

Skyline MS1 全扫描筛选

Skyline 靶向蛋白质组数据分析环境为您提供对原始质谱数据的丰富的可视化显示功能。这些可视化功能允许您优化测量的肽段和离子对以及调整整合边界等操作。Skyline 最初被开发用于利用选择反应监测（简称 SRM；亦称为多反应检测，简称 MRM）质谱进行肽段定量分析，现在它的应用范围已经扩大到从数据依赖型采集模式获得的质谱数据的 MS1 扫描中提取时间-强度色谱峰以用于肽段定量。

Skyline MS1 全扫描筛选¹ 支持导入发现蛋白质组学实验中使用数据依赖型采集（DDA）模式获得的数据集。导入新的原始数据后，Skyline 新开发的和原有功能可以帮助定量测量多个重复样本的肽段离子 MS1 信号。

本教程将包含使用 Skyline MS1 定量的以下几个方面内容：

- 为 MS1 筛选设置 Skyline 文档
- 导入原始数据，并使用从谱图库获得的保留时间信息指导 MS1 筛选期间的峰挑选
- 进一步处理 MS1 筛选的肽段，以获得多次采集重复测定下的定量信息

Skyline 旨在为靶向蛋白质组学研究提供一个独立于供应商的平台。可以导入 AB Sciex、Agilent、Bruker、Thermo-Scientific 和 Waters 等公司质谱仪产生的原始质谱数据进行 MS1 筛选。支持不同平台原始质谱数据导入的能力极大地有助于不同仪器比较和大型多中心研究。

入门指南

为了进行本教程的学习，请下载下列 ZIP 文件：

https://skyline.gs.washington.edu/tutorials/MS1Filtering_2.zip

将其中的文件解压到您的电脑文件夹，例如：

```
C:\Users\brendanx\Documents
```

这将创建一个新文件夹：

```
C:\Users\brendanx\Documents\MS1Filtering
```

这里面将包含本教程所需的所有文件。现在启动 Skyline，您将看到一个新的空白文档。

将数据依赖模式采集的质谱数据的肽段搜索结果导入 Skyline 文档

将 DDA 模式采集的质谱数据的肽段搜索结果导入 Skyline 文档最简单的方法是使用“导入肽段搜索”向导。

如果您此前在 Skyline 中处理过全扫描数据，请执行以下操作以重置本教程的全扫描设置：

- 在“设置”菜单上，单击“离子对设置”。
- 单击“全扫描”选项卡。
- 在“M1 筛选”部分，将“同位素峰值包含”字段设置为“无”。
- 在“MS/MS 筛选”部分，将“采集方法”字段设置为“无”。

“离子对设置”表单将显示如下：

离子对设置

预测 过滤器 库 仪器 全扫描

MS1 筛选(M)

同位素峰值包含(I): 无
母离子质量分析仪(P):

峰(K): 分辨率(R): 质核比

同位素标记富集(B):

MS/MS 筛选(S)

采集方法(C): 无
子离子质量分析仪(M):

隔离方案(L): 分辨率(R): 质核比

保留时间筛选

仅使用 5 分钟之内进行的MS/MS ID扫描

仅使用 5 分钟之内进行的经预测的RT扫描

包含所有匹配的扫描

确定 取消

如果您之前处理过同位素标记肽段的数据，您可能还需要进入“肽段设置 - 修饰”标签，去掉所有同位素修饰设置。

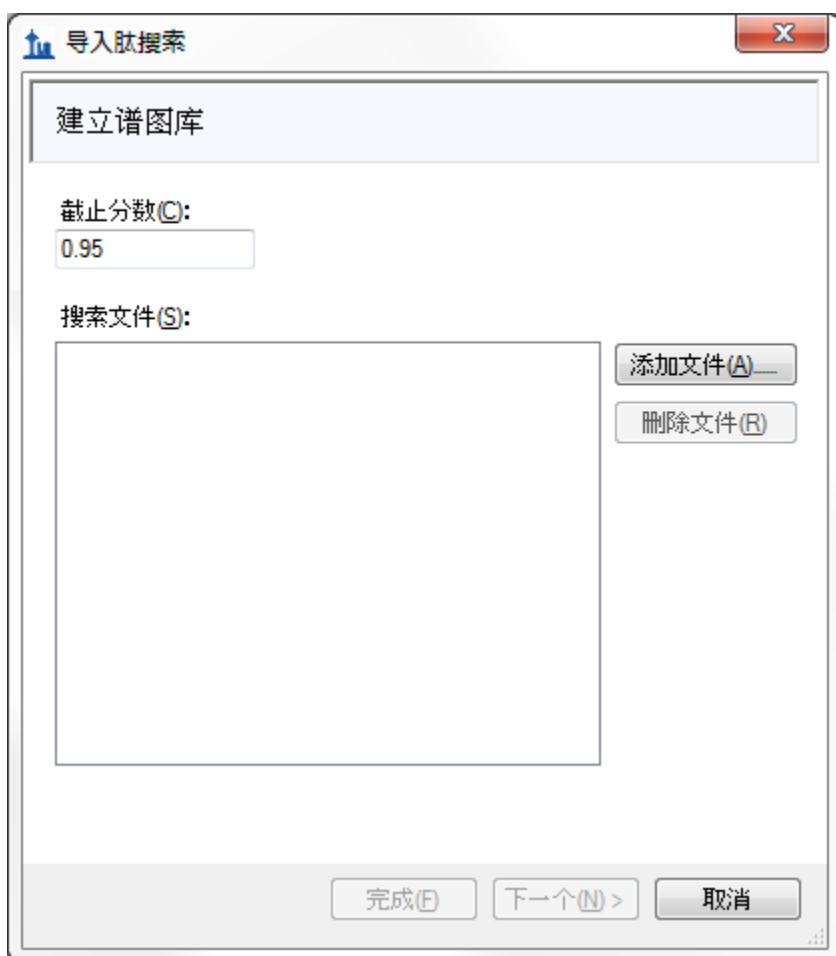
然后，执行以下操作以保存您的新文档：

- 单击工具栏上的“保存”按钮 (Ctrl-S)。
- 导航至您为本教程创建的 MS1Filtering 文件夹。
- 在“文件名”字段中输入“Ms1FilterTutorial.sky”。
- 单击“保存”按钮。

现在，按如下操作启动“导入肽段搜索”向导：

- 在“文件”菜单上选择“导入”，并单击“肽段搜索”。

Skyline 将呈现如下所示的表单：



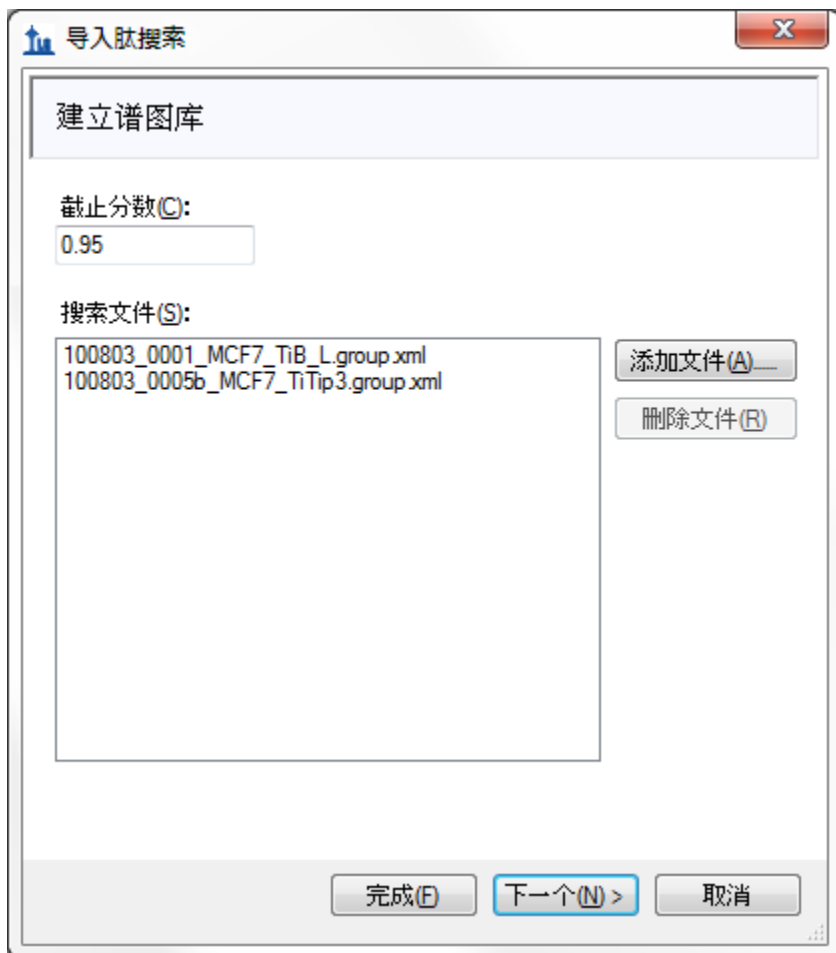
然后，您将可以使用本向导从 Skyline 支持的肽段搜索引擎的输出结果创建一个谱图库。关于其所支持的搜索引擎的完整列表，请查询“[靶向方法编辑](#)”教程。请注意，您将在本教程中使用的

数据已经被压缩至完成本教程所需的最小信息量，以便将下载的 ZIP 文件大小控制在合理的范围之内。

通过执行下列操作，将搜索结果添加至您的库：

- 单击“添加文件”按钮。
- 选择 MS1Filtering 文件夹中的两个 .group.xml 文件。
- 单击“打开”按钮。

向导视图将显示如下：



- 单击“下一步”按钮。

Skyline 将构建一个与您的 Ms1FilterTutorial.sky 文档相关的新谱图库，构建过程中将显示进度。在 MS1Filtering 文件夹中，您现在可以看到两个新文件：

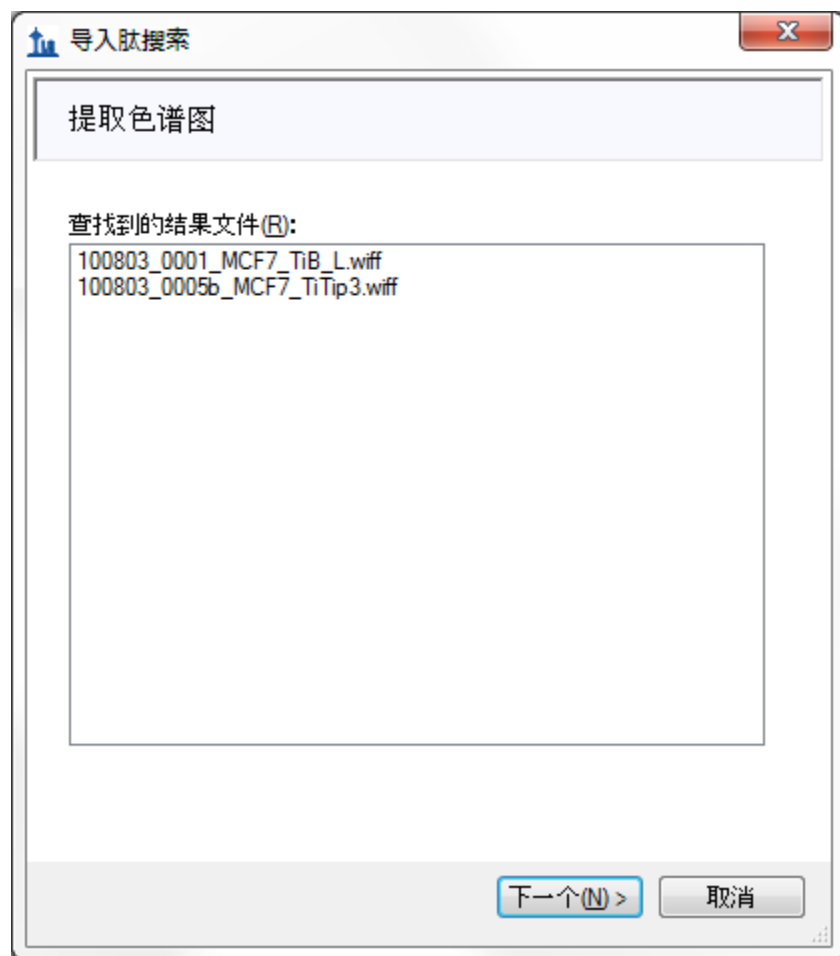
- 包含最佳匹配谱图的非冗余库“MS1FilteringTutorial.blib”。
- 包含所有匹配谱图的冗余库“MS1FilteringTutorial.redundant.blib”。

Ms1FilterTutorial.blib	10/15/2013 10:26 ...	BLIB File	803 KB
Ms1FilterTutorial.redundant.blib	10/15/2013 10:26 ...	BLIB File	1,160 KB

您还可以看到文件“MS1FilterTutorial.slc”，这是一个旨在改善库的加载时间的“Skyline 库缓存”文件。您可以删除它，Skyline 将在需要的时候重新建立这个文件。

如果您曾经使用 Skyline 构建谱图库，您可能已经习惯于根据自己的喜好命名，并将其保存在您喜欢的任何位置。在这种情况下，Skyline 将创建文档特异的谱图缓存库，与存储特定色谱图的方式非常相似。此后您还能添加更多搜索结果，以及删除搜索结果，与处理色谱图数据类似。

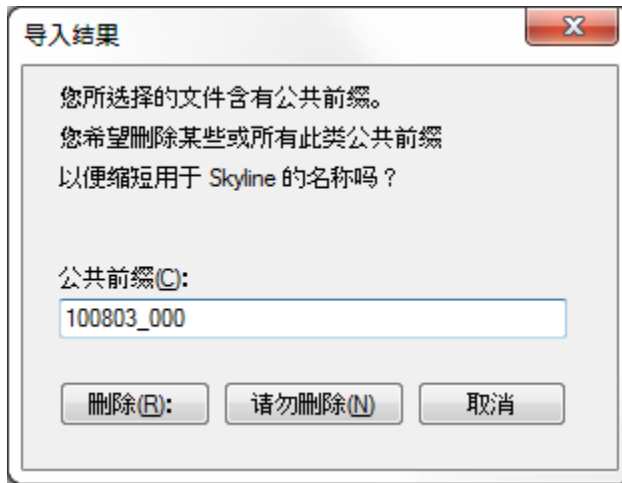
当完成库构建时，Skyline 将显示下列页面：



在这种情况下，Skyline 已找到与用于构建谱图库源文件相匹配的原始 WIFF 文件，且库似乎已包含 Skyline 将在其提取的色谱图上定位已鉴定的 MS/MS 谱图所需的保留时间信息。如果 Skyline 未能发现合适的数据文件进行色谱图提取，将需要您指定。如果构建的库无法在导入的肽段搜索文件中找到保留时间信息，Skyline 将通知您。参阅下面的“验证库保留时间信息”部分，了解 Skyline 确定谱图源文件或从肽段搜索结果文件中获得保留时间相关的故障排除的更多信息。如想继续本教程：

