Skyline MS1 全扫描筛选

Skyline 靶向蛋白质组环境为您提供您导入 Skyline 文档的原始质谱仪数据的信息可视化显示。这些显示允许您通过执行优化测量的肽段和离子对以及调整整合边界等任务来实现对数据进行操作。Skyline 最初被开发用于对选择反应监测（简称 SRM；亦称为多反应检测，简称 MRM）质谱法进行定量分析，现在它的应用范围已经扩大到从 MS1 扫描中提取时间-强度色谱图，以用于涉及数据依赖型 MS/MS 质谱分析运行的肽段定量实验。

Skyline MS1 全扫描筛选1 支持从以数据依赖型采集 (DDA) 模式操作质谱仪的发现型蛋白质组学实验导入数据集。导入原始数据后，新的和此前已有的 Skyline 功能可以在多次重复测定采集中量化肽段母离子 MS1 信号。

本教程将涵盖下列对于有效使用 Skyline MS1 筛选至关重要的方面：

* 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 筛选期间的峰拣选
* 进一步处理 MS1 筛选的肽段，获得跨采集重复测定的定量信息

Skyline 旨在为靶向蛋白质组学调查提供一个供应商中立平台。可以从仪器供应商 AB Sciex、Agilent、Bruker、Thermo-Scientific 和 Waters 导入原始数据，进行 MS1 筛选。跨各种仪器平台导入数据的能力极大地方便了跨仪器比较和大型多中心研究。

# 入门指南

如要开始本教程，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，如：

C:\Users\brendanx\Documents

这将创建一个新文件夹：

C:\Users\brendanx\Documents\MS1Filtering

其中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。

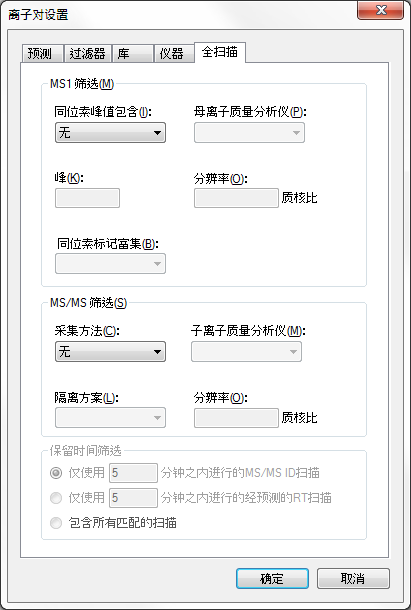
# 将数据依赖型肽搜索导入 Skyline 文档

根据数据依赖型采集 MS/MS 将肽段搜索导入 Skyline 文档的最简洁方法是使用“导入肽段搜索”向导。

如果您此前正在 Skyline 中处理全扫描数据，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：



如果您正在处理同位素标记的肽段，您可能还需要进入“肽段设置–修饰”标签，并关闭所有同位素修饰。

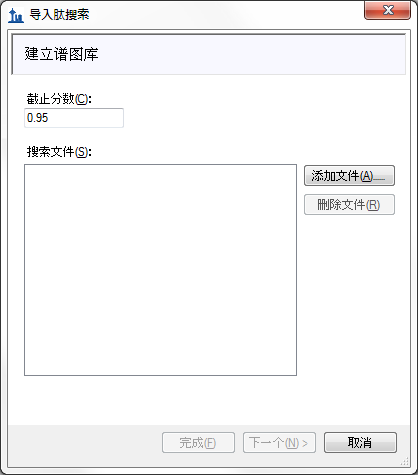
然后，执行以下操作以保存您的新文档：

* 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：

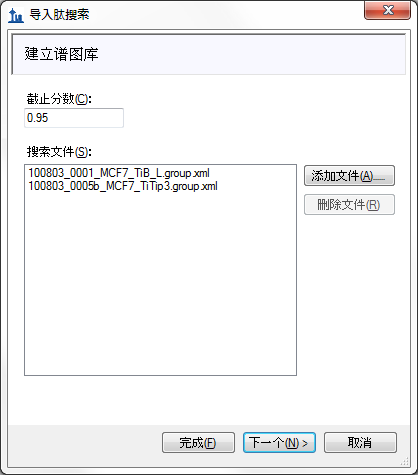


然后，您将使用本向导，从 Skyline 支持的众多肽段搜索引擎之一输出的结果创建一个谱图库。关于支持的搜索管道的全部详情，请查询“[靶向方法编辑](https://skyline.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=tutorial_method_edit)”教程。请注意，您将在本教程中使用的文件已经被压缩至完成本教程所需的最小信息量，以便将下载的 ZIP 文件大小控制在合理的范围之内。

通过执行下列操作，将包括的搜索结果添加至您的库：

* 单击“添加文件”按钮。
* 选择您为本教程创建的 MS1Filtering 文件夹中的两个 .group.xml 文件。
* 单击“打开”按钮。

向导表单现在将显示如下：



* 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 Skyline 已保存您的新谱图库的 MS1Filtering 文件夹中，您现在可以看到两个新文件：

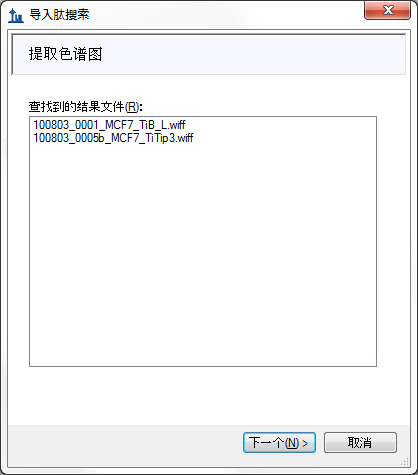
* 包含最佳匹配谱图的非冗余库“MS1FilteringTutorial.blib”。
* 包含所有匹配谱图的冗余库“MS1FilteringTutorial.redundant.blib”。



