

MASCOT Distiller

Replicate ラベルフリー定量解析チュートリアル

目次

1. はじめに:本資料でご紹介する計算内容について	2
2. 定量解析を行うデータについて	3
3. 準備 1・データ取得.....	5
4. 準備 2・データベースの設定.....	6
5. 準備 3・定量設定の定義.....	15
6. 定量計算実行 1・データ読み込み.....	22
7. 定量計算実行 2・ピーク抽出.....	26
8. 定量計算実行 3・MASCOT 検索と定量計算前の設定変更.....	28
9. 定量計算実行 4・定量計算実行と、結果表示内容について.....	36
10. 定量計算実行 5・計算後の report、export.....	46



1. はじめに:本資料でご紹介する計算内容について

この資料では、質量分析装置のデータを使ったプロテオミクス・DDA 解析において、Distiller を使って **Precursor ベースのラベルフリー計算**を実行する具体的な手順について説明しています。データは4種類で3つの technical replicates を含む計 **12 の raw データ**で構成されていて、ペプチド同定結果をもとに溶出時間がアライメントされたプリカーサーの eXtracted Ion Chromatograms (XICs)の相対強度に基づいて定量解析を行います。raw データから直接ペプチドが同定された際にはその同定個所を、raw データに同定ペプチドが見つからない場合でも global alignment の結果をもとに Precursor のピークを検出し XICs 計算に利用します。**出力内容は reference との比**となります。

資料の前半(手順2~5)では データ解析を行うための準備について、後半(手順6~10)では Distiller 上で実際に解析を行うとともに、結果の見方や、データを出力して他ソフトウェアで解析を行う方法について紹介しています。資料の構成内容については目次をご確認ください。

本資料でご紹介しているデータ解析は、弊社で公開しているブログ記事が元となっております。そちらも併せてご参照ください。

ブログ記事・英語:

<http://www.matrixscience.com/blog/global-thinking-label-free-quantitation-in-mascot-distiller-2-8.html#comments>

ブログ記事・日本語訳:

http://www.matrixscience.co.jp/blogsJ/blog_202104.html

[解析に必要な Distiller の構成]

MASCOT Distiller で定量解析を行うためには、以下の製品が必要です。

- ・MASCOT Server [インターネット試用版でなく、製品版]
- ・MASCOT Distiller [定量モジュールまで含む]

またこの資料で紹介する Distiller 定量計算は、ver.2.8.0 以降の機能を利用しています。Distiller 製品版をお持ちの方でも定量モジュールをお持ちでない方は実施できませんのでその点ご注意ください。Distiller をお持ちでない方、並びにお持ちになっても定量モジュールをお持ちでない方は試用ライセンスを提供可能です。お気軽に弊社までご連絡ください。

2. 定量解析を行うデータについて

手順 2 ではチュートリアルにおいてどのようなデータを解析するのかについて説明いたします。

本資料で実際にデータ解析を行うのは、以下論文で使用された raw データです。

Shalit T, Elinger D, Savidor A, Gabashvili A, Levin Y. J. Proteome Res. 2015, 14, 1979-1986

<https://pubs.acs.org/doi/10.1021/pr501045t>

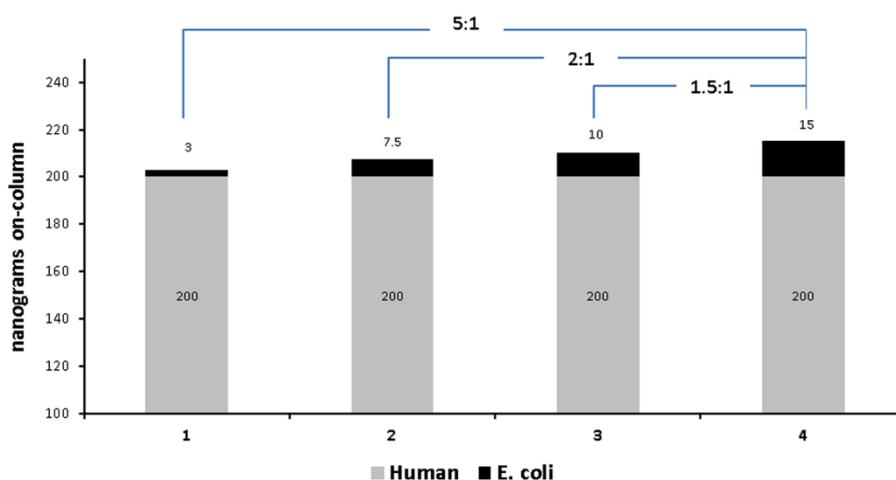
以下のサイトでデータが公開されているとともに、データや解析内容についての説明があります。raw データの取得先でもあります。

<https://www.ebi.ac.uk/pride/archive/projects/PXD001385>

本資料においても、チュートリアルの内容を理解するのに最低限必要なデータの構成並びにその中身について説明いたします。

使用データでは、200ng の HeLa 消化物にスパイクされた大腸菌 3 ng, 7.5 ng, 10 ng, 15 ng を混合したサンプル(下図)を、それぞれ3回測定しています。

実験データのデザインの意図についてですが、まず大腸菌に着目した場合、同一のタンパク質について 15ng スパイクしたサンプルとそれ以外のサンプル と比較すると、それぞれ期待される比率が 1:5(3ng), 1:2 (7.5ng), 1:1.5(10ng)となります。実際のデータがその期待された比率とどの程度ずれているかを見る事で、定量計算結果を評価する事ができます。また human(HeLa)のタンパク質に着目した場合、どのサンプル間で比較しても比率が 1:1 になる事が期待され、こちらも同様にそのずれ方から定量結果を評価する事ができます。



4 種類のサンプルについての説明。Human 並びに E.coli の混入量が記されている。

論文 Shalit T, et.al(2015) Proteome Res. 2015, 14, 1979-1986 の Fig.1 より

公開先で取得できるファイルと、サンプルの内容との関係性は以下の通りです(下表)。

File 名	Sample: Ecoli spike 量	繰り返し番号
QEP1_SpikeIn_230914_1_3ng_270914.raw	3 ng	1
QEP1_SpikeIn_230914_2_3ng_270914.raw		2
QEP1_SpikeIn_230914_3_3ng_270914.raw		3
QEP1_SpikeIn_230914_4_7-5ng_270914.raw	7.5 ng	1
QEP1_SpikeIn_230914_5_7-5ng_270914.raw		2
QEP1_SpikeIn_230914_6_7-5ng_270914.raw		3
QEP1_SpikeIn_230914_7_10ng_270914.raw	10 ng	1
QEP1_SpikeIn_230914_8_10ng_270914.raw		2
QEP1_SpikeIn_230914_9_10ng_270914.raw		3
QEP1_SpikeIn_230914_10_15ng_270914.raw	15 ng	1
QEP1_SpikeIn_230914_11_15ng_270914.raw		2
QEP1_SpikeIn_230914_12_15ng_270914.raw		3

公開されている raw ファイルとサンプルの種類について。繰り返し番号は解析に利用していません。

また検索対象としたデータベースも併せて公開されています。お手元にある human,ecoli 並びに contaminants 用データベースをそのままご利用いただいても解析に問題は生じませんが、手っ取り早く解析を行いたい場合は公開先の fasta ファイルを MASCOT にセットしてご利用ください。手順4で取得したデータベースを MASCOT で構築する方法についてご案内しています。

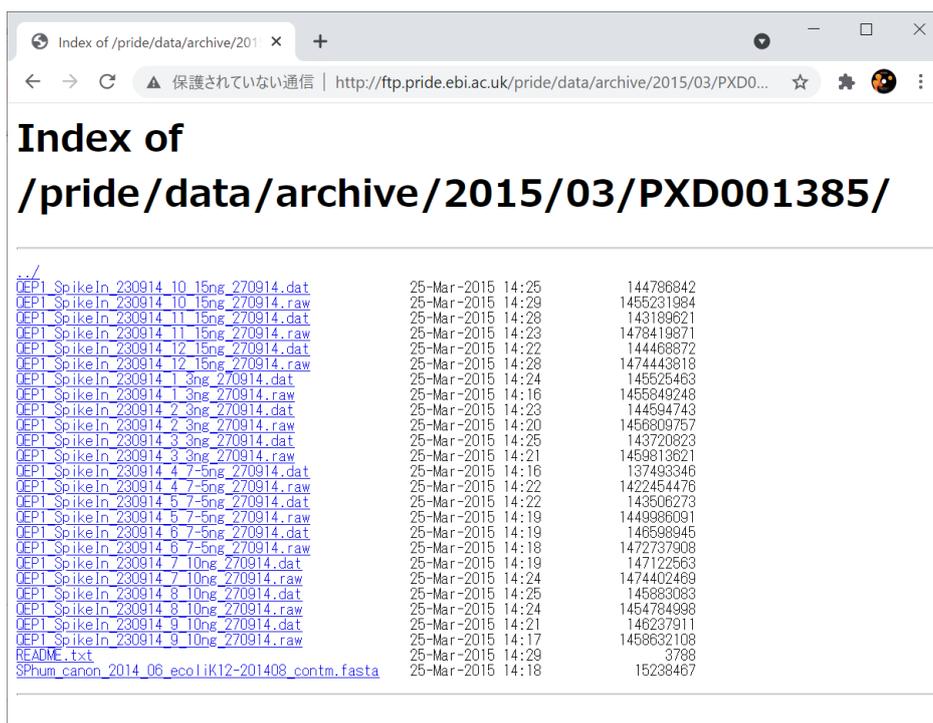
3. 準備 1・データ取得

以降 手順 3~5 では、定量解析前の準備について説明しています。

手順 3 では、手順 2で説明したデータの取得についてご案内いたします。

まず、WEB ブラウザなどでデータの公開元である以下サイトにアクセスしてください(下図)。

<http://ftp.pride.ebi.ac.uk/pride/data/archive/2015/03/PXD001385/>



データ公開先をウェブブラウザで開いた際の画面。各データについて raw と dat (MASCOT 検索結果ファイル)があるが、raw のみを取得する。必要に応じて fasta ファイルもダウンロードする。

ここで公開されている 全 raw ファイル (12 個)と、MASCOT の検索対象データベース FASTA である、“SPhum_canon_2014_06_ecoliK12-201408_contm.fasta”をダウンロードしてください。

[ダウンロードに関する参考情報]

- Raw ファイルサイズ平均 1.45 GB。12 のデータ、合計 約 18 GB
- fasta ファイル: 15MB。

ダウンロードに時間がかかるため、その間に手順 4 または 5 を進めておくことをお勧めします(手順 4 を行うには上記 fasta ファイルが必要です)。

4. 準備 2・データベースの設定

手順 4 では、MASCOT Server 上に解析用のデータベースをセットする方法についてご案内いたします。作業前にいくつかご注意・ご確認いただきたい事項があります。

- ・ご自身で準備したデータベースを利用する場合は必ずしもこの手順4でご紹介する設定を行う必要はありません。
- ・手順3でダウンロードをご案内している fasta ファイルを取得しておく必要があります。
- ・今回取得するデータベースについて、名称を“PXD001385”、FASTA 先頭行の抜き出しルールはスペース前後の区切りを利用するとします。実際には任意の名称、任意の設定でも問題ありません。

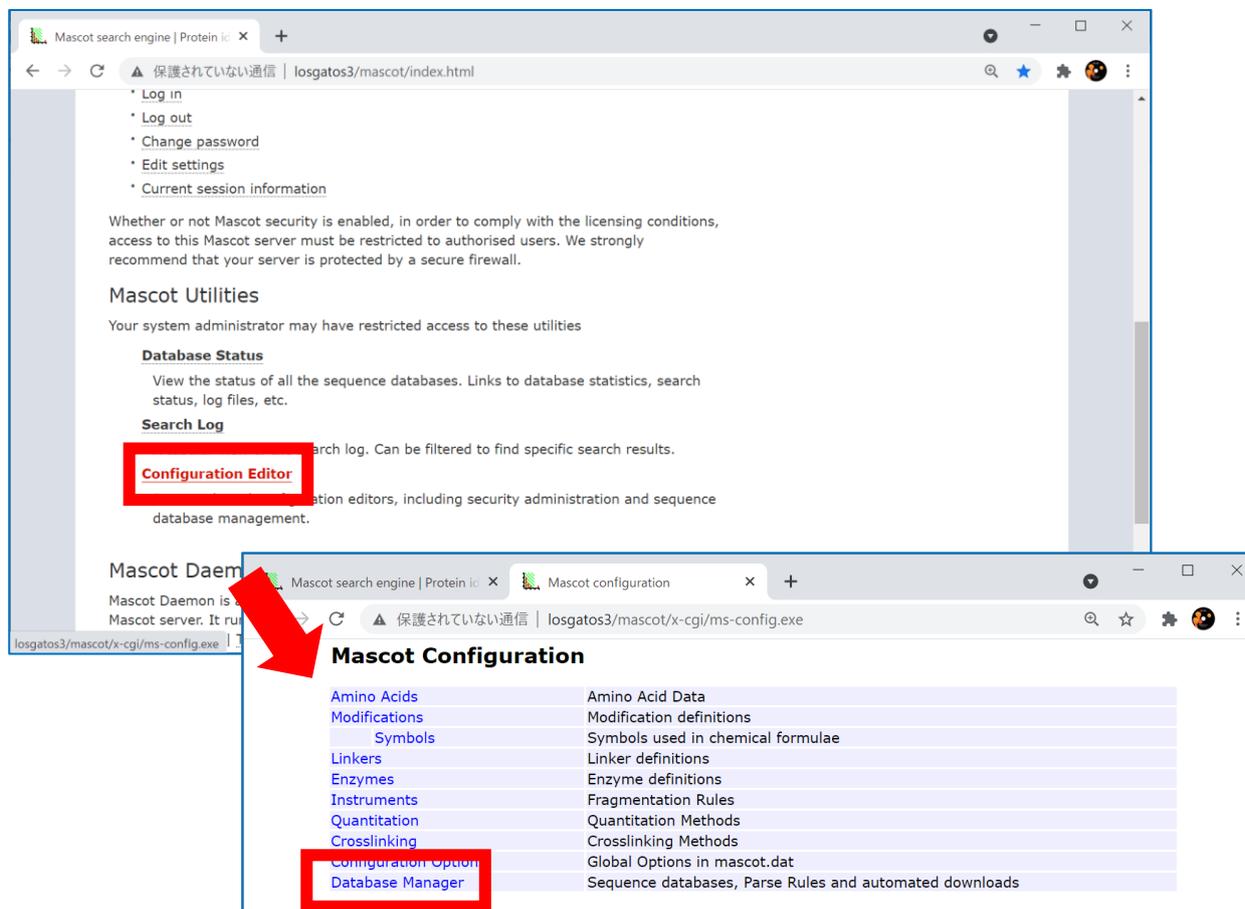
区切り例

先頭行 >sp|05100|3MG_ECOLI DNA-3-methyl adenine glycosylase

Accession : sp|05100|3MG_ECOLI

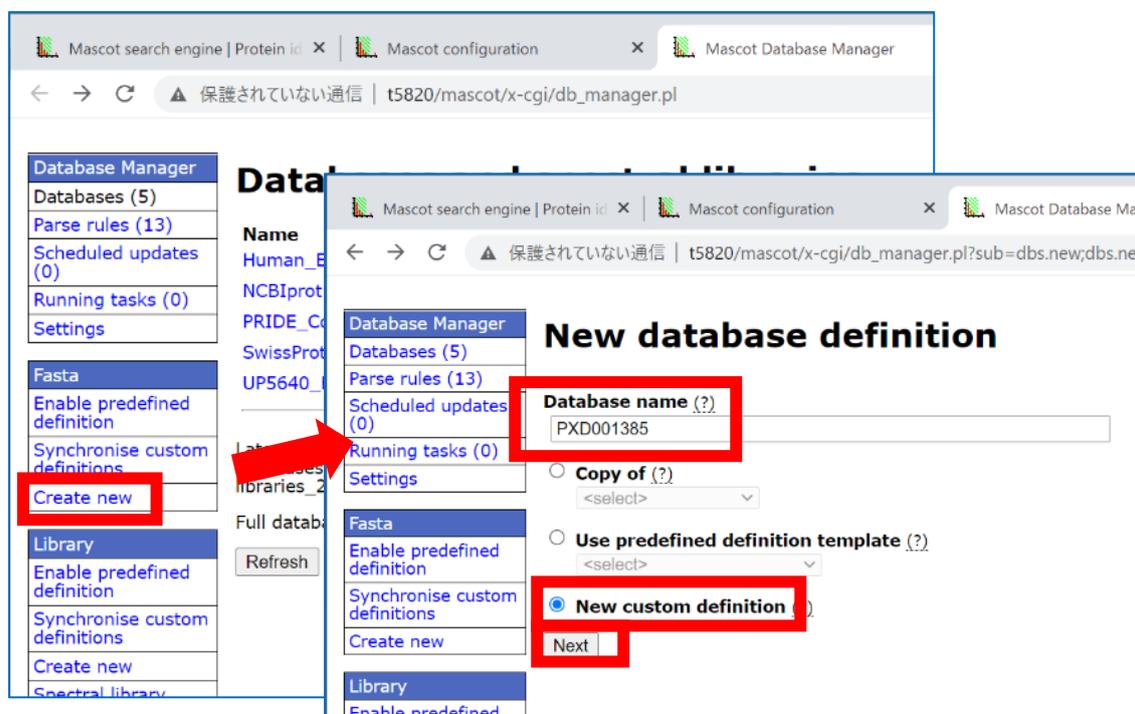
Description : DNA-3-methyl adenine glycosylase

まず、MASCOT Server の Home 画面にアクセスし、Configuration Editor -> Database Managerを開きます。

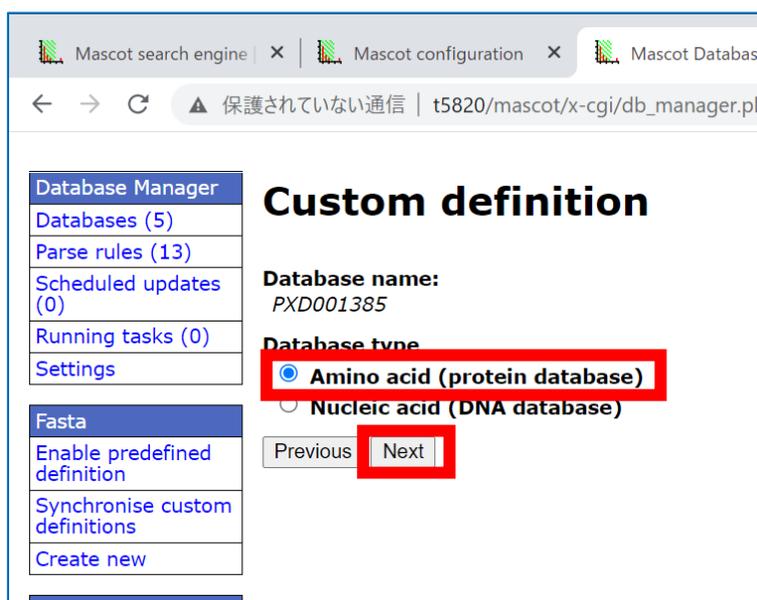


左フレームの”Fasta”にある、”**Create new**”リンクをクリックします。

続いて現れる画面で、”Database name” に”**PXD001385**”（データベースの名称）と記入し、作成の種類として”**New custom definition**”のラジオボタンを選択し、”**Next**”ボタンを押します。