- 单击“下一步”按钮。

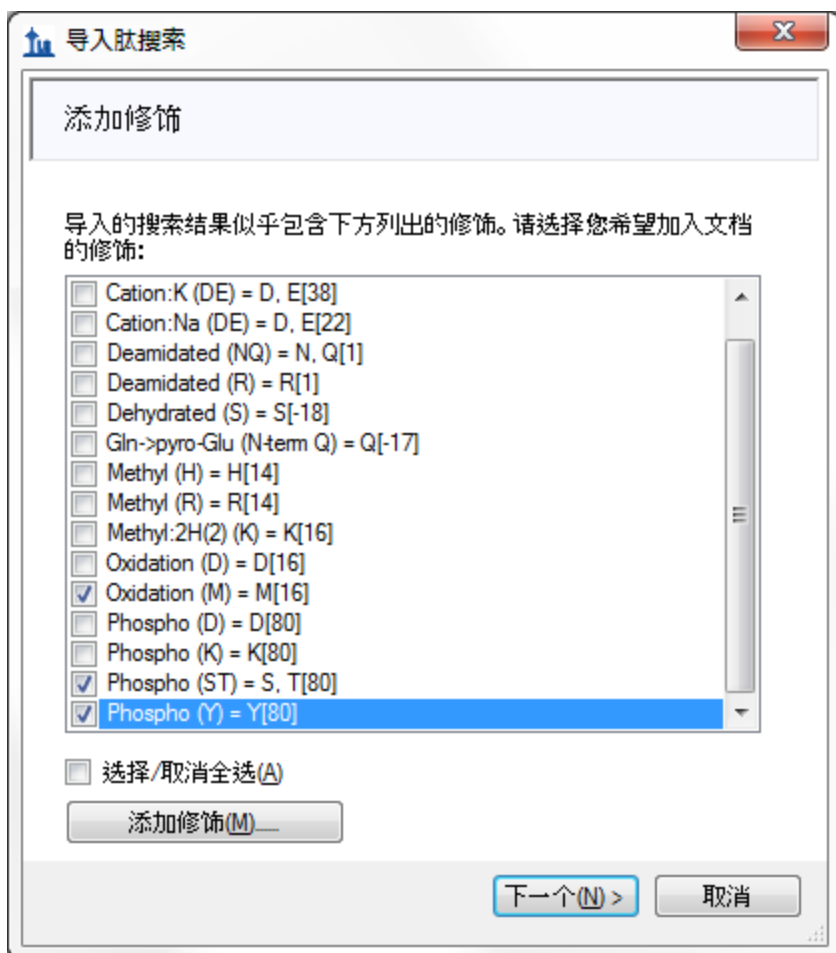
将出现一个表单，询问您如何处理两个 WIFF 文件的共同前缀：



- 单击“删除”按钮。

向导继续前进至“添加修饰”页面，该页面列出了所有搜索结果中发现的但文档中没有的氨基酸修饰。它很有可能会提示与该修饰相匹配的 Unimod 特定位点修饰。

在本教程中，您只需要“Phospho (ST)”、“Phospho (Y)”和“Oxidation (M)”修饰。勾选这些修饰如下：



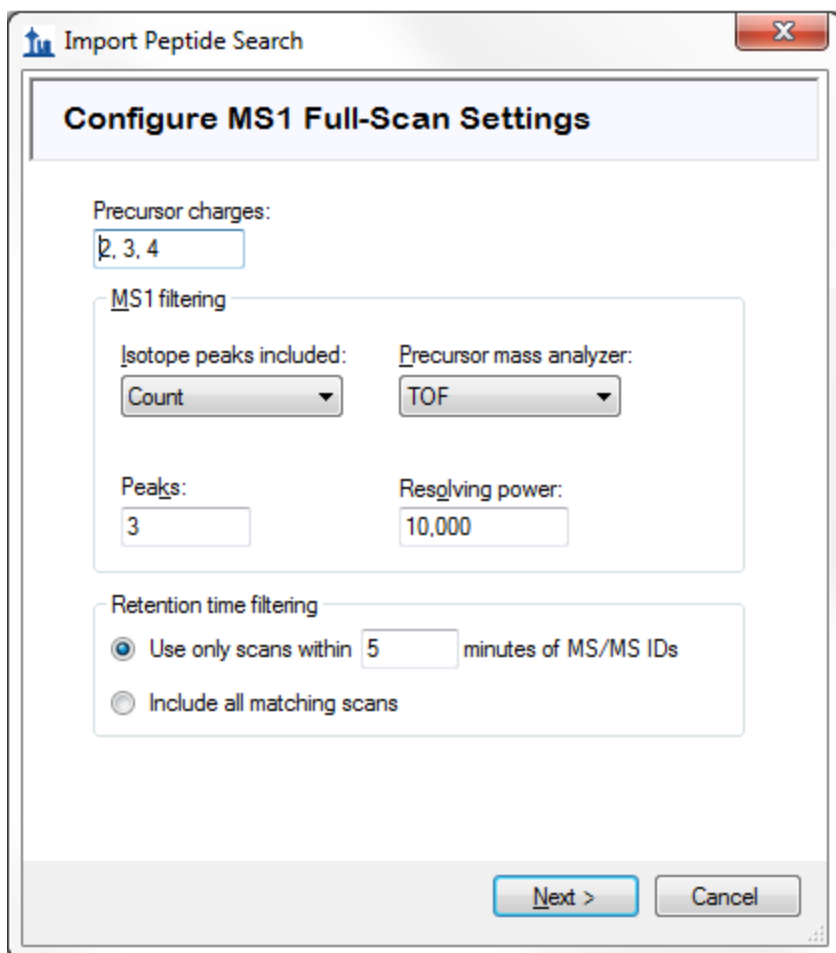
注意可能您的文档已有一个或多个这些已经定义的修饰（例如 Oxidation (M)），这种情况下列表看起来可能会有所不同。

- 单击“下一步”按钮。

向导将前进至“配置 MS1 全扫描设置”页面，您将在该页面中执行以下操作：

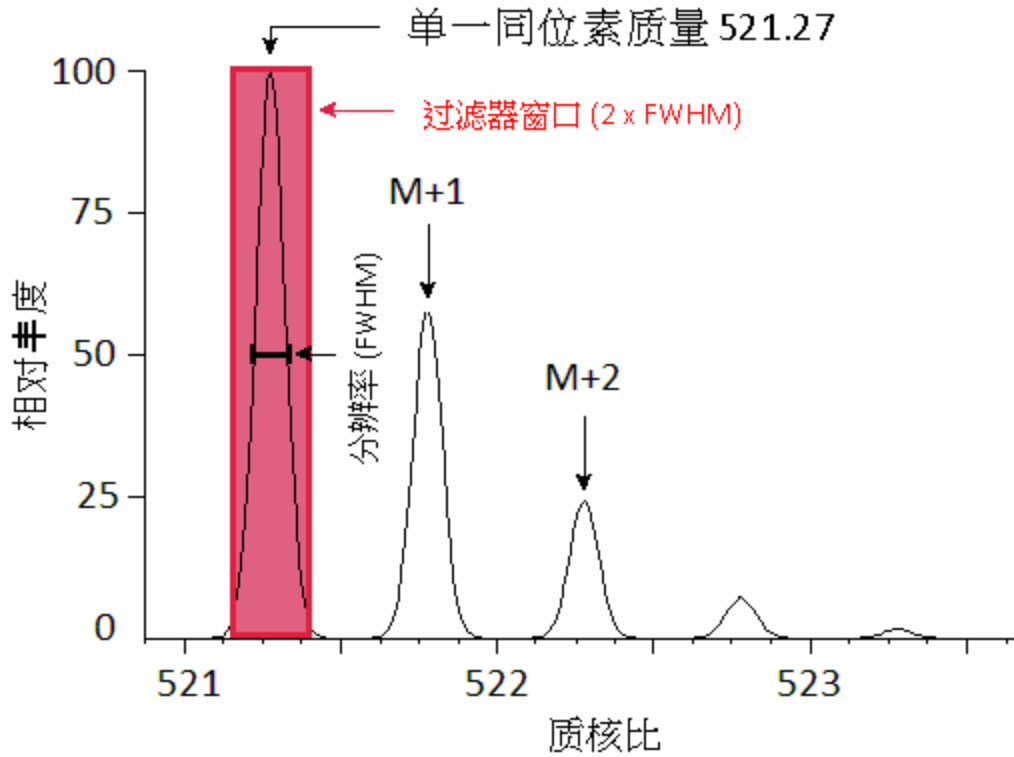
- 在“母离子电荷”字段中输入“2, 3, 4”。

本页面中所有其他字段设置为默认值，向导显示如下：

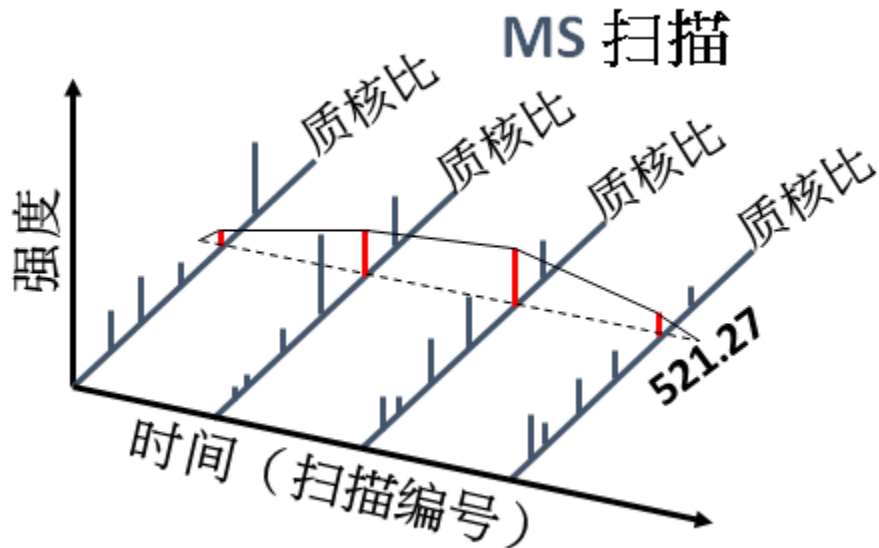


在“MS1 筛选”部分中，您将看到下列默认设置。

1. “同位素峰值包含”下拉列表中显示“计数”值。
2. “峰”字段为“3”，这将使 Skyline 从此高分辨率数据中过滤掉前 3 个同位素峰（M、M+1 和 M+2）。
3. “母离子质量分析仪”下拉列表显示“TOF”值（因为数据由 QSTAR Elite 采集的）。
4. “分辨能力”字段的默认值为“10,000”。本字段定义筛选各母离子质荷比的 MS1 筛选窗口宽度。Skyline 使用该值预测一个峰在质荷比维度上的半峰宽（FWHM），并使用 FWHM 值的两倍作为如下所示的筛选窗口。（注释：对于其他数据集和实验，可以根据仪器性能调整分辨率设置）。



在某一时间范围内提取的一系列峰将构成您在 Skyline 看到的色谱图：



“保留时间筛选”部分，注意选择“仅使用 [5] 分钟内的 MS/MS 扫描 ID”。这就意味着，对于只具有 1 个 ID 的肽段，Skyline 将围绕此 ID 提取 10 分钟的色谱图。对 3 分钟以上范围的一组 ID，Skyline 将提取 13 分钟的色谱图，即在两侧各增加 5 分钟。当某个 run 没有某一特

定肽段的 ID 结果时, Skyline 将使用其他 run 中该肽段所对应 ID 的保留时间来确定在该 run 中提取这一肽段的色谱图的时间范围。

- 单击“下一步”按钮。

此操作将进入“导入 FASTA”页面。您可以导入一个包含全部人类 SwissProt 蛋白质条目的 FASTA 文件, 以获得全面的已鉴定的肽段列表(此 MS 实验涉及人的 MCF7 乳腺癌细胞的磷酸化肽富集), 但因文件大小的关系, 我们执行下列步骤导入一个仅含 12 个人蛋白质的 FASTA 文件:

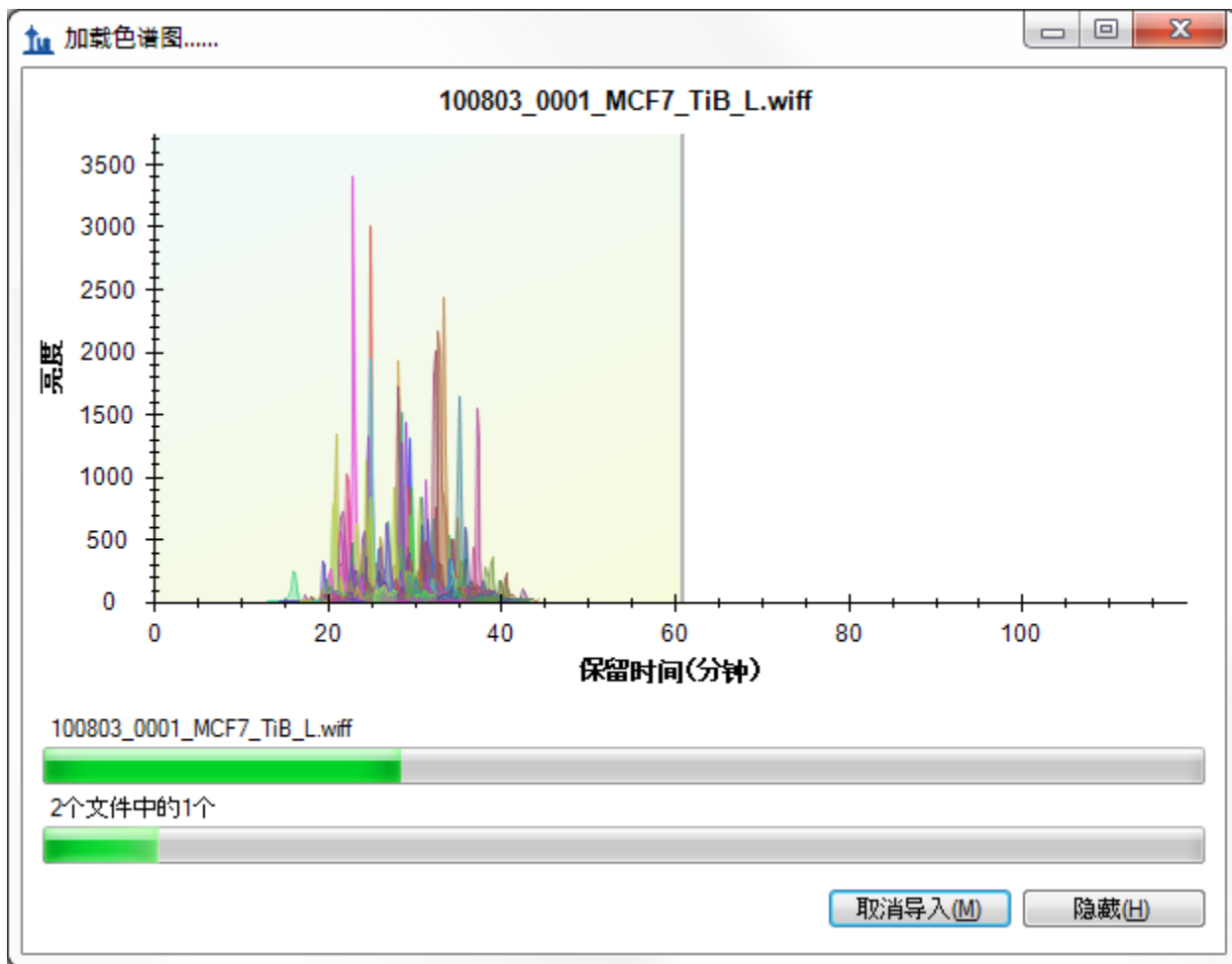
- 在“最大漏切数”下拉列表中选择“2”。
- 单击“浏览”按钮。
- 从 MS1Filtering 文件夹中选择“12_proteins.062011.fasta”文件。
- 单击“打开 FASTA”表单中的“打开”按钮。

