



タンパク質同定結果検証・整理 ソフトウェア

# Scaffold 5



- 暫定版

最終更新日 : 2021/12/08





# 目次

<b>1.はじめに</b> .....	<b>1</b>
1-1. Scaffold とは .....	1
1-2. Scaffold でのデータ読み込み .....	1
1-3. Scaffold で採用されているアルゴリズム .....	2
<b>2. データの読み込み</b> .....	<b>4</b>
2-1. 概要、データの階層構造 .....	4
2-2. MASCOT 結果ファイルを直接指定し読み込む方法 .....	6
2-2-1. dat ファイルの読み込み: validation は Scaffold 上で行う方法 .....	7
2-2-2. mzIdentML 読み込み: validation は取り込み前に行う方法 .....	15
2-3. MASCOT Server からネットワークを介して dat を取得する方法 .....	21
2-4. raw またはピークリストを直接読み込むところから始める方法 .....	28
2-4-1. 実行する検索エンジンプログラムをセットする : X! Tandem の場合 .....	28
2-4-2. 実行する検索エンジンプログラムをセットする : MSFragger の場合 .....	29
2-4-3. MSFragger または X!Tandem での検索実施とデータ読み込み方法 .....	29
2-5. 配列データベースの登録 .....	37
2-5-1. 概要 .....	37
2-5-2. Edit FASTA Databases ダイアログ .....	37
2-5-3. FASTA データベースの登録方法 .....	38
2-5-4. parse rule について(Configure Database Parser) .....	40
<b>3. Samples View</b> .....	<b>40</b>
3-1. 概要 .....	41
3-2 . Samples View 画面・表示内容 .....	41
3-2-1. Display pane: Samples table で表示する数値の設定 .....	42
3-2-2. Filtering Samples: 表示たんぱく質/ペプチドの絞り込み .....	45
3-2-3. The Samples Table: 同定タンパク質に関する情報の表示 .....	47
3-2-4. Information Panes: タンパク質・サンプルに関する追加情報 .....	49
3-3 . FDR ダッシュボード・オプションインジケータランプ .....	50
3-4 . Probability の凡例 .....	51
<b>4. Menu の各項目について</b> .....	<b>52</b>
4-1. menu の内容 説明 .....	52
4-2. ファイル保存方法・ファイルを開く方法 .....	56
4-3. sf3 ファイルの統合 (File ->merge) .....	57
4-4. sf3 ファイルのデータサイズを削減する方法 .....	58
4-5. GO の設定 .....	59
4-5-1. GOA ファイルのセット .....	59

4-5-2. 表示する GO 情報の設定 .....	62
<b>4-6. Preferences の設定内容 .....</b>	<b>66</b>
<b>4-7. Advanced Preferences の設定内容 .....</b>	<b>68</b>
<b>4-8. Pathway に関する設定 .....</b>	<b>69</b>
4-8-1. Wikipathways, Reactome と使用時のタンパク質 ID について .....	69
4-8-2. Pathway 情報の表示 [Scaffold 上] .....	69
4-8-3. Pathway 情報の表示 [外部サイト] .....	74
<b>4-9. ツールバーアイコン .....</b>	<b>76</b>
<b>5. Load Data View .....</b>	<b>77</b>
<b>5-1. 概要 .....</b>	<b>77</b>
<b>5-2. Experiment pane .....</b>	<b>78</b>
<b>5-3. BioSample tab と Load and Analyze Queue button .....</b>	<b>79</b>
<b>5-4. Information pane .....</b>	<b>80</b>
<b>6. Protein View .....</b>	<b>81</b>
<b>6-1. 概要 .....</b>	<b>81</b>
<b>6-2. Proteins pane : タンパク質に関する情報を表示 .....</b>	<b>82</b>
<b>6-3. Peptide pane : ペプチドに関する情報を表示 .....</b>	<b>83</b>
<b>6-4. Spectrum pane : タンパク質/スペクトル 関連図 .....</b>	<b>86</b>
6-4-1. Protein Sequence tab .....	86
6-4-2. Similar Proteins tab .....	87
6-4-3. Spectrum tab .....	87
6-4-4. Spectrum/Model error tab .....	88
6-4-5. Fragment table tab .....	88
<b>7. Grouping, Clustering と Similarity View .....</b>	<b>89</b>
<b>7-1. Scaffold での類似タンパク質の扱い .....</b>	<b>89</b>
<b>7-2. 表示内容の詳細 : summary 画面 .....</b>	<b>89</b>
7-2-1. Protein Grouping (same-set) .....	91
7-2-2. Protein Pairing (sub-set) .....	92
7-2-3. Protein Clustering (Family Proteins) .....	93
<b>7-3. Legacy Protein grouping .....</b>	<b>95</b>
<b>7-4. Samples View と Similarity View との関連について .....</b>	<b>97</b>
<b>7-5. Similarity View 概要 .....</b>	<b>99</b>
<b>8. Quantify View .....</b>	<b>102</b>
<b>8-1. Quantify View : 定量指標を基にしたグラフや GO, ペン図を表示 .....</b>	<b>102</b>
<b>8-2. Quantitative Value pane .....</b>	<b>103</b>
<b>8-3. Quantify Scatterplots pane .....</b>	<b>103</b>
8-3-1. Scatterplot タブ .....	104



8-3-2. <i>Stdev Scatterplot</i> タブ	105
8-3-3. <i>Volcano plot</i> タブ	106
<b>8-4. Venn Diagrams pane</b>	<b>107</b>
<b>8-5. Annotation Charts pane</b>	<b>107</b>
<b>9. Publish View</b>	<b>109</b>
9-1. Publish View : Method の文章化	109
<b>10. Statistics View</b>	<b>111</b>
10-1. Statistics View 概要 : 同定確率計算に使用したスコア分布などを表示	111
10-2. MS Sample table	112
10-3. Statistics View Upper Right Pane	113
10-3-1. <i>FDR Browser</i> タブ	113
10-3-2. <i>Peptide ROC Plots</i>	115
10-3-3. <i>Protein Probability Calculation</i> タブ	116
10-4. Multiple Search Engine Scatter Plot pane	117
10-5. Peptides Validation pane	117
<b>11. 定量手法と検定</b>	<b>118</b>
11-1. 概要	118
11-2. ラベルフリーの定量方法	118
11-2-1. <i>Spectral Counting</i>	119
11-2-2. <i>Total Ion Count</i>	120
11-2-3. <i>Precursor Ion Intensity quantitation</i>	120
11-3. Normalization について	121
11-4. 検定	122



# 1.はじめに

## 1-1. Scaffold とは

Scaffold は、DDA プロテオミクスデータをまとめ、サンプル間での比較を主な目的としたソフトウェアです。

複数の検索エンジンの結果ファイルを取り込むことが可能なほか、いくつかのフリーな検索を使って raw データ(またはピークリストファイル)から検索を行い、結果を表示させることも可能です。

取り込んだ同定ペプチドについて、Percolator を中心とした Validation 用のプログラムを適用する事ができます。

取り込んだ結果について、EXCEL などでもとめた状態とは異なりスペクトルとアサインされたペプチドの理論フラグメントピークとのマッチング状況を確認する事もできます。また Gene Ontology 情報の表示や Pathway データベースへのリンク、サンプル間の同定タンパク質比較といった、タンパク質の解析を補助する機能もついています。

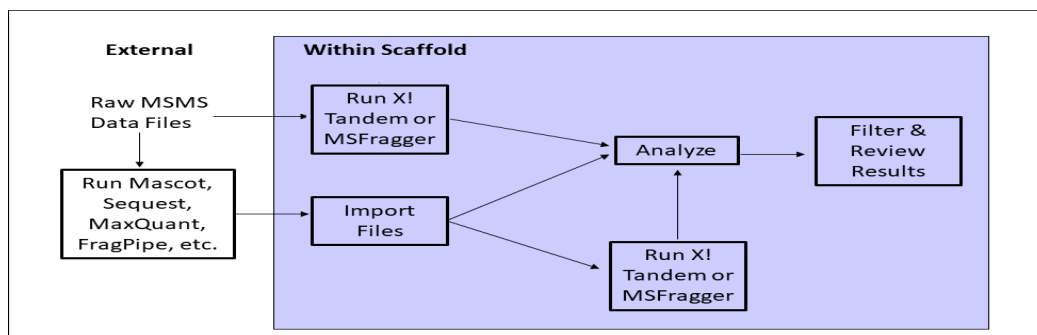
検索結果については Excel スプレッドシート、mzIdentML, などに出力できるほか、Scaffold 関連製品で取り扱う事ができる SFDB ファイル、BLIB ファイルを出力する事ができます。

結果についてはファイルに保存し、Scaffold ライセンスを所有していない共同研究先でもフリーの Viewer を使って結果を閲覧する事が可能です。

## 1-2. Scaffold でのデータ読み込み

Scaffold のデータ取り込みには大きく分けて以下の 2 種類があります。

- 検索エンジンの結果ファイルを読み込む
- 検索前の raw ファイルまたはピークリストファイルを読み込む



また、検索エンジンの結果を取り込む場合でも、改めて X!Tandem や MSFragger での検索も実行して、両者の結果を組み合わせる事も可能です。

データ取り込みについての詳細な説明は 2 章で行っています。

## 1-3. Scaffold で採用されているアルゴリズム

Scaffold では、ペプチド並びにタンパク質同定の信頼性を上げるため、以下のアルゴリズムを適用する事ができます。

- Percolator
- LFDR-based scoring system
- Peptide Prophet
- Protein Prophet

各アルゴリズムについて、少し詳しく説明します。

### ■ Percolator

取り込んだデータのペプチド-スペクトルの再評価を行うと同時に、Validation も行います。スコアだけでなく質量精度や各種情報をもとに判断します。同定スペクトルを増やす目的で、現在 幅開く利用されています。

#### 論文:

Lukas Käll, Jesse Canterbury, Jason Weston, William Stafford Noble and Michael J. MacCoss.  
Semi-supervised learning for peptide identification from shotgun proteomics datasets  
Nature Methods 4:923 – 925, November 2007

### ■ LFDR-based scoring system

バイズ統計の考えに基づいており、ナイーブバイズ分類器を使った判別スコアリングを使ってペプチド同定の検証が行われています。トレーニングデータセットの選別を繰り返し変更しながら判別スコアリングの最適化を実施しています。計算により local FDR(LFDR)を算出しその数値がそのままスコア化されています。トレーニングデータの選別は繰り返し複数回行われ、10 ある分類器のうち3つずつを対象として鍛えていきます。この繰り返し操作により誤って同定されているデータの影響が薄まり判別の正確性が増します。

他の(Percolator など)スコアリング手法同様、LFDR も他の検索エンジンが提供するスコアを取り込みます。Peptide Prophet でいう LDA や Percolator でいう SVM(Support Vector Machine)の分類器にあたるものとして、LFDR ではナイーブバイズ分類器から算出した対数尤度比を使って True Positive と False Positive ヒットの区別を判定しています。識別子の最適化にあたっては過剰適合(overfitting)による汎化不足とならないよう工夫しています。前駆体イオン(親イオン)の誤差は判別スコア計算に考慮される要素の1つというだけでなく、別の観点からもスコアに影響を与えています。Scaffold ではこのスコアについて離散型の(数え上げ方の)スコア分布でなく、より連続的なスコア分布の関数に置き換えて計算し、確率の積算を行う際にも積分計算を行っています。

## 論文:

<https://dx.doi.org/10.1021%2Facs.jproteome.5b00536>

### ■ Peptide Prophet

Peptide Prophet は検索エンジンによって算出された各ペプチドのスコアの分布を連続的な曲線の分布に変換し、同定確率を算出するために利用します。分布の変換の際には検索に利用したデータベースのサイズや検索データの特徴を利用します。この曲線化されたスコア分布を使って各スコアでの同定確率を再計算します。同定確率はそのままペプチド同定の基準にも適用されます。

## 論文:

Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.

Anal. Chem., 2002, 74 (20), pp 5383–5392 DOI: 10.1021/ac025747h

### ■ Protein Prophet

Protein Prophet は Peptide Prophet で算出された各ペプチドの同定確率をもとに、アサインされているタンパク質の同定確率を算出するアルゴリズムです。Scaffold 5 ではそのアルゴリズムをさらに調整し、シェアペプチドかオリジナルペプチドかによって計算における貢献度を変更しています。

## 論文

Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry.

Anal. Chem., 2003, 75 (17), pp 4646–4658 DOI: 10.1021/ac0341261

Scaffold では取り込んだデータに対して、上記の各手法のうちいずれかの方法が適用され評価されます。一方 Scaffold で準備されているアルゴリズムを使わず、各検索エンジンで検証の上出力された結果をそのまま取り込む方法があります。Scaffold ではこれを「Prefiltered Mode」での取り込みと呼んでいます。データの取り込み方に関する詳細は、「**2-2-2. Prefiltered Mode で検証済みの mzIdentML ファイルを Scaffold 上の検証なしに読みこむ**」をご覧ください。

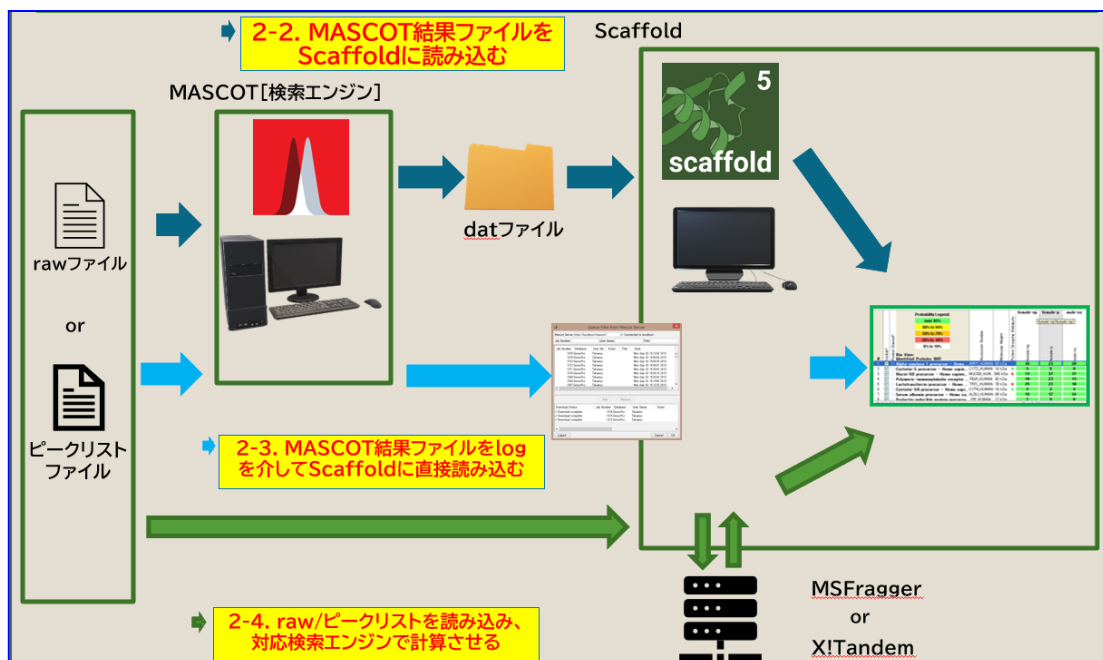
## 2. データの読み込み

### 2-1. 概要、データの階層構造

Scaffold でデータを読み込むには大きく分けて以下の3つの方法があります。

結果取り込みの方法	項目	お勧めのケース
MASCOT の検索結果ファイル(dat ファイル、又は mzIdentML ファイル) を Scaffold PC にコピーしてから Scaffold 上でファイルを取り込む	2-2	<ul style="list-style-type: none"> <li>• Scaffold と MASCOT がネットワークでつながっていない場合</li> <li>• Scaffold の validation 手法を適用しない、「Prefiltered Mode」を適用したい場合</li> </ul>
Mascot Server のログファイルを Scaffold 内で開き、該当結果を選択してネットワーク経由で取り込む	2-3	<ul style="list-style-type: none"> <li>• ネットワークがつながっていて Scaffold の validation 手法を利用する場合</li> </ul>
raw またはピークリストファイルを直接読み込み、X! Tandem または MSFragger で検索してその結果を取り込む	2-4	<ul style="list-style-type: none"> <li>• X!Tandem または MSFragger での検索から実行したい時</li> </ul>

加えて、各結果ファイルにはタンパク質の ID (Accession) 情報のみ含まれ、アミノ酸配列情報が含まれていません。Scaffold で各タンパク質のアミノ酸配列情報を見るためには、**検索に使用したデータベースファイルを Scaffold 上にも取り込む必要があります**。MASCOT の検索結果を取り込む際、該当データベースファイルからアミノ酸配列を取得し直し、Scaffold 内の表示に反映させています。**2-5** ではデータベースファイルの登録方法についてご案内します。



## [データの階層構造について]

データ取り込みの際に重要な概念として、「Category」「Sample」「dat (その他取り込み可能なテキスト/XML ファイル)」の階層構造について説明いたします。

複数の **dat** をまとめたものが **Sample**、さらに複数の **Sample** をまとめたものが **Category** です。階層構造はデータ取り込み時に定義します。



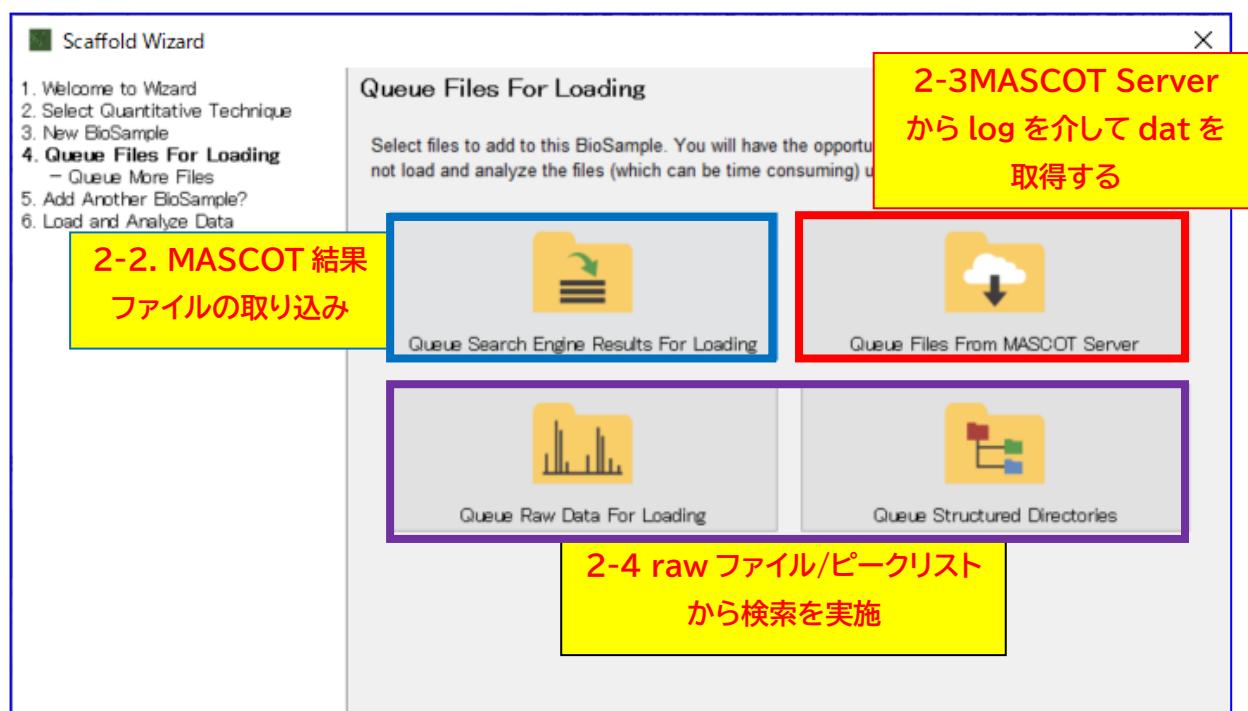
dat をまとめて表示するかどうかについては、結果表示画面「Samples」の上部にある、「**BIO**」ボタン/「**MS**」ボタンで切り替えることができます。dat をすべて表示させる形式が「**MS**」です。

#	Visible?	Starred?	MS/MS View: Identified Proteins (11)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	Int-1	Int-2	Un-1	Un-2		
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type I cytoskeletal 10 ... gi 547749 ...	60 kDa	4	6	5	7	6	11	(0)	1
2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type II cytoskeletal 1 (...gi 1346343...	66 kDa	6	4	4	6	4	4	(0)	7
3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Chain E, Leech-Derived Trypta... gi 3318722...	23 kDa	3	5	4	2	7	7	0	5
4	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type II cytoskeletal 2 e...gi 547754 ...	66 kDa	4	5	1	3	0	(0)	(0)	1
5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type I cytoskeletal 9 (...gi 8117517...	62 kDa	3	(0)	1	2	(0)	(0)	(0)	1

一方「**BIO**」ボタンを押すと sample 以下にある dat がすべてまとめられた表示がされます。

#	Visible?	Starred?	Bio View: Identified Proteins (11)	Accession Number	Weight	Protein Grouping Ambiguity	Int-1	Int-2	Un-1	Un-2
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type I cytoskeletal 10 ... gi 547749 ...	60 kDa	11	12	17	(1)		
2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Keratin, type II cytoskeletal 1 (...gi 1346343...	66 kDa	10	10	7	(7)		

Scaffold でデータを取り込む際、途中でどのように取り込むのか選択する画面が現れます。以下図では画面内の選択肢と、この章で説明する内容の対応付けをしています。

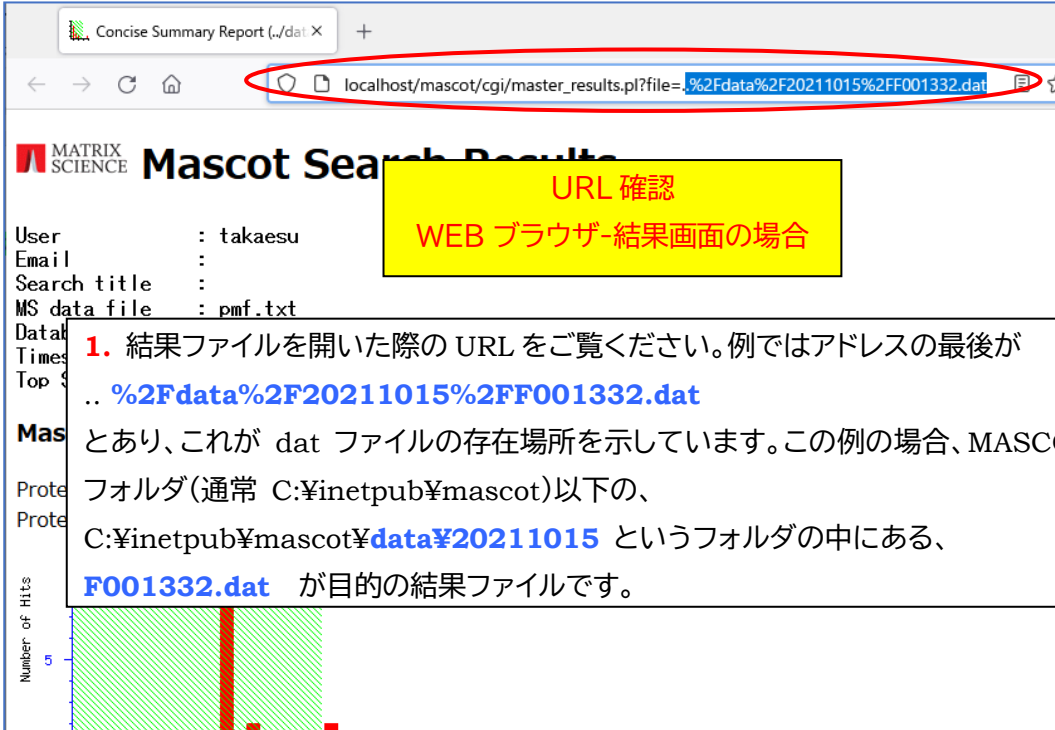


## 2-2. MASCOT 結果ファイルを直接指定し読み込む方法

MASCOT の結果を Scaffold に取り込む 2 つの方法のうちの一つ、MASCOT の検索結果ファイルを Scaffold で直接指定して読み込む方法についてご案内します。**2-2-1** では、MASCOT の結果ファイル dat の取り込みを紹介しています。取り込みたい MASCOT の結果が、どういう名称でどこにあるファイルなのか事前に確認をした上で、該当する dat ファイルを Scaffold がインストールされている PC にコピーする必要があります。また **2-2-2** では、mzIdentML ファイルの取り込みについてご案内します。例えば MASCOT において、FDR が 1% など特定の数値となるよう定めたペプチド同定基準を Scaffold 上においても適用したい場合、結果画面から mzIdentML ファイルにて出力し、Scaffold で prefiltered mode で出力時の状態から Validation をすることなく結果として取り込む事が可能です。

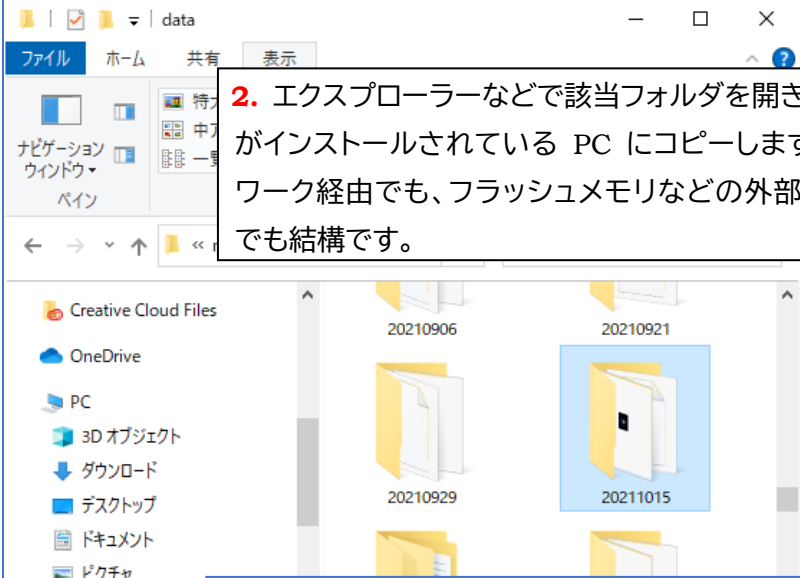
## 2-2-1. dat ファイルの読み込み: validation は Scaffold 上で行う方法

[ MASCOT Server 上の dat ファイルの場所を確認する方法 ]

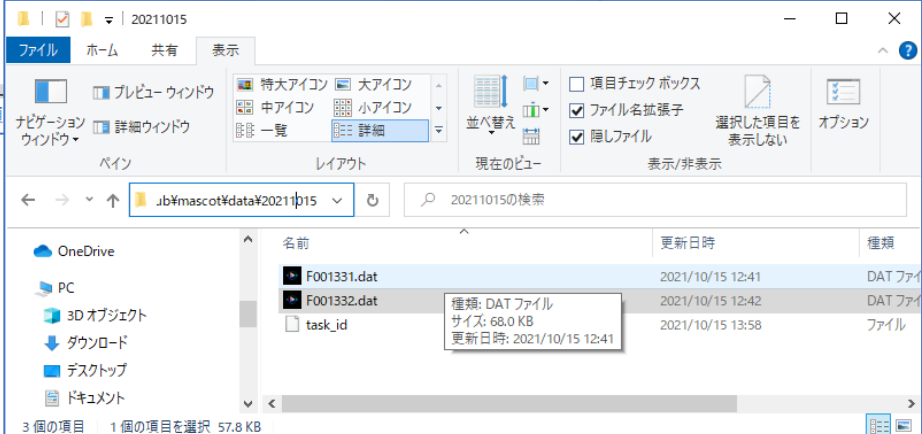


URL 確認  
WEB ブラウザ-結果画面の場合

1. 結果ファイルを開いた際の URL をご覧ください。例ではアドレスの最後が `.. %2Fdata%2F20211015%2FF001332.dat` とあり、これが dat ファイルの存在場所を示しています。この例の場合、MASCOT フォルダ(通常 `C:\inetpub\wwwroot\mascot`)以下の、`C:\inetpub\wwwroot\mascot\data\20211015` というフォルダの中にある、`F001332.dat` が目的の結果ファイルです。



2. エクスプローラーなどで該当フォルダを開き、dat ファイルを Scaffold がインストールされている PC にコピーします。ファイルコピーはネットワーク経由でも、フラッシュメモリなどの外部記憶装置経由でもどちらでも結構です。



名前	更新日時	種類
F001331.dat	2021/10/15 12:41	DAT ファイル
F001332.dat	2021/10/15 12:42	DAT ファイル
task_id	2021/10/15 13:58	ファイル

種類: DAT ファイル  
サイズ: 68.0 KB  
更新日時: 2021/10/15 12:41



**URL 確認**  
**WEB ブラウザ-Search log の場合**

**MASCOT search log**

Version: 2.7.0 - MSKK (D6RQ-7VGN-88XU-9YVD-MLUB)

Sort / filter Log File: ../logs/searches.log Start at: (-1=end, 1=start) -1 how many: 50 86 in log, 86 after filters. Data dir: GETs?:

Job#	PID	dbase	User Name	Email	Ti	In	start time	Durati
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>

**1 の別例 A.** Search log 結果一覧から dat ファイルの場所を確認する方法もあります。表示項目の中で「In」と書かれている項目のチェックボックスにチェックを入れ、左側の「Sort / Filter」 ボタンを押すと・・・

**MASCOT search log**

Version: 2.7.0 - MSKK (D6RQ-7VGN-88XU-9YVD-MLUB)

Sort / filter Log File: ../logs/searches.log Start at: (-1=end, 1=start) -1 how many: 50 86 in log, 86 after filters. Data dir: GETs?:

Job#	PID	dbase	User Name	Email	Ti	Intermediate file	start time
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<a href="#">1332</a>	3016	SwissPro	takaesu			<a href="#">../data/20211015/F001332.dat</a>	Fri Oct 15 12:...
<a href="#">1331</a>	8980	SwissPro	takaesu			<a href="#">../data/20211015/F001331.dat</a>	Fri Oct 15 12:...
<a href="#">1330</a>	7692	SwissPro	takaesu				

ファイルが置かれている相対 path とファイル名が表示されます

**URL 確認**  
**Daemon の場合**

**Task Database**

- 1: Tutorial Search 02
  - DDA-Plasma\_5ng-mL\_010.mzML
  - DDA-Plasma\_25ng-mL\_009.mzML
  - DDA-Plasma\_50ng-mL\_008.mzML
  - DDA-Plasma\_blank\_011.mzML
  - DDAp-Plasma\_5ng-mL\_014.mzML
  - DDAp-Plasma\_25ng-mL\_013.mzML
  - DDAp-Plasma\_50ng-mL\_012.mzML
- 2: Tutorial 02
- 3: Tutorial 03

Access...	Mass	Sc...	Description	Database	Result URL	Submitted	
es...	LEUTX...	22458	14	Paired-like ...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001244.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001244.dat</a>	2021/03/11 10:44
es...	DPY30...	11243	12	Protein dp...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001245.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001245.dat</a>	2021/03/11 10:46
es...	DPY30...	11243	12	Protein dp...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001246.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001246.dat</a>	2021/03/11 10:47
es...	STX3...	33134	12	Syntaxin-3 ...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001247.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001247.dat</a>	2021/03/11 10:45
es...	TVAZ2...	12278	16	T cell rece...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001248.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001248.dat</a>	2021/03/11 10:50
es...	FA50B...	38742	12	Protein FA...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001249.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001249.dat</a>	2021/03/11 10:52
es...	DPY30...	11243	11	Protein dp...	SwissProt...	<a href="http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001250.dat">http://localhost/mascot/cgi/master_results.pl?file=../data/20210311/F001250.dat</a>	2021/03/11 10:53

**1 の別例 B.** Daemon でもパスの一覧を表示させることができます。「Result URL」の項目を御覧ください。

## [ dat ファイルを指定して読み込む方法 ]

以下操作方法についてご案内します。定義するデータですが、Categoryとして Controlと Treatment, それぞれに Sample を2種類属した計4つのデータの取り込みを想定しています。

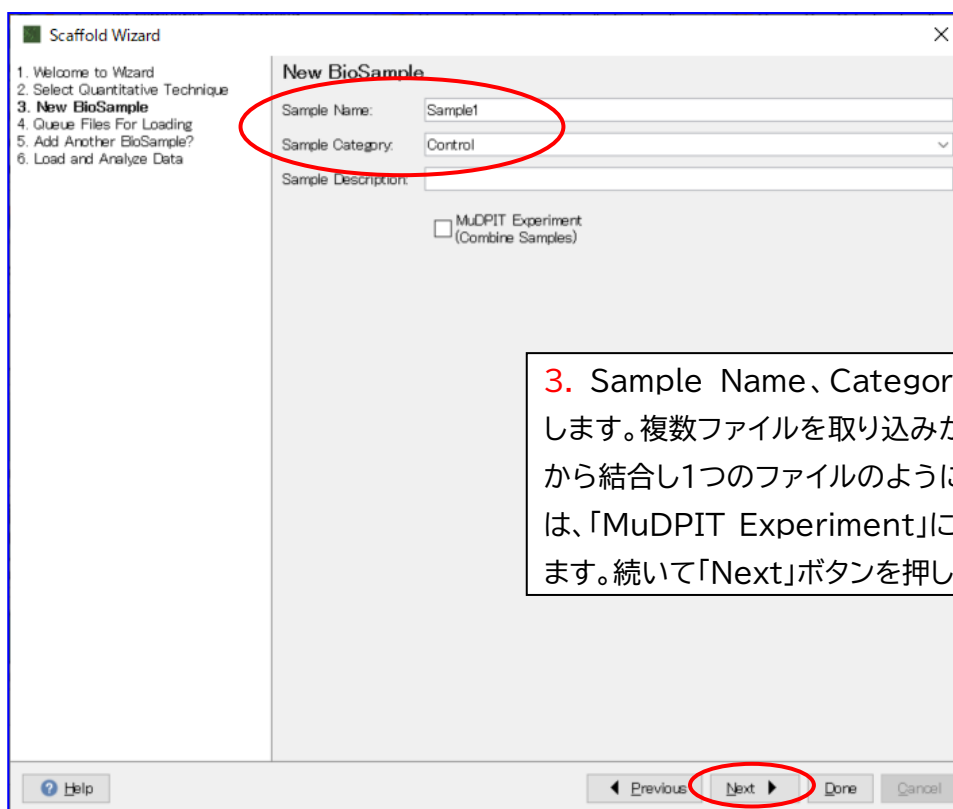
Category	Sample	dat
Control	Sample1	F001234.dat
	Sample2	F001235.dat
Treatment	Sample3	F001236.dat
	Sample4	F001237.dat

## [Control, Sample1 の指定]

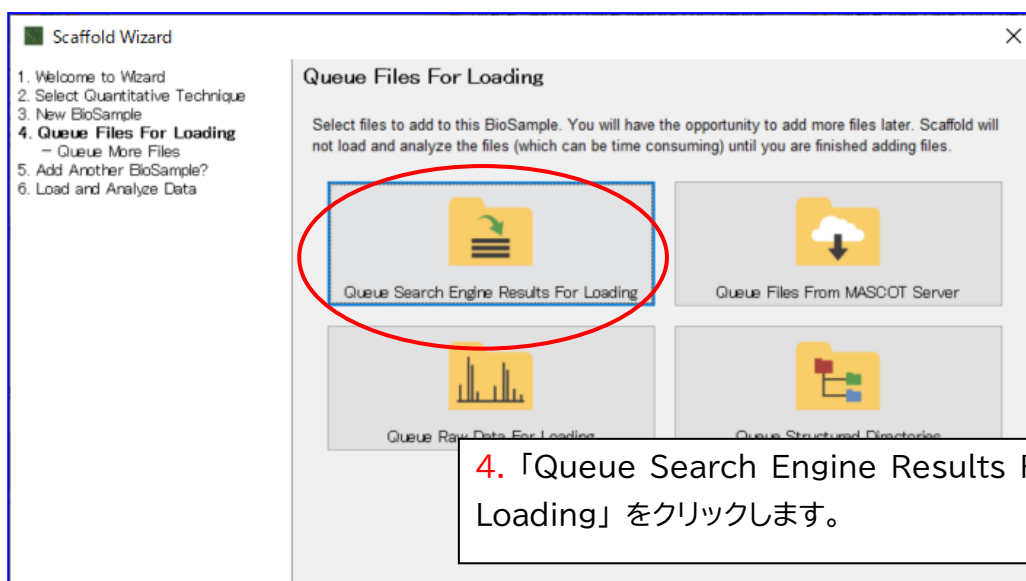
The image shows a screenshot of the Scaffold Q+S Evaluation software interface. The main window is titled 'Scaffold Q+S Evaluation - Load Data'. A red circle highlights the 'New' button in the top-left corner of the main window. A text box with the number '1' contains the instruction: '1. New ボタンを押し(または Ctrl+N) 取り込みを開始します'. Below this, a 'Scaffold Wizard' dialog box is open. The dialog box has a list of steps: '1. Welcome to Wizard', '2. Select Quantitative Technique', '3. Queue Files For Loading', '4. Add Another BioSample?', and '5. Load and Analyze Data'. The 'Select Quantitative Technique' section is active, showing a list of options: 'Spectrum Counting (Standard)', 'iTRAQ (4-plex)', 'iTRAQ (8-plex)', 'TMT (2-plex)', 'TMT (6-plex)', 'TMT (10-plex)', 'TMT (11-plex)', 'TMT (16-plex)', 'Stable Isotope Labeling (Multiplex)', and 'Precursor Intensity (Standard)'. A red circle highlights the 'Spectrum Counting (Standard)' option. A text box with the number '2' contains the instruction: '2. 定量データを含んでいない場合、"Spectral Conting(Standard)" を選択します。続いて"Next"ボタンを押します。'. At the bottom of the dialog box, a red circle highlights the 'Next' button.

1. New ボタンを押し(または Ctrl+N) 取り込みを開始します

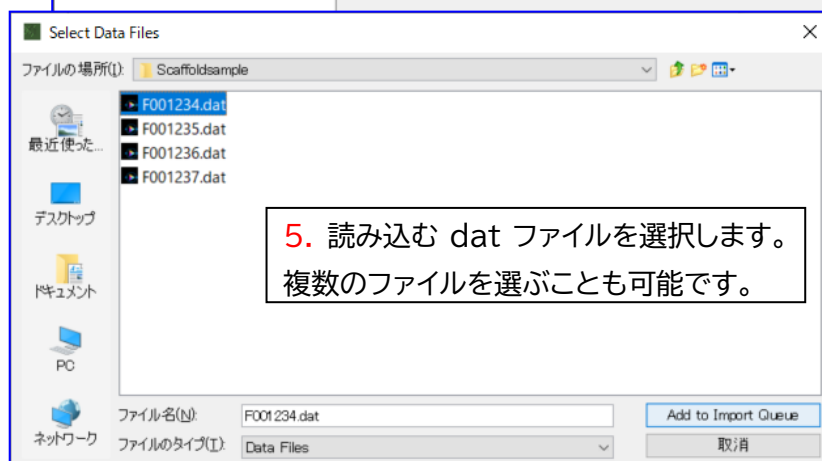
2. 定量データを含んでいない場合、"Spectral Conting(Standard)" を選択します。続いて"Next"ボタンを押します。



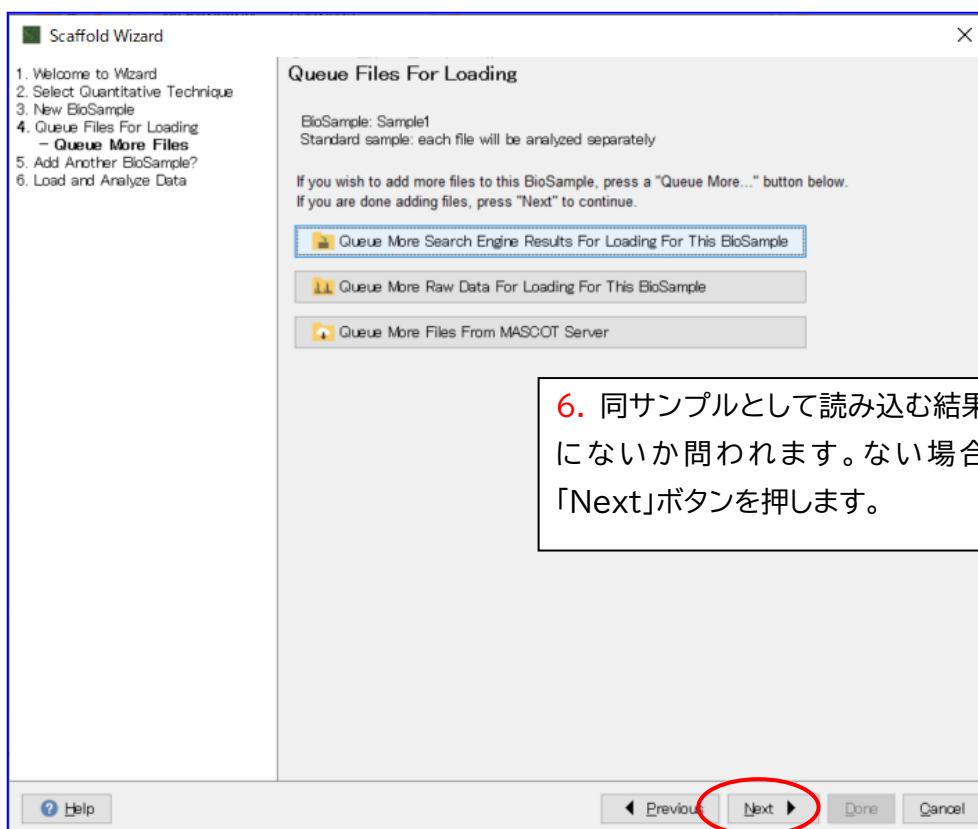
3. Sample Name、Category の欄を記入します。複数ファイルを取り込みかつ結果を最初から結合し1つのファイルのように扱いたい場合は、「MuDPIT Experiment」にチェックを入れます。続いて「Next」ボタンを押します。



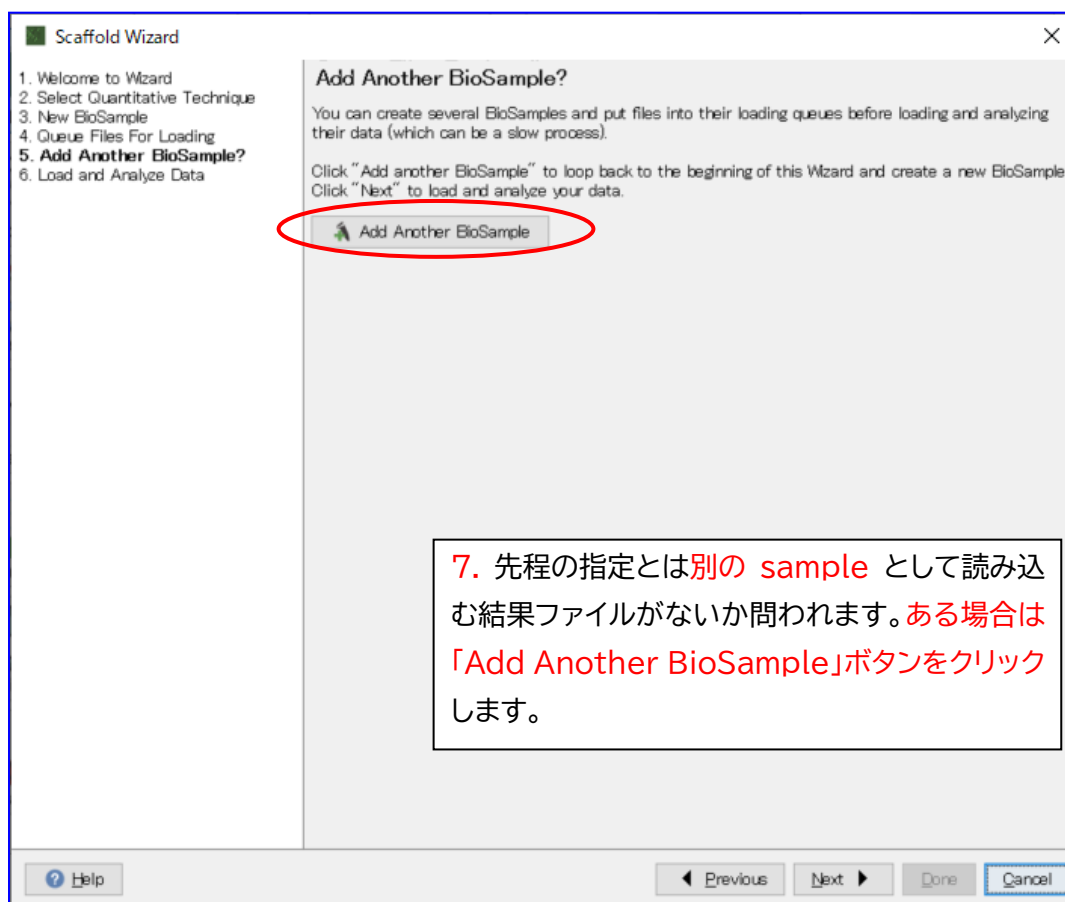
4. 「Queue Search Engine Results For Loading」をクリックします。



5. 読み込む dat ファイルを選択します。複数のファイルを選ぶことも可能です。



6. 同サンプルとして読み込む結果ファイルが更  
にないか問われます。ない場合は画面下の  
「Next」ボタンを押します。



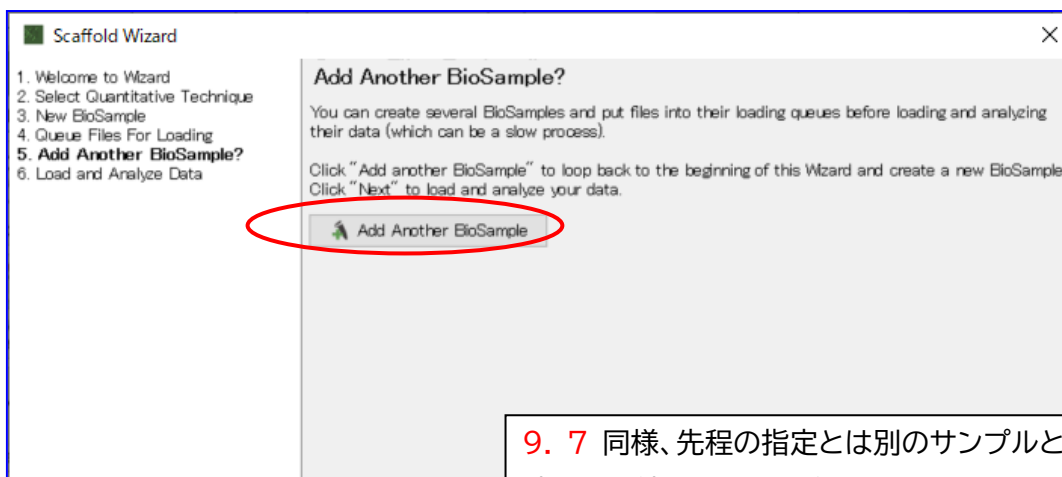
7. 先程の指定とは別の sample として読み込  
む結果ファイルがないか問われます。ある場合は  
「Add Another BioSample」ボタンをクリック  
します。

## [Control, Sample2 の指定]

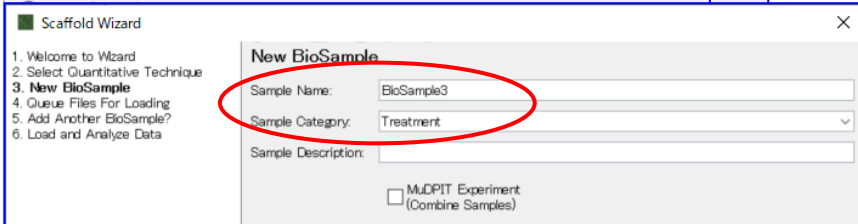
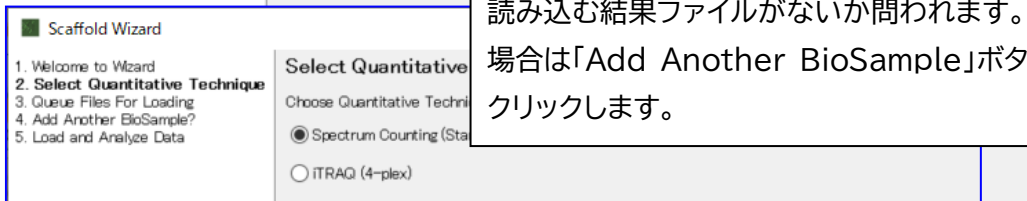
The image displays a sequence of screenshots from the Scaffold Wizard software, illustrating the process of creating a new BioSample and queuing files for loading. The screenshots are arranged in a layered fashion to show the progression of the wizard.

- Top Screenshot:** Shows the "Select Quantitative Technique" step. The "Spectrum Counting (Standard)" option is selected.
- Second Screenshot:** Shows the "New BioSample" step. The "Sample Name" field is highlighted with a red circle and contains the text "BioSample 2".
- Third Screenshot:** Shows the "Queue Files For Loading" step. A text box is overlaid on this screenshot with the following text: "8. 2~6 の操作を繰り返します。今度は Sample 名指定の際名称を変える事と、先ほどと異なる dat ファイルを選択する事に注意してください。" (Repeat steps 2-6. This time, pay attention to changing the sample name when specifying the name and selecting different .dat files than last time.)
- Fourth Screenshot:** Shows the "Select Data Files" step. The file list includes "F001234.dat", "F001235.dat", and "F001236.dat".
- Bottom Screenshot:** Shows the "Queue Files For Loading" step. The "Queue More Search Engine Results For Loading For This BioSample" button is highlighted.