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。可以删除这个文件，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 创建库作为文档特定的谱图缓存，与为您的文档存储特定色谱图的方式非常相似。此后您能添加更多搜索结果，且您能删除搜索结果，正如您可以使用这些结果处理色谱图数据一样。

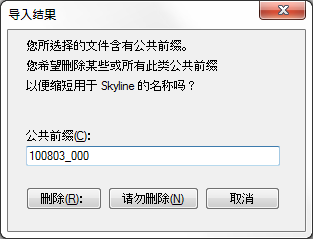
当完成库构建时，Skyline 将在向导中显示下列页面：



在这种情况下，Skyline 已找到与用于构建您的库的谱图源文件相匹配的初始 WIFF 数据文件，且库似乎已有 Skyline 将在其提取的色谱图上定位已确认的 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，将要求您进行定位。如果构建的库无法在导入的肽段搜索文件中寻找保留时间信息，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。

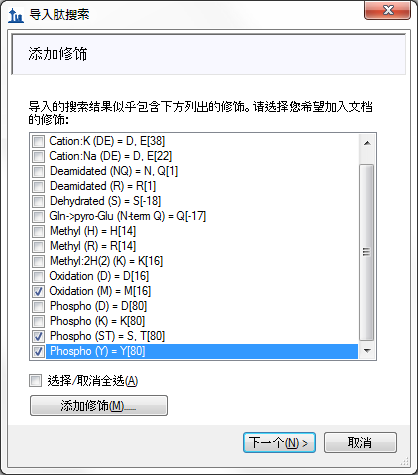
将出现一个表单，询问您如何处理两个 WIFF 文件的共同前缀：



* 单击“删除”按钮。

向导继续前进至“添加修饰”页面，该页面列出了搜索结果中发现的、文件中没有的所有氨基酸修饰。可能的情况下，将提示与搜索的氨基酸，发现的质量组合相匹配的 Unimod 部位特定修饰。

在本教程中，您只需要“Phospho (ST)”、“Phospho (Y)”和“Oxidation (M)”修饰。将其在列表中选中，向导将显示如下：



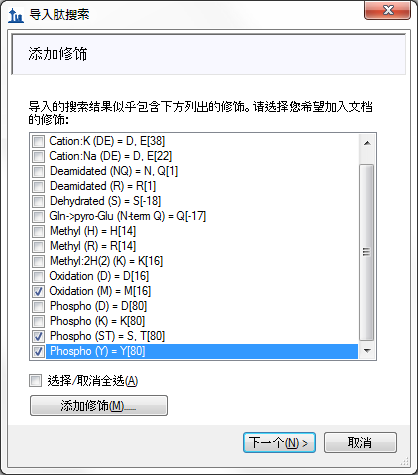
可能您的文档已有一个或多个这些定义的修饰（例如 Oxidation (M)），这种情况下列表看起来可能不同。

* 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：

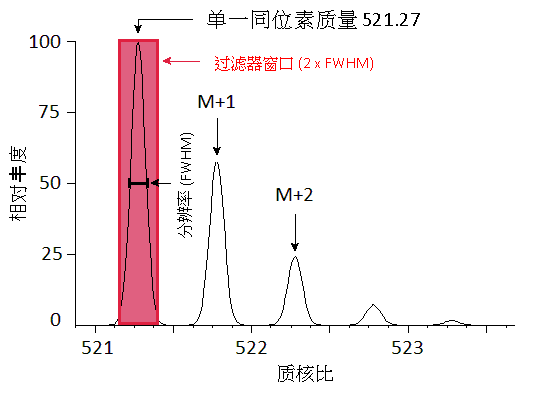
* 在“母离子电荷”字段中输入“2, 3, 4”。

本页面中所有其他字段应为本教程中可使用的默认值，向导显示如下：

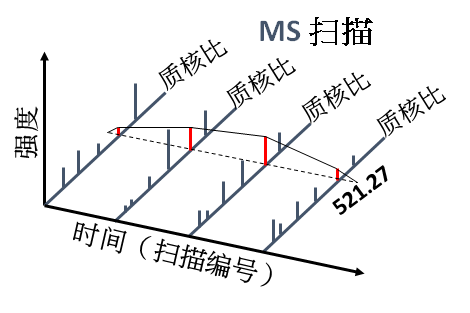


在“MS1 筛选”部分中，您将看到下列默认设置。

1. 在“同位素峰值包含”下拉列表中选择“计数”值。
2. “峰”字段包含“3”，使 Skyline 从此高分辨率数据中过滤前 3 个同位素峰（M、M+1 和 M+2）。
3. 在“母离子质量分析仪”下拉列表中，选择“TOF”值（因为数据由 QSTAR Elite 采集）。
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核比的 MS1 筛选窗口宽度。Skyline 使用该值预测质核比尺寸中一个峰的半峰宽 (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，可以根据仪器功能调整分辨率设置）。