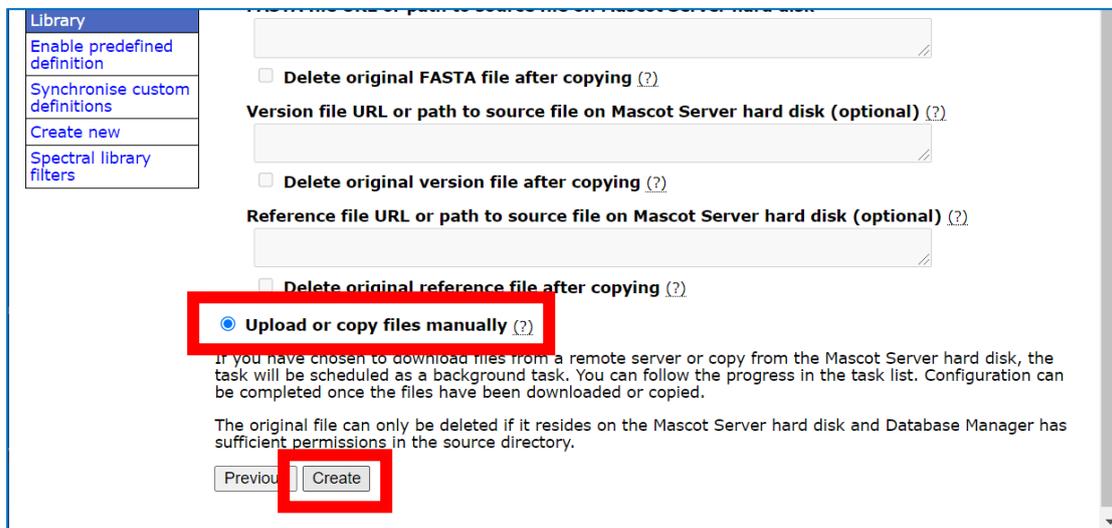


FASTA に含まれる配列の種類を選択します。ここでは ”**Amino acid (protein database)**” を選択します。選択後、”**Next**”ボタンを押します。

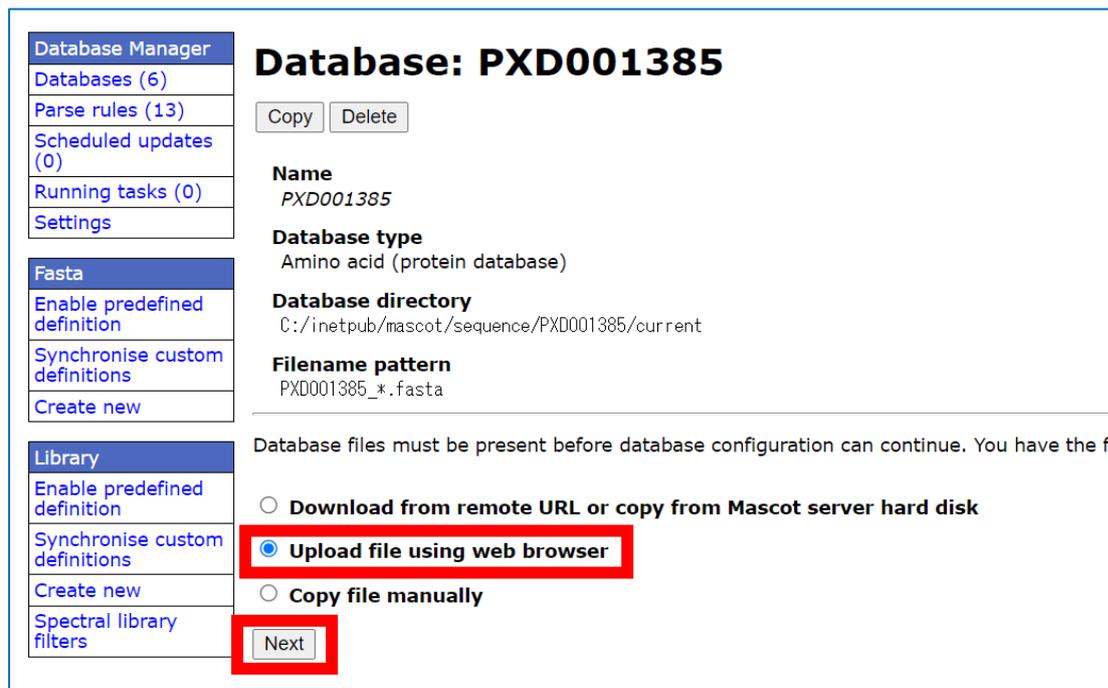


FASTA ファイルの設置方法を選択します。普段慣れている方法があればこの資料の指示通りでなくても問題ありません。ここではブラウザを介してファイルをアップロードする方法でご案内します。

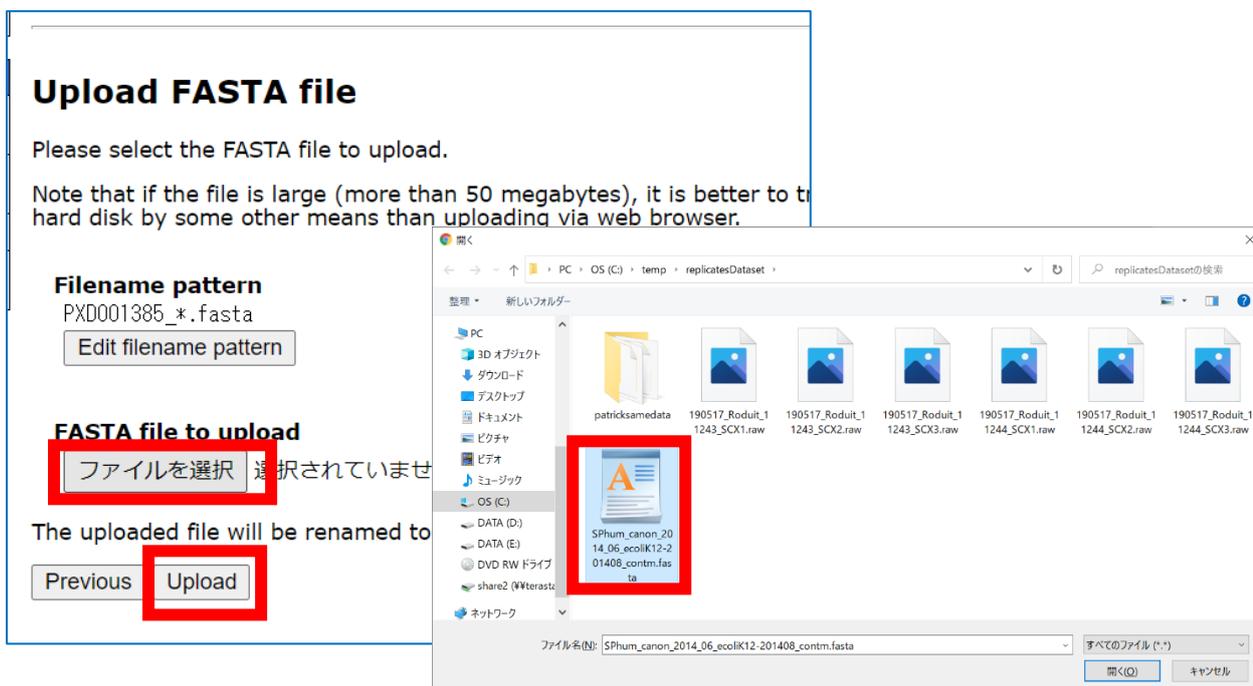
“**upload or copy file manually**”を選択し、“**Create**”ボタンを押します。



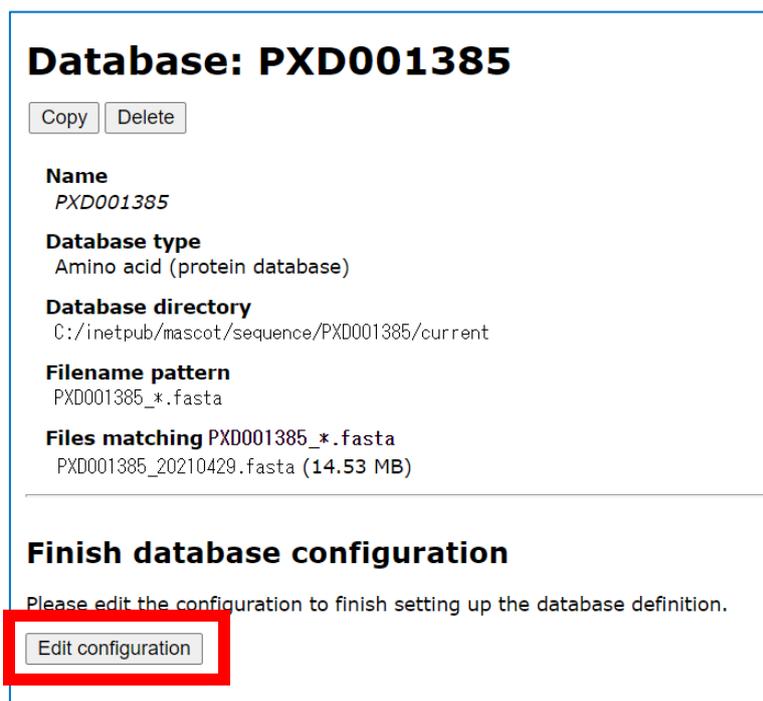
データベース設定の枠組みが構築されます。引き続きファイルを WEB ブラウザ経由で設置するため “**Upload file using web browser**”を選択し、“**Next**”ボタンを押します。



“FASTA file to upload”の下にある”ファイルを選択”ボタンを押し、公開サイトからダウンロードした fasta ファイルを選んで開きます。選択後元の画面に戻って”Upload”ボタンを押します。



ファイルがアップロードされ認識されると画面が切り替わり、”Edit configuration”ボタンが現れるのでボタンを押します。



データベースの抽出ルールを設定します。

まずは Accession (データベースの中で唯一となるような文字列・識別子)の設定を行います。

“Accession parse rule “の下にある”Choose”ボタンを押します。スペースあるいはカンマより前の識別文字をすべて選択するルール、”>¥([,]*¥)”の行を選択します。

Database configuration: PXD001385 (step 1/2)

FASTA file

Files matching PXD001385_*.fasta
 PXD001385_20210429.fasta (14.53 MB)

Accession parse rule (?)

Choose

Description parse rule (?)
 (none)

Choose

Cancel

Next

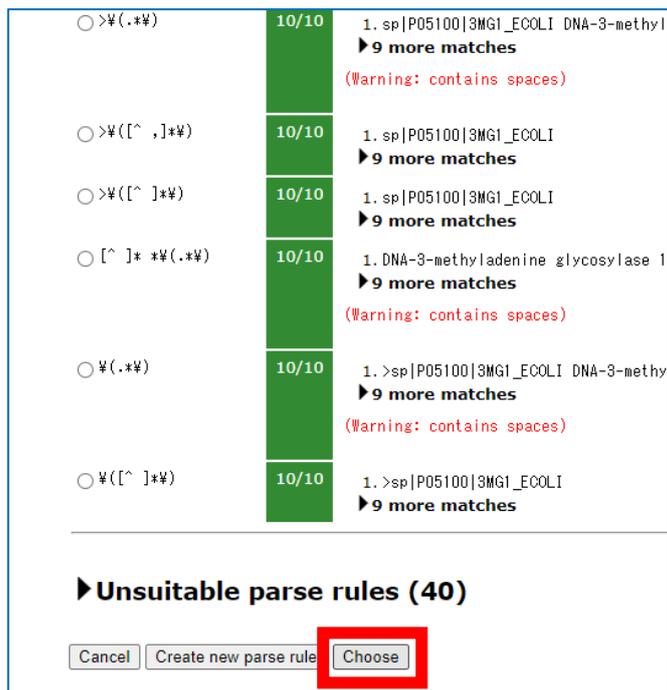
Accession and description parse rule must be selected before you can continue to the next step.

抽出ルールの候補が10現れます。候補の中にある”>¥([,]*¥)”を選びます。

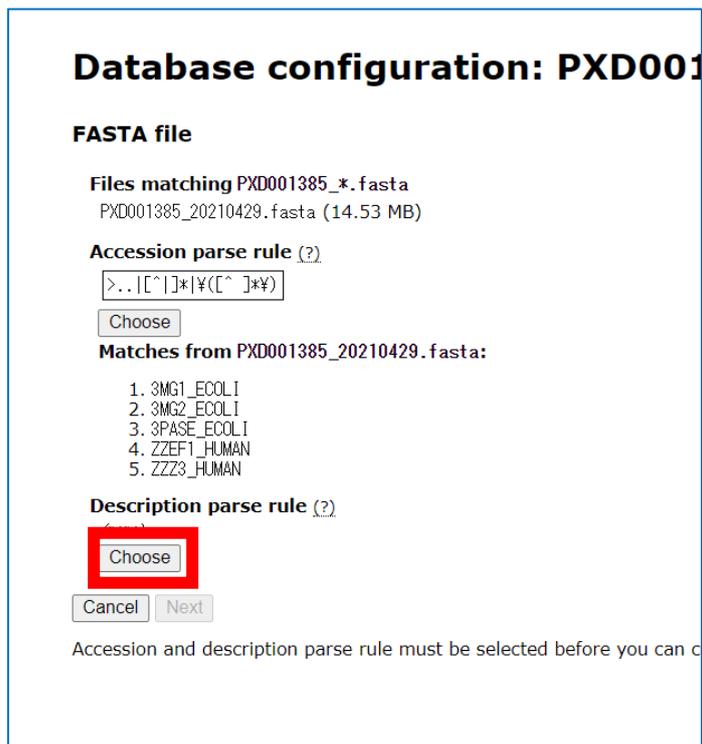
Suitable parse rules (10)

Parse rule	Match	Extracted data
<input type="radio"/> >.. [^\] * ¥([^\])*¥)	10/10	1. 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >.. ¥([^\])*¥)	10/10	1. P05100 ▶ 9 more matches
<input type="radio"/> >[^\]* ¥(.)*¥)	10/10	1. DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (stra ▶ 9 more matches (Warning: contains spaces)
<input type="radio"/> >[^\]* ¥([^\])*¥)	10/10	1. P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >¥(.)*¥)	10/10	1. sp P05100 3MG1_ECOLI DNA-3-methyladenine glycosylase 1 OS=E ▶ 9 more matches (Warning: contains spaces)
<input checked="" type="radio"/> >¥([^\ ,]*¥)	10/10	1. sp P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >¥([^\])*¥)	10/10	1. sp P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> [^\]* ¥¥(.)*¥)	10/10	1. DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (stra ▶ 9 more matches (Warning: contains spaces)
<input type="radio"/> ¥(.)*¥)	10/10	1. >sp P05100 3MG1_ECOLI DNA-3-methyladenine glycosylase 1 OS=

選択後、画面下部の”Choose”ボタンを押してください。



元の設定画面に戻ります。続いて同様の操作で”Description”の抽出ルールを設定します。”Description parse rule”の下にある”Choose”ボタンを押してください。



抽出ルールの候補が 10 現れます。スペース以降すべての文字を Description として認識する、"**>[^]* ¥(.*)**" を選択して、画面下部の"**Choose**"ボタンを押してください。

Suitable parse rules (11)

Parse rule	Match	Extracted data
<input type="radio"/> >.[^]*¥([^\s]*)	10/10	1. 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >.[^\s]*¥([^\s]*)	10/10	1. P05100 ▶ 9 more matches
<input checked="" type="radio"/> >[^]* ¥(.*)	10/10	1. DNA-3-methyladenine glycosylase 1 OS=Escherichia coli ▶ 9 more matches
<input type="radio"/> >.[^]*¥([^\s]*)	10/10	1. P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >¥(.*)	10/10	1. sp P05100 3MG1_ECOLI DNA-3-methyladenine glycosylase ▶ 9 more matches
<input type="radio"/> >¥([^\s,]*)	10/10	1. sp P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >¥([^\s]*)	10/10	1. sp P05100 3MG1_ECOLI ▶ 9 more matches
<input type="radio"/> >¥([^\s]*)	10/10	1. sp ▶ 9 more matches
<input type="radio"/> [^\s]* ¥(.*)	10/10	1. DNA-3-methyladenine glycosylase 1 OS=Escherichia coli ▶ 9 more matches
<input type="radio"/> ¥(.*)	10/10	1. >sp P05100 3MG1_ECOLI DNA-3-methyladenine glycosylase ▶ 9 more matches
<input type="radio"/> ¥([^\s]*)	10/10	1. >sp P05100 3MG1_ECOLI ▶ 9 more matches

▶ **Unsuitable parse rules (39)**

Cancel Create new parse rule **Choose**

選択後元の画面に戻ります。抽出ルールが意図しているものであることを再度確認した上で、画面下部にある"**Next**"ボタンをクリックします。

Files matching PXD001385_*.fasta
PXD001385_20210429.fasta (14.53 MB)

Accession parse rule (?)
>.[^]*¥([^\s]*)
Choose

Matches from PXD001385_20210429.fasta:

1. 3MG1_ECOLI
2. 3MG2_ECOLI
3. 3PASE_ECOLI
4. ZZEF1_HUMAN
5. ZZZ3_HUMAN

Description parse rule (?)
>[^]* ¥(.*)
Choose

Matches from PXD001385_20210429.fasta:

1. DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (strain K12)
2. DNA-3-methyladenine glycosylase 2 OS=Escherichia coli (strain K12)
3. Inorganic triphosphatase OS=Escherichia coli (strain K12) GN=...
4. Zinc finger ZZ-type and EF-hand domain-containing protein 1 OS=Escherichia coli (strain K12)
5. ZZ-type zinc finger-containing protein 3 OS=Homo sapiens GN=...

Cancel **Next**

続いて現れる画面は taxonomy 設定やタンパク質情報を取得する際に使用する設定の画面ですが、今回は特に設定しないとします。画面下部にある”**Save and finish**”ボタンを押してください。

(example)

- Copy from *EST_human* (or *EST_mouse* or *EST_others*)
<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?rettype=gb&retmode=text&db=nucleotide&tool=mascot&from=gi>
 (example)
- Copy from *NCBItr* (or *NCBIprot*)
<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?rettype=gp&retmode=text&db=protein&tool=mascot&from=gi>
 (example)
- Copy from *UP186698_X_laevis* (and similar entries)
<https://www.uniprot.org/uniprot/#ACCESSION#.txt>
 (example)

Local program (?)

Command line template

Parse rule
 (none)

To choose or edit local program as a full-text report source, please bring the database online.

