプロテインスコア、縦軸はプロテインスコアに対するタンパク質のヒット件数を示しています。閾値スコアよりも右側の領域は「Significant match」、左側の緑色斜線部分は「Random match」に相当しますので、検索によって有意にヒットしたタンパク質が得られたかどうか、また、ヒットしたタンパク質全体がどのようにスコア分布をしているかが一目でわかります。

配列データベースのタンパク質エントリ数によって閾値スコアは変化しますが、次のように期待値(検索結果ページでは Expect value として表示されます)を定義すると、配列データベースのエントリ数に関係なく、期待値 0.05 を閾値として使うことができます。すなわち、質量データのペプチドあるいはプロダクトイオンへのマッチング操作を「試行」と考えると、配列データベースのタンパク質各々に対してマッチング操作を行いますので、タンパク質エントリ数は「試行回数」に相当します。確率 P の事象を N 回試行したときにその事象が起こる平均的な回数(平均値または期待値と呼んでいます) E は、

図 22 閾値スコアを使った有意性の判断

Mascot Score Histogram

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



$$E = P \times N$$

になりますので、これをMascot検索に当てはめると次のようになります。

$$\text{平均値(期待値) } E = \text{質量データがマッチした時の確率 } P \times \text{タンパク質エントリ数 } N$$

Mascot Serverでは期待値 $E=0.05$ を統計的に有意かどうかの閾値として採用しています(閾値スコアは期待値 $E=0.05$ に対応する確率 P から求めることができます)。

たとえば、上記のスコア 60 のタンパク質の例では、エントリ数 5,000 の配列データベースに対する期待値は $5000 \times 10^{-6} = 0.005$ となり、閾値としての期待値 $E=0.05$ よりも小さいため統計的に有意であり、同定された可能性が高いことになります(統計学的に表現すれば、0.5% の確率で正しいとは言えないことになります)。一方、エントリ数 500,000 の配列データベースに対する期待値は $500000 \times 10^{-6} = 0.5$ となり、閾値としての期待値 0.05 よりも大きいため、同定された可能性は低くなります(統計学的に表現すれば、50%の確率で正しいとは言えないことになります)。

なお、SQ検索およびMIS検索では質量データとプロダクトイオンのマッチングからペプチドを同定しますので、閾値スコアや期待値を計算する際のエントリ数は、プリカーサイオン質量にマッチしたペプチドの数になります。

4-2 Identity 閾値と Homology 閾値

SQ 検索及び MIS 検索では 2 種類の閾値スコアを考慮することができます。すなわち、プリカーサイオン質量にマッチするペプチドの数を利用する「Identity 閾値」スコアと、プリカーサイオン質量にマッチするペプチドのイオンスコア分布から決定される「Homology 閾値」スコアの 2 種類です。

図 23 の上段は、ひとつの MS/MS スペクトルデータに対するイオンスコア分布を示しています。横軸はイオンスコア、縦軸はイオンスコアに対応するペプチド数です。横軸には、Homology 閾値スコア (赤い点) と Identity 閾値スコア (青い点) をプロットしています。

図 23 の下段は、上段とは別の MS/MS スペクトルデータですが、イオンスコア上位 10 件のペプチドと Homology および Identity 閾値スコアの表示例です。Identity 閾値スコアは、検索対象となるペプチドの数から計算される、期待値が 0.05 に対応する理論的な意味合い

の閾値スコアであるのに対して、Homology 閾値スコアは、検索対象となるペプチドのイオンスコア分布曲線の外れ値 (Outliner) に対応する経験的な意味合いの閾値スコアです。多くの場合、

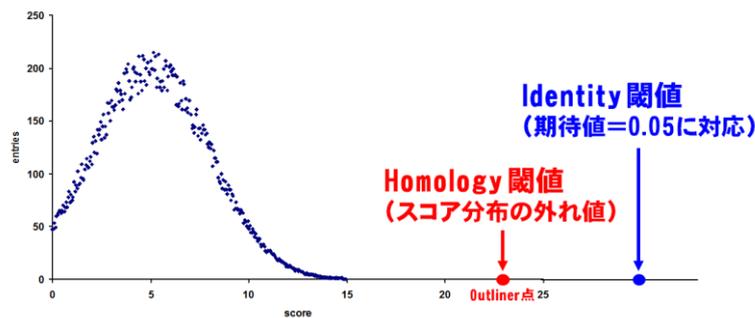
Homology 閾値スコア > Identity 閾値スコア

であり、Identity 閾値スコアを使って有意性の判定を行う場合は Homology 閾値スコアを考慮する必要はありませんが、図 23 の上段に示すように、

Homology 閾値スコア < Identity 閾値スコア

の場合は、両者の中間のイオンスコアを持つペプチドに関しては、Homology 閾値スコアを使って有意性を考察することもできますので、図 23 の下段のように、検索結果にはこれら 2 つの閾値スコア表示するようにしています。

図 23 イオンスコア分布と Homology 及び Identity 閾値



Prot	Score	Expect	Delta	Hit	Protein	Peptide
EFTU	69.3	2.9e-005	0.4765	1	EFTU_ECOLI	K.MVVTLIHPIAMDDGLR.F
Elon	15.3	7.4	0.4699			R.IITHPLGMVEDD IVTK.I
EFTU	11.3	18	0.4145			R.DLTVLIGKHPFSALR.Q
Elon	10.2	23	0.4774			R.GLGVTDNGEPIQVPVGTK.T
EFTU	9.7	26	0.5209			R.DITVYVAIDIAEFDAPK.A
Elon	9.3	29	0.4257			R.DTLDVHLVPSIFIGR.T
EFTU	8.9	32	0.4985			R.IDVEPVFTAHPTEAVR.R
Elon	7.5	45	0.4046			R.SGAIIDLVLVDSVAALVPR.A
EFTU	7.3	46	-0.5219			K.DSAGAAALAGAVGYFVLT.K.A
Elon	5.6	68	-0.4671			R.DGRDTLISSGLMTR.A

4-3 FDR : False Discovery Rate

Mascot検索によってヒットしたペプチドが有意かどうかは、閾値(スコアあるいは期待値)を使って判定します。たとえば、図24の上段は、スコアに対するペプチドの度数分布を示していますが、閾値(緑色の線)を使って、閾値の右側にある6個のブロックを有意(正解)なペプチドに、閾値の左側のブロックを有意ではない(不正解な)ペプチドに分けています。

もし、実験サンプルが既知タンパク質で構成されているとすると、図24の中段のように、Mascot検索によってヒットしたペプチドが正解(橙色)か不正解(青色)かを分けることができます。図24の下段は正解と不正解とを完全に分離した図です。

図24の下段の図では、閾値を使って①~④の4つの象限に分けることができ、それぞれ次のような意味を持っています。

- ① True Positive : 正解と判定され、実際に正解
- ② False Positive : 正解と判定され、実は不正解
- ③ True Negative : 不正解と判定され、実際に不正解
- ④ False Negative : 不正解と判定され、実は正解

実際には、実験サンプルに含まれるタンパク質は不明ですので、正解と不正解を分離することはできませんが、次のように考えると「② False Positive」に含まれるペプチドの数を把握することができ、有意なペプチドの中に含まれる不正解ペプチドの割合であるFDR(False Discovery Rate)を評価することができます。

- (1) 実在する実験サンプルの質量データを実在する配列データベースに対して検索した場合、得られた有意なペプチドには正解(True Positive)と不正解(False Positive)が混在している。
- (2) 実在する実験サンプルの質量データを実在しない配列データベース(Decoyな配列データベース)に対して検索した場合、得られた有意なペプチドには正解(True Positive)は含まれていない(無視できる)。すなわち、不正解(False Positive)のみ含まれ、この不正解の数は(1)の不正解(False Positive)の数に相当する。
- (3) (2)で得られた有意なペプチド(False Positive)の数と、(1)で得られた有意なペプチド(False Positive+True Positive)の数から次の式を使ってFDRを計算することができる。

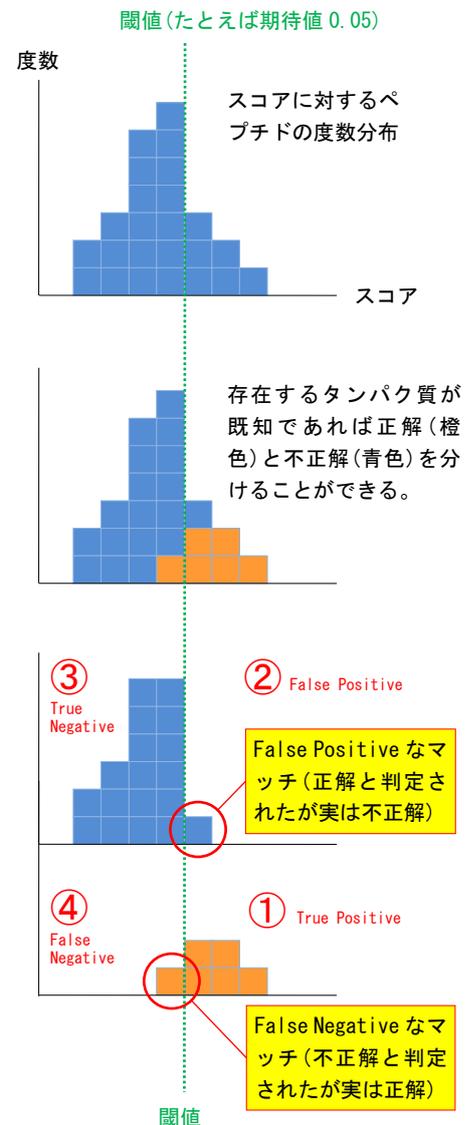
$$\begin{aligned} \text{FDR} &= \text{False Positive} / (\text{False Positive} + \text{True Positive}) \\ &= \text{Decoy配列DB由来の有意なペプチド数} / \text{Target配列DB由来の有意なペプチド数} \end{aligned}$$

Mascot Serverは、指定した配列データベース(Target配列データベースと呼んでいます)が持つアミノ酸構成比を使って、アミノ酸配列がランダムな配列データベース(Decoy配列データベースと呼んでいます)を新たに作成し、Target配列データベースとDecoy配列データベースを同時・独立に検索することにより、両者から得られた検索結果から自動的にFDRを計算し、その結果を「Summary Report」ページに表示します。

MIS検索条件設定ページ(図20)において、ページの下方にある[Decoy]をチェックし、[Start Search...]ボタンを押してMascot検索を実行してください。[Database(s)]で選択したTarget配列データベースに対するDecoy配列データベースをリアルタイムに作成し、2つの配列データベースに対して同時・独立に検索を行います。2つの配列データベースを検索しますので、通常の2倍の検索時間を要します。

図25はFDRの計算例を示しています。この計算例を表示するには次のURLにアクセスしてください。「Protein Family Summary」ページが表示されますので、フォーマットコントロールパネルの下にある▶Decoy search summary (random protein sequences)をクリックして展開してください。

図 24 検索結果の仕方



http://ホスト名/mascot/cgi/master_results_2.pl?file=F981139.dat
http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat

図 25 False Discovery Rate (FDR) の計算結果

Protein Family Summary

Filter Significance threshold p< 0.05 Max. number of families AUTO
 Ions score or expect cut-off 0 Dendrograms cut at 0
 Show Percolator scores