向导现在显示如下:



- 单击“完成”按钮。

Skyline 将为所有 FASTA 文件中的肽段添加与其相匹配的谱图目标（由肽段搜索结果确定），然后导入两个 WIFF 文件，并从中提取色谱图。您将看到如下所示的进度图：



导入完成后，在检验色谱图数据前请先验证您创建的谱图库。

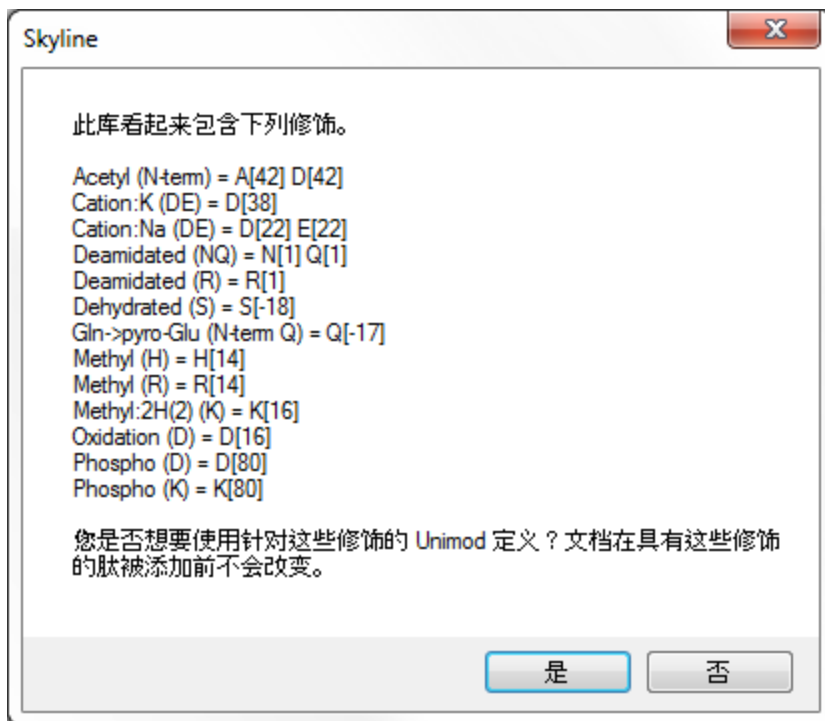
验证库保留时间信息

当您使用肽段鉴定结果构建 MS1 筛选的谱图库时，如果您未曾使用此结果完成过此类操作，您应确保谱图库里包含必须的保留时间信息，以支持下面涉及到的 Skyline 功能。使用“导入肽鉴定结果”向导的好处之一是，当您的库缺少必要的信息时，将尽早通知您。

要验证您刚创建的库包含 MS1 筛选峰提取和峰注释的保留时间信息，请执行下列步骤：

- 在“视图”菜单上单击“谱图库”。

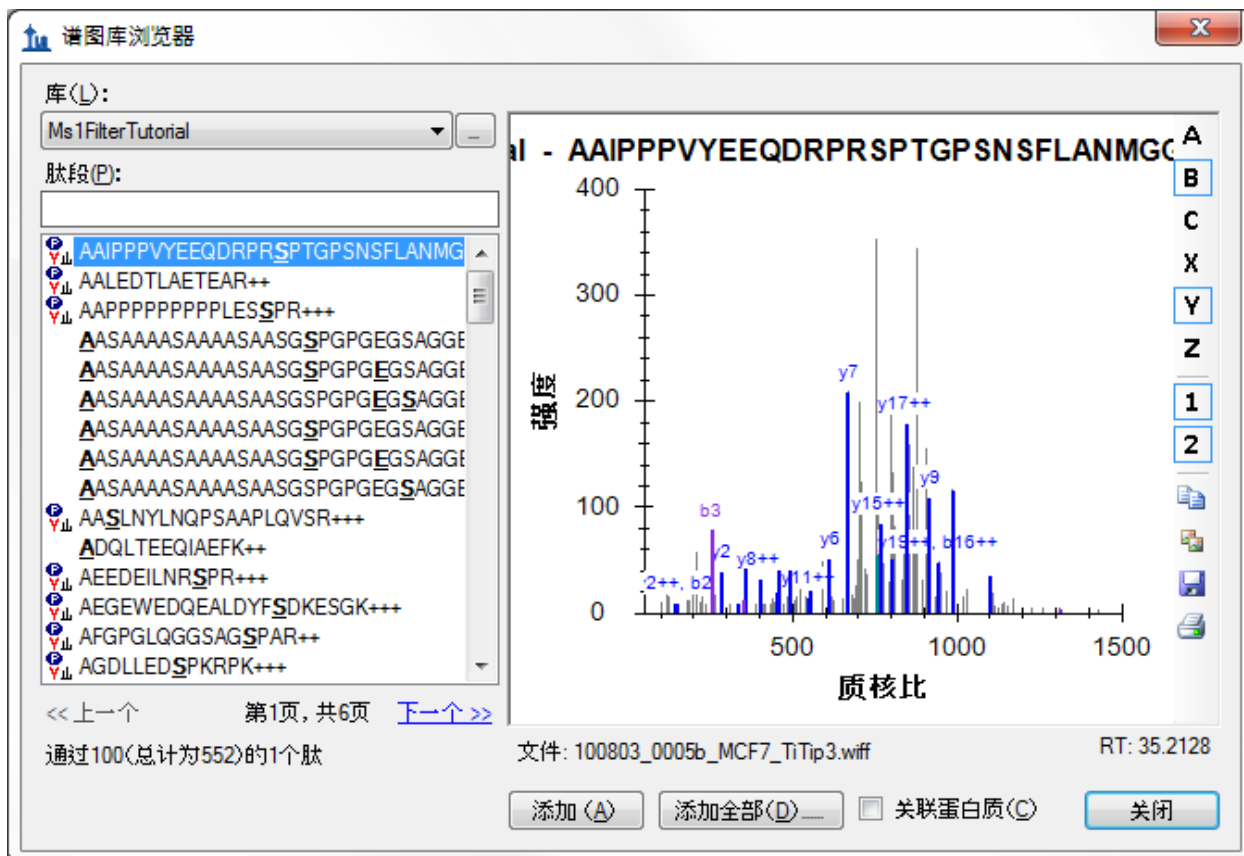
Skyline 将再次提供使用从“导入肽段搜索”向导中选择的、未加入文档的库中检测到的修饰。



现在选择在“谱图库浏览器”中使用不会将其加入目前的文档，除非您使用“谱图库浏览器”将使用这些修饰的肽段添加到您的文档。但是，这些修饰对于本教程并不重要。您可以通过下列操作不使用这些修饰而继续本教程：

- 单击“否”按钮。

将出现如下所示的“谱图库浏览器”：



在肽段列表中，序列文本左侧没有图标的肽段包含您未选择使用的任何修饰。

谱图下方，您可以看到文本“文件：100803_005b_MCF7_TiTip3.wiff”和“RT:35.2128”。

“RT”值告诉您有保留时间信息，“文件”值告诉您已经与您导入 Skyline 的文件正确关联。

“文件”值无需与您导入的文件完全匹配。通常，许多肽段搜索流程都涉及到将原始质谱数据转换为 mzXML、mzML、MGF、MG2 等格式，故 Skyline 将寻找基础名称匹配，其中“basename.mgf”与“basename.wiff”成功匹配。由于某一特定流程需要更大的灵活性，此匹配也有不敏感的情况，因此“BASENAME.mzML”将匹配“Baseline.RAW”，且在处理多点扩展时，会将

“basename.c.mzXML”与“basename.raw”匹配。然而，如果您看到一些类似“F011852.dat”或其他与您想导入 Skyline 的数据没有相同基础名称的搜索输出文件，则需要您查看您的搜索流程，并可能与 Skyline 团队合作，来解决此问题。关于具体的 Mascot .dat 文件，建议您查询 Skyline 网站的“[Mascot 搜索结果 ID 注释缺失](#)”页面。其他问题，建议您在 Skyline 支持模块（在“帮助”菜单上单击“支持”）上提问，寻求解决此类问题的帮助。

现在按向下键选择其他肽段，您将看到“文件”和“RT”值发生变化。完成检查 MS/MS 谱图及其源文件和保留时间后，执行下列操作返回到您创建的文档：

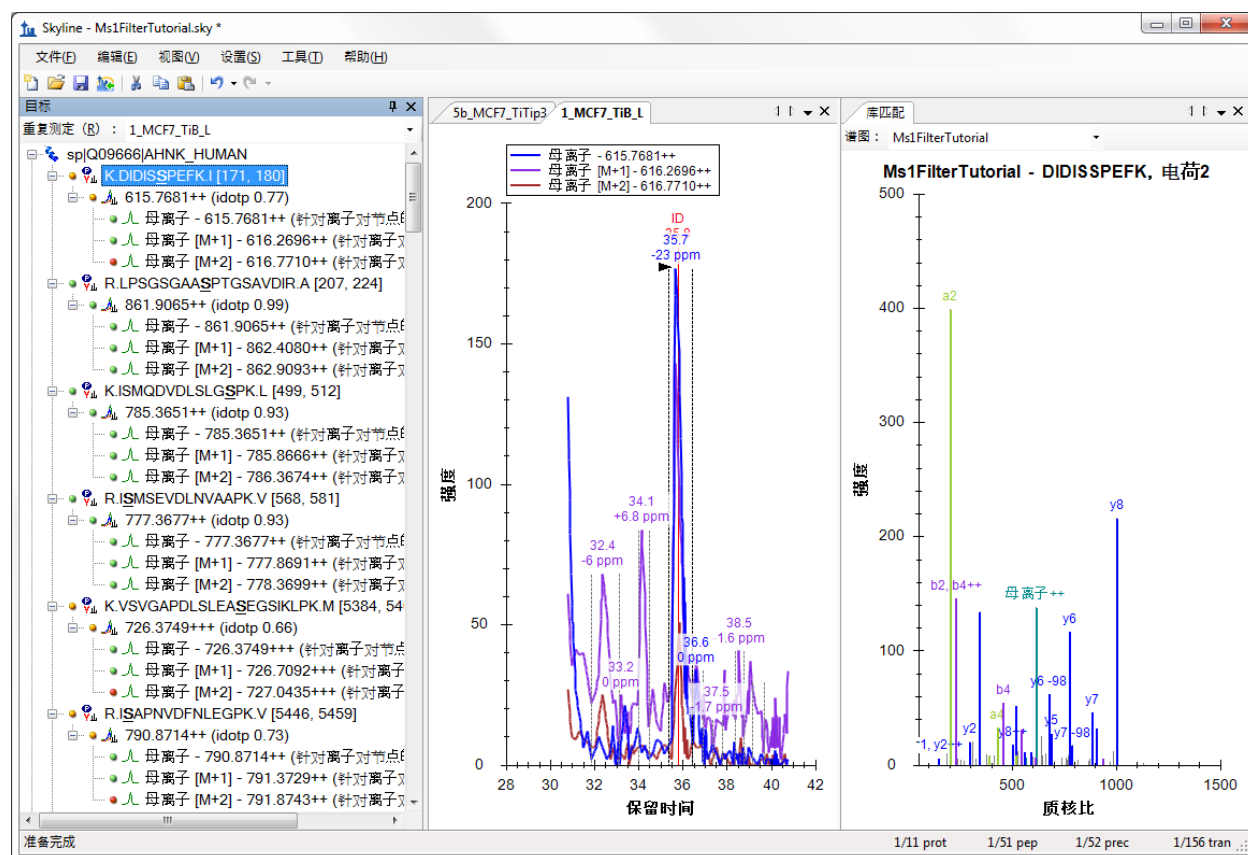
- 单击“谱图库浏览器”中的“关闭”按钮。

肽段目标、谱图和色谱图

您应在 Skyline 目标视图中看到 51 个肽段（计数见状态栏显示）。

- 单击首个磷酸肽 K.DIDISSPEFK.I 序列，将出现 MS/MS 谱图。（请注意，肽段序列中粗体、加下划线的“S”残基表示丝氨酸磷酸化）。
- 如果您未看到 MS/MS 谱图，在“视图”菜单上单击“库匹配”。
- 如果您未看到如下图所示的尽可能多的注释峰，在“视图”菜单上选择“离子类型”，并选中“A”、“B”、“Y”和“母离子”。
- 如果您未看到肽的整个色谱图，在“视图”菜单上选择“自动缩放”，并单击“无”（Shift-F11）。
- 在“编辑”菜单上，选择“全部展开”并单击“母离子”。

您的 Skyline 文档将显示如下：



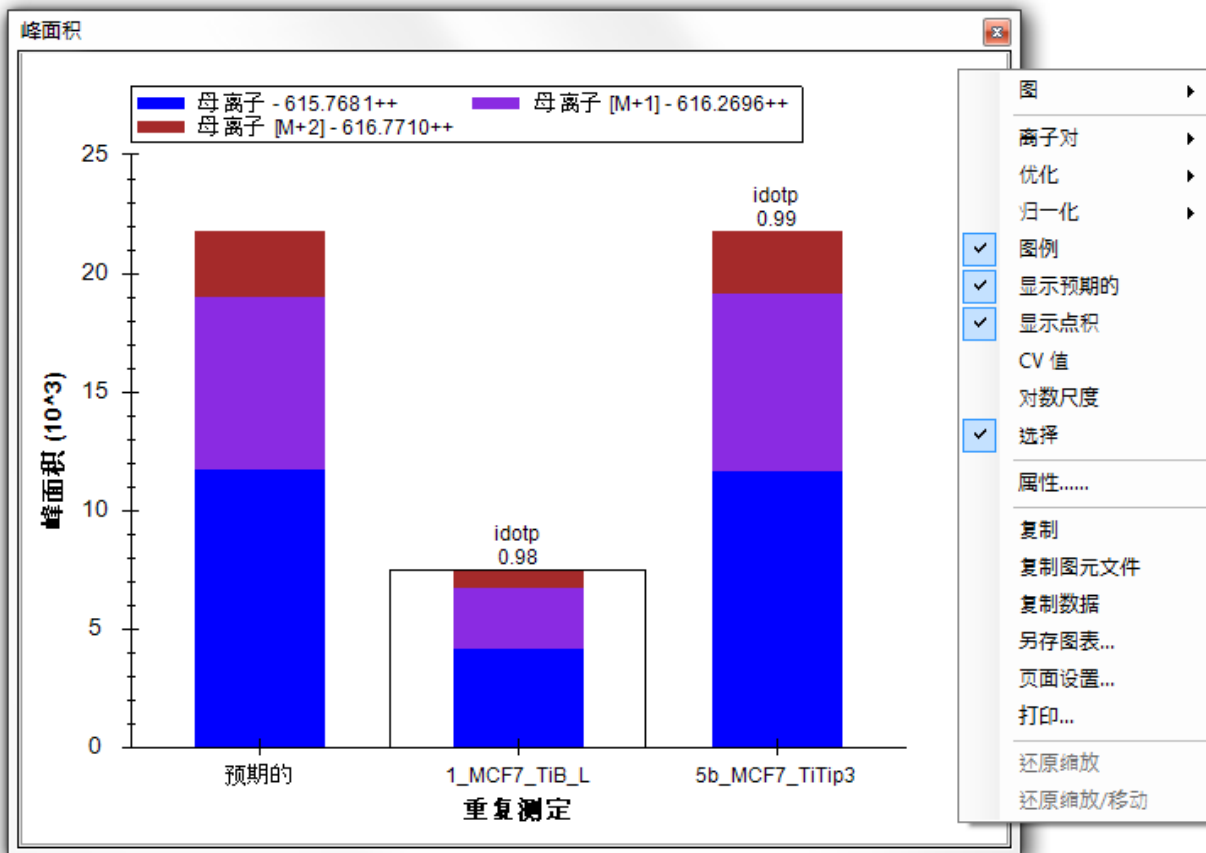
现在文档已完成 MS1 筛选已导入的两个 DDA run 所需的全部配置。由于在导入向导中选择了“仅使用 [5] 分钟内的 MS/MS 扫描 ID”设置，本视图中的色谱图长度将约为 10 分钟（31 至 41 分钟）。请注意，在针对 MS1 筛选进行 Skyline 文档设置时，您将在由三重四级杆 SRM 实验所产生的子离子离子对（例如 y-离子）的地方，看到不同的母离子同位素峰值，例如对于肽 DIDISSPEFK：母离子 - 615.7681++、母离子[M+1] - 616.2696++和母离子[M+2] - 616.7710++。