作成したデータベース画面に戻ります。画面下部にある “**Activate**” ボタンをクリックしてデータベースを使用可能な状態にしてください。

Database: PXD001385

Name
 PXD001385

Database type
 Amino acid (protein database)

Database directory
 C:/inetpub/mascot/sequence/PXD001385/current

Filename pattern
 PXD001385_*.fasta

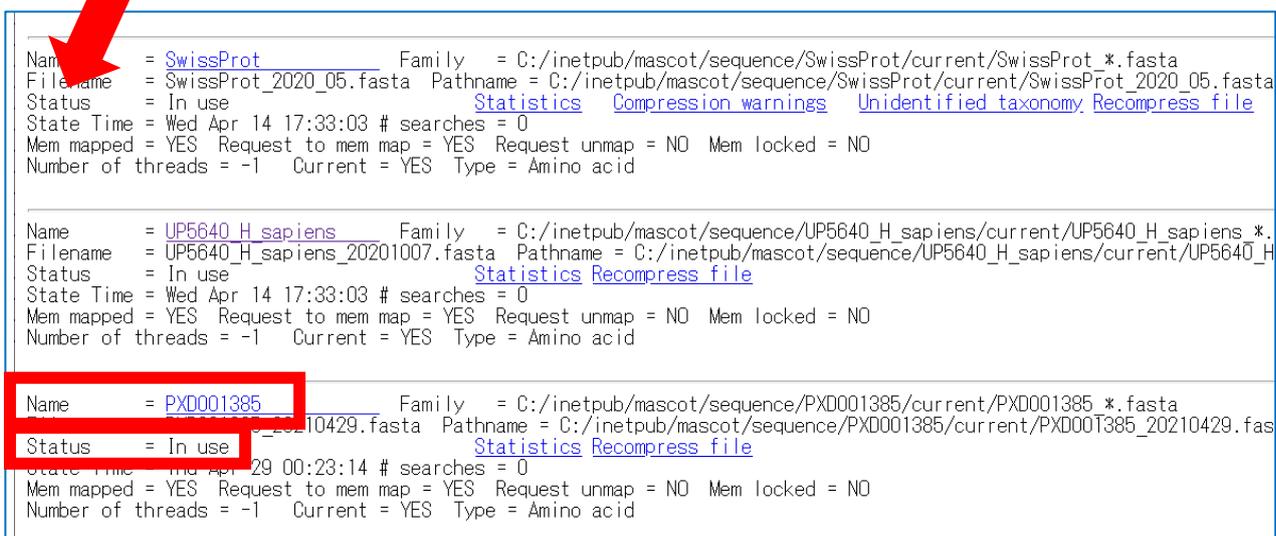
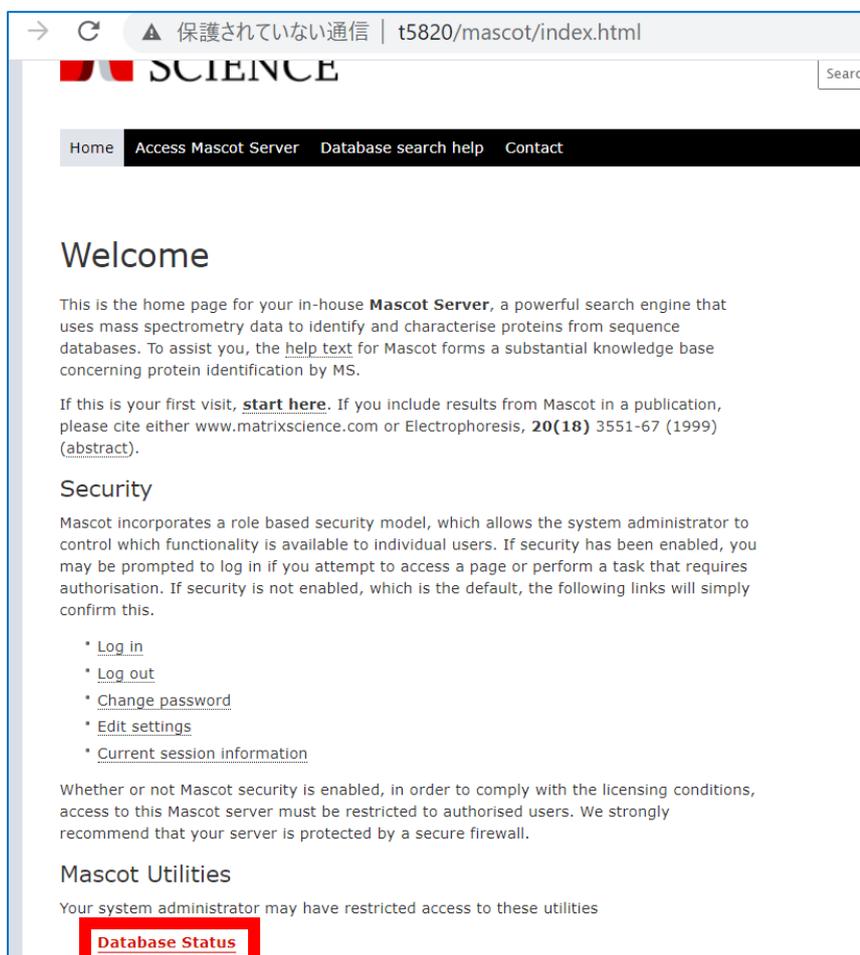
Files matching PXD001385_*.fasta
 PXD001385_20210429.fasta (14.53 MB)

Database status
 Offline

Scheduled updates
 (no schedules defined)

Main database file URL has not been defined; database updates are disabled.

データベース構築の確認は “**Database Status**” 画面にて行います。Home 画面にある “Database status” をクリックします。作成したデータベース PXD001385 が使用中のデータベースの一覧画面に表れているかどうか、現れている場合、“**Status**”項目が “**In use**” となっているか、の2点についてご確認ください。



5. 準備 3・定量設定の定義

手順 5 では、MASCOT Server 上に、今回の解析に特化した Label Free (Replicate)設定を作成します。サンプルの種類(今回の場合は 3ng,7.5ng,10ng,15ng の4種類、それぞれ3つずつの繰り返しで x3 となり計 12 個)の Components 定義作成と、結果画面に表示する ratio の定義を行います。Distiller で計算を行う際、ここで定義した定量の設定を選択して検索を実行します。

MASCOT ではあらかじめ 検索パラメータ”Quantitation”の中に、”Label Free [MD]”という設定があり、検索時にこれを選択すれば replicate ラベルフリーの定量計算を行う事ができます。しかし特に詳しい定義をせずにこの設定を利用した場合、結果として表示される各 raw ファイルの名称や表示される ratio については MASCOT Server 側で自動的に割り振られた記号・ratio となってしまいます。項目名をわかりやすくしたり、ratio を標準以外のものにしたい場合は、この手順 5 で紹介するような設定の作成をあらかじめ行ったうえで検索時に作成した項目を Quantitation から選択して検索を実行する必要があります。

Home -> Configuration Editor -> **Quantitation** と辿り、Quantitation 設定画面を開きます。

recommend that your server is protected by a secure firewall.

Mascot Utilities

Your system administrator may have restricted access to these utilities

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

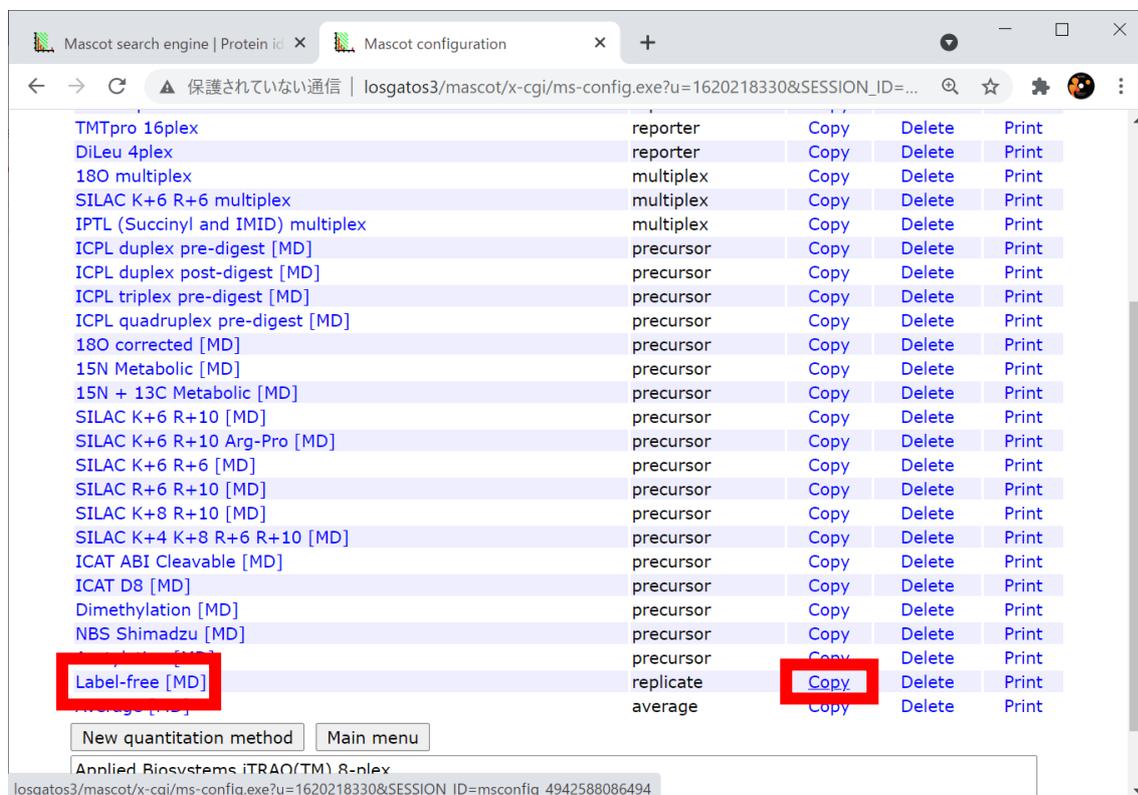
Search Log
A tabular view of the search log. Can be filtered to find specific search results.

Configuration Editor
Browser-based configuration editors, including security administration and database management.

Mascot Configuration

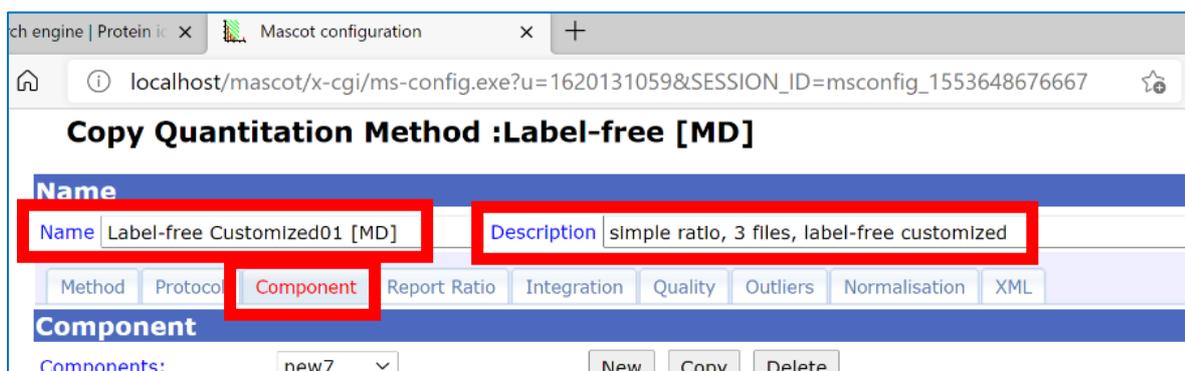
Amino Acids	Amino Acid Data
Modifications	Modification definitions
Symbols	Symbols used in chemical formulae
Linkers	Linker definitions
Enzymes	Enzyme definitions
Instruments	Fragmentation Rules
Quantitation	Quantitation Methods
Crosslinking	Crosslinking Methods
Configuration Options	Global Options in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

定量設定画面が現れます。既に MASCOT で使用可能な定量設定が準備されています。今回は Replicate 手法の基本である”Label-free[MD]”項目をカスタマイズして使用する事を目的としているため、”Label-free[MD]” 行にある “Copy” をクリックします。



Label-free[MD]設定の複製が現れます。まず、name と description を書き換えます。今回の例では name を”Label-free Customized01[MD]”, description は既存のもの後ろに”customized”という文字を追加しました。

続いて、”Component”タブをクリックします。



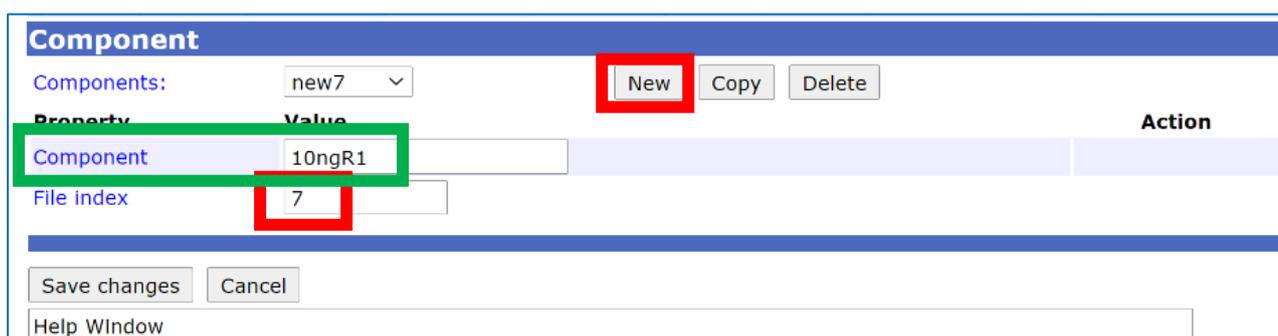
デフォルトでは C1,C2,Ref という3つの設定が作成されています。

MASCOT Server + Distiller を使った Replicates の定量計算では、計算したファイル数(サンプル数)と、この”Component”タブで定義された Component **設定数が一致している時のみ**、”Component”と”Report Ratio” タブで**設定した内容が結果に反映されます**。設定数が一致していない場合は、”Ref”, ”C1”, ”C2”・・・(データ数に準じた数の C_{n-1}) などの記号が自動的に割り振られ、すべての C_{n-1} の Ref に対する Ratio が結果に表示されます。

ここでは、12 の raw ファイルに対応する **component 名**を以下のように設定する事とします。

File 名	Sample:Ecoli spike 量	File index	名称
QEP1_SpikeIn_230914_1_3ng_270914.raw	3 ng	1	3ngR1
QEP1_SpikeIn_230914_2_3ng_270914.raw		2	3ngR2
QEP1_SpikeIn_230914_3_3ng_270914.raw		3	3ngR3
QEP1_SpikeIn_230914_4_7-5ng_270914.raw	7.5 ng	4	7p5ngR1
QEP1_SpikeIn_230914_5_7-5ng_270914.raw		5	7p5ngR2
QEP1_SpikeIn_230914_6_7-5ng_270914.raw		6	7p5ngR3
QEP1_SpikeIn_230914_7_10ng_270914.raw	10 ng	7	10ngR1
QEP1_SpikeIn_230914_8_10ng_270914.raw		8	10ngR2
QEP1_SpikeIn_230914_9_10ng_270914.raw		9	10ngR3
QEP1_SpikeIn_230914_10_15ng_270914.raw	15 ng	10	15ngR1
QEP1_SpikeIn_230914_11_15ng_270914.raw		11	15ngR2
QEP1_SpikeIn_230914_12_15ng_270914.raw		12	15ngR3

各 Component について、”Component”で名称を作成し、”File index”で取り込まれた raw ファイルの順番と Component との紐づけを行います。この”File index”は重複しないようにしてください。一つずつ作成し、新たな設定を作成するために”New”(あるいは Copy)ボタンを押して、作成を繰り返します。



すべて作成すると、Component が以下のように12個作成されます。作成後、”Report Ratio”タブをクリックします。

Copy Quantitation Method :Label-free

Name

Name Description

Method Protocol **Component** **Report Ratio** Integration Qu

Component

Components:

Property

Component

File index

Help Window

3ngR1

3ngR2

3ngR3

7p5ngR1

7p5ngR2

7p5ngR3

10ngR1

10ngR2

10ngR3

15ngR1

15ngR2

15ngR3

続いて、Distiller で表示される Ratio の内容を定義します。

表示される ratio の設定については、各 Component 名を変数とした単純な四則演算の結果のみ表示可能です。また必ず分子・分母にくる Component を指定する必要があります。

MASCOT Server の protein ratio 計算方法にもいくつかのオプションがあるものの、ご自身で細かい式を指定しつつ ratio を計算する事は Distiller 内で行う事ができません。そのような計算をご希望の方は、本資料の手順 10 で紹介するようなデータの export 機能を使用すると、タンパク質単位でなくペプチド単位で定量値(ratio でなく intensity から計算された単一の数値)が出力されますので、その数値を使ってご自身で別のソフトウェアを使って protein ratio や各種検定の計算を行ってください。

ここでは各サンプルについて、3つの繰り返し実験の単純な平均値を使った ratio を計算する、というイメージで、 $(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(3\text{ngR1}+3\text{ngR2}+3\text{ngR3})$ といったような項目を作成します。

設定する ratio は以下の3つです。

- 15ng と 3ng の ratio : $(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(3\text{ngR1}+3\text{ngR2}+3\text{ngR3})$
- 15ng と 7.5ng の ratio : $(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(7.5\text{ngR1}+7.5\text{ngR2}+7.5\text{ngR3})$
- 15ng と 10ng の ratio : $(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(10\text{ngR1}+10\text{ngR2}+10\text{ngR3})$

Report Ratio で計算式を記入します。”Report Ratio”項目は単なる名称ではなく計算式となりますので、変数となる Component 名も含めて正確に記述してください。

例: $(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(3\text{ngR1}+3\text{ngR2}+3\text{ngR3})$

Report ratio で指定した変数名と、Component を紐づけする設定を行う必要もあります。”Add Numerator” , ”Add Denominator” ボタンを駆使して Component 数に合わせて項目を呼び出すとともに、選択肢から Component を呼び出して、はっきりと定義する必要があります。

*Coefficient は係数です。設定数値が Components の定量値に掛けられます。

詳しい設定内容は以下図をご覧ください。

Property	Value	Action									
Report Ratio	$(15\text{ngR1}+15\text{ngR2}+15\text{ngR3})/(3\text{ngR1}+3\text{ngR2}+3\text{ngR3})$	Delete Report Ratio									
Numerator	<table border="1"> <tr> <td>15ngR1</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> <tr> <td>15ngR2</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> <tr> <td>15ngR3</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> </table>	15ngR1	Coefficient : 1.0	Delete	15ngR2	Coefficient : 1.0	Delete	15ngR3	Coefficient : 1.0	Delete	Add numerator
15ngR1	Coefficient : 1.0	Delete									
15ngR2	Coefficient : 1.0	Delete									
15ngR3	Coefficient : 1.0	Delete									
Denominator	<table border="1"> <tr> <td>3ngR1</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> <tr> <td>3ngR2</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> <tr> <td>3ngR3</td> <td>Coefficient : 1.0</td> <td>Delete</td> </tr> </table>	3ngR1	Coefficient : 1.0	Delete	3ngR2	Coefficient : 1.0	Delete	3ngR3	Coefficient : 1.0	Delete	Add denominator
3ngR1	Coefficient : 1.0	Delete									
3ngR2	Coefficient : 1.0	Delete									
3ngR3	Coefficient : 1.0	Delete									

作成予定の3つの ratio について設定を作成します。各 Report Ratio はすべて表示できなかったため、以下画像内の黄色い枠の中に設定内容を記しました。