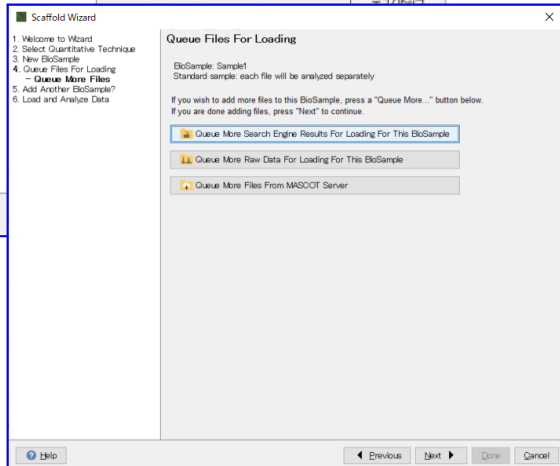
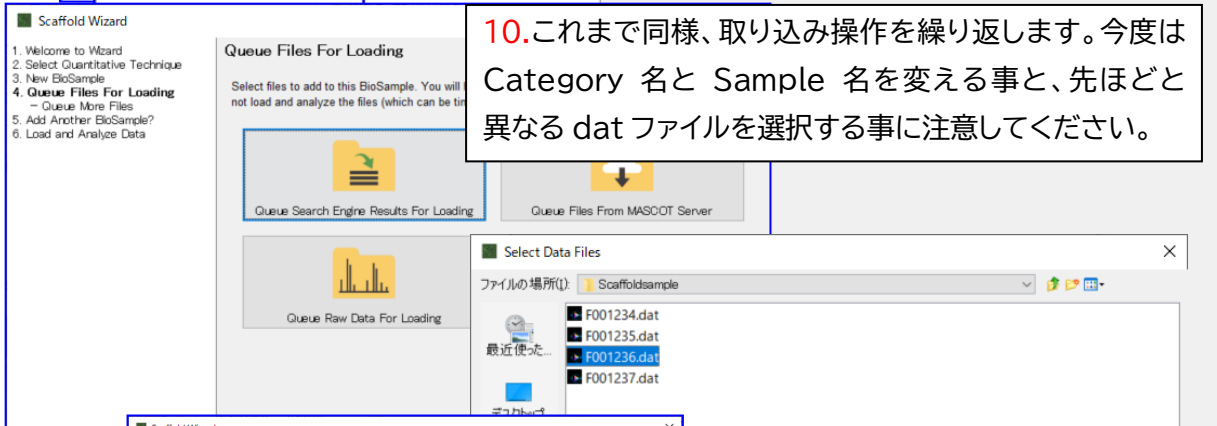
[Treatment, Sample3 の指定]



9. 7 同様、先程の指定とは別のサンプルとして読み込む結果ファイルがないか問われます。ある場合は「Add Another BioSample」ボタンをクリックします。

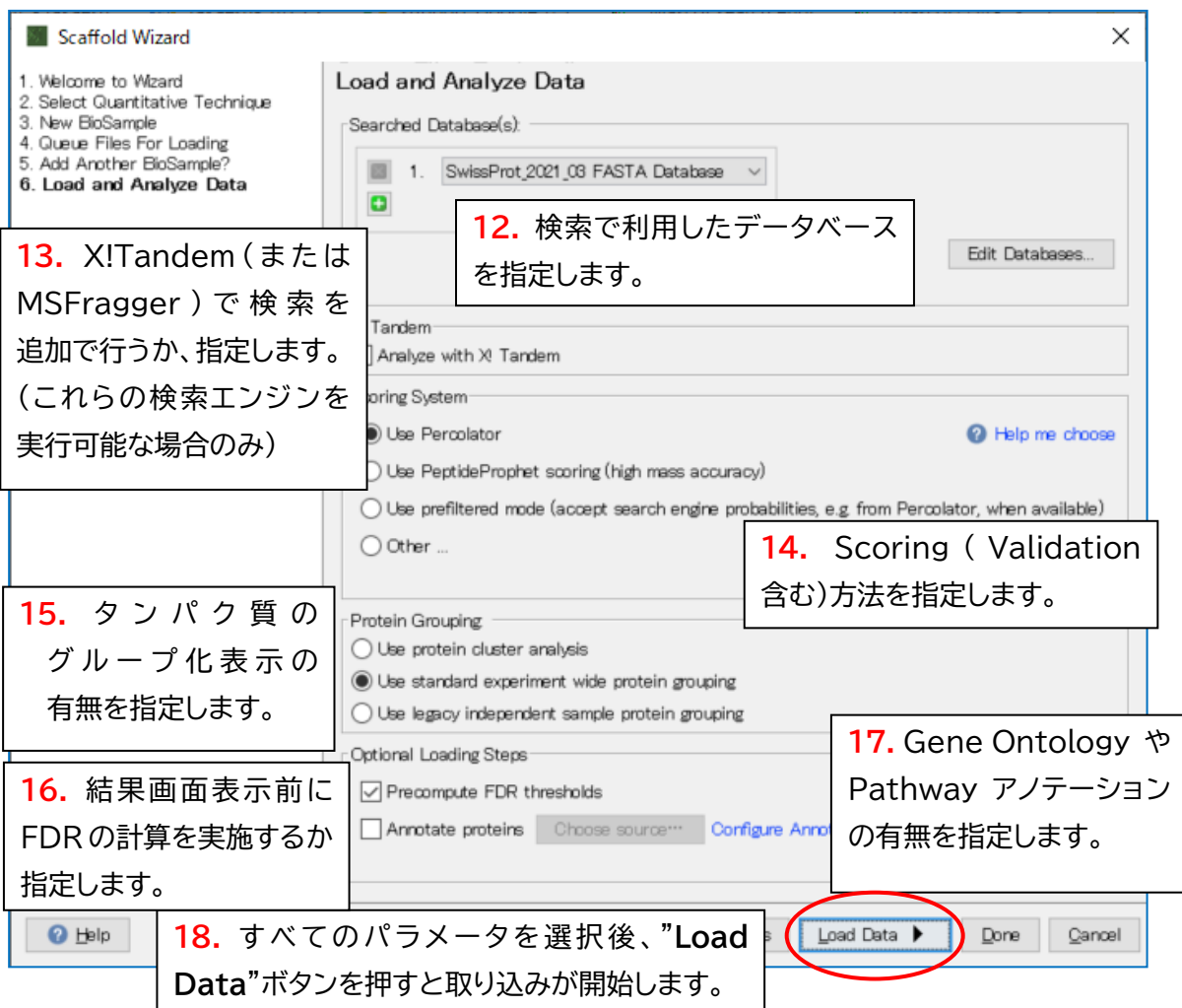
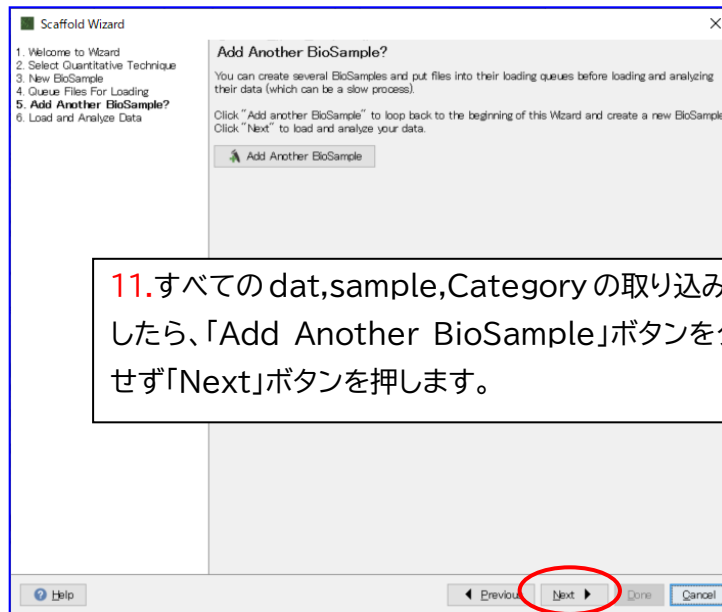


10. これまで同様、取り込み操作を繰り返します。今度は Category 名と Sample 名を変える事と、先ほどと異なる dat ファイルを選択する事に注意してください。



## [Treatment, Sample4 の指定]

これまで同様取り込み操作を繰り返します。

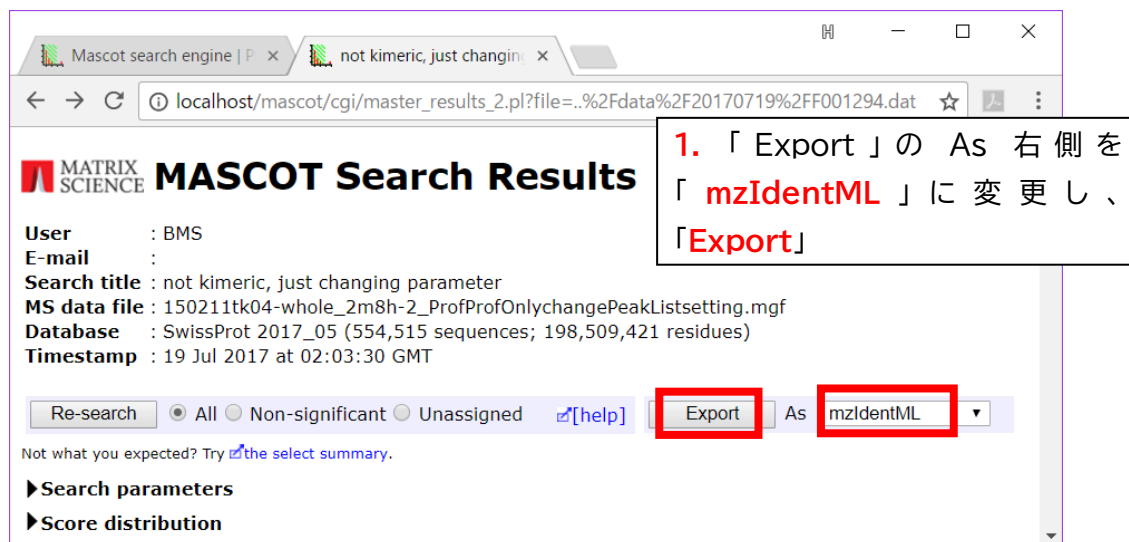


以上で dat のデータ取り込みは終了です。

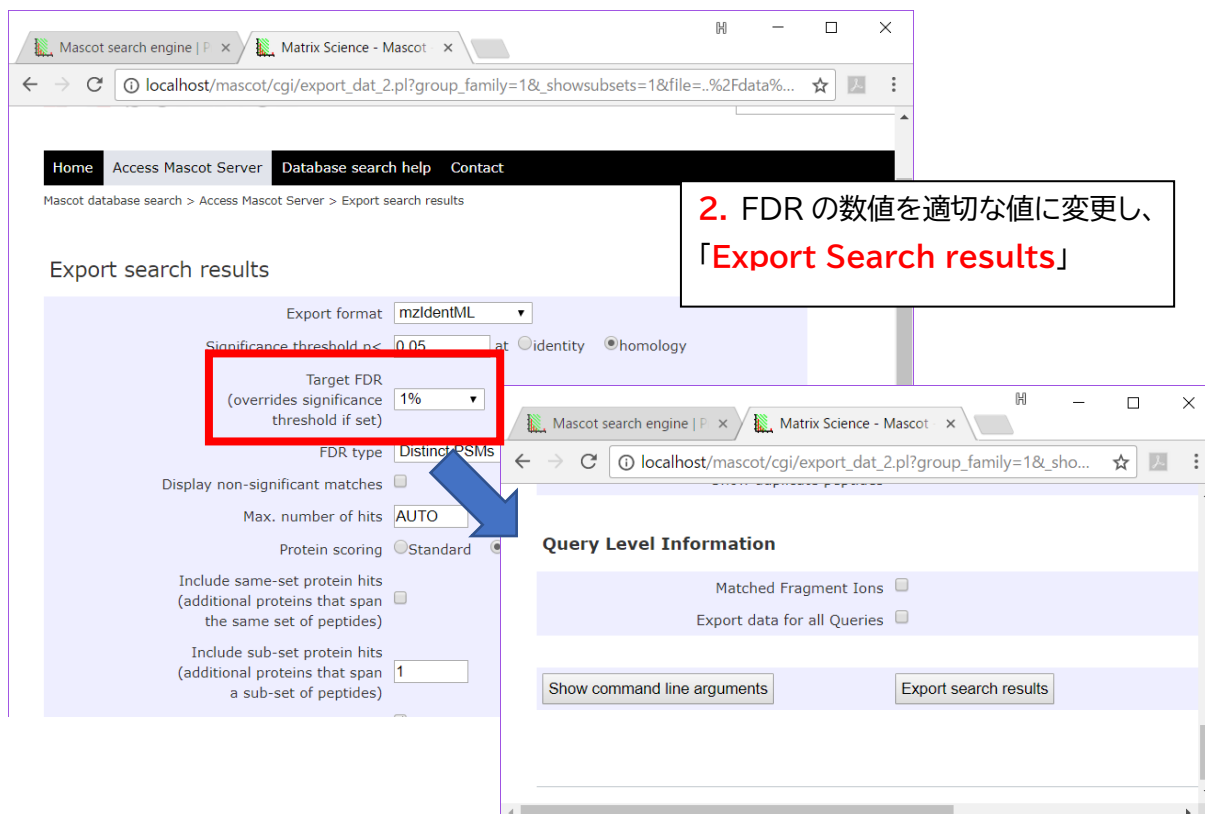
## 2-2-2. mzIdentML 読み込み: validation は取り込み前に行う方法

### [mzIdentML 並びに MGF ファイルの準備]

Scaffold で取り込みたい結果について、**MASCOT** の結果画面にて、「Export」の As 右側の選択肢を「**mzIdentML**」に変更し、「**Export**」ボタンを押します。



ファイル出力時の条件設定を行う画面に移行します。「**Target FDR**」などの数値を 1% など適切な値に変更し、画面下部の「**Export search results**」ボタンを押します。

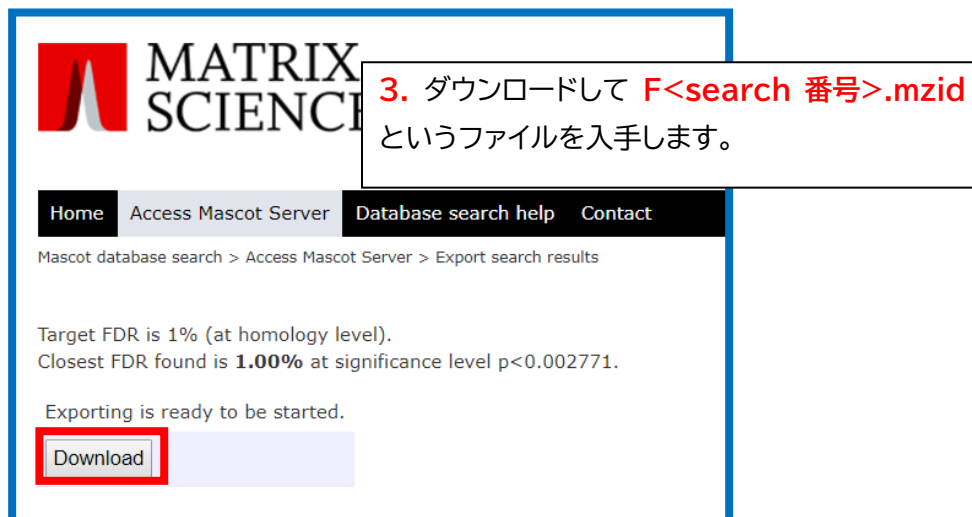




遷移した画面で「Download」ボタンを押します。出力されるファイルは、「**F(search 番号).mzid**」という名称のファイルとなり、ブラウザの既定のダウンロードフォルダなどに保存されます。

この mzIdentML ファイルの作成を、Scaffold にて取り込みたい結果で繰り返し行います\*1。

\*1 Mascot Daemon にて、検索実行と連動して自動的に mzIdentML ファイルを作成するオプションが準備されています。

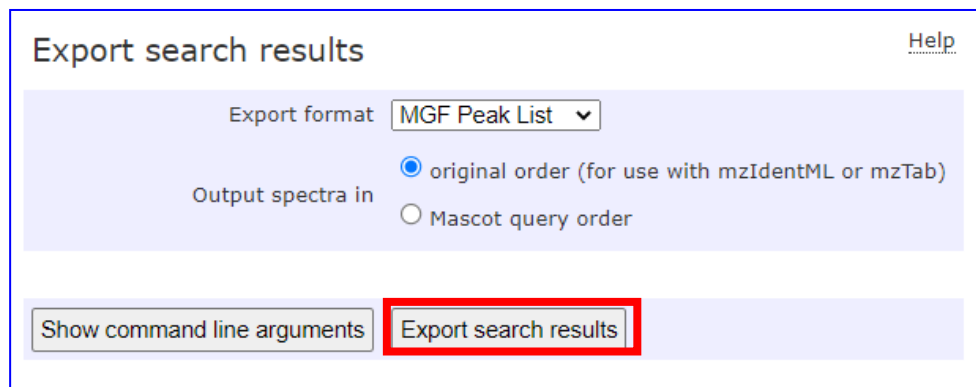


この mzid ファイルには各 query のピーク情報が含まれていません。mzid と同じフォルダ上にピークリストファイルである MGF ファイルを出力しておく必要があります。MGF の出力も結果画面の Export から可能です。

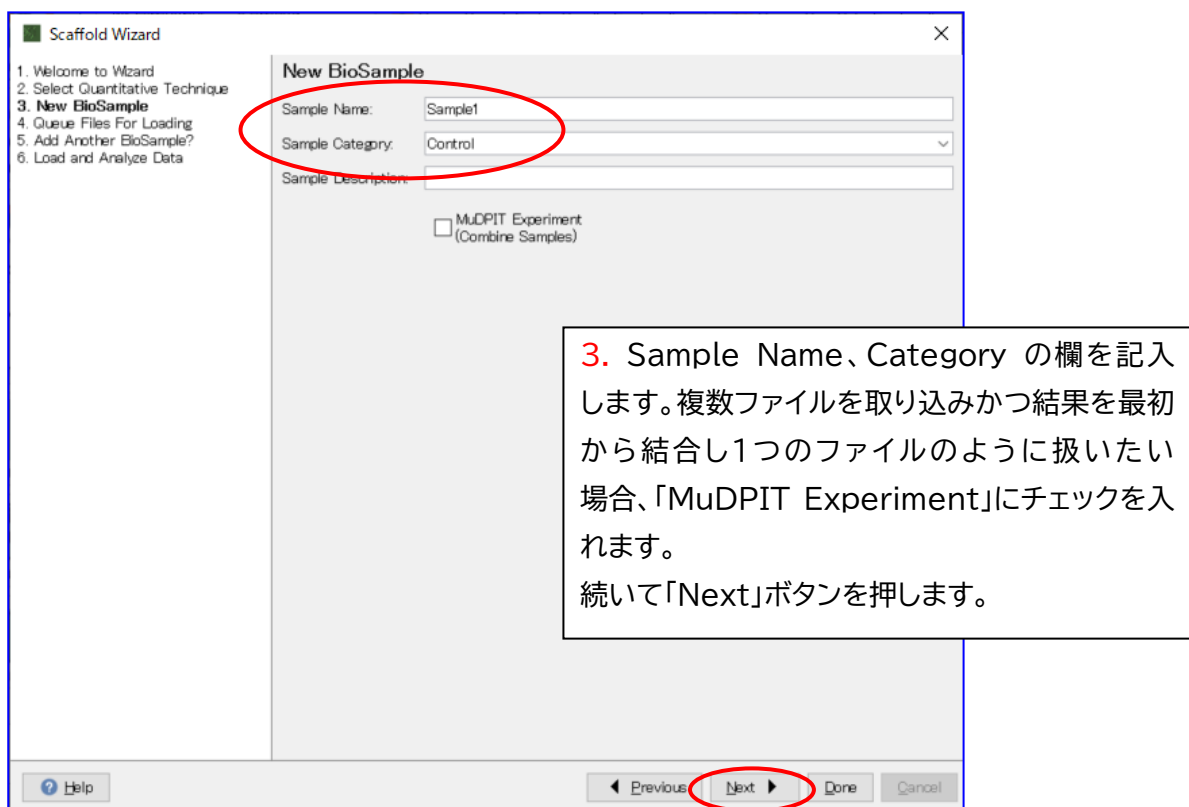
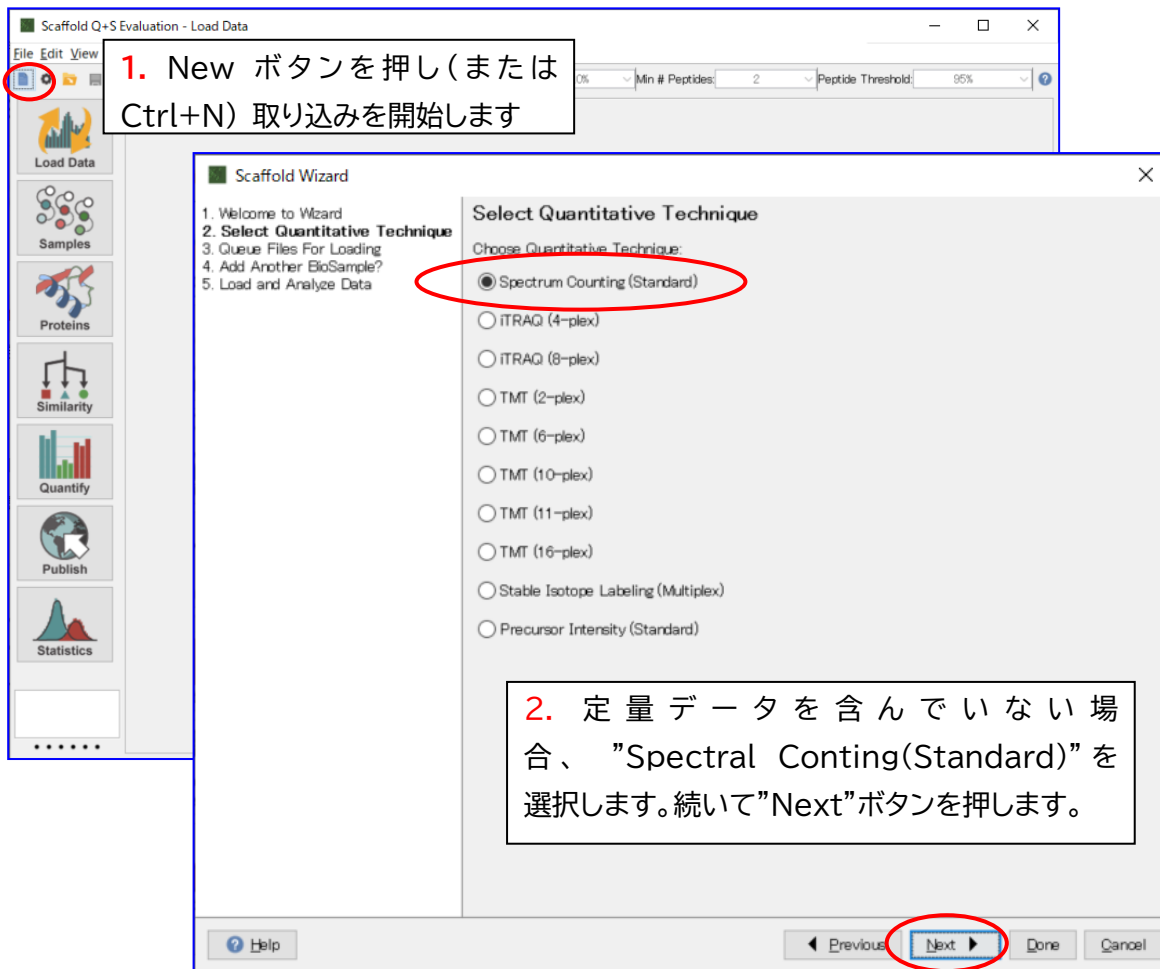
結果画面の Export format で「**MGF Peak List**」を選択すると、画面が遷移します。

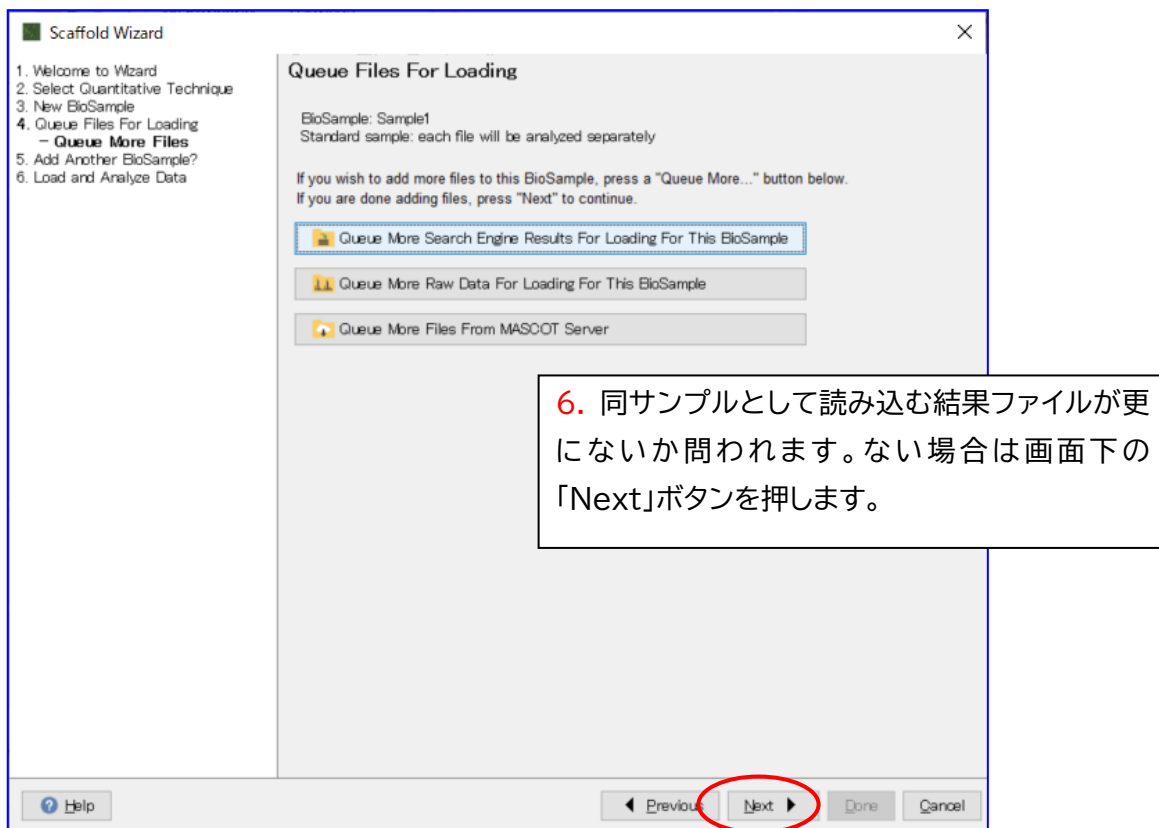
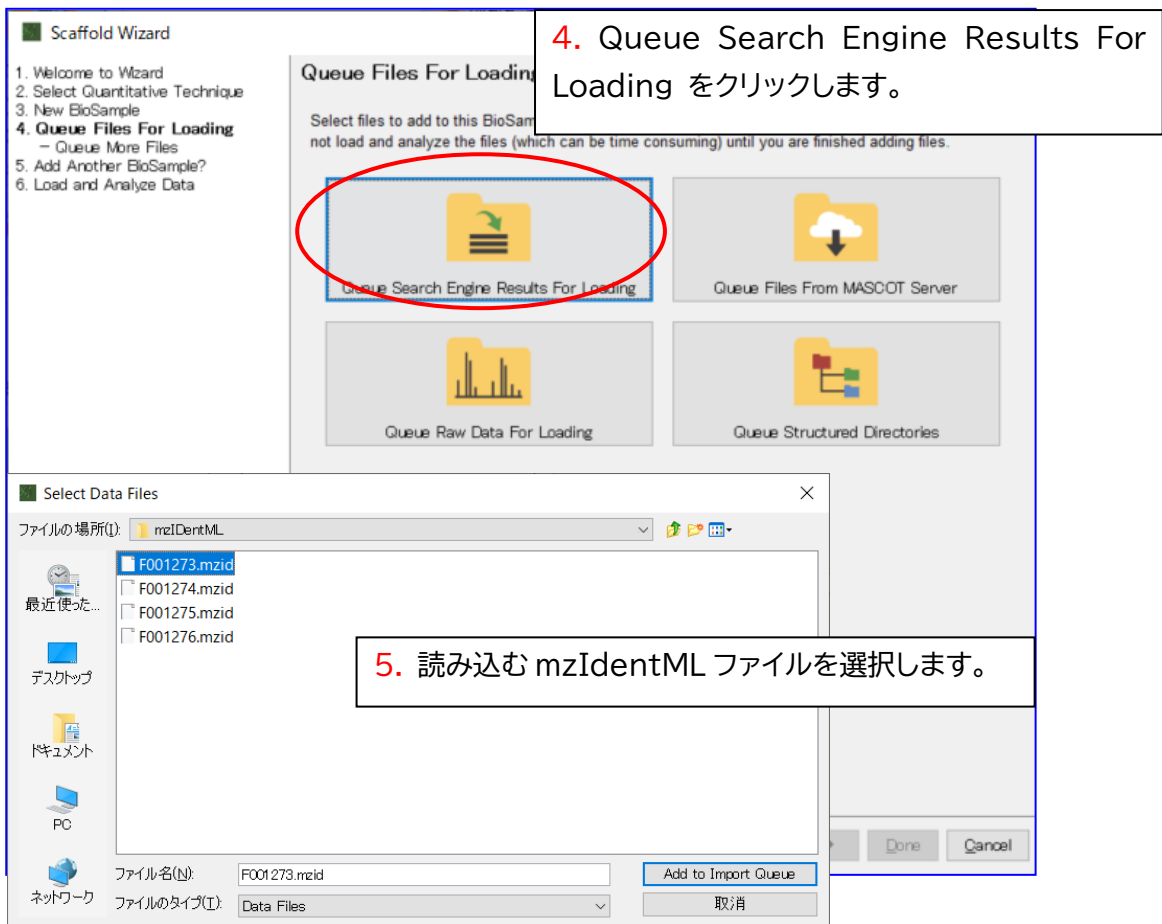


画面下の「Export search results」ボタンを押すと MGF ファイルがダウンロードされます。



すべての mzid ファイルを準備したら、Scaffold を起動しデータの取り込みを開始します。手順は以下の通りです。





7. 先程の指定とは別の **Sample** として読み込む結果ファイルがないか問われます。ある場合は「Add Another BioSample」ボタンをクリックします。

Scaffold Wizard

- Welcome to Wizard
- Select Quantitative Technique
- New BioSample
- Queue Files For Loading
- Add Another BioSample?**
- Load and Analyze Data

Add Another BioSample

You can create several BioSamples from their data (which can be time consuming) until you are finished adding files.

Click "Add another BioSample" to add more files. Click "Next" to load and analyze the files.

**Add Another BioSample**

Scaffold Wizard

- Welcome to Wizard
- Select Quantitative Technique**
- Queue Files For Loading
- Add Another BioSample?
- Load and Analyze Data

Select Quantitative Technique

Choose Quantitative Technique:

Spectrum Counting (Standard)

iTRAQ (4-plex)

Scaffold Wizard

- Welcome to Wizard
- Select Quantitative Technique
- Queue Files For Loading
- Queue Files For Loading**
- Add Another BioSample?
- Load and Analyze Data

Queue Files For Loading

Select files to add to this BioSample. You will have the opportunity to add more files later. Scaffold will not load and analyze the files (which can be time consuming) until you are finished adding files.

Queue Search Engine Results For Loading

Queue Files From MASCOT Server

Structured Directories

Scaffold Wizard

- Welcome to Wizard
- Select Quantitative Technique
- New BioSample**
- Queue Files For Loading
- Add Another BioSample?
- Load and Analyze Data

New BioSample

Sample Name: **BioSample 2**

Sample Category: Control

Sample Description:

MuDPIT Experiment (Combine Samples)

8. 2~7 の操作を繰り返します。Sample 名指定の際名称を変える事と、先ほどと異なるmzIdentML ファイルを選択する事に注意してください。

Scaffold Wizard

- Welcome to Wizard
- Select Quantitative Technique
- New BioSample
- Queue Files For Loading
- Add Another BioSample?
- Load and Analyze Data

BioSample: Sample1  
Standard sample: each file will be analyzed separately

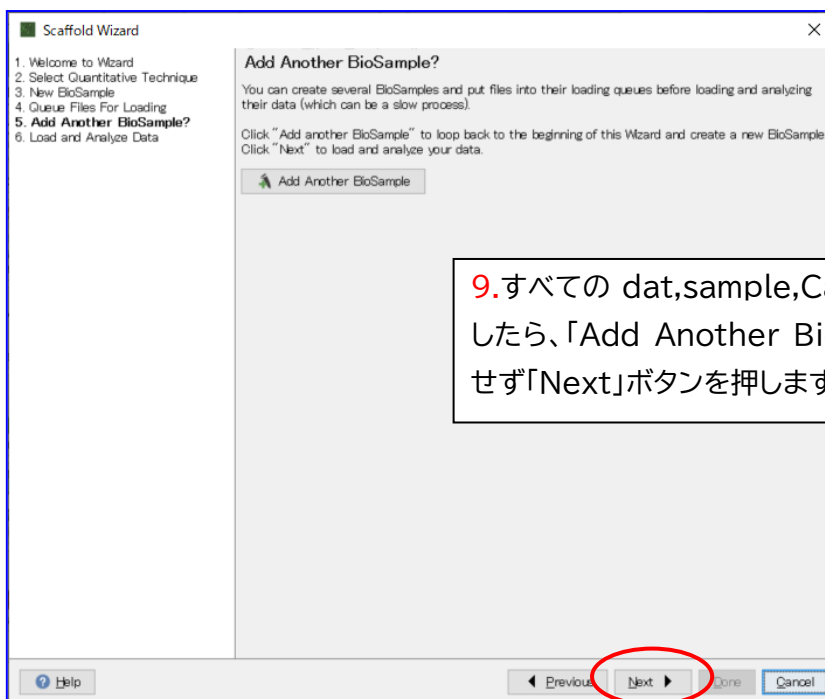
If you wish to add more files to this BioSample, press a "Queue More..." button below. If you are done adding files, press "Next" to continue.

Queue More Search Engine Results For Loading For This BioSample

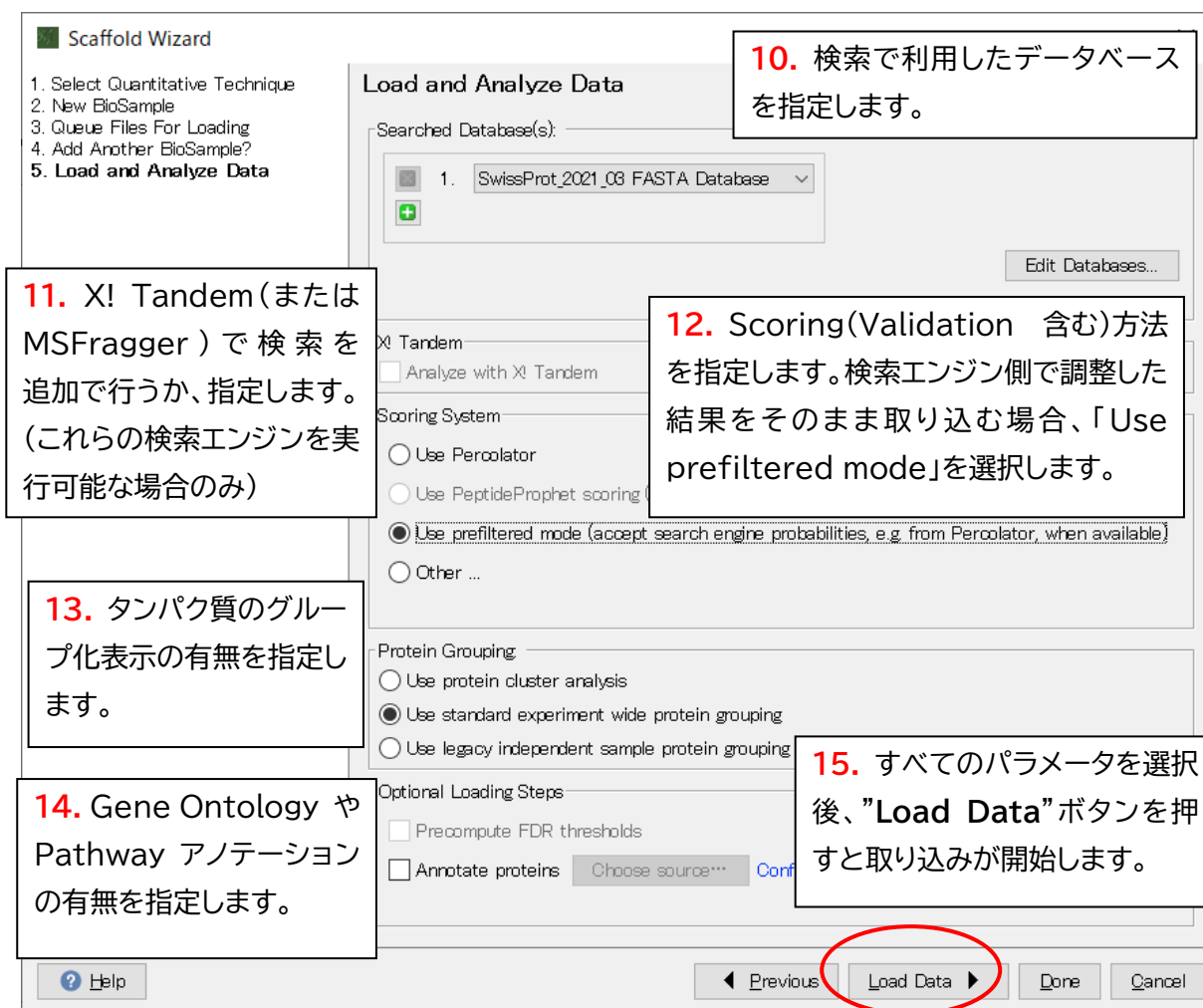
Queue More Raw Data For Loading For This BioSample

Queue More Files From MASCOT Server

Help Previous Next Done Cancel



9.すべての dat,sample,Category の取り込みが終了したら、「Add Another BioSample」ボタンをクリックせず「Next」ボタンを押します。



10. 検索で利用したデータベースを指定します。

11. X! Tandem(または MSFragger) で検索を追加で行うか、指定します。(これらの検索エンジンを実行可能な場合のみ)

12. Scoring(Validation 含む)方法を指定します。検索エンジン側で調整した結果をそのまま取り込む場合、「Use prefiltered mode」を選択します。

13. タンパク質のグループ化表示の有無を指定します。

14. Gene Ontology や Pathway アノテーションの有無を指定します。

15. すべてのパラメータを選択後、「Load Data」ボタンを押すと取り込みが開始します。

以上で mzIdentML の取り込みは終了です。

## 2-3. MASCOT Server からネットワークを介して dat を取得する方法

MASCOT の結果を Scaffold に取り込む 2 つの方法のうちの1つ、MASCOT のログファイルを開き、Mascot Server からネットワークを通じて直接 dat ファイルを取り込む方法です。

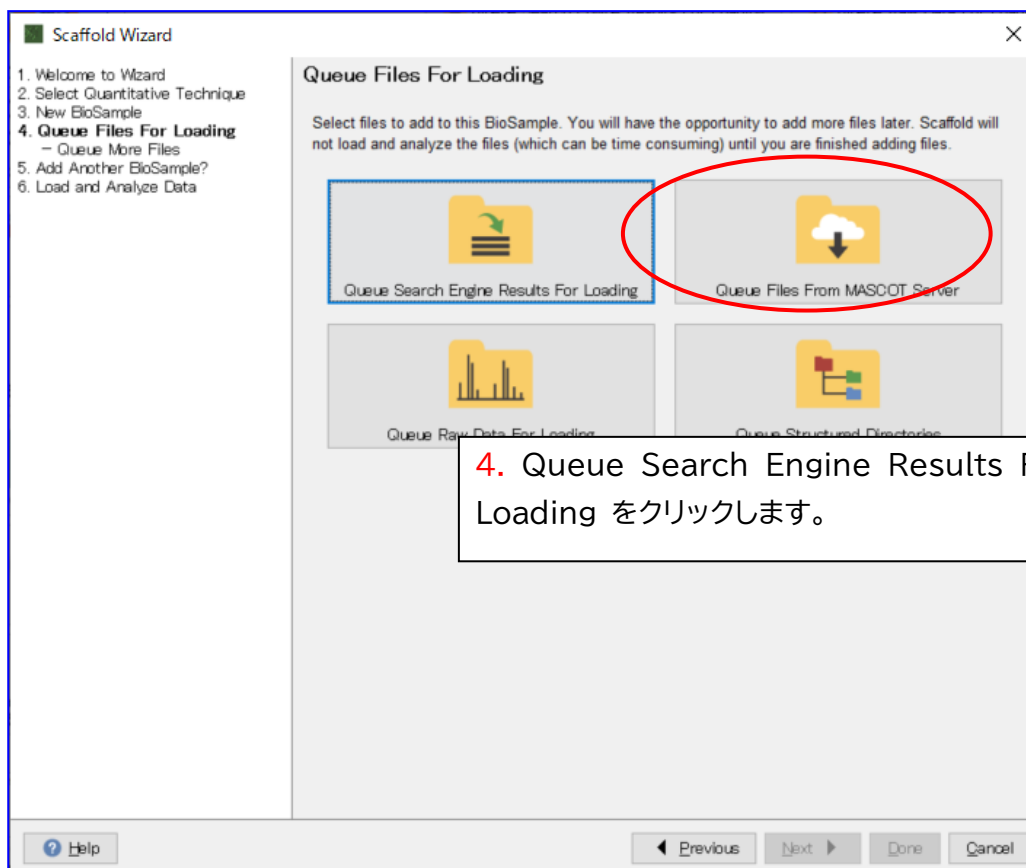
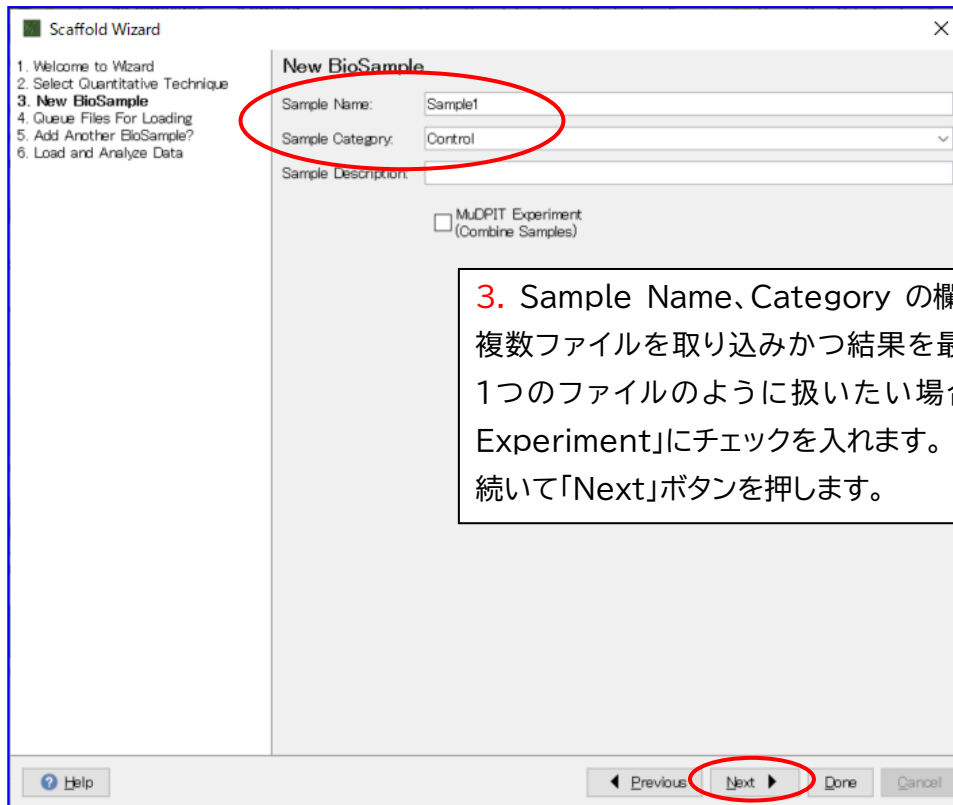
### [ Sample を Search log から指定しネットワーク経由でファイルを取得する方法 ]

Category として Control と Treatment, それぞれに Sample を2種類属したデータを想定しています。

Category	Sample	dat
Control	Sample1	F001244.dat
	Sample2	F001245.dat
Treatment	Sample3	F001246.dat
	Sample4	F001247.dat

### [Control, Sample1 の指定]

The image shows two overlapping windows from the Scaffold Q+S Evaluation software. The top window is the main application interface with a toolbar. A red circle highlights the 'Load Data' button (represented by a folder icon). A text box with a black border contains the instruction: "1. New ボタンを押し(または Ctrl+N) 取り込みを開始します". The bottom window is the 'Scaffold Wizard' dialog box. It has a list of steps on the left, with step 2, 'Select Quantitative Technique', highlighted. On the right, under 'Choose Quantitative Technique:', the radio button for 'Spectrum Counting (Standard)' is selected and circled in red. Other options include ITRAQ (4-plex), ITRAQ (8-plex), TMT (2-plex), TMT (6-plex), TMT (10-plex), TMT (11-plex), TMT (16-plex), Stable Isotope Labeling (Multiplex), and Precursor Intensity (Standard). At the bottom of the dialog, the 'Next' button is circled in red. A text box with a black border contains the instruction: "2. 定量データを含んでいない場合、"Spectral Conting(Standard)"を選択します。続いて"Next"ボタンを押します。".



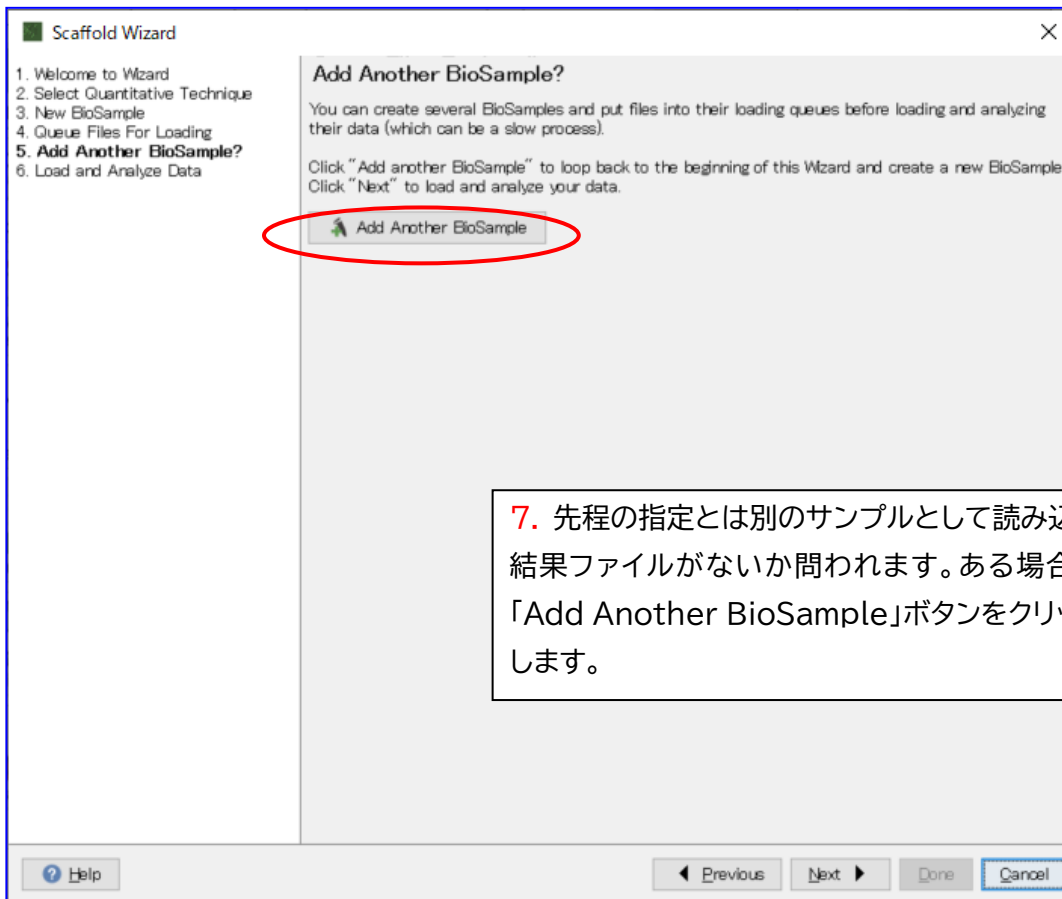
**5. 取り込む結果をログからクリックで選びます(同時に複数選択をすることも可能です)。「Add」ボタンを押すと下の欄に取り込まれた結果が表示されます。選択を終えたら「OK」ボタンを押します。**

Job...	Database	User Name	Email	Title
1243	SwissPro			MS/MS Example
1244	SwissPro	name	short me	20210310 Tutorial Sear
1245	SwissPro	name	short me	20210310 Tutorial Sear
1246	SwissPro	name	short me	20210310 Tutorial Sear
1247	SwissPro	name	short me	20210310 Tutorial Sear
1248	SwissPro	name	short me	20210310 Tutorial Sear
1249	SwissPro	name	short me	20210310 Tutorial Sear

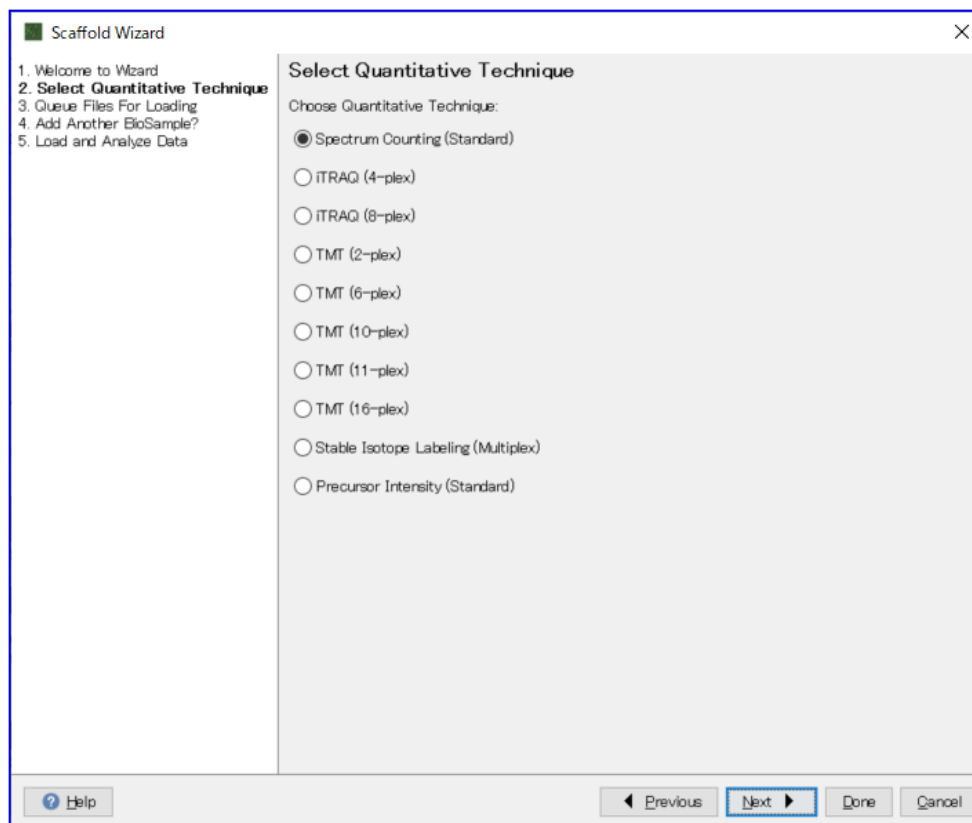
Download Status	Job Num..	Database	User Name	Email
Download complete	1244	SwissPro	name	short memo

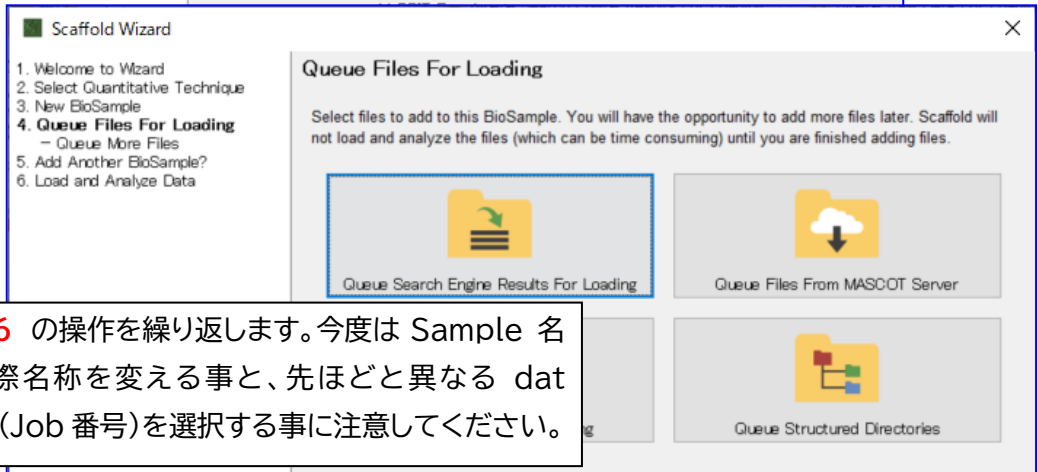
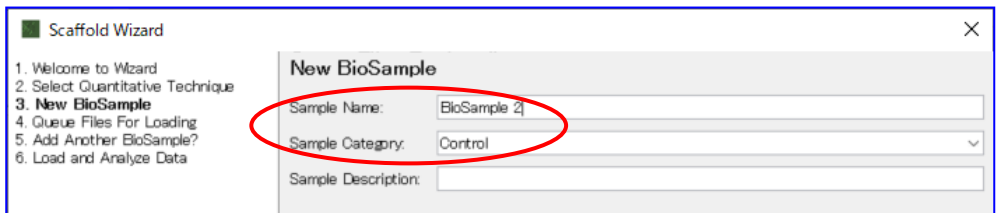
**6. 同サンプルとして読み込む結果ファイルが更がないか問われます。ない場合は画面下の「Next」ボタンを押します。**



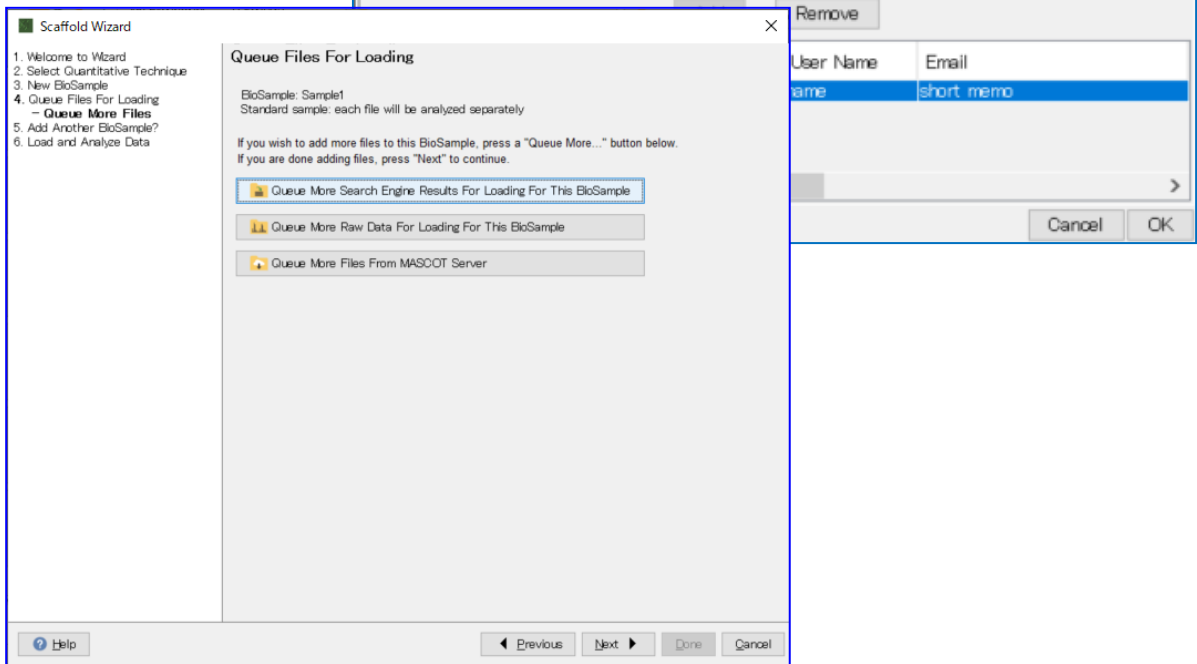
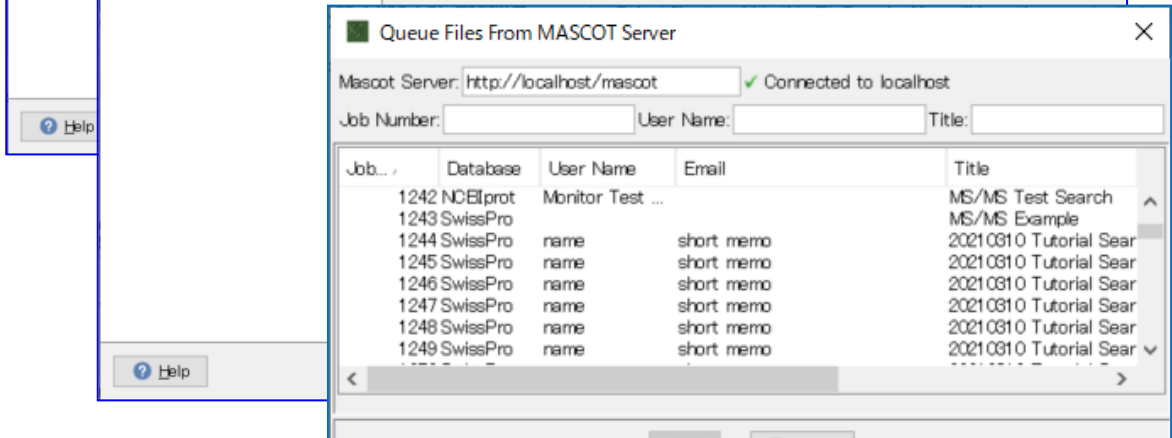


[Control, Sample2 の指定]

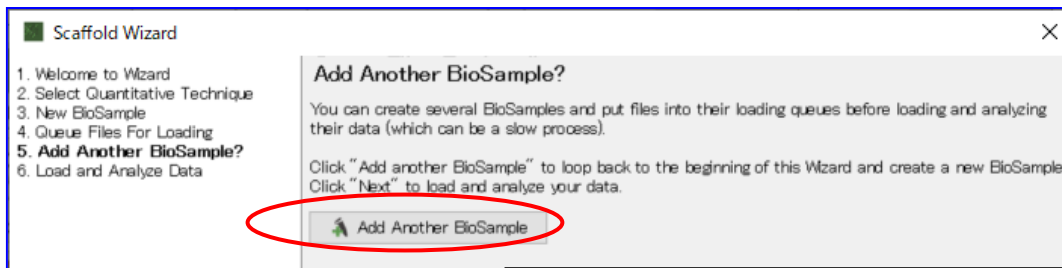




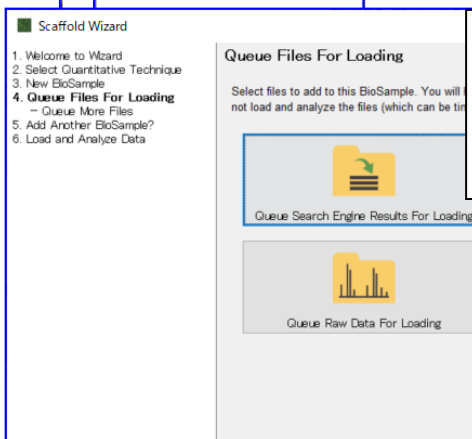
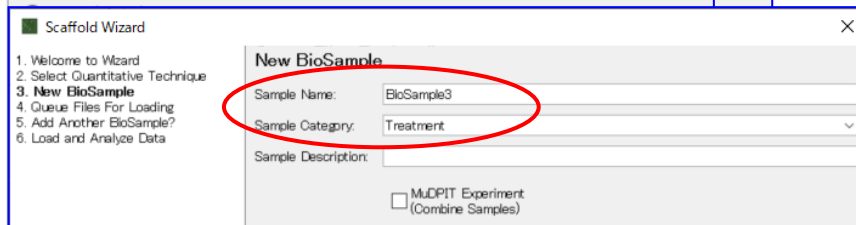
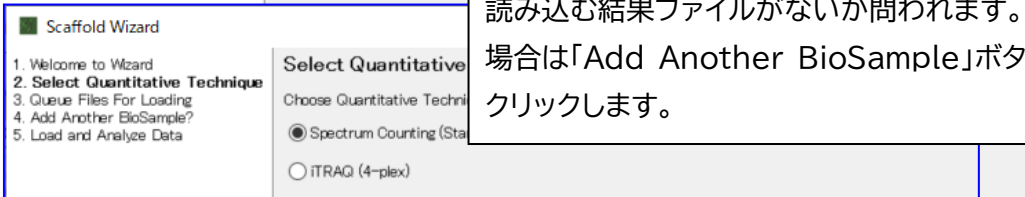
8. 2~6 の操作を繰り返します。今度は Sample 名指定の際名称を変える事と、先ほどと異なる dat ファイル(Job 番号)を選択する事に注意してください。