▼Decoy search summary (random protein sequences)

Peptide matches	in target	in Decoy	FDR
- above identity threshold	2343	96	4.10%
- above identity or homology threshold	2851	221	7.75%

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

[above identity threshold]の条件では、Target 配列 DB でヒットしたペプチドの数は 2343 個、Decoy 配列 DB では 96 個ですので、 $FDR=96 \div 2343 \times 100=4.10\%$ になります。同様に、[above identity or homology threshold]の条件では、 $FDR=221 \div 2851 \times 100=7.75\%$ になります。

Decoy 配列 DB に対する検索結果ページへのリンクになっています。

Protein Family Summary

Filter Significance threshold p< 0.06710 Max. number of families AUTO [help]
 Ions score or expect cut-off 0 Dendrograms cut at 0
 Show Percolator scores

▼Decoy search summary (random protein sequences)

Peptide matches	in target	in Decoy	FDR
- above identity threshold	2479	123	4.96%
- above identity or homology threshold	3104	359	11.57%

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

[Adjust to]の機能を使って、FDR=5%を満たす[above identity threshold]なペプチドを抽出し、それに対応するイオンスコアの閾値を求めることができます。

図25の上の例では、指定したイオンスコアの閾値条件を満たすペプチドを抽出し、FDRを求めています。すなわち、Target配列DBに対しては、期待値0.05未満に対応するIdentity閾値条件 [above identity threshold] の下でヒットしたペプチドの数は2343個、それに対してDecoy配列DBでは96個ですので、FDRは次のように計算することができます。

$$96 \div 2343 \times 100 = 4.10 \text{ (\%)}$$

同様に、Identity 閾値あるいは Homology 閾値のどちらか小さい方の値を閾値条件とした [above identity or homology threshold] の場合の FDR は次のように計算することができます。

$$221 \div 2851 \times 100 = 7.75 \text{ (\%)}$$

図25の下例では上の例とは逆に、[above identity threshold]行の[Adjust to]ボタン右で指定したFDR=5%(数字右肩の*は現在指定されていることを示しています)を満たすペプチドを抽出し、実際にはFDR=4.96%に対応するイオンスコアの閾値として 0.06710 を求めています。

Decoy配列データベースに関する詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/decoy_help.html
http://www.matrixscience.com/help/decoy_help.html

▶ 5 検索結果

5-1 検索結果ページの種類

表2に検索結果ページの種類をまとめました。検索結果ページは3種類あり、それぞれ、ヒットしたタンパク質のリスト（「Summary Report」ページ）、ヒットしたタンパク質の詳細情報（「Protein View」ページ）、ヒットしたペプチドの詳細情報（「Peptide View」ページ）を表示します。また、「Summary Report」ページは、検索方法と表示される情報により5種類に分けられます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/results_help.html

http://www.matrixscience.com/help/results_help.html

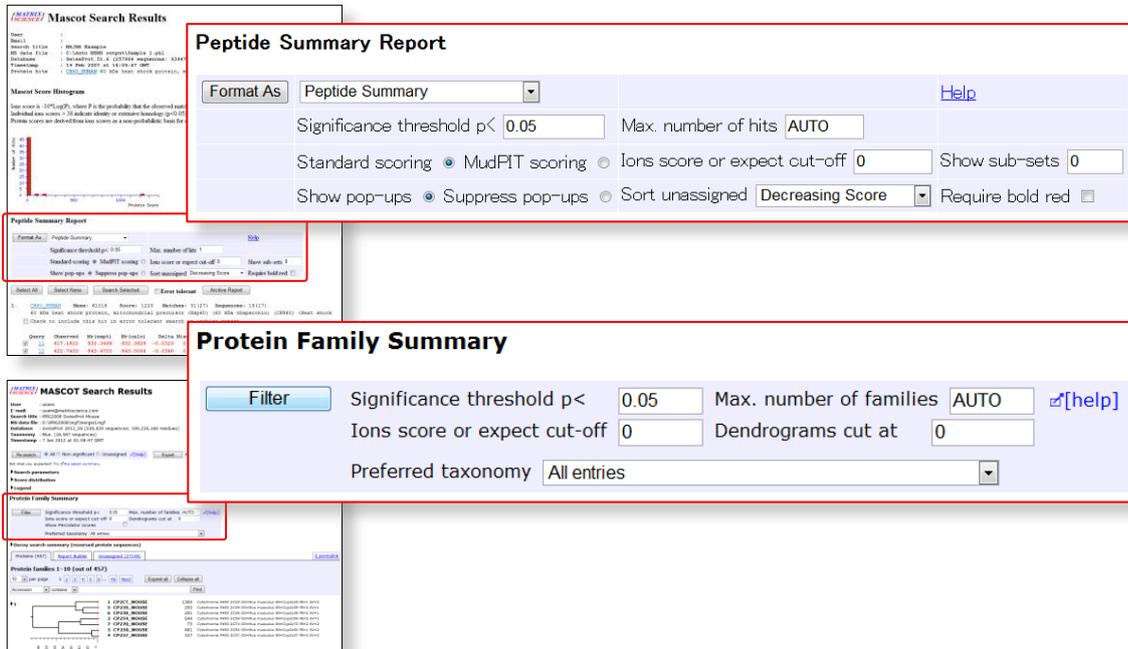
表 2 検索結果ページの種類

1. 「Summary Report」ページ（ヒットしたタンパク質をスコア順に表示します）
 - PMF 検索
 - Concise Protein Summary（最初に表示される検索結果ページ）
 - Protein Summary
 - MIS 検索
 - Protein Family Summary（MS/MS データが 300 個以上の場合に最初に表示される検索結果ページ）
 - Peptide Summary（MS/MS データが 300 個未満の場合に最初に表示される検索結果ページ）
 - SQ 検索
 - Select Summary（Peptide Summary をよりコンパクトにした検索結果ページ）
2. 「Protein View」ページ（ヒットしたタンパク質の詳細情報を個別に表示します）
3. 「Peptide View」ページ（ヒットしたペプチドの詳細情報を個別に表示します）

5-2 フォーマットコントロールパネル

図26に示すように、ヒットしたタンパク質リストを表示する「Summary Report」ページにはフォーマットコントロールパネルがあり、「Summary Report」ページの種類の切り替えや、表示される情報を調整するための条件を設定することができます。変更した条件を有効にするには[Format As]または[Filter]ボタンを押してください。

図 26 フォーマットコントロールパネル



- (1) [Format As]ボタンの右にあるプルダウンメニューから「Summary Report」の種類を選択することができます。現在の「Summary Report」の種類はプルダウンメニューのすぐ上に表示されています(図26の上段は「Peptide Summary Report」、下段は「Protein Family Report」です)。
- (2) [Significance threshold p<]で指定したプロテインスコアの期待値(0.99~1E-18の数値)よりも大きい期待値を持つタンパク質はリストから除外されます。
- (3) [Max. number of hits]で指定した数のタンパク質をリスト表示します。“AUTO”または“0”を指定した場合は、閾値スコア(MIS検索の場合は全てのペプチドのイオンスコア平均値)よりも大きなプロテインスコアを持つタンパク質リストが表示されます。
- (4) 表示されるプロテインスコアの計算方法として、[Standard scoring](帰属するペプチドのイオンスコアを積算)あるいは[MudPIT scoring](帰属するペプチドのイオンスコアのうち、閾値スコアを超過した部分のスコアを積算)を選択します。
- (5) [Ions score or expect cut-off]で指定した数値が0~1の間であれば期待値として扱い、1以上であればイオンスコアとして扱います。指定した期待値(イオンスコア)よりも大きい(小さい)ペプチドをリストから除外します。
- (6) [Show sub-sets]では0~1の数値(たとえば0.5)を指定し、代表タンパク質に帰属するペプチド数に対して、指定した割合以上の帰属ペプチド数を持つ「Sub-set」タンパク質を表示します。
- (7) タンパク質に帰属するペプチドの「クエリ番号」にマウスカーソルを載せた時に、(イオンスコアが)トップ10ペプチドのポップアップウィンドウを表示するかどうかのスイッチとして[Show pop-ups]あるいは[Suppress pop-ups]を選択します。
- (8) [Sort unassigned]では、タンパク質リストに帰属しないペプチドを並べる条件(スコア昇順あるいはスコア降順)を選択します。
- (9) [Require bold red]では、タンパク質をチェックすると、イオンスコアランクが1位のペプチド(赤文字で表示)でかつ初登場のペプチド(太文字で表示)を含むタンパク質リストを表示します。

- (10) [Dendrograms cut at]で指定したスコア以上の距離を持つタンパク質に対する樹形図が表示されます。
- (11) [Preferred taxonomy] では、タンパク質ヒットリストに優先的に表示させたい生物種名を選択してください。NCBIInrのように、ひとつのタンパク質に複数の生物種が登録されている場合、タンパク質ヒットリストにはその代表タンパク質が表示されますが、代表タンパク質ではなく、目的の生物種のタンパク質を表示させたい場合に有効です。

フォーマットコントロールに関する詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/msms_summaries_help.html#FORMAT2

http://www.matrixscience.com/help/msms_summaries_help.html#FORMAT2

5-3 タンパク質の推定

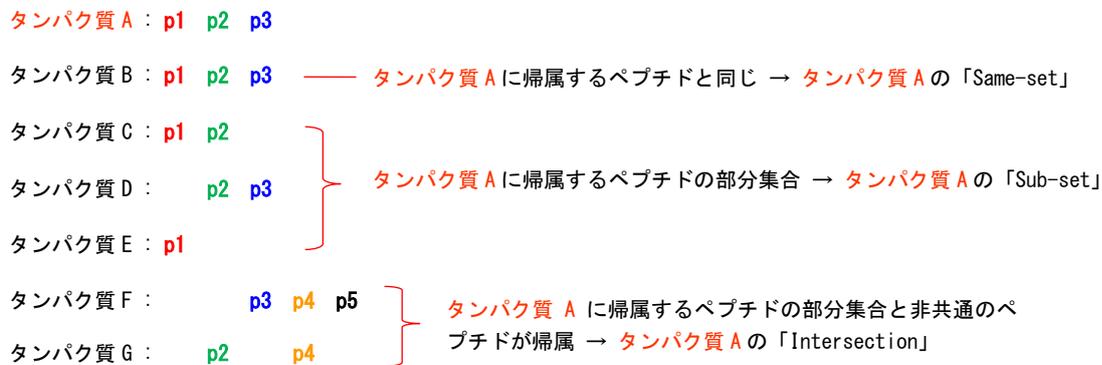
MIS検索またはSQ検索ではペプチドを同定しますので、ヒットしたペプチドが所属するタンパク質を整理することにより、実験サンプルに含まれているタンパク質を推定します。また、ヒットしたペプチドが複数のタンパク質に所属する場合は、図27に示すように、これらのタンパク質を「Same-set」、「Sub-set」、「Intersection」に分類し、検索結果ページに表示します。

図27の例では、タンパク質Aには3つのペプチド(p1、p2、p3)が所属していますが、同じペプチドが所属しているタンパク質Bはタンパク質Aの「Same-set」に分類され、(p1、p2、p3)の部分集合に当たるペプチドが所属するタンパク質C~Eはタンパク質Aの「Sub-set」に分類されます。タンパク質FとGはタンパク質Aに所属しない(非共通の)ペプチド(p4、p5)が含まれており、タンパク質Aに対する「Intersection」と呼んでいます。なお、図27に示すペプチドの所属状況から、タンパク質F以外のタンパク質については、実験サンプルに含まれているかどうかを判断することはできません。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/interpretation_help.html

http://www.matrixscience.com/help/interpretation_help.html

図27 ヒットした5種類のペプチドとタンパク質への所属例 (p1 ~ p5 はヒットしたペプチドを示します)



5-4 Concise Protein Summary (PMF)

PMF検索が終了して最初に表示されるのは図28に示す「Concise Protein Summary」ページです。ヒットしたタンパク質は質量データにマッチしたペプチドの帰属状況によりグルーピングされ、簡潔にリスト表示されます。

図 28 「Concise Protein Summary」ページ (ヒットしたタンパク質をコンパクトに表示)

The screenshot shows the Mascot Search Results page. It includes a header with the user's name (usami) and search details. A histogram shows the distribution of protein scores, with a peak around 100. Below the histogram is a table of search results, with the top hit being [OPSD_HUMAN] with a score of 102 and 11 matches. The table lists various protein entries with their mass, score, expect value, and number of matches.