要配置一些在一般情况下有所帮助的其他功能，尤其是可视化某些 MS1 筛选数据，请执行下列步骤：

- 在“设置”菜单上，确保已选中“全部合并”。

此操作将命令 Skyline 计算一个峰组中所有色谱图（此处为母离子 M、M+1 和 M+2）的整合面积，无论峰是否与其中的最大峰共洗脱。

- 在“视图”菜单上选择“峰面积”，并单击“重复间比较”。
- 在“峰面积”窗口中单击右键，选择“归一化”，并单击“无”。
- 在“峰面积”窗口中单击右键，并选中“显示预期的”和“显示点积”（将在下面解释这两项功能）。

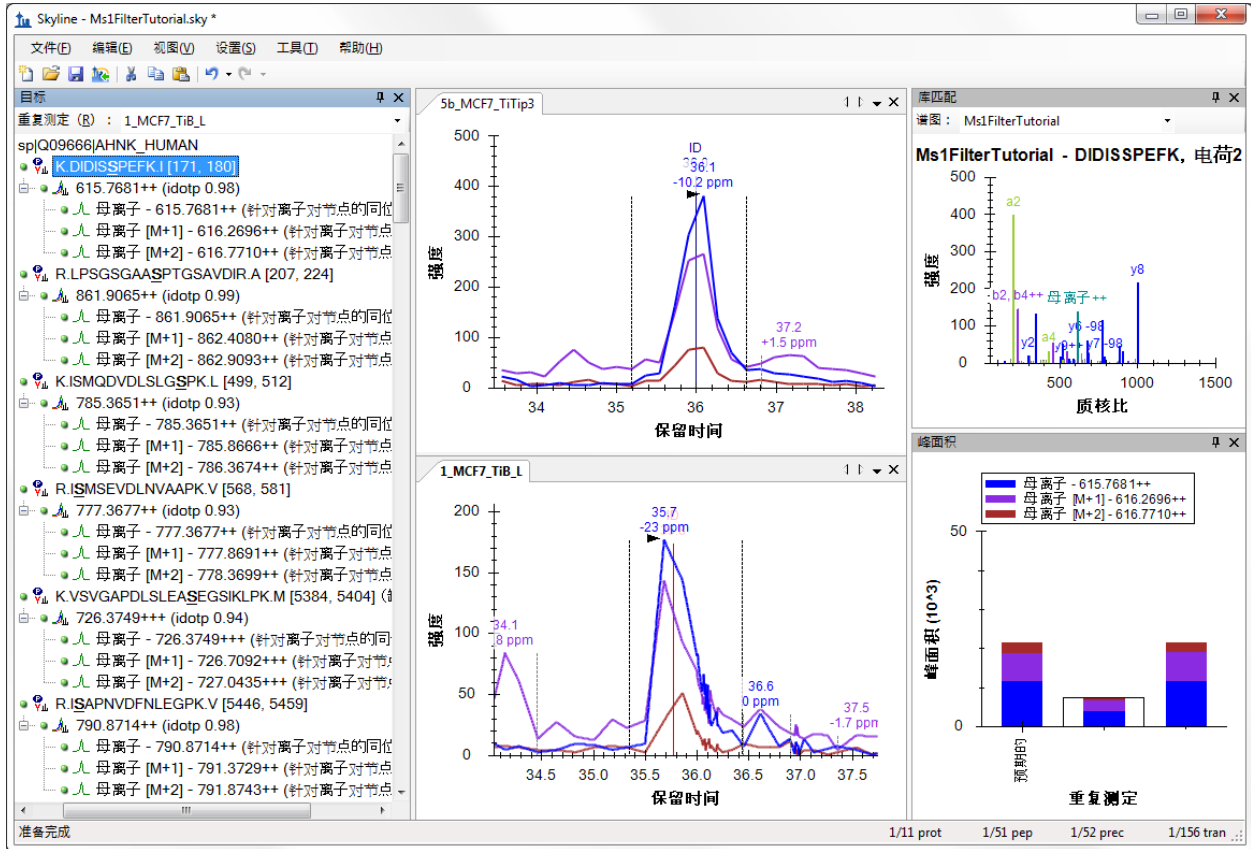


您可以通过下列操作将“峰面积”窗口停在您需要的位置：

- 单击并按住鼠标左键，然后将其拖动到您想放置的位置，也许是 Skyline 窗口的右侧边缘。
- 然后单击并拖动“库匹配”视图，并将其停在“峰面积”视图的上方。
- 在“视图”菜单上，选择“自动缩放”并单击“最佳峰值” (F11)。

- 在“视图”菜单上选择“排列图”，并单击“平铺”（Ctrl-T）。

您的 Skyline 文件将显示如下：



如果您的“库匹配”视图未正确放置，请将其移动至峰面积重复间比较视图的上方，如上所示。

色谱图视图显示所有的母离子及其同位素离子的 MS1 提取离子色谱图，包括 M（蓝色）、M+1（紫色）、M+2（棕色）。如果您使用过 Skyline 对 SRM 数据进行定量分析，可能对被选择的峰下面对应的保留时间注释较为熟悉，但是您将看到一个新的质量误差注释，这是被注释的色谱图里所有被整合的点的质量误差的加权平均值（本例中为 M 或蓝线）。如果您未看到质量误差，在色谱图视图中单击右键，并单击“质量误差”。可能这个误差并不是您所期望的现代高分辨率仪器的误差性能，但正如前面提到的，此数据源自较早的 QSTAR Elite。

您还能在提取的粒子色谱图中看到垂直线，图顶部有 ID 注释，尽管 1_MCF7_TiB_L 的垂直线在峰注释的后面。ID 代表“鉴定结果”，表示与某一特定可信肽段对应的取样的 MS/MS 谱图的保留时间。红线表示这是“库匹配”视图目前显示的谱图。如果您在上部的图中单击“ID”注释，“库匹配”视图将显示您从 5b_MCF7_TiTip3 重复中已确认的谱图，现在将保存于此您已创建的库中。您还能从“库匹配”视图窗口顶部的“谱图”下拉列表中选择重复测定的名称和保留时间（36 分钟），取代您在单击“ID”注释前已选择的非冗余库中的“最佳”谱图。您能通过单击 ID 注释或使用“谱图”下拉列表在两次采集的谱图中前后转换，可以发现他们非常相似。

查看本档中一些其他 51 个肽段前，请首先执行以下操作：

- 在“编辑”菜单上选择“全部折叠”，并单击“肽段” (Ctrl-Shift-D)。

此后，确保鼠标在“目标”视图（肽树）中，并使用向下键，每次选择一个肽段。对于前三个磷酸化肽段，您将看到它们在每个重复里都被鉴定到了一次，并且“库匹配”视图中显示它们的谱图中有少数注释为“-98” ($-\text{H}_3\text{PO}_4$) 的较为显著的中性丢失离子。

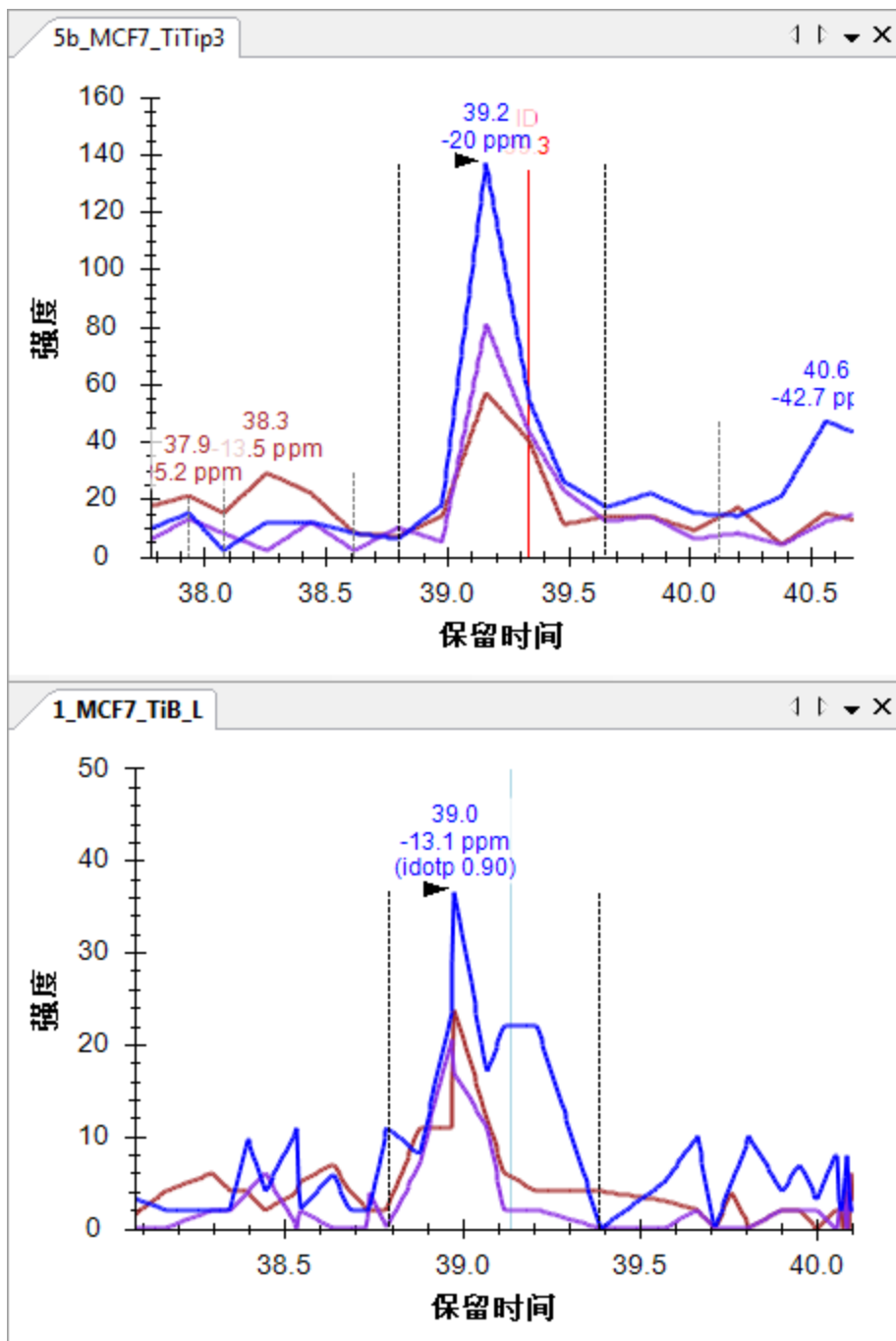
对于第四个肽 ISMSEVDLNVAAPK，您将发现仅 5b_MCF7_TiTip3 重复有 ID 注释。使用 5b_MCF7_TiTip3 中的该肽段对应 ID 的保留时间来校准 1_MCF7_TiB_L 重复中该肽段对应的峰。要查看校准的 ID，请执行以下操作：