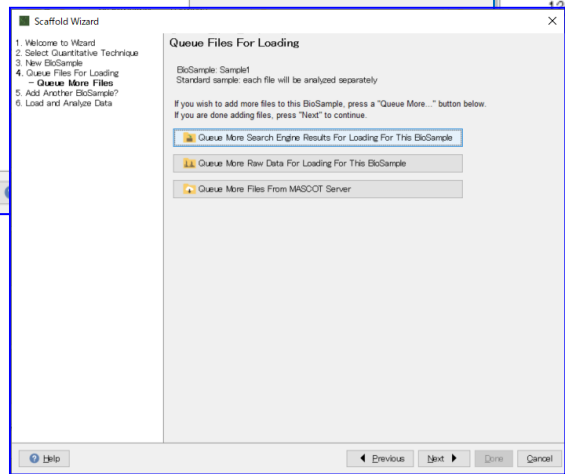
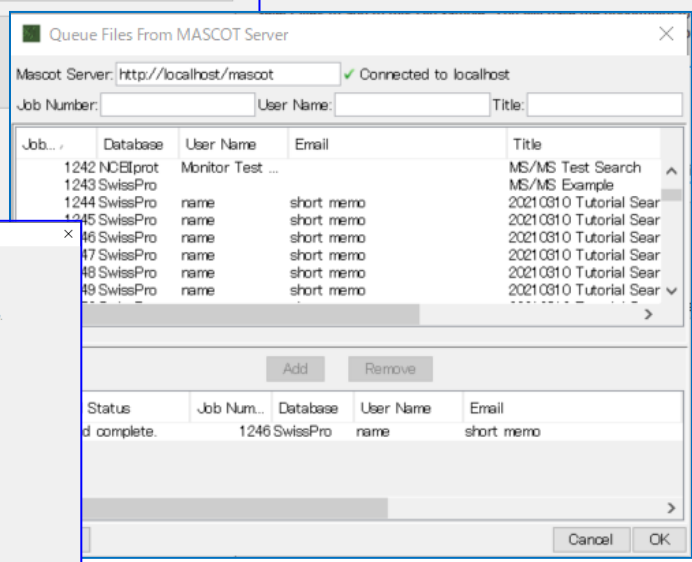
[Treatment, Sample3 の指定]



9. 7 同様、先程の指定とは別のサンプルとして読み込む結果ファイルがないか問われます。ある場合は「Add Another BioSample」ボタンをクリックします。

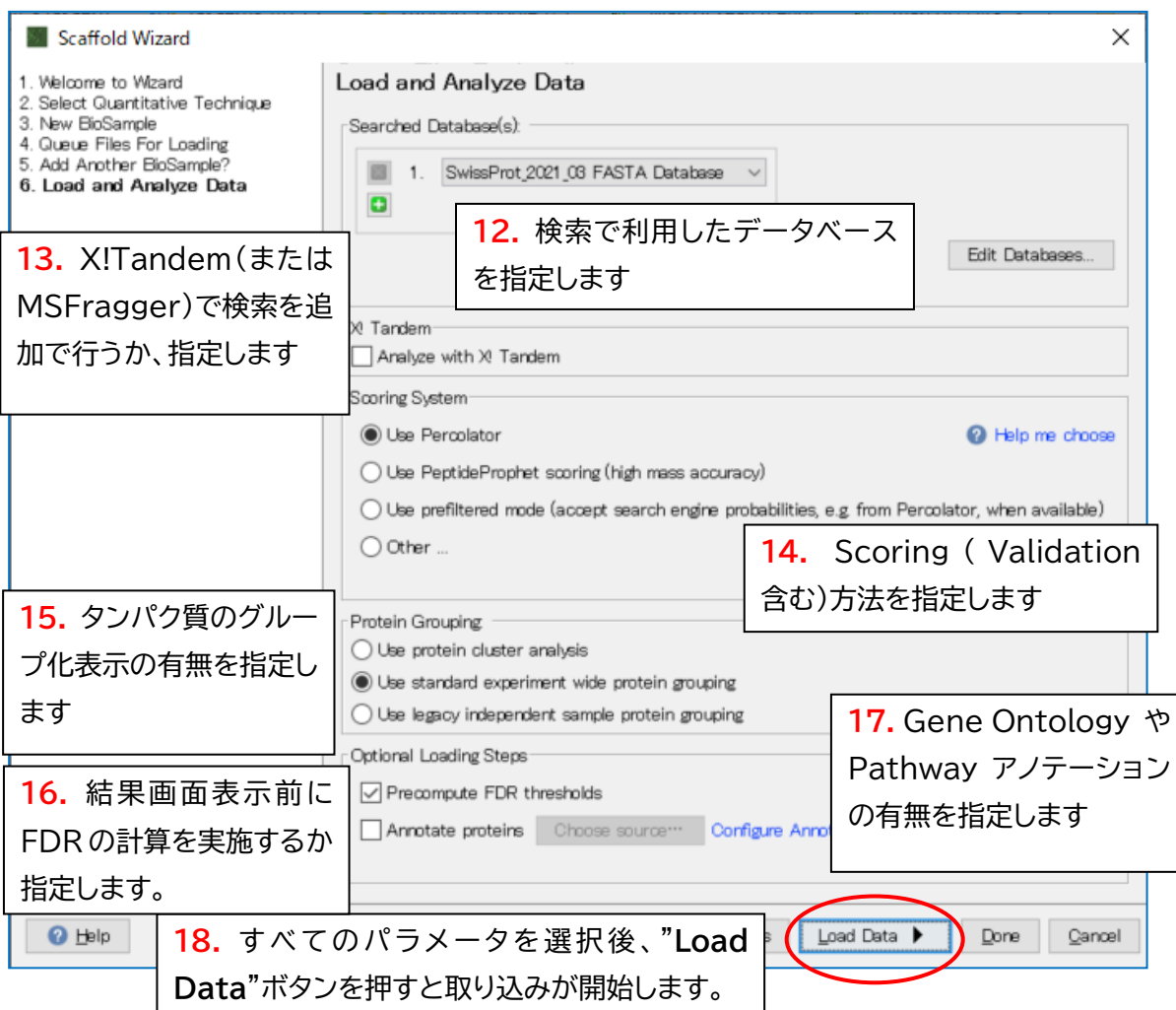
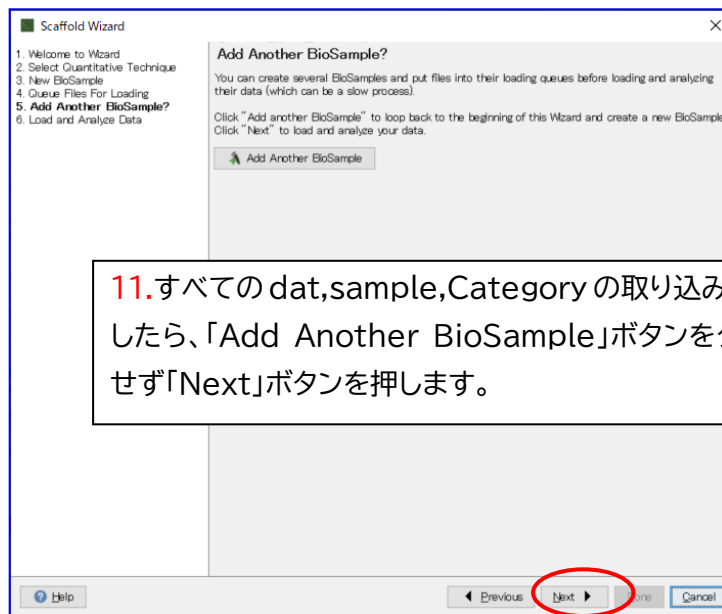


10. これまで同様、取り込み操作を繰り返します。今度は Category 名と Sample 名を変える事と、先ほどと異なる dat ファイル (Job 番号) を選択する事に注意してください。



## [Treatment, Sample4 の指定]

これまで同様取り込み操作を繰り返します。



以上で search log からデータ取り込む操作は終了です。

## 2-4. raw またはピークリストを直接読み込むところから始める方法

Scaffold 5 からは、raw データを取り込み、検索エンジン X! Tandem または MSFragger で検索を行ってその結果を Scaffold 5 上に表示させることができるようになりました。検索機能を利用したい場合、予め各検索エンジンの実行ファイルの場所を Scaffold 上から指定する設定を行う必要があります。それぞれのプログラムをご利用の際にはライセンスについてよくご確認いただき、license agreement に承諾される必要がありますのでご注意ください。また検索エンジンによって読み込み可能なファイルフォーマットが異なりますのでその点もご注意ください。

検索可能なファイルフォーマット

**X! Tandem** : mzXML, MGF, DTA, PKL などのピークリストファイル

**MSFragger** : ThermoFisher raw データ、Bruker .d ディレクトリ、mzML などの raw ファイル

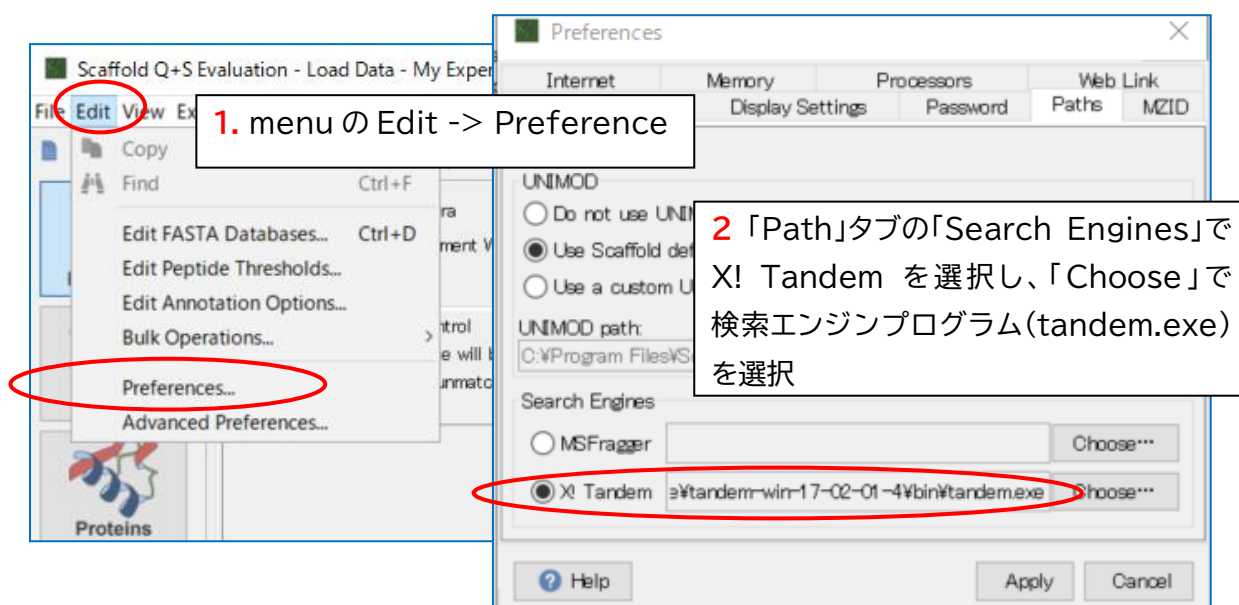
### 2-4-1. 実行する検索エンジンプログラムをセットする : X! Tandem の場合

以前までのバージョンと異なり、現バージョンでは X! Tandem がバンドルされていません。利用するためにはユーザーが X! Tandem をダウンロードして Scaffold と同じコンピューターにインストールする必要があります。

X! Tandem URL は以下にガイドがあります。

<http://www.proteomesoftware.com/search-tools>

インストールした X! Tandem を Scaffold で使用するため、以下の操作を行ってください。



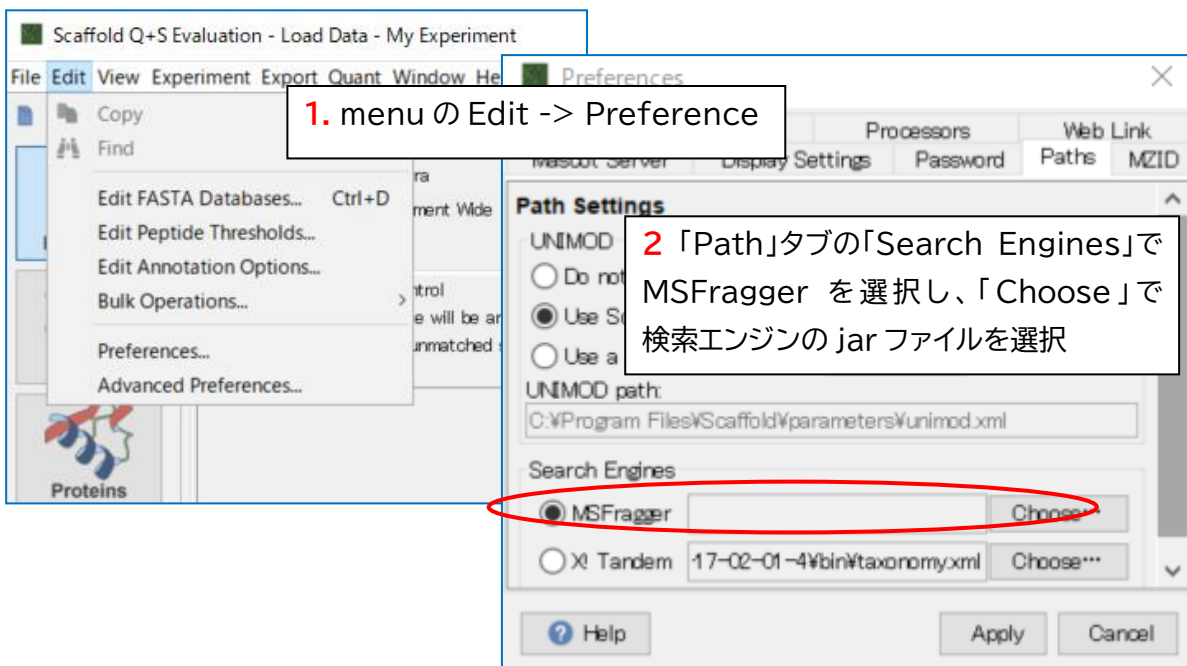
## 2-4-2. 実行する検索エンジンプログラムをセットする : MSFragger の場合

MSFragger を利用するためにはユーザーが MSFragger をダウンロードして Scaffold と同じコンピューターにインストールする必要があります。

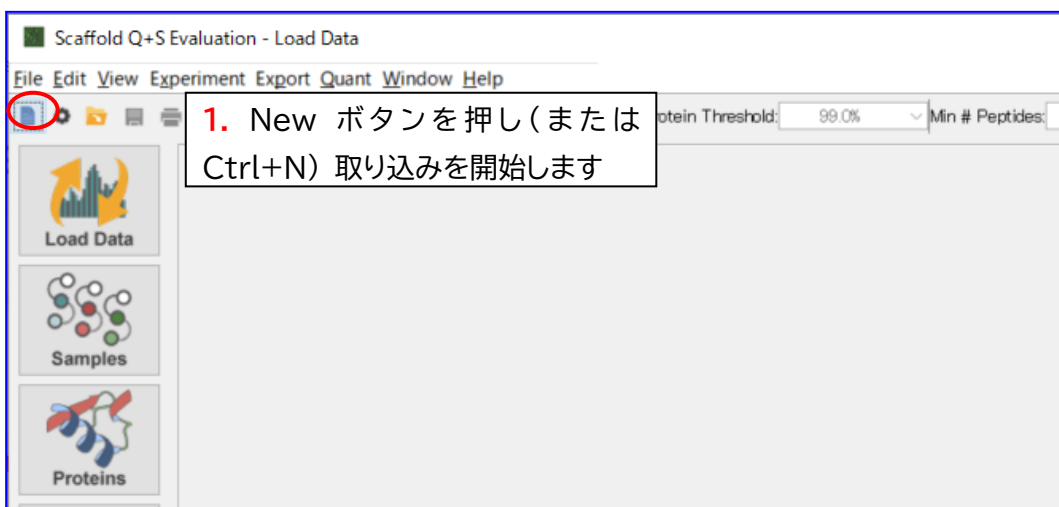
MSFragger URL は以下にガイドがあります。

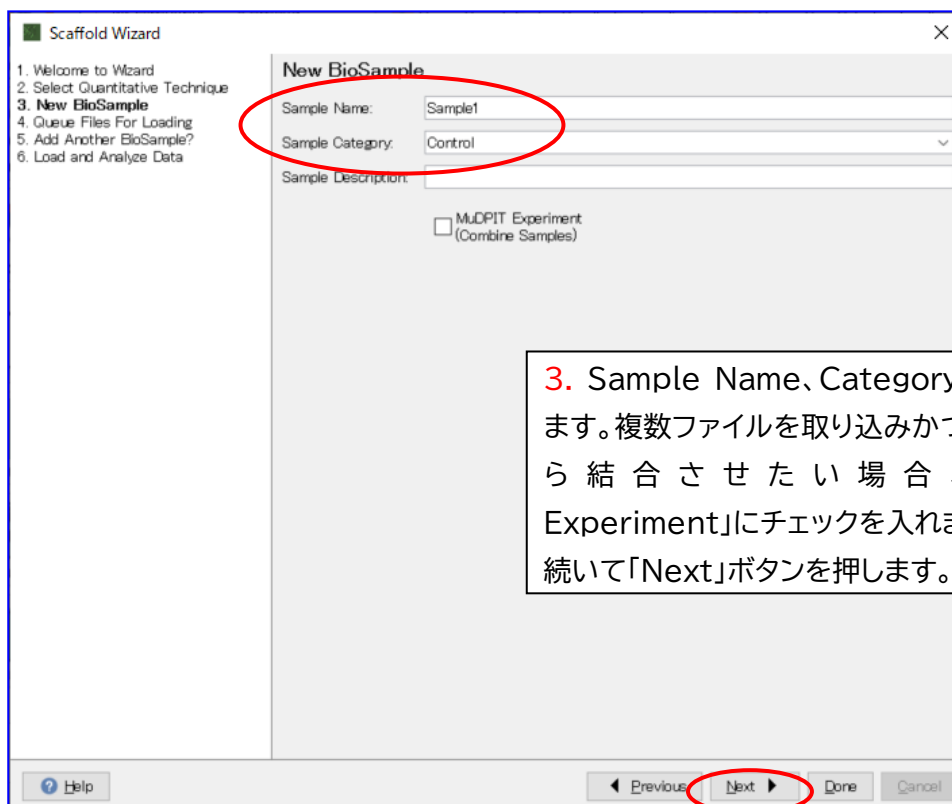
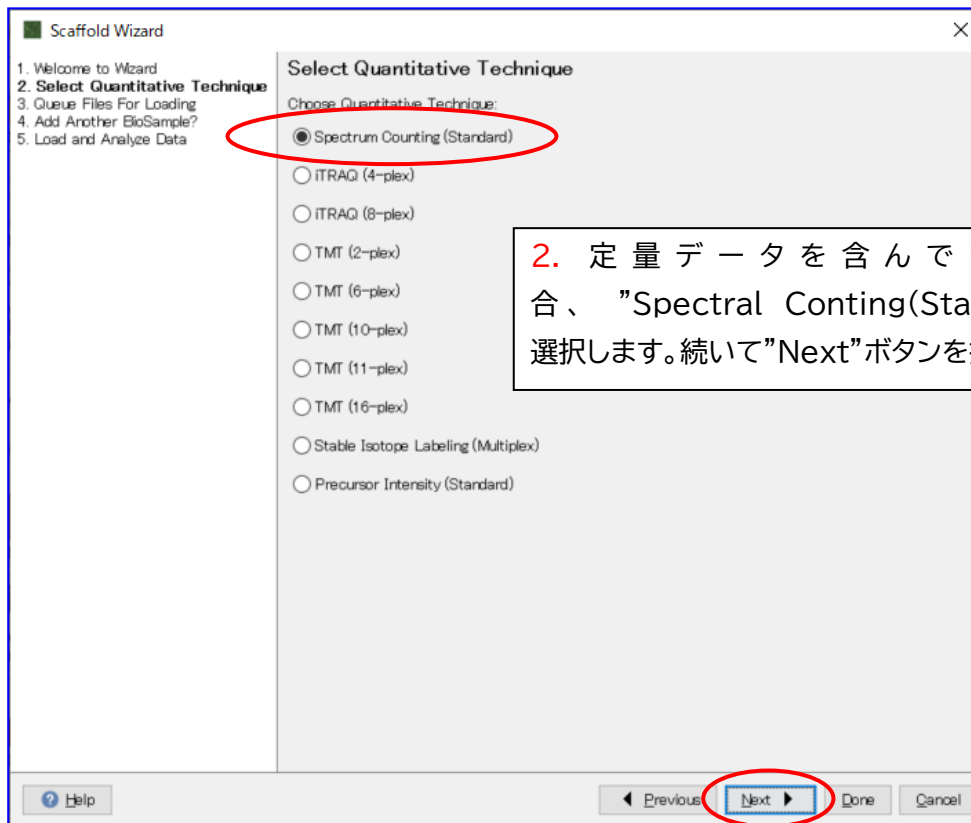
<http://www.proteomesoftware.com/search-tools>

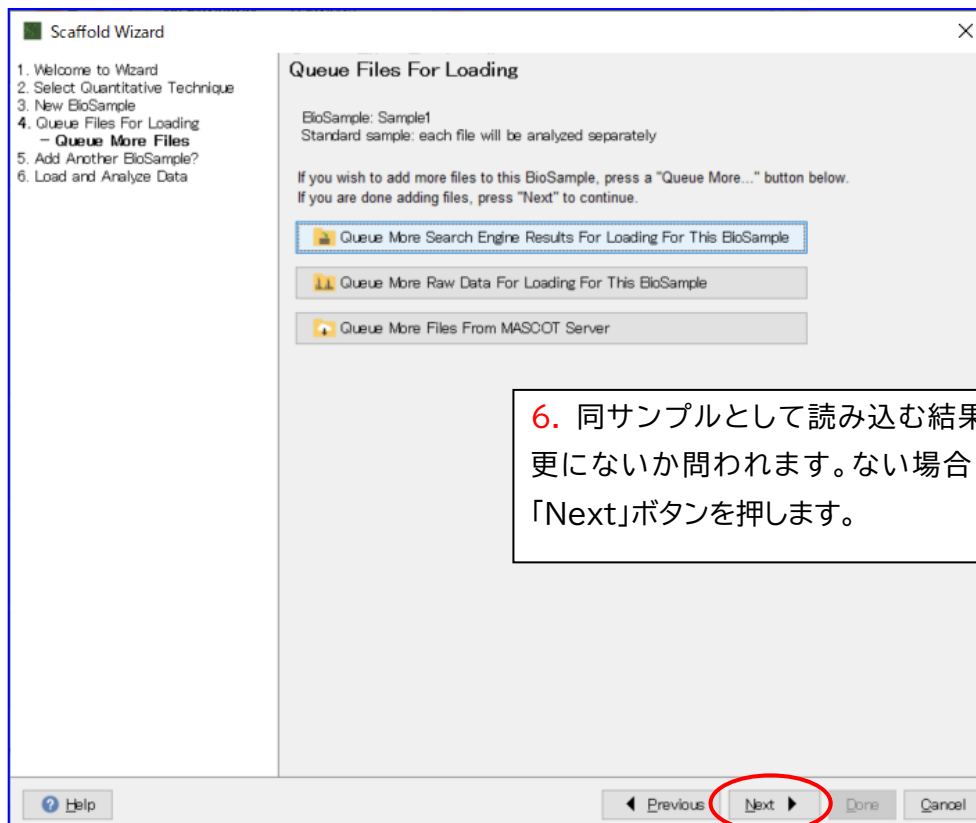
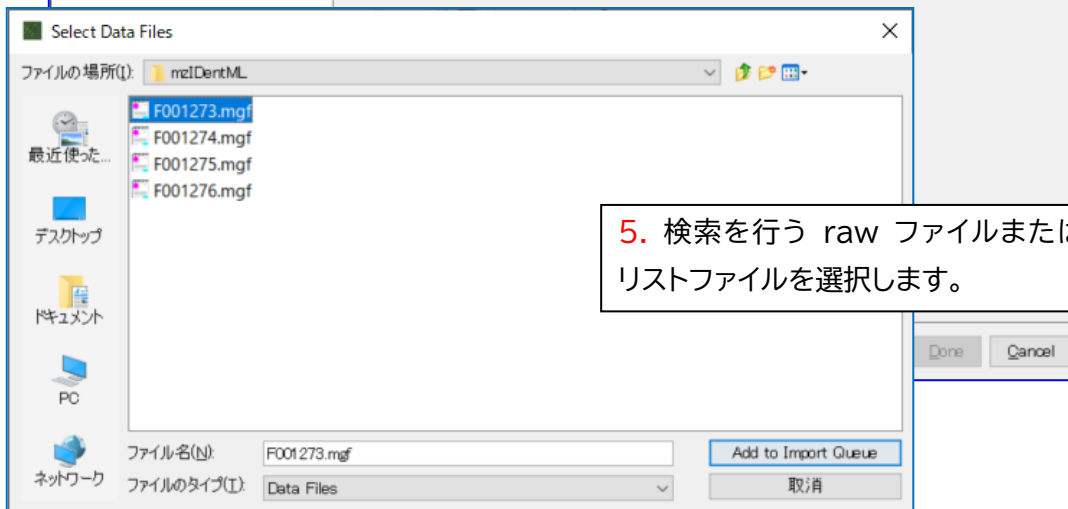
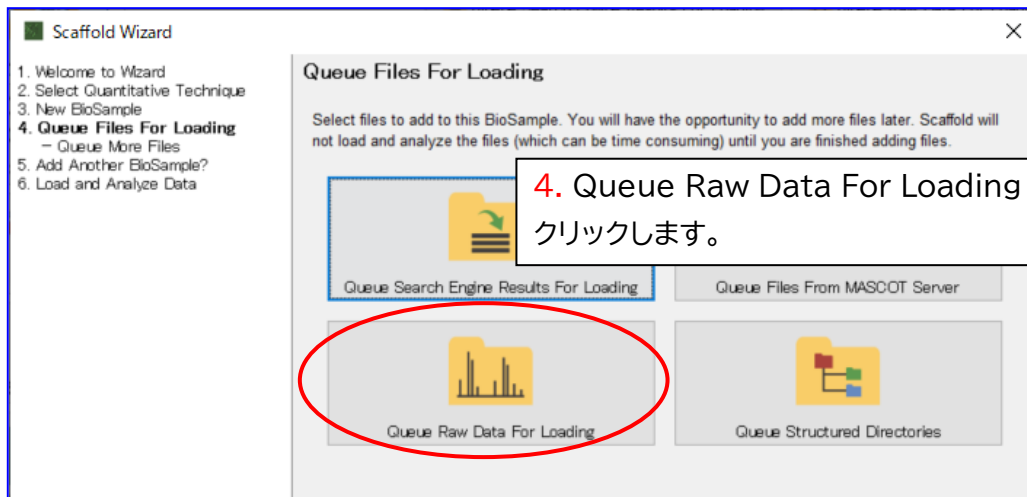
インストールした MSFragger を Scaffold で使用するため、以下の操作を行ってください。



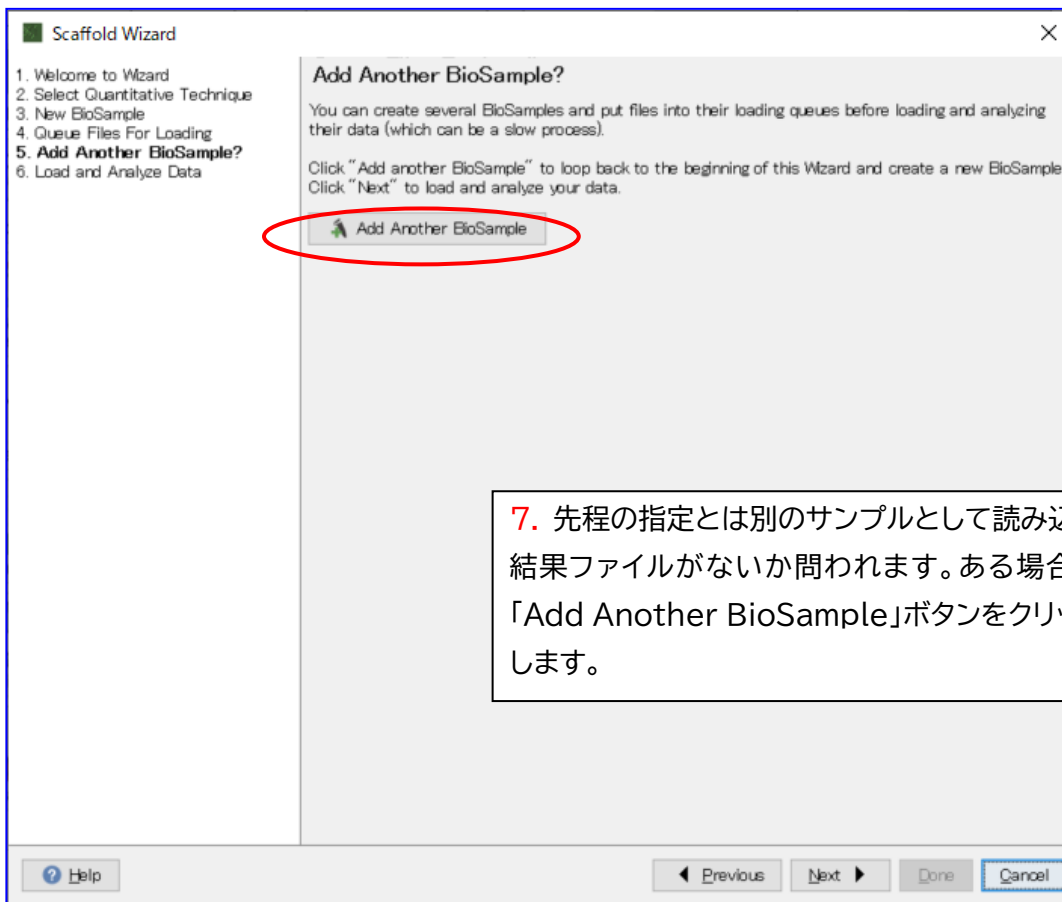
## 2-4-3. MSFragger または X!Tandem での検索実施とデータ読み込み方法



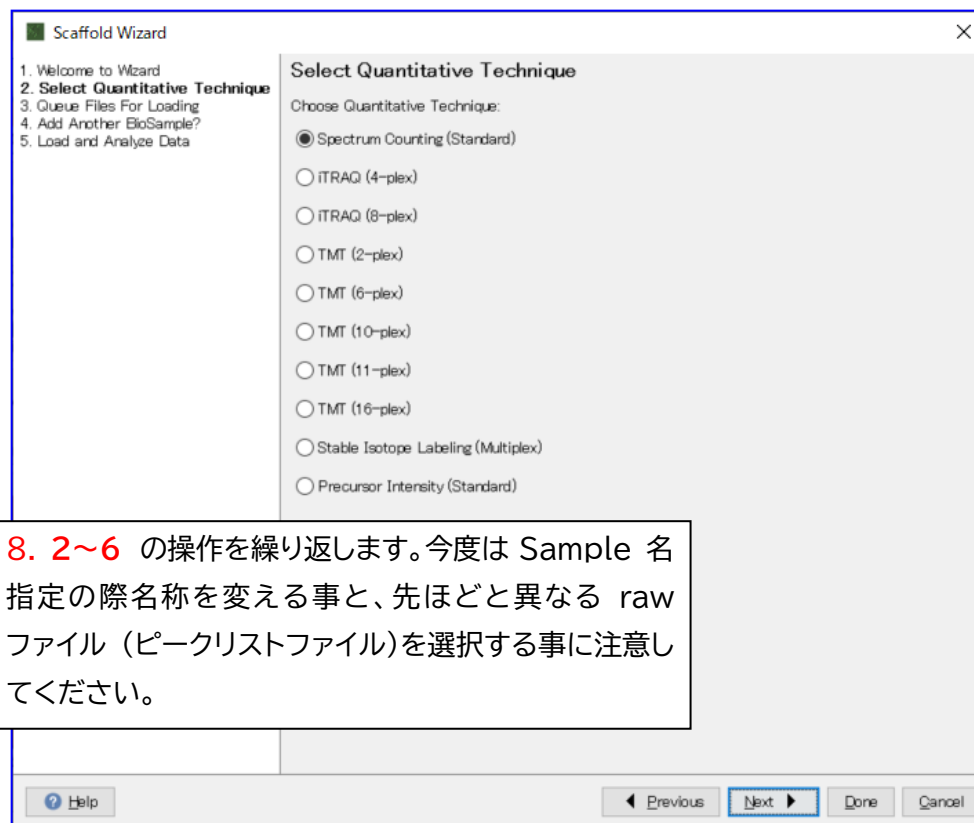




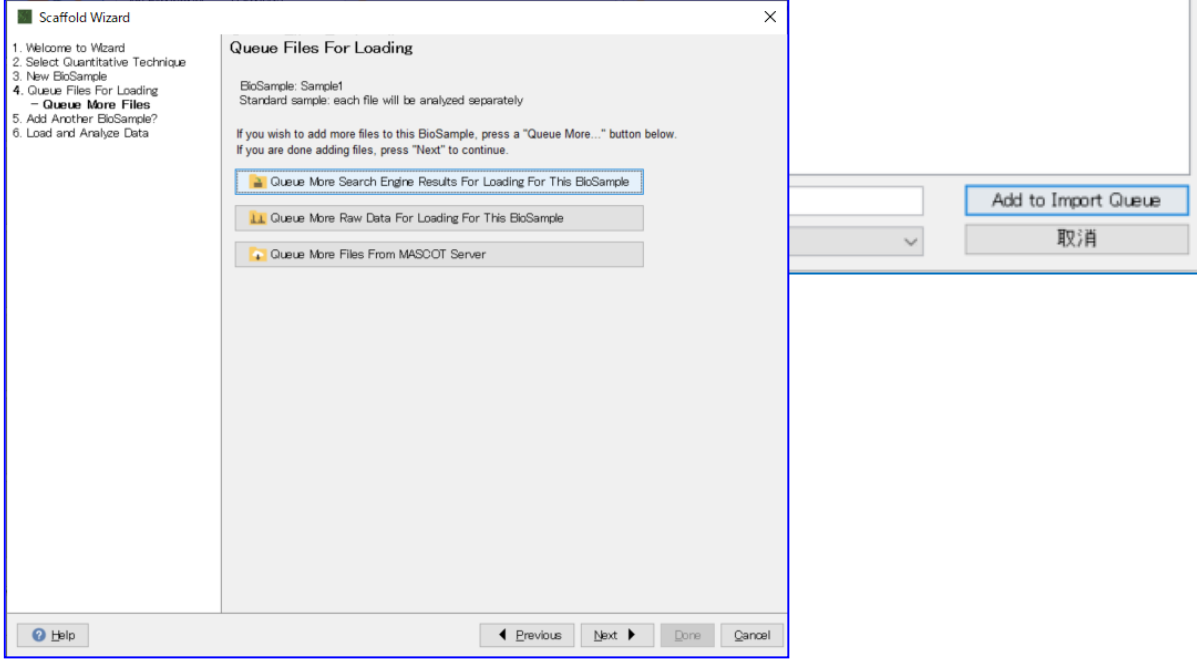
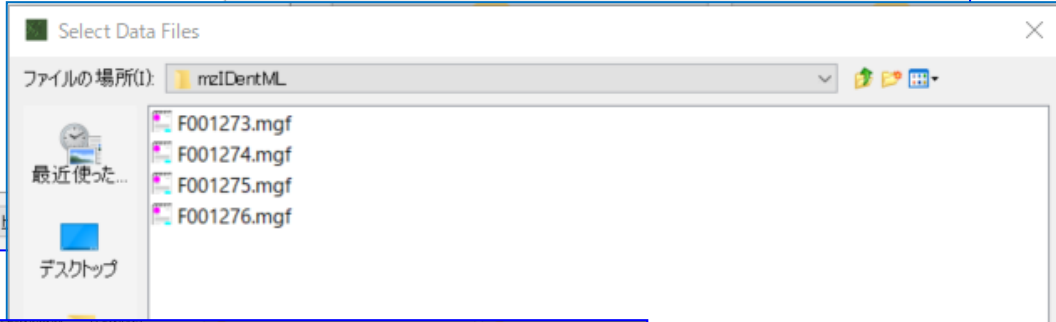
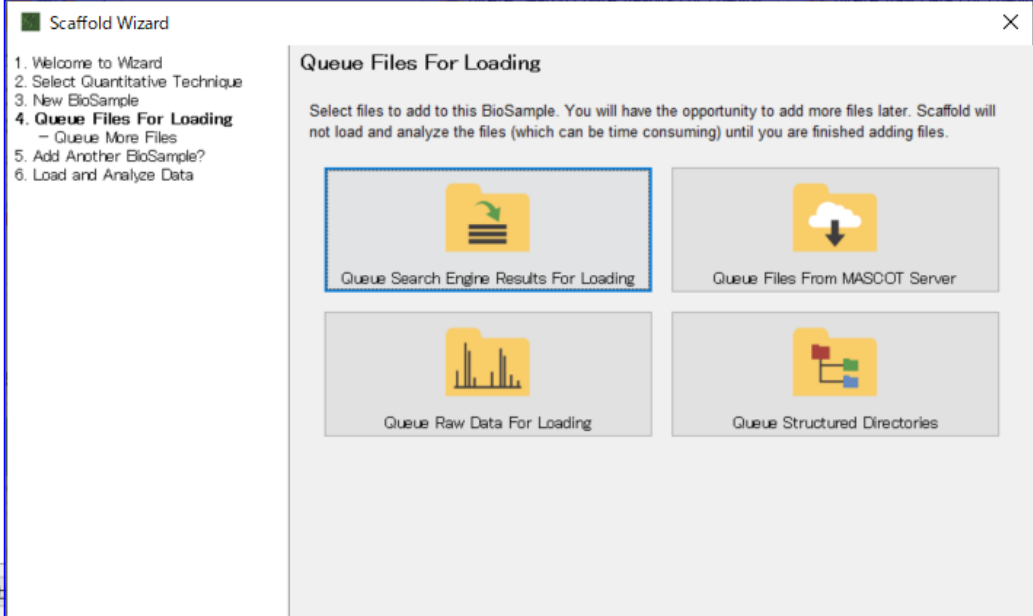
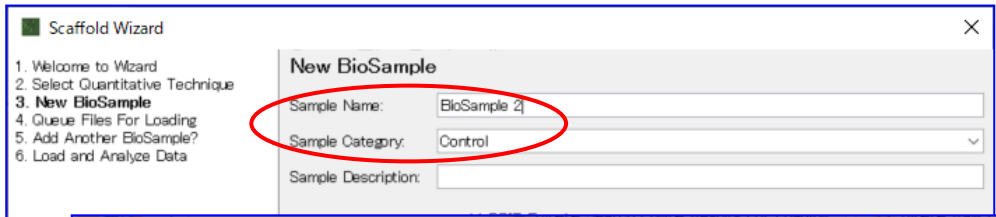


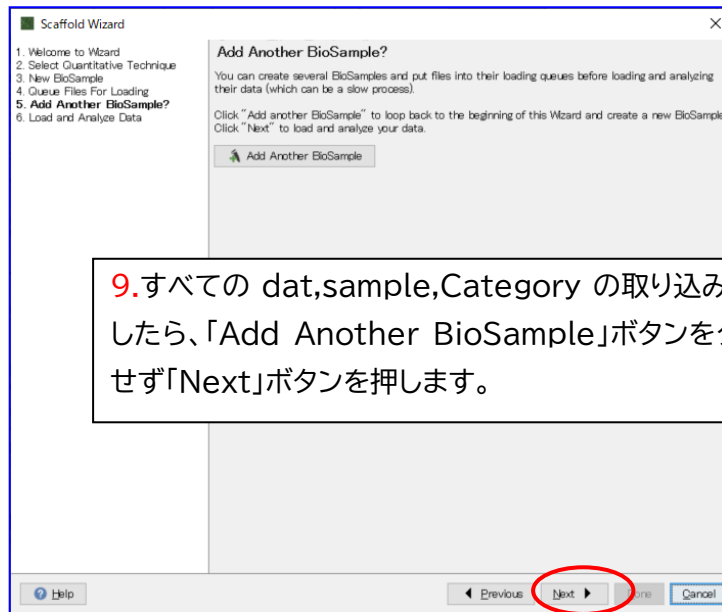


7. 先程の指定とは別のサンプルとして読み込む結果ファイルがないか問われます。ある場合は「Add Another BioSample」ボタンをクリックします。

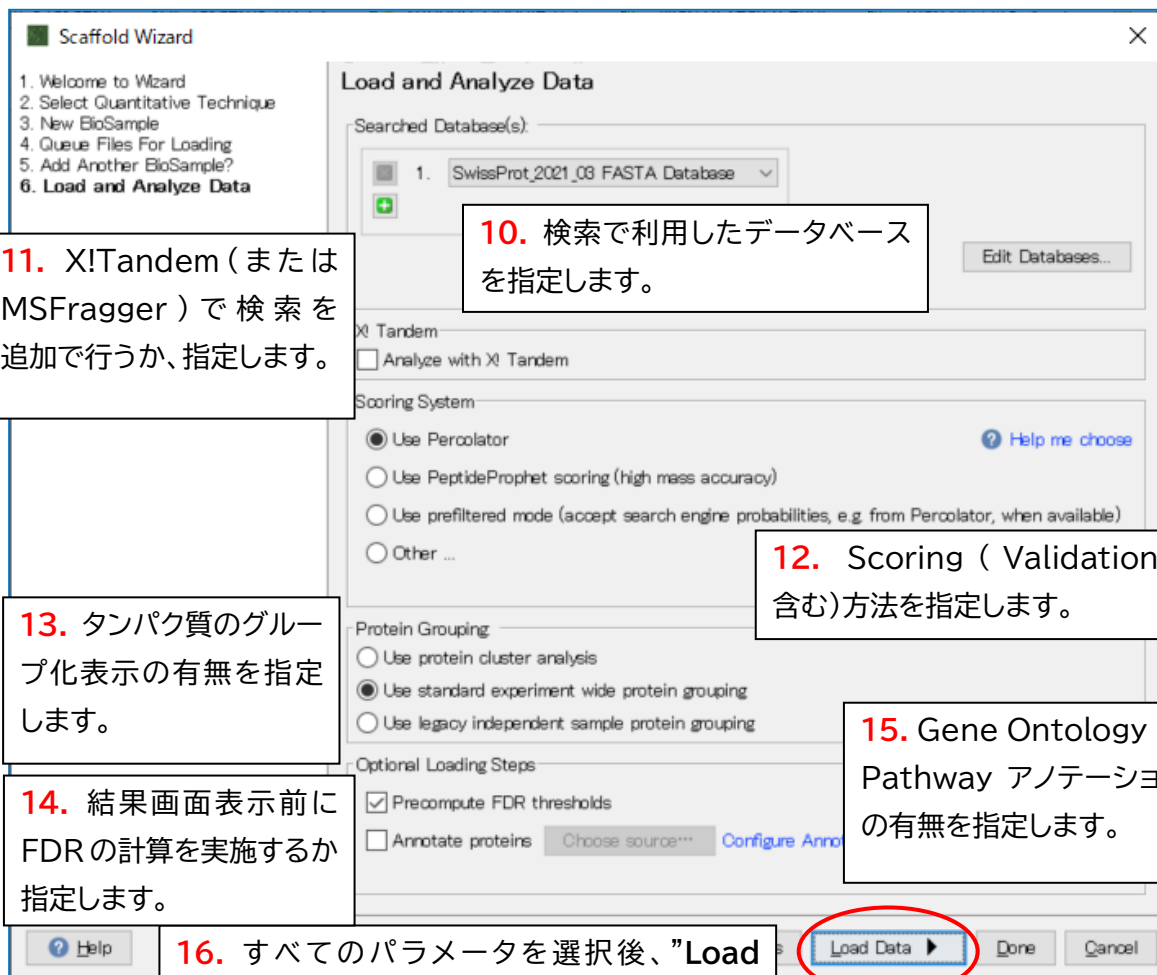


8. 2~6 の操作を繰り返します。今度は Sample 名指定の際名称を変える事と、先ほどと異なる raw ファイル (ピークリストファイル) を選択する事に注意してください。





9. すべての dat,sample,Category の取り込みが終了したら、「Add Another BioSample」ボタンをクリックせず「Next」ボタンを押します。



11. X!Tandem (または MSFragger ) で 検索を 追加で行うか、指定します。

10. 検索で利用したデータベースを指定します。

13. タンパク質のグループ化表示の有無を指定します。

12. Scoring ( Validation 含む)方法を指定します。

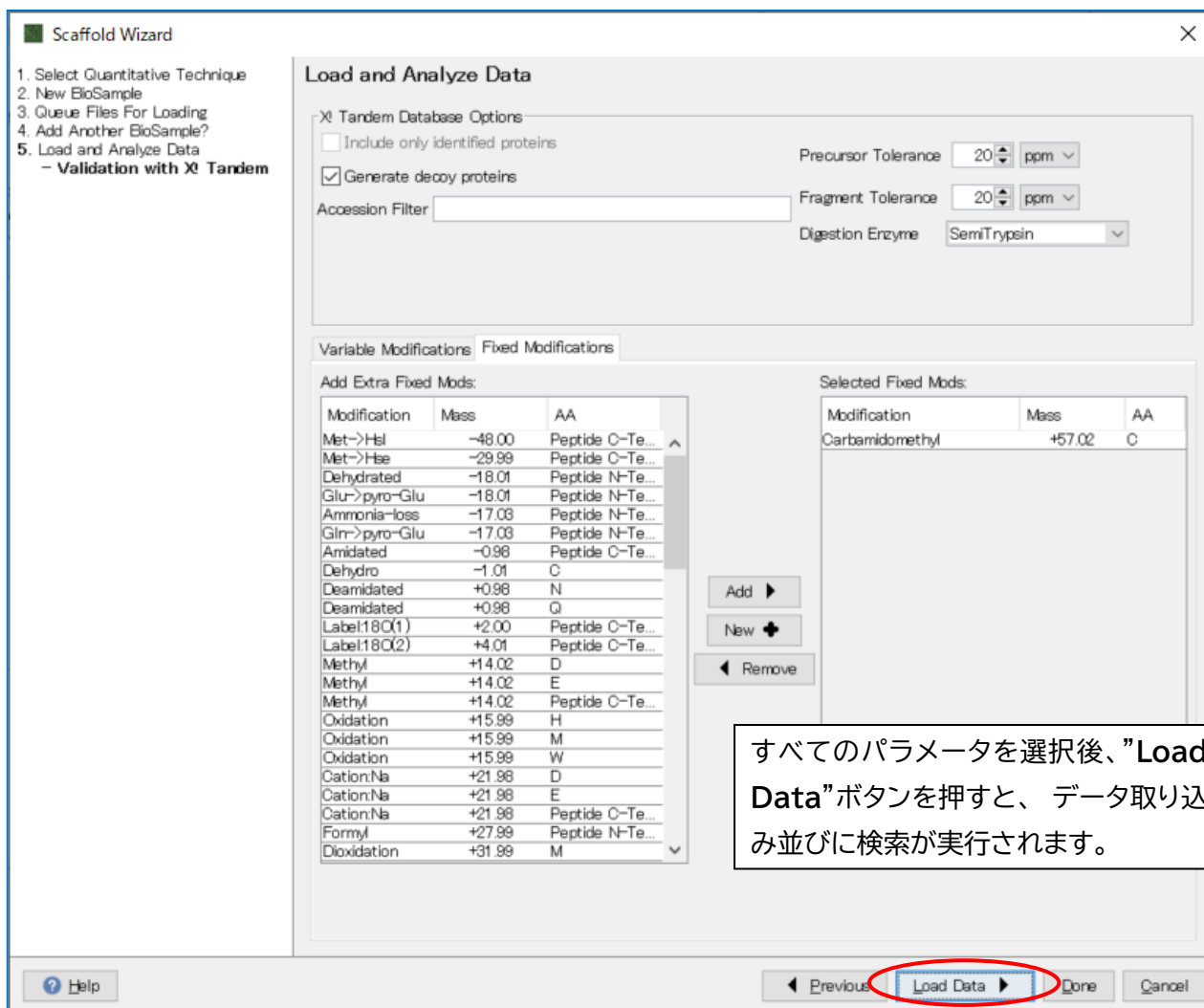
14. 結果画面表示前に FDR の計算を実施するか指定します。

15. Gene Ontology や Pathway アノテーションの有無を指定します。

16. すべてのパラメータを選択後、「Load Data」ボタンを押すと、検索パラメータ入力画面に遷移します。

## [X! Tandem の場合]

以下のように検索を実行する際に与えるパラメータ入力画面が現れます。

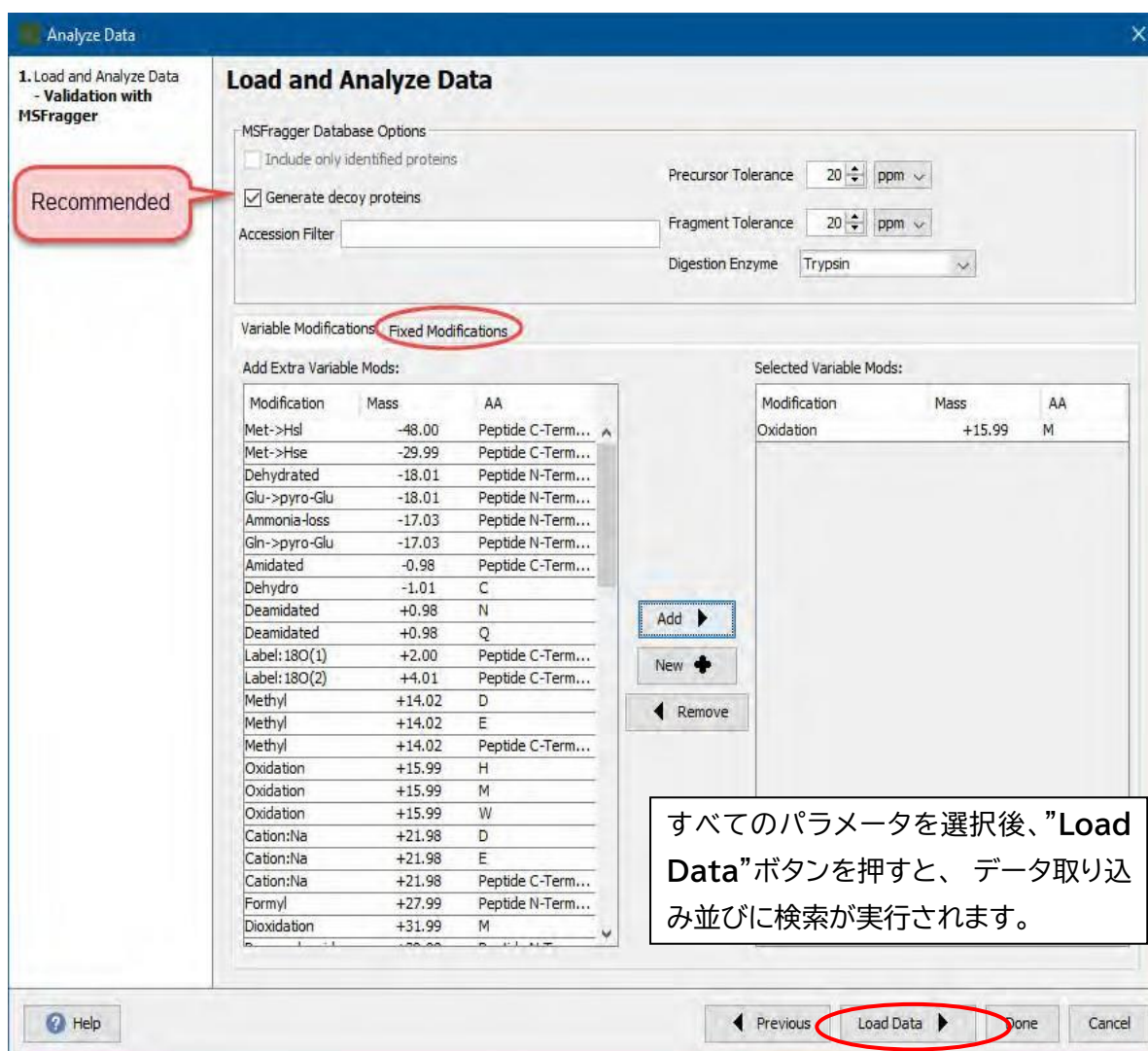


### X! Tandem Database Options

- Include only identified proteins** : 拡張検索実施のオプションだが、この画面では使用しません。
- Generate decoy proteins** : FDR 計算を行うための Decoy データベース作成を行います。
- Accession Filter** : 文字列の条件検索で、taxonomy や特定遺伝子のみを対象とする時に利用します (“\_HUMAN”などと利用します)。
- Precursor Tolerance** : Precursor の測定値と理論値の誤差
- Fragment Tolerance** : Fragment の測定値と理論値の誤差
- Digestion Enzyme** : タンパク質配列から理論ペプチドを生成する際の切断パターン。
- Modification** : 修飾。Fixed(全アミノ酸置換)と Variable (置換する/しない を両方考慮)

## [MSFragger の場合]

以下のように検索を実行する際に与えるパラメータ入力画面が現れます。



### MSFragger Database Options

- Include only identified proteins** : 拡張検索実施のオプションだが、この画面では使用しません。
- Generate decoy proteins** : FDR 計算を行うための Decoy データベース作成を行います。
- Accession Filter** : 文字列の条件検索で、taxonomy や特定遺伝子のみを対象とする時に利用します (“\_HUMAN”などと利用します)。
- Precursor Tolerance** : Precursor の測定値と理論値の誤差
- Fragment Tolerance** : Fragment の測定値と理論値の誤差
- Digestion Enzyme** : タンパク質配列から理論ペプチドを生成する際の切断パターン。
- Modification** : 修飾。Fixed(全アミノ酸置換)と Variable (置換する/しない を両方考慮)

## 2-5. 配列データベースの登録

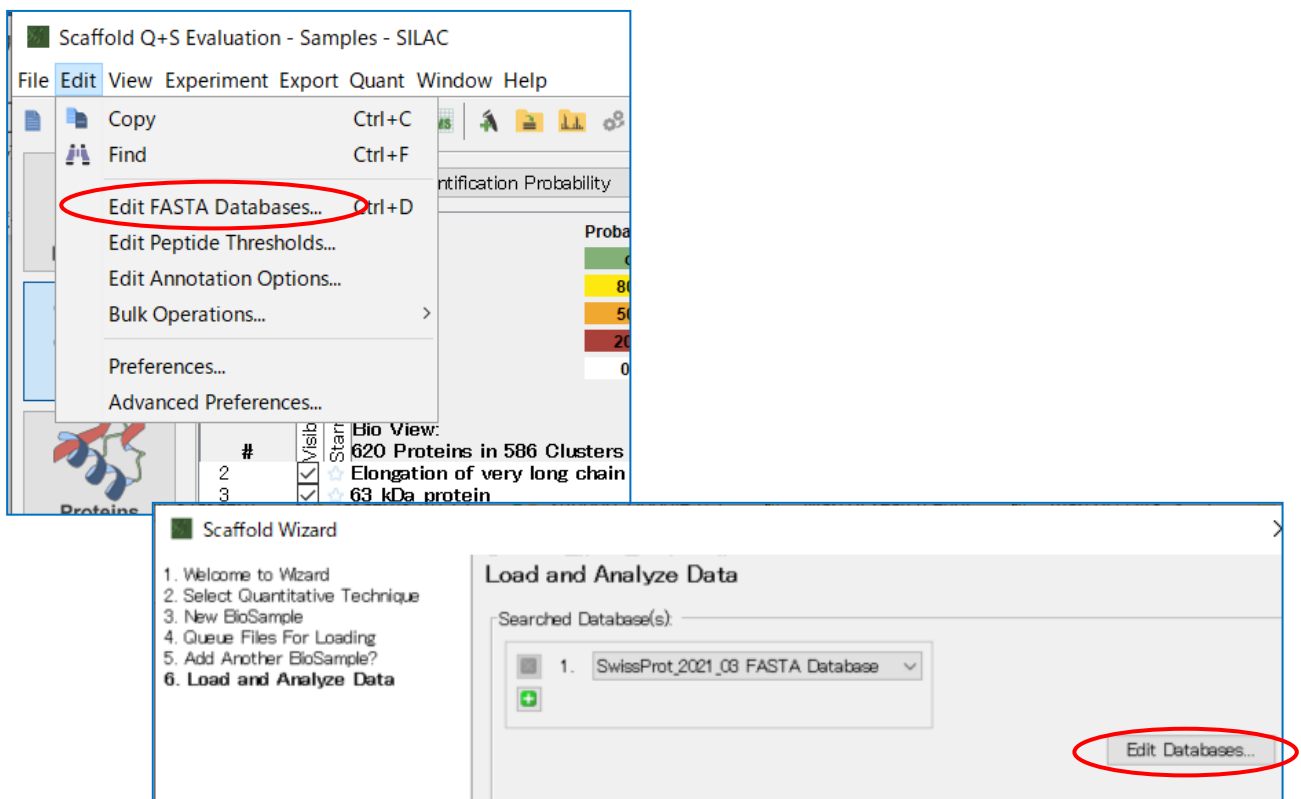
### 2-5-1. 概要

Scaffold に取り込む結果ファイルには、タンパク質の配列が含まれていません。同定タンパク質の ID(Accession)情報を基に配列情報を取得し紐づけるため、Scaffold 上で検索に利用したものと同じ配列データベースを取り込む必要があります。また FASTA データベースのセットは Scaffold で検索結果を読み込んだり検索を実行する前に行ったりする必要があります。Scaffold で取り込んだ FASTA ファイルは“Edit FASTA Databases”ダイアログで管理する事ができます。