Annotations in yellow boxes provide details about the page elements:

- ユーザ名、検索タイトル、データファイル名、検索に使用した配列 DB 情報、検索日時、トップスコアでヒットしたタンパク質名などの情報を表示します。** (User name, search title, data file name, search database information, search date, top score, and protein name of the hit are displayed.)
- 閾値スコア (70)、閾値期待値 (0.05)、スコア分布グラフを表示します。この例では、閾値 70 を超えたヒットが 1 件あることがわかります。** (Threshold score (70), threshold expect value (0.05), and score distribution graph are displayed. In this example, it can be seen that there is 1 hit exceeding the threshold of 70.)
- 検索結果ページ名、フォーマットコントロールパネル、再検索ボタン ([Re-Search All]、[Search Unmatched]) を表示します。** (Search result page name, format control panel, and search buttons ([Re-Search All], [Search Unmatched]) are displayed.)
- ヒットしたタンパク質を、代表タンパク質とその Same-set と Sub-set にグルーピングして表示します。タンパク質のアクセッション番号は「Protein View」ページへのリンクになっています。** (Hits are grouped by representative protein, Same-set, and Sub-set. The accession number of the protein is a link to the 'Protein View' page.)

図28の例では、[OPSD_HUMAN]が代表タンパク質であり、その質量は 39437 Da、プロテインスコアは 102、期待値は 3.3E-005、質量データにマッチした11個のペプチドが帰属しています。次の行の[OPSD_MACFA]以降は[OPSD_HUMAN]のSub-setに相当するタンパク質が表示されており、代表タンパク質の[OPSD_HUMAN]に帰属する11個のペプチドの「部分集合にあたるペプチド」が帰属しています。

代表タンパク質の表示件数は、有意にヒットしたタンパク質と有意ではない最大スコアを持つタンパク質の和ですが、フォーマットコントロールパネルの[Max. number of hits]の入力欄に表示させたいタンパク質数を入力し、[Format As]ボタンを押すことにより、最大で50件のタンパク質を表示させることができます。

[Significance threshold p<]入力欄に数値を入力し、[Format As]ボタンを押すことにより、閾値を変更することができます。1未満の数値を指定した場合は期待値として、1以上の数値を指定した場合はプロテインスコアとして処理されます。

「Concise Protein Summary」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/pmf_summaries_help.html#CONCISE

http://www.matrixscience.com/help/pmf_summaries_help.html#CONCISE

5-5 Protein Summary (PMF)

フォーマットコントロールパネルから[Protein Summary]を選択し、[Format As]ボタンを押すと、図29に示す「Protein Summary」ページに切り替わります。

図 29 「Protein Summary」ページ (ヒットしたタンパクをより詳細に表示)

The screenshot displays the Mascot Protein Summary page. It is divided into several sections:

- Index:** Lists two results:
 - OPSD_HUMAN** (Accession: 39437, Mass: 39437, Score: 102, Description: Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1)
 - OPSD_MACFA** (Accession: 39606, Mass: 39606, Score: 65, Description: Rhodopsin OS=Macaca fascicularis GN=RHO PE=2 SV=1)
- Results List:**
 - Item 1 (OPSD_HUMAN):** Mass: 39437, Score: 102, Expect: 3.3e-005, Matches: 11. It lists observed mass, theoretical mass (Mr(calc)), Delta, Start, End, Miss, and Peptide for various matches. The first match (903.3420) is highlighted in red.
 - Item 2 (OPSD_MACFA):** Mass: 39606, Score: 65, Expect: 0.15, Matches: 8. It lists similar data for the second protein.
 - No match to:** A list of masses that did not match any protein in the database.
- Search Parameters:**
 - Type of search: Peptide Mass Fingerprint
 - Enzyme: Trypsin
 - Fixed modifications: Carbamidomethyl (C)
 - Variable modifications: Oxidation (M)
 - Mass values: Monoisotopic
 - Protein Mass: Unrestricted
 - Peptide Mass Tolerance: ± 0.2 Da
 - Peptide Charge State: 1+
 - Max Missed Cleavages: 1
 - Number of queries: 19

Annotations in yellow boxes provide additional context:

- Index:** 閾値スコアを超えたタンパク質+番外 1 件を表示します。
- Results List (Item 1):** アクセッション番号、Mass(質量)、Score(プロテインスコア(閾値を超えた場合は赤色文字で表示))、Expect(期待値)、Matches(マッチしたペプチドの数)、タンパク質名などを表示します。
- Results List (No match to):** このタンパク質にマッチしなかった質量データを表示します。
- Results List (Item 2):** 質量にマッチしたペプチドの理論質量(Mr(calc))、Delta(質量誤差)、Start/End(配列の位置)、Miss(未消化サイト数)、Peptide(配列情報+修飾情報)を表示します。
- Search Parameters:** 検索条件を表示します。

「Protein Summary」ページではヒットしたタンパク質の内容をより詳しく見ることができます。ヒットした各々のタンパク質に関して、アクセッション番号、質量(Mass)、プロテインスコア(Score)、期待値(Expect)、質量データにマッチしたペプチド数(Matches)、タンパク質情報(タンパク質名、由来生物種など)、マッチした質量データの質量(Observed, Mr(expt))とその理論値(Mr(calc))およびそれらの差分(Delta)、マッチした質量データに対応するペプチドのタンパク質内における位置(Start, End)、未切断サイト数(Miss)、アミノ酸配列(Peptide)と修飾の情報を表示しています。また、[No match]の行にはこのタンパク質に対してマッチしなかった質量データを表示しています。

ヒットランク第1位の [OPSD_HUMAN] は、プロテインスコアは 102、期待値は閾値としての期待値 0.05 を大きく下回る 3.3E-005、スコア分布グラフを見るとヒットランク第2位以降のタンパク質群とはスコアで40程度の距離があり、検索に投入した18個の質量データのうちの11個の質量データに対応するペプチドが帰属していますので、このタンパク質が実験サンプルに含まれている可能性が高いことがわかります。また、実験から得られた情報(等電点や質量)と整合性がとれば、その可能性はさらに高くなります。

「Protein Summary」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/pmf_summaries_help.html#PROTSUM

http://www.matrixscience.com/help/pmf_summaries_help.html#PROTSUM

5-6 Protein View (PMF/MIS/SQ)

「Summary Report」ページに表示されているタンパク質のアクセッション番号をクリックすると図30に示す「Protein View」ページが表示され、ヒットしたタンパク質に関するより詳細な情報を見ることができます。

図30 「Protein View」ページ (タンパク質情報をより詳細に表示)

NCBI BLAST 検索サイトや NCBI Taxonomy Browser ページへのリンク、等電点 (pI) の理論値の他、検索条件、マッチしたアミノ酸のカバー率などの情報を表示します。

マッチしたペプチドは赤色文字で表示されます。また、「Unformatted sequence string:」の右にあるアミノ酸残基数リンクをクリックするとこのエントリ情報をテキストで表示します。

マッチしたペプチドの Start/End (配列の位置)、Delta (質量誤差)、Miss (未消化サイト数)、Peptide (配列情報+修飾情報) を表示します。なお、[Sort peptide by] ボタンで昇順または降順に並べ替えることができます。

マッチした質量の実験値と理論値の差をプロットしています。質量分析計のキャリブレーションカーブに対応します。なお、「RMS error」の値は標準偏差を示しています。

SwissProt などの配列データベースによってはタンパク質の詳細情報を表示が表示されます (設定が必要です)。

「Protein View」ページでは、そのタンパク質の詳細情報(全体アミノ酸配列とマッチした質量データがカバーしているアミノ酸配列部分、等電点理論値、アノテーション情報、NCBIサイトへのリンクなど)を見ることができます。また、質量データ(実験値)とその理論値との質量差をプロットした質量誤差グラフは、質量分析計固有のキャリブレーションカーブ(質量校正曲線)に対応します。質量誤差グラフと実際のキャリブレーションカーブが異なる場合は、このタンパク質は擬陽性(False Positive なヒット)の可能性があるので、質量データや検索条件を吟味する必要があります。

「Protein Summary」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/results_help.html#PROT

http://www.matrixscience.com/help/results_help.html#PROT

5-7 Protein Family Summary (MIS/SQ)

Mascot検索に投入したMS/MSの質量スペクトルデータ数が300以上の場合は、図31に示す「Protein Family Summary」ページが最初に表示されます。

図 31 「Protein Family Summary」ページ

MASCOT Search Results

User :
E-mail :
Search title : IPRG2008 SwissProt Mouse
MS data file : D:\IPRG2008\mgf\merged.mgf
Databases : 1: cRAP 20090731 (111 sequences; 37,180 residues)
2: SwissProt 57.14 (514,789 sequences; 181,163,771 residues)
Taxonomy : 1: (none)
2: Mus. (16,273 sequences)
Timestamp : 5 Mar 2010 at 14:04:48 GMT
Warning : No taxonomy indexes for cRAP, taxonomy 'Mus.' ignored. Searching all entries in cRAP

Re-search | All | Non-significant | Unassigned | Export | As | XML

Search parameters
Score distribution
Legend

Protein Family Summary

Filter: Significance threshold p<c 0.05 Max. number of families AUTO
Ions score or expect cut-off 0 Dendrograms cut at 0
Show Percolator scores
Preferred taxonomy All entries

Decoy search summary (random protein sequences)
Proteins (482) | Report Builder | Unassigned (26931)

Protein families 1-10 (out of 482)
10 per page | 1 2 3 4 5 6 ... 49 | Next | Expand all | Collapse all

Accession | contains | Find

1: TRY1_BOVIN 1597 TRY1_BOVIN
2: CP2CT_MOUSE 1307 Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
3: CP239_MOUSE 293 Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
4: CP238_MOUSE 202 Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
5: CP254_MOUSE 535 Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
6: CP270_MOUSE 69 Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
7: CY250_MOUSE 382 Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1

2: GRP78_MOUSE 1292 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa8 PE=1 SV=3
3: HSP7C_MOUSE 353 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
4: HSP7L_MOUSE 165 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1 PE=2 SV=4