在一段时间内提取的一系列这些强度将构成您在 Skyline 看到的色谱图：



“保留时间筛选”部分，注意选择“仅使用 [5] 分钟之内进行的 MS/MS ID 扫描”。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，在 ID 两侧各增加 5 分钟。当运行缺乏特定肽段的任何 ID 时，Skyline 将使用其他运行中 ID 的保留时间校准确定提取色谱图的时间范围。

* 单击“下一步”按钮。

此操作将把您带入向导中的“导入 FASTA”页面。您可以导入一个全部人类 SwissProt 蛋白质条目的 FASTA 文件，以获得全面确认的肽段列表（此 MS 实验涉及一个人 MCF7 乳腺癌细胞株样品和随后的磷酸化肽富集），但因文件大小的关系，您将通过执行下列步骤导入一个仅含 12 个人蛋白质的小得多的 FASTA 文件：

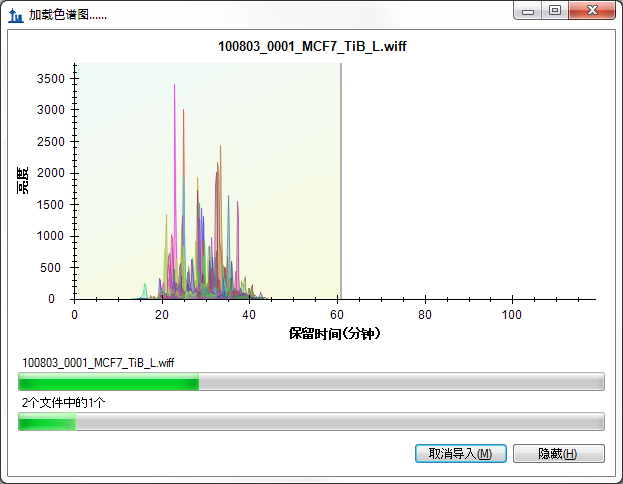
* 在“最大遗漏的酶解位点数”下拉列表中选择“2”。
* 单击“浏览”按钮。
* 从您为本教程创建的 MS1Filtering 文件夹中选择“12\_proteins.062011.fasta”文件。
* 单击“打开 FASTA”表单中的“打开”按钮。

向导现在将显示如下：



* 单击“完成”按钮。

Skyline 将从具有您导入的肽段搜索结果的匹配谱图的 FASTA 文件中添加所有肽段目标，然后开始导入两个 WIFF 文件，并从中提取色谱图。您将看到如下所示的进度图：



完成导入后，检查色谱图数据前请首先深入研究您创建的谱图库。

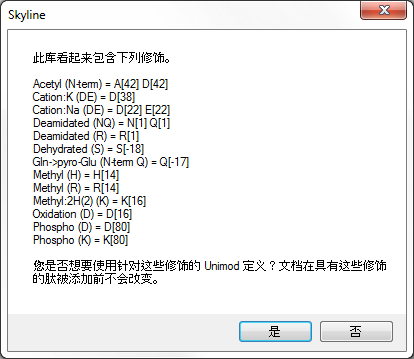
## 验证库保留时间信息

任何时候您使用肽段搜索管道的结果构建 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，您应确保结果库包含必须的保留时间信息，以支持下面解释的 Skyline 功能。 使用“导入肽搜索”向导的优势之一是，当您的库缺少必要的信息时，将提早通知您。

要验证您刚创建的库包含 MS1 筛选峰拣选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”。

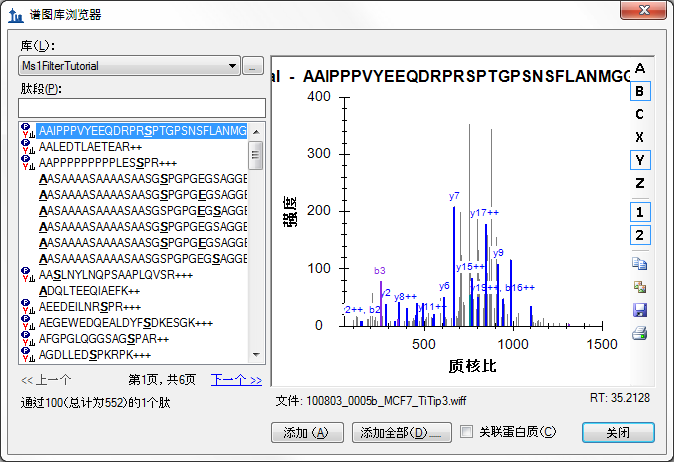
Skyline 将再次提供使用从“导入肽段搜索”向导中选择的、未加入文档的库中检测到的修饰。



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。Skyline 发现，许多肽段搜索管道都涉及将原始仪器数据转换为 mzXML、mzML、MGF、MG2 等格式。通常，Skyline 将寻找基础名称匹配，其中“basename.mgf”与“basename.wiff”成功匹配。由于管道的特定实例需要更大的灵活性，此匹配也有不敏感的情况，因此“BASENAME.mzML”将匹配“Basename.RAW”，且最终包含多点扩展的处理，因此“basename.c.mzXML”与“basename.raw”匹配。然而，如果您看到一些类似“F011852.dat”或一些其他不共享基础名称，但有您想导入Skyline的数据的搜索输出文件，则需要您查看您的搜索管道，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](https://skyline.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=mascot_missing_rt)”页面。其他问题，建议您在 Skyline 支持板（在“帮助”菜单上单击“支持”）上提问，寻求解决此类问题的帮助。

现在按下向下箭头键选择其他肽段，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，请通过执行下列操作返回到您创建的文档：

