

# Skyline 非数据依赖型采集

---

非数据依赖型采集 (DIA)<sup>1,2</sup> 是一项可用于进行大规模靶向蛋白质组学实验的先进技术。靶向采集方法, 例如选择性反应监测 (SRM) 和平行反应监测 (PRM), 仅可适用于一些无采集时序安排的肽段或在有时序安排的情况下, 适用于每次质谱仪运行下的数十至数百个肽段。与SRM相比, DIA可以检测更多数量的肽段(数千或甚至整个蛋白质组), 并且不严重影响检测方法的敏感性、选择性和再现性。DIA 的另一优势是, 待测量的肽段不需要提前或计划指定, 反而在广泛的母离子质荷比中, 任何需要的肽段的子离子色谱图均可以在采集后从 DIA 运行中提取。

2010年10月, Skyline首次用于从DIA 数据中提取色谱图。自此以来, 这一功能已得到了很大的改进。当前版本的 Skyline可用于多种常见的DIA数据分析方法和流程。Skyline 亦支持所有可以进行 DIA 的仪器, 包括 AB SCIEX, Agilent, Bruker和Waters 的 Q-TOF 仪器和Thermo 的Q-Orbitrap 仪器。

若要使 DIA 实验富有成效, 那么开始便需要在某种色谱仪和质谱仪上采用数据依赖型 (DDA) 的方法进行初始的数据的采集。如果这些初始鸟枪法的检测和DIA实验在同一仪器上进行, 这会大有帮助。但如果使用相似的多肽破碎技术和色谱分析法, 同一试验方法则亦可能在不同的仪器平台之间转移。在这些初始的 DDA 实验中, 样本可以被分离或进行简化从而达到更高的蛋白质组覆盖率。多肽-谱图匹配软件将用于分析DDA 数据, 而生成的多肽序列 (ID)、谱图和保留时间会被用来创建一个谱图库和保留时间 (iRT) 库 (在 Skyline 中) 或创建一个加长的离子对列表 (含有片段离子子集的相似信息) 或者被称为 “化验物库” (在其他工具中)。这些子离子相对丰度值和标准化的保留时间 (iRT) 可以用于任何数量的后续 DIA 实验, 但这些 DIA 实验需使用相同的仪器和保留时间校准的标准肽段。

尽管许多现有方法可以将 DDA数据的搜索结果转化为适用于DIA 分析方法的信息 (详见于 [Skyline 教程网络研讨会 #2](#) 中的 “先验知识工作流”), 但最直接、最方便的开始 DIA 分析的方法是将 DDA和 DIA 在相同的仪器上交叉进行, 并且使用 DDA 结果的谱图和保留时间来进行预测 DIA 的实验。任何一种可以创建包含更精准的子离子相对丰度值和标准化的保留时间信息的库 (被称为色谱图库, 详见于 [Panorama 色谱图库教程](#)) 都可以被用于根据已有的DIA结果来预测之后要进行的 DIA实验。

在本教程中, 您将学习到如何使用DDA/DIA 交叉运行的方法在 Skyline 中设置、导入和处理 DIA 数据 (其实本教程中也包含了替代方法的指示、更多的高级方法, 贯穿于本教程中)。

## 开始

如要开始本教程，请下载下列 ZIP 文件：

<https://skyline.gs.washington.edu/tutorials/DIA.zip>

请注意此文件非常大（下载内容 4.5 GB，解压后 6.0 GB），因为DIA实验经常会产生较大的文件，通常比 SRM 实验的文件大约大100至200倍。如果此下载内容时间太长，或如果您没有足够的磁盘空间，则您可以下载另一个较小的文件版本（下载内容 73 MB，解压后 113 MB）：

<https://skyline.gs.washington.edu/tutorials/DIASmall.zip>

如果您选择缺少质谱仪原始数据文件的较小版本，则需要跳过本教程中的一些步骤，下文将有注释。无论您选择哪个 ZIP 文件，下一步都是解压文件到您计算机中的一个文件夹，如：