- 右键单击一幅色谱图，选择“肽段 ID 次数”，并单击“校准的”（如果未选中）。

您应该看到一条蓝色的线出现在 1_MCF7_TiB_L 重复中整合峰的边界内侧。尽管如此，这种情况下峰很可能在边界的左侧。如要纠正这一问题：

- 在大约 38.8 分钟处单击 X 轴下方，拖动到大约 39.4 分钟处。

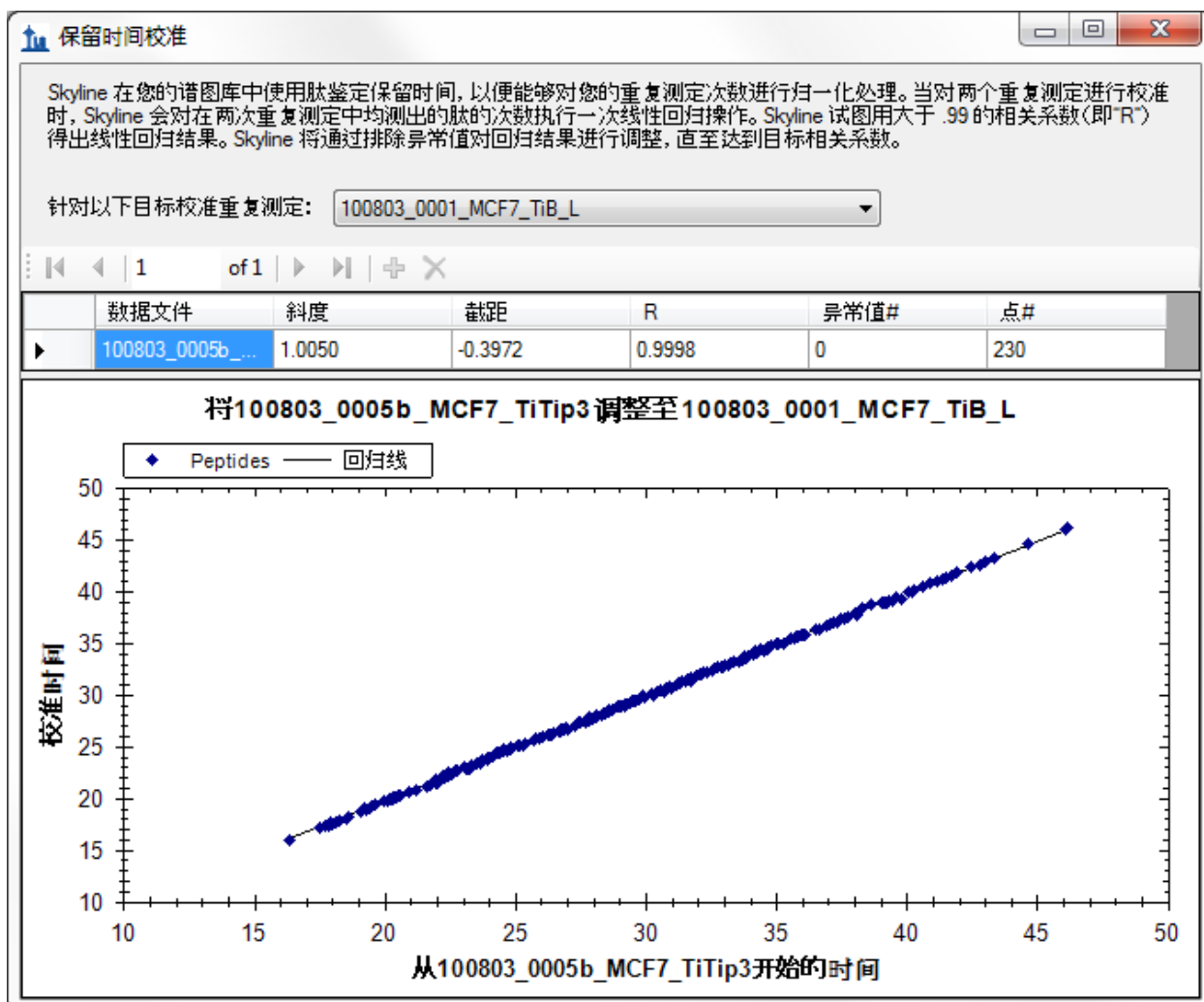
色谱图将如下所示：



为深入了解保留时间校准如何工作，请执行下列操作：

- 在“视图”菜单上选择“保留时间”，并单击“校准”。

Skyline 将呈现如下所示的窗口：



此窗口为您显示在运行间用于校准时间的线性回归。目前, Skyline 计算您的谱图库中两两谱图源文件间这样的线性回归。当出现 2 个以上 run 时, 您将看到在“校准保留时间源重复”下拉列表中出现一个以上的 run。当某个 run 中某一肽段缺少对应 ID 时, 我们就能使用这种线性回归的方法来对不同重复的 MS/MS 的保留时间进行映射, 改善峰的挑选过程。关于使用线性回归映射保留时间范围的更多详情, 请参见 [iRT 保留时间预测](#) 教程。

这种情况下, 您能看到这两个 run 间的保留时间重复性相当好, 斜率为 1.005, 截距为 -0.3972, 相关系数 (R) 为 0.9998, 无异常值。正如表单顶部段落所示, 当 R 低于 0.99 时, Skyline 将弃去异常值, 直至发现一组 R 大于 0.99 的肽段, 并使用这些肽段的保留时间所产生的线性方程。

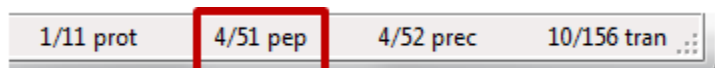
您还可能发现, 此回归的计算使用了 230 个点, 而您的文件仅包含 51 个肽段, 而且这 51 个肽段并不是被所有的 run 都鉴定到了。但是, 请记住, 您构建的库共包含 552 个肽段, 其中有许多肽段包含本文档未使用的修饰。这似乎表示两个文件中共同鉴定到了 552 个肽段中 230 个。Skyline 尝试使用所有搜索结果中都被鉴定到的肽段的 ID 来进行回归拟合。当某一肽段在某个

run 里出现多个 ID 时，Skyline 将使用最早的 ID，因为这个保留时间可能比稍晚的保留时间或甚至平均值更稳定。例如，我们发现在梯度洗脱过程中更早洗脱的肽段基本上都被鉴定到了。

- 单击“保留时间校准”窗口右上角的红色的 X，关闭该窗口。

审阅数据

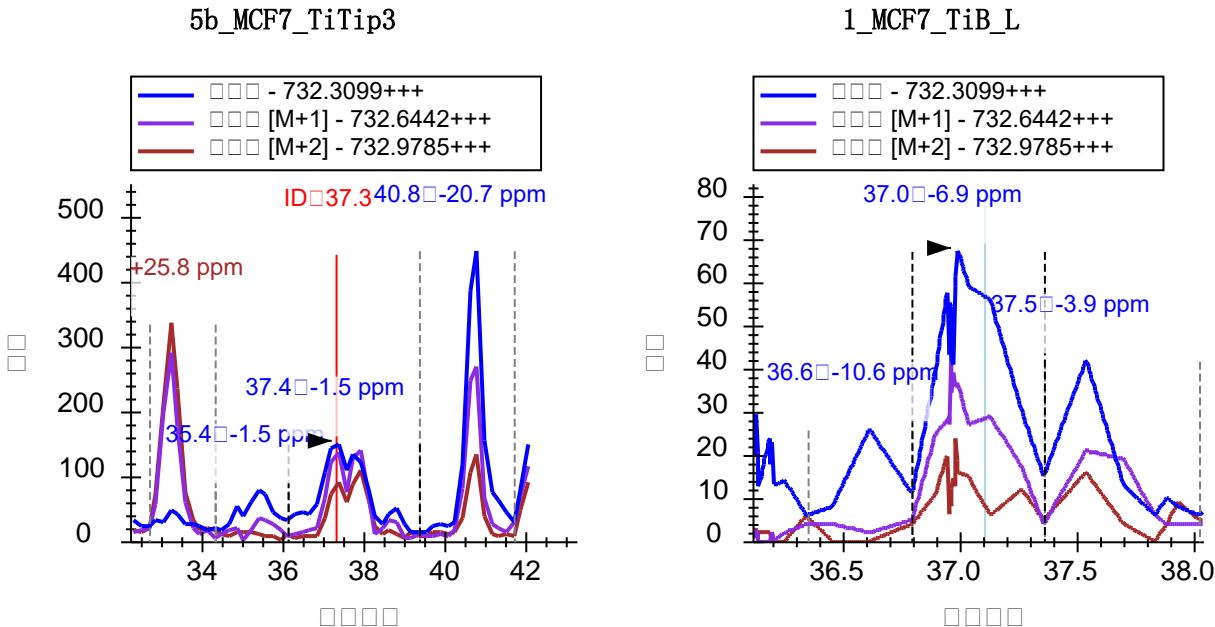
有了此基本理解，且 Skyline 按此方式配置，您现在能快速审阅本文档中的所有 51 个肽段。您只需要单击“目标”视图，并使用向下箭头依次选中每一个肽。要知道目前从 51 个肽中选择的肽段的编号，您可以查看 Skyline 窗口右下方的状态栏：



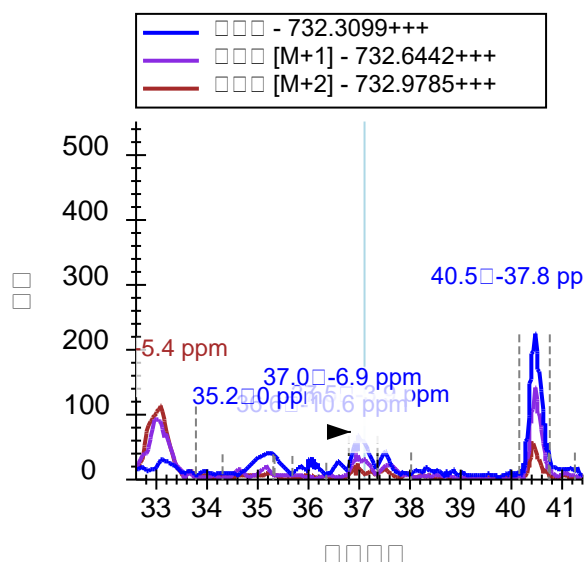
在肽段 ISMSEVDLNVAAPK 后面，您将看到有 4 个肽段，它们的峰的整合是可以接受的，虽然 VSVGAPDLSLEASEGSIKLPK 可能依据 5b_MCF7_TiTip3 略作调整。其中一部分肽段在 run 中都有 ID，另一部分肽段仅在其中一个 run 中有 ID，这里 Skyline 已使用保留时间校准挑选了正确的峰。

对色谱图的基本认识

对于第 9 个肽 SSKASLGSLEGEAEAEASSPK，它在 1_MCF7_TiB_L 中缺少对应的 ID，且峰的整合看起来有点偏：



使用您鼠标上的滚轮（朝您自己的方向向回滚动）缩小 1_MCF_TiB_L 图，直至您能看到与 5b_MCF7_TiTip3 图相同的峰：



这是处理色谱图数据的一个极其重要的方面：正如您的目标肽段可以在多个 run 中以高度相似的时间被洗脱出来，其他肽段也如此。目标（37 分钟）两侧（33 分钟和 40.5 分钟）的两个峰由两个其他肽段形成，如果这两个峰与目标肽段共同洗脱出来，则将被视为干扰。但是，如果它们不是被共同洗脱出来时，则来自于其他肽的信号就可能创建一个重复的视图，能帮助指导您定位至目标保留时间，甚至在极低的信号水平下亦能如此。使用分离度较低的方法时尤为如此，正如 MS1 筛选，因为使用色谱图提取范围内的信号您可以看到更多的肽段。

现在请通过执行下列操作，校正 1_MC7_TiB_L 的整合范围：

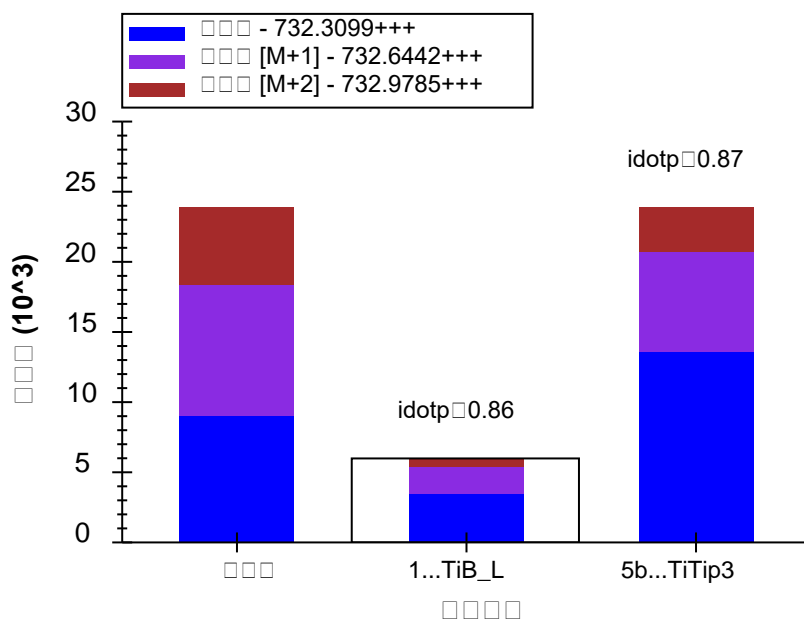
- 单击并拖动 36.5 至 38 分钟之间的保留时间轴。

您将在“峰面积”图中看到，此操作将峰的同位素分布点积值 (idotp) 从 0.87 改善为 0.9，并将质量误差从 -6.9 略微改善至 -6.5 ppm。

继续处理剩余的肽段之前，可以花些时间尝试提取的色谱图中的其他两个峰，对峰的准确性是否有所提高。40.5 分钟的峰在所有 3 个母离子 (M、M+1 和 M+2) 通道的信号都非常好，但您还能看到，3 个母离子通道中该峰的质量误差都比预期小 (-20.7 和 -37.8 ppm)。

- 单击 40.5 分钟的峰上标签。

这将使 Skyline 挑选这些峰，并在“峰面积”图中显示它们的 idotp 值低于此前选择的峰 (0.87 和 0.86，而此前为 0.96 和 0.90)：



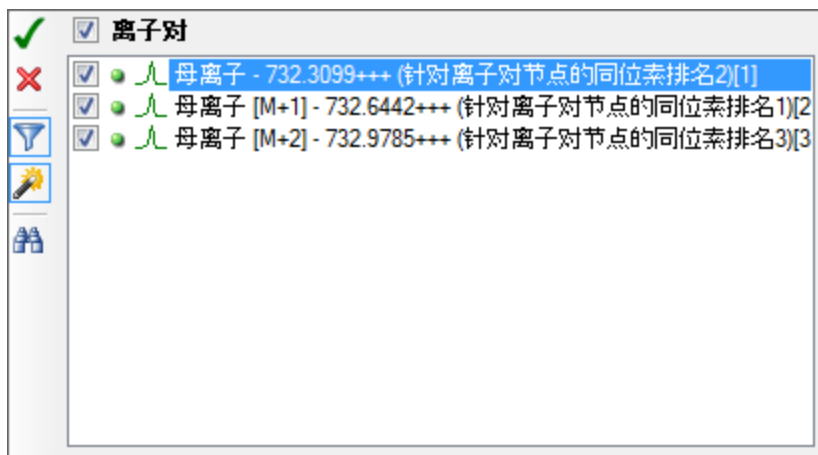
您可以从标示预测的色谱柱中看到分布，因为 M+2 和 M+3 同位峰比目标肽段的预期同位素分布强度小，这将告诉您产生此峰的肽段中碳原子比在目标肽段中的小（因此产生 ^{13}C 的机会较小），而目标肽段对就的是大约 37 分钟处的一个 ID。

转向第 33 分钟的峰，您将看到，此峰没有任何信号标明属于目标肽段的单一同位素峰，但由于有与 M+1 和 M+2 的强度非常相似的峰，其与目标肽段的 M 和 M+1 预测同位素分布相似。该肽段在 5b_MCF7_TiTip3 中的质量误差是+25.8 ppm，而当 1_MC7_TiB_L 中完全整合时质量误差为+5.6 ppm。虽然这并不像 40.5 分钟的峰那样糟糕，但平均误差达 +15.7 ppm，仍比 37 分钟峰的平均误差 -3.5 ppm 差很多。