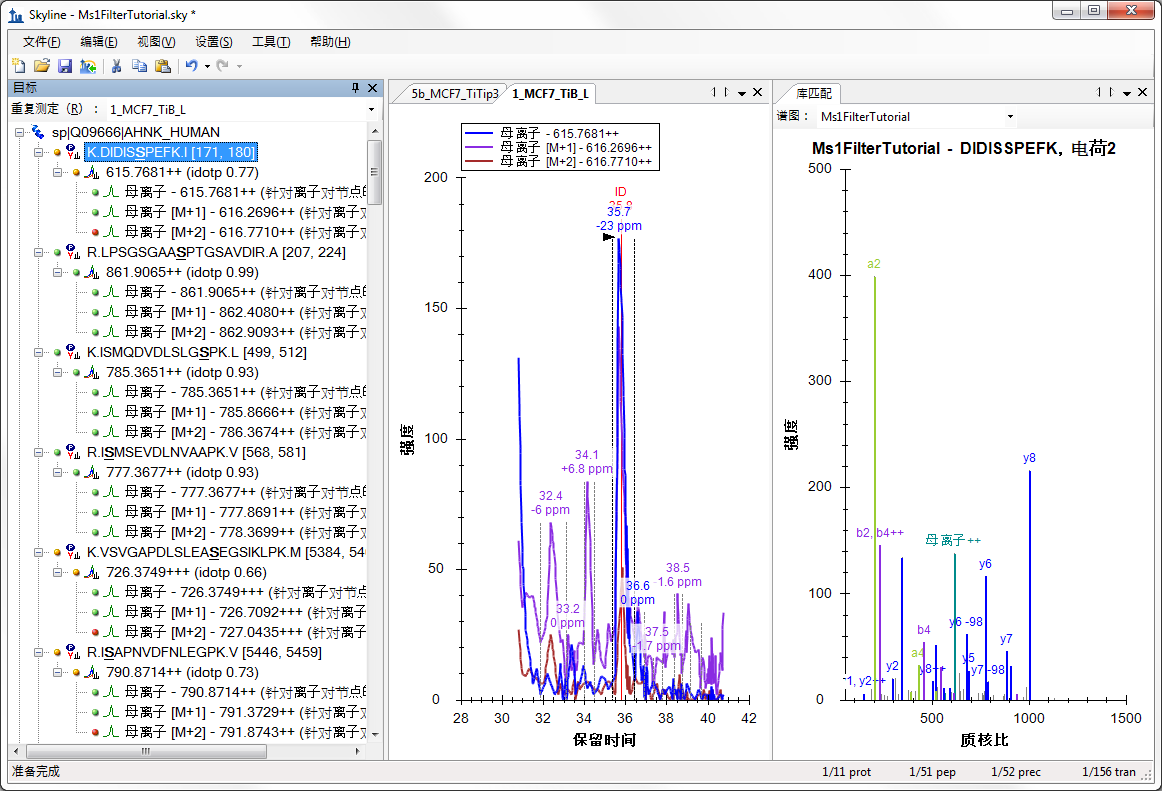
* 单击“谱图库浏览器”中的“关闭”按钮。

# 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.DIDIS**S**PEFK.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：



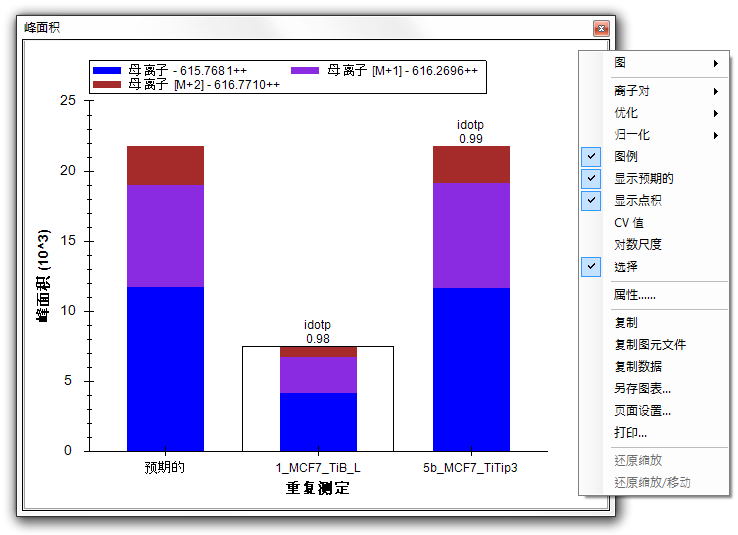
现在文档已导入两次 DDA 运行，已完成进行 MS1 筛选所需的充分配置。您将看到，由于在导入向导中选择了“仅使用 [5] 分钟之内进行的 MS/MS ID 扫描”设置，本视图中的色谱图长度约为 10 分钟（31 至 41 分钟）。请注意，在针对 MS1 筛选进行 Skyline 文档设置时，您将在看见三重四级杆 SRM 实验子离子离子对（例如 y-离子）的地方，看到不同的母离子同位素峰值，例如对于肽DIDIS**S**PEFK：母离子 - 615.7681++、母离子[M+1] - 616.2696++和母离子[M+2] - 616.7710++。