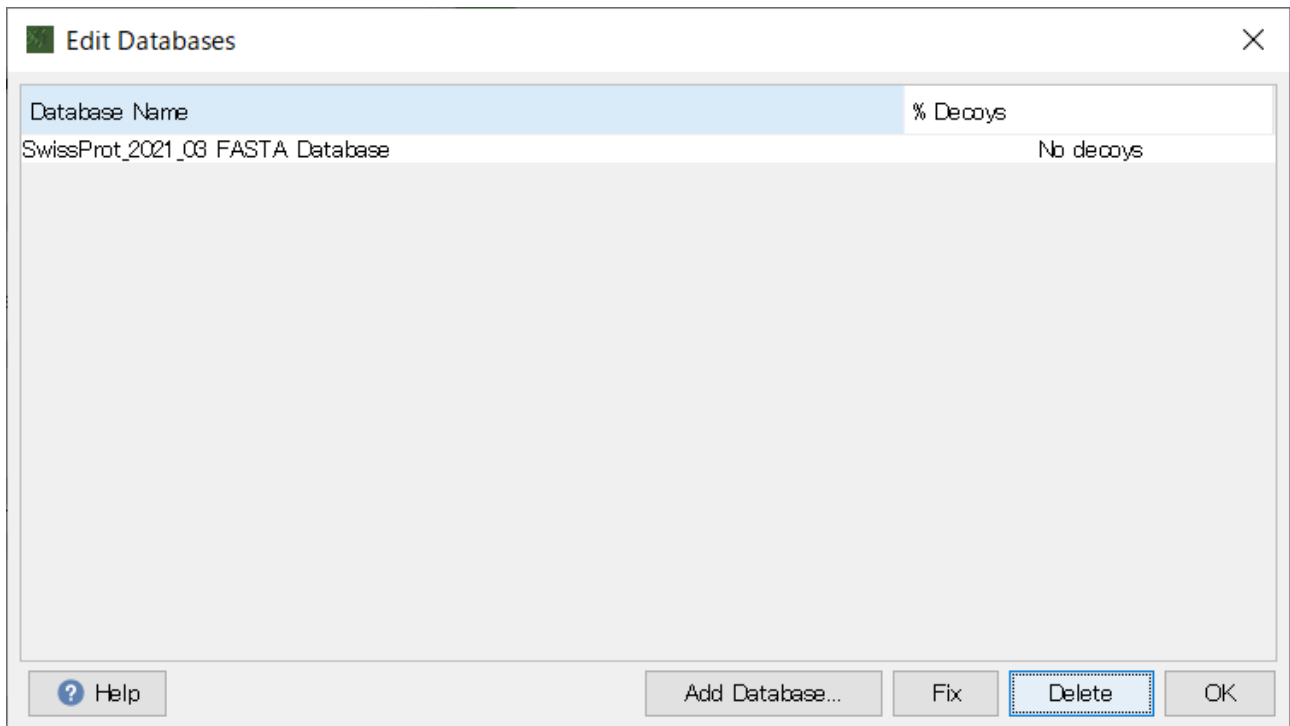
FASTA を取り込む際、FASTA 先頭行の中でどの部位が Accession(ユニークな ID)となり、またどの部位が Description(タンパク質の内容についての説明)にあたるのかを指定する必要がありますが、それを自動的に行う設定も Scaffold には準備されています。ここでは配列データベースを Scaffold に取り込ませる方法やその関連事項について説明しています。

### 2-5-2. Edit FASTA Databases ダイアログ

- Menu の Edit -> **Edit FASTA Databases**
- データ取り込みなど各所必要な時に現れる「Edit Databases」ボタン  
これら操作によりダイアログが現れます。

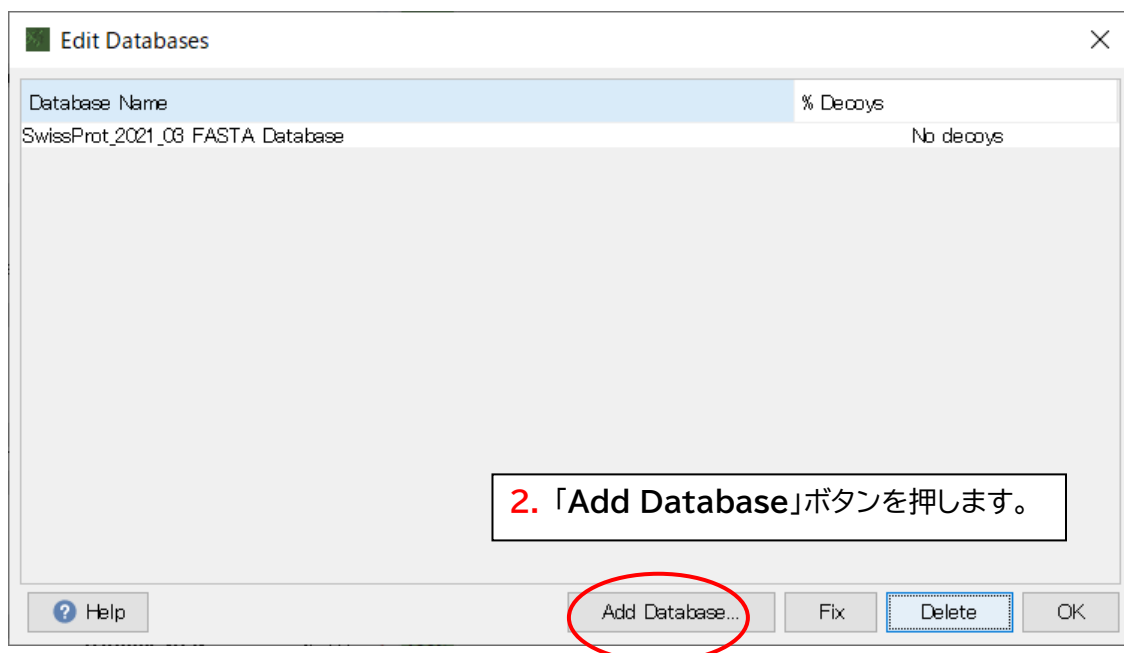


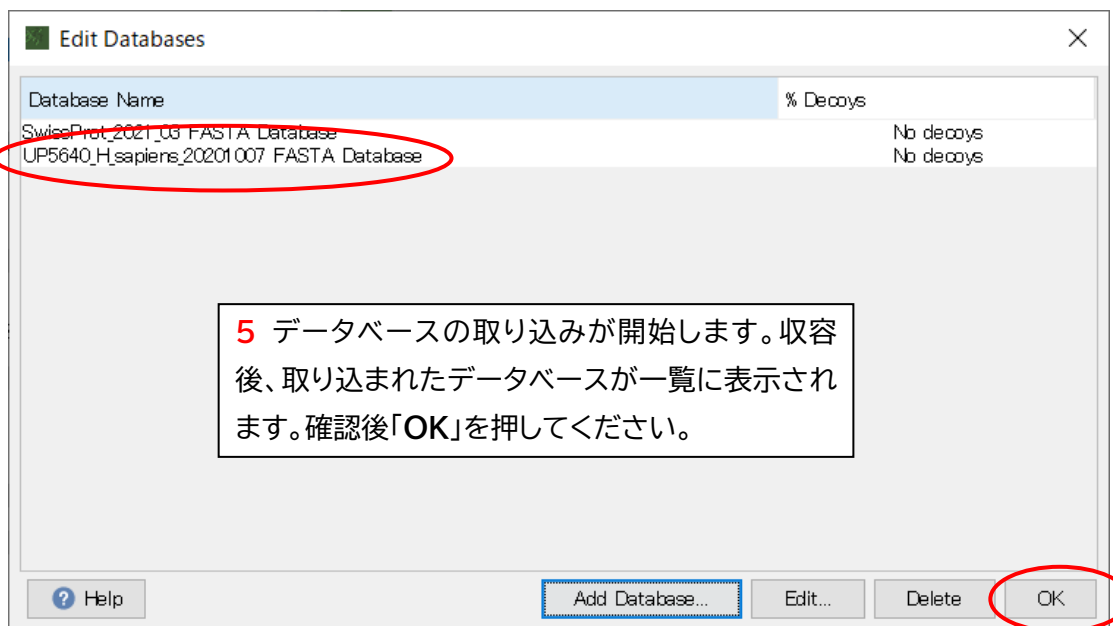
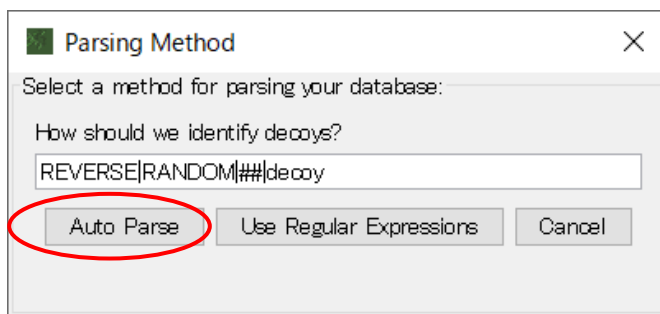
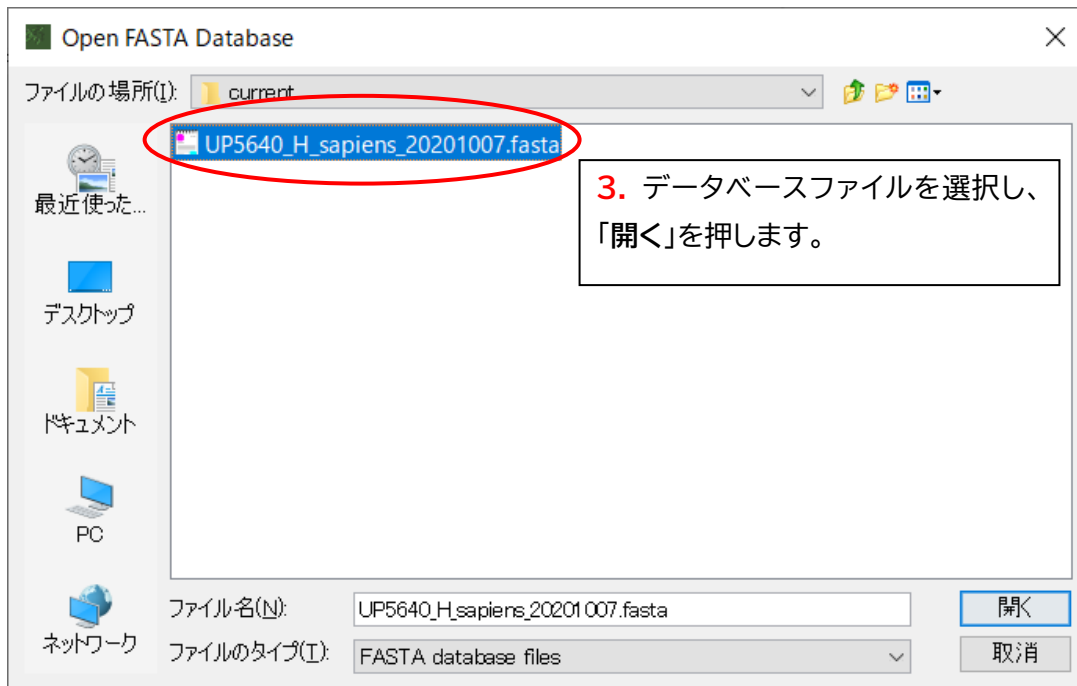
「Edit Databases」ダイアログでは登録されているデータベースの一覧、データベースの追加・変更・削除を行う事ができます。



### 2-5-3. FASTA データベースの登録方法

1. 2-5-2 で案内している方法で、「Edit Databases」ダイアログを開きます。







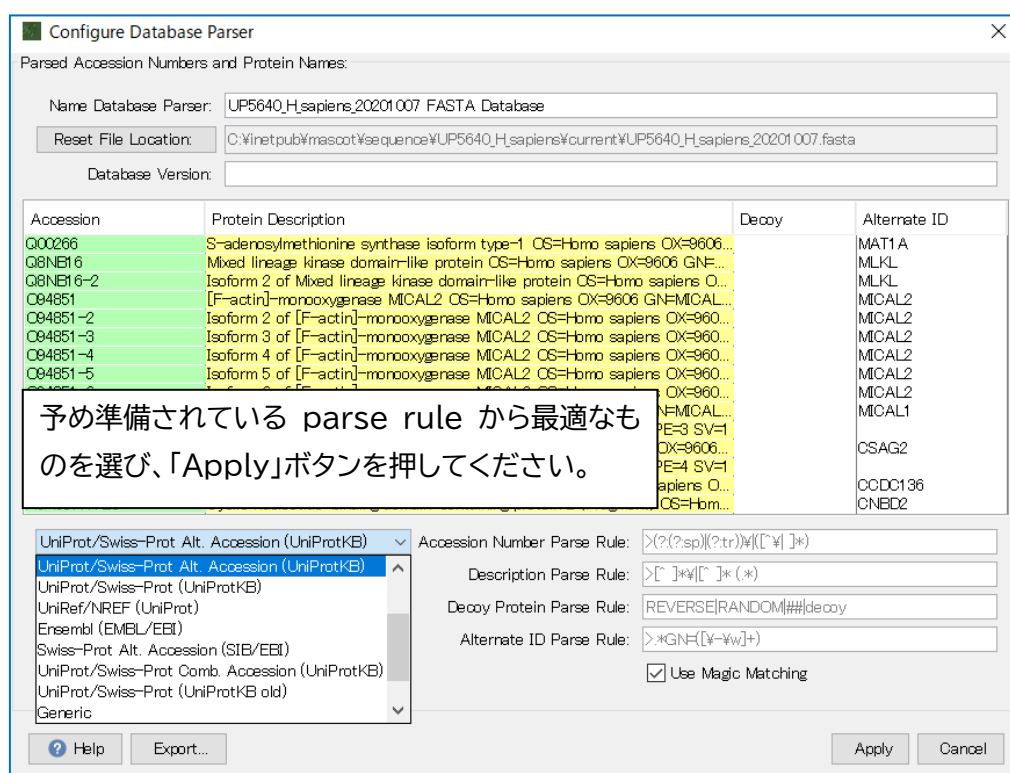
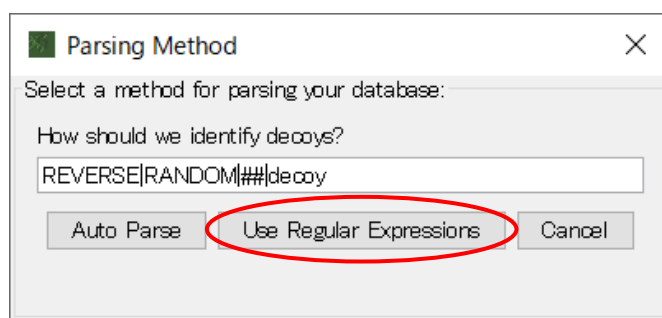
## 2-5-4. parse rule について (Configure Database Parser)

2-5-3 の parse rule に関する補足説明をいたします。

FASTA の先頭行の情報から、Accession (データベースのユニークな ID) と Description (説明部分の記述) の箇所を認識し、各エントリデータの特性とします。この時、Accession や Description の部位を認識するためのルール、文法のようなものが「parse rule」です。「正規表現」と呼ばれるようなルールを使って定義します。Accession については、結果ファイルに含まれている内容と Scaffold 上で指定した内容が一致している必要があります。

Parse rule は 2-5-3 でも説明したように、通常は「Auto Parse」を選択すれば問題ありません。しかし検索エンジン側で使用していた parse rule と、Scaffold auto parse で指定した parse rule が一致していない場合は調整する必要があります。一致しないと Scaffold で取り込んだのち各タンパク質の配列の紐づけがうまくいかず、配列や分子量などの表示が正しく行われません。

ご自身で最適な parse rule を指定したい場合、FASTA ファイル設定時の途中、「Parsing Method」選択時に、「Use Regular Expressions」ボタンを押します(下図)。



## 3. Samples View

### 3-1. 概要

Scaffold では各種機能を持つ View(画面)があり、左側にそれらの View を切り替えるためのスイッチがあります。

データを取り込み後、最初に表示されるのが「**Samples**」View です。

各サンプル別の同定タンパク質やアサインされたペプチドの数、同定されたタンパク質の Gene Ontology 情報を一覧で表示する画面です。Scaffold のデータ検証は主にこの Samples View で行います。

The screenshot displays the Scaffold software interface for the 'Samples - Label-Free' view. The left sidebar contains navigation buttons for 'Load Data', 'Samples' (highlighted with a red box), 'Proteins', 'Similarity', 'Quantify', and 'Publish'. The main window shows a table of protein clusters with columns for '#', 'Visible?', 'Starred?', 'Protein Name', 'Accession Number', 'Molecular Weight', 'Protein Grouping Ambiguity', 'Taxonomy', and 'Gene Ontology'. A 'Probability Legend' is visible above the table, indicating color-coded probability ranges for protein identification. The table lists 18 rows of data, including Hemoglobin subunit alpha from various species. The bottom of the screen shows 'Protein Information' and 'Gene Ontology' sections.

3章では **Samples View** について、その見方・使い方を説明いたします。  
英文マニュアルの「**Chapter.6 Sample View**」に該当する内容です。

### 3-2 . Samples View 画面・表示内容

Samples view ではサンプル毎に、同定タンパク質／アサインされたペプチドの数／Gene Ontology などの情報を表示しています。以降、表示内容について説明いたします。

## [全体図]

下図内の番号(①など)が、以降の説明の小項目(「3-2-1」など)と連動しています。

The screenshot shows the Scaffold software interface with several callouts:

- ① The Display pane :** samples table で表示する数値の設定 (Setting the numerical values to be displayed in the samples table).
- ② Filtering Samples** 表示タンパク質/ペプチドの絞り込み条件 (Filtering conditions for displayed proteins/peptides).
- ③ The Samples Table** 同定タンパク質に関する情報を表示 (Displaying information about identified proteins).
- ④ Information Panes** 類似タンパク質情報、GO 情報など (Similar protein information, GO information, etc.).

The interface includes a menu bar (File, Edit, View, etc.), a toolbar, and a main display area with a table of protein data. The table has columns for protein name, accession number, molecular weight, and various GO terms. A 'Display Options' dropdown is set to 'Total Unique Peptide Count'. A 'Probability Legend' shows color-coded bars for confidence levels: over 95% (green), 80% to 94% (yellow), 50% to 79% (orange), 20% to 49% (red), and 0% to 19% (grey). The table shows several rows of hemoglobin subunit alpha data from various species like Spalax leucodon, Mus musculus, and Ondata zibethicus. Below the table are 'Protein Information', 'Gene Ontology', and 'Sample Information' panes.

以降、各項目について説明をしていきます。

### 3-2-1. Display pane: Samples table で表示する数値の設定

各サンプルの列で表示される数字について、表示内容を切り替えます。主に、同定の確からしさや定量に関連する数値を評価できる数値を選択することができます。

Display Options の選択項目は次頁表の通りです。

項目	説明
Protein Identification Probability	タンパク質の同定確率。タンパク質にアサインされたペプチドの同定確率(Peptide prophet) を基に Protein Prophet で計算 されます。
Percent Coverage	タンパク質全長(残基数)に対し、アサインされたペプチドの残基数の割合。
Percentage of Total Spectra	サンプル内全スペクトルに対して、タンパク質にアサインされたスペクトルの数の割合。
Exclusive Unique Peptide Count	<p>下記説明文参照。用語の構成として、  最初： <b>Exclusive</b> か <b>Total</b>  その後ろに  <b>Unique peptide / Unique spectrum / Spectrum</b>  <b>Count</b> という形式となっています。</p>
Total Unique Peptide Count	
Exclusive Unique Spectrum Count	
Total Unique Spectrum Count	
Exclusive Spectrum Count	
Total Spectrum Count	
Quantitative Value	Total Spectrum Count を、各サンプルの全同定スペクトル数を元に Normalize した指標値。表示内容を別の定量指標に変更可能 ( <b>Experiment -&gt; Quantitative Analysis</b> の 「 <b>Use Normalization</b> 」->「 <b>Quantitative Method</b> 」)

説明補足:

#### ■ **Exclusive** と **Total**

・**Exclusive** は、該当タンパク質(グループ)のみに存在するペプチド/スペクトル。それに対し **Total** は、**Exclusive** に加え他のグループ【類似タンパク質群】でもアサインされているペプチド/スペクトルを加えた数となります。従って、数は

**Total**  $\geq$  **Exclusive** となります。

#### ■ **Unique peptides / Unique spectra / Spectrum**

Scaffold の用語において「Unique」は、他の同定タンパク質と共通してアサインされているペプチドであるかどうかは一切関係ありません。同じ配列にマッチしている複数のスペクトルについて、1つにまとめて1つと数え上げるか、それともばらばらのままで数え上げるかの違いです。3つの項目は次頁のような違いがあります。

#### ・**Unique Peptides**

タンパク質にアサインされたペプチド数。修飾のあるなしに関わらず、また電荷の相違に関わらず、同一配列のペプチドは1つとして数えます。

## •Unique Spectra

Unique peptides と似ていますが、修飾が異なる場合や電荷が異なる場合は別のものとして数えます。

## •Spectrum

同一ペプチド配列として同定されたスペクトルすべてを1つにまとめずバラバラに数えます。

従って数は

Spectrum  $\cong$  Unique Spectra  $\cong$  Unique Peptides となります。

### 「 Req Mods 」:

特定の修飾基がついているペプチドのみを表示するためのフィルター

### 「 Search 」:

すべてのタンパク質 / ペプチド データを対象としたキーワード検索で検索条件に該当するタンパク質 / ペプチドのみを表示するためのフィルター。入力欄にキーワードを入れて検索する単純検索のほか、虫眼鏡アイコンをクリックすると現れる詳細検索画面があります(下図)。

**Configure Advanced Protein Filter**

Search for proteins based on accession number, protein name, peptide sequences (or sub sequences) and spectrum ID names. All searches support regular expressions. For example, you can find possible CaMKII phosphorylation sites by searching for peptides with the "R.[ST]" motif. Please visit [this site](#) for more help with forming regular expressions.

Warning:  
Searching by spectrum name can be extremely slow. It works best on small data sets.

Matching **All** of the text filters:

Accession/Protein Name:

Protein Sequence Motif:

Identified Peptide Sequence:

Spectrum Name (SLOW):

Category Filtering Options

Present in **All** of the Categories:

Absent from the Categories:

Quantitative Profile Filtering Options  Filter by: **Significant**

Compared Categories:  High Categories:  Low Categories:

Starred as...

With taxonomy:

With **Any** of these GO Terms:

### 3-2-2. Filtering Samples: 表示たんぱく質/ペプチドの絞り込み

Protein Threshold: 99.0%	Min # Peptides: 2	Peptide Threshold: 95%
--------------------------	-------------------	------------------------

#### [Protein Threshold]

タンパク質の同定確率(Protein Identification Probability)、または Protein FDR による表示タンパク質の絞り込みができます。同定確率は アルゴリズムProtein Prophetにより計算される数値です。Protein FDR は%の後ろに「FDR」とついている項目(下図)で、decoyデータベースに対する検索を行った時のみ現れます。(Protein FDRの計算方法は少し特殊です。英文マニュアルをご参照ください)

Protein Threshold: 99.0%	Min # Peptides: 2	Peptide Threshold: 95%
--------------------------	-------------------	------------------------

- 20.0%
- 50.0%
- 80.0%
- 90.0%
- 95.0%
- 99.0%**
- 99.9%
- 1.0% FDR
- 2.0% FDR
- 3.0% FDR
- 5.0% FDR
- 10.0% FDR

#### [Min # Peptides]

タンパク質にアサインされた同定ペプチド数の最低数値を指定できます。1~5 までの数値を設定できます。

#### [Peptide threshold]

ペプチドの同定確率 (Peptide Identification Probability)または FDRのパーセントによる表示ペプチドの絞り込みができます。数値はアルゴリズム LFDR(local False Discovery Rate)またはPeptide Prophet により計算されます。ここで定義した閾値の基準を上回るペプチドのみが、タンパク質の同定や Display options などで表示されるペプチドの数、Coverage の数値計算に使われます。

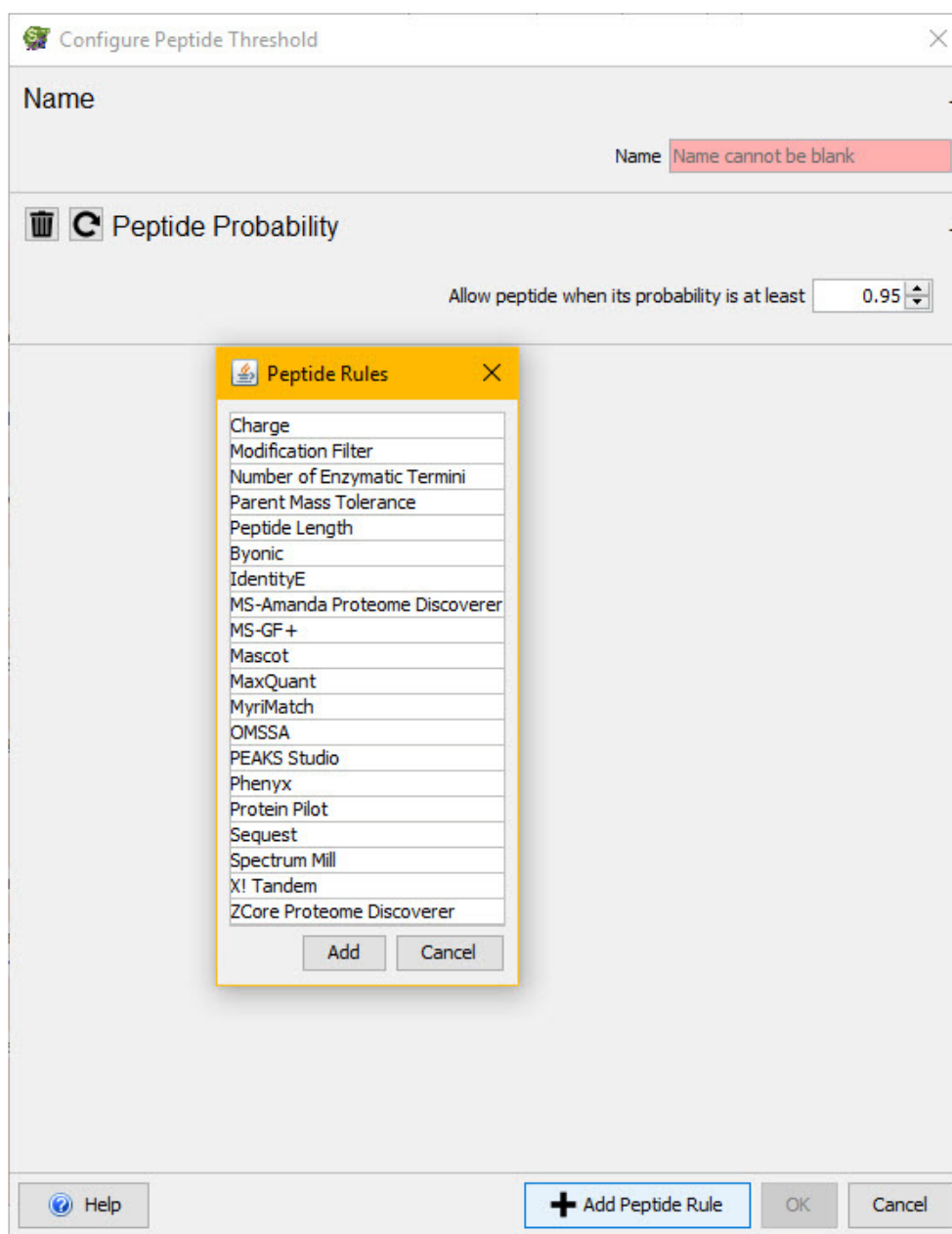
同定基準は、予め提示されている選択肢の他、「**Custom**」を選択する事でユーザーが独自に設定をすることができます。Custom を選択した場合、Protein Threshold の基準は使用できなくなります。

Protein Threshold: 99.0%	Min # Peptides: 2	Peptide Threshold: 95%
--------------------------	-------------------	------------------------

- 0%
- 20%
- 50%
- 80%
- 90%
- 95%**
- Custom...
- 0.1% FDR
- 0.5% FDR
- 1.0% FDR
- 2.0% FDR
- 5.0% FDR

“Custom”では様々な条件を指定する事ができます。

- Peptide Probability (最初から入っていますがゴミ箱アイコンをクリックして消す事ができます)
- 電荷
- Precursor の誤差
- ペプチドの長さ
- 修飾の種類
- Missed Cleavage数
- 各検索エンジンから得られるパラメータ [スコアなど]

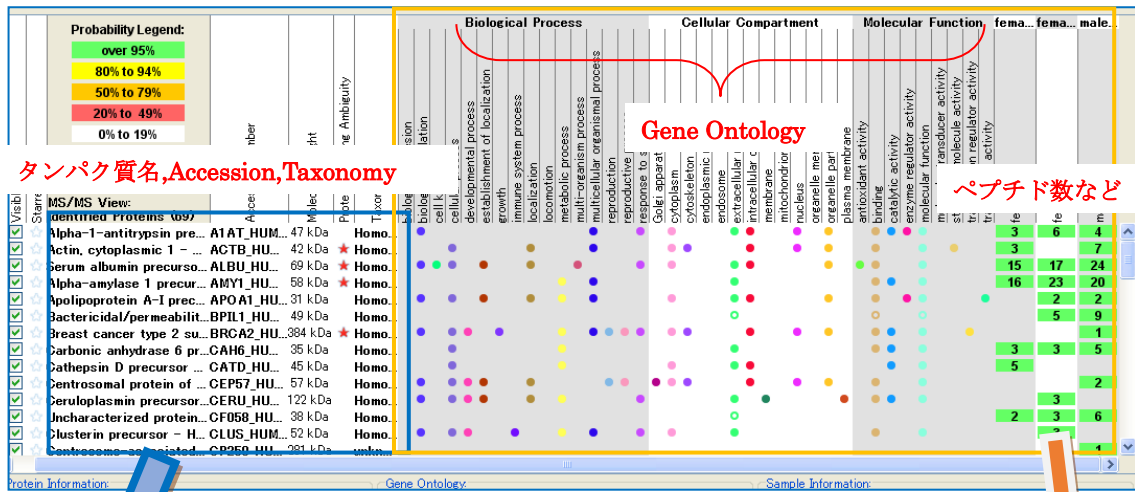


複数の条件を組み合わせて作成する事も可能です。



### 3-2-3. The Samples Table: 同定タンパク質に関する情報の表示

Samples Tableでは各サンプルで同定されたタンパク質の情報が表示されます。



タンパク質名, Accession, Taxonomy

ペプチド数など

Bio View 又は MS/MS View

タンパク質総数  
タンパク質名

Accession

分子量

生物種

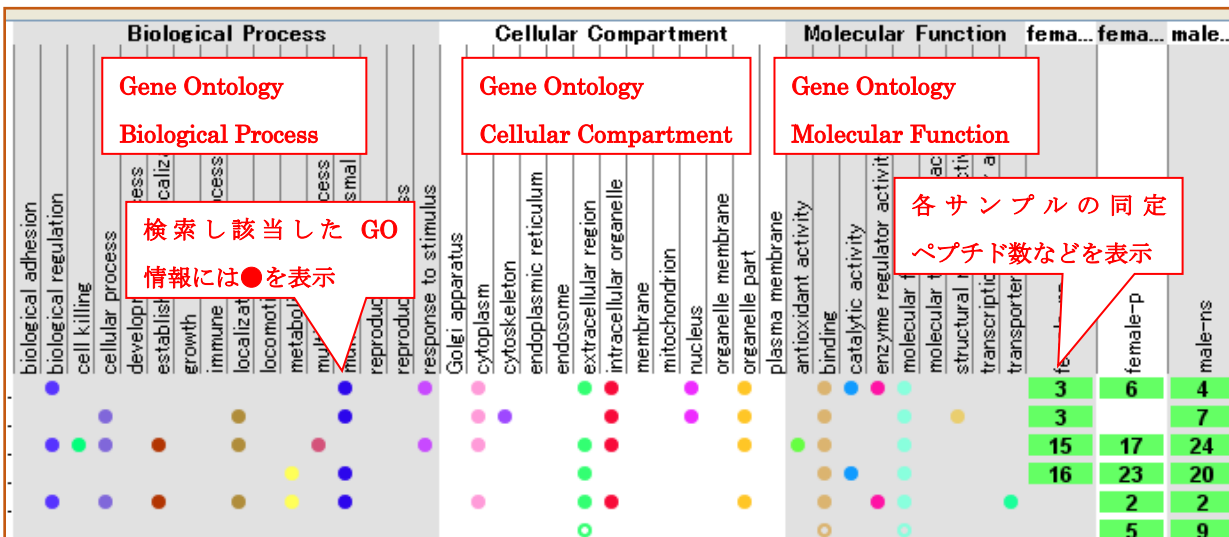
Visible?	Starred?	MS/MS View: Identified Proteins (69)	Accession Numb	Molecular Weight	Protein Grouping Ambiguity	Taxonomy
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Alpha-1-antitrypsin pre... A1AT_HUM...	47 kDa	Homo..		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Actin, cytoplasmic 1 - ... ACTB_HU...	42 kDa	★ Homo..		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Serum albumin precursor... ALBU HU...	69 kDa	★ Homo..		

Gene Ontology  
Biological Process  
検索し該当した GO 情報には●を表示

Gene Ontology  
Cellular Compartment

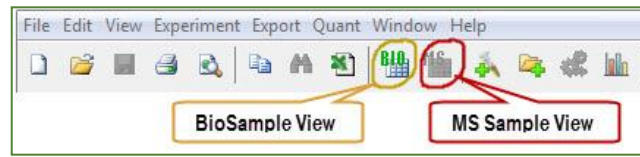
Gene Ontology  
Molecular Function

各サンプルの同定 ペプチド数などを表示

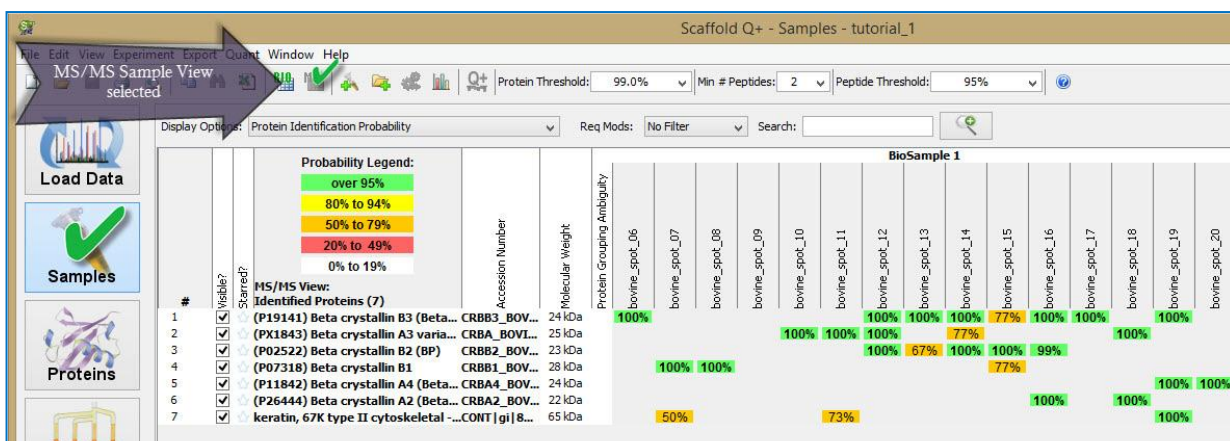
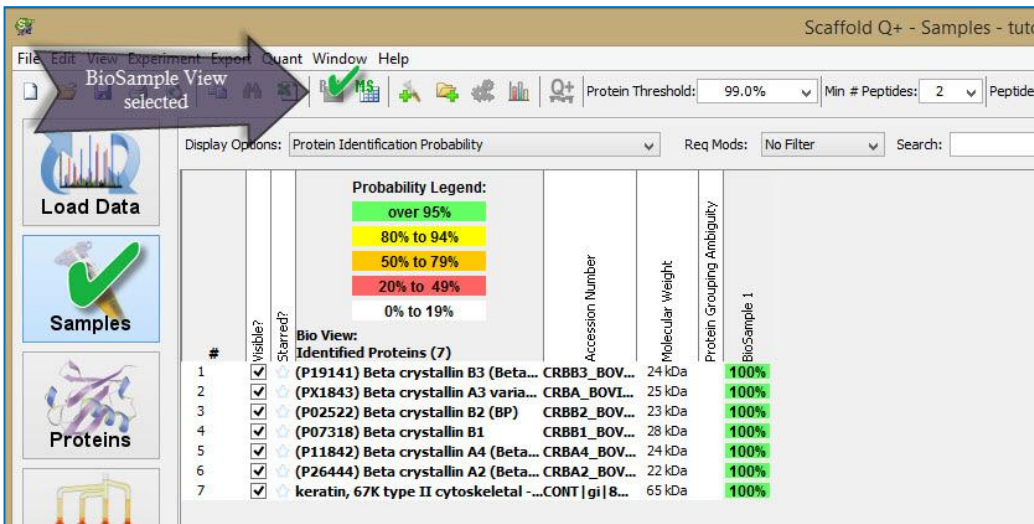




Samples View での表示は、アイコンの BioSample View / MS Sample View ボタンの切り替えによって内容を変更することができます(下図)。



BioSample View では、予め定義された BioSample 単位にデータがまとめられます。一方 MS Sample View を選ぶと、まとめられていた MS Sample が展開して表示されるようになります(下図)



また、表内の各列をクリックすると、その項目での降順／昇順／並び替え解除 にて各列を並び替えることができます。

**Probability Legend:**

- over 95% (Green)
- 80% to 94% (Yellow)
- 50% to 79% (Orange)
- 20% to 49% (Red)
- 0% to 19% (Light Red)

**MS/MS View: Identified Proteins (69)**

Protein	Accession Number	Molecular Weight	Protein Grouping Ambiguity	Taxonomy	Biological Process	Cellular Compartment	Molecular Function	fema... female-tp	fema... female-P	male... male-ms
Alpha-1-antitrypsin pre...	A1AT_HUM...	47 kDa		Homo...	biological adhesion			3	6	4
Actin, cytoplasmic 1	ACTB_HU...	42 kDa	*	Homo...	cellular process			3	7	7
Serum albumin precurs...	ALBU_HU...	69 kDa	*	Homo...	establishment of localization			15	17	24
Alpha-amylase 1 precu...	AMY1_HU...	58 kDa	*	Homo...	growth			16	23	20
Apolipoprotein A-1 pr...	APOA1_HU...	31 kDa		Homo...	immune system process				2	2
Bactericidal/permeabil...	BP1L1_HU...	49 kDa		Homo...	localization				5	9
Breast cancer type 2 su...	BRCA2_HU...	384 kDa	*	Homo...	locomotion					1
Carbonic anhydrase 6 pr...	CAH6_HU...	35 kDa		Homo...	metabolic process			3	3	5
Cathepsin D precursor	CATD_HU...	45 kDa		Homo...	multicellular organismal process			5		

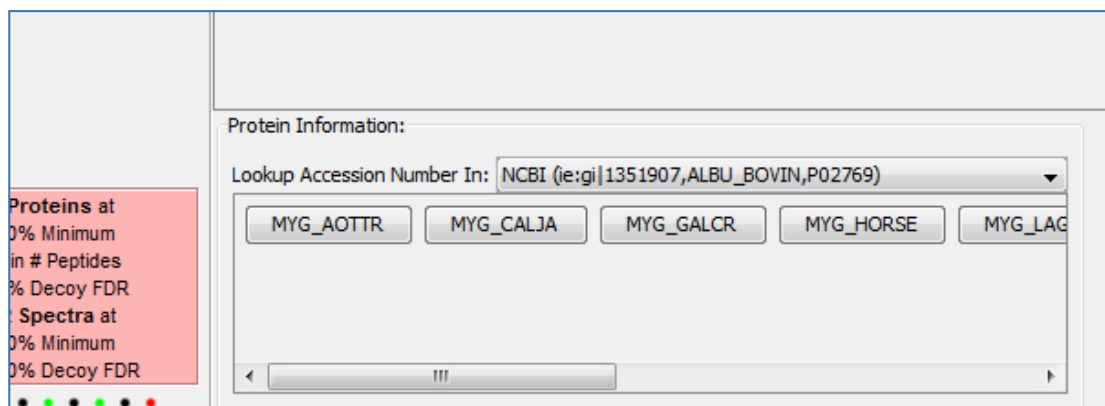
### 3-2-4. Information Panes:タンパク質・サンプルに関する追加情報

画面下部には Samples table とは異なる情報を表示する Information Panes があります(下図)。左から順に、「**Protein Information pane**」「**Annotation pane**」「**Sample Information pane**」から構成されています。



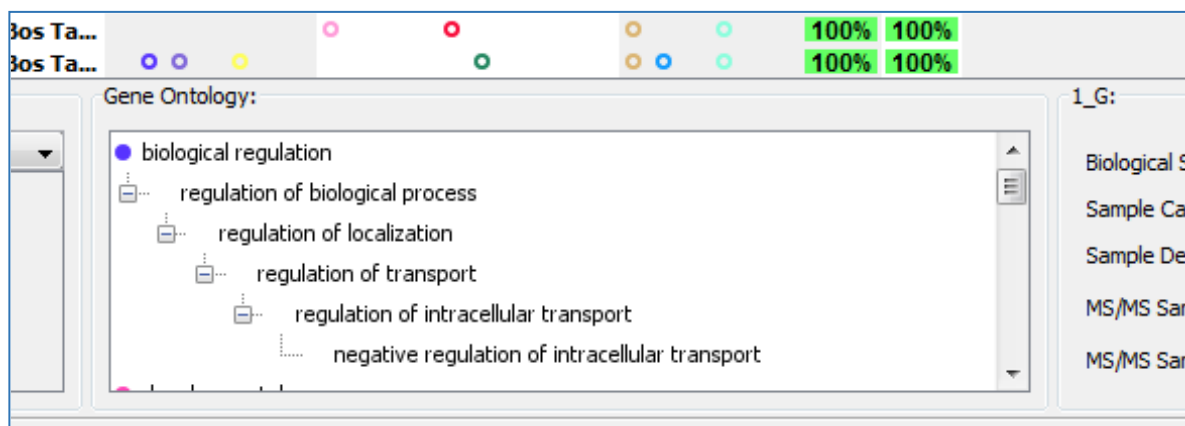
#### Protein Information pane

左下の「Protein Information pane」では、選択しているタンパク質について、ファミリー(Scaffold では、MASCOT でいう subset と sameset に該当する)タンパク質の ID が並んで表示されます (下図)。また「Lookup Accession Number In」で接続先のサイトを選択してからボタンを押すと、該当サイトにおけるエントリー情報をブラウザで表示させることができます。



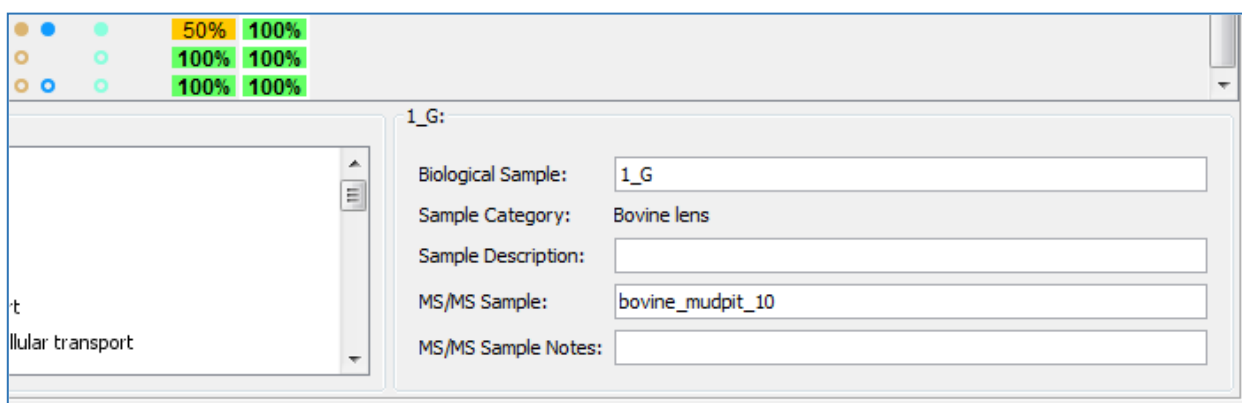
## Annotation pane

真ん中下の「Gene Ontology pane」では、選択しているタンパク質の Gene Ontology 階層構造の情報が表示されます(下図)。気になる GO に関して選択しダブルクリックすると、Gene Ontology の用語説明のページが WEB ブラウザで開きます。



## Sample Information pane

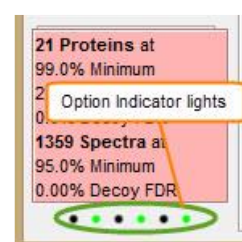
画面右下は、Sample に関する情報を表示する「Sample Information pane」です。sample table 内で、各 sample(Biosample または MS Sample)の列の何かの項目を選択すると、その sample に関する情報が表示されます(下図)。



## 3-3 . FDR ダッシュボード・オプションインジケータランプ

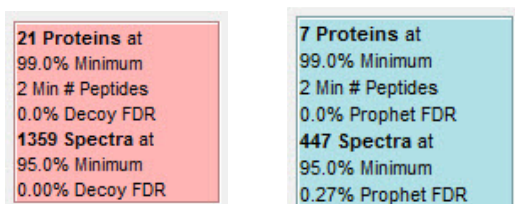
画面一番左側の下に、FDR の計算内容に関する表示と、表示に関するオプションで何が選択されているかを一目で判別するインジケータがついています(右図)。

FDR ダッシュボードには、同定基準を超えるタンパク質数、スペクトル数並びに



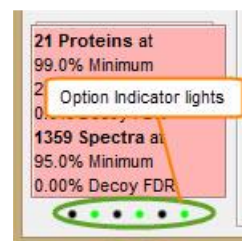
それぞれのフィルターリング条件が表示されます。

解析対象の検索が Decoy データベースに対しての検索も行い FDR の計算ができる時、ダッシュボードの色が赤(下図左)になります。decoy の計算をしていない場合、青色(下図右)となります。



FDR ダッシュボードの下にある6つの○は、表示のオプション選択状況のオン/オフを緑丸/黒丸で表しています(下図)。各丸の内容は左から順に以下の内容です。丸にカーソルを合わせると説明が表示されます。

- Show less <5% probability
- Show lower Scoring Matches
- Show entire protein Clusters
- Use Protein Cluster Analysis
- Use Independent Sample Grouping strategy
- Scoring Scheme (LFDR:緑, Peptide Prophet Advanced : オレンジ、Peptide Prophet:黒)



インジケーターの上にカーソルを合わせる事で、選択内容を文字で確認する事ができます。なお上記リストの上の3項目は、Menu の View にて設定内容を切り替える事が可能です。

### 3-4 . Probability の凡例

MS Sample 並びに BioSample の各セルについている色は、Sample 画面の表左上にあるタンパク質の同定確率(Probability)に基づいています(下図)。タンパク質の同定確率は、Protein Prophet によって計算された数値です。

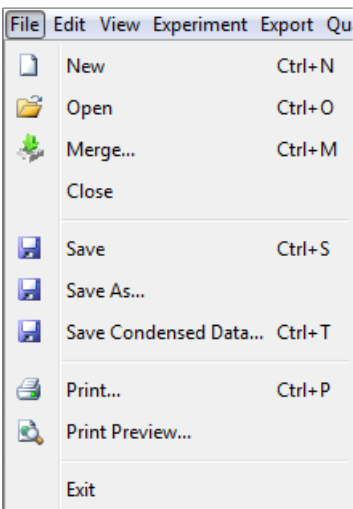
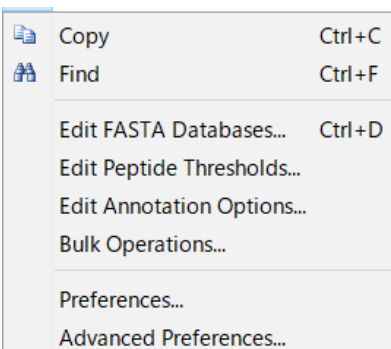
#	Visible?	Starred?	MS/MS View: Identified Proteins (7)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	BioSample 1														
							bovine_spot_06	bovine_spot_07	bovine_spot_08	bovine_spot_09	bovine_spot_10	bovine_spot_11	bovine_spot_12	bovine_spot_13	bovine_spot_14	bovine_spot_15					
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(P19141) Beta crystallin B3 (Beta...	CRBB3_BOV...	24 kDa		100%														
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(PX1843) Beta crystallin A3 varia...	CRBA_BOVL...	25 kDa																
3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(P02522) Beta crystallin B2 (BP)	CRBB2_BOV...	23 kDa																
4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(P07318) Beta crystallin B1	CRBB1_BOV...	28 kDa		100%	100%													
5	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(P11842) Beta crystallin A4 (Beta...	CRBA4_BOV...	24 kDa																
6	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(P26444) Beta crystallin A2 (Beta...	CRBA2_BOV...	22 kDa																
7	<input checked="" type="checkbox"/>	<input type="checkbox"/>	keratin, 67K type II cytoskeletal ...	CONT   gi   8...	65 kDa		50%														

## 4. Menu の各項目について

この章では、Scaffold のメニュー選択内容と各項目で設定できる内容について説明しています。最初に一通り説明し、それ以降の項目で各選択項目の中でより詳しい解説が必要な情報についてピックアップしています。









### 4-1. menu の内容 説明

File Edit View Experiment Export Quant Window Help

メニュー	コマンド
<p><b>File</b></p> 	<ul style="list-style-type: none"> <li>・<b>New</b> - ファイル作成ウィザードを起動します。詳細は 2 章「MASCOT 結果取り込み」をご覧ください。</li> <li>・<b>Open</b> - Scaffold のファイル(.sf3)を開きます。</li> <li>・<b>Merge</b> - Scaffold 上で .sf3 ファイルを統合します。詳細は「4-3.sf3 ファイルの統合(merge)」をご覧ください。</li> <li>・<b>Close</b> - 今開いている sf3 ファイルを閉じます</li> <li>・<b>Save</b> - 今開いている sf3 ファイルを保存します</li> <li>・<b>Save As</b> - 今開いている sf3 ファイルを別名で保存します。</li> <li>・<b>Save Condensed Data</b> - 今開いている sf3 ファイルを、保存 対象を変更しながら別名保存します。詳細は「4-4.データサイズ 間引き方法」をご覧ください。</li> <li>・<b>Print</b> - 現在開いている view 画面を印刷します。</li> <li>・<b>Print Preview</b> - 印刷の preview を表示します。</li> <li>・<b>Exit</b> - Scaffold を終了します。</li> </ul>
<p><b>Edit</b></p> 	<ul style="list-style-type: none"> <li>・<b>Copy</b> - 選択時に開いている view のデータをそのままクリップボードにコピーします。タブ区切りのデータとなります。</li> <li>・<b>Find</b> - 検索用のダイアログを開き view から該当項目を探します。</li> <li>・<b>Edit FASTA Database</b> - FASTA データベースを追加します。詳細は「2-4. 配列データベースの登録」をご覧ください。</li> <li>・<b>Edit Peptide Threshold</b> - ペプチドの閾値を設定します。詳細は「3-2-2. Filtering Samples」内の Peptide Threshold 項目をご覧ください。</li> <li>・<b>Edit Annotation Options</b>-GO に関する設定画面を開きます。詳細は「4-5.GO の設定」をご覧ください。</li> <li>・<b>Bulk Operation</b> - [Samples view のみ] 選択項目をまとめて設定変更します。</li> <li>・<b>Preferences</b> - Scaffold の設定変更。詳細は「4-6.Preferences 設定」をご覧ください。</li> <li>・<b>Advanced Preferences</b> - Scaffold の設定変更。詳細は「4-7.Advanced Preferences」をご覧ください。</li> </ul>



メニュー	コマンド
<p><b>View</b></p> 	<ul style="list-style-type: none"> <li>・<b>Navigation Pane</b> – View 切り替えパネル(画面左側) 表示/非表示</li> <li>・<b>Switch Sample View</b> – BioSample / MS Sample 表示切替</li> <li>・<b>Switch Display Options</b> – Display Options を切替。詳細は「3-2-1.Display options」をご覧ください。</li> <li>・<b>Show Entire Protein Clusters</b> – タンパク質のクラスターに含まれる、Protein probability が設定基準値以下のメンバー タンパク質について表示/非表示 するかの切替</li> <li>・<b>Show Lower Scoring Matches</b> -Protein probability が設定基準値以下の場合に display option で定められた数値を各サンプルのセルに表示/非表示 するかの切替</li> <li>・<b>Show &lt;5% Probabilities</b> – Peptide Probability のとても 低いペプチドについて 表示/非表示 するかの切替</li> <li>・<b>Show Sample Notes</b> – Samples 画面下段の「Information pane」について 表示/非表示 するかの切替</li> <li>・<b>Show Hidden Annotations</b> – Hidden と定義したタンパク質の表示/非表示 するかの切替</li> <li>・<b>Navigate</b> – 画面内にタブがある view 内の pane について、選択タブを切替</li> </ul>
<p><b>Experiment</b></p> 	<ul style="list-style-type: none"> <li>・<b>Edit Experiment</b> – 選択中の MS Sample(experiment)に ついて、名称などを設定するダイアログが現れます</li> <li>・<b>Edit BioSample</b> – 選択中の BioSample の編集。「Load Data」view→sample タブ選択右クリック→「Edit BioSample」と同じ 操作(ダイアログ出現)</li> <li>・<b>Add BioSample</b> – BioSample を追加。「Load Data」view→ 「Add BioSample」と同じ操作(ダイアログ出現)</li> <li>・<b>Delete BioSample</b> – BioSample を削除。「Load Data」view→ BioSample のタブ選択右クリック→「Delete BioSample」と同じ操作(ダイアログ出現)</li> <li>・<b>Queue Files For Loading</b> – Experiment を追加。「Load Data」view→「Queue Files For Loading」と同じ操作(ダイアログ出現)</li> <li>・<b>Queue Files From Mascot Server For Loading</b> – 1つ上の「Queue Files for Loading」コマンドと似ていますが、dat ファイルを MASCOT Server のログから選択しネットワーク経由で取得。</li> <li>・<b>Queue Structured Directories For Loading</b> – 2 つ上の「Queue Files for Loading」と似ていますが、取得対象がファイルでなくフォルダ内のファイル群である場合に使用。「Load Data」 view→「Queue Structured Directories」と同じ操作(ダイアログ出現)</li> </ul>