C:\Users\damodei\Documents

这将创建一个新文件夹：

C:\Users\damodei\Documents\DIA

这里会包含本教程需要的文件。你可以在资源管理器中双击或在运行的 Skyline 文档中从**文件**菜单（Ctrl+O）中选择**打开**，从而打开此文件夹中的 DIABlack.sky 文件。

## 设置 DIA 隔离方案

您打开的 Skyline 文档应该是完全空白的，没有离子对、没有质谱仪数据、没有特殊设置。为了阐述使用Skyline分析 DIA 数据的完整过程，您需要从头创建可适用于 DIA数据分析的 Skyline 文档，包括填补必要的设置、离子对、谱图库和保留时间等信息。

如果您要使用 DIA 和 DDA 交叉的方法进行实验，那么您需首先为 DIA 和 DDA 实验来设置仪器。要达到此目的，您仅需简单照顾到一下您可能感兴趣的任何目标肽段，如保证母离子质荷比范围包括您的目标。DDA的实验就交给您自己完成，但 Skyline 可以通过定义“隔离方案”（母离子隔离窗口生成片段以使用串联质谱 MS/MS 的模式）从而帮助您设置 DIA 方法。即使您已经收集了 DIA 数据，您也需要定义您所使用的隔离方案以便于 Skyline 处理您的 DIA 数据。按以下步骤可定义本教程实验的隔离方案：

- 在**设置**菜单中单击**离子对设置**。
- 单击**全扫描**标签。
- 在**采集方法**下拉列表中，选择“DIA”以使 Skyline 知道您将导入 DIA 数据。
- 在**已包含的同位素峰值**下拉列表中，选择“计数”。
- 在**峰值**字段，输入“3”。

对大部分胰蛋白酶多肽，前 3 个同位素峰值是强度最高的。您也可以使用基于基准（强度最高的）同位素峰值的百分比作为强度阈值，但这些设置已经是适用于胰蛋白酶肽的合理默认设置。

- 将母离子质量分析仪和子离子质量分析仪设置为“Orbitrap”（轨道阱）。（此数据系收集于 Q-Exactive，其中 MS1 和 MS2 扫描已使用 Orbitrap 执行。）
- 将 MS1 筛选下的分辨力设置为在“200”时“35000”，MS/MS 筛选下设置为在“200”时“17500”。

离子对设置分类应看起来像：

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): 计数

母离子质量分析仪(P): Orbitrap

峰(K): 3

分辨力(Q): 35000

位于(A): 200 质荷比

同位素标记富集(B): 默认

串联质谱过滤(S)

采集方法(C): DIA

子离子质量分析仪(M): Orbitrap

隔离方案(L):

分辨力(Q): 17500

位于(I): 200 质荷比

保留时间筛选

仅使用 5 MS/MS ID 的分钟

仅使用 5 预测保留时间的分钟

包含所有匹配的扫描

确定 取消

现在您已设置好了基本的全扫描仪器参数。下一步，您需要指定 DIA 隔离方案，或在进行 DIA 实验时仪器循环通过的母离子质荷比范围。例如，在本教程中的数据集中，Q Exactive 仪器使用了一个关于母离子质荷比范围的循环，该范围开始为 500 至 520 质荷比，然后为 520 至 540 质荷比，一直到 880 至 900 质荷比（或从 500 至 900 质荷比的 20 个连续 20 质荷比窗口），接着便重复了此循环。按以下步骤可在 Skyline 中指定此隔离方案：

- 在**隔离方案**下拉列表菜单中选择<添加...>。

编辑隔离方案分类应看起来像如下示例：

编辑隔离方案

名称(N):

确定

取消

使用结果数据隔离目标(U)

隔离宽度(W): 2 质荷比

去卷积(D): 无

非对称(A)

预指定隔离窗口(P)

计算(C)-

图形(G)-

开始	结束

去卷积(D):

边距(N): 无

每次扫描的窗口(W):

指定目标(I)

- 选择**预指定隔离窗口**。

这将启动一张网格，您可以指定隔离窗口。您可以在网格中手动输入窗口边界，但因为此在例中窗口边界的循环具有高度规律性（从 500 至 900 质荷比，每次增量为 20 质荷比），所以有更快的方法来指定边界：