Report Ratio			
Property	$(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)$		
Report Ratio	$(15ngR1+15ngR2+15ngR3)$		
Numerator	15ngR1	Coefficient : 1.0	Delete
	15ngR2	Coefficient : 1.0	Delete
	15ngR3	Coefficient : 1.0	Delete
Denominator	3ngR1	Coefficient : 1.0	Delete
	3ngR2	Coefficient : 1.0	Delete
	3ngR3	Coefficient : 1.0	Delete
<hr/>			
Property	$(15ngR1+15ngR2+15ngR3)/(7p5ngR1+7p5ngR2+7p5ngR3)$		
Report Ratio	$(15ngR1+15ngR2+15ngR3)$		
Numerator	15ngR1	Coefficient : 1.0	Delete
	15ngR2	Coefficient : 1.0	Delete
	15ngR3	Coefficient : 1.0	Delete
Denominator	7p5ngR1	Coefficient : 1.0	Delete
	7p5ngR2	Coefficient : 1.0	Delete
	7p5ngR3	Coefficient : 1.0	Delete
<hr/>			
Property	$(15ngR1+15ngR2+15ngR3)/(10ngR1+10ngR2+10ngR3)$		
Report Ratio	$(15ngR1+15ngR2+15ngR3)$		
Numerator	15ngR1	Coefficient : 1.0	Delete
	15ngR2	Coefficient : 1.0	Delete
	15ngR3	Coefficient : 1.0	Delete
Denominator	10ngR1	Coefficient : 1.0	Delete
	10ngR2	Coefficient : 1.0	Delete
	10ngR3	Coefficient : 1.0	Delete
New Report Ratio			
<div style="border: 1px solid red; display: inline-block; padding: 2px;">Save changes</div> Cancel			

設定後 **Save changes** ボタンを押すと今回設定した全体の定義が保存されます。

[参考・タンパク質を使った Normalization について]

以下設定内容は normalization を希望される方のみご利用ください。

参照論文では 変動のない human のタンパク質、特に値が真ん中程度にあるタンパク質 20 個を全体データの補正・Normalization に利用しています。この資料では使用例として、参照論文にリストアップされたタンパク質 20 を使って normalization する設定をご案内いたします。

参照論文が利用したタンパク質 20 については、

<https://pubs.acs.org/doi/10.1021/pr501045t>

supporting info にある 3 つめの補足資料(excel ファイル)にある定量結果表をご参照ください。

表の中に” House keeping protein? (1=yes, 0=no)”というタイトルの列があり、行の並びの真ん中あたりに この設定項目が 1 になっているタンパク質が normalization に使用されたタンパク質です。

今回の計算で同様にタンパク質による normalization を希望される場合、下図のようにして

「**Normalization**」タブ内で、**使用するタンパク質の accession を指定してください。**

なお、この normalization はここで必ずしも設定しなくとも結構です。例えば一度計算を実行後、その結果を見た上で改めて normalization に使用するタンパク質を設定して計算させることも可能です。

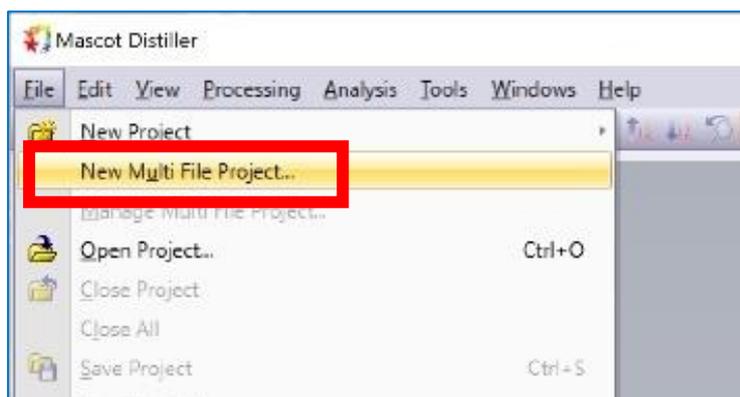
Property	Value	Action
Normalisation method	average	
Normalisation Proteins	sp P78406 RAE1L_HUMAN	Delete
	sp O14561 ACPM_HUMAN	Delete
	sp P15924 DESP_HUMAN	Delete
	sp Q8NBJ5 GT251_HUMAN	Delete
	sp P10606 COX5B_HUMAN	Delete
	sp O43242 PSMD3_HUMAN	Delete
	sp P63151 2ABA_HUMAN	Delete
	sp O00233 PSMD9_HUMAN	Delete
	sp Q9Y3C8 UFC1_HUMAN	Delete
	sp P24752 THIL_HUMAN	Delete
	sp O75153 CLU_HUMAN	Delete
	sp Q96CT7 CC124_HUMAN	Delete
	sp Q99426 TBCB_HUMAN	Delete
	sp Q03252 LMNB2_HUMAN	Delete
	sp Q9Y5K5 UCHL5_HUMAN	Delete
	sp Q4VC31 CCD58_HUMAN	Delete
	sp P68400 CSK21_HUMAN	Delete
	sp P12081 SYHC_HUMAN	Delete

すべての設定が終わったら、画面下部にある”Save changes”ボタンを押して設定を終了してください。

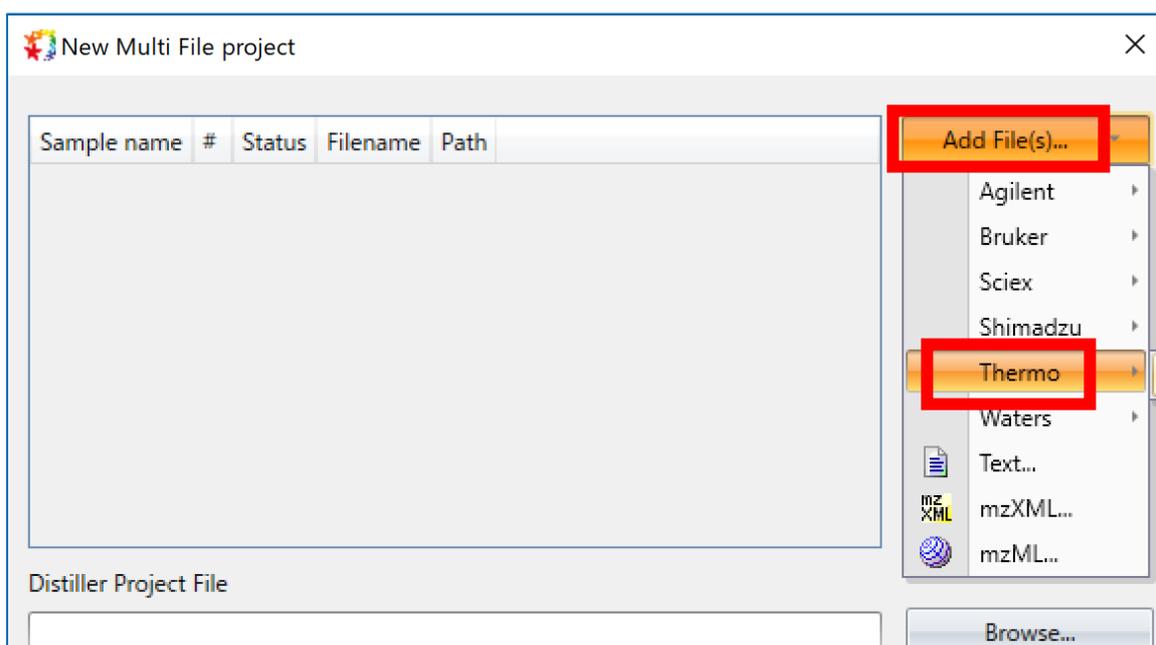
6. 定量計算実行 1・データ読み込み

手順 2～5 で準備が整ったところで実際に定量計算を行います。手順 6 では Distiller 上で raw データを読み込む操作を行います。

まず MASCOT Distiller を起動します。続いて、Menu の File から”New Multi File Project”を選択します。

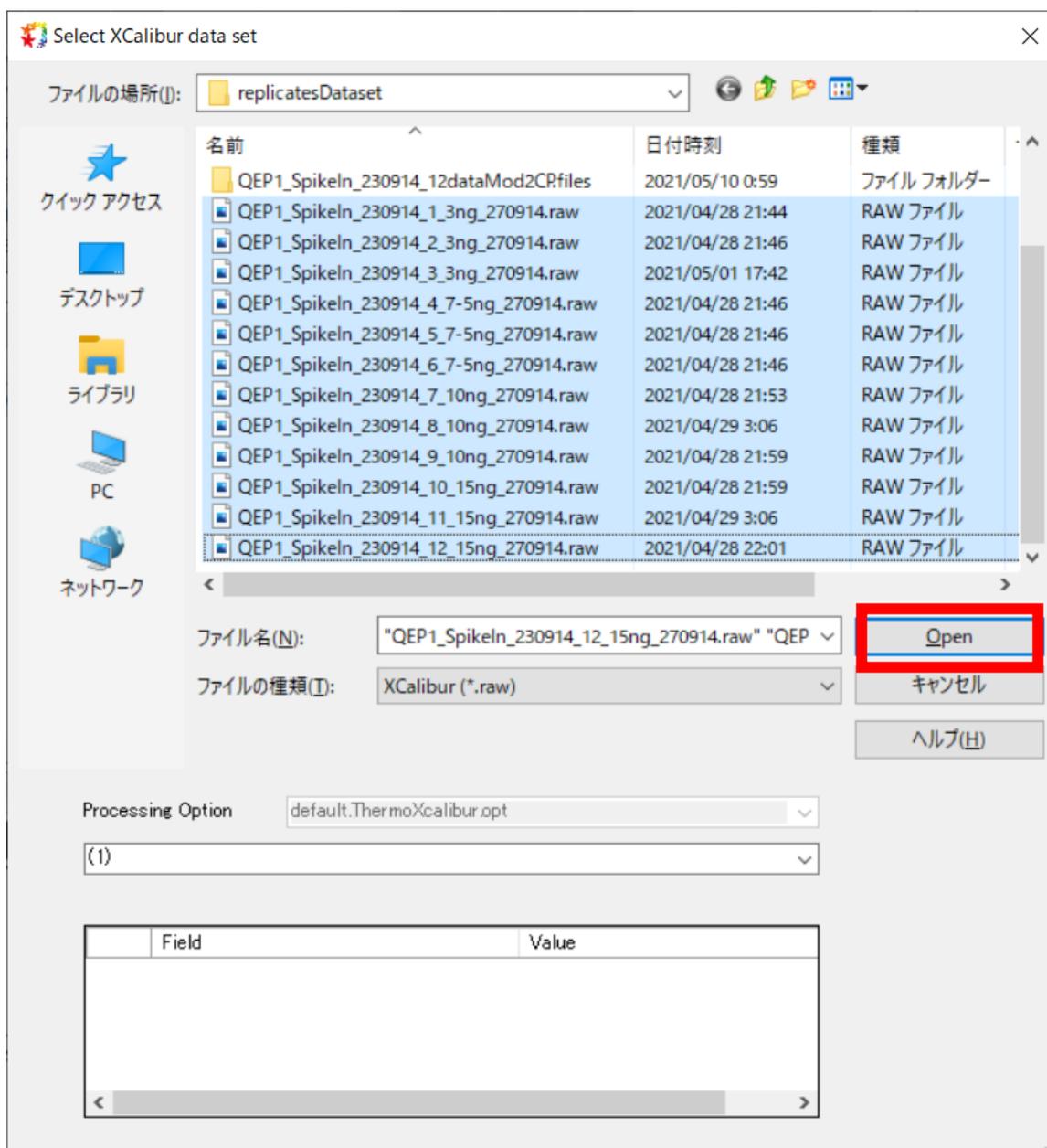


ダイアログ右側の **Add Files** をクリックします。ここで、raw データに合わせたメーカー名並びにソフトウェア名を選択します。今回の PXD001385 のデータは、ThermoFisher Scientific 社の装置で測定されたデータです。Add Files -> Thermo -> Xcalibur を選択します。



12 の raw ファイルを選択します。定量設定時に定めた File index はファイルの読み込み順と対応しています。後で再調整する必要が無いよう、Components の index 番号と raw ファイルの読み込み順が同じになるよう、raw ファイルを選択して取り込むのが望ましいです。しかしながら対応付けに失敗しても後で変更可能ですので、よくわからない場合は順番を気にせず選んでいただければ OK です。

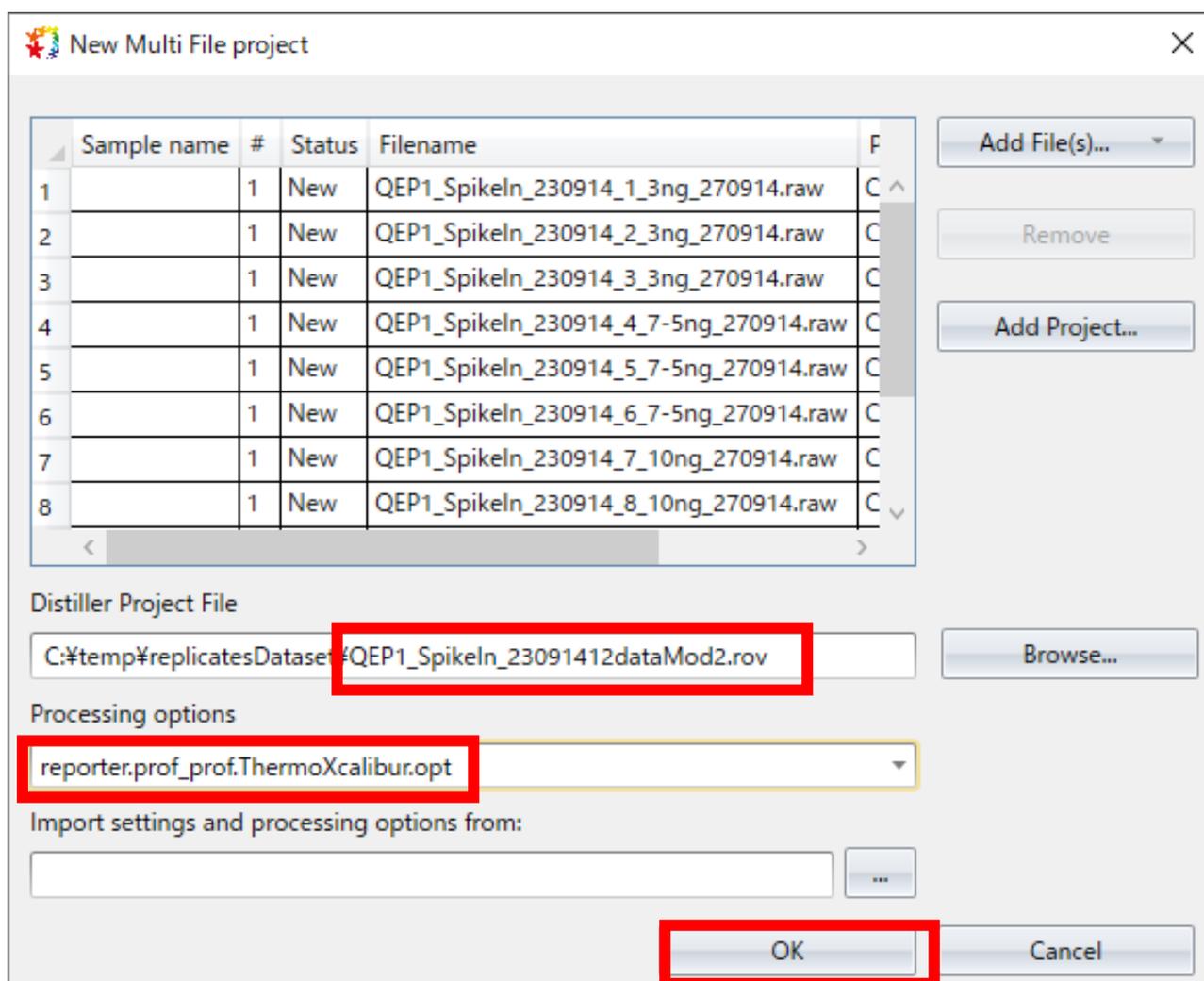
ファイルを選択して”Open”ボタンを押します。



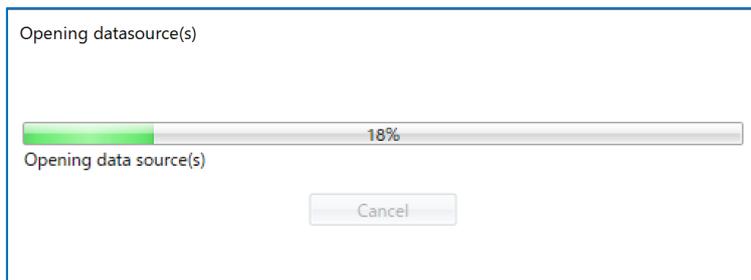
元のダイアログに戻ります。さらに2か所の変更を行います。

1か所目は **Distiller Project File**, 全体ファイル名です。Distiller project file (拡張子が .rov のファイル) ですが、デフォルトでは読み込んだ最初の raw ファイルに基づいて名づけられており、混同してしまう事も多いので変更されることをお勧めします。今回は raw ファイル名の後半部分にあたる個所を削除し各データに共通する個所を残した上で、"12dataMod2"という文字を加えています。

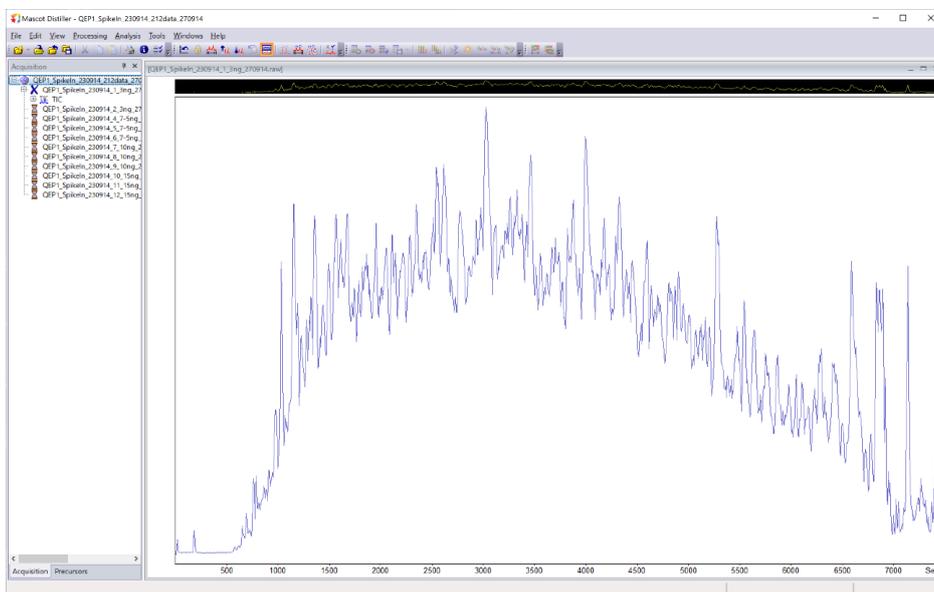
2か所目は **"Processing Options"** です。Distiller では各装置・各手法に合わせてデータ処理に適したデフォルト設定値を準備していますが、それを適用するためにデータ読み込み時に opt ファイルを選択します。装置・測定手法に近い名称の opt ファイルを選択してください。今回の例では多少名称と実行内容が合いませんが、**reporter.prof.prof.ThermoXcalibur.opt** を選択しています。