2: CYB5_MOUSE 1228 Cytochrome b5 OS=Mus musculus GN=Cy5a PE=1 SV=2

4.1 2: CYB5_MOUSE 1228 16817 63 (46) 5.00 Cytochrome b5 OS=Mus musculus GN=Cy5a PE=1 SV=2

63 peptide matches (13 non-duplicate, 50 duplicate)
Auto-fit to window

Query Dupes	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
40493	513.8603	1025.7060	1025.6001	0.1059	36	0.018	1	U	K.VYDLTK.F
4438	520.9165	1039.8184	1039.5212	0.2972	3	1.4	4	U	-MAGQSDK.D + Oxidation (M)
15915	704.4691	1406.9236	1407.8230	-0.8994	11	1.5	5	U	K.STWIISEK.V
15982	470.4068	1408.1987	1407.8230	0.3756	9	0.0073	1	U	K.STWIISEK.V

Score > 32 indicates identity

Query Dupes	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
18020	738.0084	1474.0022	1473.7959	0.2064	74	3.5e-006	1	U	K.YTLLLEIQK.H
18026	492.3889	1474.1449	1473.7959	0.3491	6	2	9	U	K.YTLLLEIQK.H
18994	755.5320	1509.0494	1509.8186	-0.7692	3	1	9	U	-MAGQSDK.V.Y
15552	787.0462	1572.0719	1571.8065	0.2654	12	5.9e-007	1	U	K.YTIIISLHSDK.S
15643	525.0961	1572.2366	1571.8065	0.4301	57	4.6e-006	1	U	K.YTIIISLHSDK.S
22351	552.4631	1694.9676	1694.8437	0.1240	46	0.00035	1	U	K.FLSEHFGSEVLR.E
22386	828.5314	1655.0483	1654.8437	0.2047	62	4.0e-005	1	U	K.FLSEHFGSEVLR.E
30413	1175.6897	2349.3647	2349.0227	0.3420	100	5e-009	1	U	R.EQAQGDATERFESVQVSTDAK.E
30415	784.1341	2349.3805	2349.0227	0.3578	104	1.7e-009	1	U	R.EQAQGDATERFESVQVSTDAK.E

1 subset or intersection (1 subset protein in total)

5: PD1A1_MOUSE 1116 Protein disulfide-isomerase OS=Mus musculus GN=Pdi4b PE=1 SV=1
2: TXND3_MOUSE 46 Thioredoxin domain-containing protein 3 OS=Mus musculus GN=Txndc3 PE=1 SV=2

6: CP1A2_MOUSE 1048 Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1

7: RDH7_MOUSE 1023 Retinol dehydrogenase 7 OS=Mus musculus GN=Rdh7 PE=2 SV=1
2: H1706_MOUSE 612 Hydroxysteroid 17-beta dehydrogenase 6 OS=Mus musculus GN=Hsd17b6 PE=2 SV=1

8: ENPL_MOUSE 1015 Endoplasmic reticulum protein OS=Mus musculus GN=Hsp90b1 PE=1 SV=2

9: MGST1_MOUSE 833 Mitochondrial glutathione S-transferase OS=Mus musculus GN=Mgst1 PE=1 SV=1

10: RL7A_MOUSE 771 60S ribosomal protein L7a OS=Mus musculus GN=Rpl7a PE=1 SV=1

10 per page | 1 2 3 4 5 6 ... 49 | Next | Expand all | Collapse all

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

ユーザ名、検索タイトル、データファイル名、検索に使用した配列 DB 情報、検索日時などの情報を表示します。

再検索 ([Re-search] ボタン)、検索結果のファイル出力 ([Export] ボタン)、「Select Summary」ページへの移動、検索条件やスコア分布グラフの表示 (▶アイコンをクリック)、閾値スコアや表示グループ数などの設定、Decoy 検索結果の表示などを行うことができます。

タブを選択して表示を切り替えることができます。

検索結果全体からキーワード検索することができます。

共通するペプチドを持つタンパク質の類似度を示す樹形図、アクセス番号、プロテインスコア、アノテーションを表示します。

▶アイコンをクリックすることにより、Same-set および Sub-set タンパク質、タンパク質に属するペプチド情報、トップ 10 ペプチドなどの情報を表示します。なお、イオンスコアランク 1 位のペプチドは赤色文字で、イオンスコアランク 2 位以下のペプチドは黒色文字で、期待値が 0.05 よりも小さいペプチドは太文字で表示されます。

1 ページ当たりのタンパク質ファミリーの表示数の指定、表示したいページ番号の指定、表示情報の展開・格納を行うことができます。

「Protein Family Summary」ページでは、ヒットしたタンパク質はプロテインスコア順に表示されますが、有意にヒットしたペプチドを共通項として持つ類似タンパク質が存在する場合はそれらをひとつのグループ (Protein Family) にまとめ、互いの類似度を示す樹形図とともに表示されます。また、検索結果をブラウザの 1 ~ 2 ページ分の量になるようにコンパクトにまとめ、より詳細な情報は対応する項目名の前にある ▶アイコンをクリックして展開し、必要な時に表示するようにしているため、大規

模な質量データに対する検索結果もスピーディに表示させることができます。なお、初めての表示の際に表示内容をインデックス化するため、その処理に若干時間を要しますが、2回目以降は即座に表示されます。

なお、図31に示した「Protein Family Summary」ページの例は次のURLにアクセスし、ご覧ください。

http://ホスト名/mascot/cgi/master_results_2.pl?file=F981139.dat

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat

有意にヒットしたペプチドを共通項として持つ類似タンパク質は次の操作を通じて収集し、Protein Familyとしてグルーピングします。ひとつのProtein Familyに属するタンパク質メンバーの類似度は、非共通のペプチドのスコアを使ってメンバー間の距離を定義・計算し、階層的クラスタ処理を行って求めています。

- (1) プロテインスコア順のタンパク質リストを作成する。
- (2) (1)のリストから、最大プロテインスコアのタンパク質を選択し、このタンパク質に帰属するHomology閾値スコア以上のペプチドを抽出する。
- (3) (2)で抽出したペプチドを含むタンパク質を(1)のリストから抜き出して「Protein Family」のメンバーとし、それらを(1)のリストから除く。
- (4) (3)のタンパク質に帰属するHomology閾値スコア以上のペプチドを抽出し、(1)のリストからこれらのペプチドを含むタンパク質を抜き出して「Protein Family」のメンバーとし、それらを(1)のリストから除く。
- (5) 抽出できるタンパク質が無くなるまでこの操作を繰り返す。
- (6) 新たな「Protein Family」を作成するために同様な操作を行い、Homology閾値スコア以上のペプチドが無くなるまで繰り返す。
- (7) 「Protein Family」のメンバーとして抽出したタンパク質のペプチドをアミノ酸配列が重複しないように整理する(修飾、電荷、最大スコアによる重複を除く)。
- (8) (7)のタンパク質をSame-setとその他(Sub-set、Intersection)に分ける。
- (9) タンパク質メンバー間の距離(非共通ペプチドに関して、Homology/Identityの小さい方の閾値スコア超過分の積算)を求め、階層的クラスタ処理を行う。

タンパク質間の距離は、図32に示したように、プロテインスコアが大きい方のタンパク質(タンパク質F)から見て、プロテインスコアが小さい方のタンパク質(タンパク質G)に存在する非共通なペプチド(p2)を選択し、そのスコアの閾値スコア(Homology閾値とIdentity閾値の小さい方のスコア、この例では31)からの超過分スコアを積算して求めています。この例では対応するペプチドは p2 のひとつ

図 32 タンパク質間の距離の求め方

タンパク質F : p3 p4 p5

タンパク質G : p2 p4

プロテインスコアF > Gとすると p2 の閾値スコア超過分が F と G の距離

Query	Dupes	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	1	2	Peptide
113.1	2	488.2970	974.5794	974.5518	0.0276	34	0.03	1	U			R.SVDRIR.L
113.2	2	605.0467	1208.0787	1207.7056	0.3731	57	0.0002	1	U			R.GLFIDDK.G
	1	605.0112	1208.0078	1207.7056	0.3022	(50)	0.0011	1	U			R.GLFIDDK.G
		590.4919	1768.4540	1767.8550	0.5990	71	1.6e-006	1	U			K.DYGVYLEDSSQTLR.L

“Score > 31 indicates identity” → 34-31=3 → クラスタリング

とつでするので、タンパク質FとGの距離は 34-31=3 になります。このようにして求めたタンパク質間距離を使って階層的クラスタ処理を行い、その結果をタンパク質の類似度を示す樹形図として表示しています。

「Protein Family Summary」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/help/msms_summaries_help.html#FAMILY

http://www.matrixscience.com/help/msms_summaries_help.html#FAMILY

5-8 Peptide Summary (MIS/SQ)

Mascot検索に投入したMS/MSの質量スペクトルデータ数が300未満の時は「Peptide Summary」ページが最初に表示されます(図33の例を表示するには次のURLにアクセスし、フォーマットコントロールパネルで[**Ions score or expect cut-off**]の値を 0.05 に、[**Max number of hits**]の値を 2 にセットし、[**Format As**]ボタンを押してください)。ヒットしたタンパク質はプロテインスコア順に表示され、タンパク質に帰属するペプチドはMS/MSの質量スペクトルデータ情報とともに表示されます。

http://ホスト名/mascot/cgi/master_results.pl?file=../data/F981123.dat

http://www.matrixscience.com/cgi/master_results.pl?file=../data/F981123.dat

図 33 「Peptide Summary」ページ

Mascot Search Results

User: []
 Search title: MS/MS Example
 MS data file: C:\Users\MSM\output\Sample 1.pkl
 Database: SwissProt 51.6 (27954 sequences; 93847433 residues)
 Timestamp: 19 Feb 2007 at 14:09:47 GMT
 Protein hits: CH60_HUMAN 60 kDa heat shock protein, mitochondrial precursor (P9660) (60 kDa chaperonin) (C9960) (Heat shock protein 60 kDa class B group 1)
 CH60_DROME 60 kDa heat shock protein, mitochondrial precursor (P9660) (60 kDa chaperonin) (C9960) (Heat shock protein 60 kDa class B group 1)

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Individual ions scores > 38 indicate identity or extensive homology (p < 0.05).
 Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Peptide Summary Report

Format As: Peptide Summary
 Significance threshold p < 0.05 Max. number of hits 2
 Standard scoring # Mascot scoring Ions score or expect cut-off 0.05 Show sub-sets 0
 Show pop-ups # Suppress pop-ups Sort unassigned Decreasing Score Require bold red