- 单击**计算**按钮来显示**筹划隔离方案**分类。
- 在**起始质荷比**字段输入“500”。
- 在**终止质荷比**字段输入“900”。
- 在**窗口宽度**字段输入“20”。
- 勾选**优化窗口布局**复选框。

这样将在肽段不常出现的区域中布置窗口边界。如果 Q1 隔离在边界<sup>3</sup>具有合理的高效率，这种方法则会减少在 SWATH<sup>2</sup> 初始页面中推荐的重叠窗口（在任一边均有 0.5 质荷比边距）。如果您的 DIA 数据已被采集，那么在此处定义的隔离方案需反映出已用于数据采集的仪器设置，这点十分重要。

**筹划隔离方案**分类应看起来像：



筹划隔离方案

起始质荷比(S): 500 终止质荷比(E): 900 确定

窗口宽度(W): 20 窗口类型(W): 测量 取消

窗口计数(W): 20

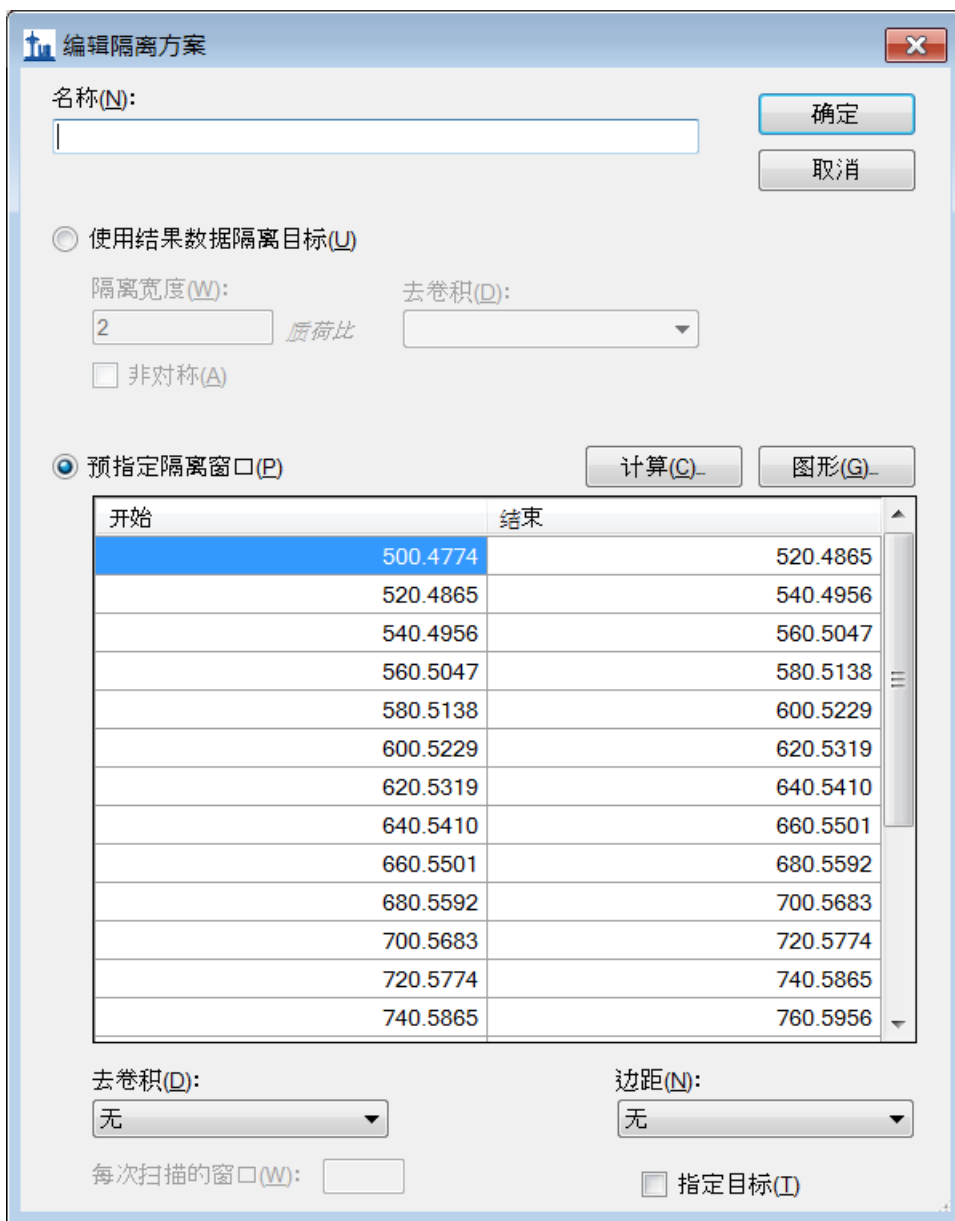