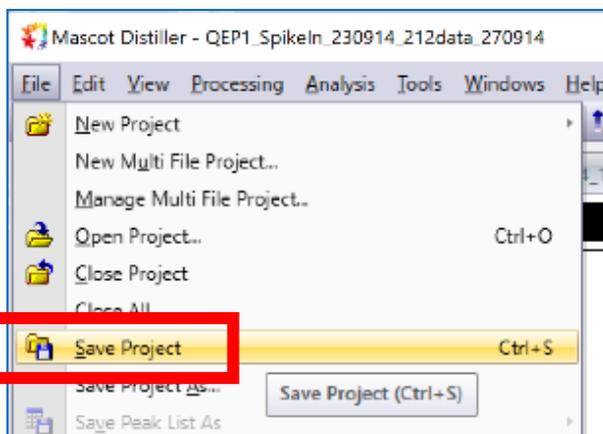
項目選択後、"OK"ボタンを押すとファイルを開き始めます。



ファイル読み込みが終了すると TIC 画面が表示されます。



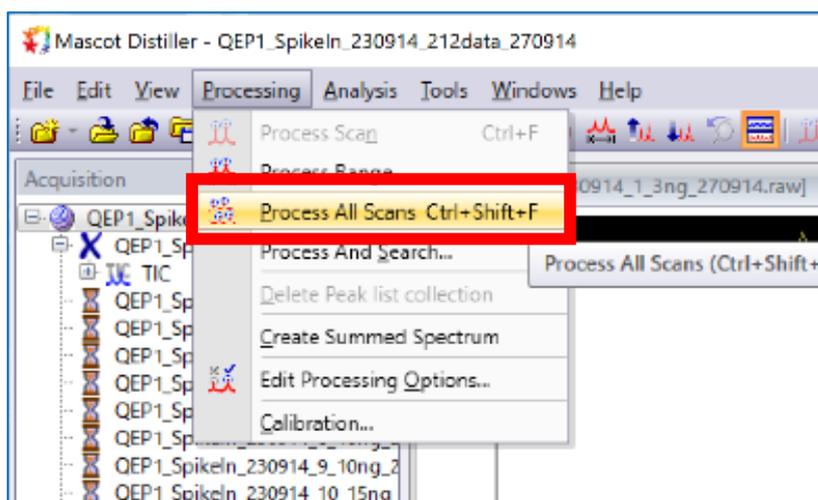
各ステップにおいて、次のステップに進む前にデータを Save しておくことをお勧めいたします。



7. 定量計算実行 2・ピーク抽出

読み込んだ各 raw データに対してピーク抽出処理を行います。MASCOT Server 側では raw データをそのまま受け付ける事はできません。検索前にスペクトルからノイズを除去し、ペプチド/フラグメントのピークに該当する箇所を抜き出してまとめた入力データを作成する必要があります。その処理の事をこの資料では「ピーク抽出」と呼んでいます。

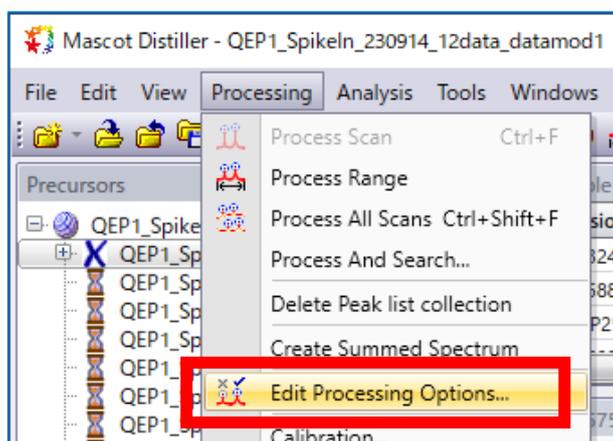
Menu の Processing から”**Process All Scans**”を選択すると、すべての raw データに対してピーク抽出を行います。



[関連項目 : Peak Processing option について]

以下は希望される方だけの操作です。ピーク抽出はあらかじめ定められた設定をもとに計算されます。設定は、データ取り込み時の opt ファイル内に定められていますが、データ読み出しの後に変更する事ができます。設定内容並びに設定を変更する方法をご紹介します。必要に応じて実行してください。

Menu の Processing から”**Edit Processing Options**”を選択してください。



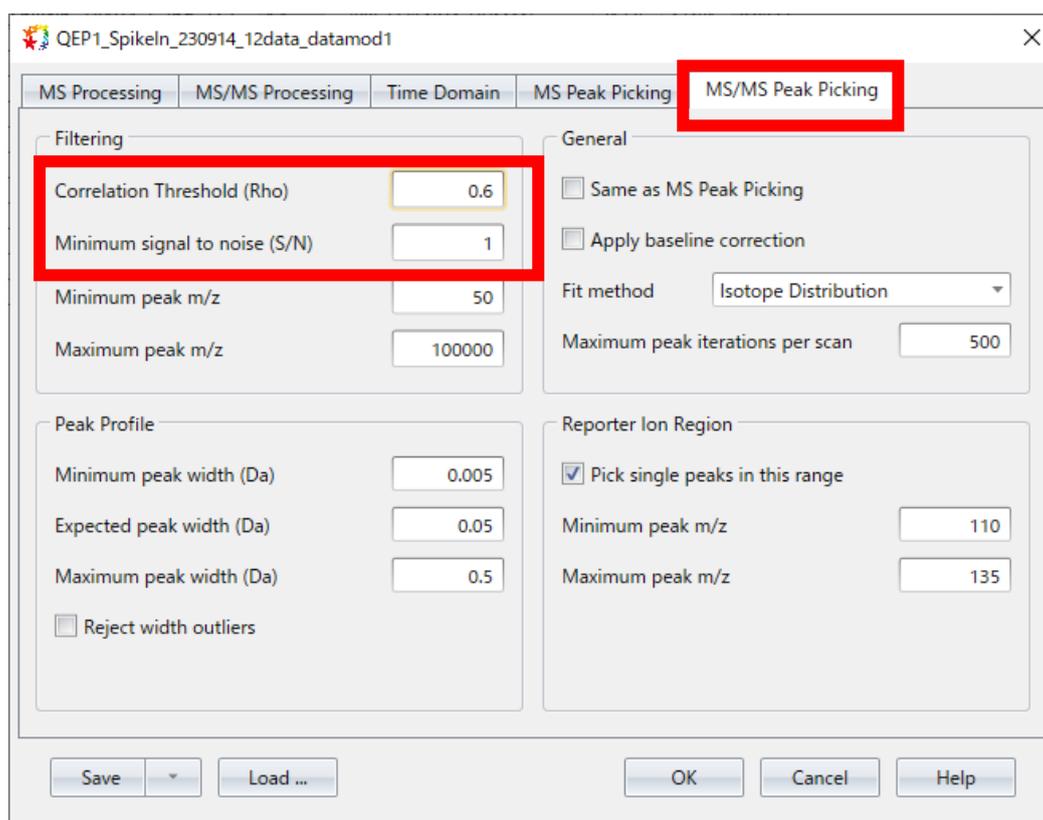
Processing options は5つのタブから構成されています。

タブ	設定内容
MS Processing	MS スペクトルデータの特性・データの質について
MS/MS Processing	MS/MS スペクトルデータの特性・データの質について
Time Domain	溶出時間と関連した、同一ペプチドに対する認識と処理
MS Peak Picking	MS のピークの認識と選別(抽出)条件
MS/MS Peak Picking	MS/MS のピークの認識と選別(抽出)条件

各タブに設定項目の詳細については、ダイアログ右下の”Help”をご覧ください。

ここでは抽出ピークの調整に特にわかりやすく影響するパラメータ 2 つについてのみピックアップしてご紹介します。抽出ピークの内容が変わると検索結果も変わることがあります。

MS/MS Peak Picking タブをご覧ください。



Correlation Threshold: 理論同位体スペクトルと実測データの類似度、相関係数。1 に近いと理論値の形状に似ていることが求められ、0 に近ければ似ていなくてもよくなります

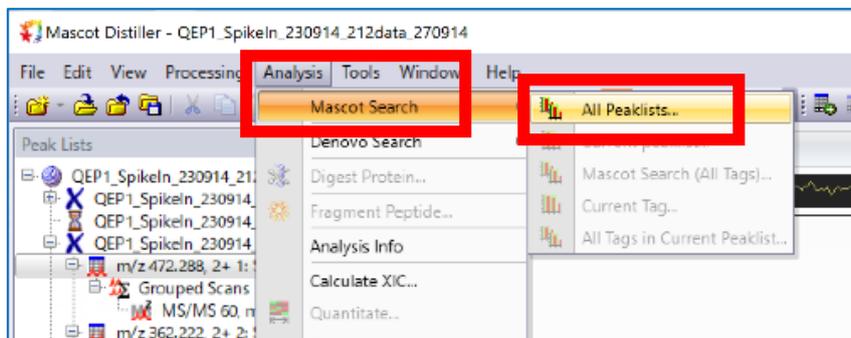
Minimum signal to noise (S/N) : S/N。値が低いほど選定条件が緩く対象が広がります。

条件を定めたら”Save”をして、Process Peak Scan を実行してください。

8. 定量計算実行 3・MASCOT 検索と定量計算前の設定変更

ピーク抽出が終わりでしたら、引き続き MASCOT Server での検索を行います。

Distiller の menu Analysis -> Mascot Search -> **All Peaklists** を選択してください。



検索パラメータ入力画面が現れます。今回は以下画像のように各種パラメータを設定しました。

特にご注意くださいなのが”Quantitation”項目です。**手順 4 にてあらかじめデータに合った Quantitation 設定を作成した場合はその作成項目を、そうでない場合は”Label Free[MD]”**を選択してください。設定値は以下の通りです。設定内容は次頁図をご覧ください。

設定項目	設定値	備考
Database	PXD001385	手元のデータベースを使用してもよい
Allow up to	1	独自に判断した数値。元の論文データ投稿サイトには指定がない。「2」でも良い
Quantitation	Label free Customized[MD]	上述の通り
Fixed Modification	Carbamidemethyl(C)	
Variable Modification	Oxidation(M)	
Peptide tol.	10 ppm	論文やサイトで使用されている数字
MS/MS tol.	0.02 Da	論文やサイトで使用されている数字
# ¹³ C	1	独自に判断した数値。「2」でも良い
Instrument	Default	独自に判断した設定値。装置に合わせたものに変更してもよい
Decoy	チェックを入れる	FDR 計算のため必須

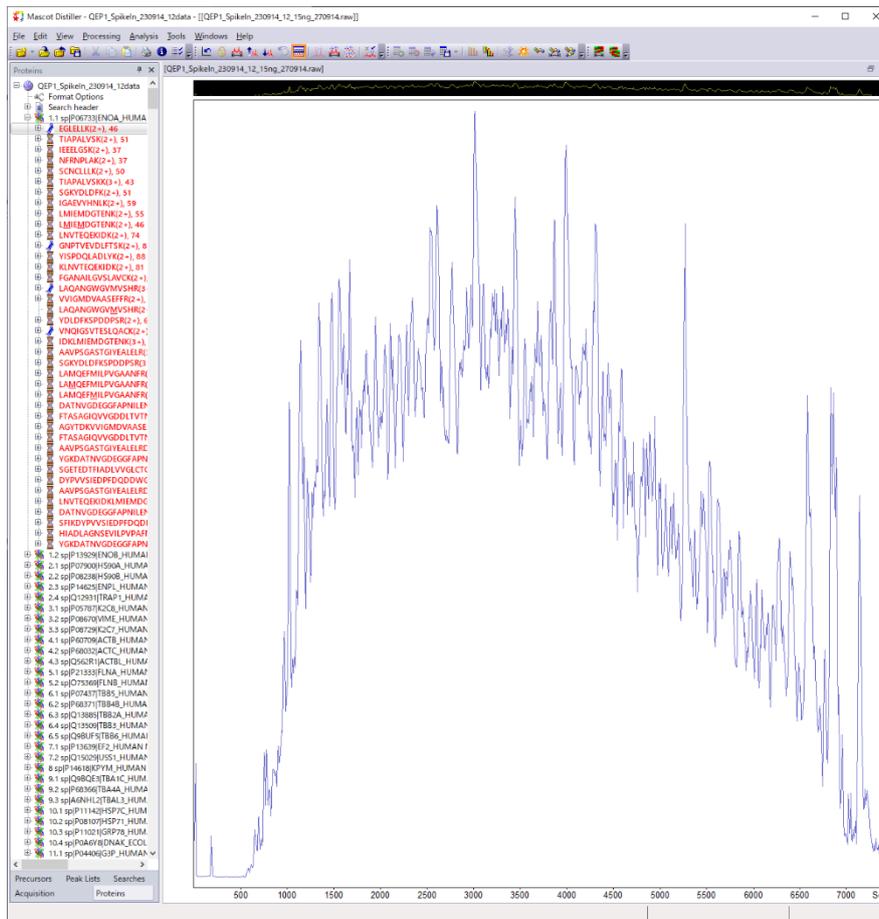
画面下部の” **Start Search**”ボタンを押すと検索を実行します。

The screenshot shows the MASCOT MS/MS Ions Search interface. The window title is "Mascot Search (localhost)". The main heading is "MASCOT MS/MS Ions Search". The interface is a form with various input fields and dropdown menus. The "Start Search ..." button is highlighted with a red box. The "Reset Form" button is also visible. The "Cancel" button is at the bottom right.

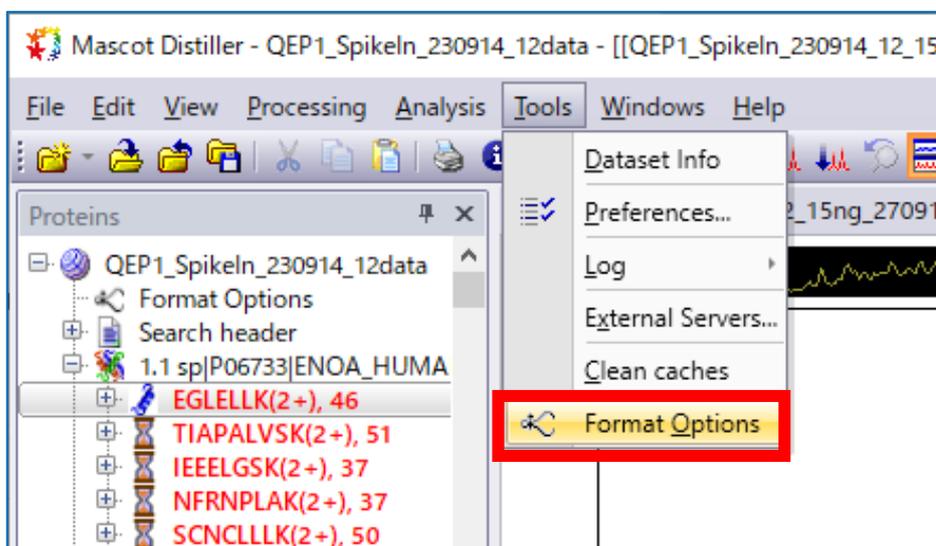
Your name	takaesu	Email	
Search title	QEP1_SpikeIn_230914_12data		
Database(s)	PXD001385 (AA)	> <	Nucleic acid (NA) Human_EST Amino acid (AA) NCBIprot SwissProt UP5640_H_sapiens Spectral library (SL) PRIDE_Contaminants
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Quantitation	Label-free Customized01 [MD]		
Crosslinking	None		
Fixed modifications	Carbamidomethyl (C)	> <	Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Carbamidomethyl (N-term) Carbamyl (K) Carbamyl (N-term) Carboxymethyl (C) Cation:Na (C-term)
	Display all modifications	<input type="checkbox"/>	
Variable modifications	Oxidation (M)	> <	
Peptide tol. ±	10 ppm	# ¹³C	1
MS/MS tol. ±	0.02 Da	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Peptide charge	2+	Precursor	m/z
Data file		Error tolerant	<input type="checkbox"/>
Data format	Mascot generic	Report top	AUTO hits
Instrument	Default		
Decoy	<input checked="" type="checkbox"/>		
	Start Search ...		Reset Form
			Cancel

[次頁に続きます]

検索が終わると、Distiller の左フレームに同定タンパク質一覧が表示されます。



続いて、同定タンパク質の選定条件について変更したり、定量計算の基準を変更したりします。Menu の Tools から”Format Options”を選択します。



Format Options は3つのタブから構成されています。

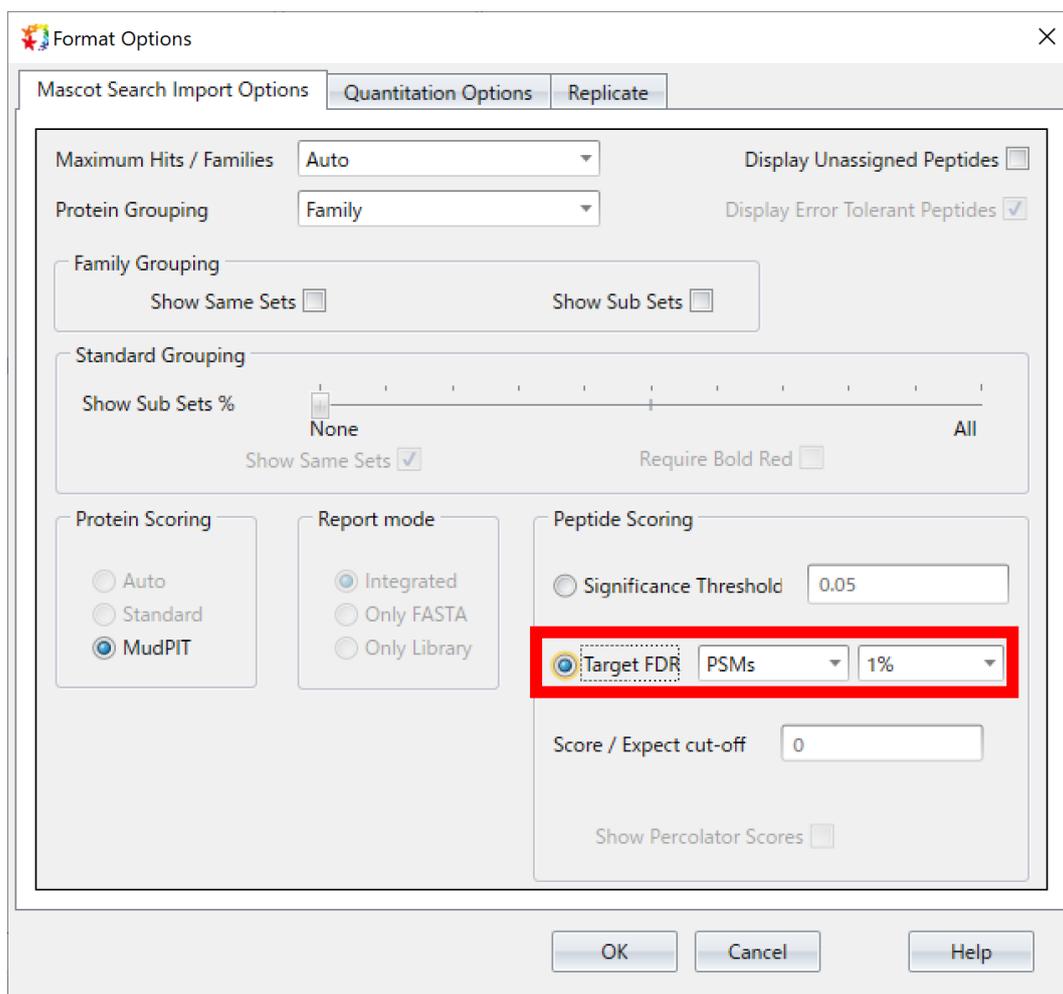
- Mascot Search Import Options
- Quantitation Options
- Replicates (定量手法 Replicates の時のみ現れます)