1. CH60_HUMAN Mass: 41014 Score: 1176 Matches: 27(27) Sequences: 17(17)
 60 kDa heat shock protein, mitochondrial precursor (P9660) (60 kDa chaperonin) (C9960) (Heat shock protein 60 kDa class B group 1)
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calcd)	Delta Miss Score	Expect	Rank	Uniqs	Peptide
11	417.1822	832.3498	832.3828	-0.0329	0	45	0.016	R.AKFGSDDE.E
12	422.7433	841.6300	841.6066	-0.0266	0	46	0.013	R.VKDFQVDE.D
13	451.2499	900.4853	900.5280	-0.0428	0	52	0.0039	K.LADQVALE.V
14	456.7806	911.5467	911.5904	-0.0337	0	59	0.00056	R.VKQVQVVE.A
15	460.7447	920.4768	920.5036	-0.0288	0	49	0.013	R.VTDAIAATE.A
16	595.7855	1189.5565	1189.6012	-0.0447	0	57	0.0013	R.EKSIIDAMG.E
17	603.7720	1205.5294	1205.5562	-0.0268	0	60	0.00049	R.RKLIISDAG.E + Oxidation (M)
18	608.2059	1214.4850	1214.4857	-0.0045	0	78	2.4e-005	R.RKQVQVDE.D
19	617.2857	1232.5560	1232.5885	-0.0316	0	81	4e-006	R.VKQVQVDE.D
20	672.8370	1343.6605	1343.7085	-0.0480	0	64	0.00018	R.VTVEIQKQVVE.V
21	714.8884	1427.7423	1427.8028	-0.0435	0	62	0.00014	R.QNGLADVALE.E
22	714.8888	1427.7720	1427.8028	-0.0327	0	170	2.1e-005	R.QNGLADVALE.E
23	722.8849	1443.7532	1443.8007	-0.0455	0	75	1.2e-005	R.QNGLADVALE.E + Oxidation (M)
24	722.8834	1443.7722	1443.8007	-0.0285	0	170	2.2e-005	R.QNGLADVALE.E + Oxidation (M)
25	722.8843	1443.7142	1443.7490	-0.0348	0	50	4.2e-007	R.VTDELEISDAG.E
26	740.8441	1513.6777	1513.7439	-0.0662	0	89	4.7e-007	K.TLDELEISDAG.E + Oxidation (M)
27	840.3291	1917.9420	1918.0436	-0.1010	0	102	2.1e-008	R.EKSIIDVALEISDAG.E
28	940.0327	1943.0569	1943.0566	-0.0027	0	493	3.3e-007	R.EKSIIDVALEISDAG.E
29	1019.5104	2037.0047	2037.0150	-0.0087	0	52	0.0015	R.LQRIISDQVVEDEK.E
30	1057.0337	2112.0929	2112.1323	-0.0394	0	116	6.9e-010	R.ALMSQVQLADVALEISDAG.E
31	1085.0399	2198.0433	2198.1370	-0.0919	0	170	1.7e-005	R.ALMSQVQLADVALEISDAG.E + Oxidation (M)
32	1078.0417	2184.0809	2184.1221	-0.0412	0	490	3.3e-007	R.ALMSQVQLADVALEISDAG.E + 2 Oxidation (M)
33	789.1040	2364.2948	2364.3264	-0.0296	0	56	0.00025	R.RPVLVILAEVDQVRLATVFLVM.L
34	1183.1570	2364.2994	2364.3264	-0.0270	0	60	5.3e-005	R.RPVLVILAEVDQVRLATVFLVM.L
35	789.1034	2364.3063	2364.3264	-0.0261	0	95	6.3e-008	R.RPVLVILAEVDQVRLATVFLVM.L
36	808.1322	2481.3748	2481.3942	-0.0194	0	48	0.0003	R.TALLDAAGVALLTAVVPTVVE.E
37	854.0588	2559.1545	2559.2413	-0.0868	0	75	5.1e-006	K.LVQVQVNTREAGDQVTTAVLAD.

2. CH60_DROME Mass: 40771 Score: 162 Matches: 3(3) Sequences: 3(3)
 60 kDa heat shock protein, mitochondrial precursor (P9660) (60 kDa chaperonin) (C9960) (Heat shock protein 60 kDa class B group 1)
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calcd)	Delta Miss Score	Expect	Rank	Uniqs	Peptide
11	417.1822	832.3499	832.3829	-0.0329	0	45	0.016	R.AKFGSDDE.E
12	417.2857	1232.5569	1232.5885	-0.0316	0	42	0.032	U R.VGSDSEVDEK.E
13	834.0588	2559.1548	2559.2413	-0.0868	0	75	5.1e-006	R.LVQVQVNTREAGDQVTTAVLAD.

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

Query	Observed	Mr(expt)	Mr(calcd)	Delta Miss Score	Expect	Rank	Uniqs	Peptide
11	430.7328	859.4510	859.4897	-0.0327	0	36	0.15	IFAEIAR.E + Oxidation (M)
12	714.3849	1426.7153	1426.8143	-0.0990	1	31	0.32	KAKFLKDELLE.L
13	828.1238	2481.3840	2481.4047	-0.0247	0	24	0.43	TALLDAAGVALLTAVVPTVVE.E
14	1065.0523	2120.1100	2120.1372	-0.0172	0	26	0.63	R.ALMSQVQLADVALEISDAG.E + Oxidation (M)
15	445.2283	892.4421	892.4538	-0.1118	0	25	1.2	LADQVLE
16	745.2842	1461.2869	1461.3071	-0.0218	1	15	24	SDSTRA
17	662.2756	661.2883	661.3261	-0.0388	1	14	20	ENNAHE
18	1038.5031	3112.4873	3112.5023	-0.0150	0	13	6.7	RGALTAGVDFPDESLTDLRDEVDQVDEK.E + Oxidation (M)
19	1101.6217	3106.4144	3106.4032	0.0132	0	12	31	QLADVALE
20	660.4172	661.4059	661.3237	0.0802	0	12	23	ALAGEE
21	714.3725	713.3652	713.4072	-0.0419	0	11	43	LADQVE
22	673.2495	670.2422	672.3555	-0.0133	0	11	71	AKVDVQ
23	1101.5366	3100.5093	3100.5349	-0.0256	0	9	68	DEFTDSEK
24	1099.0947	2196.1749	2196.1862	-0.0113	1	8	35	LRASVETLLSAGQVDEK.E
25	645.3526	1280.4926	1280.7397	-0.0431	0	8	67	VVPPVQDQALSR
26	642.3526	1280.4926	1280.4630	0.0297	1	7	78	DEFTDSEK
27	963.8379	1325.6612	1325.7667	-0.1055	1	7	70	TFLVQVQVDEK.E
28	1048.5615	2095.1085	2095.0659	0.0426	1	6	70	ALREILKQVPPVQVDEK.E
29	747.0261	2238.0884	2238.0773	0.0091	1	5	77	VNSAFATLLSAGQVDEK.E
30	1020.9179	2029.9613	2029.9170	-0.0508	1	4	1.1e+002	RGLSADVALEISDAG.E + Oxidation (M)
31	932.3444	931.3571	931.4658	-0.1087	1	4	2e+002	TDSTSEK
32	500.2580	499.2487	499.3005	-0.0519	0	4	1.2e+002	LADQV
33	745.3840	1456.7034	1456.6507	0.1207	0	3	1.6e+002	R.LVQVQVNTREAGDQVTTAVLAD.E + 2 Oxidation (M)
34	933.4390	932.4917	932.4498	0.0419	1	3	2.7e+002	SDRQVVE + Oxidation (M)
35	884.4009	1770.7792	1770.8924	-0.0962	1	2	2e+002	DEVALDQVQVADQVDEK.E
36	747.4125	746.4052	746.3623	0.0419	0	2	1.6e+002	SDQVVE + Oxidation (M)
37	930.7030	929.4907	929.5284	0.1463	1	2	2.2e+002	LQVQVDEK.E
38	711.3647	710.3774	710.3711	-0.0127	0	1	2.3e+002	QSDSEK
39	930.4821	929.4908	929.5243	0.1213	1	1	3.4e+002	DEFTDSEK
40	949.5507	1897.0868	1897.0343	0.0502	1	0	2.3e+002	LLEKQVPPVQVDEK.E

Search Parameters

Type of search: MS/MS Ion Search
 Enzyme: Trypsin
 Variable modifications: Oxidation (M)
 Mass values: Monoisotopic
 Protein Mass: Unrestricted
 Peptide Mass Tolerance: 0.2 Da
 Fragment Mass Tolerance: 0.2 Da
 Max Missed Cleavages: 1
 Instrument type: ESI-QQQ-TOP
 Number of queries: 67

ユーザ名、検索タイトル、データファイル名、検索に使用した配列 DB 情報、検索日時、有意なプロテインスコアでヒットしたタンパク質名などの情報を表示します。

イオンスコア閾値(この例では 38)、ヒットしたタンパク質のスコア分布グラフを表示します。

フォーマットコントロールパネルでは「Summary Report」ページの種類の切り替えや、表示される情報量を調整するための条件を設定することができます。[Search Selected]ボタンで再検索を行うことができます。

ヒットしたタンパク質をタンパク質名などの基本情報とそれに帰属するペプチドとともにプロテインスコア順に表示します。クエリ番号にマウスカーソルを合わせるとマッチしたイオンスコア上位 10 件のペプチドが表示されます(下表)。

Top scoring peptide matches to query 11
 Score greater than 27 indicates identity
 Score greater than 40 indicates homology

Score	Expect	Delta	Hit	Protein	Peptide
45.4	0.016	-0.0329	1+	CH60_HUMAN	APFGDNR
11.3	42	0.0001		K.APFMNDR.R	
10.4	52	0.0001		K.ANAPSCDR.L	
10.1	55	-0.0727		K.AFGSSRMK.T	
8.8	74	-0.0801		K.APVLIMR.M	
8.8	74	-0.0801		K.APVLIMR.D	
7.4	1e+002	-0.0540		K.SASMIADR.R	
7.4	1e+002	-0.0177		K.AVGDSDNR.G	
7.2	1e+002	-0.0541		R.ITSPGTGGR.L	
6.9	1.2e+002	-0.0727		K.ATVACGGR.L	

ヒットリスト表示されていないタンパク質(この例では「CH60_HUMAN」と「CH60_DROME」)に帰属していないペプチドをリスト表示します。

この検索で使用した検索条件を表示します。

「Peptide Summary」ページにはヒットしたタンパク質各々について次の内容が記載されています。図33の例では、フォーマットコントロールパネルで設定した条件に対して2つのタンパク質(「CH60_HUMAN」と「CH60_DROME」)がヒットしていることを示しており、たとえば「CH60_HUMAN」については次の情報が表示されています。

ヒット番号「1.」、タンパク質のアクセッション番号「CH60_HUMAN」、質量「Mass: 61016」、プロテインスコア「Score: 1176」、ペプチドにマッチしたMS/MSの質量スペクトルデータ数(カッコ内の数はイオンスコアの閾値を超えたMS/MSの質量スペクトルデータ数)「Matches: 27 (27)」、ヒットしたペプチド数(カッコ内の数は閾値を超えたペプチド数)「Sequences: 17 (17)」、アノテーション「60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock)」

また、ヒットしたタンパク質に帰属するペプチドは次の項目とともにリスト表示されます。

Query : MS/MSの質量スペクトルデータのクエリ番号(「Peptide View」ページにリンクしています)
 Observed : プリカーサイオン質量の実験値
 Mr (expt) : プリカーサイオンの中性質量(実験値)
 Mr (calc) : プリカーサイオンの中性質量(理論値)
 Delta : プリカーサイオン質量の実験値と理論値の差
 Miss : 未切断サイト数
 Score : イオンスコア(カッコ付のものは、他により高いイオンスコアを持つペプチドが存在することを示しています)
 Expect : 期待値
 Rank : スコアランク(「4-2 Identity閾値とHomology閾値」をご覧ください)
 Unique : このタンパク質にのみ存在するUniqueなペプチドの場合は **U** を記載
 Peptide : ペプチドのアミノ酸配列