去卷积(D): 无 边距(M): 无

每次扫描的窗口(D): 边距宽度(G):

优化窗口布局(P)  生成目标(I)

- 单击**确定**按钮。

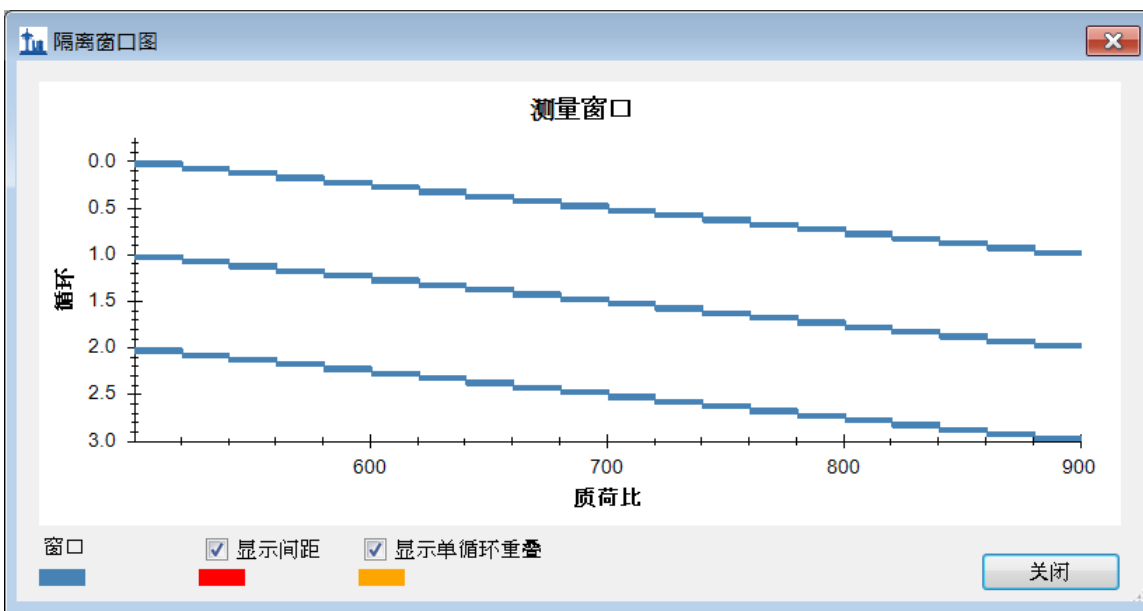
Skyline 将自动填补需要覆盖 500 至 900 质荷比（间隔 20 质荷比）的 20 个窗口边界。  
**编辑隔离方案**分类应看起来像：



您可以通过Skyline 看到母离子质荷比范围隔离随着时间的变化，这样方便检查您输入的内容是否正确：

- 单击您刚才单击过的**计算**按钮旁边的**图形**按钮。

您会看到随着时间前进的隔离窗口循环图形，y 轴显示的是循环/时间，x 轴显示的是质荷比：



- 单击**关闭**按钮。
- 在**名称**字段输入此隔离方案的名称“500 至 900，增量 20”。
- 单击**编辑隔离方案**分类中的**确定**按钮。
- 单击**离子对设置**分类中的**确定**按钮。