以降、各タブ内の設定について補足説明をします。

MASCOT Search Import Options

ファミリータンパク質や subset/sameset タンパク質の表示などシェアペプチドに関連した内容、並びに同定タンパク質/同定ペプチドの基準を設定することができます。

参照論文では ペプチドの同定基準として FDR 1%の設定を行っていました。この資料でも同様の設定をいたします。”Peptide Scoring”の選択肢で **“Target FDR”を、PSMsの1% と設定**しました。“PSMs” の選択肢には他に”Sequences”もあります。同一ペプチドにマッチした query に対するカウントが異なり、Sequence の場合同一ペプチドにマッチしている query を1つにまとめてカウントします。お好みの方をご利用ください。(ご利用の際、選択内容を明記すればどちらも使用が認められています)。



Quantitation Options

MASCOT Server で行った Quantitation の設定内容を引き継いでおり、ご希望により変更が可能です。以下、計算に関連が深い項目に関してのみ簡単な説明を記します。各項目の詳細は”Help”ボタンを押してご確認ください。

Format Options

Mascot Search Import Options Quantitation Options Replicate

Family range

All Families

All protein families from PXD001385

Range

Enter Family numbers and/or Family number ranges separated by commas. For example 3,4,6-10,13

Quality

Correlation Threshold 0.8

Std. Err. Threshold 999

Fraction Threshold 0.5

Protein Ratio Type median

Normalisation average

Outlier removal auto

Peptide threshold

Type at least homology

Value 0.05

Changing any of these options will require the Quantitation Report to be re-generated. All Options

OK Cancel Help

“Quality”

XIC 計算の際、該当ペプチドの Precursor ピークを探索したり前後の時間に探索範囲を拡張したりする際に利用するパラメータです。「手順 9. 定量計算実行4・定量計算実行と、結果表示内容について」(P. 36~) も併せてご参照ください。

Correlation Threshold

理論スペクトルと実測スペクトルの同位体クラスター形状に関してマッチングを行った際の相関係数です。1に近いほど実測スペクトルが理論スペクトルの形状に近い必要があり、逆に 0 に近いほど形状が離れていても問題ありません。1に近いほど厳しく、0に近いほど緩い基準となります。

Std Err.Threshold

各時間の ratio について、最小二乗法フィッティングの直線との差(標準誤差)です。前頁図の設定値のように”999”などであれば実質このパラメータは利用しないことを意味しています。

Fraction Threshold

ペプチドの同位体クラスターピークが存在する m/z 領域に、それ以外のピークが存在するかどうかを表す指標です。チェックの有無でパラメータとして使用するかどうかが変わります。

Fraction とは (同位体クラスターピークの面積) / (領域全体のピークの面積) の事を意味します。値が 1 に近いほどペプチドの同位体クラスターピーク以外の存在を許さず、逆に 0 に近いほど他のピークを許容します。

ここで表示されている項目以外にもパラメータは存在します。詳細は”All Options”ボタンを押して内容をご確認ください。また MASCOT の HELP ページ、Quantitation の”Configuration”の項目も併せてご覧ください。

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

“Protein Ratio Type”

タンパク質にアサインされたペプチドの ratio から、タンパク質の ratio をどのように計算するかについて定義します。average, median, weighted(peptide の intensity をもとにした重みづけを加味した average)の3種類があります。

“Normalisation”

ペプチドの ratio を定めた基準を使って正規化します。

“Outlier removal”

ratio について外れ値を検出し処理をする方法について選択します。

Normalization や Outlier removal に関する設定については、以下 URL も併せてご参照ください。

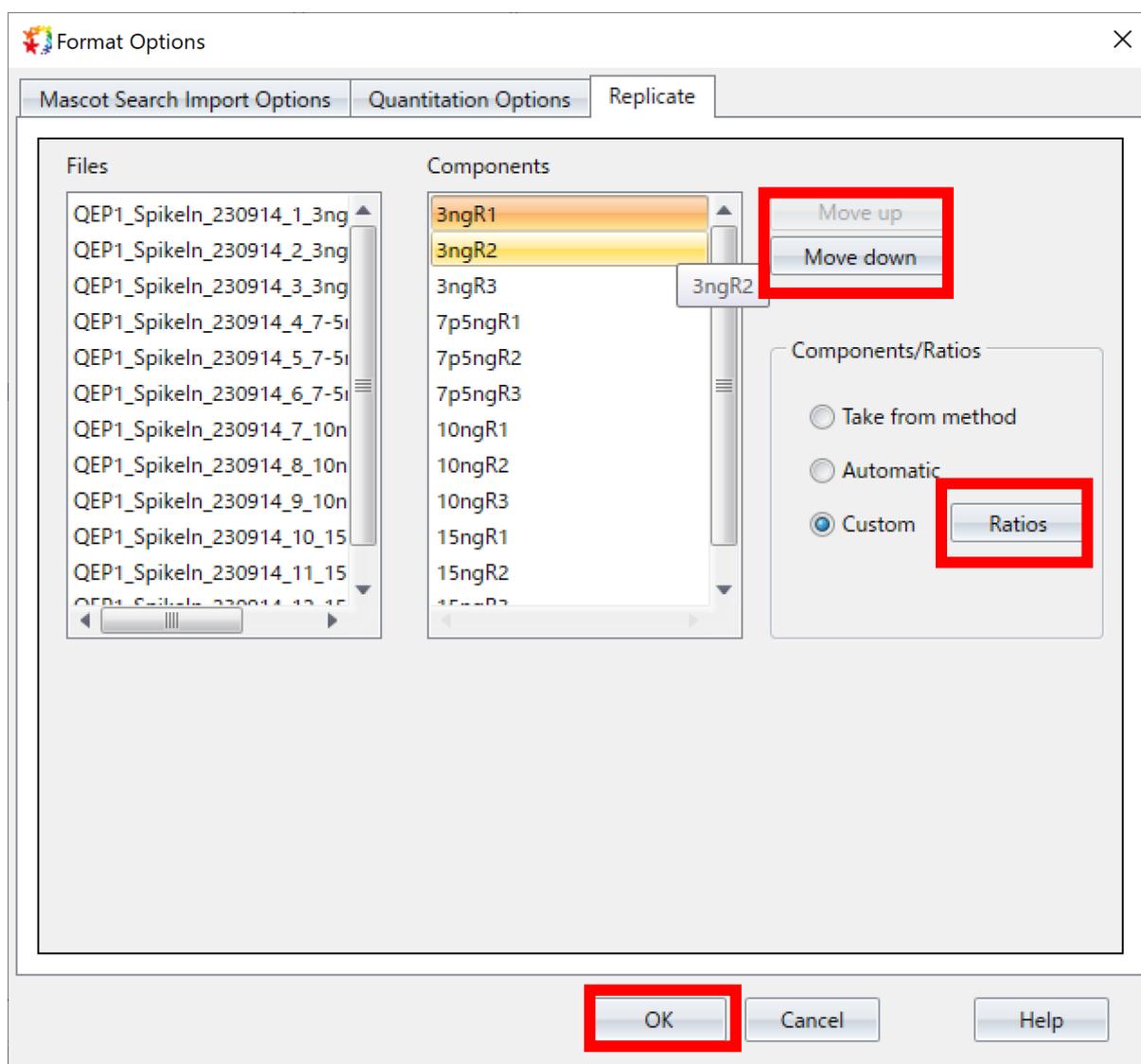
http://www.matrixscience.com/help/quant_statistics_help.html

Replicate

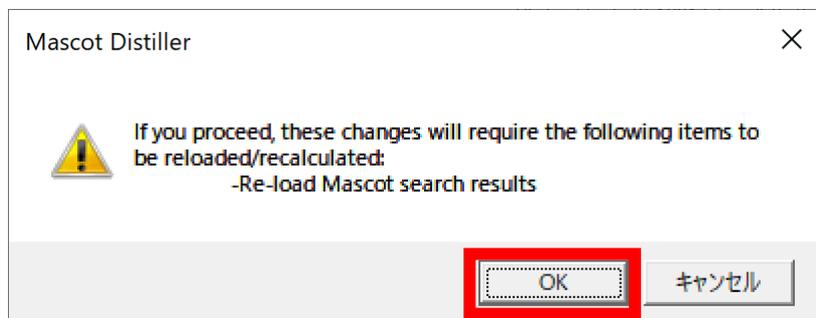
raw ファイルと component の紐づけを確認できるほか、表示する ratio についても定義を確認したり変更したりする事ができます。

Files で表示されている raw ファイル名と、Components の名称が正しく対応付けられているか確認してください。紐づけに意図と異なる個所がある場合、右側の”Move up” ”Move down”を押して Components の並び順を変え、対応内容を変更してください。

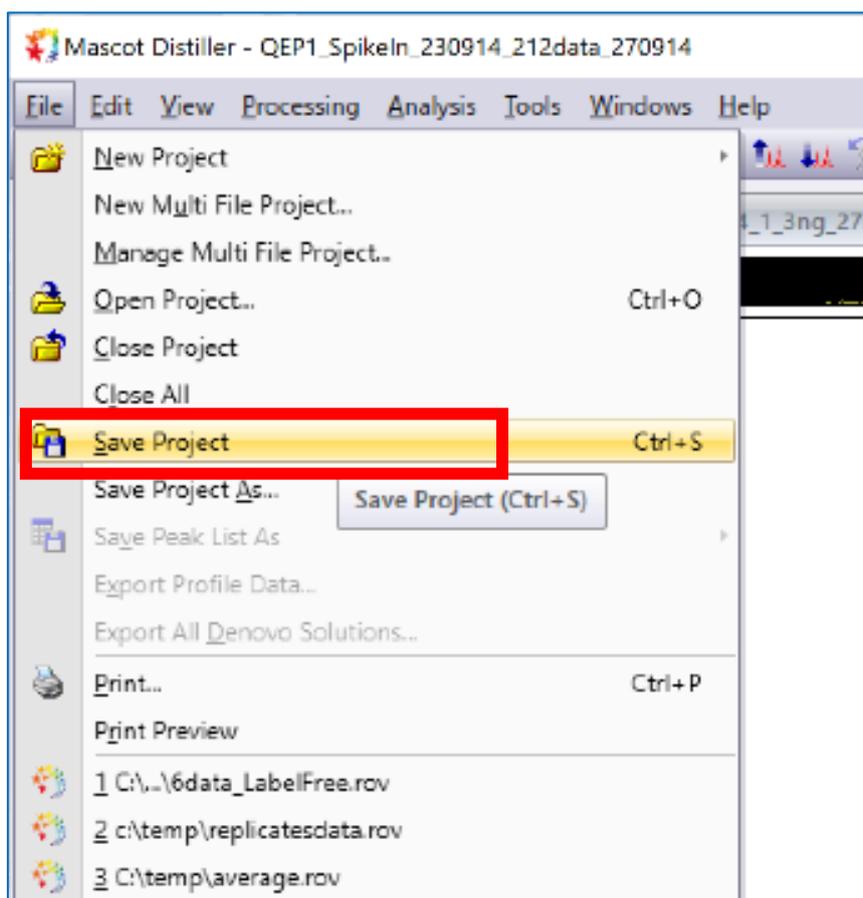
また Custom ボタンの隣、”Ratios”ボタンを押すと、表示される Ratio を変更する事ができます。設定方法は MASCOT Server の Quantitation で行った”Report Ratio”と同じです。詳しくは手順 4 をご覧ください。



すべての設定を変更して”OK”ボタンを押すと、結果内容の再度取り込みを実行する旨警告文が現れます。特に問題ありませんのでそのまま”OK”ボタンを押してください。引き続き、同定結果を取得し直す処理が行われます。



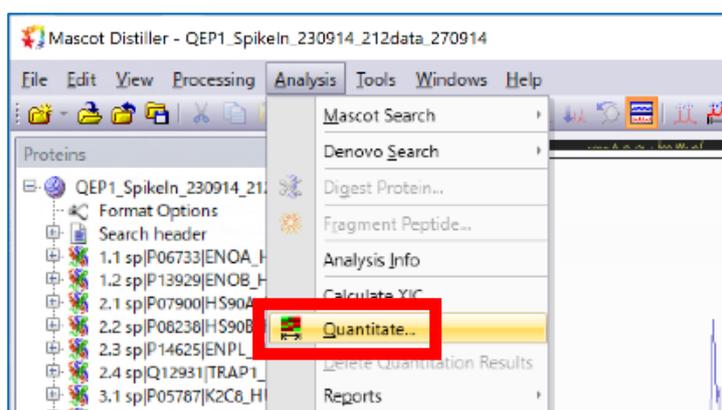
この後実行する定量計算にはとても時間がかかります。また場合によっては途中で失敗してソフトウェアが落ちてしまう事もあります。定量計算を行う前に、必ず **project ファイル(.rov ファイル)** を保存するようにしてください。



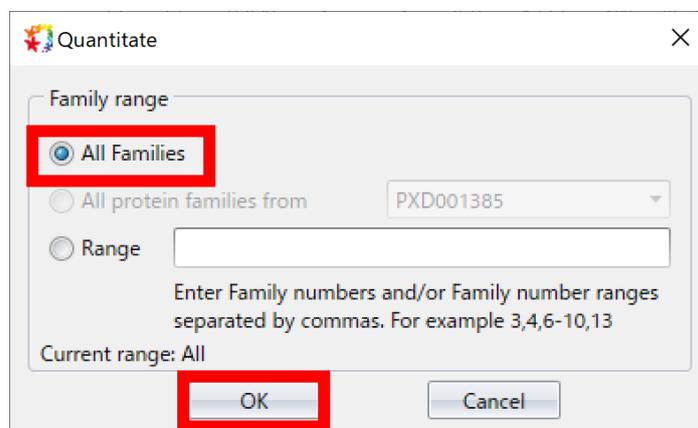
9. 定量計算実行 4・定量計算実行と、結果表示内容について

同定ペプチドの情報をもとに各 raw データで XIC 計算を行い、さらにアサインされているペプチドの情報から各タンパク質の定量数値を計算します。

menu の Analysis で、”Quantitate”を選択します。



Quantitate ダイアログが現れます。計算対象となるタンパク質を選択します。ここではすべてのタンパク質で計算を行うという事で、”All Families”を選択し、”OK”ボタンを押します。



定量の計算が実施され、画面中央～右側に定量結果が表示されます。

[計算時間とコンピュータのスペックに関連する補足]

replicates 定量計算には非常に時間がかかります。今回のサンプルデータ (4x3=12 の raw ファイル)において、弊社推奨のスペックを持つコンピュータでも **24~48 時間程度**かかります。ただし ver.2.8 へのアップグレードにより従来よりも計算に必要なメモリ量は大幅に減っています。このデータについては 30GB 程度のメモリを搭載したコンピュータであれば 問題なく計算が実行可能です。

以降、定量計算結果の検証に関連する事項について説明いたします。

表示された結果の一番上、“**Quantitation Table**”は、タンパク質の定量について、ユーザーが独自に定めた、あるいは自動的に定められた比が一覧に表示されます。

Quantitation table でタンパク質を選択すると、その下に“**Matches Table**”が現れます。“**Matches Table**”には、Quantitation table で選択したタンパク質にアサインされているペプチドについて、ratio やそれに関連する数字と共に表示されます。

さらに“**Matches Table**”でペプチドを選択すると、Distiller にて“**TIC Window**”と呼んでいる、TIC 並びに該当のペプチド部分の XICs が全サンプルを対象に表示されます。

The screenshot displays the MASCOT Distiller interface with three key components highlighted in red boxes:

- Quantitation Table:** A table with columns for Accession, Score, Mass, and various protein identifiers. It lists identified proteins and their corresponding scores and masses.
- Matches Table:** A table showing peptide sequences (e.g., RGLELEK, TIAPALVSK) and their associated ratios and scores across different samples.
- TIC Window (XICs 情報含む):** A chromatogram showing Total Ion Chromatogram (TIC) and extracted ion chromatograms (XICs) for the selected peptides, plotted against retention time.