なお、リストされているペプチドは文字種および文字色によって次の意味を持っています。

- (1) **赤色文字** : イオンスコアランク 1 位のペプチド
- (2) **黒色文字** : イオンスコアランク 2 位以下のペプチド
- (3) **太文字** : ペプチドリストに初めて登場したクエリ
- (4) **細文字** : すでにペプチドリストに登場しているクエリ

たとえばクエリ番号**11**の例では、ヒット番号1の「CH60_HUMAN」に帰属するペプチド **K. APGFGDNR. K** にマッチしたクエリとして初登場しており、また、この配列は「**11**」の上にマウスカーソルを合わせて表示されるイオンスコアランクが第1位ですので太文字の赤色文字で表示されていますが、ヒット番号2の「CH60_DROME」においては、すでにヒット番号1の「CH60_HUMAN」のところで登場しているクエリになりますので、細文字の赤色文字で表示されています。

クエリ番号**27**の例では、ヒット番号1の「CH60_HUMAN」においては **K. VGGTSDVEVNEK. K** のペプチドにマッチし、初登場クエリおよびイオンスコアランク1位のペプチドですので太い赤色文字で表示されていますが、ヒット番号2の「CH60_DROME」においては、イオンスコアランク第2位の **R. VGGSSSEVNEK. K** にマッチしていますので、細い黒色文字で表示されています。

太い赤色文字のペプチド(初登場かつイオンスコアランク第1位)を含む、同定された可能性がより高いタンパク質だけを表示させたい場合は、コントロールパネルの[Require bold red]をチェックし、[Format As]ボタンを押してください。

「Peptide matches not assigned to protein hits: (no details means no match)」のブロックには、ヒットしたタンパク質(この例では「CH60_HUMAN」および「CH60_DROME」)に帰属しないペプチドをリスト表示しています。

「Peptide Summary」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/msms_summaries_help.html#PEPSUM

http://www.matrixscience.com/help/msms_summaries_help.html#PEPSUM

5-10 Peptide View (MIS/SQ)

「Summary Report」ページのクエリ番号あるいは「Protein View」ページのイオンスコアは図35に示す「Peptide View」ページにリンクしています。

図 35 「Peptide View」ページ

(MATRIX) Mascot Search Results

Peptide View

MS/MS Fragmentation of **TVIEQSWGSPK**
 Found in **CH60_HUMAN** in SwissProt, 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock protein 60 kDa class B member 1) (HSPD6)

Match to Query 31: 1343.660496 from(672.8375242*)
 Data file C:\Auto-MSMS-output\Sample 1.pld

Click mouse within plot area to zoom in by factor of two about that point
 Or, Plot from 0 to 2000 Da Full range

Label all possible matches Label matches used for scoring

Show Y-axis

Monoisotopic mass of neutral peptide Mr(calc): 1343.7085
 Ions Score: 64 Expect: 0.00016
 Matches: 25/115 Fragment ions using 52 most intense peaks (help)

#	b	b ⁺	b*	b ⁺ *	b ⁰	b ⁰⁺	Seq.	y	y ⁺	y*	y ⁺ *	y ⁰	y ⁰⁺	#
1	102.0550	51.5311			84.0444	42.5258	T	1243.6681	622.3377	1226.6416	613.8244	1225.6575	613.3324	12
2	201.1234	101.0653			183.1128	92.0600	V	1144.5997	572.8035	1127.5732	564.2902	1126.5891	563.7982	10
3	314.2074	157.6074			296.1969	148.6021	I	1031.5156	516.2615	1014.4891	507.7482	1013.5051	507.2562	9
4	427.2915	214.1494			409.2809	205.1441	E	918.4336	459.7194	901.4050	451.2061	900.4210	450.7141	8
5	536.3341	278.6707			538.3235	269.6654	E	789.2899	395.1981	772.3624	386.6849	771.3784	386.1928	7
6	684.3927	342.7000	667.3661	334.1867	666.3821	333.6947	Q	661.2304	331.1688	644.3039	322.6556	643.3198	322.1636	6
7	771.4247	386.2160	754.3981	377.7027	753.4141	377.2107	S	574.2984	287.6528	557.2718	279.1396	556.2878	278.6475	5
8	957.5040	479.2556	940.4775	470.7424	939.4934	470.2504	W	388.2191	194.6132	371.1925	186.0999	370.2085	185.6079	4
9	1014.5255	507.7664	997.4989	499.2531	996.5149	498.7611	G	244.1656	122.5864	227.1390	114.0731	226.1530	113.0863	3
10	1101.5575	551.2824	1084.5310	542.7691	1083.5469	542.2771	S	147.1128	74.0600	130.0863	65.5468	129.0600	64.0000	2
11	1198.6103	599.8088	1181.5837	591.2955	1180.5997	590.8035	P							1
12							K							1

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

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
64.4	1343.7085	-0.0480	TVIEQSWGSPK	
23.0	1343.6681	-0.0076	VVDLQESSQPSR	Phospho S4 84.56%
22.8	1343.8176	-0.1571	SHIKQALPYAK	Phospho S6 14.73%
15.7	1343.6681	-0.0076	KALNENDGDIQK	
15.4	1343.7409	-0.0804	TVLQEGNSQK	
15.2	1343.7449	-0.0844	TVLWKVDDAVAK	
14.0	1343.6932	-0.0327	VDVIQKAEADK	
13.5	1343.7296	-0.0691	VTDGETLQGIQK	
13.0	1343.7409	-0.0804	TVLVESSGDDGIRK	
12.1	1343.7045	-0.0440	VATIGVQQSINAK	

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

MS/MSの質量スペクトルデータ情報、ヒットしたペプチドとそれが所属するタンパク質情報を表示しています。

MS/MSの質量スペクトルのどのピークにどのプロダクトイオンが対応しているかを確認することができます。グラフ内でクリックするとm/z範囲が拡大されます(拡大表示させたいm/z領域を指定することもできます)。グラフ及びプロダクトイオン表にはイオンスコアに貢献しているプロダクトイオンが表示されていますが、[Label all possible matches]を選択すると、数値的に可能な全てのマッチングが表示されます。[Show Y-axis]をチェックすると、信号強度を示すy軸が表示されます。質量誤差グラフ(実験値と理論値の差)はキャリブレーションカーブに対応します。

NCBIのblast検索ページにリンクしています。[Web gateway]リンクをクリックすると他の検索サイトへのリンクページが表示されます。

MS/MSの質量スペクトルデータにマッチした、イオンスコア上位10個のペプチドを表示しています。アミノ酸配列部分をクリックすると、「Peptide View」ページはそのペプチドに関する内容に書き換えられます。もし、ペプチドが修飾されており、複数の修飾サイトが存在する場合は、図12及び下に示すように[Site Analysis]項が加わり、どの修飾サイトが正しいかを確率的な数値で表示します。

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	KMEDSVGCLTAEVVK	
15.5	1846.9230	-0.0161	GAYTIQHPVLGLEIK	
14.3	1846.7735	0.0355	INKRLNLIKRPQHSDDPAQIMGYKNN	

MD-scoreは7、20となるため、S4リン酸化ペプチドが最も可能性が高く、約84%の確率で正しいことを示しています。「Sequence」項のリンクをクリックするとプロダクトイオンスペクトルのピークに対するイオンシリーの帰属状況の違いを確認することができます。

Score	Mr(calc)	Delta	Sequence	Site Analysis
84.5	3541.7900	0.0191	KRYGASAGNVGDEGGVAPNIQTAEFALDLIVDAIK	Deamidated N9 99.79%
57.2	3541.7900	0.0191	KRYGASAGNVGDEGGVAPNIQTAEFALDLIVDAIK	Deamidated N19 0.19%
47.9	3541.7900	0.0191	KRYGASAGNVGDEGGVAPNIQTAEFALDLIVDAIK	Deamidated Q21 0.02%
14.3	3541.7735	0.0355	INKRLNLIKRPQHSDDPAQIMGYKNN	
14.3	3541.7735	0.0355	INKRLNLIKRPQHSDDPAQIMGYKNN	
13.5	3541.7470	0.0620	ENEVDRKRVNVEDEMTKLRVIMQNILKNN	

MD-scoreは27、36と大きいため、N9サイトの脱アミド化ペプチドが100%に近い確率で正しいことを示しています。

Mascot Serverは、プリカーサイオン質量にマッチしたペプチドを対象として、より強度の高いピークとプロダクトイオンをマッチングさせながら、最もイオンスコアが高くなるピークの組合せ検索し、イオンスコア順に上位10件のペプチドを検索結果として残します。

プロダクトイオン表には、イオンスコア計算に使用されたプロダクトイオンは赤色文字で表示されていますが、**赤色斜体太文字**はイオンスコアに大きく貢献しているマッチング、**赤色太文字**は偶然ではないマッチング、**赤色細文字**は偶然の可能性のあるマッチングを示しています。

イオンスコア上位10件のペプチドは「Peptide View」ページ下方に表形式で表示されています。図35に示されている例では、スコアランク1位の TVIIEQSWGSPK のイオンスコアは64.4、2位の VVDLQESSQPSR のイオンスコアは23.0であり、その差は41ですので、マッチングの確率としては約10000倍の差があることになります。この例のように、一般的には、ランク1位と2位以降のイオンスコアの差は非常に大きく、実在するMS/MSの質量スペクトルが高いイオンスコアでマッチするペプチドは一つであり、その他の低いイオンスコアのペプチドに対しては偶然にマッチしていることがわかります。

ペプチドが修飾されており、修飾サイトが複数存在する場合は、どの修飾サイトが正しいのかを、Mascot Delta score (当該ペプチドのイオンスコアの差、MD-scoreと呼んでいます)から計算される確率を「Site Analysis」項に表示しています。デフォルト設定では、MD-scoreが10の場合は91%の確率に対応します。図35下方にある2つの表では、リン酸化と脱アミド化の例を示しています。リン酸化の例では、S4サイトのリン酸化ペプチドが84%の確率で正しいことを示しています。また、脱アミド化の例では、N9サイトの脱アミド化ペプチドがほぼ100%の確率で正しいことを示しています。

「Peptide View」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/msms_summaries_help.html#PEPSUM
http://www.matrixscience.com/help/msms_summaries_help.html#PEPSUM

「Site Analysis」に関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/pt_mods_help.html#SITE
http://www.matrixscience.com/help/pt_mods_help.html#SITE

5-11 検索結果の出力 (PMF/MIS/SQ)

フォーマットコントロールパネルから [Export Search Results] を選択し、[Format As] ボタンを押すと図 36 に示す「Export search results」ページが表示されます。

検索結果の内容を様々な書式 (XML、CSV、pepXML、mzIdentML、DTASelect、Mascot DAT File、MGF Peak List) でファイルに出力することができます。たとえば、[Export format] から「CSV」を選択し、出力したい項目をチェックした後、ページ最下方の [Export search results] ボタンを押してください。検索ジョブ#をファイル名として持つ CSV 形式のファイルが、Windows XP の場合は C:\temp フォルダに、Windows 7 の場合は download フォルダに保存されます。CSV ファイルは表計算ソフトウェアで開くことができます。