如想更全面了解 33 分钟峰的同位素分布的问题，请执行下列操作：

- 单击“目标”视图中肽元素左侧的“+”图标以展开。
- 将鼠标光标悬停在 732.3099+++ 母离子元素上方。
- 单击母离子元素右侧的向下箭头。

Skyline 将显示如下所示的弹出窗口：



如果您仅看到这三个母离子对：

- 单击漏斗图标，删除离子对筛选。

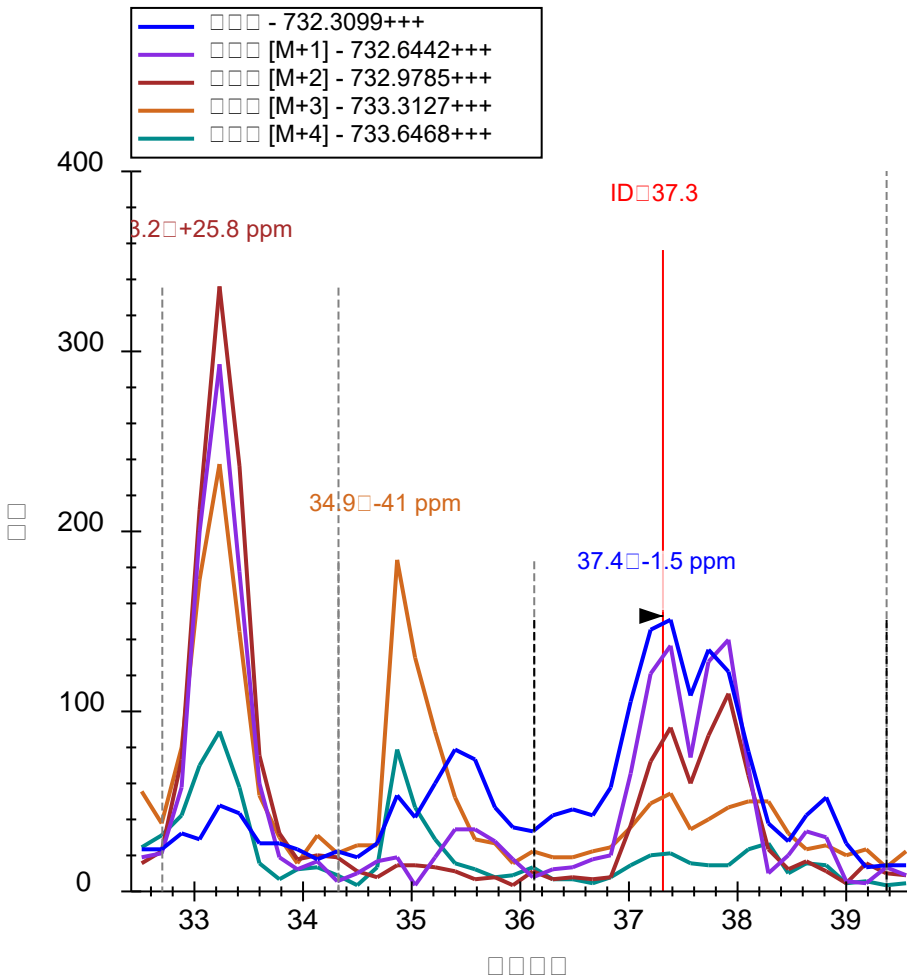
这将使 Skyline 显示此肽段母离子所有可能的离子对：



绿色虚线表示 Skyline 已获得色谱图数据的离子对。Skyline 自动提取同位素分布中至少 1% 强度的色谱图。此外，Skyline 始终提取 M-1 的色谱图，因为挑选正确的峰在这一质荷比一般不会产生信号。

- 选中 M+3 和 M+4 离子对。
- 单击左上部的绿色复选框，或按 Enter 键。

这个操作将在图中增加 M+3 和 M+4 的色谱图，您将在 33 分钟看到比 37 分钟更多的已鉴定的峰。在使用 Skyline 处理任何类型的色谱图数据时，可靠的保留时间重复性的作用相当重要，无论怎么强调也不为过。



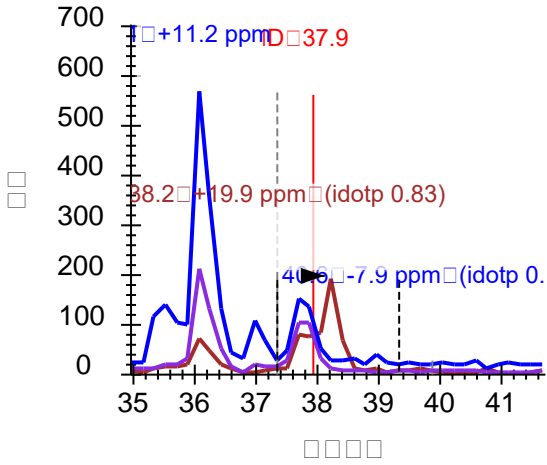
您现在可以相当确信，33 分钟时的峰确由另一种肽段产生，其原子组成与目标肽段非常相似，同时也带有三个电荷，不同仅仅在于其单一同位素峰质量比目标的大 1 Dalton。

- 单击撤销操作键 (Ctrl-Z) 3 次，直至返回原始整合校准。

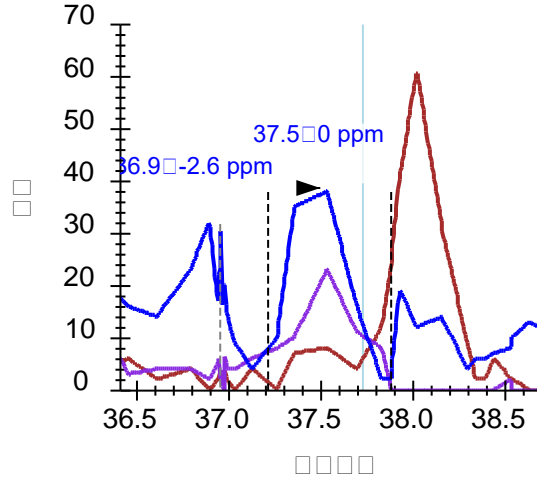
检测并解读干扰

接下来，有了这些能解读 Skyline 中色谱数据的工具，您不必花费太大力气就可以找到被真正干扰的第一个肽段。您将看到双磷酸化肽 ASLGSLEGEAEAEASSPPKGK 的色谱图如下所示：

5b_MCF7_TiTip3



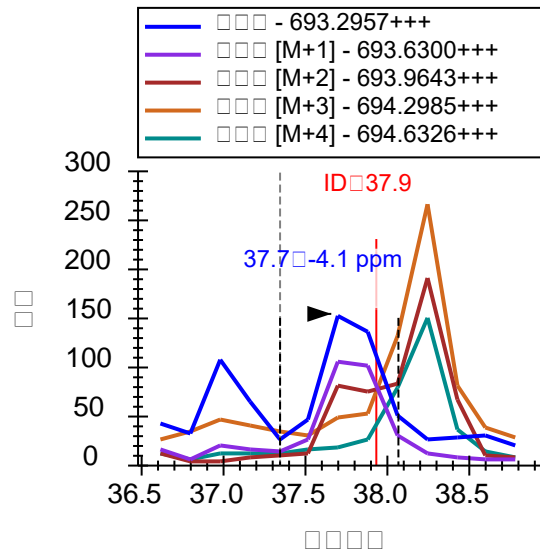
1_MCF7_TiB_L



同样地，5b_MCF7_TiTip3 有该肽段对应的 ID，但在 1_MCF7_TiB_L 中没有。我们将依据 5b_MCF7_TiTip3 中的 ID 校准 1_MCF7_TiB_L 的峰。M+2 色谱图中该峰显示出几乎不受其右侧峰的干扰，丰度最大的峰质量误差为 0 ppm。如果您使用鼠标滚轮再次缩小，您将看到两图均在 36 分钟含有一个非常相似的峰，质量误差分别为 +11.2 和 +9.6 ppm，idotp 值分别为 0.78 和 0.76（您可从“峰面积”视图中看到，单击保留时间注释选择峰）。

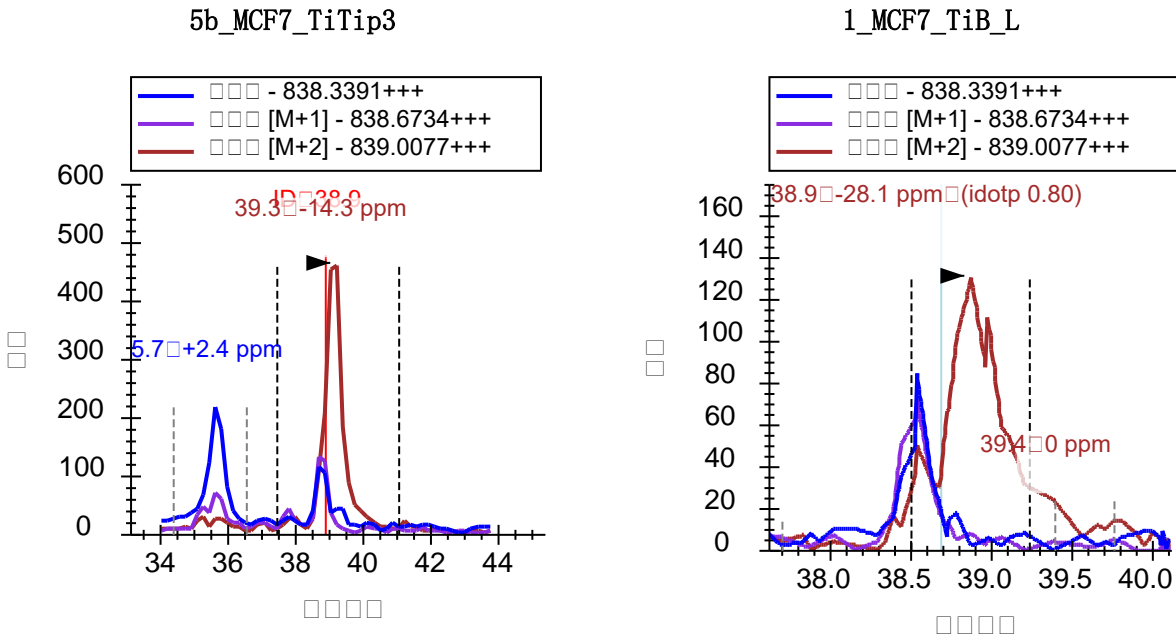
5b_MCF7TiTip3 的整合边界实际包括 M+2 的干扰，事实上本色谱图中的其他峰足够接近，以至于即使非常小心的人工整合，也似乎不太可能完全排除其信号。如果您经过尝试，您可以用 0.94 idotp 和 -4.1 ppm 质量误差得到整合的峰。

您可以使用相同的技术添加 M+3 和 M+4，并发现干扰峰可能由另一个带 3 个电荷和质量大 2 Dalton 的肽段产生。

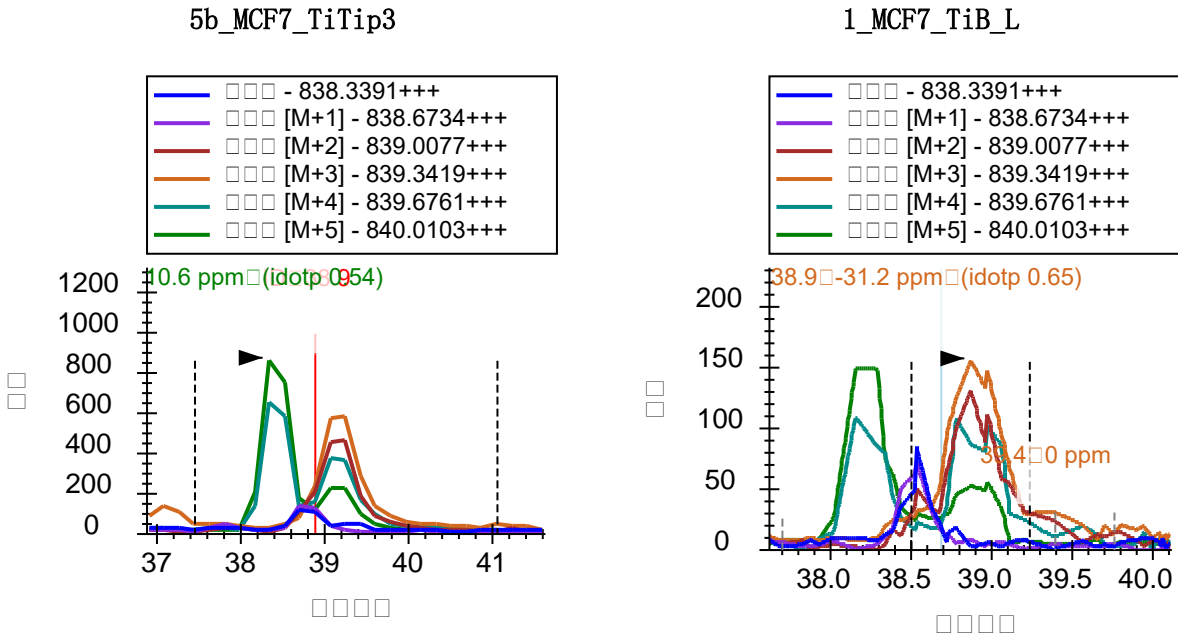


- 单击“撤销”按钮恢复此变化，并继续探索。

在肽段 AEGWEDQEALDYFSDKESGK 处，您将发现一个更强的干扰，其信号甚至更难排除。



为此肽段添加 M+3、M+4 和 M+5 色谱图，可以看到在母离子空间内的此特定质量和保留时间结合的非常紧密：



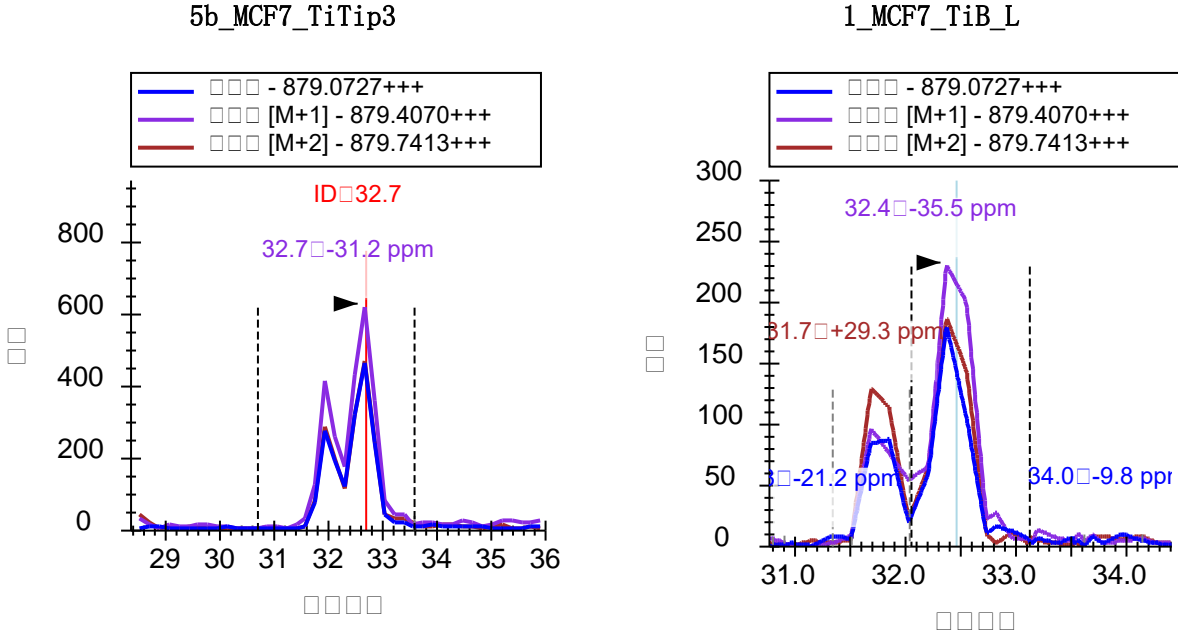
为获得此肽段无干扰的整合峰，您必须删除除 M 和 M+1 之外的所有色谱图。现在执行此操作，然后适当调整整合边界。到目前为止，您可能希望得到具有更大分离度的方法，但您实际上仅通过 MS1 扫描就得到了许多有用的定量数据。您希望尽可能地得到排名最高且没有明显干扰的母离子以便于定量统计分析。在已知可信的峰的鉴定的情况下，此操作可以最大程度限制您从此教程数据中看到的这种干扰的影响。