メニュー	コマンド
<p><b>Experiment</b> 続き</p>	<ul style="list-style-type: none"> <li>・<b>Apply New Database</b> – データベースの追加。詳しくは 「2-4. 配列データベースの登録」を参照してください。</li> <li>・<b>Extract Alternate IDs</b> – 検索時の「Accession」とは別に 定められた ID に関する設定。Pathway など外部データベースとの連携の際に使用する ID の抜き出しルールを選択します。</li> <li>・<b>Apply Protein Annotation Preferences</b> - 類似タンパク質 グループの中で、どのタンパク質を代表として選択し表示するかに関する設定。正規表現で設定します。</li> <li>・<b>Load and Analyze Queue</b> - 「Load Data」View にて Experiment と BioSample の紐づけが完了しているものの、Experiment の 取り込みが未完了な状態の場合、取り込みを開始します。</li> <li>・<b>Reset Peptide Validation</b> – 主に Proteins View で、手動で変更した peptide valid チェック(同定ペプチドとして認識する/ しない の切り替え)をリセットします。</li> <li>・<b>Add or Edit Annotations</b> – GO または Pathway 情報の付与に関する設定を行います。詳しくは「4-5. GO の設定」、並びに「4-8. Pathway の設定」をご覧ください。</li> <li>・<b>Quantitative Analysis</b> – 定量解析(検定)を行います。詳細は 11 章の「定量手法と検定」をご覧ください。</li> </ul>
<p><b>Export</b></p> <div style="border: 1px solid red; padding: 5px; margin-bottom: 10px;"> <ul style="list-style-type: none"> <li>Subset Database...</li> <li>Spectra...</li> <li>BLIB...</li> <li>ProtXML...</li> <li>mzIdentML...</li> <li>SFDB...</li> <li>Scaffold Batch...</li> <li>Scaffold Batch Archive...</li> </ul> </div> <ul style="list-style-type: none"> <li> Publication Report...</li> <li> Samples Report...</li> <li> Spectrum Reports &gt;</li> <li> Peptide Reports &gt;</li> <li> Protein Reports &gt;</li> <li> PSEA-Quant Report...</li> </ul> <hr/> <ul style="list-style-type: none"> <li> Current View...</li> <li> Complete...</li> </ul>	<ul style="list-style-type: none"> <li>・<b>Subset Database</b> – 結果に出てきたタンパク質のうち特定の条件を満たすタンパク質のみを fasta フォーマットで出力します。</li> <li>・<b>Spectra</b> – 特定の条件を満たす ピークリストを出力します。</li> <li>・<b>BLIB</b> – blib フォーマットで同定結果の中かピークリストを出力します。skyline や Scaffold DIA などでも使用可能です。</li> <li>・<b>ProtXML</b> – protXML フォーマットで解析結果を出力します</li> <li>・<b>mzIdentML</b> – mzIdentML フォーマットで解析結果を出力します</li> <li>・<b>SFDB</b> – Scaffold PerSPECtives(別売) フォーマットでファイルを保存します。</li> <li>・<b>Scaffold Batch / Scaffold Batch Archive</b> ともに Scaffold Batch(別売)で使用するフォーマットファイルを保存します。</li> </ul> <p>[次頁に続く]</p>

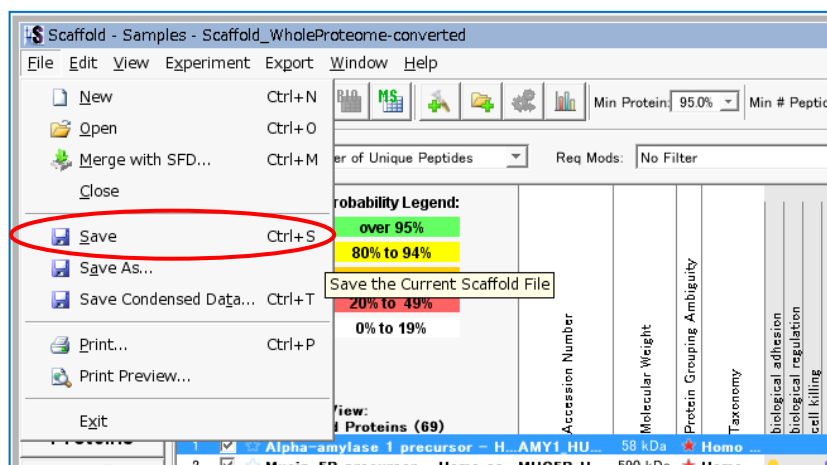
メニュー	コマンド
<p><b>Export</b> 続き</p> 	<p>以降は <b>Excel</b> で開く事ができる <b>CSV</b> フォーマットでの出力です</p> <ul style="list-style-type: none"> <li>・<b>Publication Report</b> – Publish view で表示される内容と同じ情報</li> <li>・<b>Samples Reports</b> – Samples view で表示される内容と同じ情報で出力します。出力時に、表示する数字を Display option から選択できるほか、sameset/subset のタンパク質も併せて出力するかを選ぶことができます。</li> <li>・<b>Spectrum Reports</b> – 検索結果をスペクトル単位で出力します。Spectrum の数字をそのまま出力するか、定量指標を出力するか選ぶことができます。</li> <li>・<b>Peptide Reports</b> – 検索結果をペプチド単位で出力します。Protein View の Peptide pane のデータ一覧に該当します。</li> <li>・<b>Protein Reports</b> – 検索結果をタンパク質単位で出力。Protein View の Protein pane のデータ一覧 に該当します。</li> <li>・<b>PSEA-Quant Report</b> – PSEA-Quant 解析(同定タンパク質について機能別のグループ化を行う解析)のレポートを出力します。</li> <li>・<b>Current View</b> – 現在みている View の情報を出力(*View によっては選択できません)</li> <li>・<b>Complete</b> - 様々な解析データをまとめてフォルダに出力します。 <ul style="list-style-type: none"> <li>- sample 画面の display option を変えたパターンで出力</li> <li>- タンパク質毎に Peptide Report と同内容の情報が出力</li> </ul> </li> </ul>
<p><b>Quant</b></p> 	<p>(「Q+」または「Q+S」モジュール購入時で該当データを読み込み時のみ選択可能)</p> <ul style="list-style-type: none"> <li>・<b>Launch Q+ Quantitation Browser</b> – Q+モジュールを起動</li> <li>・<b>Edit Quantitative Method / Purity Correction</b> – 定量解析に関連する値の補正に関する設定</li> </ul>
<p><b>Window</b></p> 	<ul style="list-style-type: none"> <li>・<b>Emphasize</b> – 選択中の view が複数の panel から構成される時、大きく表示したい箇所を emphasize で選択してください。</li> <li>・<b>view 名</b> - view 名と同じ項目を選択する事で View 画面の切り替えができます。画面左側「Navigation pane」で各 View のパネルをクリックするのと同じです。(表示されているショートカットキーでも画面を切り替える事ができます。)</li> </ul>



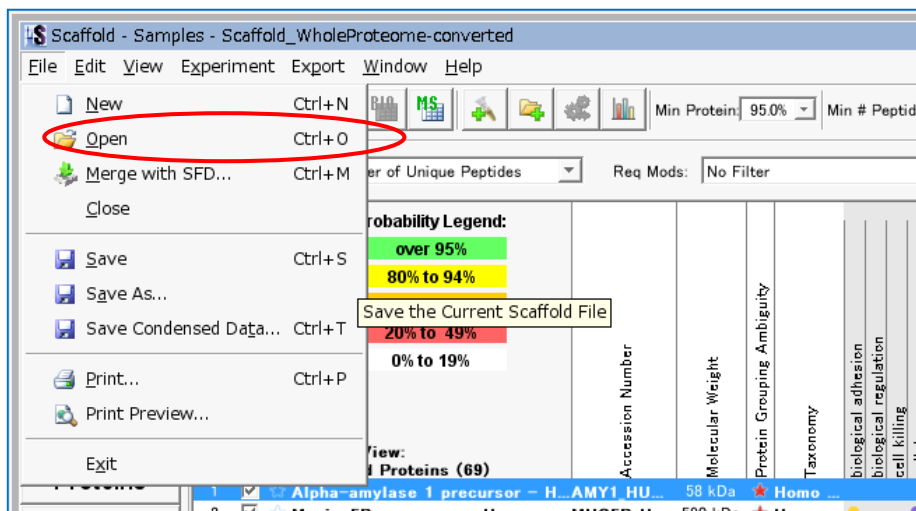
メニュー	コマンド
<p><b>Help</b></p>  <p>The screenshot shows the 'Help' menu with the following items: Help on Current View..., Help Contents, Scaffold User's Guide, Scaffold Q+ User's Guide, Open Demo Files, Scaffold FAQs/Resource Center, Show Log Files, Referencing Scaffold, Upgrade License Key..., and About Scaffold.</p>	<ul style="list-style-type: none"> <li>• <b>Help on Current View</b> – 現在開いている view に関して online help 画面が開きます。</li> <li>• <b>Help Contents</b> – online help が開きます。</li> <li>• <b>Scaffold User's Guide</b> – PDF の Scaffold マニュアルが開きます。</li> <li>• <b>Scaffold Q+ User's Guide</b> – PDF の Scaffold Q+ マニュアルが開きます。</li> <li>• <b>Open Demo Files</b> – 予め準備されている demo ファイルを開く事ができます。</li> <li>• <b>Scaffold FAQs/Resource Center</b> – Proteome Software 社のサイトにある FAQ や tutorial 用の Resource にアクセスできます。</li> <li>• <b>Show Log Files</b> – errorlog, output log が格納されているフォルダを開きます。</li> <li>• <b>Referencing Scaffold</b> – Scaffold 関連の reference (論文)に関する情報にアクセスできます。</li> <li>• <b>Upgrade License Key</b> – license を再入力できます。</li> <li>• <b>About Scaffold</b> – Scaffold のバージョンや、コピーライトに関する情報などにアクセスできます。</li> </ul>
<p><b>IdentityE</b></p>	<ul style="list-style-type: none"> <li>• <b>Quantitation Options</b> – identityE に関するオプションを指定</li> <li>• <b>Export IdentityE report</b> – IdentityE のレポート同様、解析対象ペプチド情報のタブ区切りデータを出力します。</li> </ul>

## 4-2. ファイル保存方法・ファイルを開く方法

File -> Save (又は Ctrl キー + S)でファイルを保存できます。Scaffold ファイルの拡張子は「**sf3**」です。



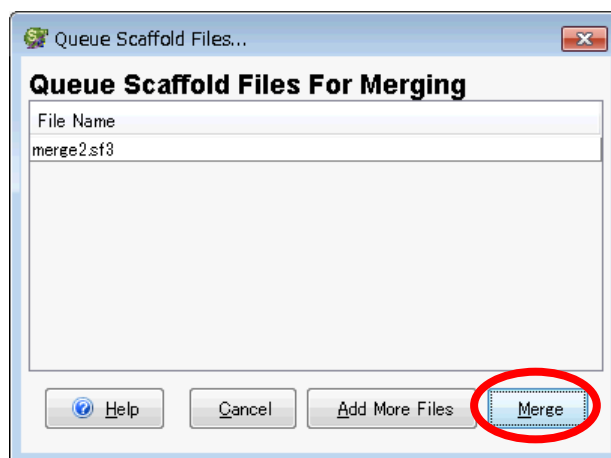
OS が日本語環境の場合ファイルできるだけデスクトップ上でなく別の場所に置くようにしてください。保存したファイルを開くには、File -> Open (Ctrl + O)とします。旧バージョンのファイル(拡張子 sf3)を開く場合、フォーマット変換の過程が入るので多少時間がかかります。ご注意ください。



### 4-3. sf3 ファイルの統合 (File ->merge)

別々に作成された sf3 ファイルを、Scaffold 上で統合することができます。

統合する時はまず一方の sf3 ファイルを Scaffold 上で開きます。File -> merge とするとファイル 選択画面になるので、もう一方の sf3 ファイルを選択します。すると右図のような Merge 操作のダイアログが現れます。ダイアログの「Merge」ボタンを押すとファイルの統合が行われます。



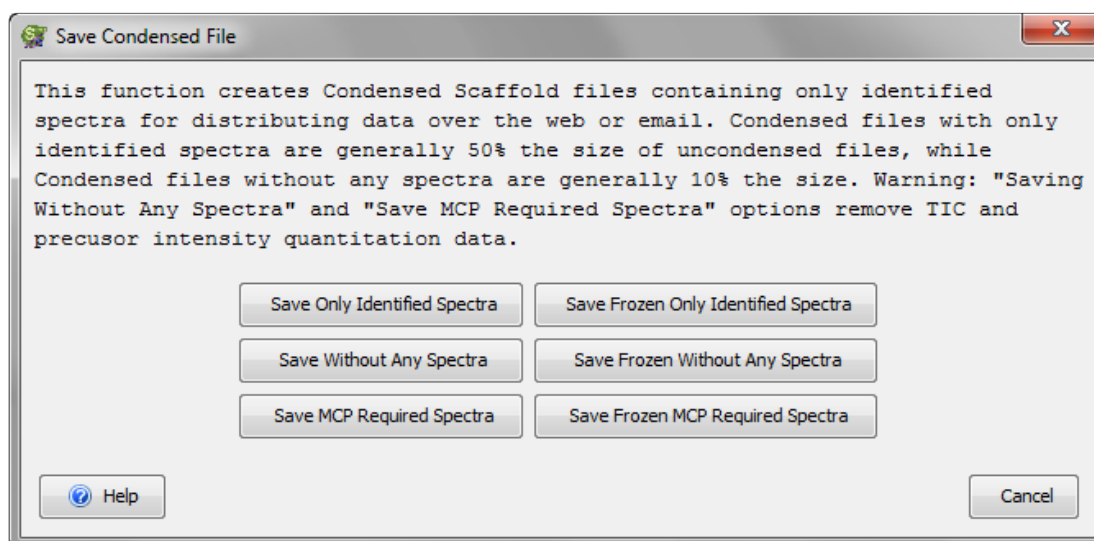
なお、Merge の逆、すなわちデータから一部のデータを削除するのは「Load Data」view から行います。詳細は「5-3.BioSample tab と Load and Analyze Queue button」をご覧ください。

## 4-4. sf3 ファイルのデータサイズを削減する方法

Scaffold で読み込む dat ファイルのサイズが大きくなり処理が遅くなったりできなくなる場合、sf3 ファイルが内部に持っている MS/MS のスペクトルデータを捨て、sf3 ファイル全体のサイズを小さくすることで対処できます。

\* ここで指定する操作でファイルを保存すると、データが完全に捨てられてしまい元の状態に戻す事ができません。試す場合などは必ずファイルのバックアップを取ってから行ってください。

File -> Save condensed data と選択すると、選択肢のダイアログが現れます(下図)。



選択肢の内容は以下の通りです。なお、「**Frozen**」とついているものは、ついていないものに加えてファイル情報を内部で圧縮しており、ファイルサイズはより小さくなりますが処理は遅くなります。

### •Save Only Identified Spectra

### •Save Frozen Identified Spectra

同定基準を超えるスコアを持つデータのみスペクトルデータを保存し残りのデータは破棄します。

### •Save Without Any Spectra

### •Save Frozen Without Any Spectra

同定基準を超えるものも含め、すべてのスペクトルデータを破棄します。各ペプチドの配列やスコアに関する情報は残ります。

### •Save MCP Required Spectra

### •Save Frozen MCP Required Spectra

MCP で提出が求められる基準に基づいたもので、基本的には Only Identified Spectra と同じです。

処理によりファイルサイズが減少します。目安として、「**Only Identified Spectra**」が元データの 50%、「**Without Any Spectra**」が元データの 90%を削除します。また Frozen の圧縮では 1/2～ 1/3 となります。

ある解析例では以下のようにファイルサイズが縮小されます。

項目	ファイルサイズ
元のデータ	1,337,646 KB
<b>Only Identified Spectra</b>	457,167 KB
<b>Without any spectra</b>	79,657 KB
<b>Save MCP Required Spectra</b>	457,134 KB
<b>Frozen Save only Identified Spectra</b>	212,247 KB
<b>Frozen Without Any Spectra</b>	36,271 KB

## 4-5. GO の設定

Scaffold の Samples ウィンドウで Gene Ontology の情報を表示させる事ができます。Gene Ontology の情報は NCBI から取得することができるほか、EBI にて公開されている Gene Ontology Annotation Database ファイルを Scaffold がインストールされている PC にダウンロードし、そのファイルから取得する方法もあります。ローカルのファイルをいったんセットすると、インターネットを介した取得より短時間で情報を取得できるほか、インターネット接続がない／回線速度が遅い 環境でも問題なく利用できます。以降、「GOA ファイルのセット」「表示する GO の設定」「GO 情報を付与、表示する方法」についてご案内します。

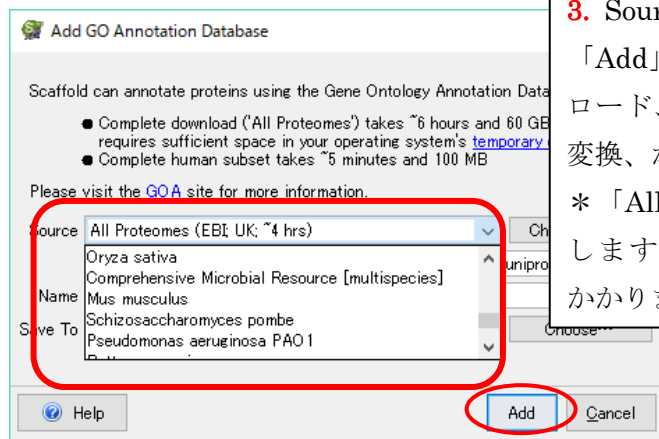
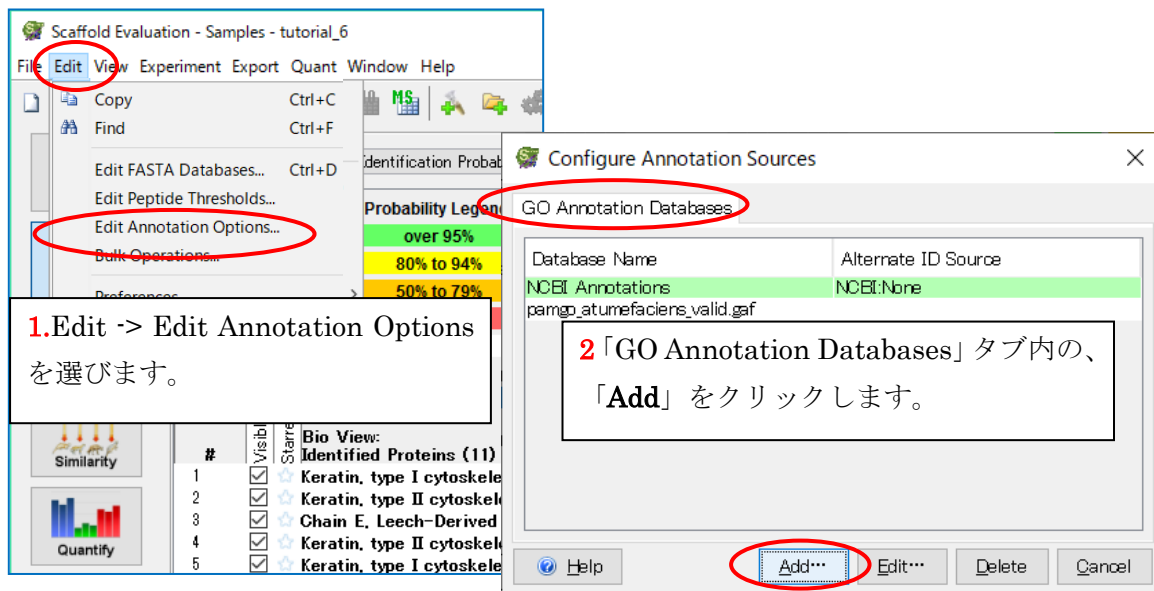
### 4-5-1. GOA ファイルのセット

Gene Ontology Annotation Database を Scaffold にセットする方法について記します。操作は GO のファイルが自動で取得できる場合と、自動取得を試みた結果何らかの理由で操作が完了しない場合にファイルを別途取得してセットする場合の 2 種類をご案内いたします。

[次頁に続きます]

## [ファイル取得を自動で行う場合]

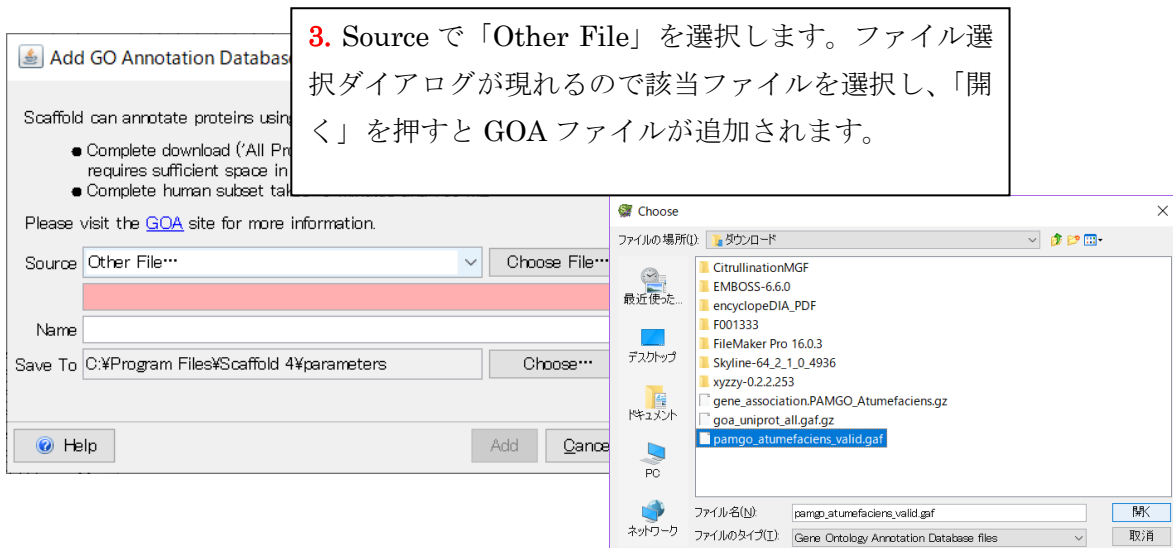
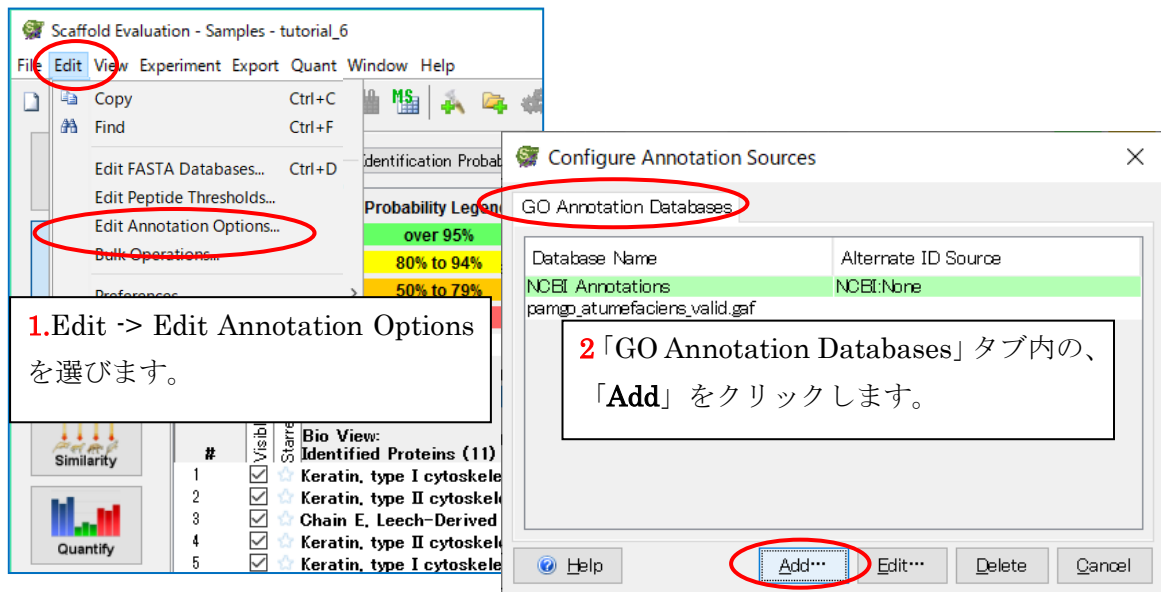
Edit -> 「Edit Annotation Options」 -> 「GO Annotation Database」にて「Add」ボタンを押します。現れるダイアログの「Source」項目で使用予定の生物種を選択し、「Add」ボタンを押します。(下図)



## [ファイルを別途準備してセットする場合]

前述の [ファイル取得を自動で行う場合] にて、ファイル取得が一定時間内に終わらなかったなど何らかの理由で実行完了しなかった場合、ファイル取得を別途行ってそのファイルを読み込むことで問題を回避することができます。

Edit -> 「Edit GO Term Options」 -> 「GO Annotation Database」にて使用予定の生物種を選択し、「Add」ボタンを押します(下図)。



## 4-5-2. 表示する GO 情報の設定

GO は「Biological process」「Cellular component」「Molecular function」の3つについて、それぞれ階層構造で構成されています。Scaffold では GO 情報を表示させることができますが、どの階層のどの項目について表示するかは以下 2 つの選択肢があります。

- PSEA-Quant を適用し得られた Gene Ontology を適用
- 自身で選択

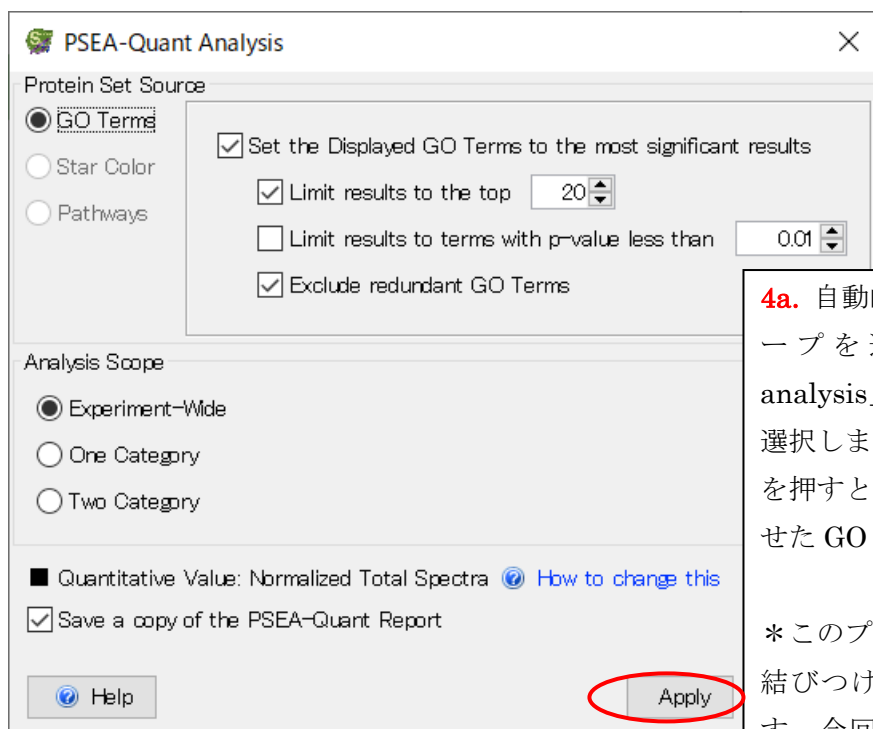
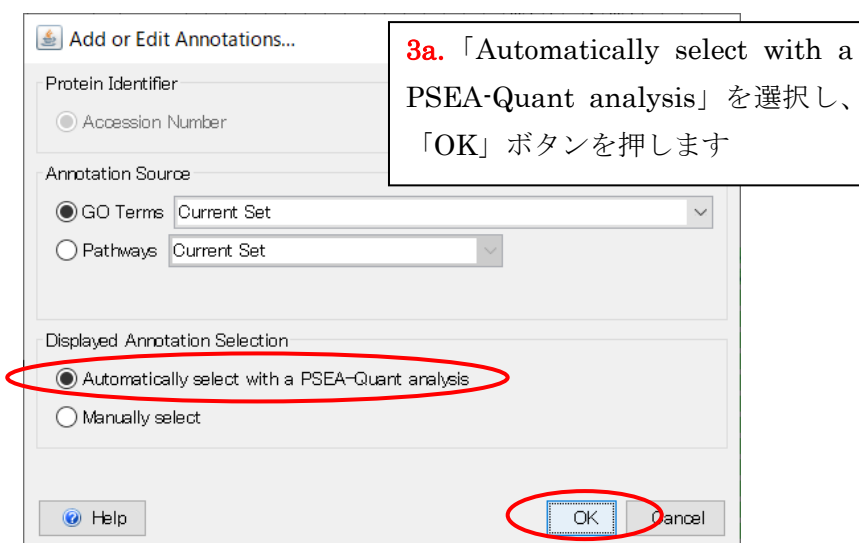
ここでは表示内容に関する設定方法をご案内します。

### [設定画面の表示]

1. Experiment -> Add or Edit Annotations を選びます。

2. Annotation Source で GO Terms を選択します。続けて、GO の表示項目を選択します (以下続きます)

[a. Automatically select with a PSEA-Quant analysis を選択する場合]

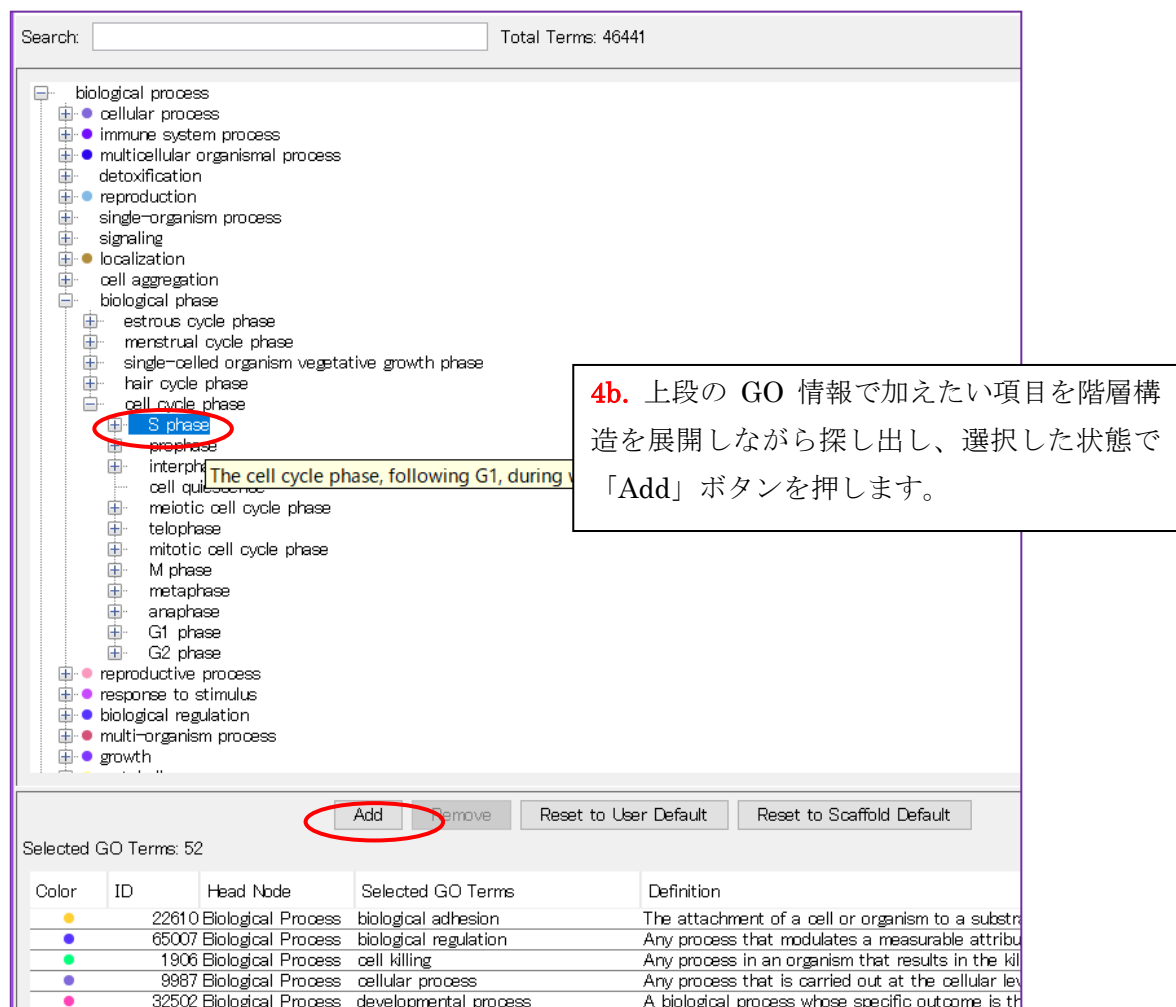
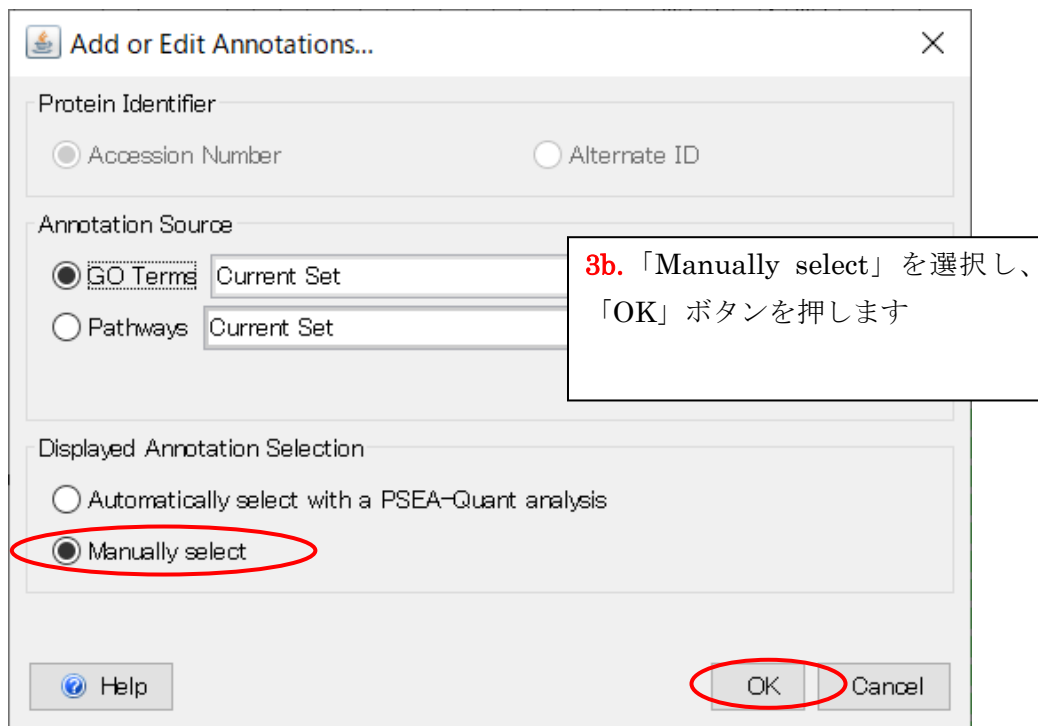


4a. 自動的に、表示に最適な GO グループを選択する 「PSEA-Quant analysis」 実行に関するパラメータを選択します。設定後、「Apply」ボタンを押すと設定が適用され、設定に合わせた GO 情報が表示されます。

\*このプログラムは本来定量解析との結びつける目的で開発されたものです。今回の選択肢でも Quantitative Value と結び付けた Report 出力に関するオプションを選択する事が出来ます。



**[b.Manual Select を選択する場合:表示項目を増やす操作]**



Selected GO Terms: 53

Color	ID	Head Node	Selected GO Terms	Definition
●	51320 Biological Process	S phase	S phase	The cell cycle phase, following G1, during which DNA synthesis takes place.
●	22610 Biological Process	biological adhesion	biological adhesion	The attachment of a cell or organism to a substrate, another cell, or other o...
●	65007 Biological Process	biological regulation	biological regulation	Any process that modulates a measurable attribute of any biological process...
●	1906 Biological Process	cell killing	cell killing	Any process in an organism that results in the killing of its own cells or those...
●	9987 Biological Process	cellular process	cellular process	Any process that is carried out at the cellular level, but not necessarily restr...
●	32502 Biological Process	developmental process	developmental process	A biological process whose specific outcome is the progression of an integrat...
●	51234 Biological Process	establishment of localization	establishment of localization	Any process that localizes a substance or cellular component. This may occur...
●	40007 Biological Process	growth	growth	The increase in size or mass of an entire organism, a part of an organism or ...
●	2376 Biological Process	immune system process	immune system process	Any process involved in the development or functioning of the immune syste...
●	51179 Biological Process	localization	localization	Any process in which a cell, a substance, or a cellular entity, such as a prot...
●	40011 Biological Process	locomotion	locomotion	Self-propelled movement of a cell or organism from one location to another...
●	8152 Biological Process	metabolic process	metabolic process	The chemical reactions and pathways, including anabolism and catabolism, by...
●	51704 Biological Process	multicellular process	multicellular process	A biological process which involves another organism of the same or differ...

5b. 下段の表示リストに該当項目が加わります。

### [Manual Select を選択する場合:表示項目を減らす操作]

Selected GO Terms: 53

Color	ID	Head Node	Selected GO Terms	Definition
●	51320 Biological Process	S phase	S phase	The cell cycle phase, following G1, during which DNA synthesis takes place.
●	22610 Biological Process	biological adhesion	biological adhesion	The attachment of a cell or organism to a substrate, another cell, or other o...
●	65007 Biological Process	biological regulation	biological regulation	Any process that modulates a measurable attribute of any biological process...
●	1906 Biological Process	cell killing	cell killing	Any process in an organism that results in the killing of its own cells or those...
●	9987 Biological Process	cellular process	cellular process	Any process that is carried out at the cellular level, but not necessarily restr...
●	32502 Biological Process	developmental process	developmental process	A biological process whose specific outcome is the progression of an integrat...
●	51234 Biological Process	establishment of localization	establishment of localization	Any process that localizes a substance or cellular component. This may occur...
●	40007 Biological Process	growth	growth	The increase in size or mass of an entire organism, a part of an organism or ...
●	2376 Biological Process	immune system process	immune system process	Any process involved in the development or functioning of the immune syste...
●	51179 Biological Process	localization	localization	Any process in which a cell, a substance, or a cellular entity, such as a prot...
●	40011 Biological Process	locomotion	locomotion	Self-propelled movement of a cell or organism from one location to another...
●	8152 Biological Process	metabolic process	metabolic process	The chemical reactions and pathways, including anabolism and catabolism, by...
●	51704 Biological Process	multicellular process	multicellular process	A biological process which involves another organism of the same or differ...

6b. 下段の GO 情報で削除したい項目を選択し「Remove」ボタンを押します。

7b. 下段の表示リストから該当項目が削除されます。

Selected GO Terms: 52

Color	ID	Head Node	Selected GO Terms	Definition
●	22610 Biological Process	biological adhesion	biological adhesion	The attachment of a cell or organism to a substrate, another cell, or other o...
●	65007 Biological Process	biological regulation	biological regulation	Any process that modulates a measurable attribute of any biological process...
●	1906 Biological Process	cell killing	cell killing	Any process in an organism that results in the killing of its own cells or those...
●	9987 Biological Process	cellular process	cellular process	Any process that is carried out at the cellular level, but not necessarily restr...
●	32502 Biological Process	developmental process	developmental process	A biological process whose specific outcome is the progression of an integrat...
●	51234 Biological Process	establishment of localization	establishment of localization	Any process that localizes a substance or cellular component. This may occur...
●	40007 Biological Process	growth	growth	The increase in size or mass of an entire organism, a part of an organism or ...

## 4-6. Preferences の設定内容

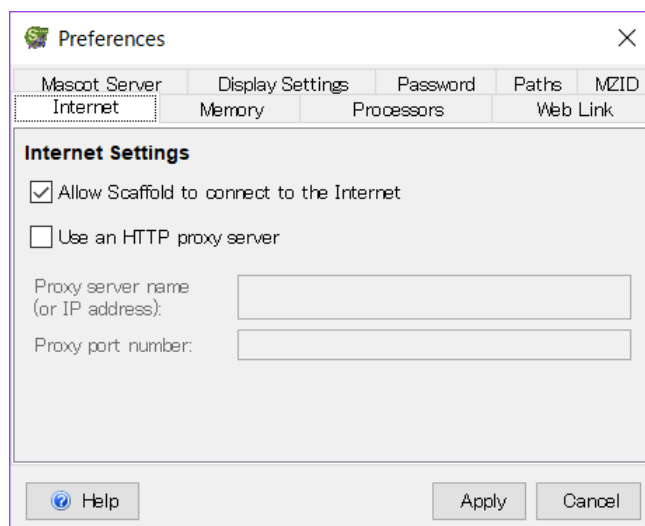
Scaffold 動作に関する設定が集まっている「Preferences」ダイアログは、メニューの「Edit」→「Preferences」と選択する事で開く事ができます。Preferences ダイアログには、以下の計9つのタブから構成されています。

「Internet」「Memory」「Processors」「Web Link」「Mascot Server」「Display Settings」「Password」「Paths Settings」「MZID」

以下、各タブで設定できることについて説明しています。

### □ Internet

Internet 接続する/しない、Proxy サーバーの設定を行う事ができます(右図)。



### □ Memory

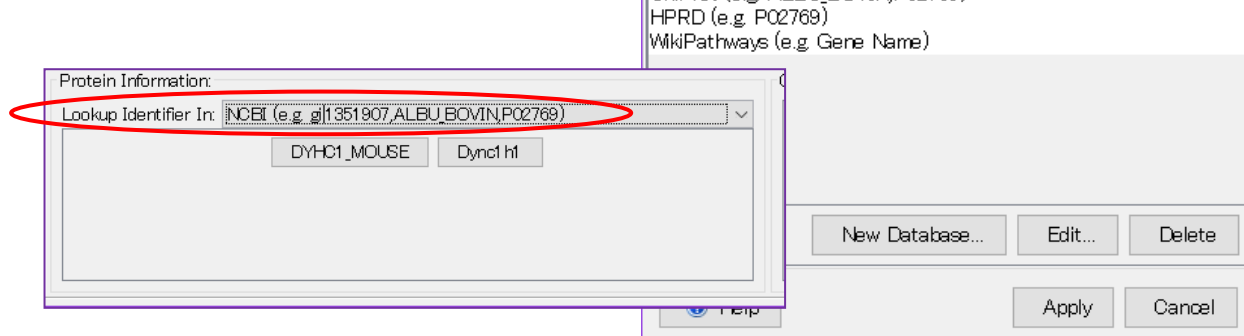
Scaffold で使用する事ができる Memory の最大値を設定する事ができます。Scaffold の動作が遅い時、設定値を上げる事で改善される事があります。設定変更の内容を反映させるためには、Scaffold ソフトウェアの再起動を行う必要があります。

### □ Processors

Scaffold で使用するコア数を指定する事ができます。ただし Scaffold 自体は使用可能なスレッド上限は2で、それ以上の設定は X!Tandem の検索時にのみ有効です。

### □ Web Link

Samples View の画面下部、Protein Information (下図、右図)で表示されるタンパク質データベースへのリンクに関する設定ができます。



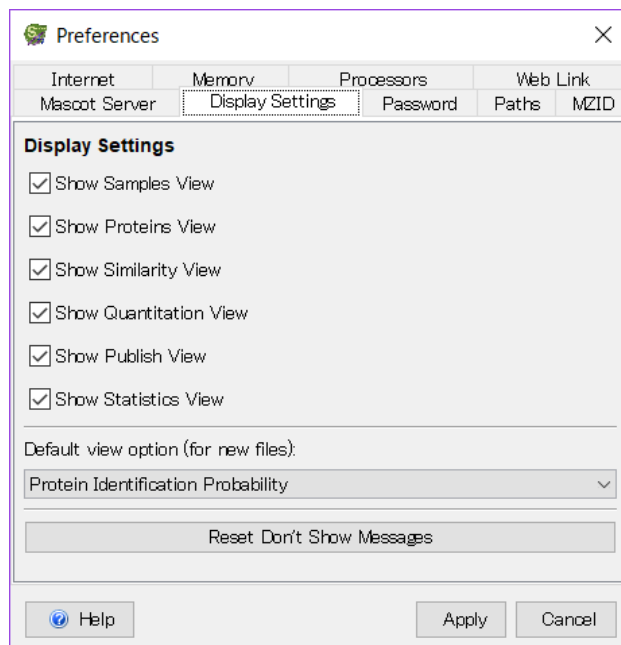
## □ Mascot Server

log ファイルを参照する MASCOT Server に 関する設定を行う事ができます。

## □ Display Settings

利用できる View を選択する事ができます(右図)。チェックを外すと画面左の Navigation pane から消え、さらに menu の Window でも選択する事ができなくなります。

またデータ取り込みやファイルオープン時に最初に開く View に関する設定も行う事ができます。



## □ Password

ファイル単位で各種操作にパスワードを設定し、パスワードを入力しないと幾つかの操作ができないようにします(右図)。設定に関する各項目についての説明は以下の通りです。

### - Use Password

パスワード使用の ON/OFF、並びにパスワードを設定します。

### - Protect Exporting Spectra

スペクトルを出力する際パスワードを要求します。

### - Protect Resetting Thresholds

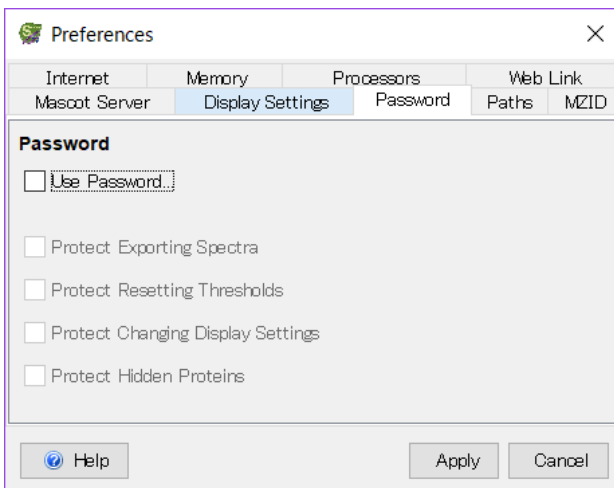
Threshold を変更する際パスワードを要求します。

### - Protect Changing Display Settings

表示 View 画面の変更をする際にパスワードを要求します。

### - Protect Hidden Proteins

Hidden protein の設定を変更する際にパスワードを要求します。



## □ Path

修飾設定で利用しているファイル unimod に関する設定を行います。

### - Do not use UNIMOD

結果ファイルに入っている修飾に関する情報をそのまま利用します。

### - Use Scaffold default UNIMOD

Scaffold が自身で持っている UNIMOD ファイルを利用します。

### - Use a custom UNIMOD file

ユーザーが準備した UNIMOD ファイルを利用します。

## □ MZID

MZID にタンパク質の配列情報が含まれている場合、MZID に含まれている方のタンパク質配列を優先して使用するよう設定することができます。

## 4-7. Advanced Preferences の設定内容

Scaffold が結果検証のアルゴリズムで使う検索エンジンのスコアについて定義することができます。設定内容により同定基準に変動があり、結果も変わってきます。SequestとMASCOT用の設定がありますが、Sequest(Proteome Discoverer) 用の設定はバージョンが 1.2 より古い場合のみ適用を検討するもので、多くの人は当てはまりません。

### [Sequest]

バージョンが 1.2 より古い場合、以下の資料をご参照の上設定値を検討してください。ver.1.3 以上の場合、設定画面にあるように最適な設定を「Auto-Detect」します。

・PDF(英文) マニュアル

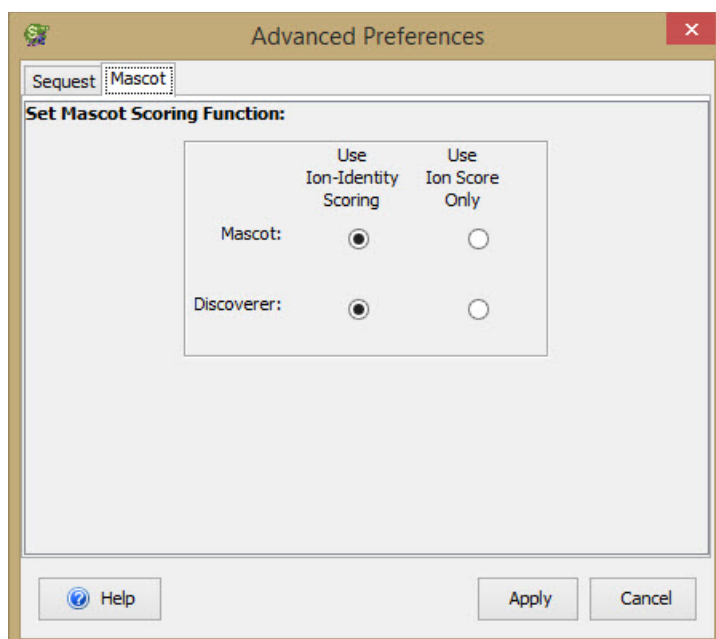
Chapter.4 Scaffold's Main Window -> Advanced Preferences -> Sequest tab

・Online help

Main menu component ->Advanced Preferences -> Sequest tab

### [MASCOT]

「Ion - Identity Scoring も利用」するか「Ion Score のみ使用」するかを選択することができます。設定は、MASCOT から来た結果か、Proteome Discoverer から来た結果かでさらに分かれています。



## 4-8. Pathway に関する設定

Scaffold の Samples ウィンドウで Pathway の情報を表示させる事ができます。Pathway の情報は「Wikipathways」または「Reactome」サイト から取得する事ができます。以降、「データベースの使用条件 (タンパク質 ID)」、「適用の実施方法」「お勧めの Pathway 情報表示」についてご案内します。

### 4-8-1. Wikipathways, Reactome と使用時のタンパク質 ID について

WikiPathways, Reactome はそれぞれ反応パスウェイについてまとめられたデータベース[サイト]です。詳細は以下情報をご参照ください。

#### Wikipathways

Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR. (2009) Mining Biological Pathways Using WikiPathways Web Services. PLoS ONE 4(7)

#### Reactome

Fabregat et al. 2017 PMID: 28249561

Wikipathways を利用するためには、「gene name」情報が必要です。Scaffold にて gene name は FASTA ファイル内の情報、または NCBI のサイトから取得します。gene name 情報の取得並びに表示させるには、「Experiment」->「**Extract Alternate IDs**」を実施する必要があります。

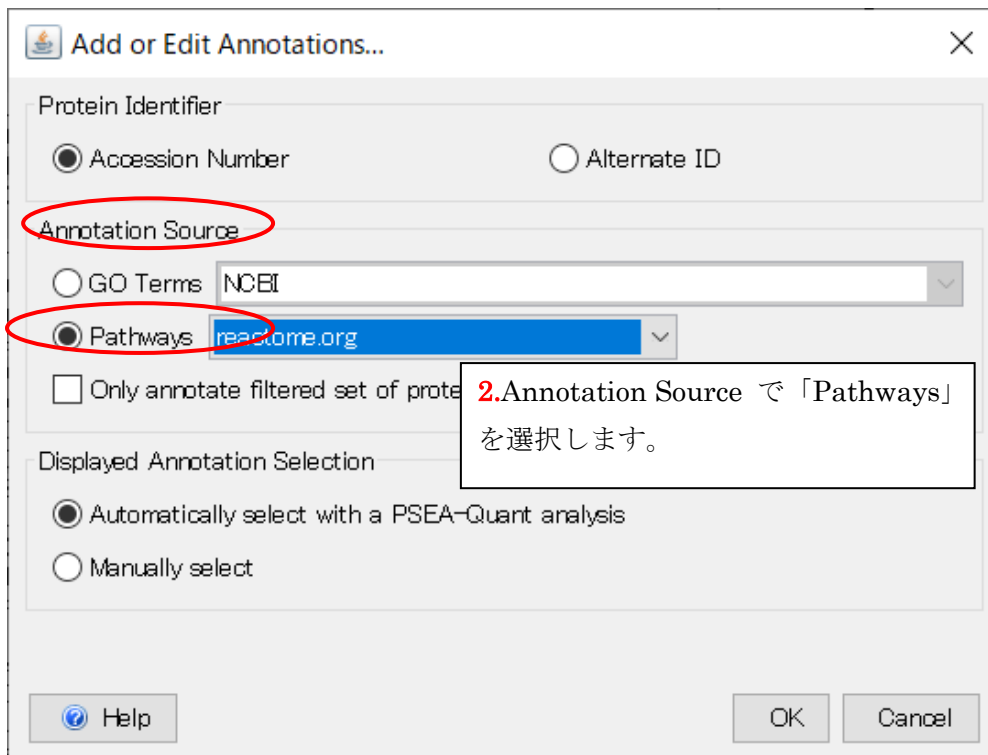
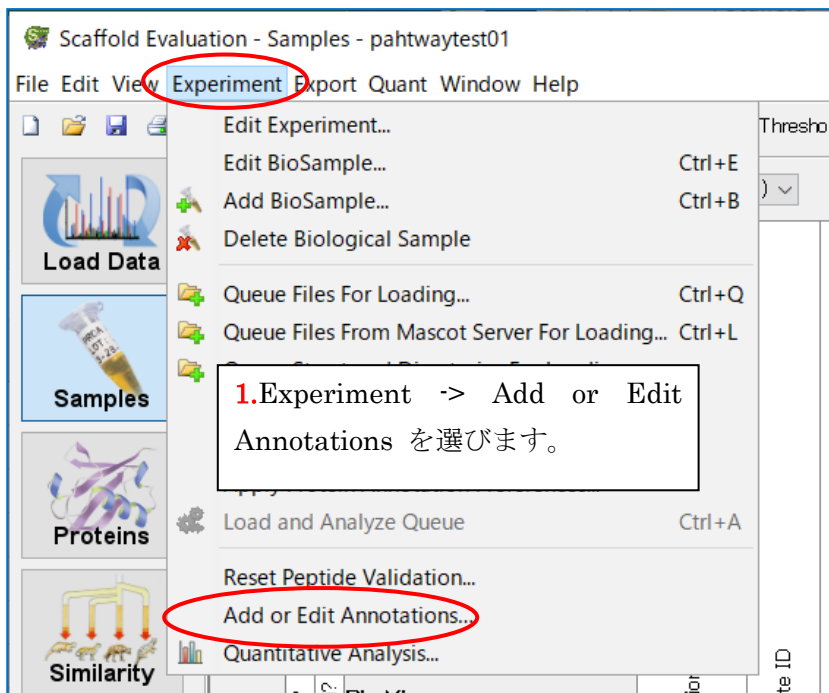
一方、Reactome データベースを利用する際も **Uniprot の Accession** (例:P02769 など。'ALBU\_BOVIN などの表記ではない)情報が必要となります。そのため、Reactome 使用を前提とする場合、**検索時から Accession (P02769 など)が ID となっているデータベースに対する検索結果を Scaffold に取り込むように注意してください。**

### 4-8-2. Pathway 情報の表示 [Scaffold 上]

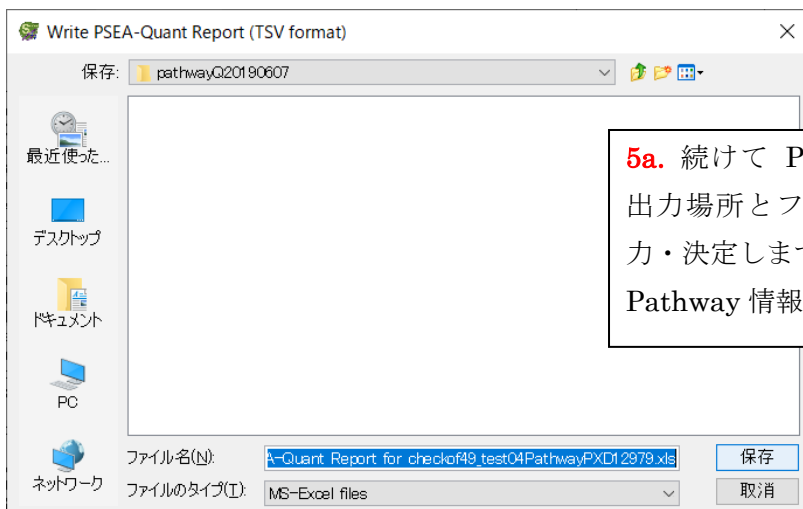
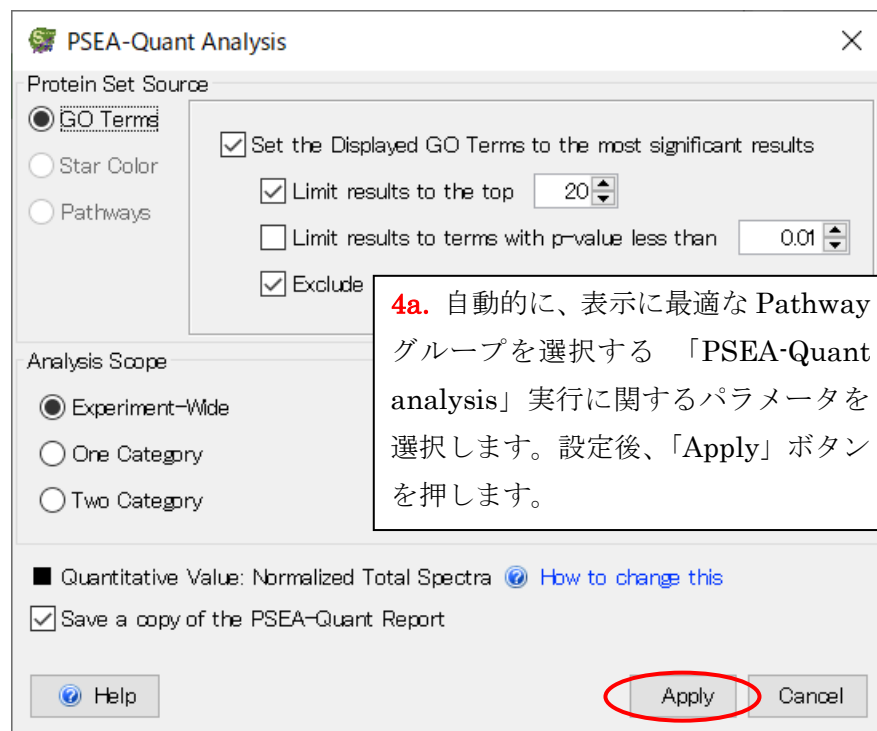
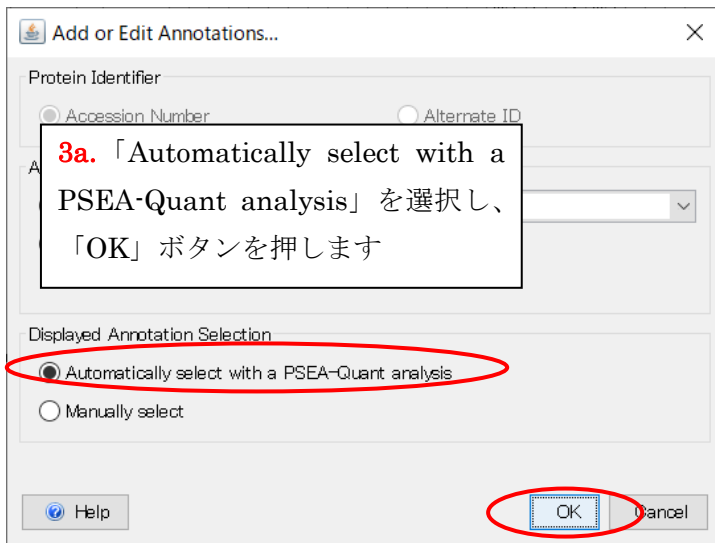
Pathwayに関する情報を Scaffold の Samples 画面で表示させる事ができます。ただし Scaffold 上では該当タンパク質が含まれる Pathway の種類が表示されるのみです。タンパク質が含まれる Pathway 一覧や、特定 Pathway の全体図とその中で該当タンパク質が占める位置の確認については、外部サイトの情報を確認する必要があります。外部サイトの参照については、「**4-8-3. Pathway 情報の表示 [外部サイト]**」をご参照ください。

次頁以降、Scaffold 上で Pathway 情報を表示させる操作方法についてご案内しています。

[設定画面の表示]



[a. Automatically select with a PSEA-Quant analysis を選択する場合]





Scaffold Evaluation - Samples - checko49\_test04PathwayPXD12979

File Edit View Experiment Export Quant Window Help

Protein Threshold: 99.0% Min # Peptides: 5 Peptide Threshold: 0%

Display Options: Total Unique Peptide Count Req Mods: No Filter Search: Advanced

Probability Legend:

- over 95%
- 80% to 94%
- 50% to 79%
- 20% to 49%
- 0% to 19%

6a. Pathway 情報が表示されます。

2373 Proteins in 2114 Clusters  
With 2295 Filtered Out

#	Protein Name	Accession Number	Molecular Weight	Pathway
2	Clathrin heavy chain 1	CLH1_MOUSE	192 kDa	
3	ATP-citrate synthase	ACLY_MOUSE	120 kDa	
4	Cluster of Tubulin beta-5 chain	TUBB5_MOUSE [8]	50 kDa	
5	Cluster of Ras-related protein Rab14	RAB14_MOUSE [18]	24 kDa	
6	Endoplasmic reticulum chaperone BiP	ENPL_MOUSE	92 kDa	
7	Spectrin alpha chain, non-erythrocytic 1	SPTN1_MOUSE	285 kDa	
8	Cluster of heat shock cognate 71 kDa protein	HSPA8_MOUSE [5]	71 kDa	
9	Myosin-9	MYH9_MOUSE	226 kDa	
10	Cluster of Alpha-actinin-4	ACTN4_MOUSE [4]	105 kDa	
12	Ras GTPase-activating-like protein IQGAP1	IQGA1_MOUSE	189 kDa	
13	Transitional endoplasmic reticulum ATPase	TERA_MOUSE	89 kDa	
14	Spectrin beta chain, non-erythrocytic 1	SPTB1_MOUSE	274 kDa	
15	Neutral alpha-glucosidase A3	GARB3_MOUSE	107 kDa	
16	Hypoxia up-regulated protein 1	HYOU1_MOUSE	111 kDa	
17	UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1_MOUSE	176 kDa	

Protein Information: Lookup Identifier In: NCEI (e.g. gl|1351907|ALBU\_BOVIN|P02769) DYNC1\_MOUSE Dync1 hl

Annotation: Sample Information: Biological Sample: Sample Category: Sample Description: MS/MS Sample: MS/MS Sample Notes:

**[b.Manual Select を選択する場合:表示項目を増やす操作]**

Add or Edit Annotations...