「Export search results」ページに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/export_help.html

http://www.matrixscience.com/help/export_help.html

図 36 「Export Search Results」ページ

出力したい書式を選択してください。エクセルなどの表計算ソフトウェアで閲覧したい場合は「CSV」を選択してください。

検索条件、タンパク質情報、ペプチド情報、質量データ情報等さまざまな項目があります。出力したい項目をチェックしてください。

[Export search results] ボタンを押すとファイルに出力することができます。

5-12 Report Builder (MIS/SQ)

[Protein Family Summary]ページの[Report Builder]タブをクリックすると、図37に示す[Report Builder]ページが表示されます。[Report Builder]ページでは[Proteins]タブで表示されている代表タンパク質に関して、プロテインスコアやアクセッション番号などの項目を自由に組み合わせて、また、項目に対するフィルタ条件(たとえばプロテインスコア50以上など)を指定して、簡潔なタンパク質リストを作成することができます。作成したタンパク質リストはCSVファイルとして出力することができます。

「Report Builder」タブに関するより詳しい内容は次のページをご覧ください。

http://ホスト名/mascot/help/msms_summaries_help.html#BUILDER

http://www.matrixscience.com/help/msms_summaries_help.html#BUILDER

図 37 「Protein Family Summary : Report Builder」 ページ

The screenshot shows the 'Report Builder' tab in the Mascot Search Results interface. The page is titled 'MASCOT Search Results' and includes search parameters, a legend, and a protein family summary. The 'Report Builder' section is highlighted with a red box and a yellow callout box that says: 「Report Builder」タブを選択してください。

Below the 'Report Builder' section, there are two columns: 'Enabled' and 'Available'. The 'Available' column contains a list of protein hits with their accession numbers and descriptions. A red arrow points to the 'Available' column with a yellow callout box: ▶アイコンをクリックして[Columns:]を展開してください。[Custom]ラジオボタンを選択すると、[Enabled]と[Available]間で[→]および[←]ボタンを使って選択項目を移動することができます。また、[↑]および[↓]ボタンを使って選択項目の並び位置を変更することができます。[Apply]ボタンを使って操作内容を有効にしてください。

Below the 'Available' column, there are filter conditions. A red arrow points to the filter section with a yellow callout box: ▶アイコンをクリックして[Filters:]を展開してください。2つのプルダウンから項目と条件を選択し、値を入力して[Filter]ボタンを押してください。同様な操作を繰り返して[Update]ボタンを押すことにより、論理条件を加味した新たなフィルタ条件を追加することができます。

At the bottom of the page, there is a table of protein hits. A red arrow points to the 'Export as CSV' button with a yellow callout box: CSV ファイルを出力します。

At the bottom right of the page, there is a yellow callout box: [Columns:]と[Filters:]の条件でタンパク質リストが作成されます。項目をクリックすると斜体文字に変化し、昇順・降順のリストを表示することができます。

▶ 6 質量分析計システムとの連携

質量分析計システムが出力する「生の質量データファイル(RAWファイル)」は、同位体やノイズの質量ピークの他、LCの保持時間やスキャン番号など様々な情報を含んでいます。また、生の質量データファイルの書式は質量分析装置メーカーや機種によってまちまちです。Mascot検索ではモノアイソトピックなピークで構成されたピークリストを使いますので、生の質量データ(ファイル)を処理してピークリスト(ファイル)を作成する必要があります。

質量分析計に付随する質量データ解析用ソフトウェアはピーク抽出機能をサポートしていますので、Mascot検索に必要なピークリストファイルを作成することができ、また同時に、Mascot Serverのクライアントソフトウェアとして動作するものもあります。質量データ解析用ソフトウェアとMascot Serverの連携方法やピークリストファイルの作成方法については次のページの「Instrument Specific Tips」をご覧ください。

http://ホスト名/mascot/help_index.html

http://www.matrixscience.com/help_index.html

なお、弊社が提供する Mascot Distiller は精度良くかつ高速にピーク抽出処理を行うことができるソフトウェアです。質量分析計システムが出力する全ての種類の生の質量データファイルの書式をサポートしており、Mascot Serverのクライアントソフトウェアとして動作しますので、より簡便に質量分析計システムとMascot Serverを連携させることができます。詳しくは次のページをご覧ください。

<http://www.matrixscience.com/distiller.html>

http://www.matrixscience.jp/pdf/MascotDistiller_quickstart.pdf

6-1 アジレント・テクノロジー

質量分析計に付随する質量データ解析用ソフトウェア(MussHunter)を使って生の質量データファイル(拡張子は .d)からピーク抽出処理を行い、Mascot 検索用の MGF 形式のピークリストファイルを作成することができます。

6-2 エービー・サイエックス

(1) Analyst

Analyst の[Script]メニューに登録されている [mascot.dll] は Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。また、mascot.dll は Mascot Daemon の[Data import filter]から呼び出すことができますので、Mascot Daemon と組み合わせることにより Mascot 検索を自動化することもできます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_analyst.html

http://www.matrixscience.com/help/instruments_analyst.html

なお、mascot.dll の最新版は、

http://www.matrixscience.com/help/instruments_analyst.html

からダウンロードすることができます。

(2) Data Explorer

Data Explorer のサポートサイトからダウンロードできる Visual Basic マクロを使って Mascot Server に接続することができます。また、Data Explorer は Mascot Daemon の[Data import filter]から呼び出すことができますので、Mascot Daemon と組み合わせることにより Mascot 検索を自動化することもできます。詳しくは、次のページをご覧ください。

http://ホスト名/mascot/help/instruments_data_explorer.html

http://www.matrixscience.com/help/instruments_data_explorer.html

(3) AB SCIEX 4000/5000 シリーズ (TOF/TOF)

GPS Explorer および TS2Mascot ユーティリティは Mascot Server のクライアントソフトウェアとして動作しますので、Oracle データベースにアクセスし、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。また、MS Data Converter は Oracle データベースにアクセスし、Mascot 検索用の MGF 形式のピークリストファイルを作成することができます。詳しくは、次のページをご覧ください。

http://ホスト名/mascot/help/instruments_4000.html
http://www.matrixscience.com/help/instruments_4000.html

6-3 島津製作所

LAUNCHPAD は Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。

6-4 サーマフィッシャーサイエンティフィック

(1) Bioworks

Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_xcalibur.html#BIOWORKS
http://www.matrixscience.com/help/instruments_xcalibur.html#BIOWORKS

(2) Proteome Discoverer

Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_xcalibur.html#PROTD
http://www.matrixscience.com/help/instruments_xcalibur.html#PROTD

(3) lcq_dta_shell

生の質量データファイル(拡張子は .RAW)からピーク抽出処理を行う extract_msn を Mascot Server PC にインストールすることにより、ピークリストの作成から Mascot Server への検索投入までを連続的に実行する lcq_dta_shell ページを利用することができます。また、extract_msn は Mascot Daemon の [Data import filter] から呼び出すことができますので、Mascot Daemon と組み合わせることにより Mascot 検索を自動化することもできます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_xcalibur.html#LCQ_DTA
http://www.matrixscience.com/help/instruments_xcalibur.html#LCQ_DTA
http://www.matrixscience.jp/pdf/jap_settingsOfMs+MdForXcaliburRawFile.pdf

(4) その他

生の質量データファイル(拡張子は .RAW)からピーク抽出処理を行うプログラムがいくつか公開されています。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_xcalibur.html#EXTRACT
http://www.matrixscience.com/help/instruments_xcalibur.html#EXTRACT

6-5 日本ウォーターズ

(1) ProteinLynx Global Server (PLGS)

PLGD は Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを一括して実行することができます。

(2) MassLynx

MassLynx を使って Mascot 検索用のピークリストファイルファイル(拡張子は .pk1)を作成することができます。また、MassLynx sample list は Mascot Daemon の [Data import filter] から呼び出すことができますので、Mascot Daemon と組み合わせることにより Mascot 検索を自動化することもできます。詳しくは次のページをご覧ください。

http://ホスト名/mascot/help/instruments_masslynx.html

http://www.matrixscience.com/help/instruments_masslynx.html

6-6 日本電子

質量分析計に付随する質量データ解析用ソフトウェアを使って生の質量データファイルからピーク抽出処理を行い、Mascot 検索用の MGF 形式のピークリストファイルを作成することができます。

6-7 日立ハイテクノロジーズ

質量分析計に付随する質量データ解析用ソフトウェアを使って生の質量データファイルからピーク抽出処理を行い、Mascot 検索用の MGF 形式のピークリストファイルを作成することができます。

6-8 ブルカー・ダルトニクス

BioTools は Mascot Server のクライアントソフトウェアとして動作しますので、ピークリストの作成から Mascot Server への検索投入までを連続的に実行することができます。

7 Mascot検索の自動化 : Mascot Daemon

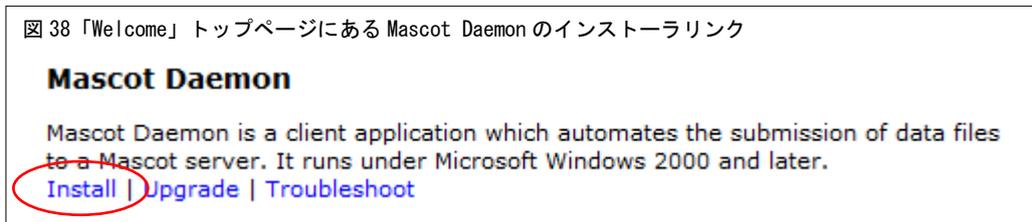
Webブラウザを利用すると、検索条件の設定やMascot検索の実行などをインタラクティブに操作することができます。一方で、同じ検索条件で100件の質量データをMascot検索にかけたいときは、PCに付きっきりで、100回の同じような操作をインタラクティブに行う必要があります。

Mascot DaemonはMascot検索作業の自動化を支援するためのクライアントソフトウェアです。たとえば、100件の質量データを一つずつ自動的にMascot検索にかけたり、生の質量データをピーク抽出プログラムで処理した後にMascot検索にかけるようなプロセスを定義し、実行することができます。Mascot Daemon は Mascot Serverに対するクライアントソフトウェアとして動作しますので、Webブラウザと同じように、Mascot Server PCや、Mascot Server PCとネットワークで接続されたPCにインストールすることができます。

7-1 インストール

Mascot Daemon のインストーラはMascot Server からダウンロードすることができます。Mascot Server の「Welcome」トップ

図 38 「Welcome」トップページにある Mascot Daemon のインストーラリンク



ページ(<http://ホスト名/mascot/>)を開き、ページの中程にある[Install]リンクをクリックしてください(図 38)。「Installing Mascot Daemon」のページが開きますので、インストール手順を読み進み、「5. Install Mascot Daemon by clicking on this link」の次にある[Daemon.ms]リンクをクリックしてください。[ファイルのダウンロード]のダイアログが表示されますので、[実行]ボタンを押し、表示されるダイアログの内容をに従ってインストール進めてください。