3つの表並びにグラフについて、さらに詳しく説明します。

[Quantitation table]

	Accession	Score	Mass	(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)	SD(geo)	#	(15ngR1+15ngR2+15ngR3)/(7p5ngR1+7p5ngR2+7p5ngR3)	(15ngR1+15ngR2+15ngR3)/(10ngR1+10ngR2+10ngR3)
<input checked="" type="checkbox"/>	1.1 sp P06733 ENOA_HUMAN	33306	47481	1.1327	1.0902	34		
<input type="checkbox"/>	1.2 sp P13020 ENOA_HUMAN	9940	47299	1.1704	1.1012	5		
<input type="checkbox"/>	1.3 sp P13020 ENOA_HUMAN	18698	85006	1.1109	1.1024	45		
<input type="checkbox"/>	2.2 sp P08238 HS90B_HUMAN	17983	83554	1.1230	1.0404	45		
<input type="checkbox"/>	2.3 sp P14625 ENPL_HUMAN	7026	92696	1.1133	1.0544	29		
<input type="checkbox"/>	2.4 sp Q12931 TRAP1_HUMAN	2003	80345	1.1043	1.0652	9		
<input type="checkbox"/>	3.1 sp P05787 K2C8_HUMAN	17641	53671	1.1180	1.1617	48		
<input type="checkbox"/>	3.2 sp P08670 VIME_HUMAN	11853	53676	1.1082	1.0447	38		
<input type="checkbox"/>	3.3 sp P08729 K2C7_HUMAN	8882	51411	1.1129	1.0392	31		
<input type="checkbox"/>	4.1 sp P07437 TB85_HUMAN	15839	50095	1.1377	1.1118	28		

各タンパク質について、あらかじめ定めた定量の計算結果 (ratio) を表示します。この資料では手順にて表示する ratio を以下の3通りに定めたため、その内容が表示されています。

- $(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)$
- $(15ngR1+15ngR2+15ngR3)/(7p5ngR1+7p5ngR2+7p5ngR3)$
- $(15ngR1+15ngR2+15ngR3)/(10ngR1+10ngR2+10ngR3)$

事前に定義していない場合、各 Component について Ref(erence) との単純な比が表示されます。

SD(geo) は幾何標準偏差です。(アサインされたペプチドの定量値から求められる) タンパク質の定量計算方法によっては表示されないことがあります。

は定量計算に用いられたペプチドデータの数です。

SD(geo) と # は設定した ratio 毎に数値が表示されます。

現在選択しているタンパク質については行の先頭にペンマーク が表示されます。選択されているタンパク質について、Quantitation table のすぐ下に "Matches table" が表示されます。

[Matches table]

z	Sequence	X	(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)	Std.Err	X	(15ngR1+15ngR2+15ngR3)/(7p5ngR1+7p5ngR2+7p5ngR3)	Std.Err	X	(15ngR1+15ngR2+15ngR3)/(10ngR1+10ngR2+10ngR3)	Std.Err	Fraction	Correlation	Intensity	Modifications
2+	YISPDQLADLYK	✓	1.1284	0.0057	✓	0.9756	0.0095	✓	1.017	0.0057	0.9777	0.9984	323306106	
4+	YGKDATNVGDEGGFAPN...	✓	1.1189	0.0149	✓	1.0040	0.0277	✓	1.055	0.0194	0.9649	0.9826	29218246	
3+	YGKDATNVGDEGGFAPN...	✓	1.1346	0.2646	✓	0.9466	0.0919	✓	1.055	0.0561	0.9116	0.9293	7663089	
3+	YLDLDFKSPDDPSR	✓	1.0623	0.1130	✓	0.9862	0.1756	✓	0.988	0.0628	0.8374	0.9756	12275965	
2+	YLDLDFKSPDDPSR	✓	1.0693	0.4321	✓	1.0086	0.2506	✓	0.9847	0.1903	0.8302	0.9374	1051777	
3+	VVIGMDVAASEFFR	✓	2.4034	0.9701	✓	1.0478	1.5315	✓	1.2112	1.3873	0.6144	0.9374	1051777	
2+	VVIGMDVAASEFFR	✓	1.1600	0.1075	✓	1.0842	0.0726	✓	1.0218	0.0728	0.5645	0.9374	1051777	
3+	VNQGVSVTSLQACK	✓	1.0870	0.2995	✓	0.8909	0.2687	✓	1.0541	0.1936	0.7588	0.9618	9039708	

説明する領域が広いので、表の左側、右側に分けて説明いたします。

拡大左側

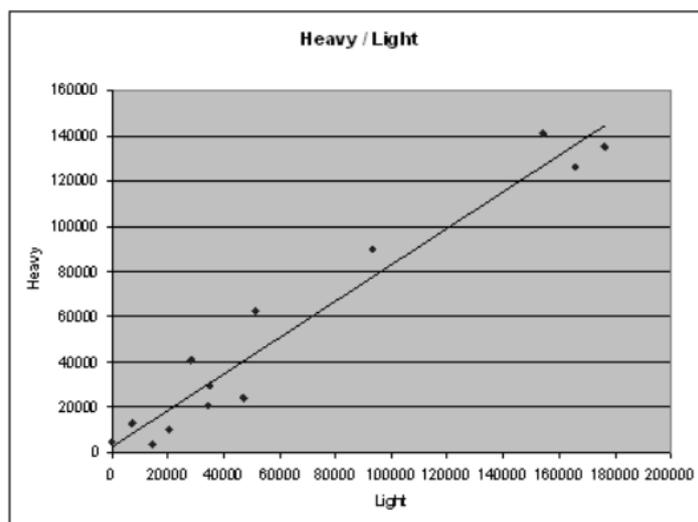
Matches [sp]P06733[ENOA_HUMAN]		ratio	Std.Err	X		
z	Sequence	X	(15ngR1+15ngR2+15ngR3)/(3ngR1+3ngR2+3ngR3)	Std.Err	X	(1
2+	YISPDQLADLYK	✓	1.1284	0.0057	✓	
4+	YGKDATNVGDEGGFAPN...	✓	1.1189	0.0149	✓	
3+	YGKDATNVGDEGGFAPN...	✓	1.1346	0.2646	✓	
3+	YDLDFKSPDDPSR	✓	1.0623	0.1130	✓	

タンパク質にアサインされた各ペプチドについて、定量値の比(**ratio**)が表示されます。表示される ratio はユーザーの定義に基づき、上記のような 1 組の表示だけでなく設定した ratio 分の情報が表示されます。Quantitation の選択項目で事前に定義していない場合や、Quantitation の選択項目と raw ファイル数が合わない場合、各 Component について Ref(erence)との単純な比が表示されます。すべての表示比に対してそれぞれ、“Std.Err”、“X”という名称の項目も併せて表示されます。

Std.Err は、標準誤差(Standard error)の略です。

Ratio を計算する際にばらつきが大きいデータは計算に利用しないよう閾値を定める事ができます。MASCOT Distiller の定量計算では、XIC の計算をするか、あるいはペプチドの ratio をタンパク質の ratio 計算に利用するかについて、3つの観点から検証しています。そのうちの1つが StdErr.です。

Std.Err は計算されたペプチドの比率に対する各比の標準誤差です。ペプチドの比率は XIC ピークにおける各プリカーサイオンの強度に対して最小二乗法フィッティングにより求めています。フィッティングして得られた直線の傾きが ratio となりますが、この時この直線に対する標準誤差が設定値よりも大きいペプチドの比率はタンパク質の比率計算から除外されます。(下図例は ラベルフリー定量でなく SILAC のもので、2 サンプルの名称がそれぞれ Light, Heavy となっていますのでご注意ください。)



“X”は、タンパク質の定量計算に該当ペプチドが使用されているかを表しており、チェックが入っている場合計算に利用されていることを示しています。

拡大右側

	Fraction	Correlation	Intensity	
Std.Err	Fraction	Correlation	Intensity	Modifications
0.0057	0.9777	0.9984	323306106	
0.0194	0.9649	0.9826	29218246	
0.0561	0.9116	0.9293	7663089	
0.0628	0.8374	0.9756	12275965	

各ペプチドに対して定量に関連する評価値が2つ、結果に表示されています。“Fraction”と“Correlation”です。これらはペプチド自体が定量計算に用いられるかの判定をするだけでなく、前駆体が検出され前後の保持時間に対しての探索を継続する際の基準にもなっています。

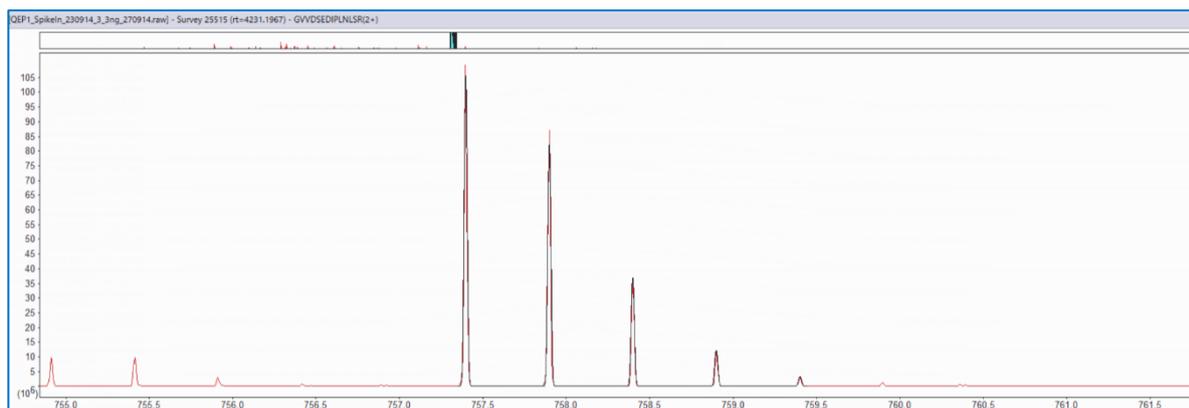
Fraction : ピーク領域の面積比(Fraction of peak area)

Precursor のピーク付近(Scan Window,次頁以降の図も参照の事)の全ピークの面積に対して、計算された理論クラスターピークの面積が占める割合です。理論クラスターピークは実測データの中でモノアイソトピックピークに位置するピークの情報から計算されます。Scan Window に対して、precursor の同位体クラスター以外のピークが多く混ざっているデータ、もともと全然関係のなかったピークを誤って同位体クラスターと認識しているようなデータを排除するための目安となります。値が大きい(1 に近い)ほど理論値と合致するピークだけが指定領域内に現れている事が求められ、逆に小さいほど理論値以外のピークの存在を許容します。

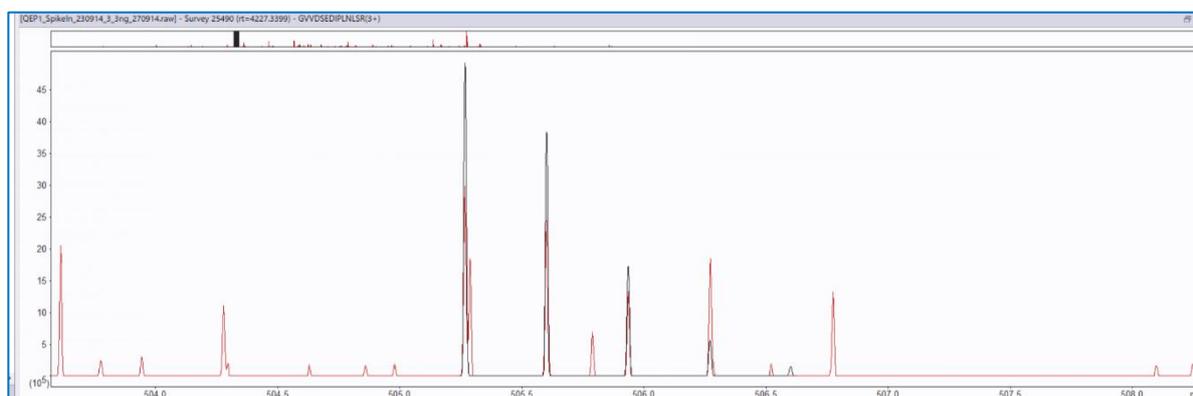
理論スペクトルと実測スペクトルについては、Distiller 上でペプチドを選択した際に次頁のような図が表示されますのでそちらをご参照ください。さらに、右クリック→ ”SPP/DPP Quantitation info” のオン/オフ により理論値の表示形式が変わります。

次頁図例では、赤線が実際のスペクトルデータ、黒線が計算された理論クラスターピークです。Fraction の数値が低いほど、黒線と重なっている箇所以外の赤線のスペクトルが描く面積が増える事がわかります。

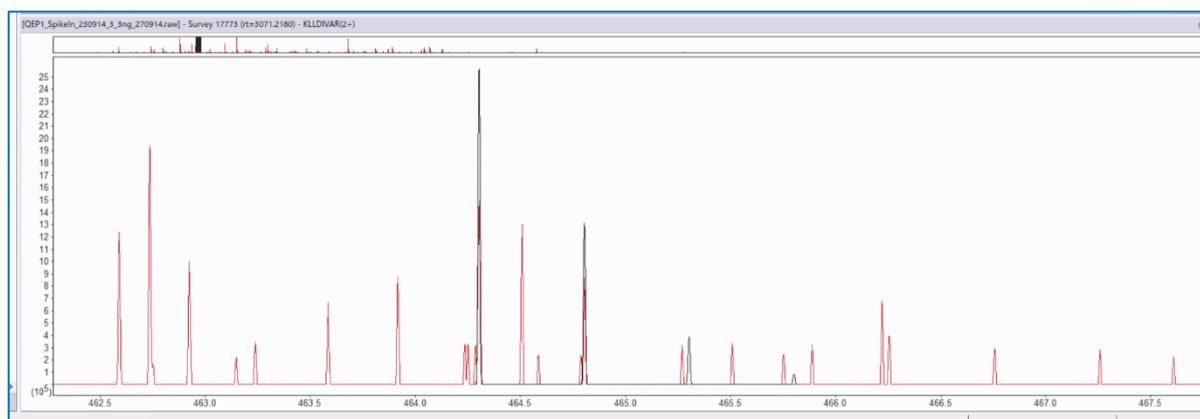
Fractionの数値 0.99 の例



Fractionの数値 0.50 の例



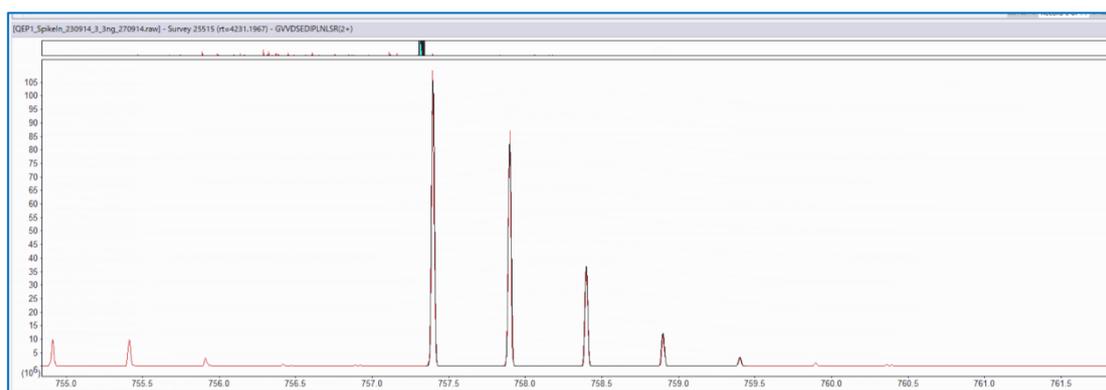
Fractionの数値 0.26 の例



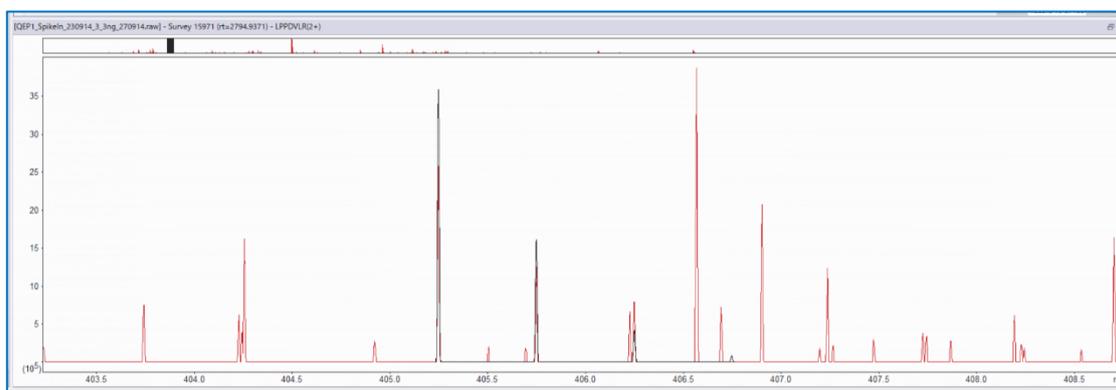
Correlation coefficient(相関係数)

実験データのピークと理論データのピークを重ね合わせた際の相関係数です。1 に近いほど一致度が高い必要があり、逆に 0 に近くなると一致度が低くても許容します。

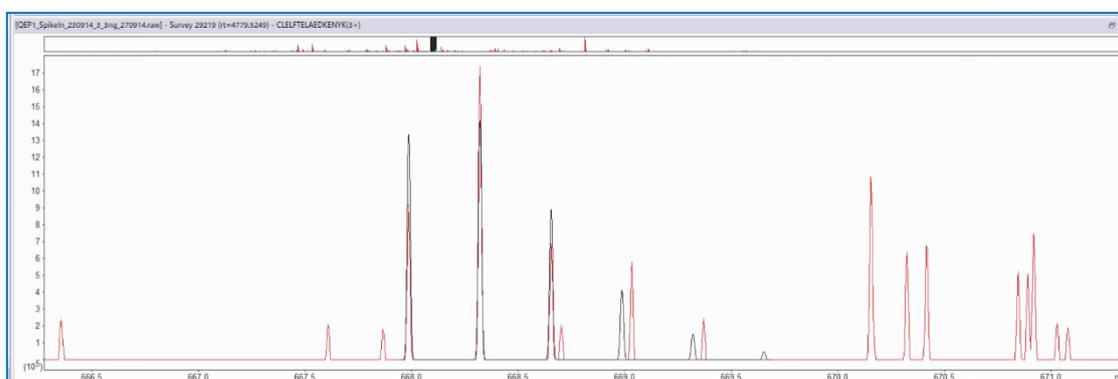
Correlation 0.999 (Fraction 0.991)の例



Correlation 0.984 (Fraction 0.461)の例

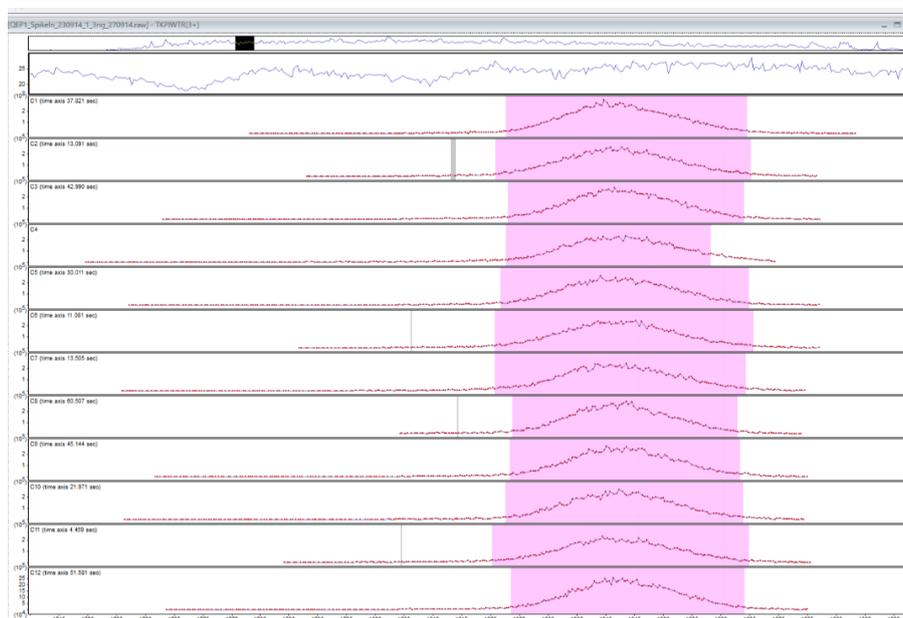


Correlation 0.778 (Fraction 0.500)の例、定量計算には使われていない

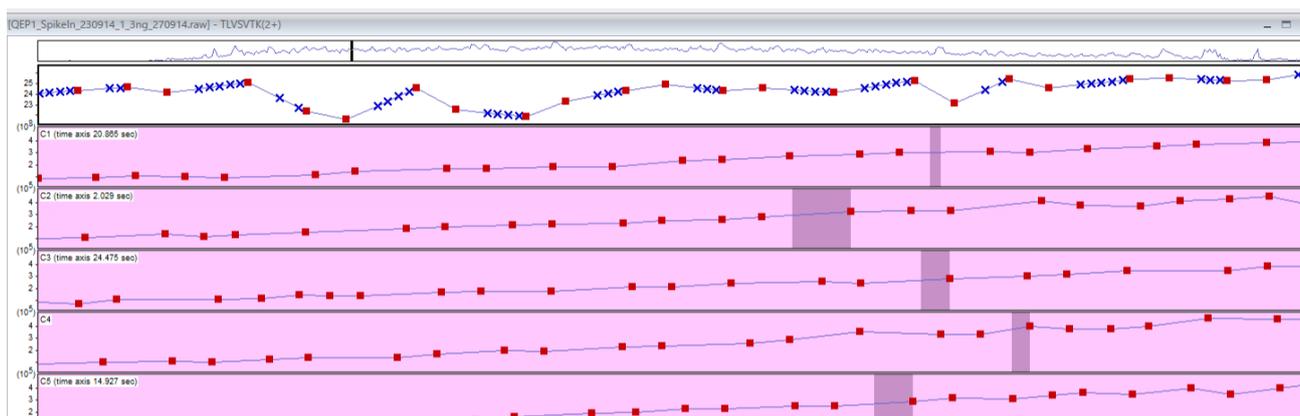


[TIC Window]