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，无论峰是否与最大峰共同洗脱。

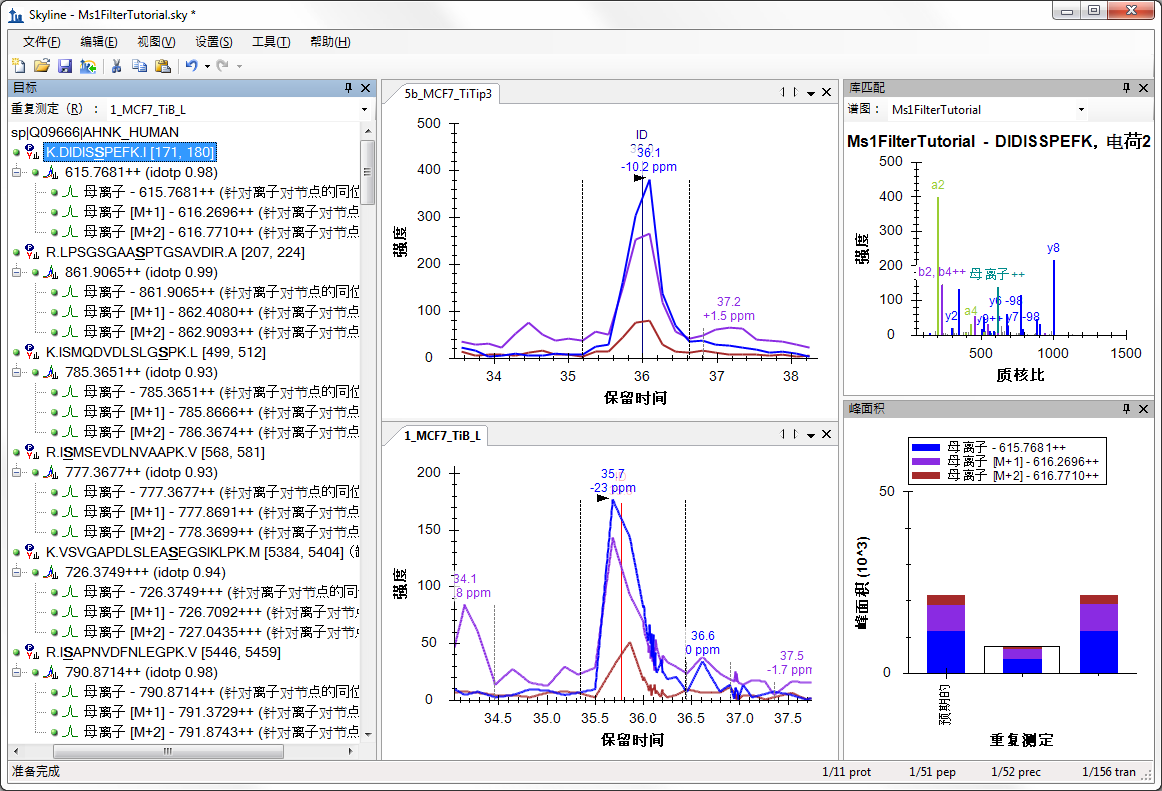
* 在“视图”菜单上选择“峰面积”，并单击“重复测定比较”。
* 在“峰面积”窗口中单击右键，选择“归一化”，并单击“无”。
* 在“峰面积”窗口中单击右键，并选中“显示预期的”和“显示点积”（将在下面解释这两项功能）。



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测定视图上方，如上所示。

色谱图视图显示 MS1 提取离子色谱图所有的母离子同位素离子 M（蓝色）、M+1（紫色）、M+2（棕色）。如果您已使用 Skyline 进行 SRM，可能对选择的峰下面的保留时间注释熟悉，您将看到一个新的质量误差注释，是跨带注释的色谱图所有整合点的质量误差权重平均值（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”。您无法预测现代高分辨率仪器的质量准确性，但正如前面提到的，此数据源自较早的 QSTAR Elite。

您将还能在提取的粒子色谱图中看到垂直线，图顶部有 ID 注释，虽然 1\_MCF7\_TiB\_L 的垂直线在峰注释的后面。ID 代表“鉴定”，显示取样的 MS/MS 谱图的保留时间，以及此特定的肽的信任识别。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 重复测定中已确认的谱图，现在将保存于此前导入期间您已创建的库中。还将从“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”下拉列表在两个采集的谱图中前后转换，并看到他们非常相似。

查看本文档中一些其他 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” (Ctrl-Shift-D)。

此后，确保焦点在“目标”视图（肽树）中，并开始使用您键盘上的向下箭头键，每次选择一个肽。对于前三个磷肽，您将看到每次重复测定时确认的每个结果，'“库匹配”谱图中有少数突出的中性丢失离子注释为“-98”(-H3PO4)。

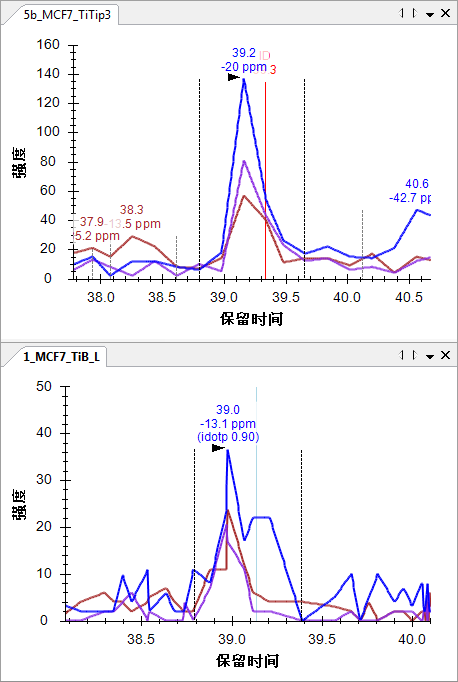
对于第四个肽 I**S**MSEVDLNVAAPK，您将发现仅 5b\_MCF7\_TiTip3 重复测定有一个 ID 注释。使用 5b\_MCF7\_TiTip3 中的 ID 保留时间校准从 1\_MCF7\_TiB\_L 重复测定中选择峰。要查看校准的 ID，请执行以下操作：