Protein Identifier

Accession Number  Alternate ID

3b. 「Manually select」 を選択し、「OK」 ボタンを押します

Pathways  Current Set

Displayed Annotation Selection

Automatically select with a PSEA-Quant analysis

Manually select

Help OK Cancel

**Configure Reactome**

Resource: UniProt

Species

- Alphapapillomavirus 9
- Arenicola marina
- Bacillus anthracis
- Bos taurus
- Caenorhabditis elegans
- Candida albicans
- Canis familiaris
- Cavia porcellus
- Cercopithecus aethiops
- Chlamydia trachomatis
- Chlorocebus sabaeus
- Clostridium botulinum
- Clostridium botfringens
- Clostridium tetani
- Corynephage beta
- Cowpox virus
- Oricetulus griseus
- Oribidia fasciculata

Search for species:\*

**4b.** 表示させたい Pathway 項目を手動で選択し、チェックを入れます。必要に応じてダイアログ下部にある検索ウィンドウなどを使用します。すべて選択後、「OK」ボタンを押します。

OK Cancel

Scaffold Evaluation - Samples - checkof49\_test04PathwayPXD12979

File Edit View Experiment Export Quant Window Help

Protein Threshold: 99.0% Min # Peptides: 5 Peptide Threshold: OK

Display Options: Total Unique Peptide Count Req Mods: No Filter Search: Advanced

**5b.** Pathway 情報が表示されます。

Bio View: 2373 Proteins in 2114 Clusters With 2295 Filtered Out

#	Protein Name	Accession Number	Alternate ID	Molecular Weight
2	Clathrin heavy chain 1 OS=Mus musculus OX=10090 GN=Cltc PE=1 SV=3	CLH1_MOUSE	Cltc	192 kDa
3	ATP-citrate synthase OS=Mus musculus OX=10090 GN=Acly PE=1 SV=1	ACLY_MOUSE	Acly	120 kDa
4	Cluster of Tubulin beta-5 chain OS=Mus musculus OX=10090 GN=Tubt5 PE=1 SV=1 (TBE5_MOUSE)	TBE5_MOUSE [8]	Tubt5	50 kDa
5	Cluster of Ras-related protein Rab-14 OS=Mus musculus OX=10090 GN=Rab14 PE=1 SV=3 (RAB14_MOUSE)	RAB14_MOUSE [16]	Rab14	24 kDa
6	Endoplasmic reticulum chaperone BIP OS=Mus musculus OX=10090 GN=Hspa5 PE=1 SV=2	ENPL_MOUSE	Hsp90b1	92 kDa
7	Endoplasmic reticulum chaperone BIP OS=Mus musculus OX=10090 GN=Hspa5 PE=1 SV=3	BIP_MOUSE	Hspa5	72 kDa
8	Spectrin alpha chain, non-erythrocytic 1 OS=Mus musculus OX=10090 GN=Sptnl PE=1 SV=4	SPTN1_MOUSE	Sptnl	265 kDa
9	Myosin-9 OS=Mus musculus OX=10090 GN=Myh9 PE=1 SV=4	MYH9_MOUSE	Myh9	226 kDa
10	Cluster of Heat shock cognate 71 kDa protein OS=Mus musculus OX=10090 GN=Hspa8 PE=1 SV=1 (HSPA8_MOUSE)	HSP70_MOUSE [5]	Hspa8	71 kDa
11	Cluster of Alpha-actinin-4 OS=Mus musculus OX=10090 GN=Actn4 PE=1 SV=1 (ACTN4_MOUSE)	ACTN4_MOUSE [4]	Actn4	105 kDa
12	Ras GTPase-activating-like protein IGGAP1 OS=Mus musculus OX=10090 GN=Iggap1 PE=1 SV=2	IGGA1_MOUSE	Iggap1	189 kDa
13	Transitional endoplasmic reticulum ATPase OS=Mus musculus OX=10090 GN=Vcp PE=1 SV=4	TERA_MOUSE	Vcp	89 kDa
14	Spectrin beta chain, non-erythrocytic 1 OS=Mus musculus OX=10090 GN=Sptbn1 PE=1 SV=2	SPTB2_MOUSE	Sptbn1	274 kDa
15	Neutral alpha-glucosidase AB OS=Mus musculus OX=10090 GN=Garb1 PE=1 SV=1	GAMAB_MOUSE	Garb1	107 kDa
16	Hypoxia up-regulated protein 1 OS=Mus musculus OX=10090 GN=Hyoul PE=1 SV=1	HYOUI_MOUSE	Hyoul	111 kDa
17	UDP-glucose glycoprotein glucosyltransferase 1 OS=Mus musculus OX=10090 GN=Uggt1 PE=1 SV=4	UGG1_MOUSE	Uggt1	176 kDa

Protein Information: Lookup Identifier In: NCB1 (e.g. g|1351907|ALBU\_BOVIN|P02769) [DYHC1\_MOUSE] [Dyrd1\_H]

Annotation: [Empty]

Sample Information: Biological Sample: [Empty] Sample Category: [Empty] Sample Description: [Empty] MS/MS Sample: [Empty] MS/MS Sample Notes: [Empty]

2373 Proteins at 99.0% Minimum  
5 Min # Peptides  
6.3% Decoy FDR  
214497 Spectra at 0.0% Minimum  
1.23% Decoy FDR

### 4-8-3. Pathway 情報の表示 [外部サイト]

Scaffold 上では各タンパク質が含まれる Pathway 情報が表示されるのみにとどまります。Pathway の詳細情報については、公開元の WikiPathways または Reactome サイト上で確認することになります。両サイトへのリンクは Scaffold の Samples 画面下部、「Protein Information」pane と「Annotation」pane にあります。

The screenshot shows the Scaffold interface with a list of proteins. The 'Protein Information' pane is highlighted with a red box and contains the following information:

- Lookup Identifier In: WikiPathways (e.g. Gene Name)
- Buttons: BIP\_MOUSE, Hspa5

The 'Annotation' pane is also highlighted with a red box and contains the following information:

- Pathways
- Prion disease pathway (Homo sapiens)
- Photodynamic therapy-induced unfolded protein response (Homo sapiens)
- ATF6 (ATF6-alpha) activates chaperone genes (Homo sapiens)

This is a close-up view of the panes shown in the previous screenshot. The 'Protein Information' pane shows the 'Lookup Identifier In' dropdown set to 'WikiPathways (e.g. Gene Name)' and two buttons labeled 'BIP\_MOUSE' and 'Hspa5'. The 'Annotation' pane shows a list of pathways: 'Prion disease pathway (Homo sapiens)', 'Photodynamic therapy-induced unfolded protein response (Homo sapiens)', and 'ATF6 (ATF6-alpha) activates chaperone genes (Homo sapiens)'.

左側の「Protein Information」にある gene name (または Accession) ボタンをクリックすると、該当タンパク質でサイト上にて検索した Pathway 一覧が表示されます(下図)。

The screenshot shows the WikiPathways search results page. The search bar contains 'Hspa5' and the results show 11 pathways found. The first three pathways are highlighted:

- Photodynamic therapy-induced unfolded protein response (Homo sapiens)
- ATF6 (ATF6-alpha) activates chaperone genes (Homo sapiens)
- Unfolded Protein Response (UPR) (Homo sapiens)

また真ん中の「Annotation」paneには、検索項目でマッチした Pathway の項目リストがあり、それぞれのハイパーリンクをクリックすると該当 Pathway が表示されます。

Pathway の中で該当タンパク質がどの位置に存在するかについては、表示された WEB ページ内の検索で該当タンパク質名を探してください。

Search: HSPA5 1 / 1

ページ内で該当タンパク質を検索

pathway discussion view source

### Prion disease pathway (Homo sapiens)

Lot van de Wouw, Friederike Ehrhart, Kristina Hanspers, Egon Willighagen, et al.

Regulators of PRNP			
PAX5	POU2F2	NFKB1	
EP300	CHD2	RAD21	CTCF SPI1
MEF2C	BATF	SMC3	STAT3 RFX5
TBP	RXRA	BCL11A	EBF1 IRF4

タンパク質が強調表示される

RESET

「Protein Information」または「Annotation」 pane についての詳細は、「3-2-4. Information Panes: タンパク質・サンプルに関する追加情報」も併せてご確認ください。

[次頁に続く]

## 4-9. ツールバーアイコン

使用頻度が高い項目については、画面左上にアイコンが準備されています。ほとんどのアイコンはメニューに同じ項目があります。

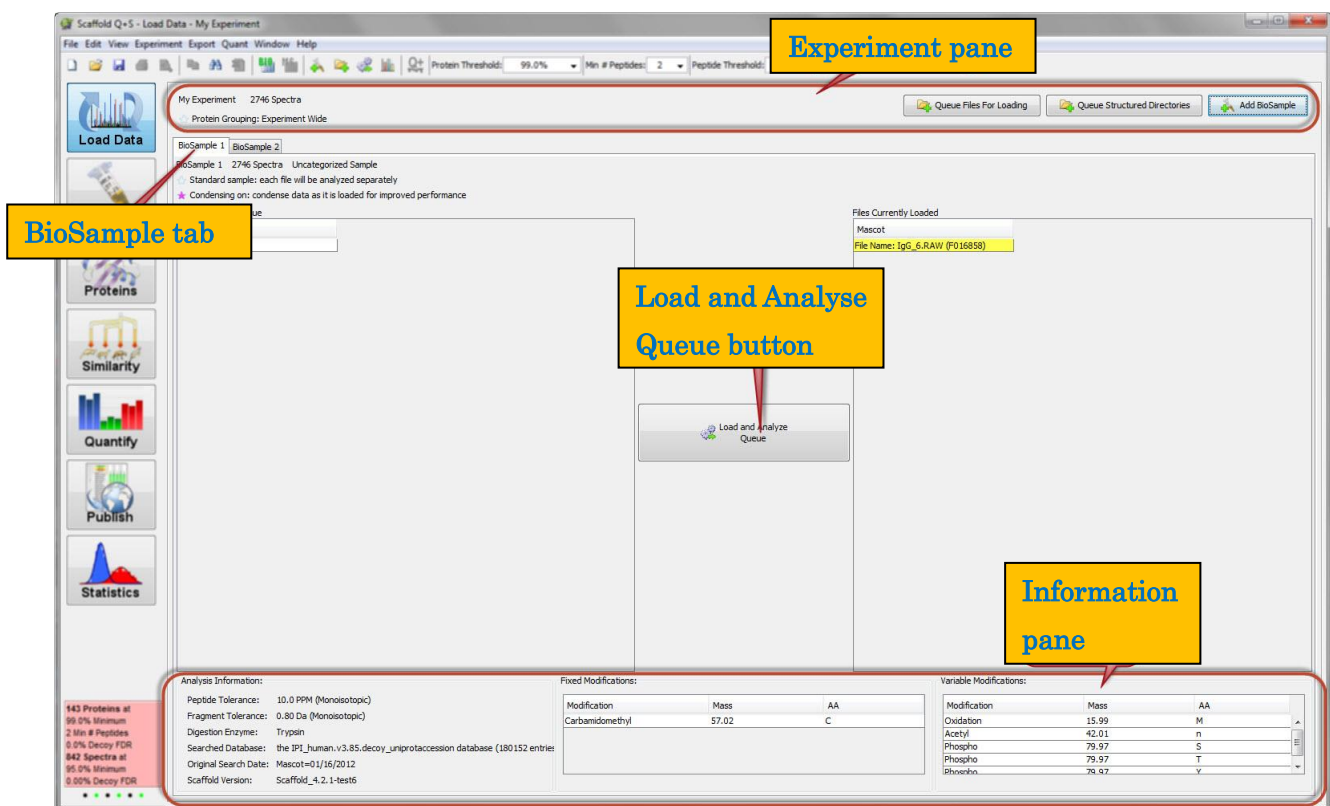
アイコン	説明
	・ <b>New</b> - ファイル作成ウィザードを起動します。詳細は 2 章「MASCOT 結果取り込み」をご覧ください。
	・ <b>Bach Job Queue Dialog</b> - Scaffold Batch の実行ダイアログを表示。このアイコンは Scaffold Batch のライセンス購入時のみ表示されます。
	・ <b>Open</b> - Scaffold のファイル(.sf3)を開きます。
	・ <b>Save</b> - 今開いている sf3 ファイルを保存します
	・ <b>Print</b> - 現在開いている view 画面を印刷します。
	・ <b>Print Preview</b> - 印刷の preview を表示します。
	・ <b>Copy</b> - 選択時に開いている view のデータをそのままクリップボードにコピーします。タブ区切りのデータとなります。
	・ <b>Find</b> - 検索用のダイアログを開き view から該当項目を探します。
	・ <b>Excel</b> - 現在の View 画面情報について、CSV フォーマットで出力。Export- current view と同じ。
	・ <b>BioSample Summarization level</b> - データを BioSample 単位にまとめた表示に切り替えます。
	・ <b>MS/MS Sample Summarization level</b> - データを Experiment (MS/MS Sample) 単位にまとめた表示に切り替えます。
	・ <b>Add BioSample</b> - BioSample を追加。「Load Data」view→「Add BioSample」と同じ操作(ダイアログ出現)
	・ <b>Queue Files For Loading</b> - Experiment を追加。「Load Data」view→「Queue Files For Loading」と同じ操作(ダイアログ出現)
	・ <b>Load and Analyze Queue</b> - 「Load Data」View にて Experiment と BioSample の紐づけが完了しているものの、Experiment の取り込みが未完了な状態の場合、取り込みを開始します
	・ <b>Quantitative Analysis</b> - 定量解析(検定)を行います。詳細は「」
	・ <b>Scaffold Q+/Q+S</b> - Q+/Q+S モジュールを起動します。
	・ <b>Help</b> - Online help を起動します。

## 5. Load Data View

### 5-1. 概要

Scaffold では各種機能を持つ View があり、画面左側にそれらの View を切り替えるためのスイッチがあります。5 章では「Load Data」View について説明しています。

「Load Data」View では現在取り込んでいるデータの BioSample に関して MS Sample データを追加/削除 したり、BioSample 自体を追加/削除 する事ができます。



以降 Load Data View 画面について、上図で示す4つのパーツ、

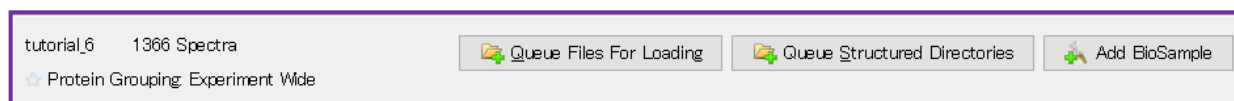
- **Experiment pane**
- **BioSample tabs** と、**Load and Analyse Queue button**
- **Information pane**

に分けて説明していきます。

## 5-2. Experiment pane

View 画面上部、「Experiment pane」(下図)では Scaffold で取り込んでいるデータに関する概要を確認する事ができるほか、BioSample\*や MS Sample の追加を行う事ができます。

\* **BioSample** … (MS) sample データをまとめた、1 つ上の階層単位。詳細は「**2-1. 概要、結果ファイルの階層構造**」をご確認ください。



Experiment pane 左側は Experiment の名称 (sf3 ファイルの名称)や取り込んだスペクトルの総数、タンパク質のグループ化の設定が表示されています。

右側には3つのボタンがあります。

### • Queue Files For Loading

現在選択されている BioSample に、新たに MS Sample を加えます。加える MS Sample が 1 ファイルの時に使用します。

### • Queue Structured Directories

現在選択されている BioSample に、新たに MS Sample を加えます。加える MS Sample がフォルダ構造になっている時に使用します。

### • Add BioSample

BioSample を追加する時に利用します。

以下内容の関連項目として、2 章の「MASCOT 結果取り込み」も併せてご参照ください。

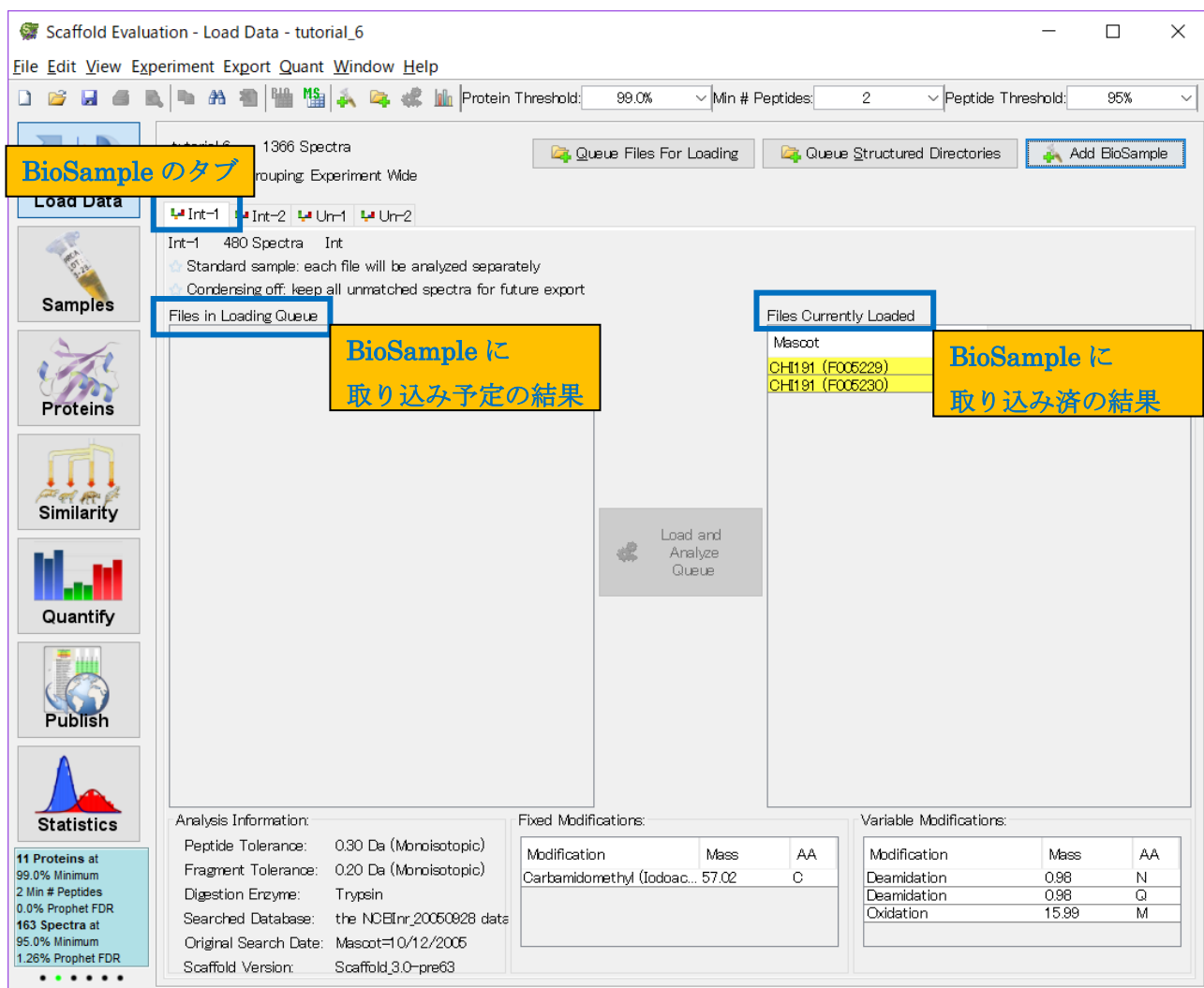
- MS Sample, BioSample, Category (**2-1. 概要、結果ファイルの階層構造**)
- データの取り込み操作 (**2-2, 2-3, 2-4**)



### 5-3. BioSample tab と Load and Analyze Queue button

Load Data View の主体部分では、BioSample 別にタブで構成された画面が表示されています(下図)。タブには BioSample 名が、そのすぐ下にはスペクトル数と(BioSample より上位に位置する)Category 名が表示されています。その下の行には以下2つのオプションの現状について表示されています。

- データのまとめ方(標準か MUDPIT 設定か)
- スペクトルデータの扱い (condense 設定)



その下に続いている左右二つの表ですが、画面の右側の表「Files Currently Loaded」が現在 BioSample に取り込み済みの MS Sample を表しています。一方左側の表「Files in Loading Queue」は、追加はされていないが追加されるように設定されている MS Sample を表しています。左側に項目がある状態で、真ん中のボタン「**Load and Analyze Queue**」を押すと、データの取り込みを開始します。

左右どちらの表に含まれる MS Sample も、右クリック→Remove とすることで取り除く事ができます。特に右側の「Files Currently Loaded」のデータを取り除くと、各種再計算が自動的に実行されます。



## 5-4. Information pane

Load Data View の下部にある表示が「Information pane」です。

左から順に「**Analysis Information**」「**Fixed Modifications**」「**Variable Modifications**」の sub pane から構成されています。

検索条件	修飾 (fixed)	修飾 (variable)																		
<b>Analysis Information:</b> Peptide Tolerance: 0.30 Da (Monoisotopic) Fragment Tolerance: 0.20 Da (Monoisotopic) Digestion Enzyme: Trypsin Searched Database: the NCBI nr_20050928 datat Original Search Date: Mascot=10/12/2005 Scaffold Version: Scaffold_3.0-pre63	<b>Fixed Modifications:</b> <table border="1"><thead><tr><th>Modification</th><th>Mass</th><th>AA</th></tr></thead><tbody><tr><td>Carbamidomethyl (Iodoac...</td><td>57.02</td><td>C</td></tr></tbody></table>	Modification	Mass	AA	Carbamidomethyl (Iodoac...	57.02	C	<b>Variable Modifications:</b> <table border="1"><thead><tr><th>Modification</th><th>Mass</th><th>AA</th></tr></thead><tbody><tr><td>Deamidation</td><td>0.98</td><td>N</td></tr><tr><td>Deamidation</td><td>0.98</td><td>Q</td></tr><tr><td>Oxidation</td><td>15.99</td><td>M</td></tr></tbody></table>	Modification	Mass	AA	Deamidation	0.98	N	Deamidation	0.98	Q	Oxidation	15.99	M
Modification	Mass	AA																		
Carbamidomethyl (Iodoac...	57.02	C																		
Modification	Mass	AA																		
Deamidation	0.98	N																		
Deamidation	0.98	Q																		
Oxidation	15.99	M																		

**Analysis Information** では、検索に関する条件が表示されています。

**Fixed Modifications** 並びに **Variable Modifications** では、検索時に指定した修飾に関する情報が記されています。

各種表示は、BioSample tab 内の Files Currently Loaded で MS Sample が指定されている場合、選択 Sample での情報が表示されます。何も選択されていない場合はすべての sample について該当項目が列挙される形で表示されます(ただし通常は検索条件がすべて同じデータを取り込むケースがほとんどであると考えます)。

# 6. Protein View

## 6-1. 概要

Scaffold では各種機能を持つ View があり、画面左側にそれらの View を切り替えるためのスイッチがあります。

6 章では「Proteins」View について説明いたします。

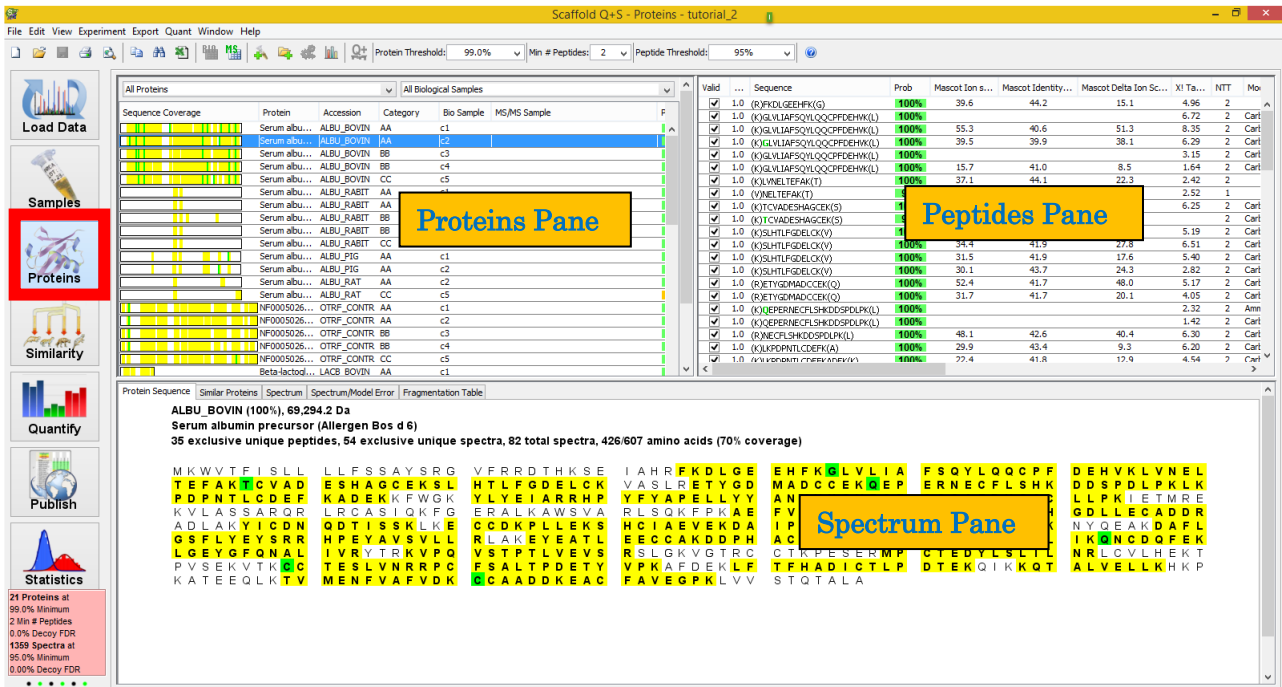
各タンパク質に関する詳細な情報、例えばタンパク質全長に対してアサインされた同定ペプチドの分布状況であったり、マッチしたペプチドの MS/MS マススペクトルとのマッチング状況を確認したりといった事は、「Proteins」View で行います。

Proteins View は、主に3つのパーツで構成されています。

左上の「Proteins Pane」では同定タンパク質に関する様々な情報を表示しています。

右上の「Peptides Pane」では、同定タンパク質にアサインされているペプチドに関する様々な情報を表示して、Proteins Pane の選択項目と連動しています。

画面下部の「Spectrum Pane」では、Proteins/Peptides pane で選択している内容に関連する各種図/グラフ が表示されます。以降各 pane についてより詳しく説明しています。





## •MS/MS Sample

データの MS sample 名

## •m.w.

タンパク質の質量

## •Prob.

protein probability の数値。prefiltered mode で取り込んだ場合は表示されません。

## •%Spec.

Sample View の Display option にもある、「Percentage of all Spectra」

## •#Pep.

Sample View の Display option にもある、「Exclusive unique peptide count」

## •#Unique

Sample View の Display option にもある、「Exclusive unique spectrum count」

## •#Spec

タンパク質のグループ化(クラスタリング)設定をしている場合、Display option の「Total spectrum Count」を表します。グループ化設定をしていない場合、Display option の「Exclusive Spectrum Count」を表します。列の項目名にカーソルを合わせるとどちらが表示されているか、確認する事ができます。

## •%Cov.

Sequence Coverage (%)

## 6-3. Peptide pane : ペプチドに関する情報を表示

Protein View の右上、Peptides pane では、左上で選択された protein にアサインされたスペクトルデータに関する情報が表示されています(下図)。

選択中のタンパク質について...

Valid	Assigned	Sequence	Prob	Misc.	Misc.	Misc.	MT	Modifications	Observed	Actual M.	Other	Delta	Delta	Rate	Intensity	TIC	Start	Stop	# Cl.	Other Pr.	Spectrum ID
✓	✓	(R)KAEKLEK	100%	612	468	00	2		599.59	1,200.00	2	-0.043	-39		4381	109	172	0		Sum of 3 scan	
✓	✓	(R)MPELTLHA	100%	632	473	00	2		714.30	1,202.59	2	-0.006	-6.0		4681	288	266	0		Sum of 3 scan	
✓	✓	(R)SQ	100%	675	468	00	2		693.28	1,204.69	2	-0.005	-18		2240	323	323	0		Sum of 3 scan	
✓	✓	(R)QK	100%	579	468	00	2		693.30	1,204.59	2	-0.046	-39		16940	323	323	0		Sum of 3 scan	
✓	✓	(R)QK	100%	450	467	00	2		747.28	1,462.62	2	-0.111	-79		13510	323	244	0		Sum of 3 scan	
✓	✓	(R)QAEWPKNS	100%	482	466	00	2		585.25	1,108.44	2	-0.041	-57		1818	326	343	0		Sum of 3 scan	
✓	✓	(R)LENGDTKRG	100%	481	470	00	2		593.30	1,194.59	2	0.004	7.2		1218	442	450	0		Scan 4796 (RT)	

ペプチドの情報を表示

Protein Sequence: Similar Proteins: Spectrum: Spectrum/Model Error: Fragmentation Table  
gls47748 (100%) 99.502 3 Da  
Keratin, type I cytoskeletal 10 (Cytokeratin-10) (CK-10) (Keratin-10) (K10)  
6 exclusive unique peptides, 6 exclusive unique spectra, 8 total spectra, 51/593 amino acids (8% coverage)

```
MSVRYSSSKH YSSRSRSGGG GGGCGGGGG VSSLRISSK GSLGGGFSFG GFSGGFSFRG SSSGGCGFGS SGGYGGLGFF GGGSFHSGY SSSFGGSYGG SFGGGNFGGG  
SFGGGSFDD GFGGDFDGG FGGDFDGGD LLSGNERVTM GNLDNDLASY LGHVRALTEEL NYELKQK LKE WYKHHNSHG GERPRYSKYV GTIDDDLNQI LRLFTDANK  
LQIDNARLA ADFFLYEY EALRDSVEA DINGLRWLD SELTTRADLE MQLSLELE AYKHNHFE EMDLRAVSTG DVVEMHAA P GVDLTDLNLN MRGDTLELAK  
QNRKDAEAWF NEEKELTTE IDNIEQISS YKSEITELR NVDALEI ELO SQLALKOSLE ASLAETEGRY CVQLSOLQAD ISALEEQLQO IRAETECQNT EYQQLDIDKI  
RLKELKGLDTRW SLLLEGGSSG GGGGGGSSFG GQYGGGSSGQ GSSGGGSSGQ GSSGGGSSGQ GQYGGGSSGQ GNGGGSSSGQ HGGSSSSGGY GGSSSGGGGQ  
YGGSSSGGGS SSGGGYGGGS SSGGHSKSSS GSVGESSSKG PRY
```

表の各項目は以下の通りです。

Valid	Assigned	Sequence	Prob	Masc...	Masc...	Masc...	NTT	Modifications	Obs
✓	✓	(R)ALEESNYELEGK(I)	100%	67.8	46.8	0.0	2		6
✓	✓	(R)ALEESNYELEGK(I)	100%	66.2	46.8	0.0	2		6
✓	✓	(R)VLDELTLTK(A)	100%	63.2	47.3	0.0	2		5
✓	✓	(R)SQYEQLAEQNR(K)	100%	67.5	46.8	0.0	2		6
✓	✓	(R)SQYEQLAEQNR(K)	100%	57.9	46.8	0.0	2		6
✓	✓	(R)SQYEQLAEQNRK(D)	96%	42.0	46.7	0.0	2		7
✓	✓	(K)DAEAWFNEK(S)	97%	48.2	46.6	0.0	2		5
✓	✓	(R)LENEIQTYR(S)	97%	48.1	47.0	0.0	2		5

#### •Valid

チェックが入っているデータは protein probability の計算に使用されます。取り込み時のデフォルトでは peptide probability の threshold を満たすすべてのデータにチェックが入っています。

#### •Weight/Assigned

単純なグループ化が適用されている場合は該当ペプチドがユニークだと緑のチェック、共有ペプチドだと赤の十字で表示されます。Clustering (類似タンパク質のグループ化)が適用されている時には該当データの weight の数値を表します。ユニークなら 1,シェアならシェア状況に応じて数値が小さくなります。

#### •Sequence

ペプチド配列を表します。前と後ろの()に囲われた部分はペプチド直前あるいは直後のアミノ酸を表します。

#### •Prob.

peptide probability. Prefiltered mode で取り込まれた場合は表示されません。

#### •Search engine scores

検索エンジンごとに表示内容が異なります。

- **SEQUEST** : Xcorr と DeltaCn
- **Mascot** : Ion score、Identity score、Delta Ion Score
- **X! Tandem** : Expect 値(log)

#### •NTT

missed cleavage あるいは Number of Tryptic Termini (同じ意味ですが検索エンジンなどで用語が異なります)。

#### •Modifications

修飾情報

Observed	Actual M...	Char...	Delta...	Delta...	Rete...	Intensity	TIC	Start	Stop	# Ot...	Other Pr...	Spectrum ID
691.30	1,380.59	2	-0.049	-36			27770	166	177	0		Sum of 3 sca
691.31	1,380.60	2	-0.043	-31			4081	166	177	0		Sum of 3 sca
516.30	1,090.59	2	-0.0062	-6.0			4651	258	266	0		Sum of 3 sca
683.31	1,364.61	2	-0.025	-18			3240	323	333	0		Sum of 3 sca
683.30	1,364.59	2	-0.046	-33			16660	323	333	0		Sum of 3 sca
747.32	1,492.62	2	-0.11	-73			13510	323	334	0		Sum of 3 sca
555.23	1,108.44	2	-0.041	-37			1618	335	343	0		Sum of 3 sca
583.30	1,164.59	2	0.0084	7.2			1218	442	450	0		Scan 4796 (r

- **Observed**

スペクトルデータ側のペプチドの  $m/z$

- **Actual Mass**

ペプチドの質量

- **Charge**

ペプチドの電荷

- **Delta Da**

ペプチド質量の差の Da、実測値 - 理論値

- **Delta ppm**

ペプチド質量の差の Da、(実測値 - 理論値) / (実測値)

- **Retention Time**

LC の保持時間 (秒)。データによっては表示されない

- **Intensity**

ペプチドの Precursor スペクトルでの intensity (面積)。データによっては表示されない。

- **TIC**

MS/MS 各ピークの intensity の和。

- **Start**

ペプチド先頭部の、タンパク質全長における位置 (残基番号)。

- **Stop**

ペプチド末端部の、タンパク質全長における位置 (残基番号)。

- **# Other Proteins**

シェアペプチドの場合、シェアされているタンパク質の数。

- **Other Proteins**

シェアされているタンパク質の Accession。

- **Spectrum ID**

スペクトルの名称。

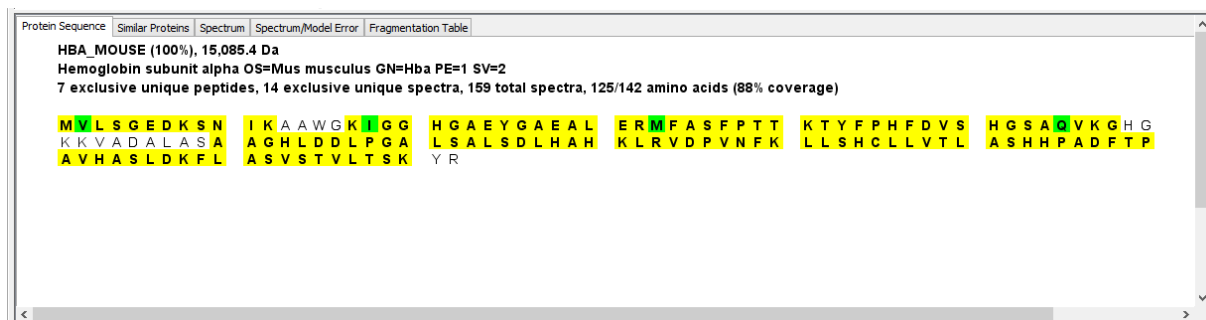
## 6-4. Spectrum pane :タンパク質/スペクトル 関連図

Spectrum pane (英文マニュアルでは Protein Sequence pane)では、proteins pane や peptides pane で選択しているタンパク質やスペクトルに関連する図を表示させることができます。以下5つのタブから構成されています。

- Protein Sequence tab
- Similar Proteins tab
- Spectrum tab
- Spectrum/Model Error tab
- Fragmentation Table tab

以下、各タブの画面について説明しています。

### 6-4-1. Protein Sequence tab



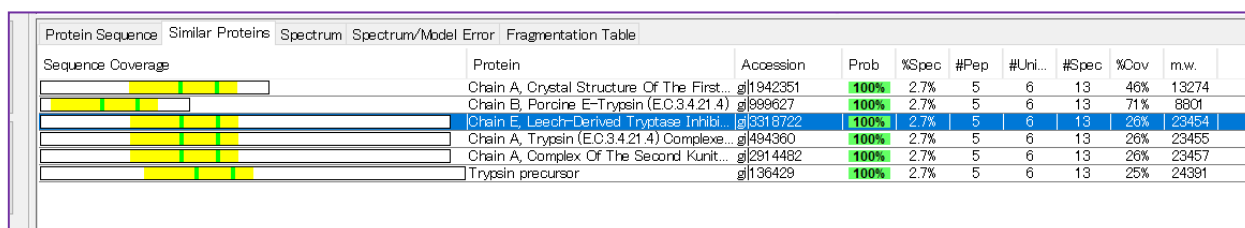
Protein Sequence tab では、選択しているタンパク質の全長に対して、アサインされたペプチドがどこに位置するのかを表しています。黄色い部分はペプチドがマッチしている箇所、緑のアミノ酸は修飾を受けていることを表します。また配列の上部にはタンパク質の Accession や Description、質量、アサインされたペプチド/スペクトル 数、coverage (%)なども表示されます。

またタブ内で右クリックを選択する事で、画像として保存したり関連情報をクリップボードにコピーしたりする事もできます。

[次頁に続きます]

## 6-4-2. Similar Proteins tab

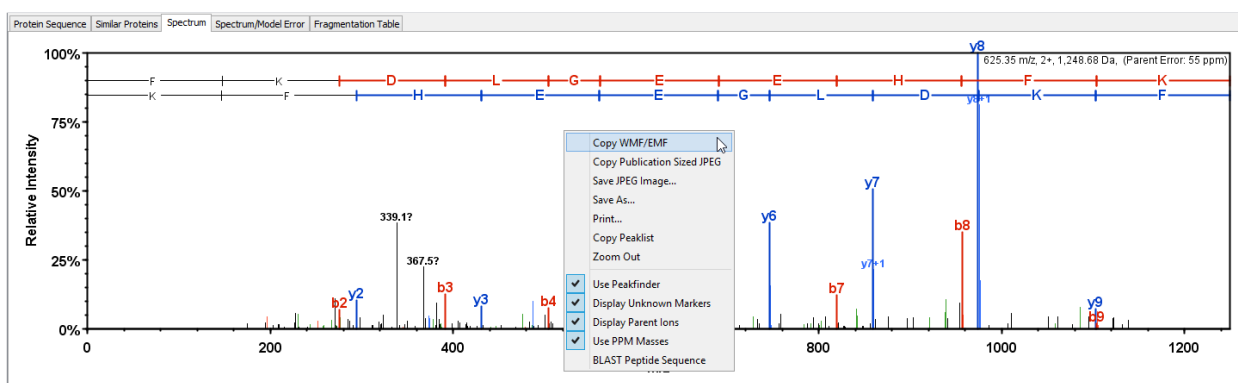
Similar Proteins tab では、選択しているタンパク質と同様のペプチドマッチをしているタンパク質 (グループ) について、その類似度を確認する事ができます (下図)。グループ内のタンパク質すべてについて、左上の Proteins pane と基本的に同じ情報が表示されます。グループ内のタンパク質において、ペプチドの重なり具合などを確認する事ができます。



Protein	Accession	Prob	%Spec	#Pep	#Uni...	#Spec	%Cov	m.w.
Chain A, Crystal Structure Of The First...	gi1942351	100%	2.7%	5	6	13	46%	13274
Chain B, Porcine E-Trypsin (E.C.3.4.21.4)	gi398627	100%	2.7%	5	6	13	71%	8801
Chain E, Leech-Derived Trypsin Inhibi...	gi3318722	100%	2.7%	5	6	13	26%	23454
Chain A, Trypsin (E.C.3.4.21.4) Comple...	gi494360	100%	2.7%	5	6	13	26%	23455
Chain A, Complex Of The Second Kunit...	gi2914482	100%	2.7%	5	6	13	26%	23457
Trypsin precursor	gi136429	100%	2.7%	5	6	13	25%	24381

## 6-4-3. Spectrum tab

Spectrum tab では、画面右上の Peptides pane で選択中のスペクトルデータについて、スペクトルベースで理論値とのマッチング状況を確認する事ができます (下図)。



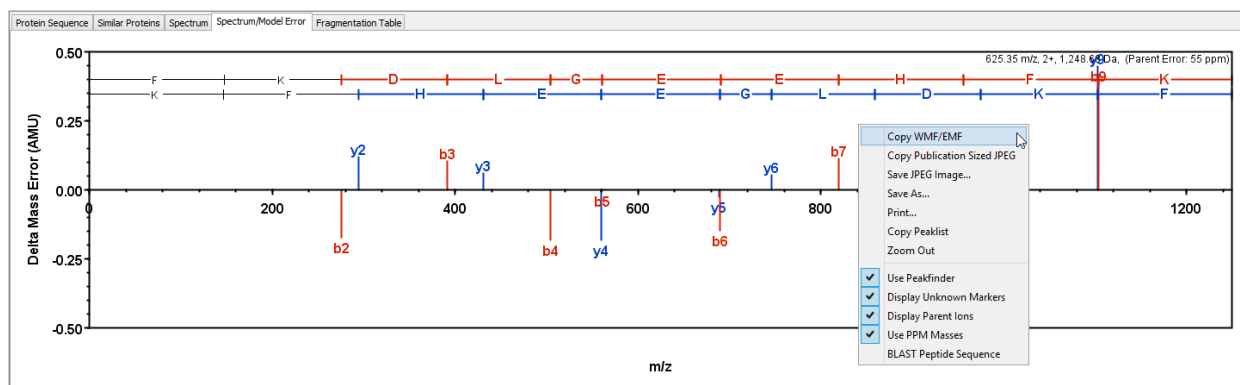
スペクトルのグラフはインタラクティブに操作/表示 できます。ドラッグ&ドロップで特定領域を拡大したり、拡大後にシングルクリックする事で元の表示に戻したりする事ができます。またタブ内で右クリックを選択する事で、画像として保存したりピークリスト情報をクリップボードにコピーしたり、配列を BLAST 検索したりする事もできます。

[次頁に続きます]



## 6-4-4. Spectrum/Model error tab

Spectrum/Model Error tab では、右上の Peptides pane で選択しているスペクトルデータについて、理論値と実測値との誤差を確認することができます(下図)。縦軸が誤差、横軸はフラグメントの質量(Da)です。



またタブ内で右クリックを選択する事で、画像として保存したりピークリスト情報をクリップボードにコピーしたり、配列を BLAST 検索したりする事もできます。

## 6-4-5. Fragment table tab

Fragment table tab では、右上の Peptides pane で選択しているスペクトルデータについて、理論値ベースで理論値とのマッチング状況を確認することができます(下図)。色が塗られているところがマッチした箇所です。「+2H」は 2 価、「-NH3」「-H2O」は脱アミノ/脱水を表します。またタブ内で右クリックを選択する事で、画像として保存したりデータをクリップボードにコピーしたり、CSV ファイルに出力することができます。

Protein Sequence											Similar Proteins	Spectrum	Spectrum/Model Error	Fragmentation Table
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y				
1	114.1	57.5			I	2,284.2	1,142.6	2,267.1	2,266.2	20				
2	227.2	114.1			I	2,171.1	1,086.0	2,154.1	2,153.1	19				
3	328.2	164.6		310.2	T	2,058.0	1,029.5	2,041.0	2,040.0	18				
4	465.3	233.1		447.3	H	1,956.9	979.0	1,939.9	1,938.9	17				
5	562.3	281.7		544.3	P	1,819.9	910.4	1,802.9	1,801.9	16				
6	676.4	338.7	659.4	658.4	N	1,722.8	861.9	1,705.8	1,704.8	15				
7	823.4	412.2	806.4	805.4	F	1,608.8	804.9	1,591.8	1,590.8	14				
8	938.5	469.7	921.4	920.5	NH1	1,461.7	731.4	1,444.7	1,443.7	13				
9	995.5	498.3	978.5	977.5	G	1,346.7	673.9	1,329.7	1,328.7	12				
10	1,109.5	555.3	1,092.5	1,091.5	N	1,289.7	645.3	1,272.7	1,271.7	11				
11	1,210.6	605.8	1,193.6	1,192.6	T	1,175.6	588.3	1,158.6	1,157.6	10				
12	1,323.7	662.3	1,306.6	1,305.7	L	1,074.6	537.8	1,057.6	1,056.6	9				
13	1,438.7	719.9	1,421.7	1,420.7	D	961.5	481.3	944.5	943.5	8				
14	1,552.7	776.9	1,535.7	1,534.7	N	846.5	423.7	829.4	828.5	7				
15	1,667.8	834.4	1,650.7	1,649.8	D	732.4	366.7	715.4	714.4	6				
16	1,780.9	890.9	1,763.8	1,762.8	I	617.4	309.2	600.4	600.4	5				
17	1,911.9	956.4	1,894.9	1,893.9	M	504.3	252.7	487.3	487.3	4				
18	2,025.0	1,013.0	2,007.9	2,007.0	L	373.3	187.1	356.3	356.3	3				
19	2,138.1	1,069.5	2,121.0	2,120.0	I	260.2	130.6	243.2	243.2	2				
20	2,284.2	1,142.6	2,267.1	2,266.2	K	147.1	74.1	130.1	130.1	1				

## 7. Grouping, Clustering と Similarity View

Scaffold では各種機能を持つ View があり、画面左側にそれらの View を切り替えるためのスイッチがあります。7 章では「**Similarity**」View について説明しています。また Similarity と関連がある内容として、Scaffold にて Group 化、Clustering 化する際のルールについても併せて説明しています。

### 7-1. Scaffold での類似タンパク質の扱い

質量分析データベースのプロテオミクスの解析ではスペクトルデータを元にペプチド配列を同定します。そしてそのペプチドがどのタンパク質に含まれる配列と同じかという情報を元に同定タンパク質をリストアップします。検索対象のデータベース中に、同定ペプチドをシェアするタンパク質が複数存在する事は頻繁に起こります。

ヒットしたペプチドの組み合わせが全く同じ、mascot で「**same-set**」と呼んでいる組み合わせを、Scaffold では「**protein group**」と呼んでいます。Samples リストの中で Accession の表示の後ろに( ) がありその中に + 数字、と表示されているものがこれに該当します(下図)。

Cytochrome b-c1 complex subunit 8 OS=Mus musculus GN=U...	OCRB_MOUSE	10 kDa	LK3 transgenic mice	●	●
40S ribosomal protein S29 OS=Bos taurus GN=RPS29 PE=3 SV...	RS29_BOVIN (+6)	7 kDa	Bos Taurus	○	○
Hornerin OS=Homo sapiens GN=HRNR PE=1 SV=2	HORN_HUMAN	282 kDa	Homo sapiens	●	●

same-set より下位、ヒットしたペプチドにオリジナルのペプチドがなく、他により多くのペプチドがマッチしているタンパク質が存在するケースを MASCOT では「**sub-set**」と呼んでいます。Scaffold では sub-set のデータは Samples View の同定タンパク質リストに現れず、後述するように Similarity View の「No group」に一緒くたにされてしまいます。

一方シェアするペプチドを持ちながら、他のタンパク質にはアサインされていないユニークなペプチドをもつタンパク質もあります。MASCOT ではこれらタンパク質はすべて「**Family Protein**」としてまとめられます。取り込み時のグループ化アルゴリズムの選択にもよりますが、Scaffold ではユニークペプチドのみならずシェアペプチドの確からしさに基づいて、クラスター(Cluster)としてまとめられるケースと、リスト中の他のタンパク質にシェアペプチドが存在する事を示すにとどまるケースがあります。

以降 7 章では、グループ化や Clustering の手法について説明しています。

### 7-2. 表示内容の詳細 : summary 画面

Scaffold 4 以降では「Share peptide Grouping」というグループ化・クラスター化のアルゴリズムを適用する事ができます。Scaffold 3 以前で扱われていた内容については「Legacy Protein Grouping」と呼んでいます。

「Share peptide Grouping」の方法は、従来の方法で求められていた、「高い同定確率であるペプチドが

『ユニーク』に存在する」という条件適用にこだわると同定タンパク質リストから抜け落ちてしまうようなタンパク質を救済する事を第一の目的としたアルゴリズムです。

データ取り込み時に「Protein Grouping」というオプションで「Use protein cluster analysis」という選択肢を選ぶことでこのアルゴリズムが適用可能です(下図)。

**Load and Analyze Data**

Searched Database:  
uniprot\_sprot\_mouse\_20121129 FASTA Database (2)  
 Use non-default forward/decoy ratio: No Decoys  
Add New Database

X! Tandem:  
 Analyze with X! Tandem

Scoring System:  
 Use LFDR scoring (all instruments)  
 Use legacy PeptideProphet scoring (high mass accuracy)  
 Use legacy PeptideProphet scoring (standard)

Protein Grouping:  
 Use protein cluster analysis  
 Use standard experiment wide protein grouping  
 Use legacy independent sample protein grouping

Protein Annotations:  
 Don't annotate (No download required)  
 Fetch GO annotations remotely (UniProt, EPI, NCBI; ~20 mins every time)  
[Configure GO Source](#)

◀ Previous Load Data ▶ Done Cancel

他の選択肢を選んだ場合でも、データを取り込んだ後に変更することができます。

メニューの Experiment -> Edit Experiment にて現れるダイアログにて、ラジオボタン上の 選択肢にて「use protein cluster analysis」の項目を選び、「Apply」ボタンを押す事で変更可能です(右図)。

**Edit Experiment**

Experiment Description:

Protein Grouping:  
 Use protein cluster analysis  
 Use standard experiment wide protein grouping  
 Use legacy independent sample protein grouping

Help Apply Cancel

以降、このグループ化並びにクラスター化の アルゴリズムの説明のため、内容をさらに以下3つのパートにわけて説明します。

- 7-2-1. Protein Grouping (same-set)
- 7-2-2. Protein Paring (sub-set)
- 7-2-3. Protein Clustering (Family protein)

## 7-2-1. Protein Grouping (same-set)

ヒットしたペプチドの組み合わせが全く同じ、mascot で「**same-set**」と呼んでいる組み合わせを、 Scaffold では「**protein group**」と呼んでいます。前述のように、group に複数のタンパク質を含む場合、 Samples view にて Accession の後ろに()と+数字、と表示されます(下図)。

Starred?	Accession	Alternate ID	Molecular Wt	Protein Grp	MSG081_ME	MSG081_F
	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex su... NDUFA9_MOUSE	Ndufa9	48 kDa		13	6
	Arginine-tRNA ligase, cytoplasmic OS=Mus musculus OX=10090 GN=Calm1 PE... CALM1_MOUSE (+2)	Calm1	17 kDa		7	7
	Calmodulin-1 OS=Mus musculus OX=10090 GN=Calm1 PE... CALM1_MOUSE	Calm1	17 kDa		7	7
	Adenylyl cyclase-associated protein 1 OS=Mus musculus OX=10090 GN=Cap1 PE... CAP1_MOUSE	Cap1	52 kDa		10	9
	Elongation factor Tu, mitochondrial OS=Mus musculus OX=10090 GN=EFTU_MOUSE	Eftu	42 kDa		9	10
	Cluster of Calcium/calmodulin-dependent protein kinase... KCC2B_MOUSE [4]	Kcc2b	42 kDa		6	10

Accession Number のところをクリックすると、group に含まれるタンパク質が表示され、その中から選択する事でリストに表示される Accession を変更することができます(下左図)。また Samples View 下部の「Protein Information pane」でもグループに属するタンパク質を確認することができます(下右図)

The screenshot shows two parts of the software interface. On the left, a dropdown menu is open, showing a list of protein accessions: CALM1\_MOUSE, CALM1\_MOUSE, CALM2\_MOUSE, and CALM3\_MOUSE. The first two entries are highlighted in blue. On the right, the 'Protein Information' pane is visible. It has a 'Lookup Identifier In:' field set to 'NCBI (e.g. gi|1351907,ALBU,BCVIN,P02769)'. Below this, three buttons are displayed: CALM1\_MOUSE, CALM2\_MOUSE, and CALM3\_MOUSE, all of which are highlighted with a red box. The background shows a list of protein entries with checkboxes and accession numbers.