インストールが終了したら、[スタート]→[すべてのプログラム]→[Mascot]→[Mascot Daemon]をクリックして Mascot Daemon を起動してください。初めて Mascot Daemon を起動すると、Mascot Server の URL を指定するための [Preferences] ダイアログが表示されますので、例にならって Mascot Server の URL (<http://ホスト名/mascot/cgi/>) を入力し、[Save]ボタンを押ししてください。Mascot Server と通信ができることが確認できた時点で Mascot Daemon が起動します。なお、Windows 7 の場合は次の操作を行った後に Mascot Daemon を起動し、Mascot Server の URL を指定してください。

[C:\Program Files (x86)\Matrix Science\Mascot Daemon\Daemon.exe]を右クリック→[プロパティ]→[互換性]タブ→[特権レベル]の「管理者としてこのプログラムを実行する」をチェック

7-2 マニュアル

次の書類をご覧ください。

http://www.matrixscience.jp/pdf/jap_mdaemon_manual_2_4.pdf

また、英文ヘルプページ([Help]メニュー→[Mascot Daemon Help ...])も併せてご覧ください。

▶ 8 Mascot Serverの管理

Mascot Serverは、配列データベース検索に関連するプログラムやそれらのプログラムを制御する設定ファイルなどで構成された、複雑で巨大なソフトウェアです。Mascot Serverは極めて安定に動作するサーバソフトウェアですが、利用状況やエラーログなどを定期的にチェックすることにより、より安定した状態で利用・運用することができます。

8-1 Welcome トップページ

図39に示したように、「Welcome」トップページにはMascot検索を実行するための[Mascot]リンクを始めとして、配列データベースの稼働状況を確認することができる[Database Status]リンク、過去の検索結果にアクセスすることができる[Search Log]リンク、Mascot Serverの設定を編集するためのページを集めた[Configuration Editor]リンク、ヘルプページやマニュアル類へのリンクなど、様々なリンクがあります。

より新しいサポート情報は弊社ホームページに掲載されています。[WHAT'S NEW]、[HELP]、[SUPPORT]ページにアクセスし、対応する項目をご覧ください。弊社ホームページへは「Welcome」トップページ左上にある弊社ロゴをクリックするか、次のURLにアクセスしてください。

<http://www.matrixscience.com>

Mascot Serverで使われている検索エンジンやユーティリティプログラムの多くは単体で動作させることができますので、コマンドプロンプトから直接実行したり、自作のプログラムから呼び出して利用することができます。詳しい仕様については次の英文マニュアル「Setup & Installation Manual」の[Program Reference]の項をご覧ください。

<http://ホスト名/mascot/pdf/manual.pdf>

8-2 検索ログ：Search Log

Mascot Serverは、実行されたMascot検索に対して連続的な番号(検索Job#と呼んでいます)を付与し、ユーザ名や検索タイトルとともに検索ログとして記録しています。「Welcome」トップページの[Search Log]リンクをクリックするか、次のURLにアクセスしてください。図40に示す「MASCOT search log」ページが表示されます。

<http://ホスト名/mascot/x-cgi/ms-review.exe>

[Job#]欄の番号をクリックすると、それに対応する検索結果ページが表示されます。

各項目名の下にあるラジオボタン(○)は、昇順(1)あるいは降順(-1)に並べ替える際に選択します。選択後に [Sort/filter] ボタンを押してください。

各項目名の下にあるチェックボックス(□)は、チェックされた項目の全内容を表示させる際に選択します。チェックした後に [Sort/filter] ボタンを押してください。

図 39 「Welcome」トップページのリンク



チェックボックスの下にある枠は文字列の入力ウィンドウになっています。ここに入力した文字列にマッチする内容を持つ検索ログが表示されます。マッチさせたい文字列を入力し、[Sort/filter]ボタンを押してください。なお、検索条件設定ページの[Your name]、[Email]、[Search title]で入力した内容は[User Name]、[Email]および[Title]欄にそのまま反映されます。

図 40 Mascot search log ページ



「MASCOT search log」ページは C:\inetpub\mascot\logs\searches.log ファイルの内容を反映していますが、バックアップしたログファイルと検索結果ファイルを指定し「MASCOT search log」ページに表示させることもできます。また、各項目の表示文字数は C:\inetpub\mascot\config\mascot.dat ファイルの[Options]セクションにある[ReviewColWidths]で指定されており、テキストエディタまたは[Configuration Editor]を利用して変更することができます。

8-3 配列データベース稼働状況 : Database Status

「Welcome」トップページの [Database Status] リンクをクリックするか、次の URL にアクセスしてください。図 41 に示す「MASCOT search status page」ページが表示されます。

<http://ホスト名/mascot/x-cgi/ms-status.exe>

Mascot Server のバージョン、ライセンス情報、利用可能なプロセッサ数の他、現在有効になっている配列データベース毎に、現在使用されている配列データベースファイル、現在の稼働状況 (Status 行が「In use」であれば使用可能)、検索状況 ([配列データベース名] リンク)、統計情報 ([Statistics] リンク : 残基数、エントリ数、最も大きなエントリの残基数、残基の出現数、各生物種に含まれるエントリ数) などを確認することができます。

また、検索ログ ([Searches log] リンク)、エラーログ ([error log] リンク) にアクセスすることができます。

図 41 Mascot search status page ページ



8-4 設定値変更 : Configuration Editor

Mascot Serverは、原子の質量、修飾、消化酵素、イオンシリーズ、配列データベース、定量解析手法に関係する様々な設定値を持っています。これらの設定値は、`C:\inetpub\wwwroot\mascot\config` フォルダに存在する設定ファイルの中で定義されており、Mascot Configuration Editorを利用して閲覧・編集することができます。「Welcome」トップページの[Configuration Editor]リンクをクリックするか、次のURLにアクセスしてください。図42に示す「Mascot Configuration」ページが表示されます。

<http://ホスト名/mascot/x-cgi/ms-config.exe>

たとえば、新たな修飾を追加したい場合は[Modifications]リンクをクリックしてください。「Mascot Configuration: Modification」ページが表示されますので、[Add new modification] ボタンを押し、修飾情報を定義してください。

Mascot Configuration Editorを使って次の設定値を閲覧・編集することができます。

(1) Elements

元素の名称とモノアイソトピックおよび平均質量を編集することができます。

(2) Amino Acids

アミノ酸情報を閲覧することができます。

(3) Modifications

修飾情報の閲覧・編集の他、追加・削除することができます。

(4) Symbols

原子情報を閲覧することができます。

(5) Enzymes

消化酵素情報の閲覧・編集の他、追加・削除することができます。

(6) Instruments

MIS検索で使用するイオンシリーズの組合せの閲覧・編集の他、追加・削除することができます。

(7) Quantitation

定量解析手法の閲覧・編集の他、追加・削除することができます。

(8) Configuration Option

mascot.datのOptionsセクションの設定値を閲覧・編集することができます。

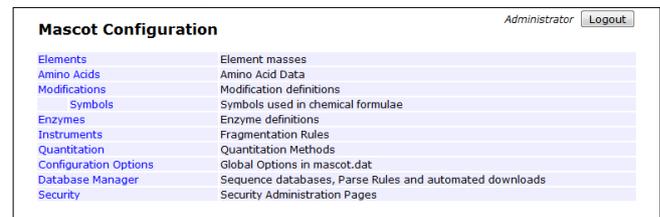
(9) Database Manager

配列データベースの設定内容の閲覧・編集の他、追加・削除することができます。

(10) Security

Mascot security機能が有効になっている場合に表示されます。ロールに基づくユーザ権限を管理することができます。

図 42 Mascot Configuration ページ



8-5 セキュリティ

Mascot Server はセキュリティロールに基づくユーザ管理機能をサポートしており、Mascot Server へのアクセス(ユーザ名とパスワードを要求されます)、Mascot 検索の許可、配列データベースの利用許可、検索ログの閲覧許可など、様々な項目に対してアクセス制限を付与することができます。

詳しくは次のマニュアルをご覧ください。

<http://ホスト名/mascot/pdf/manual.pdf>

http://www.matrixscience.jp/pdf/jap_security.pdf

8-6 検索条件のデフォルト設定

WebブラウザのCookie機能を利用して検索条件の初期値(デフォルト値)を設定することができます。「Welcome」トップページの[Mascot]リンク→[Search Form Defaults]リンクの順にクリックするか、次のURLにアクセスしてください。図43に示す「Set Mascot search form defaults」ページが表示されます。

http://ホスト名/mascot/cgi/form_defaults.pl

初期値として設定したい項目や数値を選択または入力し、ページ下方にある[Save defaults as cookie...]ボタンを押してください。

「Your search form defaults have been saved as cookie」のメッセージが表示され、検索条件の初期値として設定されます。

なお、この設定はWebブラウザがサポートしているCookieの機能を利用しているため、Mascot Server側ではなく、クライアントソフトウェアであるWebブラウザ側のCookieの設定として記録されます。従いまして、この設定は異なるログオンユーザや異なるコンピュータ上で起動したWebブラウザには反映されませんので、Webブラウザ毎に設定してください。

図 43 Set Mascot search form defaults ページ

8-7 エラーログ

Mascot Serverの動作に関わるエラーは、C:\inetub\macot\logs\errorlog.txtファイルにエラーログとして記録されています。エラーログを閲覧するにはこのファイルをノートパッドのようなテキストエディタで開くか、「Welcome」トップページの[Database Status]リンク→[error log]リンクをクリックしてください。

エラーは次のような書式で出力されます。

```
Error [M00422 - Job -16 - X00308:compress] - Wed Aug 17 15:08:09 2011
- Accession [gi|323714265] is a duplicate, database file name: nr_human_20110817x.fasta
```

エラー出力の先頭に記載されている「M00422」はエラー番号です。エラー番号の内容とトラブルシューティングを確認するには、「Welcome」トップページの[Database Status]リンク→[Error message descriptions]リンクをクリックしてください。

8-8 バックアップ

Mascot Serverは `C:\inetpub\mascot` フォルダにインストールされ、図16 に示すようなフォルダ構造を持っています。Mascot Server PC本体の不測のトラブルによる情報消失に備え、定期的に `C:\inetpub\mascot` フォルダを別の記憶装置にバックアップしてください。トラブルが起きた場合、このバックアップを使ってトラブル発生前の状態に戻すことができます。バックアップ先記憶装置の記憶容量等の問題ですべてのフォルダのバックアップが難しい場合は、少なくとも `config`、`data`、`logs` フォルダをバックアップするようにしてください。このバックアップを使ってトラブル発生前と同じような動作環境に戻すことができます。

バックアッププログラムはフリーのものも含めていろいろありますが、次の書類はマイクロソフトが提供する高機能バックアッププログラムである「robocopy」プログラムと、Windowsがサポートしている「タスクスケジューラ」を組み合わせ、指定した時間にバックアップを実行するための設定手順を紹介しています。

http://www.matrixscience.jp/pdf/jap_robocopy_settings.pdf

2012年6月 第10版

Copyright 2004 - 2012 Matrix Science Ltd. , マトリックスサイエンス株式会社

本書の一部あるいは全部について、マトリックスサイエンス株式会社から文書による許可を得ずに、いかなる方法においても無断で複写、複製することを禁じます。



マトリックスサイエンス株式会社

電子メール : support-jp@matrixscience.com

電話 : 03-5807-7895 ファクシミリ:03-5807-7896

住所 : 〒101-0021 東京都千代田区外神田 6-10-12 KN ビル 3 階