* 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 重复测定峰整合边界内侧。尽管如此，这种情况下峰很可能在整合边界的左侧。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。

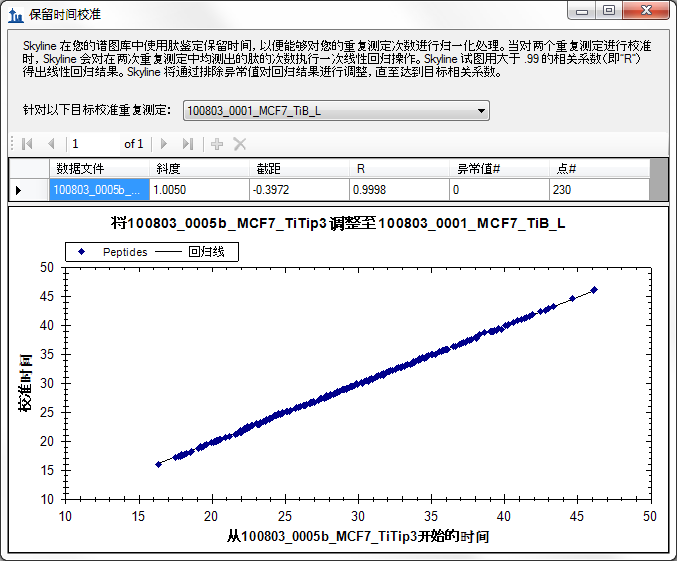
色谱图将如下所示：



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 计算您的谱图库中每个谱图源文件和每个其他谱图源文件间这样的线性回归。当出现 2 个以上运行时，您将看到每个运行下出现一行，在“针对以下目标校准重复测定”下拉列表中选择的运行除外。如上所示，当一次运行中未显示 ID 时，能使用产生的线性方程映射运行间的 MS/MS ID 时间，改善峰拣选。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**](https://skyline.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=tutorial_irt) 教程。

这种情况下，您能看到这些两次运行间的保留时间再现性相当好，斜度为 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，并使用产生的线性方程。

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，各运行中并未确认所有肽段。但是，请记住，您构建的库共包含 552 个肽，其中有许多肽具有本文档未使用的修饰。这似乎显示，两个文件确认了 552 个肽中的 230 个肽。Skyline 尝试使用本回归的两项搜索结果文件中存在的所有 ID。当单次运行出现多个 ID 时，Skyline 将使用最早的 ID 保留时间，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，我们已看到梯度洗脱过程中再次确认了早洗脱出的肽的实例。

* 单击“保留时间校准”窗口右上角的红色的 X，关闭该窗口。

# 观察数据

有了此基本理解，且 Skyline 按此方式配置，您现在能快速观察本文档中的所有 51 个肽段。现在为达到此目的，只需单击“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 个肽中选择的肽量，您可以查看 Skyline 窗口右下方的状态栏：



肽 I**S**MSEVDLNVAAPK 之后，您将看到 4 个峰积分可接受的肽，虽然 VSVGAPDLSLEA**S**EGSIKLPK 可能依据5b\_MCF7\_TiTip3 略作调整。部分肽在两次运行中都有 ID，部分肽仅在一次运行中有 ID，且 Skyline 已使用校准拣选了正确的峰。

## 对色谱图横向显示的基本认识

对于第 9 个肽 SSKA**S**LG**S**LEGEAEAEASSPK，1\_MCF7\_TiB\_L 运行缺少自身的 ID，且整合看起来有点偏：

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

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：正如预期您的目标肽可以在多次运行时以高度相似的时间被洗脱出来，其他肽也如此。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形成，如果这两个峰与目标肽共同洗脱出来，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。使用选择性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预期看到更多的肽。

现在请通过执行下列操作，纠正 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。

继续处理剩余的肽之前，花些时间考虑提取的色谱图中提取的其他两个峰。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 拣选这些峰，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）：



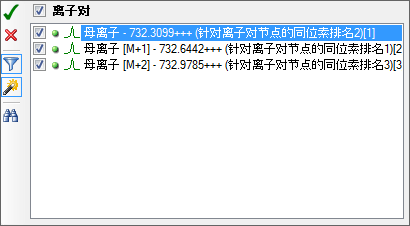
您可以从标示预测的色谱柱中看到分布，因为 M+2 和 M+3 峰比目标肽段的预期同位素分布小，这将告诉您产生此峰的肽中碳原子比在大约37分钟处也有一个 ID 的目标肽段少（且因此得到 13C 的机会较小）。

转向第 33 分钟的峰，您将看到，此峰缺少目标肽的单一同位素质核比的任何信号峰，但由于拥有与 M+1 和 M+2 的强度非常相似的峰，其与目标肽的 M 和 M+1 预测同位素分布相似。 5b\_MCF7\_TiTip3 的质量误差是+25.8 ppm，且当 1\_MC7\_TiB\_L 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。