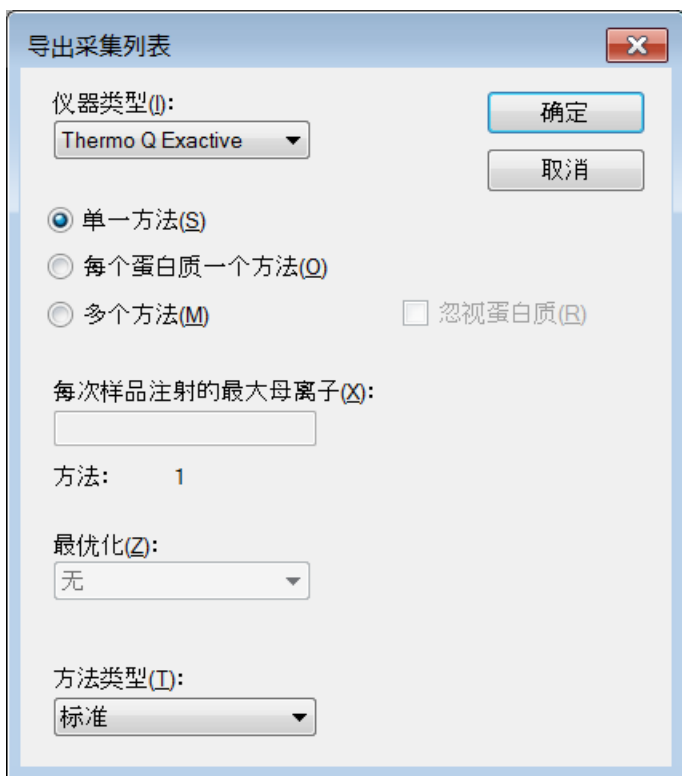
请注意，在 DIA 仪器中不需要一组特定的离子对 (SRM) 或母离子 (PRM) 才能测量，所以您的文档实际上现在已有了进行一次 DIA 实验的所需要的信息（尽管仍没有目标）。按以下步骤可将 DIA 隔离方案导出到仪器中：

- 在**文件**菜单上选择**导出**，然后单击**隔离列表**。

导出**隔离列表**分类将显示出来，您可以选择隔离列表导出的格式。

- 在**仪器类型**下拉列表中选择 **Thermo Q Exactive**。

此分类应看起来像：



- 单击**确定**按钮。
- 在出现的保存分类中，导航至您为本教程创建的文件夹。
- 在**文件名称**字段输入名称“DIA\_tutorial\_isolation\_list.csv”。
- 单击**保存**按钮。

打开您保存的文件。此文件应看起来像：



	A	B	C	D
1	510.4819			
2	530.491			
3	550.5001			
4	570.5092			
5	590.5183			
6	610.5274			
7	630.5365			
8	650.5456			
9	670.5547			
10	690.5638			
11	710.5729			
12	730.582			
13	750.5911			
14	770.6002			
15	790.6093			
16	810.6183			
17	830.6274			
18	850.6365			
19	870.6456			
20	890.6547			
21				

此隔离方案系采用 Thermo Q Exactive 格式（本教程的数据采集于此），但 Skyline 也可以导出适用于其他几种仪器类型的格式。

如果这是真实的 DIA 实验，则您将可以将这份隔离列表文件导入到您的仪器上进行 DIA 的数据采集。或者，您也可以简单地在仪器软件中手动的指定隔离方案。数据采集的其他方法参数（例如 MS/MS 隔离宽度和分辨率）将在方法文件中手动设置。在本教程中，我们已向您提供了 DIA 结果，但当您在进行到“导入质谱数据”之前不需要实际导入它们。