グループに属するタンパク質と、そのタンパク質に帰属するペプチドについて着目すると、ペプチドは以下の2種類に大別することができます。

- シェアペプチド ... 複数タンパク質にアサインされるペプチド
- ユニークなペプチド ... 単一のタンパク質にアサインされているペプチド

シェアペプチドについてはタンパク質によって配分比率が計算され、確率の計算などにはその配分比率が適用されます。配分比率自体はタンパク質の存在が確からしいほど高くなるような計算式になっていて、主にユニークなペプチドを使って計算します。

タンパク質 A における ペプチド p の配分比率 Weight  $W(p,A)$  は、以下の式から算出します。

$$W(p, A) = \frac{PE_{excl}(A)}{\sum_{All(B \supseteq p)} PE_{excl}(B)}$$

分子(numerator)である  $PE_{excl}(A)$  とは、Aにアサインされているすべてのユニークペプチド X の同定確率  $P_X$  を足し合わせたものです(下式)。

$$PE_{excl}(A) = \sum_{X \subset A} P_X$$

分母(denominator)が意味するところは標準化です。 $PE_{excl}(A)$ と同様ユニークペプチドの probability 和を、ペプチド p がアサインされているすべてのタンパク質でさらに足し合わせ、その数字で割ると いう標準化を行っています。すなわち、他のタンパク質に比べユニークペプチドの数が多く その probability が高いと配分比率も高くなるようになっています。計算された配分比率は Similarity View の画面内で表示されます(下図)。

Index	Peptide	Prob	Exclusive To	Valid	Cluster of Myoglobin (MYG_HORSE)						Cluste... No Group	
					MYG_HORSE (+1)	MYG_CASFI	MYG_GALCR (+2)	MYG_OCHPR	MYG_ORYAF	MYG_RABIT		MYG_GLOME (+2)
1	ADIAGHGQEVLR	100%		<input checked="" type="checkbox"/>	0.76	0.02		0.11		0.11		
2	ALELFR	33%		<input checked="" type="checkbox"/>	0.63	0.01	0.09	0.09	0.09		0.09	—
3	ETLEKFDKFKNLKSEDEMKGS...	100%	Myoglobin	<input checked="" type="checkbox"/>				1.00				
4	GDFGADAQGAMTK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							
5	GLSDGEWQQVNLVWVK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							
6	HGTVLTALGGILK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							

## 7-2-2. Protein Pairing(sub-set)

same-set のデータは group としてまとめられますが、ペプチドのアサイン状況が sub-set, 包含関係的に下位に位置する状況、の場合、group にはまとめられず Similarity View でのみ確認する事ができます。Similarity View の画面右側、「No Group」にまとめられています(次頁図)。

Histone H2A type 1-B/E OS=Homo sapiens GN=HIST1H2AB PE=1 SV=2

Index	Peptide	Prob	Exclusive To	Valid	Cluste... H2A1B_HUMAN (+14)	Cluste... H2A1D_HUMAN (+18)	Cluste... * H2AV_BOVIN (+21)	Cluste... * H2AZ_CANAL (+1)	No Group
1	ATIAGGGVIPHIHK	52%	Histone H2...	<input checked="" type="checkbox"/>			1.00		
2	HLQLAIR	80%		<input checked="" type="checkbox"/>	0.39	0.39	0.21	0.01	
3	HLQLAIRNDEELNK	100%		<input checked="" type="checkbox"/>	0.50	0.50			
4	NDEELNKLLGK	100%	Histone H2...	<input checked="" type="checkbox"/>			1.00		
5	NDEELNKLLGR	100%		<input checked="" type="checkbox"/>	1.00				
6	VTIAQGGVLPNIQAVLLPK	100%		<input checked="" type="checkbox"/>	0.50	0.50			

### 7-2-3. Protein Clustering (Family Proteins)

MASCOT では、同定基準を超えるシェアペプチドが1つでも存在すれば Family Protein としてまとめられます(下記条件 1)。一方 Scaffold の Protein Clustering のアルゴリズムではそれよりも厳しい条件(下記条件 2)があり、シェアペプチドが一定の基準を超えていなければなりません。

クラスター化のルールは以下の通りです

\* 英文マニュアルと記述の構成が少し異なりますのでご注意ください

1. シェアペプチドの probability 値の和が 95% 以上である
2. シェアペプチドの probability 値の和が、シェア並びにユニークペプチドの probability 値の和に対して 50% 以上である。
3. 1,2 の条件をすべて満たすタンパク質同士で Family が構成される

なお、Cluster 形成を検討する段階では Samples Filter の各種条件は検討されません。Cluster 形成後、filter が適用され、Samples 画面に表示されるかどうかが決まります。下記資料に、計算例も含めた説明がごさいます。

<https://proteomesoftware.zendesk.com/hc/en-us/articles/115001221723-Protein-Grouping-and-Clustering-in-Scaffold>

Cluster は、Samples 画面にて Row number の横に十字アイコンが表示され、名称も「Cluster of ~」と表示されています(下図)。Accession の後ろには、Cluster に属するタンパク質の数も[N]の形で表示されています。

294	<input checked="" type="checkbox"/>	Cambridge 1 OS=Mus musculus OX=10090 GN=Cambridge 1... CAP1_MOUSE
295	<input checked="" type="checkbox"/>	Adenyl cyclase-associated protein 1 OS=Mus musculus ... CAP1_MOUSE
296	<input checked="" type="checkbox"/>	Elongation factor Tu, mitochondrial OS=Mus musculus OX... EFTU_MOUSE
297	<input checked="" type="checkbox"/>	Cluster of Calcium/calmodulin-dependent protein kinase ... KCC2B_MOUSE [2]
298	<input checked="" type="checkbox"/>	Transport and Golgi organization protein 1 homolog OS=... TGO1_MOUSE
299	<input checked="" type="checkbox"/>	Cluster of Creatine kinase B-type OS=Mus musculus OX=... KCRB_MOUSE
300	<input checked="" type="checkbox"/>	Aldo-keto reductase family 1 member A1 OS=Mus muscul... AK1A1_MOUSE
301	<input checked="" type="checkbox"/>	Dynammin-1-like protein OS=Mus musculus OX=10090 GN=... DNM1L_MOUSE



Cluster の Accession number のところをクリックすると、Cluster で表示する代表タンパク質の Accession に関して選択する事ができます。

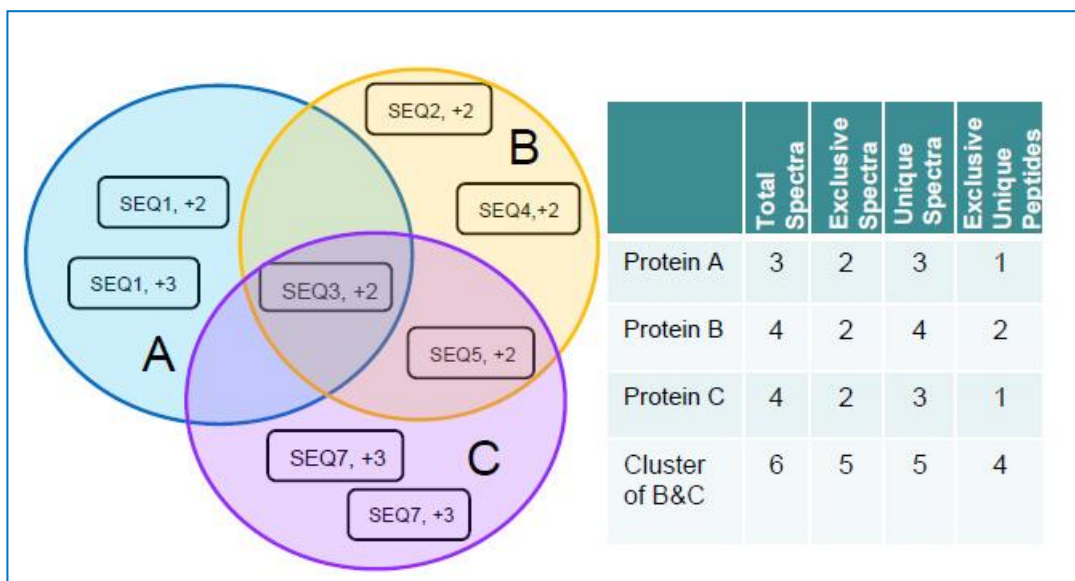
十字ボタンをクリックするとその下に展開し、Cluster に属するタンパク質が一覧で表示されます(下図)。

295	✓	☆ Adenylyl cyclase-associated protein 1 OS=Mus musculus ... CAP1_MOUSE
296	✓	Elongation factor 1 $\alpha$ , mitochondrial OS=Mus musculus OX=... EFTU_MOUSE
297	✓	Cluster of Calcium/calmodulin-dependent protein kinase ... KCC2B_MOUSE [2]
297.1	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2B_MOUSE
297.2	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2D_MOUSE
298	✓	☆ Transport and Golgi organization protein 1 homolog OS=... TGO1_MOUSE
299	✓	Cluster of Creatine kinase B-type OS=Mus musculus OX=... KCRB_MOUSE
300	✓	☆ Aldo-keto reductase family 1 member A1 OS=Mus muscul... AK1A1_MOUSE

メニューの View -> Show Entire Protein Clusters を選択すると、Samples 画面の各種 Filtering 条件を満たさないタンパク質がグレーアウトの形で表示されます(下図)。ここからも、Family の形成には Samples の Filtering 条件は関係ないことがわかります。

297	✓	Cluster of Calcium/calmodulin-dependent protein kinase ... KCC2B_MOUSE [4]
297.1	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2B_MOUSE
297.2	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2G_MOUSE
297.3	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2D_MOUSE
297.4	✓	☆ Calcium/calmodulin-dependent protein kinase type II ... KCC2A_MOUSE
298	✓	☆ Transport and Golgi organization protein 1 homolog OS=... TGO1_MOUSE
299	✓	Cluster of Creatine kinase B-type OS=Mus musculus OX=... KCRB_MOUSE [2]
300	✓	☆ Aldo-keto reductase family 1 member A1 OS=Mus muscul... AK1A1_MOUSE

Samples View などに表示される数字について、Cluster では個々のタンパク質とは内容が異なります。例えば下図をご覧ください。丸と A,B,C がタンパク質、丸の中の小さい四角の表示がスペクトルを表し、タンパク質 B と C がクラスターを構成しているとします。B の Total Spectra は 4、C も 4 ですが B と C のクラスターでは和の 8 ではなく 6 であることに注意してください。シェアされているペプチドはまとめて 1つとカウントされます。







ペプチド単位でデータを眺めた際、シェアされているタンパク質の中で最も probability の和が大きいタンパク質に帰属させます。値が同じタンパク質が2つ以上ある場合はそのすべてに帰属させます。下図では緑に塗られた箇所が帰属する位置です。

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Peptide	Exclusive To	Valid	B 157831969	B 109055949	B 11513851	B 23565563	B 73089799	B 191092898	B 13279011	B 30585149	B 146747386	B 1237572	B 169158861
2	AKWYPEVR		FALSE	9%	9%	9%			9%	9%	9%	9%	9%	
3	CVVVG DGAVGK		FALSE	28%		28%	28%		28%	28%	28%			
4	DDKDTIEK		TRUE	73%	73%	73%				73%	73%	73%	73%	73%
5	GSPQAIK	Chain A, Small G-Protein	TRUE	75%										
6	IISAMQTIKCVVVG DGAVGK		TRUE											
7	KLTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%	95%	95%	95%	95%	95%			95%	95%	
8	LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE							95%	95%			
9	LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%	95%	95%	95%	95%	95%			95%	95%	
10	LVPITYPQGLAMAK		TRUE											
11	TVFDEAIR		TRUE	95%	95%	95%	95%	95%	95%	95%	95%			95%
12	VDSKPVNLGLWDTAGQEDYDR		TRUE											92%
13				433%	358%	358%	285%	285%	285%	263%	263%	263%	263%	260%

続いて、緑に塗られたペプチドがないタンパク質をリストから除きます(下図)。

	B	C	D	E	F	G	H	I	J
1				Group 1	Group 2		Group 3		
2	Peptide	Exclusive To	Valid	B 157831969	B 13279011	B 30585149	B 169158861		
3	AKWYPEVR		FALSE	9%	9%	9%			
4	CVVVG DGAVGK		FALSE	28%	28%	28%			
5	DDKDTIEK		TRUE	73%	73%	73%	73%		
6	GSPQAIK	Chain A, Small G-Protein	TRUE	75%					
7	IISAMQTIKCVVVG DGAVGK		TRUE						
8	KLTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%					
9	LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%			
10	LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%					
11	LVPITYPQGLAMAK		TRUE						
12	TVFDEAIR		TRUE	95%	95%	95%	95%		
13	VDSKPVNLGLWDTAGQEDYDR		TRUE				92%		
14				433%	263%	263%	260%		

最後に、緑に塗られたペプチドの中で probability が95%未満のものしかないタンパク質をリストから除きます(下図)。

	A	B	C	D	E	F	G	H
1					Group 1	Group 2		
2	Index	Peptide	Exclusive To	Valid	B 157831969	B 13279011	B 30585149	
3	1	AKWYPEVR		FALSE	9%	9%	9%	
4	2	CVVVG DGAVGK		FALSE	28%	28%	28%	
5	3	DDKDTIEK		TRUE	73%	73%	73%	
6	4	GSPQAIK	Chain A, Small G-Protein	TRUE	75%			
7	5	IISAMQTIKCVVVG DGAVGK		TRUE				
8	6	KLTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%			
9	7	LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%	
10	8	LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%			
11	9	LVPITYPQGLAMAK		TRUE				
12	10	TVFDEAIR		TRUE	95%	95%	95%	
13	11	VDSKPVNLGLWDTAGQEDYDR		TRUE				
14					433%	263%	263%	

最終的に残ったものが Legacy proteins group でまとめられる「group」となります。

## 7-4. Samples View と Similarity View との関連について

類似(Group,Cluster)タンパク質の類似状況と、**Samples View** 並びに **Similarity View** でそれらがどのように表示されるかについて、改めて説明します。

group [same-set] のタンパク質は、**Samples view** で代表タンパク質1つにまとめられて表示されます(下図)。Accession Number の後ろに(+数字)の形で表示され、カーソルを合わせると画面下部の Protein Information pane に group に属する他のタンパク質の情報を確認することができます。

#	Visible?	Starred?	Protein Name	Accession Number	Alternate ID
1260	<input type="checkbox"/>	<input type="checkbox"/>	DENN domain-containing protein 4C OS=Mus musculus O...	DEN4C_MOUSE	Dennd4c
1261	<input type="checkbox"/>	<input type="checkbox"/>	Sorbitol dehydrogenase OS=Mus musculus OX=10090 GN=...	DHSO_MOUSE	Sord
1262	<input type="checkbox"/>	<input type="checkbox"/>	FAS-associated factor 1 OS=Mus musculus OX=10090 G...	FAF1_MOUSE	Faf1
1263	<input type="checkbox"/>	<input type="checkbox"/>	Histone H2A.V OS=Mus musculus OX=10090 GN=H2afv P...	H2AV_MOUSE (+1)	H2afv
1264	<input type="checkbox"/>	<input type="checkbox"/>	Hemoxygenase 2 OS=Mus musculus OX=10090 GN=Hmo...	HMOX2_MOUSE	Hmox2
1265	<input type="checkbox"/>	<input type="checkbox"/>	Importin-8 OS=Mus musculus OX=10090 GN=Ipo8 PE=1 S...	IPO8_MOUSE	Ipo8
1266	<input type="checkbox"/>	<input type="checkbox"/>	Protein NipSnap homolog 1 OS=Mus musculus OX=10090 ...	NIP51_MOUSE	Nipsnap1
1267	<input type="checkbox"/>	<input type="checkbox"/>	Nuclear pore complex protein Nup98-Nup96 OS=Mus mus...	NUP98_MOUSE	Nup98
1268	<input type="checkbox"/>	<input type="checkbox"/>	Adenylosuccinate synthetase isozyme 2 OS=Mus musculu...	PURA2_MOUSE	Adss
1269	<input type="checkbox"/>	<input type="checkbox"/>	Sideroflexin-3 OS=Mus musculus OX=10090 GN=Sfxn3 PE...	SFXN3_MOUSE	Sfxn3

MASCOTでいう sub-set のタンパク質は、samples view には全く表示されません。**Similarity View** の「No group」欄にまとめて表示されます(下図)。

Index	Peptide	Prob	Exclusive To	Valid	Cluste... (+14)	Cluste... (+18)	Cluste... (+21)	Cluste... (+1)	No Group
1	ATIAGGVIPHIHK	52%	Histone H2...	<input type="checkbox"/>			1.00		H2A1A_HUMAN (+25)
2	HLQLAIR	80%		<input type="checkbox"/>	0.39	0.39	0.21	0.01	H2A1_CANAL (+30)
3	HLQLAIRNDEELNK	100%		<input type="checkbox"/>	0.50	0.50			H2A1_CANAL (+30)
4	NDEELNKLLGK	100%	Histone H2...	<input type="checkbox"/>		1.00			H2A1_CANAL (+30)
5	NDEELNKLLGR	100%	Histone H2...	<input type="checkbox"/>	1.00				H2A1_CANAL (+30)
6	VTIAQGGVLPNIQAVLLPK	100%		<input type="checkbox"/>	0.50	0.50			H2A1_ASHGO (+40)

Family protein に属するような、シェアペプチドを持ちつつユニークなペプチドを持つ場合で、シェアペプチドの確からしさもある程度保証されている場合、**Samples View** には「Cluster」としてまとめられて表示されます(下図)。

295	✓	Adenylyl cyclase-associated protein 1 OS=Mus musculus ... CAP1_MOUSE
296	✓	Elongation factor Tu, mitochondrial OS=Mus musculus OX_FETU_MOUSE
297	✓	Cluster of Calcium/calmodulin-dependent protein kinase ... KCC2B_MOUSE [2]
297.1	✓	Calcium/calmodulin-dependent protein kinase type II ... KCC2B_MOUSE
297.2	✓	Calcium/calmodulin-dependent protein kinase type II ... KCC2D_MOUSE
298	✓	Transport and Golgi organization protein 1 homolog OS=...
299	+	Cluster of Creatine kinase B-type OS=Mus musculus OX=...
300	✓	Aldo-keto reductase family 1 member A1 OS=Mus muscul...

一方 シェアペプチドの確度が基準を満たさない場合や、古いバージョンの Grouping アルゴリズムを使った場合、Samples View において特にタンパク質をまとめた表示は行いませんが、**Similarity View** にてその重複度合いをチェックする事ができます。

Cluster も含め、シェアペプチドをもつタンパク質は **Samples View** にて「Protein Grouping Ambiguity」列の星印がついています(下図)。

Display Options: Total Unique Peptide Count		Req Mods: No Filter	Search:						
<b>Probability Legend:</b> <span style="background-color: #90EE90; border: 1px solid black; display: inline-block; width: 15px; height: 10px;"></span> over 95% <span style="background-color: #FFD700; border: 1px solid black; display: inline-block; width: 15px; height: 10px;"></span> 80% to 94% <span style="background-color: #FFA500; border: 1px solid black; display: inline-block; width: 15px; height: 10px;"></span> 50% to 79% <span style="background-color: #FF6347; border: 1px solid black; display: inline-block; width: 15px; height: 10px;"></span> 20% to 49% <span style="background-color: #FFB6C1; border: 1px solid black; display: inline-block; width: 15px; height: 10px;"></span> 0% to 19%									
#	Visible?	Starred?	Bio View: 2793 Proteins in 2524 Clusters With 12 Decoys and 6 Filtered Out	Accession Number	Alternate ID	Molecular Weight	Protein Grouping Ambiguity	MSG911_... MSG911_MariaRohm	MSG969b_... MSG969b_Fashcroft
1	✓	✓	Cytoplasmic dynein 1 heavy chain 1 OS=Mus musculus O...	DYHC1_MOUSE	Dync1h1	532 kD	93	71	
2	✓	✓	ATP-citrate synthase OS=Mus musculus OX=10090 GN=A...	ACLY_MOUSE	Acly	120 kD	48	35	
3	✓	✓	Cluster of Tubulin beta-5 chain OS=Mus musculus OX=10...	TBB5_MOUSE [...	Tubb5	50 kD	43	36	
4	✓	✓	Clathrin heavy chain 1 OS=Mus musculus OX=10090 GN=...	CLH1_MOUSE	Cltc	192 kD	55	43	
5	+	✓	Cluster of Ras-related protein Rab-14 OS=Mus musculus...	RAB14_MOUSE...	Rab14	24 kD	68	58	
6	✓	✓	Endoplasmic reticulum chaperone Hsp90b1 OS=Mus musculus OX=10090 GN=Hsp90b1 P...	ENPL_MOUSE	Hsp90b1	92 kD	37	31	
7	✓	✓	Endoplasmic reticulum chaperone BIP OS=Mus musculus ...	BIP_MOUSE	Hspa5	72 kD	42	33	
8	✓	✓	Spectrin alpha chain, non-erythrocytic 1 OS=Mus muscul...	SPTN1_MOUSE	Sptant	285 kD	60	50	

星印をダブルクリックすると 該当タンパク質に関する **Similarity view** に切り替わります(下図)。

Cluster of Tubulin beta-5 chain OS=Mus musculus OX=10090 GN=Tubb5 PE=1 SV=1 (TBB5_MOUSE)												
Index	Peptide	Prob	Exclusive	Valid	TBB5_MOUSE	* TBB1_MOUSE	TBB2A_MOUSE	TBB2E_MOUSE	TBB3_MOUSE	TBB4A_MOUSE	TBB4E_MOUSE	TBB6_MOUSE
1	ALVDLEPGTMDSVR	100%		✓	0.36		0.07	0.07	0.50			
2	ALTVPELTQQMFDAK	100%		✓					0.47	0.40	0.07	0.07
3	ALTVPELTQQMFDSK	100%		✓			0.50	0.50				
4	ALTVPELTQQVFDAK	100%	Tubulin bet...	✓	1.00							
5	AVLDLEPGTMDSVR	100%		✓						0.85	0.15	
6	EVDEQMLAIQSK	100%	Tubulin bet...	✓				1.00				
7	EVDEQMLNQGSK	100%		✓	0.62		0.13	0.13			0.13	
8	EVDEQMLSVQSK	100%	Tubulin bet...	✓						1.00		
9	FPGQLNADLR	100%		✓	0.22	0.02	0.05	0.04	0.31	0.26	0.05	0.05
10	FPGQLNADLRK	95%		✓	0.22	0.02	0.05	0.04	0.31	0.26	0.05	0.05

Similarity view で内容をチェックしたことのタンパク質は Samples View で星の色が緑に代わります(下図)。切り替わる前の星の色は赤色です。

Display Options: Total Unique Peptide Count    Req Mods: No Filter    Search:

Probability Legend:

- over 95% (Green)
- 80% to 94% (Yellow)
- 50% to 79% (Orange)
- 20% to 49% (Red)
- 0% to 19% (Light Red)

Bio View:  
2793 Proteins in 2524 Clusters  
With 12 Decoys and 6 Filtered Out

#	Visible?	Starred?	Accession Number	Alternate ID	Molecular Weight	Protein Grouping Ambiguity	MSG911_... MSG011_MariaRehm	MSG969b_... MSG069b_FAshcroft
1	<input type="checkbox"/>	<input type="checkbox"/>	DYHC1_MOUSE	Dync1h1	532 kDa	93	71	
2	<input type="checkbox"/>	<input type="checkbox"/>	ACLY_MOUSE	Acly	120 kDa	48	35	
3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Tubb5_MOUSE	Tubb5	50 kDa	43	36	
4	<input type="checkbox"/>	<input type="checkbox"/>	CLH1_MOUSE	Cltc	192 kDa	55	43	
5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	RAB14_MOUSE	Rab14	24 kDa	60	50	
6	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	ENPL_MOUSE	Hsp90b1	92 kDa	21	19	
7	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	SPTM1_MOUSE	Hspa5	72 kDa	20	18	
8	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	MYH9_MOUSE	Sptan1	285 kDa	52	51	
9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	ACTB_MOUSE	Actb	226 kDa	21	19	
10	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Tuba1a_MOUSE	Tuba1a	42 kDa	20	18	
11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	HSP70_MOUSE	Hsp70	50 kDa	20	18	
12	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			71 kDa	36	30	

Protein Information:    Gene Ontology:    Sample Information:

## 7-5. Similarity View 概要

「Similarity」View では、特定タンパク質とマッチング内容が類似する別タンパク質について、ペプチドの重なり具合などをチェックすることができます(下図)。

Scaffold Q+S - Similarity - tutorial\_2

File Edit View Experiment Export Quant Window Help

Protein Threshold: 99.0%    Min # Peptides: 2    Peptide Threshold: 95%

Serum albumin precursor (Allergen Bos d 6)

Cluster of Serum albumin precursor (Allergen ... No Group)

Protein Pulldown list

Index	Protein	Similarity	ALBU_BOVIN (+)	ALBU_CANFA	ALBU_PIG	ALBU_HGT	ALBU_RAT	FETA_HORSE	ALBU_SHEEP	ALBU_MKINU	ALBU_MOUSE	ALBU_FELCA	ALBU_HUMAN	ALBU_HORSE
3	APFETETK	100%	1.00											
4	AIPENLPLTADFAEDKDKVCK	100%	1.00											
5	CCADDKKEACFAVEGPK	100%	1.00											
6	LCCTESLVNR	100%	0.98		0.02									
7	CDNQDTISSK	82%	1.00											
8	DAPLGSFLYR	100%	1.00											
9	DADPENLPLTADFAEDKDKVCK	100%	1.00											
10	DDRPHACYTVFDK	100%	1.00											
11	DLGEEHFK	100%	1.00											
12	ECCDPLLEK	100%	0.94		0.02	0.04								
13	ECCQKLLKCADRR	100%	0.89	0.00	0.02	0.04								
14	ECCQKLLKCADRADLAK	100%	0.84	0.00	0.02	0.04								
15	ENFVAFVDK	29%	1.00											
16	FTYGFK	26%	1.00											

Similarity table

Identifications

Sequence	Prob	Masc...	Masc...	Masc...	XI Ta...	NTT	Modifications	Observed	Actual Mass	Charge	Delta ...	Delta ...	Rate...	Intensity	TIC	Spectrum ID	Category	Bio Sample	MS/MS Sa...
(Q)CCTESLVNR(E)	100%	43.0	43.9	29.8	2.00	2	Carbamidomethyl...	569.77	1,157.52	2	0.033	29		2.195E7	ScaffoldIDNumbe...	AA	c1	control_071...	
(Q)CCTESLVNR(?)	73%	40.8	42.1	29.6		2	Carbamidomethyl...	570.06	1,138.11	2	0.62	550		2.414E7	ScaffoldIDNumbe...	AA	c2	control_071...	
(Q)CCTESLVNR(E)	100%				2.51	2	Ammonia-loss (-1...	1,122.71	1,121.70	1	1.2	1100		2.369E7	ScaffoldIDNumbe...	AA	c2	control_071...	
(Q)CCTESLVNR(E)	97%	33.5	43.8	12.4		2	Carbamidomethyl...	569.73	1,137.45	2	-0.037	-33		2.203E7	ScaffoldIDNumbe...	BB	c3	control_071...	
(Q)CCTESLVNR(E)	100%	39.0	43.8	27.8	2.70	2	Carbamidomethyl...	569.73	1,137.45	2	-0.037	-33		2.08E7	ScaffoldIDNumbe...	BB	c4	control_071...	
(Q)CCTESLVNR(E)	100%	44.9	43.3	29.5	1.92	2	Carbamidomethyl...	569.70	1,137.38	2	-0.11	-94		1.804E7	ScaffoldIDNumbe...	CC	c5	control_071...	

Identification pane

21 Proteins at 99.0% Minimum  
2 Min # Peptides  
0.0% Decoy FDR  
1359 Spectra at 95.0% Minimum  
0.00% Decoy FDR

「**Similarity**」View 画面は主に3つのパーツに分かれています。

画面上部にある「**Protein Pulldown list**」は、現在選択中のタンパク質を表します。

「**Similarity table**」は類似タンパク質とのペプチドの重複度合いを確認する事ができます。

「**Identification pane**」では、「Similarity table」にて選択しているペプチドに関して関連情報を表示する欄で、「Proteins View」画面の右上「peptides pane」や、下部「Spectrum pane」と同じ内容です。項目の詳細については、Protein View 内画面の各説明「**6-3 peptide pane**」「**6-4.Spectrum pane**」をご参照ください。

本資料では、メインとなる表示「Similarity table」についてのみ説明しています。

**Similarity** view の画面上段から中段にかけて表示されている画面が「**Similarity table**」です(下図)。類似タンパク質について、どのペプチドがシェアされているのかを確認する際に利用します。

Index	Peptide	Prob	Exclusive To	valid	MYG_HORSE (+1)	MYG_CASFI	MYG_GALCR (+2)	MYG_OCHPR	MYG_ORYAF	MYG_RABIT	MYG_GLOME (+2)	MYG_ELEMA (+2)
1	ADIAGHGQEVLR	100%		<input checked="" type="checkbox"/>	0.76	0.02		0.11		0.11		
2	ALELFR	33%		<input checked="" type="checkbox"/>	0.63	0.01	0.09	0.09	0.09		0.09	
3	ETLEKFDKFKNLKSEDEMKGS...	100%	Myoglobin	<input checked="" type="checkbox"/>				1.00				
4	GDFGADAQGAMTK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							
5	GLSDGEWQQVLNVWGK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							
6	HGTWLTALGGILK	100%	Myoglobin	<input checked="" type="checkbox"/>	1.00							

各列の項目はそれぞれ以下の通りです。

- **Index** - ペプチドの通し番号
- **Peptide** - ペプチド配列
- **Prob.** - ペプチドの同定確率。prefiltered mode で取り込んだ場合、すべて「99%」と表示
- **Exclusive to** - ユニークペプチドかどうか。ユニークな場合、タンパク質名が表示
- **Valid** - ペプチドを解析に利用しているかどうか。ユーザーが自らクリックし、使用するかどうかを経能する事もできます。

その次の列からはタンパク質になります。タンパク質はユニークなタンパク質毎に、あるいはクラスターごとに色分けで表示されています。”Exclusive to” 列にタンパク質名の表示と着色があり、最上列のタンパク質の色と連動しています。

ペプチドとタンパク質が交わる各セルに表示される数字は、新しいグルーピングアルゴリズムが採用されている時は Weight の数値が表示されます。Legacy protein grouping の時は Probability が表示され

ます(右下図)。ペプチドがアサインされているタンパク質のセルのみ数字が表示されます。

タンパク質の Accession の下に\*がある場合、ペプチドの同定基準を満たしユニークなペプチドが存在するものの、タンパク質の同定基準を満たせず現段階の Samples のリストには表示されていないタンパク質であることを示しています (\*右図の ALBU\_PIG など)

Index	Peptide	Exclusive To	Valid	Serum albumi...	Seru...	Seru...	Seru...	No Group
				ALBU_BOVIN	ALBU_CONTR	* ALBU_PIG	ALBU_RABIT	* ALBU_RAT
1	AADKDNCFATEGPNLVARSK...	Serum albu...	<input checked="" type="checkbox"/>				95%	
2	AATITK		<input checked="" type="checkbox"/>					
3	AEFVEVTK	Serum albu...	<input checked="" type="checkbox"/>	100%	100%			
4	AIPENLPPLTADF AEDKDVCK	Serum albu...	<input checked="" type="checkbox"/>	100%	100%			
5	CCAADDKEACFAVEGPK	Serum albu...	<input checked="" type="checkbox"/>	100%	100%			
6	CCTESLVNR		<input checked="" type="checkbox"/>	100%	100%	(100%)		(100%)
7	CDNQDTISSK	Serum albu...	<input checked="" type="checkbox"/>	62%	62%			
8	DAFLGSFLYEYSR	Serum albu...	<input checked="" type="checkbox"/>	100%	100%			
9	DAIPFNI PPI TADFAFDKDVCK	Serum albu...	<input checked="" type="checkbox"/>	100%	100%			

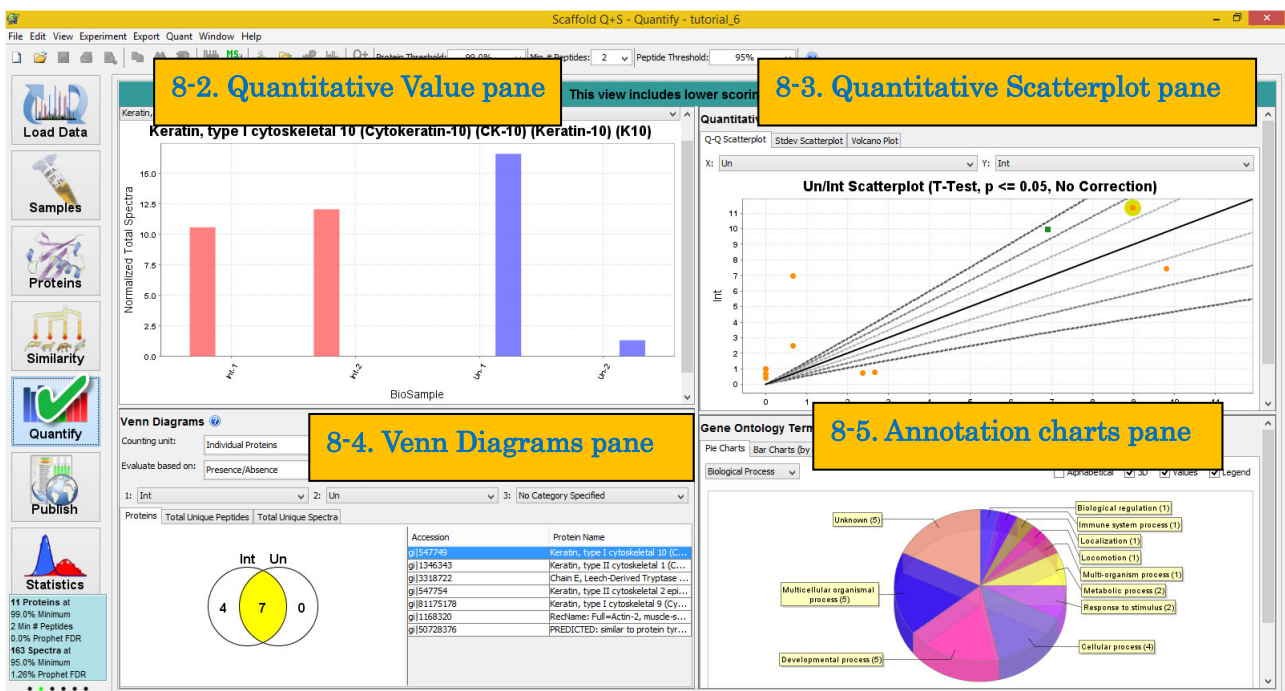


## 8. Quantify View

### 8-1. Quantify View : 定量指標を基にしたグラフや GO,ベン図を表示

Scaffold では各種機能を持つ View(画面)があり、左側にそれらの View を切り替えるためのスイッチがあります。定量値に関するグラフを表示したり、同定結果のベン図や Gene Ontology に関する図・グラフを表示したりする事ができるのが「**Quantify**」View です(下図)。

Quantify View は大きく分けると4つのパートから構成されています。



#### • Quantitative Value pane (左上)

選択中のタンパク質について、サンプル 別の定量に関連する値(Quantitative value)の比較ができる棒グラフが表示されます。

#### • Quantitative Scatterplots pane (右上)

サンプル間で各タンパク質の定量値に関する散布図や volcano plot が表示されます。

#### • Venn Diagrams pane (左下)

同定タンパク質/ペプチドに関してサンプル別に比較ができるベン図が表示されます。またベン図の各エリアに属するタンパク質/ペプチドを確認する事ができます。

#### • Annotation charts pane (右下)

Samples に表示されているタンパク質の Gene Ontology の項目情報や Pathway の項目情報について

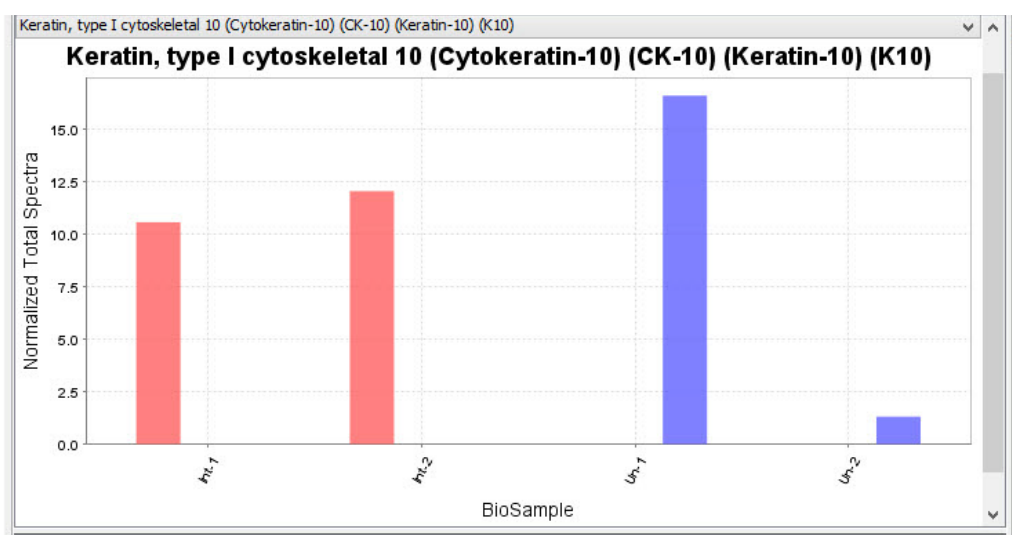
まとめた円グラフなどが表示されます。

Quantify 内の表示内容は連動しており、例えば Quantitative Value pane のプルダウンでタンパク質を選択すると Quantitative Scatterplot pane や Venn Diagrams pane で該当タンパク質が強調表示されます。

以降、各 pane について説明しています。

## 8-2. Quantitative Value pane

画面左上が Quantitative Value pane です。縦軸が Quantitative value で各サンプル別の棒グラフが表示されます。縦軸の値は Experiment -> Quantitative Analysis で選択している「Quantitative Method」の選択肢に連動しています。選択肢内容については、「11-2.ラベルフリーの定量方法」をご参照ください。



グラフ内で右クリックをすると、各種画像ファイルへのエクスポートや印刷、データのテキストコピーなどの選択肢が現れます。

## 8-3. Quantify Scatterplots pane

画面右上が Quantitative Scatterplots pane です。定量値について、サンプル間で比較して確認したい際に利用します。以下の3つのタブから構成されています。

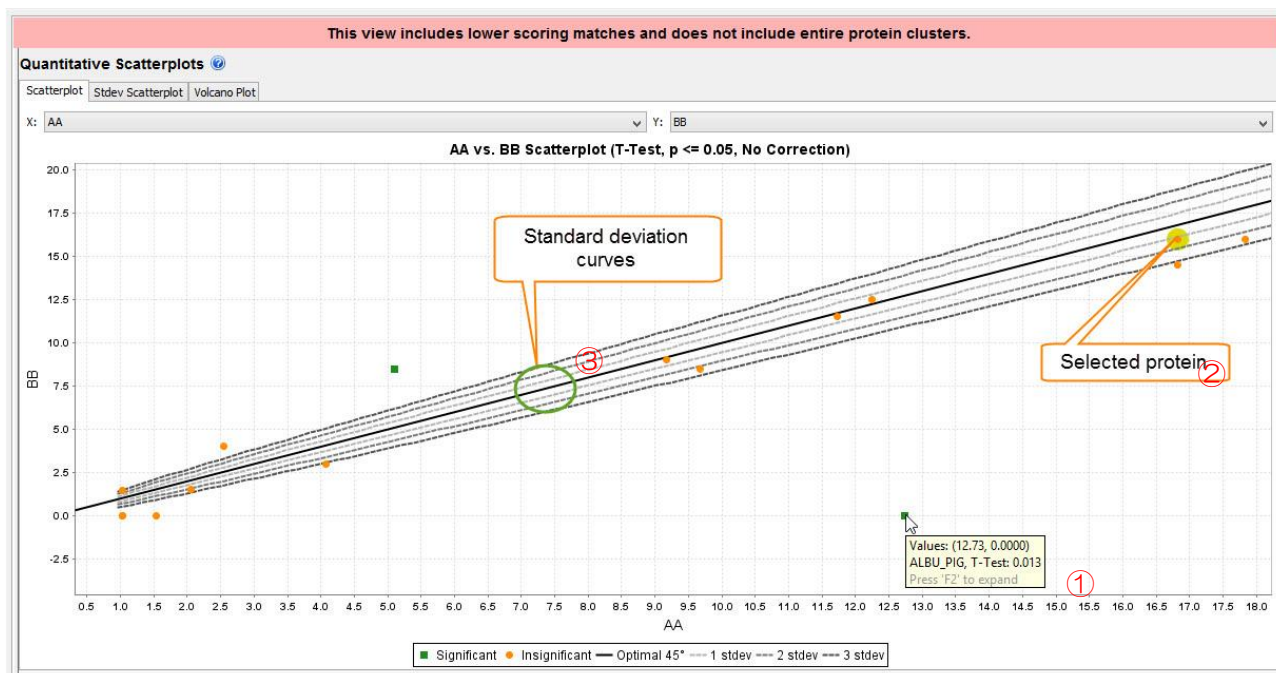
- 8-3-1. Scatterplot タブ
- 8-3-2. Stdev Scatterplot タブ
- 8-3-3. Volcano Plot タブ

以降、Quantify Scatterplots pane 内の各タブについて説明します。



### 8-3-1.Scatterplot タブ

X軸とY軸に、チェックしたいサンプルをプルダウンから選択します。それぞれの軸の値は Quantitative value を表し、各点はタンパク質に該当します(下図)。Quantitative Value pane と同じく、Quantitative value が具体的に何を示しているかは Experiment -> Quantitative Analysis で選択している「Quantitative Method」の選択肢に連動しています(同ダイアログ内の「Use Normalization」や「Minimum Value」選択とも連動しています)。



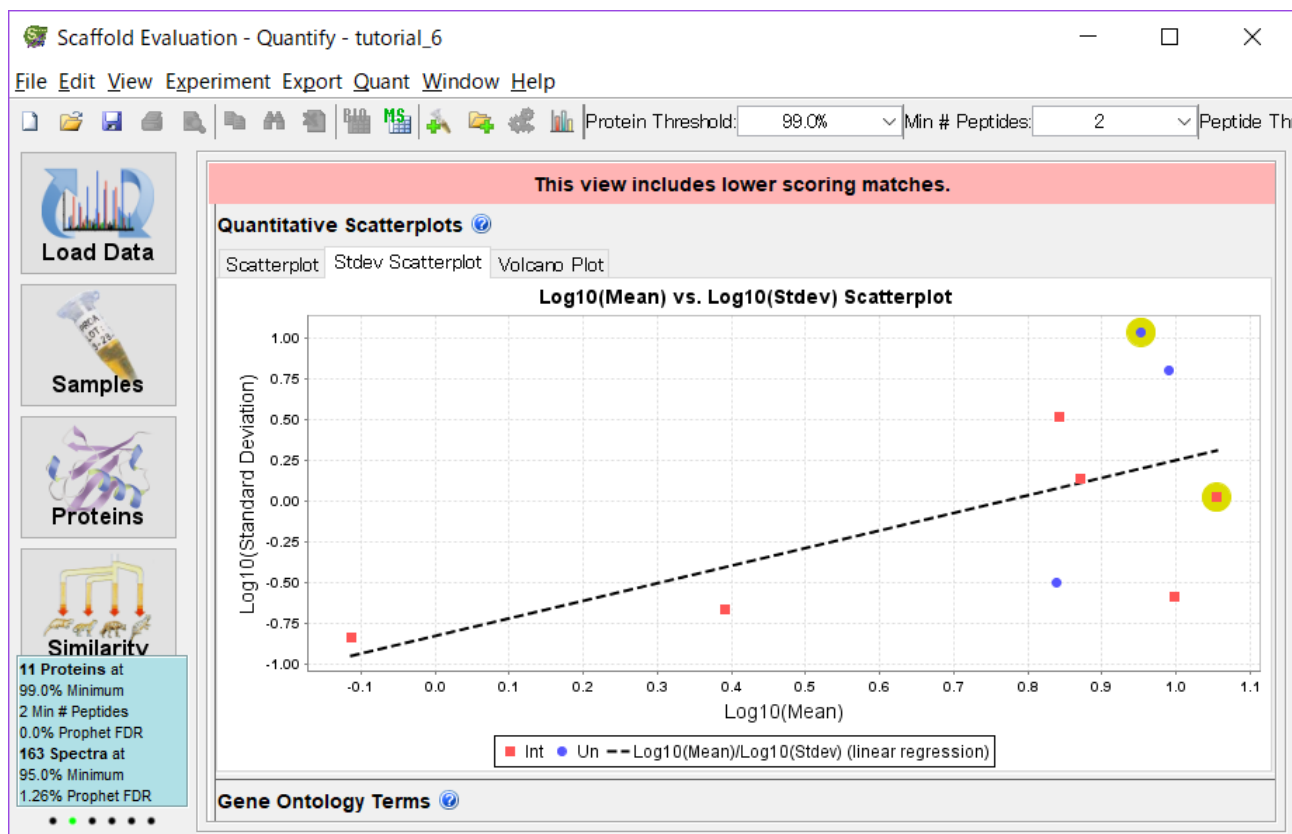
グラフ内の各点について、カーソルを合わせると縦軸と横軸の値、並びに Accession の情報が表示されます(上図内①)。またクリックするとタンパク質を選択したことになり(上図内②)、そのタンパク質の情報について左上の Quantitative Value pane での図表示や、左下の Venn Diagrams でのハイライト表示と連動します。

また、図内には  $y=x$  の直線が引かれる(上図内③の実線)ほか、 $\pm\sigma$ 、 $\pm2\sigma$ 、 $\pm3\sigma$ に該当す線も引かれます。

さらに検定を実施している場合、有意基準に達しているかどうかで各点の色が変化します。検定でなく Coefficient of Variance (変動係数)が選択されていた場合は 1 以上の時、Fold Change が選択されていた場合は 2 以上、または 0.5 以下の時に「有意」としています。

### 8-3-2.Stdev Scatterplot タブ

Stdev Scatterplot タブでは、Coefficient of variance (または variation、変動係数:CV)の評価に利用するグラフです。Quantitative Value のばらつき具合を見ています。各点はタンパク質を表し、横軸が Quantitative Value の(カテゴリー内)全サンプルにおける Mean について  $\text{Log}_{10}$  となったもの、縦軸が同様の数値の SD です(次頁図)。もし検定が行われている場合は検定の対象として選択している sample のみ、表示対象とされます。



グラフ内の点線は各プロットから計算された回帰直線です。データのばらつきが大きなタンパク質を識別し、定量の検証に使うかどうかを判断する事もできます。

### 8-3-3.Volcano plot タブ

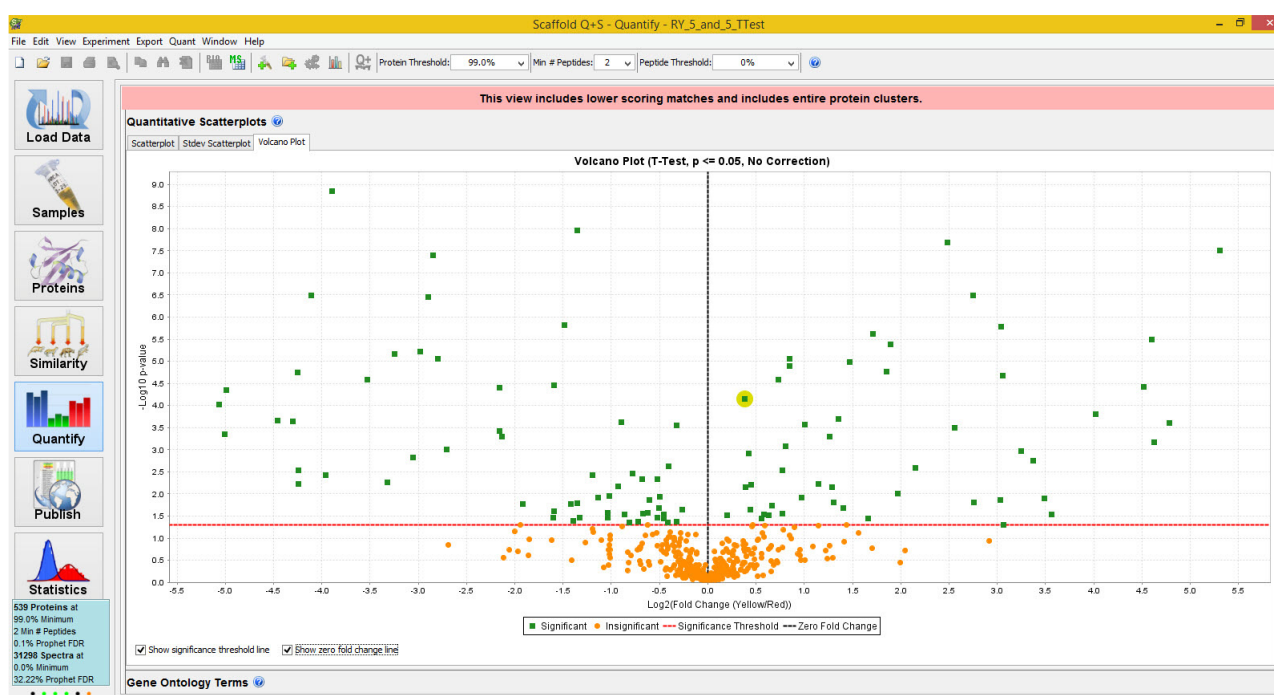
Volcano plot はタンパク質のサンプル間での変動を見るのに便利なグラフです。前提として、対象サンプルを選択した上での検定を実施する必要があります。

X 軸は Scatter plot タブで選択中の2カテゴリーにおける fold change の Log です。fold change とは、この場合単純に比(割り算)の事をさし、その値に対して  $\text{Log}_2$  をとっています\*。

\*各所で用いられる用語のばらつきによっては、本グラフの横軸である、比の  $\text{Log}_2$  の事を Fold change と呼称する場合があります。

Y 軸は検定の結果計算された各タンパク質の p-value について  $\text{Log}_{10}$  をとりマイナスをかけたものです。

従って、X 軸については左右の方向にいけばいくほど、Y 軸については上に行けば行くほど該当タンパク質が選択サンプル間で大きく変動していることを意味します(次頁図)。

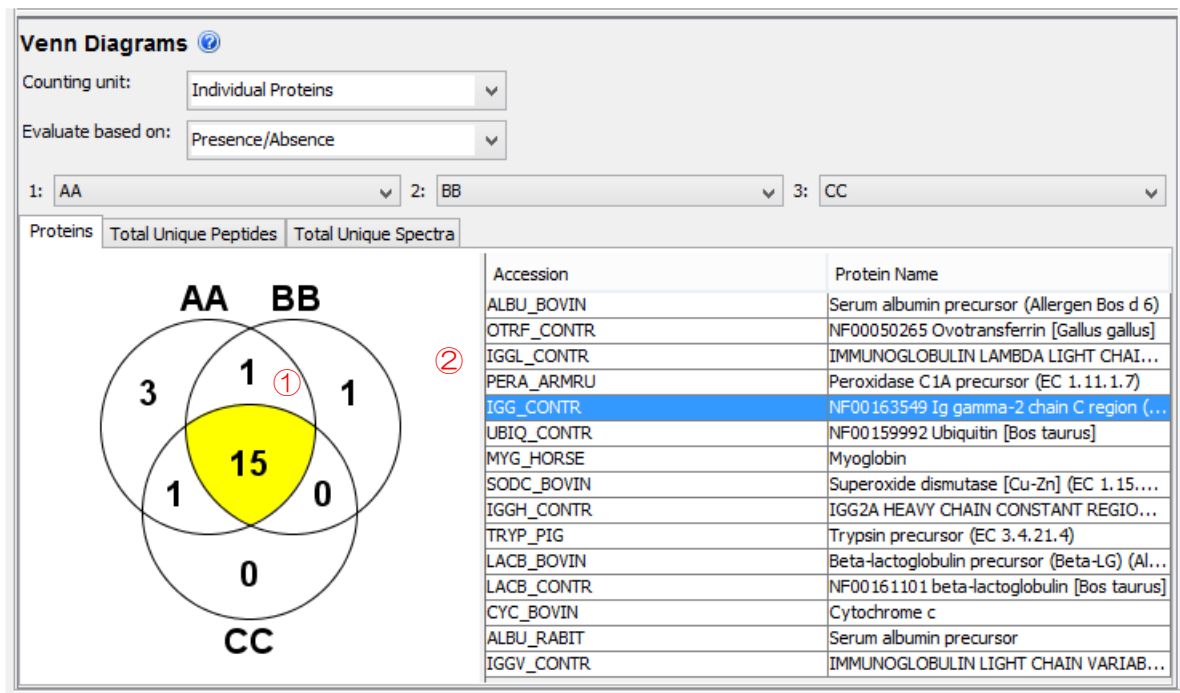


他のグラフ同様、各点はタンパク質を表し、点にカーソルを合わせると accession 並びに X/Y 軸の値が表示されます。またクリックするとそのタンパク質の情報について左上の Quantitative Value pane での図表示や、左下の Venn Diagrams でのハイライト表示と連動します。