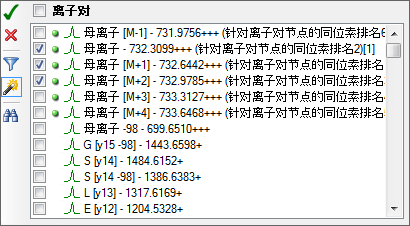
Skyline 将显示如下所示的弹出窗口：



如果您仅看到这三个母离子离子对：

* 单击漏斗图标，删除离子对筛选。

这将使 Skyline 显示此肽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 自动提取预示整个分布中至少有 1% 同位素分布的所有峰的色谱图。此外，Skyline 始终提取 M-1 的色谱图，因为正确拣选没有干扰的峰将在此质核比时一般不会产生信号。

* 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。

这样操作将向图中增加 M+3 和 M+4 的色谱图，且您将看到 33 分钟峰的信号多于已确认的 37 分钟峰。在使用 Skyline 处理任何类型的色谱图数据时，可靠的保留时间再现性的作用相当重要，无论怎么强调也不为过。



您现在可以相当确信，33 分钟时的峰由另一种肽产生，其原子组成与目标非常相似，并带有三个电荷，虽然其单一同位素质量比目标大 1 道尔顿。

* 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，您不必舍近求远即可寻找具有真正干扰的第一个肽。您将看到双磷酸化肽 ASLG**S**LEGEAEAEAS**S**PKGK 的色谱图如下所示：

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

此外，5b\_MCF7\_TiTip3 有一个肽的 ID 并未出现在 1\_MCF7\_TiB\_L 中。依据 5b\_MCF7\_TiTip3 中的 ID 校准选择 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积分，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 质量误差得到峰值整合。

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 个道尔顿的肽产生。



* 单击“撤销”按钮恢复此变化，并继续探索。

AEGEWEDQEALDYF**S**DKESGK 进一步降解为两个肽，您将发现一个更强的干扰，其信号甚至更难排除。

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

为此肽添加 M+3、M+4 和 M+5 色谱图，说明母离子空间的此特定质量和保留时间组合的紧密程度：

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

为获得此肽无干扰的整合峰，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选择性的方法，但您实际上仅通过 MS1 扫描就得到了许多有用的定量数据。关于定量统计，您可能希望限定自己以得到等级最高、但没有明显干扰的母离子。配合可接受的峰的鉴别，此操作可能限定类似您从此教程数据中看到的干扰产生的影响。

继续审查肽，您将仅需要一项小的整合调整，直至您达到肽 22ALVEFESNPEETREPG**S**PPSVQR。

## 肽的不同修饰形式

此处您将发现文档包含 ALVEFESNPEETREPG**S**PPSVQR 和下面的 ALVEFESNPEETREPGSPP**S**VQR，两者的母离子质核比都为 879.0727。本例中，先导蛋白的搜索引擎已鉴别到前者出现在 5b\_MCF7\_TiTip3 中，后者出现在 1\_MCF7\_TiB\_L 中，但色谱图澄清鉴别的两者为相同的峰，均大约出现在 32.5 分钟。

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

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

“1\_MCF7\_TiB\_L”中，同位素分布和质量误差令两者成为两个峰，在两个实例中均出现在 31.5 至 33 分钟之间，两者看起来与 5b\_MCF7\_TiTip3 中的峰有更多不同，但这种差异可能仅由方差所致。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，或磷亚型可能有重叠洗脱情况。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

继续裂解至肽 25，您将发现 YGPADVEDTTGSGATDSKDDDDIDLFG**S**DDEEESEEAKR，本文档中最长的首个带 4 个电荷的肽母离子。由于该肽分子量大，其同位素分布与分子量较小的、带 2 个电荷的肽差异很大，甚至与您一直关注的分子量较大、带 3 个电荷的肽差异也很大。现在预期不含 13C 原子的单一同位素肽，其出现频率小于 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，预期分布如下所示：



正如您一直做的，使用离子对拣选列表，您能增加从 M+3 至 M+7 的色谱图，所有的色谱图都应包含大于整体同位素分布 1% 的峰，并看到 idotp 值仍非常高，为 0.98：

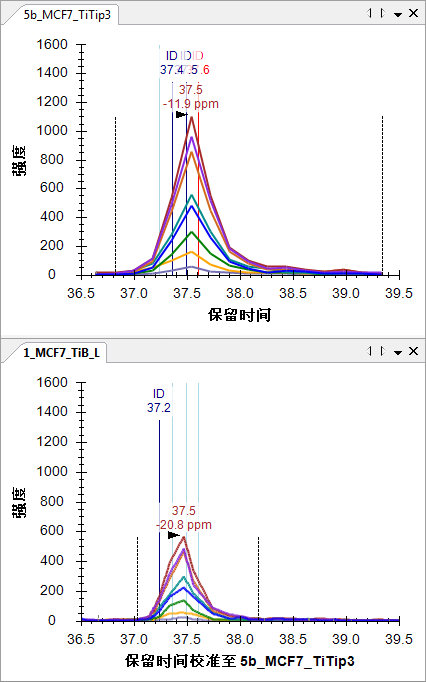


在色谱图中，您可以发现，这是文档中唯一在单次运行中得到多次确认的肽（5b\_MCF7\_TiTip3 中的 3）。

您可以执行下列操作，将色谱图调整在相同的尺度内，更容易解释重复测定间如何校准这些 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，而不仅仅是当前的肽。）
* 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：



您能看到校准使 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，审查搜索引擎确认的此肽的谱图，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。对相同肽的不同运行谱图进行判断可能需要一些想像力。