Matches でペプチドを選択した際表示される TIC Window で、各データに対する XICs も併せて表示されています(下図)。各 raw データに対して、桃色の領域が XICs 計算されている箇所を意味します。横軸の保持時間データは alignment されています。



各 raw データに対して XICs 計算対象となった領域を見比べる事ができます。デフォルトで表示される状態は保持時間全体が表示対象となっているため詳細が確認できませんが、領域をドラッグ&ドロップ操作で拡大して表示させることもできます(下図)。

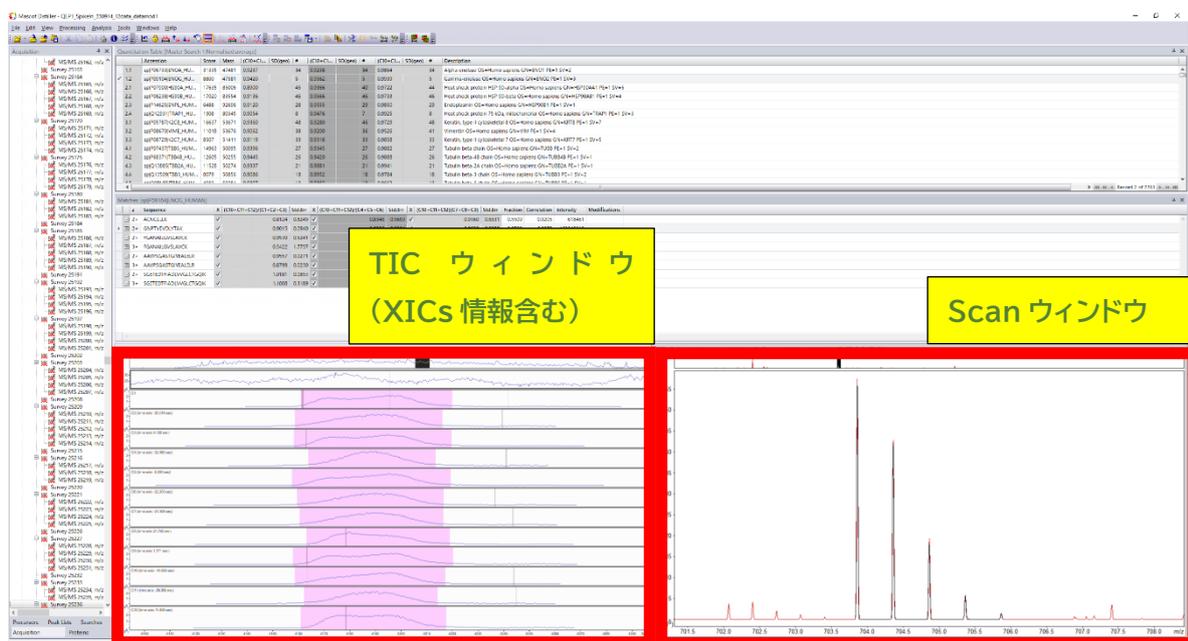
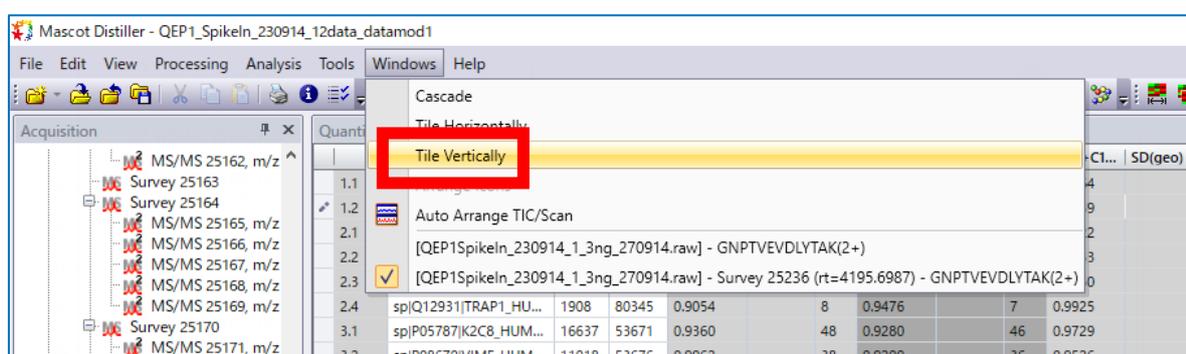


表示内容は以下の通りです。

- 赤い■ Precursor Scan
- 青い× MS/MS Scan
- 桃色の領域 XICs を計算している範囲
- 紫の領域 同定ペプチドの MS/MS スペクトルが検出された領域

Quantitation table の各項目は並び替えて表示させることができます。また ratio や normalization, peptide ratio から protein ratio を計算させる方法などについては、menu の Tools -> Format Options -> Quantitation Options にある各項目を変更する事で調整可能です。設定値を変更しながら、結果がどうなるのかを確認していくことができます。

また、TIC ウィンドウと Scan ウィンドウはどちらも表示に広い領域が必要で、場合によってはデフォルト表示に適さないことがあります。特に両ウィンドウを同時に見たい場合、Windows -> Tile Vertically を選択する事で両ウィンドウが横並びになり見やすくなる場合がありますのでお試しください。



[結果でチェックしていく内容について]

「手順 2.定量解析を行うデータについて」でご紹介しているように、解析対象のデータはタンパク質の量比は human の protein ですべて **1:1** に、Ecoli の protein で **1:5, 1:2, 1:1.5** になる事が期待されたデータです。

まずは Distiller にて表示設定をした 3 種類の ratio が、human 並びに Ecoli で上記のような期待される値にどれくらい近いのかご確認ください。データは各種項目で並び替えが可能です。

Distiller 上でいくつかの解析とグラフ表示が可能です。詳細は**手順 10 に示す report 機能**についてご覧ください。PCA でのサンプル分散状況の確認をしたり、クラスター解析、volcano plot などでの変動タンパク質に関する解析を行ったりすることが可能です。

ただし Distiller 上で行う事ができる操作・検証は限られています。必要に応じて、**手順 10 の export 機能、あるいは report 機能の一部**でデータを出力して外部のプログラムで読み込ませたうえで、期待値からのずれている度合いを表す数値をご自身で計算して頂く事も推奨いたします。

また本解析は同じサンプルについて replicate の 3 回繰り返しという構成です。その構成の特性を活かして MASCOT 側が提供している方法以外の protein ratio 計算方法を適用したい、繰り返し内容を活用した検定を行いたい、などの希望もあるかと思えます。Distiller で表示される定量結果は ratio ですが、report 機能にある出力項目「**table-peptides-int**」で各ペプチドの intensity を出力させることもできます。ご希望の方はその出力内容を使ってそれ以降の計算を実行してください。

10. 定量計算実行 5・計算後の report、export

Distiller の標準機能では raw データを取り込み、定量データの算出に関する計算個所の確認や簡単な結果出力までを行う事ができます。

より複雑な解析を行う場合、Distiller にはプログラム言語 Python と連動する機能があり、各種統計解析やグラフ表示をさせる事ができます。これらを report 機能と呼んでいます。

またデータを XML やテキストファイル、クリップボードに出力させることもできます。これを export 機能と呼んでいます。

手順 10 では report 機能と export 機能についてご紹介します。

結果の report : 各種グラフ表示・結果画面を別形態で出力・表示

MASCOT Distiller では算出された定量値からさらにレポートを作成する機能を有しています。ver.2.8 以降の Distiller ではプログラム言語 Python との連携によってレポート機能を実現しています。レポート機能は、レポートする内容を定義している Python プログラムと、プログラムに与えるデータの入力に使用する XML ファイルから構成されています。report 関連のファイルは、Distiller インストールフォルダの”reports”フォルダ内に格納されています。またデータの定義方法については、distiller_report_definition_1.xsd (Distiller インストールフォルダの”schema”フォルダにあります)に記述されています。

定義した内容は menu の Analysis -> **Reports** に一覧表示されます。Distiller ではあらかじめ作成済の reports 機能がいくつかあり、最初からこれらの機能を使用する事ができます(次頁図)。また内容について python プログラムや XML ファイルを書き換えて調整する事も可能です。

さらに、必要に応じてユーザーが report 機能を追加したり、カスタマイズしたりする事が可能です

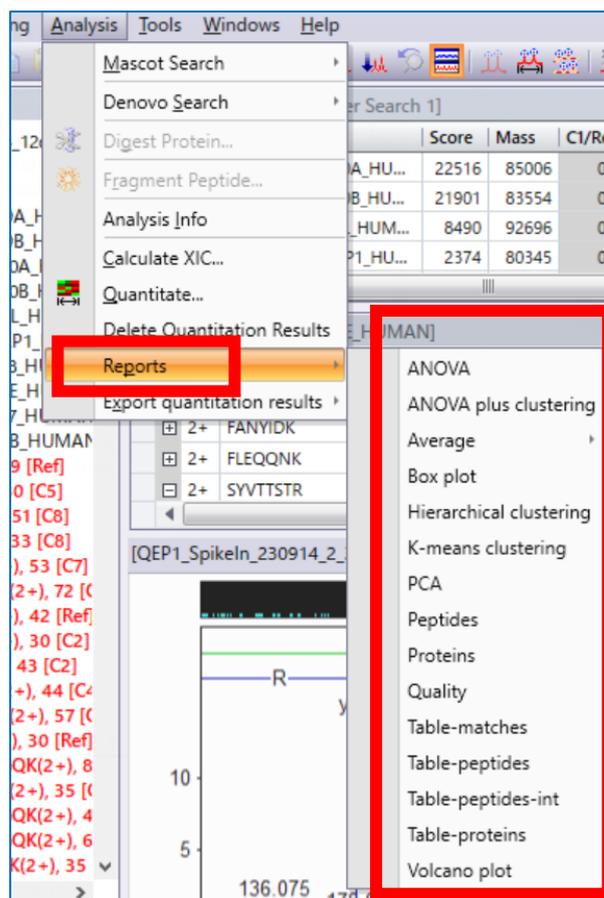
[次頁に続きます]

[統計解析・グラフ表示]

- ANOVA
- ANOVA plus clustering
- Hierarchical clustering
- K-mean clustering
- PCA
- Volcano plot

[定量数値を HTML や CSV ファイルとして出力]

- Peptides
- Proteins
- Quality
- table-matches
- table-peptides
- **table-peptides-int**
- table-proteins



* 特に、**table-peptides-int** は、ご自身でペプチドの intensity 情報から protein の定量情報を算出する事を希望される際よく使われるフォーマットです。

Hit	Member	Accession	Score	Mass	pepMatch	z	Sequence	Fraction	Correlation	Modification	Ident(C1)	Start(C1)	End(C1)	Intensity(C1)	Ident(C2)	Start(C2)	End(C2)	Intensity(C2)	Ident(C3)
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	1	2	EGLELLK	0.9907	0.9983		True	3361	3390	1.293e+07	True	3381	3409	1.306e+07	True
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	2	2	TIAPALVSK	0.8654	0.9983		True	2602	2652	6.358e+07	True	2621	2670	6.157e+07	True
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	3	2	IEEELGSK	0.6041	0.9981		True	1324	1357	1.247e+06	False	1339	1373	1.278e+06	False
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	4	2	NFRNPLAK	0.9050	0.9986		True	2192	2232	2.639e+07	True	2214	2258	2.758e+07	False
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	5	2	SONGLEEK	0.9946	0.9983		True	2615	2677	1.758e+07	True	2634	2699	7.653e+07	True
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	6	3	TIAPALVSKK	0.7026	0.9556		False	2056	2090	3.852e+05	False	2080	2122	4.01e+05	False
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	7	3	SGKYDLDFK	0.9726	0.9948		True	2405	2444	7.191e+06	True	2425	2465	1.1e+06	True
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	8	2	SGKYDLDFK	0.7722	0.9959		True	2405	2444	1.028e+07	True	2424	2465	1.15e+07	True
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	9	3	IGAEVYHNLK	0.8777	0.9984		False	2101	2139	5.665e+07	True	2120	2158	1.496e+07	False
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	10	2	IGAEVYHNLK	0.9937	0.9977		False	2101	2140	5.848e+07	False	2120	2158	1.852e+07	False

Hit	Member	Accession	Score	Mass	pepMatch	z	Sequence
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	1	2	EGLELLK
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	2	2	TIAPALVSK
1	1	sp P06733 ENOA_HUMAN	3.133e+04	47481	3	2	IEEELGSK

ons	Ident(C1)	Start(C1)	End(C1)	Intensity(C1)	Ident(C2)	Start(C2)	End(C2)	Intensity(C2)	Ide
	True	3361	3390	1.293e+07	True	3381	3409	1.306e+07	Tru
	True	2602	2652	6.358e+07	True	2621	2670	6.157e+07	Tru
	True	1324	1357	1.247e+06	False	1339	1373	1.278e+06	Fa

結果の export : テキストデータ出力

解析した数値を出力して利用する場合、前述の report で出力させる方法の他にさらにあと2種類の方法でファイル出力をさせる事ができます。

A. “Export quantitation results” と、 B. “Copy quantitation table”です。

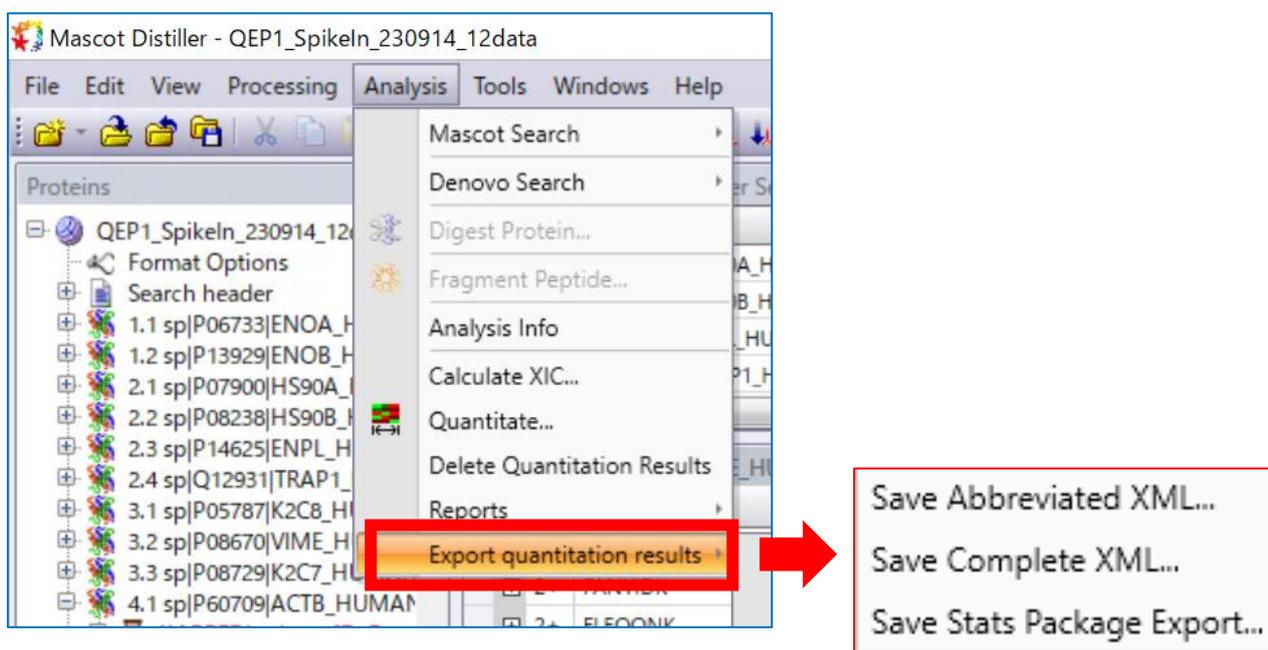
A. “Export quantitation results”

Menu の analysis -> **Export quantitation results** にて、3種類の形式でファイルを出力する事ができます。

“Save Abbreviated/Complete XML は、定量データをXML形式でまとめている出力ファイルで、主に他アプリケーションと XML 形式でのデータやり取りをする際に利用します。

実例として、Scaffold Q+S とのデータやり取りなどが挙げられます。

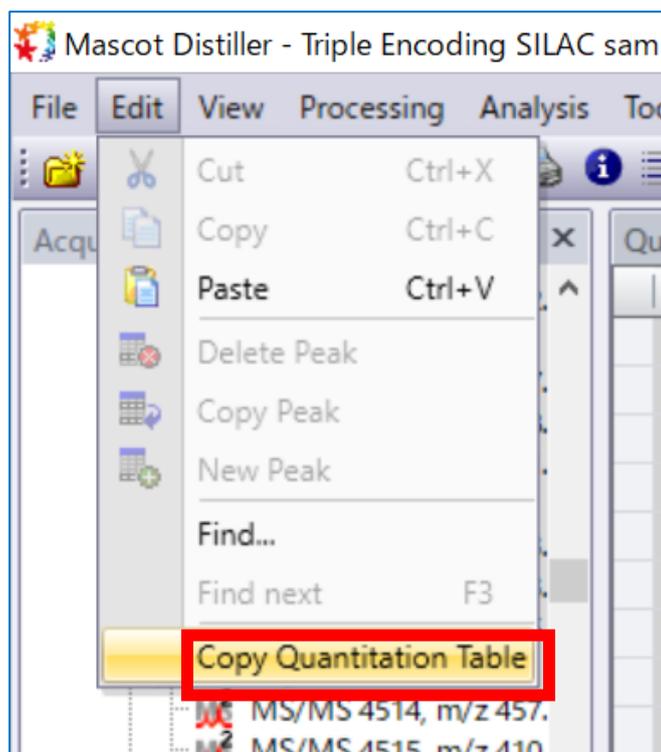
一方、“Save Stats Package export”では、query(matches)、peptide, protein, protein group レベルでの定量値をタブ区切りファイルで出力し、それらを圧縮したファイルが得られます。タブ区切りファイルを読み込めるソフトウェアにご利用ください。



B. Copy Quantitation table

手っ取り早く Distiller の Quantitation table 情報を他のアプリケーションで利用するために、クリップボードに Quantitation table 情報をコピーさせることができます。

Menu の Edit -> **Copy Quantitation table** を選択する事で、クリップボードにデータがコピーされます。



お問い合わせ先

マトリックスサイエンス株式会社

support-jp@matrixscience.com

03-5807-7895

〒110-0015 東京都台東区東上野 1-6-10 ARTビル1階