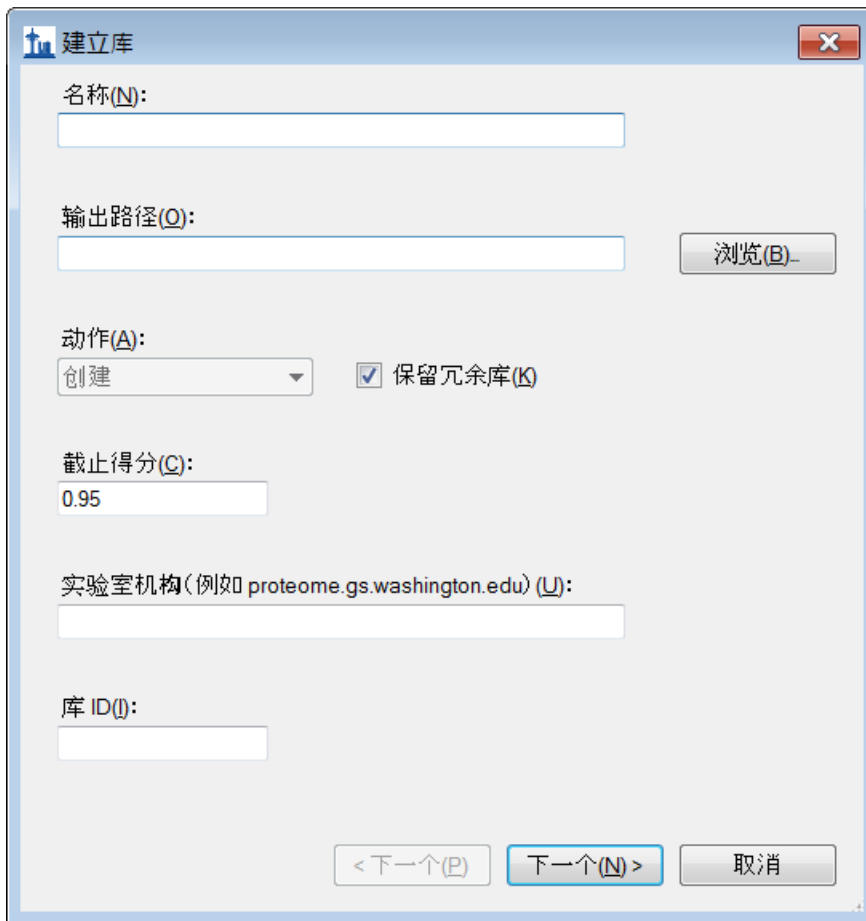
### 在 DDA 运行中设置谱图库

假设现在您所有的 DDA 和 DIA 的实验均已完成（实际上它们均用于本教程），那么您可以开始使用如 X! Tandem 等搜索引擎将 DDA 数据中的 MS/MS 谱图与肽段匹配，从而开

始数据分析。如果您运行了 Peptide Prophet 和 Trans-Proteomic Pipeline (TPP) 软件, 则将出现一系列 .xtan.xml 文件或 .pep.xml 文件。在本教程中, 如果您下载了 DIA.zip (不是 DIASmall.zip) 文件, 那么您将得到用于 DDA 运行的一个单独 .pep.xml 文件和初始的原始 DDA 数据文件的一个 .mzXML 文件转换 (804 MB)。配置 Skyline 文档用于分析相关 DIA 运行的第一步是, 将这些搜索结果导入至 Skyline, 创建一个包括 MS/MS 谱图和采集这些结果的保留时间的谱图库。按以下步骤可导入 DDA 搜索结果:

- 在**设置**菜单上单击**肽设置**。
- 在**肽设置**分类中, 单击**库**标签。
- 单击**构建**按钮来构建一个谱图库。

Skyline 会展示如下的**构建库**分类:



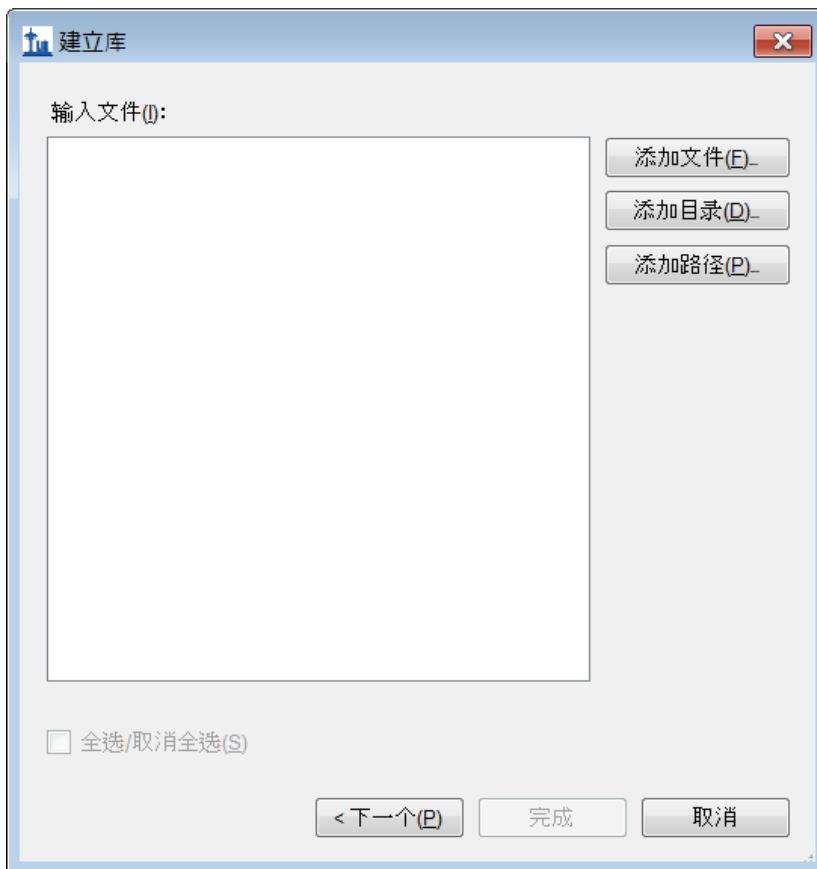
- 在**名称**字段, 输入您将创建的库的名称 “DIA\_tutorial\_library”。
- 在**实验室机构**字段中输入 “proteome.gs.washington.edu”。
- 单击**浏览**按钮。
- 确保 “DIA\_tutorial\_library.blib” 文件位于您提取教程文件的目录中。
- 单击**保存**按钮。

- 确保其他字段如上显示，即**截止得分**为 0.95，**保留冗余库**方框为未选中状态。

针对此数据组，将一个**截止得分**字段设置为“0.95”意味着将在 PeptideProphet 中得分为 0.95 或更高的肽段谱图将会被保留，这是因为 DDA 数据系使用 TPP 进行了处理（在用 SEQUEST 进行了肽段谱图匹配后）。对于其他使用 q 值或期待得分的肽段谱图匹配工具来说，当零分为最好、1 分为最差时，那么截止得分等于  $1 - q$  值或者期待得分。这样 0.95 就表示  $\leq 0.05$ 。针对可重复使用的库，您可以使用像 0.99 这样更严格的截止得分（即相应的 q 值代表  $\leq 0.01$  或 1% 错误发现率）。只有在您想在未来添加更多匹配谱图的结果至您的库中，或者您要从已加入库中的数据中提取色谱图并想查看所有肽段匹配的原始谱图时，冗余库才是必要的。

- 单击**下一步**按钮。

构建库向导会前进至如下所示的下一页：



- 单击**添加文件**按钮。
- 在出现的**添加输入文件**分类中，从您保存本教程的目录中选择以下文件：
  - interact-20130311\_DDA\_Pit01.pep.xml

这个文件包括单个 DDA 运行中匹配肽段谱图的结果。在真实的实验中，您将在质谱仪中执行 DDA 数据采集，然后通过一个搜索引擎处理 DDA 数据，从而生成一个或多个这些文件（通常有一个文件匹配 TPP 生成的 pepXML）。这些文件已在这里为您提供。请注意，原始的 DDA 运行数据文件（转换为 mzXML）interact-20130311\_DDA\_Pit01.mzXML 亦存在于同一文件夹中。您不必导入这个文件才能提取色谱图，但这个文件需要存在才能使库建立者找到这个库的 MS/MS 谱图，而这个谱图不存在于 .pep.xml 文件中。其他匹配谱图的管道输出，例如 Mascot DAT 文件、Proteome Discoverer MSF 文件和 X!Tandem 本地 XML 文件，均在一个单独输出文件中包括了所有必要的信息。