X 軸の 0 の値、すなわち変動がないラインについてはグラフ表示のオプションで変更することができます。また同様に Y 軸の Significant のラインについても変更可能です。

## 8-4. Venn Diagrams pane

Quantify 画面の左下、「Venn Diagrams pane」では最大3つのカテゴリーを対象にベン図を書かせることができます。「Proteins」「Total Unique Peptides」「Total Unique Spectra」3つのタブから構成され、それぞれタンパク質/ペプチド/スペクトル数 についてベン図を描く事ができます(下図)。



ベン図内の各領域を選択しクリックすると黄色のハイライト表示がされ(上図①)、選択領域に属するタンパク質/ペプチド/スペクトル の一覧がその右側に表示されます。表内にて青色でハイライト(上図②)を受けているのは現在選択されているタンパク質で、表内の行をダブルクリックする事でタンパク質を切り替える事もできます。

また表内のエリアをダブルクリックすると Samples View に切り替わり、その時に選択エリアに 属するタンパク質だけがフィルターリングで残された状態となっています。元の状態に戻すには Samples 画面上部の「Search」のところにある文字列を削除し空欄にします。

グラフの上の「Counting Unit」では、数え上げる対象をクラスター単位とするか個々のタンパク質とするか変更することができます。データ取り込み時のグループ化オプションによっては表示されません。

また「Evaluate based on」は対象とするタンパク質についてのオプションで、「Presence/Absence」は Samples 表での有無で判断するのに対し、「Quantitative profile」は検定に使用されたかどうかで判断されます。従って「Quantitative profile」は検定を行った時にしか選択する事ができません。

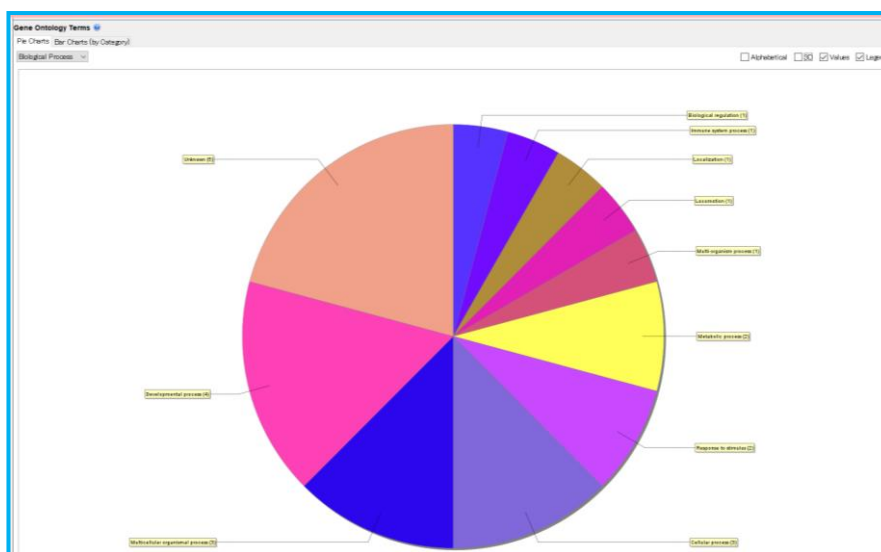
## 8-5. Annotation Charts pane

Quantify View の右下、「Annotation Charts pane」では、各タンパク質にアサインされた Gene

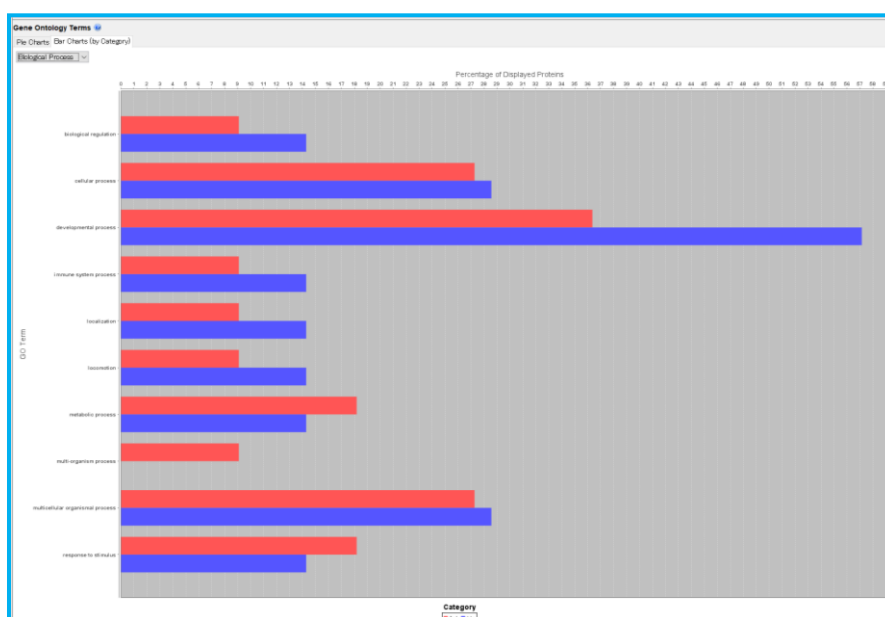
Ontology または Pathway 情報に関してまとめた図を表示します。それぞれ、2つのタブ「Pie Charts」「Bar Charts」から構成されます。

例えば Gene Ontology では「Biological Process」「Cellular Component」「Molecular Function」から構成されていますが、画面上部のプルダウンから該当項目を選ぶことで表示内容を切り替える事ができます。

「Pie Charts」タブ では Samples で表示されているすべてのタンパク質について Gene Ontology/Pathway 情報の項目を数え上げ、それを円グラフにて表示しています(下図)。



一方同様の情報を棒グラフにて表示しているのが「Bar Charts」です(下図)。



## 9. Publish View

### 9-1. Publish View : Method の文章化

Scaffold では各種機能を持つ View(画面)があり、左側にそれらの View を切り替えるためのスイッチがあります。

9 章では「**Publish**」View について説明しています。各種論文投稿時、データ解析に関する説明文を記載する必要がありますが、その目的に役立つのが Publish View 画面です。

Scaffold Evaluation - Publish - Label-Free

File Edit View Experiment Export Quant Window Help

Protein Threshold: 99.0% Min # Peptides: 2 Peptide Threshold: 95%

Experiment Methods

Parameter	Value
Experiment:	Label-Free
Peak List Generator:	
Version:	
Charge States Cal...	
Deisotoped:	
Textual Annotation:	
Database Set:	1 Database
Database Name:	the uniprot_sprot_2010_08 database
Version:	
Taxonomy:	All Entries
Number of Pr...	519348
Does database co...	
Search Engine Set:	2 Search Engines
Search Engine:	Mascot
Version:	2.4.0
Samples:	All Samples
Fragment Tole...	0.020 Da (Monoisotopic)
Parent Tolera...	10.0 PPM (Monoisotopic)
Fixed Modifica...	+57 on C (Carbamidomethyl)
Variable Modif...	+1 on NQ (Deamidated), +16 on M...
Database:	the uniprot_sprot_2010_08 databa...
Digestion Enzy...	stricttrypsin
Max Missed Cl...	2
Probability Mo...	
qe2_10122...	LFDR Model, Classifier data: Baye...
qe2_10122...	LFDR Model, Classifier data: Baye...
Search Engine:	X! Tandem
Version:	CYCLONE (2010.12.01.1)
Samples:	All Samples
Fragment Tole...	0.020 Da (Monoisotopic)
Parent Tolera...	10.0 PPM (Monoisotopic)
Fixed Modifica...	+57 on C (Carbamidomethyl)
Variable Modif...	-18 on Peptide N-Terminal (Glu...
Database:	the uniprot_sprot_2010_08 databa...
Digestion Enzy...	stricttrypsin
Max Missed Cl...	2
Probability Mo...	
qe2_10122...	Peptide Prophet with Delta Mass ...
qe2_10122...	Peptide Prophet with Delta Mass ...
Scaffold:	Version: Scaffold_4.8.9
Modification Meta...	1541 modifications
Source:	C:\Program Files\Scaffold\4.0-exp...
Comment:	
Protein Grouping ...	Experiment-wide grouping with bin...
Peptide Thresholds:	95.0% minimum
Protein Thresholds:	99.0% minimum and 2 peptides min...
Peptide FDR:	0.0% (Decoy)
Protein FDR:	0.0% (Decoy)
GO Annotation S...	gene_association.goa_uniprot-large ...
Alternate ID Sour...	

315 Proteins at 99.0% Minimum  
2 Min # Peptides  
0.0% Decoy FDR  
3591 Spectra at 95.0% Minimum  
0.00% Decoy FDR

EXPORT PROTEIN REPORT EXPORT PEPTIDE REPORT EXPORT PUBLICATION REPORT

DATABASE SEARCHING-- Tandem mass spectra were extracted by [unknown] version [unknown]. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.4.0) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). Mascot was set up to search the uniprot\_sprot\_2010\_09 database (unknown version, 519348 entries) assuming the digestion enzyme stricttrypsin. X! Tandem was set up to search the uniprot\_sprot\_2010\_09 database (unknown version, 519348 entries) also assuming stricttrypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Mascot and X! Tandem as a fixed modification. Deamidated of asparagine and glutamine, oxidation of methionine and acetyl of the n-terminus were specified in Mascot as variable modifications. Glu->pyro-Glu of the n-terminus, ammonia-loss of the n-terminus, gln->pyro-Glu of the n-terminus, deamidated of asparagine and glutamine, oxidation of methionine and acetyl of the n-terminus were specified in X! Tandem as variable modifications.

CRITERIA FOR PROTEIN IDENTIFICATION-- Scaffold (version Scaffold\_4.8.9, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability. Peptide Probabilities from Mascot were assigned by the Scaffold Local FDR algorithm. Peptide Probabilities from X! Tandem were assigned by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5393-92) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Al et al Anal. Chem. 2003;75(17):4646-58). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins were annotated with GO terms from gene\_association.goa\_uniprot-large (downloaded Feb 25, 2013). (Ashburner, M et al Nat. Genet. 2000;25(1):25-9).



画面左側が、解析内容について項目別にまとめられた表です。取り込んだデータの内容に合わせて項目が自動入力されます。項目前に鍵マークがついている欄はユーザーによる書き換えができません。

赤い強調表示されている項目はユーザーの手入力による穴埋めが促されています。そのうち入力欄に「V」となっている箇所については、クリックする事で選択肢が現れユーザーが選択をする形になっています。

左側の表に入力された内容に合わせて、右側の説明文が自動的に編集されます。右側の説明文についてはドラッグ&ドロップで文章を選択できるほか、ショートカットキーなどでテキストをコピーする事ができます。ユーザーはこの文章を出発点とし書き換えたものを論文投稿などでの method に利用する事ができます \*1。

右側下部には情報を EXCEL で読むことができる形式で出力できるボタンがついています。各ボタンによって出力される内容は、メニューの「Export」にて選択できる内容と同じです。詳細は「4-1.メニューの内容 説明一覧」の「Export」欄をご覧ください。

**\*1**

Scaffold を利用した解析においては、論文投稿時に以下論文を参照してください。

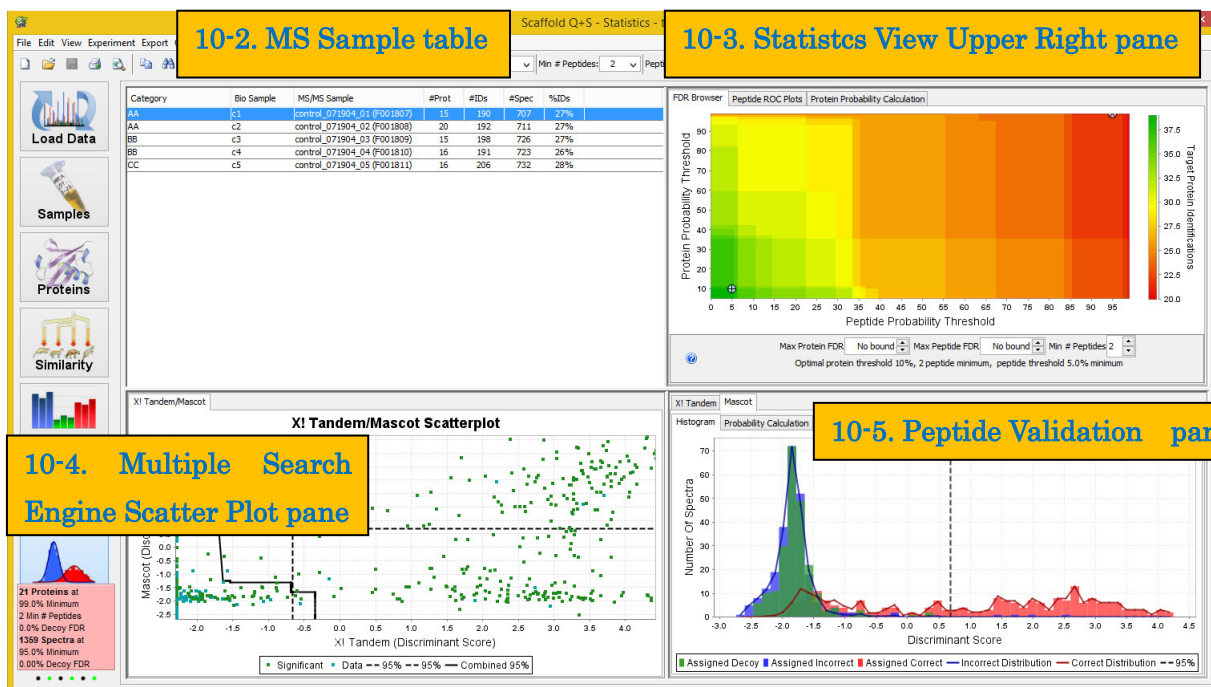
<https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200900437>

## 10. Statistics View

### 10-1. Statistics View 概要：同定確率計算に使用したスコア分布などを表示

Scaffold では各種機能を持つ View(画面)があり、左側にそれらの View を切り替えるためのスイッチがあります。

「**Statistics**」View では、主にペプチド同定やタンパク質同定のアルゴリズムに関連するデータ・グラフを表示する事ができます。従って、データを Prefiltered mode で取り込んだ場合は何も表示されません。



Statistics View は主に4つの画面から構成されています。

左上、**MS Sample table** は、Experiment 内の MS Sample に関する情報をまとめた表です。

右上、**Statistics View Upper Right pane** は、「FDR Browser tab」「Peptide RPC Plots tab」「Protein Probability Calculation tab」3つのタブから構成されていて、ペプチドやタンパク質の probability 計算に直結する数値を確認できるグラフです。

左下、**Multiple Search Engine Scatter Plot pane** は、複数の検索エンジンでのスコアを比較する散布図です。

右下、**Peptides Validation pane** は、検索エンジンでのスコア分布と適用アルゴリズムでの分布を同時に確認できるグラフです。

左上の MS Sample table で選択されている項目について、他の pane のグラフ・図が表示されます。選択の変更によってグラフもインタラクティブに表示が切り替わります。また画面上部の各種フィルターの選択内容にもインタラクティブに対応します。



## 10-2. MS Sample table

Statistics View の左上、MS Sample table は、MS Sample に関する情報がまとめられています (下図)。表示される情報は以下の通りです。

Category	Bio Sample	MS/MS Sample	#Prot	#IDs	#Spec	%IDs
AA	c1	control_071904_01 (F001807)	15	190	737	27%
AA	c2	control_071904_02 (F001808)	19	192	711	27%
BB	c3	control_071904_03 (F001809)	15	198	726	27%
BB	c4	control_071904_04 (F001810)	16	194	723	26%
CC	c5	control_071904_05 (F001811)	15	205	732	28%

- **Category** - 属する Category 名
- **BioSample** - 属する BioSample 名
- **MS/MS sample** - MS Sample 名
- **#Prot** - 同定タンパク質数
- **#IDs** - ユニークな同定スペクトル数。MS/MS Sample View モードでの、Exclusive Spectrum Counts の和(下図、黄色の枠で囲まれた箇所を参照)。またクラスターモードの時にも個々のタンパク質についてカウントされます。
- **#Spec** (ユニークでないものも含む)同定スペクトル数。
- **%IDs** 同定スペクトル中に占めるユニークスペクトルの割合。#IDs / #Spec。

MS/MS Sample View Selected

Display Option: Exclusive Spectrum Counts

MS Spectra table

Category	Bio Sample	MS/MS Sample	#Prot	#IDs	#Spec	%IDs
AA	c1	control_071904_01 (F001807)	15	190	737	27%
AA	c2	control_071904_02 (F001808)	19	192	711	27%
BB	c3	control_071904_03 (F001809)	15	198	726	27%
BB	c4	control_071904_04 (F001810)	16	194	723	26%
CC	c5	control_071904_05 (F001811)	15	205	732	28%

#ID corresponds to the Sum of values squared in yellow for the MS Sample

## 10-3. Statistics View Upper Right Pane

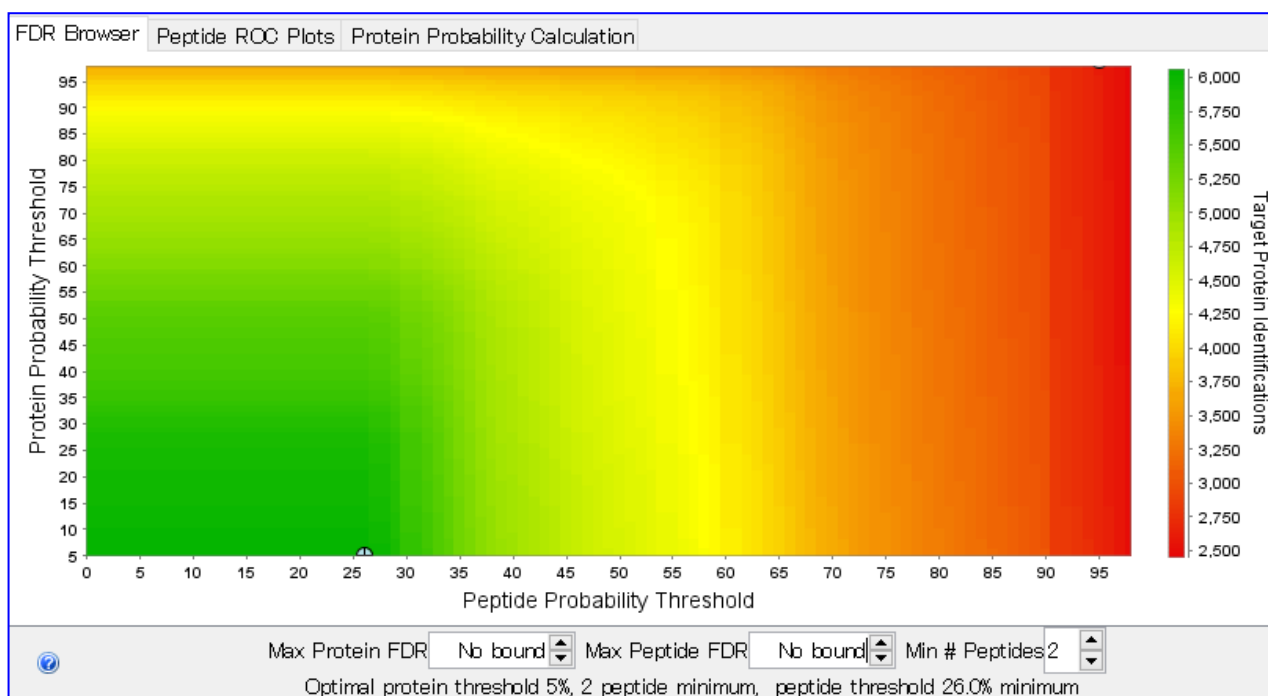
Statistics View の右上の pane には公式に名称を付けておりませんが、主にペプチドやタンパク質の同定の検証に関連する情報を提供しています。以下3つのタブから構成されています。

- **10-3-1. FDR Browser** タブ
- **10-3.2 Peptide ROC Plots** タブ
- **10-3-3. Protein Probability Calculation** タブ

以降、1つ1つのタブについて説明しています。

### 10-3-1. FDR Browser タブ

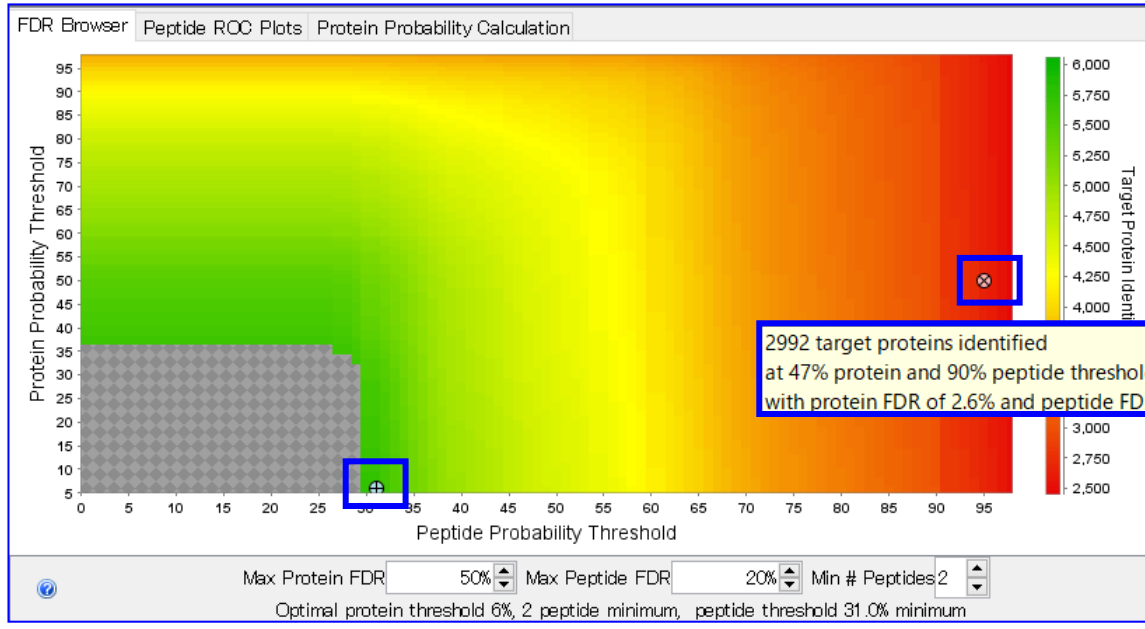
通常データベースに加え decoy データベースへの検索も実施している時のみ現れるタブです。Peptide Probability と Protein Probability の組み合わせにより、基準を超えるタンパク質がいくつになるかをヒートマップで表しています(下図)。赤がタンパク質の数が少なく緑が多くなっていますが、基本的に赤の領域ほど確度の高いタンパク質の領域です。ヒートマップの描写はグラフ下の設定 数値と連動しています。設定値には「**Max. Protein FDR**」「**Max Peptide FDR**」「**Min # Peptides**」の3種類があります。



このグラフの一番の目的は、Peptide Probability と Protein Probability の組み合わせが同定タンパク質数にどのように影響を及ぼすのかを確認する事です。解析の目的に応じて Peptide, Protein の Probability の Threshold を変更し状況に合わせた同定タンパク質リストを作成する事ができます。

表示されているヒートマップ上にカーソルを合わせると、その位置における Peptide Probability, Protein Probability、基準を超えるタンパク質数とともに、Peptide や Protein の FDR の値も表示され

ます(下図、青枠)。また、グラフの中には条件によりプロットが現れる事がります。青い○は現在の条件の中で最も拾い上げるタンパク質が多くなり、かつ各基準値の値が高くなる位置を表します。ピンクの○は現在 Filter として設定しているアサインペプチド数とグラフ内の Min # Peptides が 同一時のみ表示され、filter の設定箇所を表します。

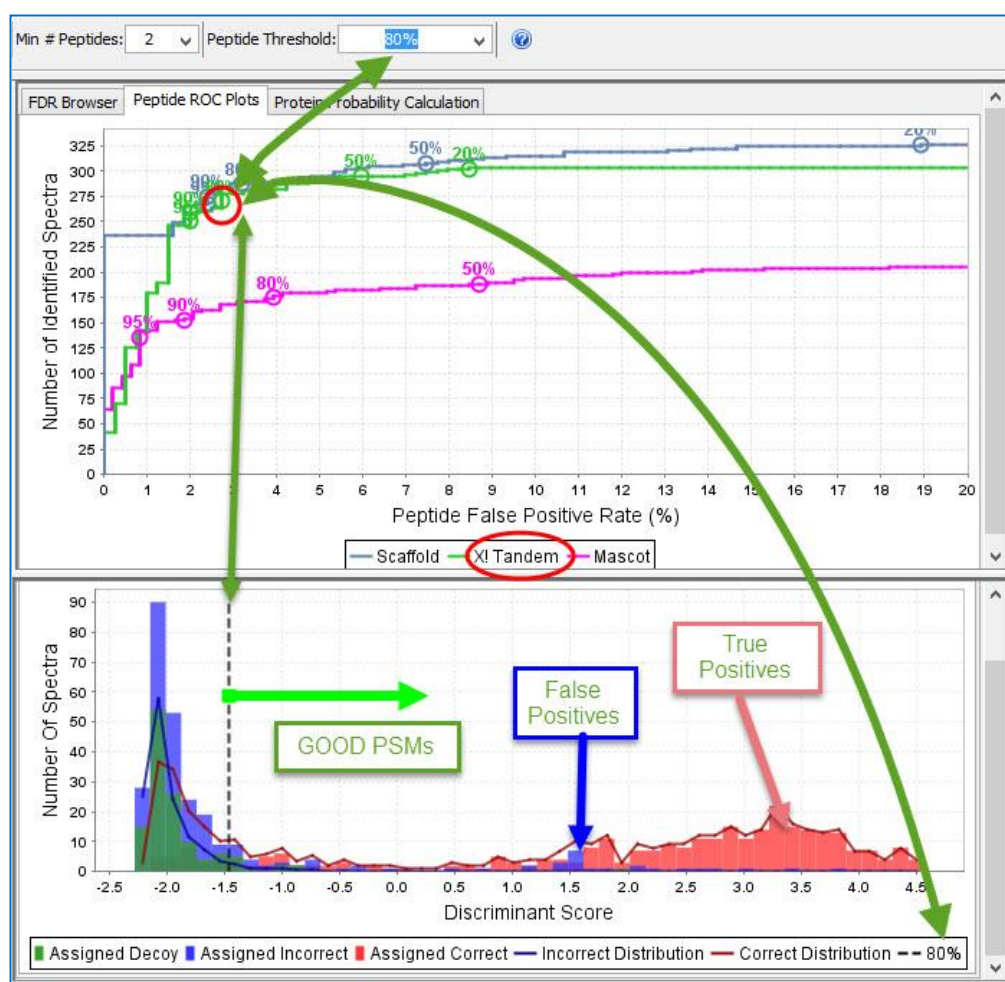


## 10-3-2. Peptide ROC Plots

Peptide ROC Plot タブでは、ROC カーブの図が表示されます。ペプチド同定において、同定基準と、Sensitivity / Specificity のバランスを確認するために利用します。

横軸に偽陽性率(1-特異度 (陰性を陰性と判断する率))、縦軸にはここでは同定ペプチド数をとったグラフです。\* 通常は Sensitivity (陽性を要請と判断する率)とする事が多いです。各判定方法 (ペプチド同定アルゴリズム)の、各同定基準値において、False Positive Rate と同定ペプチド数がそれぞれいくつかを表す線が描かれています(次頁図)。Statistics View の左上、MS Sample pane にて選択されている MS Sample についてのグラフが表示されます。

グラフの中の Sensitivity や Specificity の数値については、右下のタブ「Peptide Validation pane」の表示内容から計算されています。具体的には、Correct を Positive、Incorrect を Negative と認定してそれぞれのスコア分布から、各スコアにおける Sensitivity や Specificity が計算されます。

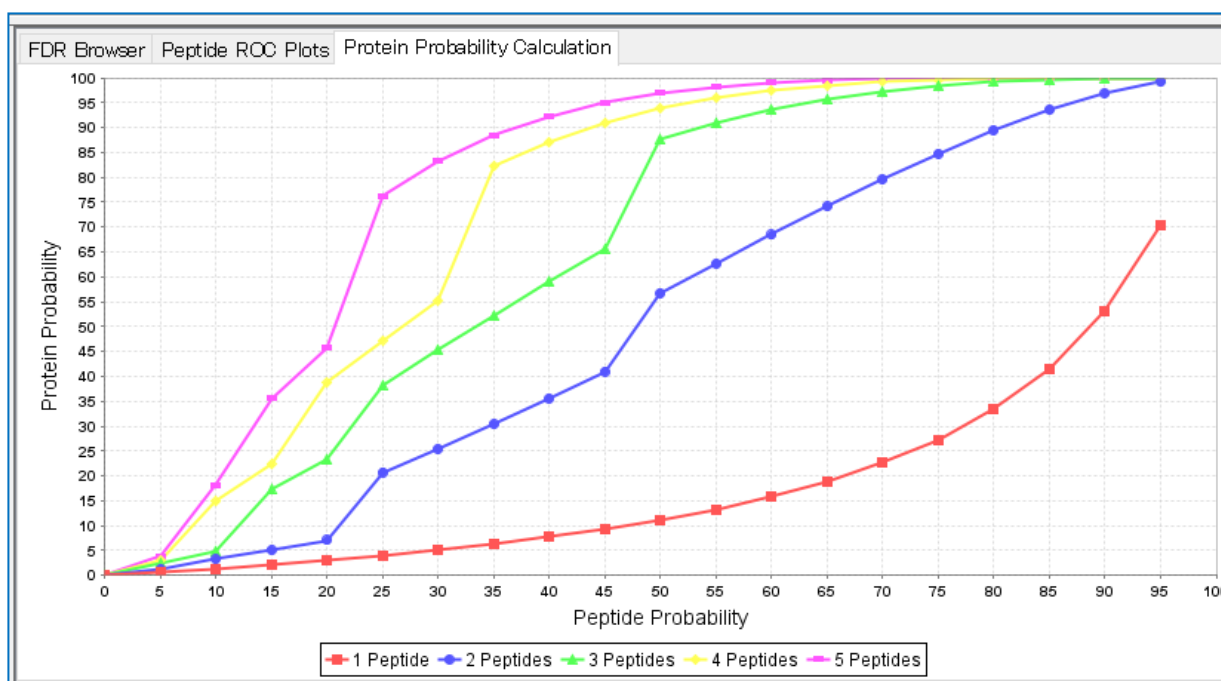


本来の ROC カーブは縦軸が Sensitivity で、 $y=x$  の直線を引いてその直線と各プロットがどれくらい離れているかが有効な判定方法かを評価する目安となりますが、本ソフトウェアのグラフでは縦軸の値が異なるためそのような見方をすることができません。よりグラフの左上方向に曲線が向かっている判定方法が優秀である、といった見方だけが可能です。

### 10-3-3. Protein Probability Calculation タブ

Protein Probability Calculation タブは、アサインされたペプチド数別に Peptide probability と Protein Probability の関係性を表したグラフです。検索を行った query 数が少ない場合は表示されません。Statistics View の左上、MS Sample pane にて選択されている MS Sample についてのグラフが表示されます。

Scaffold で採用されている、タンパク質同定確率を表す「Protein Probability」は、そのタンパク質にアサインされているペプチドの「Peptide Probability」から計算されています。計算式には検索のパラメータや FASTA のデータベースサイズが関係しています。



グラフを見ると、1 アサインペプチドの場合には peptide probability が 95%でも protein probability が十分な値ではない(上図例では 70%)である事がわかります。逆に 2 つ以上のペプチドがアサインされているタンパク質では、peptide probability が十分な値でなくても protein probability が高い値になる事がわかります。

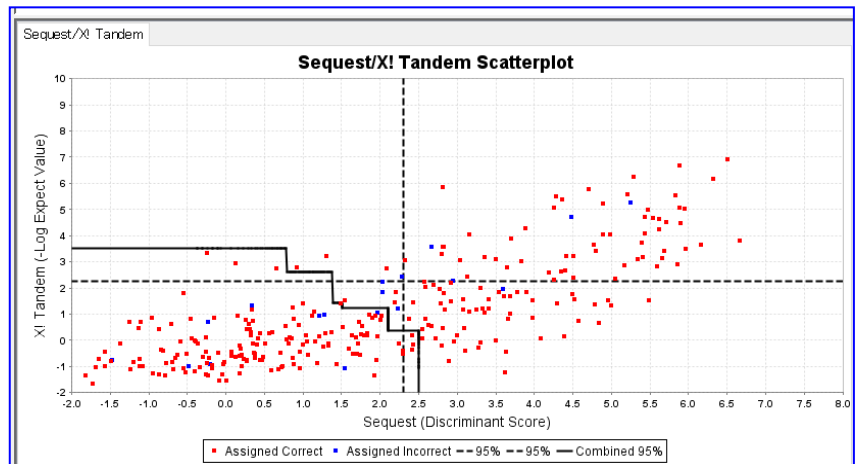
[次頁に続きます]



## 10-4. Multiple Search Engine Scatter Plot pane

Statistics View の左下、「Multiple Search Engine Scatter Plot pane」では、同じ query に対して複数の検索エンジンを適用した場合の、スコアの違いをプロット化したグラフが表示されます(下図)。Statistics View の左上、MS Sample pane にて選択している MS Sample についてのグラフが表示されます。

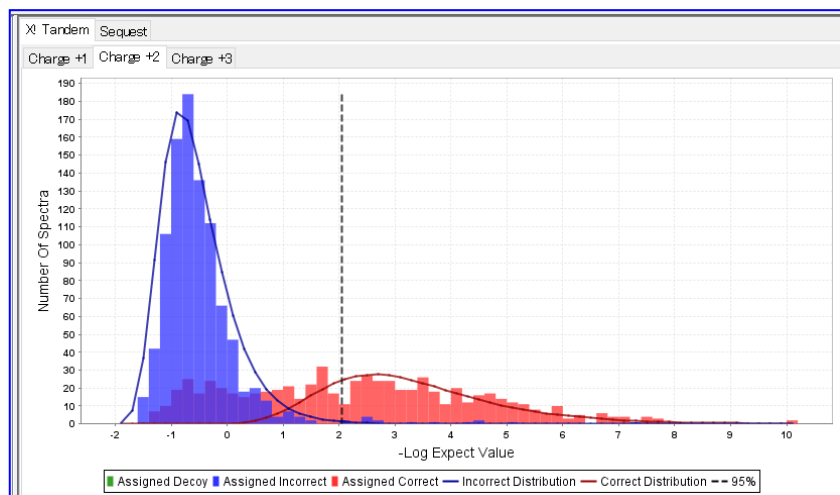
二軸それぞれに検索エンジンのスコアが配置され、それぞれの(同定)基準値が横軸/縦軸に点線で表示されます。また各プロットは correct に属するデータと incorrect に属するデータで色分けされます。さらに、Scaffold の peptide 同定確率を判定するプログラム peptide prophet のラインが "combined 95%" と表記されている実線としてグラフに表示されています。



## 10-5. Peptides Validation pane

Statistics View 画面の右下、Peptides Validation pane では、各検索エンジンのスコア分布から peptide probability 算出に利用するスコア分布をどのように変換しているか確認できるグラフが表示されます。

評価アルゴリズムとして Peptide Prophet が選択されていた場合、各検索エンジンにおいて、電荷別のスコア分布が表示されます。横軸にはマッチング内容を評価するスコアが、縦軸には個数が表された棒グラフが表示されます(下図)。棒グラフは incorrect と correct 別に色分けがされており、あるスコア位置での両社の割合がそのまま peptide probability として評価されます。



## 11. 定量手法と検定

### 11-1. 概要

Scaffold では、いくつかの Label Free の定量手法 (Spectrum Counting と MS1 イオン強度ベースの定量)に対応しています。また定量の数値を元にした検定も実施する事ができます。この章では以下の内容で、Scaffold で対応している定量手法と検定について説明しています。

11-2. ラベルフリーの定量手法

11-3. Normalization について

11-4. 検定

### 11-2. ラベルフリーの定量方法

Scaffold で対応しているラベルフリーの定量手法は大きく分けると以下の3つの手法に対応しています。

- **Spectral Counting [11-2-1]**

各タンパク質にアサインされている、同定スペクトル数(または同定ペプチド数)を元にした定量指標です。

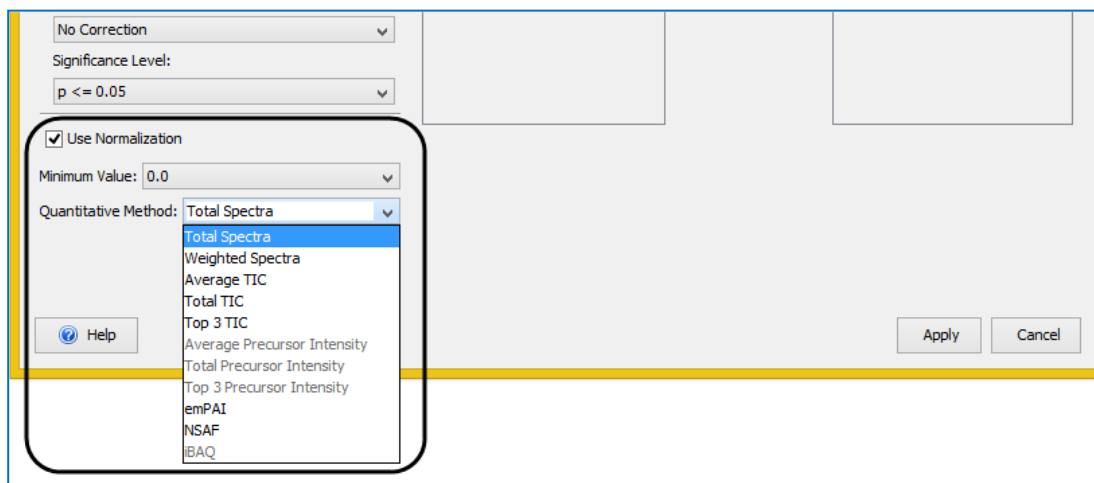
- **Total Ion Count [11-2-2]**

MS2 のスペクトルの intensity 情報を定量情報として利用します。

- **Precursor Ion Intensity quantitation [11-2-3]**

(利用できる場合のみ) MS1 のペプチドの Intensity 情報を定量情報として利用します。

Scaffold では、各手法に分類される選択肢がそれぞれ複数存在します。以降で各選択肢について説明しています。



## 11-2-1. Spectral Counting

Spectral Counting は各タンパク質にアサインされている、同定スペクトル数(または同定ペプチド数)を元にした定量指標です。Scaffold では以下の4項目が選択可能です。

- **total Spectra**
- **Weighted Spectra**
- **emPAI**
- **NSAF**

以下、各手法について説明します。

### ■ total Spectra

タンパク質にアサインされた同定スペクトルの数です。Normalization (11-3 をご参照ください)も実施されています。

### ■ Weighted Spectra

タンパク質にアサインされた同定スペクトルのうち、他のタンパク質にもアサインされているシェアされたスペクトルについては、そのシェアされている割合により小さな数字としてカウントした数です。詳細は英文マニュアルの「7-2-1. Protein Grouping (sameset)」内に記述されている weight の計算式などをご参照ください。Normalization(11-3 をご参照ください)も実施されています。

### ■ emPAI

タンパク質にアサインされた同定ペプチド数を、タンパク質の大きさで有利/不利 が生じないよう標準化処理をされた定量指標です。

emPAI は Exponentially Modified Protein Abundance Index の略です。

PAI 自体は以下の式のように、タンパク質にアサインされたペプチド数( $N_{\text{observed}}$ )を、そのタンパク質から得られる理論ペプチド数( $N_{\text{observable}}$ )で割った数値です。

$$PAI = \frac{N_{\text{observed}}}{N_{\text{observable}}}$$

\*元の論文では保持時間や各種条件によりペプチドを数え上げる際にフィルターリング条件も含んでいます。

emPAI は、PAI を使って以下のようにあらわす事ができます。

$$emPAI = 10^{PAI} - 1$$

Scaffold では、計算の簡易化を目的に論文で適用されているルールとは異なるルールで  $N_{\text{observed}}$  や  $N_{\text{observable}}$  をカウントしています。Matrix Science 社で使用している以下ルールと基本的に同じです。

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

レアケースにおいて Scaffold でしか採用していない独自ルールがありますが詳細は英文マニュアルをご参照ください



## ■ NSAF

NSAF もタンパク質の大きさで有利/不利 が生じないように標準化処理をし、さらに同定スペクトル数全数との比を考慮した定量指標です。NSAF は Normalized Spectral Abundance Factor の略です。

論文で提示されている NSAF の計算方法ですが、まず Number of spectra (論文では SpC と表記)をタンパク質の全長(L)で割った、SpC/L が「SAF」です。解析別に SAF の値をすべて足した値で各 SAF を割って標準化(Normalized)したのが NSAF です。

Scaffold では SpC のところで「number of exclusive spectra」を適用しています。また normalization は「**11-3. Normalization** について」で説明する方法が採用されています。

### 11-2-2. Total Ion Count

Total Ion Count MS2 のスペクトルの intensity 情報を定量情報として利用します。以下の 3 項目が選択可能です。

- **Average TIC**
- **Total TIC**
- **Top Three TIC**

以下、各手法について説明します。

#### ■ Average TIC

query の MS/MS スペクトル内のすべてのピークについて、その強度の平均値を定量情報として利用します。

#### ■ Total TIC

query の MS/MS スペクトル内のすべてのピークについて、その強度の和を定量情報として利用します。

#### ■ Top Three TIC

query の MS/MS スペクトル内のすべてのピークの中で強度の強い top3 について、その強度の和を定量情報として利用します。

### 11-2-3. Precursor Ion Intensity quantitation

Precursor Ion Intensity quantitation は MS1 のペプチドの Intensity 情報を定量情報として利用 します。以下の 4 項目が選択可能です。

- **Average Precursor Intensity**
- **Total Precursor Intensity**
- **Total Three Precursor Intensity**
- **iBAQ**

この定量手法を利用するためには、Scaffold の取り込み前に別のソフトウェアで手法に合わせた 解析を実施する必要がある他、その時の解析ファイルを別途取り込ませる必要があります。詳細は 以下の資料あるいは英文マニュアルの 14 章をご参照ください。

[http://www.proteomesoftware.com/pdf/loading\\_search\\_engine\\_results\\_into\\_scaffold.pdf](http://www.proteomesoftware.com/pdf/loading_search_engine_results_into_scaffold.pdf)

以降、各手法について説明します。

### ■ Average Precursor Intensity

タンパク質にアサインされているペプチドの定量値について、幾何平均を定量情報として利用 します。

### ■ Total Precursor Intensity

タンパク質にアサインされているペプチドの定量値について、すべての和を定量情報として利用 します。

### ■ Total Three Precursor Intensity

タンパク質にアサインされているペプチドの定量値について、上位3つの和を定量情報として利用 します。データ数が3つ以下の場合、すべての和を利用します。

### ■ iBAQ

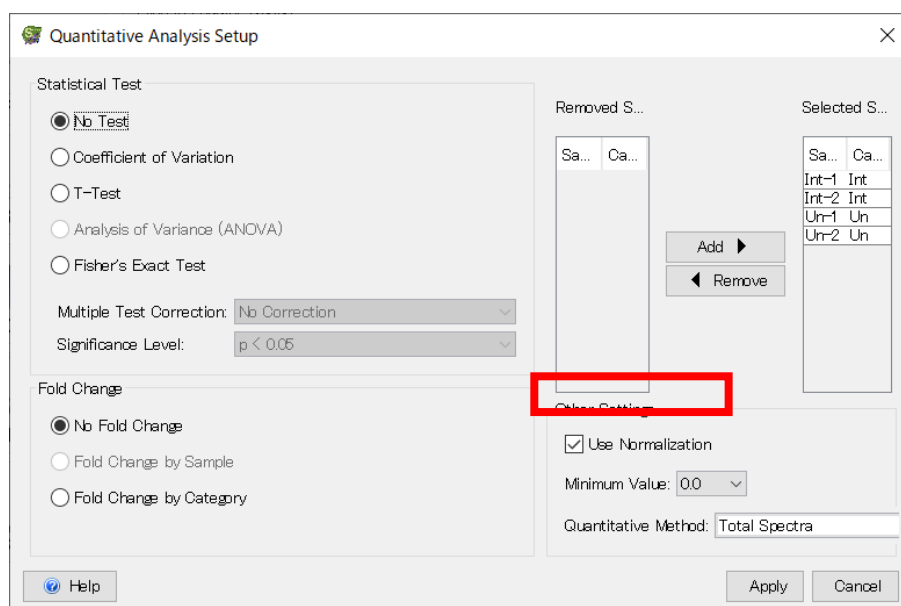
MS1 intensity ベースの定量に、emPAI の概念を組み合わせたような計算です。ペプチドの定量値の和を、該当タンパク質の理論ペプチド数で割ります。

## 11-3. Normalization について

Scaffold の定量計算では、サンプル間の誤差を吸収するための標準化処理が行われます。その前提となる考えが、「各サンプル (MS Sample) で同定されるタンパク質の全量は変わらない」というものです。Spectrum Counting なら同定スペクトル数の和が、MS1 や MS2 の Intensity ならその総和の 値が全サンプルで同じであるという前提の元、平均より多い(大きい)場合は少なく(小さく)なる ように、少ない(小さい)場合は多く(大き)くなるように調整されます。従って、サンプル間の タンパク質全量が等しくないと思われるケースでは Scaffold の Normalization を適用しない方が良いでしょう。

同様に、ダイナミックレンジの中で全量が少ないタンパク質は Normalization の影響を大きく受けて差が出やすくなり、本来の値からずれる事も大きくなる点に注意してください。(Scaffold の場合、intensity の値が小さいタンパク質やアサインされたスペクトル数が少ないタンパク質の事をさします。)

Normalization を実行する / しない、については、menu の Experiment -> Quantitative analysis で開くダイアログの右下、「Other Setting」の中に「Use Normalization」という項目があります(右図)。



なお、欠損値(Missing values)については、予め定めた最小値 (Minimum value)に置き換えて標準化計算の処理をします。Minimum value はデフォルトで 0 に設定されています。また設定した 最小値より小さな値も最小値に置き換えられてから計算されます。

## 11-4. 検定

Scaffold では算出された定量値について各種検定を実施できます。検定の計算時に利用される数値は、ペプチド・タンパク質などの各種フィルターリング条件を満たすものだけが対象となります。

Quantitative Analysis Setup

Statistical Test

No Test

Coefficient of Variation

T-Test

Analysis of Variance (ANOVA)

Fisher's Exact Test

Multiple Test Correction: No Correction

Significance Level:  $p < 0.05$

Fold Change

No Fold Change

Fold Change by Sample

Fold Change by Category

Removed Sam... Selected Sa...

Sam...	Cat...

Add ►

◀ Remove

Sa...	Cat...
Int-1	Int
Int-2	Int
Un-1	Un
Un-2	Un

Other Settings

Use Normalization

Minimum Value: 0.0

Quantitative Method: Total Spectra

Help Apply Cancel

検定とは別に、Fold Change (比)の値を表示させることができます。

利用可能な検定は選択しているカテゴリー数やカテゴリー内のデータ数などで変更され、合わない手法についてはグレイアウトされる事があります。

- Coefficient of Variance or Coefficient of Variation
- T-test
- ANOVA
- Fisher's Exact Test

検定を実施すると、Scaffold 上でいくつかの変化があります。

■Samples View : 検定の結果、有意性を示す数値(p-value など)と、変動の様子を表す列が表示されるようになります。

■Quantify View :Quantitative Value pane にて、検定の有意性と関連した値がグラフに表示されるようになります。

Fold Change 並びに各検定項目について、以降で説明いたします。

### ■ Coefficient of Variance or Coefficient of Variation

CV、変動係数は標準偏差を平均で割ったもので、データのばらつき度合を表します。値が大きいほどばらつきが大きい事を意味します。Scaffold では %表示となっています。値が大きな場合、サンプルのうち最低 1 つの値が大きすぎていて、値を(グラフなどで)見比べる事で変動のあったサンプルを特定していく事ができます。

### ■ T-Test

T 検定では 2 つのカテゴリー間の平均に有意差があるかを検証しています。Scaffold では両側検定を行っています。T 検定を実施すると、p-value が Samples 画面にレポートされます。T 検定を実施するにはカテゴリーが 2 つ以上選択されている必要があります。

### ■ ANOVA

ANOVA (ANalysis Of VAriance、分散分析)では、3 つ以上のカテゴリー間の平均に有意差があるかを検証しています。ANOVA は様々な手法がありますが、Scaffold で採用されているのはシンプルな手法のみです。ANOVA を実施すると p-value が Samples 画面にレポートされます。p-value が小さい場合、CV による評価同様少なくともカテゴリーの中のどれか1つが他と異なる事を示しますが、具体的な内容はわかりません。値をグラフなどで見比べる事により変動のあったサンプルを特定していくことができます。

### ■ Fisher's Exact Test

Fishers Exact Test (フィッシャーの正確確率検定、フィッシャーの直接確率検定)は、T 検定同様 2 サンプル間の比較により両サンプル間に差異があるかを判定する手法です。T 検定との選択における大まかな目安として、カテゴリー内の繰り返しが少ない場合(2 回など)に F 検定を適用する事を お勧めします。フィッシャーの正確確率検定を実施すると p value が Samples 画面にレポートされ、その小ささで差異があるかを判断します。

p-value の算出に関しては、以下論文で議論され構築されたモデルに基づいています。

[論文]

Zhang, B., VerBerkmoes, N.C., Langston, M. A., Uberbacher, E., Hettich, R. L., Samatova, N. F.

*Detecting differential and correlated protein expression in label-free shotgun proteomics*

J. Proteome Res., 2006, 5 (11), pp 2909–2918. DOI: 10.1021/pr0600273

## ■ Multiple Test Correction と Significance Level

検定の選択画面の下に、多重検定補正の実施に関する選択肢と、実施する場合の閾値を設定する画面があります。

Statistical Test

No Test

Coefficient of Variation

T-Test

Analysis of Variance (ANOVA)

Fisher's Exact Test

Multiple Test Correction: No Correction

Significance Level:  $p < 0.05$

No Correction

No Correction

Benjamini-Hochberg (Recommended)

Bonferroni

Holm Step-down

Hochberg Step-up

多重検定の際に擬陽性が増加してしまう事を考慮し、別途評価基準を設けて当初の目的に合わせた新たな基準の線引きをする、というのがこの手法の目的です。Scaffold では以下の手法が準備されています。

- **Benjamini-Hochberg**
- **Borferroni**
- **Holm Step-down**
- **Hochberg Step-up**

Significance Level は、多重検定補正においてはよく「ファミリーワイズエラー率(FWER と表記される事が多い)」と呼ばれます。それぞれの手法について、以下簡単に補足説明を加えます(説明の都合により項目の順番を変更して説明しています)。

### ◇ **Borferroni**

この手の手法で最も単純で初期の頃に提唱された手法です。設定したい p-value に対し、検定の数で割り算をします。しかしこの数値の場合基準が厳しくなりすぎる(有意基準に満たないものが多くなってしまふ)という事で、現在ではこの手法は選択されず別の手法が使われる事が多くなっています。

### ◇ **Hochberg のステップアップと Holm のステップダウン**

Bonferroni の不等式をベースとしていますが、様々な点を考慮し基準が厳しくなりすぎないような補正が追加されています。同定タンパク質数が少ない場合はこの手法が向いています。

詳細は英文マニュアル Appendix E の「Holm's and Hochberg's Techniques to Control the Familywise Error Rate」をご覧ください。

### ◇ **Benjamini-Hochberg**

タンパク質の数が少なすぎない場合は基本的にこちらを選択してご利用ください。Hochberg のステップアップや Holm のステップダウン同様、厳しすぎない程度に補正されます。Scaffold に おける計算では以下の論文で提唱されたモデルを利用しています。

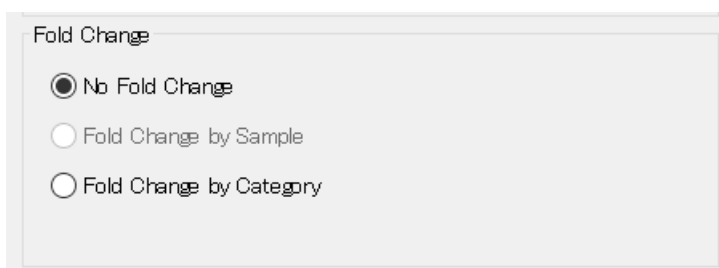
## [論文]

Benjamini Y. and Hochberg Y. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing* Journal of the Royal Statistical Society, Series B (Methodological), 1995, Vol.57, No. 1: 289-300

## ■ Fold Change

複数のサンプル間での比較を行うための検定をご紹介してきましたが、もっと単純な、何かの値をベース(Reference)に比をとって表示する「Fold Change」を表示させることもできます。

Biosample が2つしかない場合のみ「Fold Change by Sample」が選択できます。また、Quantitative Analysis にてカテゴリーを 2 つだけ選択している時に「Fold Change by Category」が選択できます(下図)。



Fold Change

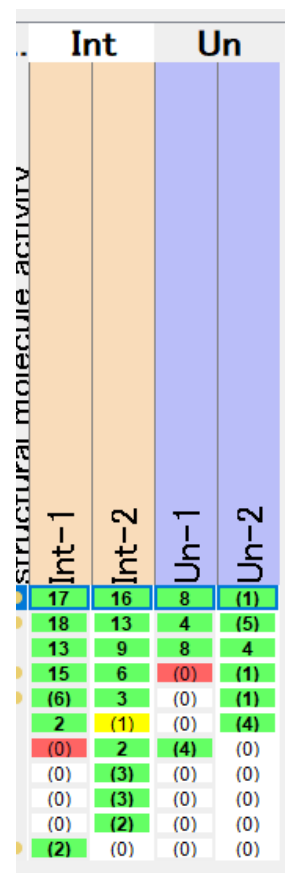
No Fold Change

Fold Change by Sample

Fold Change by Category

「Fold Change」を設定した場合、分母となるカテゴリーを選択し、もう一方との比が「Fold Change」列に表示されます。Samples 画面では分母に選ばれた方がページに、分子の方が薄紫色に色が塗られて表示されます(右図)。

Missing value を 0 に置き換えた場合などで分母が 0 であった場合、Fold Change の値は「INF」と表示されます。



	Int	Un		
STRUCTURAL MOLECULE ACTIVITY				
Int-1				
Int-2				
Un-1				
Un-2				
	17	16	8	(1)
	18	13	4	(5)
	13	9	8	4
	15	6	(0)	(1)
	(6)	3	(0)	(1)
	2	(1)	(0)	(4)
	(0)	2	(4)	(0)
	(0)	(3)	(0)	(0)
	(0)	(3)	(0)	(0)
	(0)	(2)	(0)	(0)
	(2)	(0)	(0)	(0)

## 技術サポート

ご質問等ありましたら弊社技術サポートにご連絡ください。

電子メール : [support-jp@matrixscience.com](mailto:support-jp@matrixscience.com)

電 話 : 03-5807-7897