继续审查肽段，当您到达第 22 号肽段 ALVEFESNPEETREPGSPPSVQR 时，您需要对整合峰进行一个小的调整。

肽的不同修饰形式

此处您将发现文档包含 ALVEFESNPEETREPGSPPSVQR 和下面的 ALVEFESNPEETREPGSPPSVQR，两者的母离子质荷比都为 879.0727。本例中，搜索引擎 Protein Pilot 鉴别到前者出现在 5b_MCF7_TiTip3 中，后者出现在 1_MCF7_TiB_L 中，但色谱图清晰地显示两者鉴定到的峰均出现在大约 32.5 分钟附近。

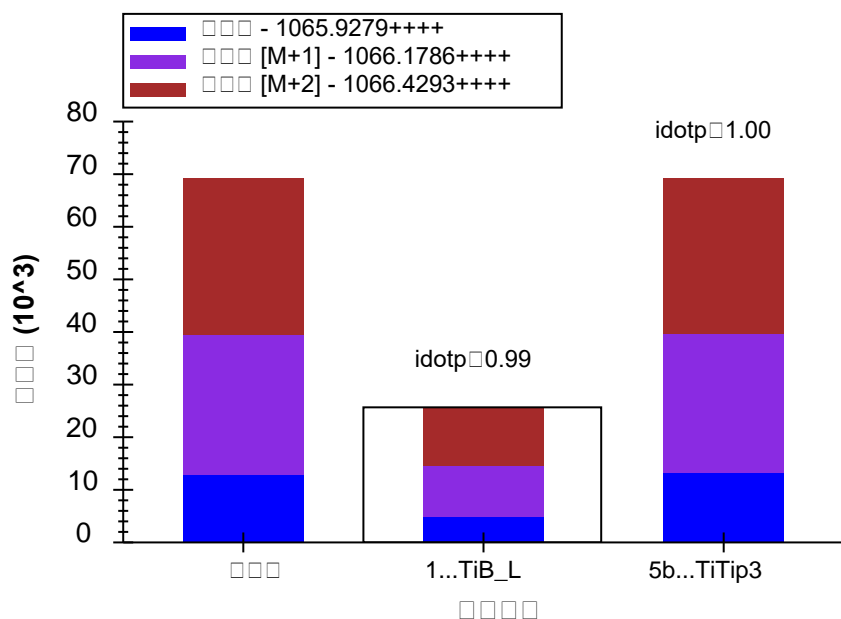
更为有趣的是，您可以看到两个峰实际上非常接近，有相同的质荷比，至少同位素分布非常相似。



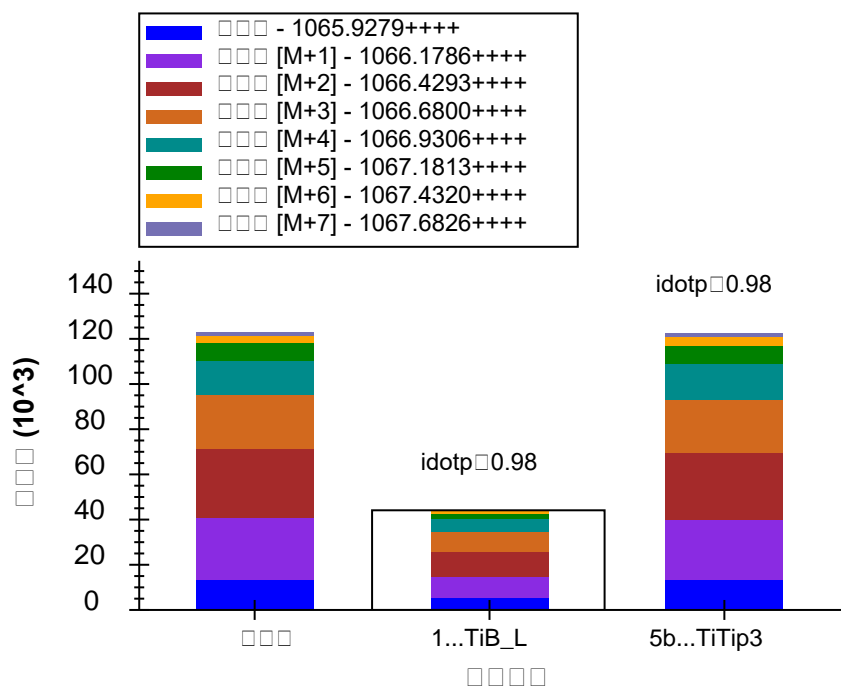
“1_MCF7_TiB_L”中，同位素分布和质量误差使两个峰看起来比 5b_MCF7_TiTip3 中的两个峰更不同，尽管它们在两个实例中均出现在 31.5 至 33 分钟之间，但这种差异可能仅由方差导致。增加 M+3、M+4 和 M+5，您能看到两个峰保持大于 0.9 个 idotp 值（再次分别整合和看“峰面积”视图，并“撤销”）。由于此肽有 4 个不同的可能磷酸化位点，两个峰可能是同一个肽的不同单一磷酸化状态，或磷亚型可能有共洗脱情况。建议在 MS1 筛选中（搜索引擎输出之外）详细评估潜在的异型体。

更多数据分析工具

继续向下至第 25 个肽段，您将发现 YGPADVEDTTGSGATDSKDDDDIDLFGSDDEEESSEEAKR，本文档中最长的首个带 4 个电荷的肽段母离子。由于该肽段分子量大，其同位素分布与分子量较小的、带 2 个电荷的肽段差异很大，甚至与您一直关注的分子量较大、带 3 个电荷的肽段差异也很大。现在不含 ^{13}C 原子的单一同位素肽段，其期望出现的频率小于 M+1 和 M+2 型离子。您可以看到这些色谱图中发生这种情况，其产生的 idotp 值分别为 1.0 和 0.99，且与预期分布一致，如下所示：



正如您一直做的，使用离子对列表，您能挑选并增加从 M+3 至 M+7 的色谱图，所有的色谱图都应包含强度大于整体同位素分布 1% 的峰，并看到 idotp 值仍非常高，为 0.98：

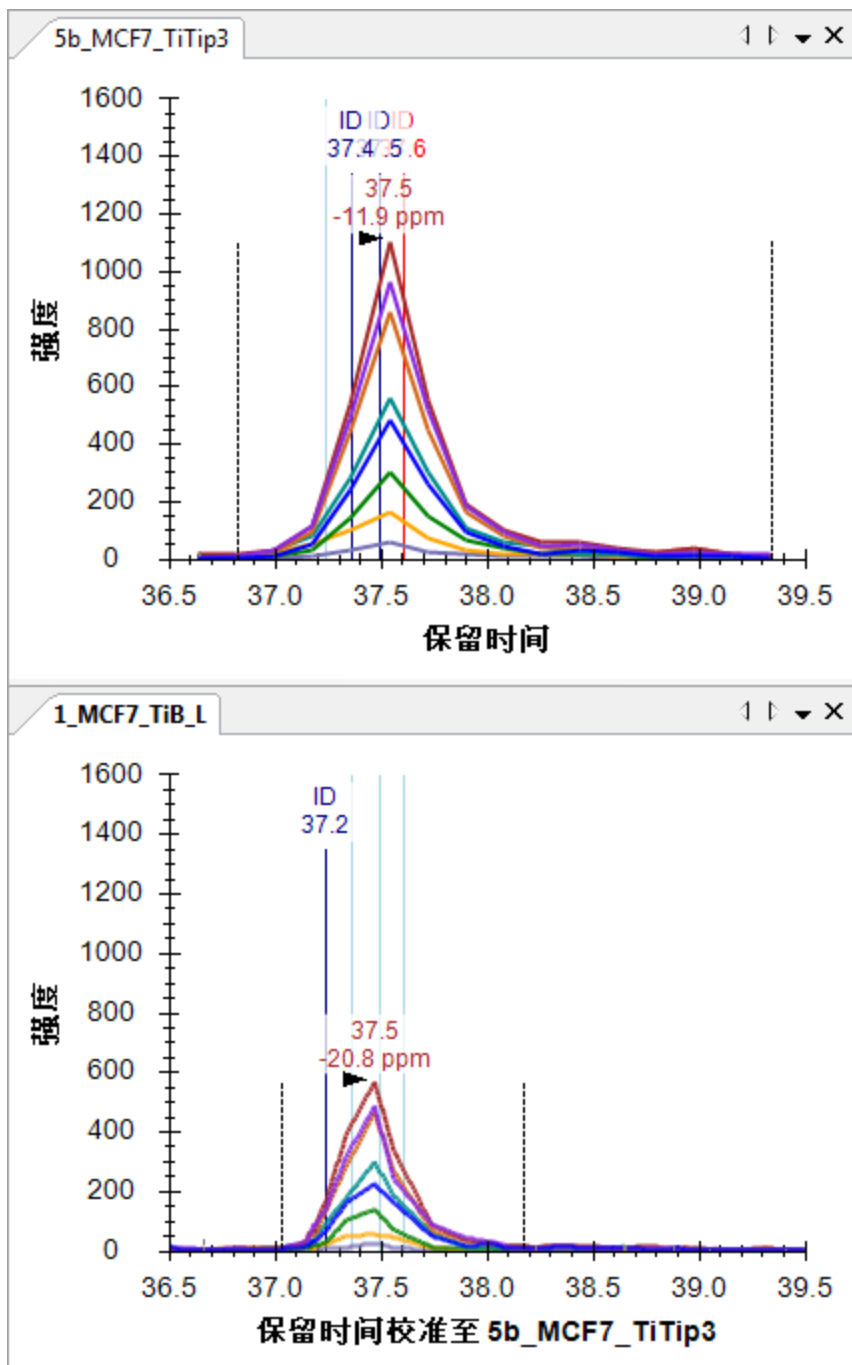


在色谱图中，您可以发现，这是文档中唯一在单个 run 中被鉴定到最多次的肽段（在 5b_MCF7_TiTip3 被鉴定到 3 次）。

您可以执行下列操作，将色谱图调整在相同的尺度内，更容易解释在重复间如何校准这些 ID：

- 右键单击色谱图，然后选中“同步缩放”。
- 右键单击色谱图，然后取消选中“自动调整 Y-轴”。
- 右键单击 5b_MCF7_TiTip3 色谱图，并选中“将时间校准至 100807_0005b_MCF7_TiTip3”。（请注意，在再次关闭之前，此操作将校准本数据集的所有肽段，而不仅仅是当前的肽段。）
- 将鼠标光标悬停在绘图区，朝您自己的方向向回滚动鼠标滚轮，略微缩小图。
- 单击并拖动 5b_MCF7_TiTip3 中整合范围周围的一个窄的矩形区域。

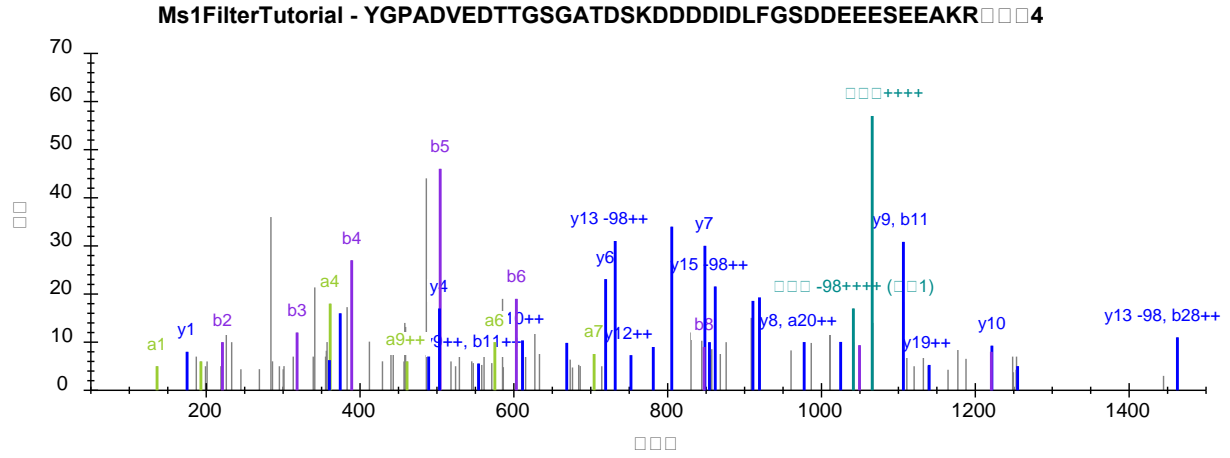
色谱图现在将显示如下：



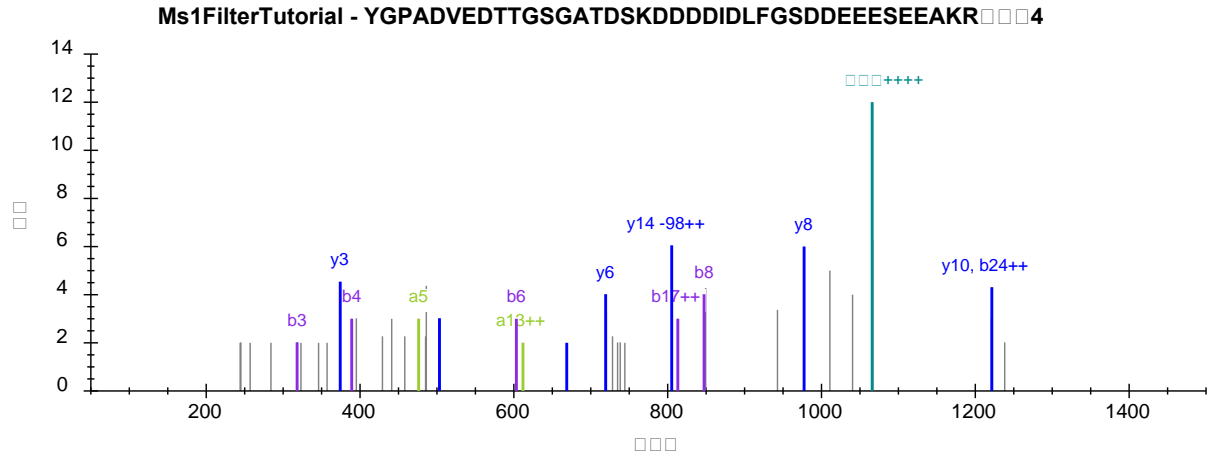
您能看到这一操作使 ID 排列非常清晰，并使峰校准良好。在关闭 Y 轴自动调整的情况下进行同步缩放，可以让您感受到各个峰的相对高度。

您可以单击色谱图上的 ID 注释，审查搜索引擎鉴定到此肽段的谱图，或您可以单击“库匹配”视图顶部的下拉列表，并使用箭头键对匹配的谱图向上翻页和向下翻页。对相同肽段的不同 run 之间的谱图进行判断可能需要一些想像力。

5b_MCF7_TiTip3 (37.61 分钟)