**5b\_MCF7\_TiTip3（37.61 分钟）**



**1\_MCF\_TiB\_L（37.03 分钟）**



但是您应该非常确信，两次运行的色谱图峰都能测量相同的肽分子。

继续考虑肽 27，GVVDSEDLPLNISR，您将发现整合需要调整：

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

开启“同步缩放”，您可以放大查看整合边界相距很远的峰：

* 单击 1\_MCF7\_TiB\_L 色谱图绘图区域。
* 在任一方向移动鼠标滚轮。

5b\_MCF7\_TiTip3 图将缩放至与 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，您将看到两次运行的峰面积现在共约 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 DQVANSAFVER 具有另一个关注的干扰：

|  |  |
| --- | --- |
| **5b\_MCF7\_TiTip3** | **1\_MCF7\_TiB\_L** |

在 1\_MCF7\_TiB\_L 中，在 24.5 分钟确认到肽。然而，两次重复测定中，您将看到在大约 25 分钟时出现强干扰峰。虽然，此干扰峰仅出现在 M 和 M+2 色谱图中。由于目标肽带 2 个电荷，这将告诉您干扰肽带 1 个电荷。5b\_MCF\_TiTip3 中的目标信号非常弱，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。

在两种重复测定采集中的这种严重干扰情况下，应仔细观察误差侧，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，您可能需要使用一种选择性更大的方法，比如靶向 MS/MS 或 SRM。

本文件中其余的 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，但您可以随时忽视并继续下一部分：

1. ETERA**S**PIK**M**DLAPSK (31 & 32) – 对相同的蛋白质重复检测两次（删除一次）
2. K**T**GSYGALAEITASK & KTG**S**YGALAEITASK (34 & 35) – 峰给出两个磷位点
3. TPSPKEEDEEPE**S**PPEKK (41) – 不佳色谱法漏整合（缩放，调整整合）
4. KEK**T**PELPEPSVK (46) – 对 M+1 和 M+2 有干扰
5. EK**T**PELPEPSVK (47) – 对 M+2 有干扰
6. VPKPEPIPEPKEP**S**PEKNSK & VPKPEPIPEPKEPSPEKN**S**K (49 & 50) – 峰给出两个磷位点
7. KETE**S**EAEDNLDDLEK (51) – 对带 1 个电荷的肽的 M+1 有干扰

# 将色谱图缓存文件最小化

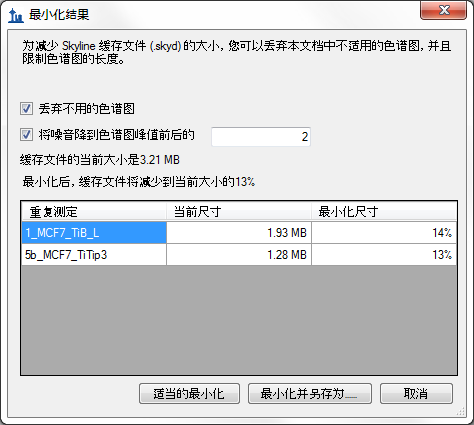
完成将现在的 50 个肽加入您的文档后，这些肽整合得相当好。继续下一步操作前，保存当前的文档：

* 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，使文件尽可能小以易于共享：

* 在“编辑”菜单上单击“管理结果”。
* 单击“最小化”按钮。
* 选中“将噪音降到色谱图出峰值前后的”复选框。
* 在该字段中输入“2”，指定“将噪音降到色谱图出峰值前后”的分钟数。
* 按 Tab 键进入重新评估尺寸减少量的表单。

“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，您可以查看本新文档中肽的色谱图，并看到色谱图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，您可以共享色谱图作为原稿的共享数据。您可能还希望可以在线利用原始数据，但最小化的 Skyline 文档将以很小的下载成本为您提供您数据的丰富视图。

# 用于 MS1 筛选的内含物列表法导出

如上所述，多次重复测定研究使用 DDA 显示 MS/MS 的采样过疏，且每次采集重复测定中，并非所有肽都有 MS/MS 鉴定。MS1 筛选可以使用此前描述的 RT 校准克服此问题。然而，当您已从初始纯度发现中删除已知的、甚至您希望检测的多种肽中的一种相对大量的目标肽时，您能使用 Skyline 为您的 DDA 实验导出一份内含物列表法，实现一种被成为“准确内含物质量筛选”的方法2。内含物列表方法应增加您对关注的肽取样进行非定向 DDA 的掌控机会。

目前，Skyline 可以为 AB SCIEX 和 Thermo 仪器导出内含物列表法，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法，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。减少窗口重叠的重要性不及 SRM。
* 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，除非您实际拥有已安装支持供应商（AB SCIEX Analyst 或 Thermo Xcalibur）的供应商仪器软件的系统。对于来自 Skyline 的所有方法导出，建议您在计划运行您的方法的仪器控制电脑上运行的 Skyline 实例中执行导出功能。因为，即使您的实验室拥有一台支持的仪器，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

您已通过本教程学习了使用 Skyline 从您的 DDA 实验数据的 MS1 扫描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 教程和教学视频中呈现的材料。虽然使用从 MS1 扫描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 筛选会继续迅速改善。

# 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 ACETYLATION AND PHOSPHORYLATION.*Mol Cell Proteomics* **11,** 202–214 (2012).

2. Jaffe, J. D. *et al.*Accurate Inclusion Mass Screening.*Mol Cell Proteomics* **7,** 1952–1962 (2008).