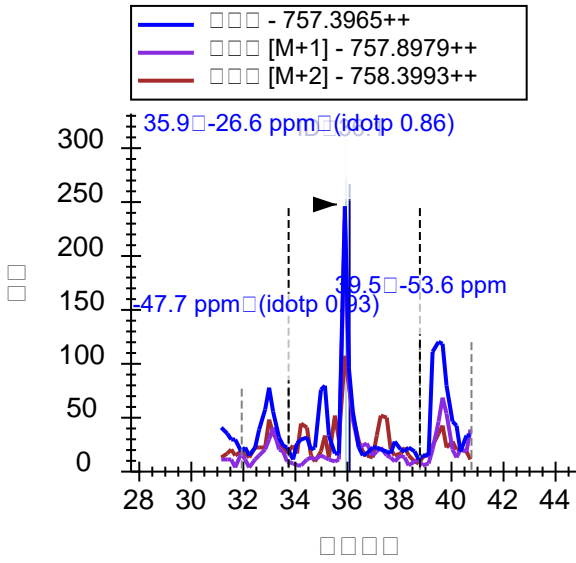
1_MCF7_TiB_L (37.03 分钟)



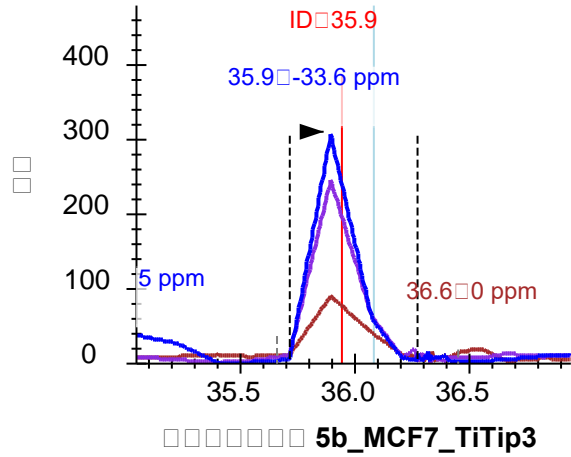
但是您应该非常确信，两个 run 之间的色谱图峰都能测量相同的肽段分子。

继续向下选择至第 27 号肽段 GVVDSEDLPLNISR，您将发现峰的整合需要调整：

5b_MCF7_TiTip3



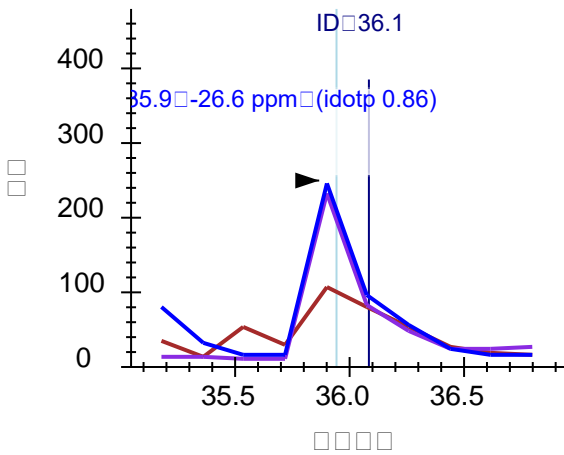
1_MCF7_TiB_L



开启“同步缩放”，您可以放大查看边界相距很远的峰：

- 单击 1_MCF7_TiB_L 色谱图绘图区域。
- 在任一方向移动鼠标滚轮。

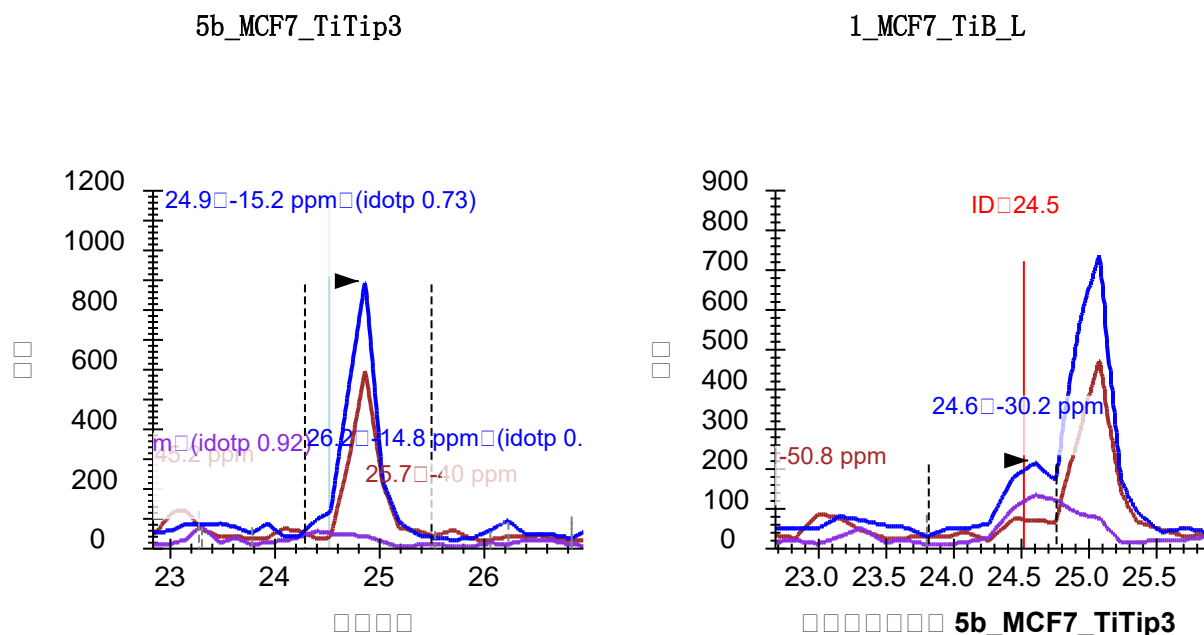
将 5b_MCF7_TiTip3 图缩放至与 1_MCF7_TiB_L 相同的尺度：



此操作将允许您通过单击和拖拽 35.7 至 36.5 分钟保留时间轴的下方从而轻易地重置整合边界。峰面积视图中，您将看到两个 run 的峰面积现在共约 8,000 至 10,000，且 5b_MCF7_TiTip3 的 idotp 值已从 0.86 上升至 0.97。

与干扰有关的更多有趣实验

下一个肽段 DQVANSAFVER 有另一个有意思的干扰肽段：



在 1_MCF7_TiB_L 中，该肽段在 24.5 分钟被鉴定到。然而，在两次重复中，您将看到在大约 25 分钟时出现强干扰峰。尽管此干扰峰仅出现在 M 和 M+2 色谱图中，但它仍表示这是一个干扰肽段。由于目标肽段带 2 个电荷，这将告诉您干扰肽段带 1 个电荷。而 5b_MCF7_TiTip3 中的目标肽段信号非常弱，且干扰非常强，因此很难看到目标峰，甚至在 M+1 色谱图中也是如此。

在两个重复中的都有这种严重干扰的情况下，我们应非常仔细确定错误峰提取的边界，并从 MS1 量化中去除此肽。如果您确实希望测量这种特定的肽段，您可能需要使用一种分离度更大的方法，比如靶向 MS/MS 或 SRM。

本文件中剩下的 7 个问题大多您已看到过，您现在希望的是使用好的工具理解并解决这些问题，下面将枚举那些关注的、但无法忽视的问题，但您可以随时忽略并继续下一部分：

1. ETERASPIKMDLAPSK (31 & 32) - 在同一的蛋白质重复出现两次（删除一次）
2. KTGSYGALAEITASK & KTGSSYGALAEITASK (34 & 35) - 对应峰给出两个磷酸化位点
3. TPSPKEEDEEPESPPEKK (41) - 效果不好的色谱图整合方法漏掉了应该包含的峰（缩放，调整整合）
4. KEKTPELPEPSVK (46) - 在同位素峰 M+1 和 M+2 上有干扰
5. EKTPELPEPSVK (47) - 在同位素峰 M+2 上有干扰

6. VPKPEPIPEPKESSPEKNSK & VPKPEPIPEPKESPEKNSK (49 & 50) - 峰给出两个磷酸化位点
7. KETESEAEDNLDDLEK (51) - 带 1 个电荷的肽段在同位素峰 M+1 上有干扰

将色谱图缓存文件最小化

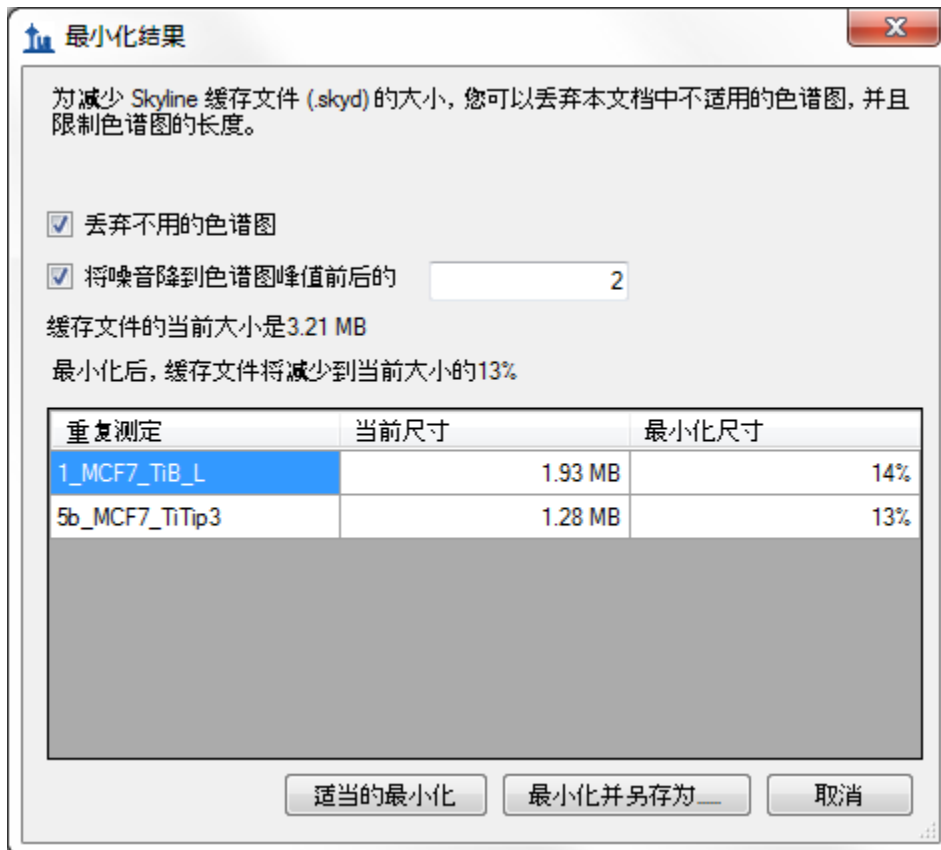
完成以上步骤并将现在的 50 个肽段加入您的文档后，这些肽段的峰将被整合得相当好。继续下一步操作前，保存当前的文档：

- 在“文件”菜单上单击“保存”(ctrl-S)。

随后，执行下列步骤，消除本文档多余的色谱图数据，使文件尽可能小以便于共享：

- 在“编辑”菜单上单击“管理结果”。
- 单击“最小化”按钮。
- 选中“将噪音限制在色谱峰前后的”复选框。
- 在该字段中输入“2”，指定“将噪音限制在色谱峰前后”的分钟数。
- 按 Tab 键进入重新评估文件大小减少量的表单。

“最小化结果”表单现在将显示如下：



表单显示，预期此操作可将缓存文件的尺寸从大约 3.24 MB 减少为当前尺寸的 13%，即大约 420 K。

- 单击“最小化并另存为……”按钮。
 - 在“另存为”表单的“文件名”字段中，输入名称“Ms1FilteringTutorial-2min.sky”。
1. 单击“保存”按钮。

如果您再按下 Shift-F11 进行缩小，您可以查看新文档中肽段的色谱图，并看到色谱图仅在超出整合边界两个方向各 2 分钟的范围内延伸。

色谱图最小化在为大型实验创建文档时可能非常有用，您可以将其当作某一文章的附件数据来与别人共享。您可能还希望可以在线得到原始数据，但最小化的 Skyline 文档将以很小的下载成本为您提供数据的丰富视图。

用列表法输出用于 MS1 筛选的内含物

如上所述，使用 DDA 的多重复研究显示对 MS/MS 的采样过疏，而且并非所有肽在每个重复中都有 MS/MS 鉴定。MS1 筛选通过此前描述的 RT 校准方法克服此问题。然而，当您想从初始的纯发现物中了解到有多少甚至是相当数量的肽段是目标肽段，您能使用 Skyline 为您的 DDA 实验导出一份内含物列表，实现一种被成为“准确内含物质量筛选”的方法²。内含物列表输出应增加您对关注的肽段在非定向 DDA 实验里采样成功的机会。

目前，Skyline 可以为 AB SCIEX 和 Thermo 仪器导出内含物列表，我们正在使用的是 Agilent 和 Waters。为了从教程的 Skyline 文档中为后续的 MS1 筛选导出一份内含物列表，您将执行下列步骤：

- 在“设置”菜单上，单击“肽段设置”。
- 单击“预测”选项卡。
- 选中“在出现时使用实际检测到的保留时间”复选框。
- 在“时间窗口”字段中输入预期色谱稳定性的适当窗口值，例如“10”分钟。减少窗口重叠的在 SRM 中更为重要。
- 单击“确定”按钮。
- 在“文件”菜单上选择“导出”，并单击“方法”。
- 在“仪器类型”下列列表中选择“AB SCIEX TOF”。
- 在“方法类型”下拉列表中选择“预定的”。
- 在“模板文件”字段中输入 QSTAR 系统采集方法模板文件的路径……

就是说，您能尽可能多地使用本教程的仪器方法导出，除非您拥有已安装供应商支持（AB SCIEX Analyst 或 Thermo Xcalibur）的软件系统。对于来自 Skyline 的所有导出方法，建议您在计划运行您的方法的仪器的控制电脑上运行 Skyline 并执行导出功能。因为，即使您的实验室拥有一

台能支持 Skyline 的仪器，您仍有可能不会在这台仪器上执行本教程的操作，当您需要时应由您完成上述步骤。

结论

通过本教程您已经学习了如何使用 Skyline 从您的 DDA 实验产生的 MS1 数据中提取定量信息的最基本和最重要的功能。幸运的是，大多数此前已有的 Skyline 功能仍同样适用于 MS1 提取的色谱图和最初为其设计的 SRM 色谱图。因此，建议您花大量时间理解其他 Skyline 教程和教学视频中演示的材料。虽然使用从 MS1 数据提取色谱图峰面积的想法由来已久，但是 MS1 筛选对于 Skyline 而言仍是一个相对较新的功能。因此，您可以预期 MS1 筛选会继续并且快速地得到改善。

补充

如果您希望查看原始 MS1 筛选文献中实际处理的数据集¹，请访问下列链接，您可以从中下载如上所述的最小化文档：

http://proteome.gs.washington.edu/supplementary_data/MS1_Filtering/minimized/

您可以浏览上一层目录以查看完整的 Skyline 文档和原始数据。

参考文献

1. Schilling, B. *et al.* Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline APPLICATION TO PROTEIN ACETYLTATION AND PHOSPHORYLTATION. *Mol Cell Proteomics* **11**, 202 - 214 (2012).
2. Jaffe, J. D. *et al.* Accurate Inclusion Mass Screening. *Mol Cell Proteomics* **7**, 1952 - 1962 (2008).