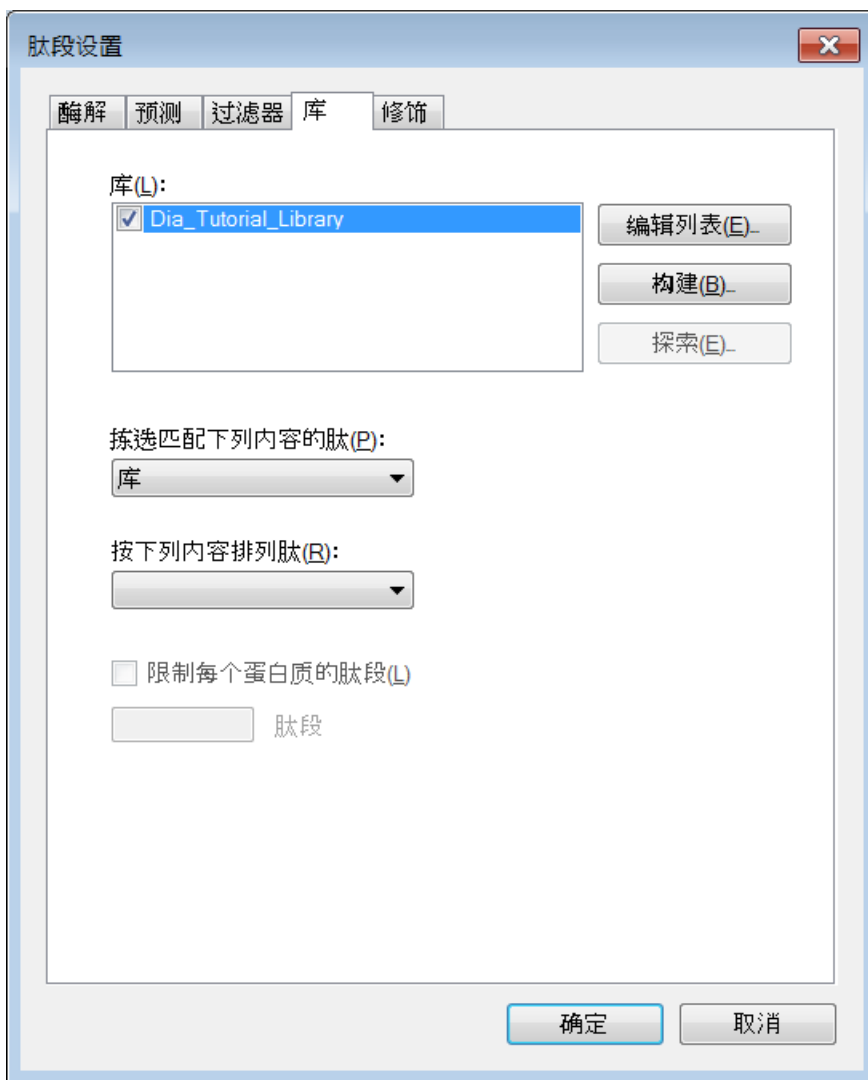
- 单击**打开**来向您正建立的库添加这些文件。
- 单击**构建库**分类中的**完成**来构建库。

**注释：**如果您下载的是小版本的教程，那么您会看到一条错误消息，这是因为小版本的教程缺失原始的 DDA .mzXML 文件。如果您下载了小版本，那么您应现在打开 Skyline 文档“DIASetup.sky”，这里包括已成功加载的谱图库。然后跳到下一部分。

如果您下载了完整版，那么您会在**肽设置**分类中看到您正构建的库出现在**库**列表中。

- 勾选“DIA\_tutorial\_library”左边的复选框，将它加入以在分析 DIA 数据集时使用。

**肽设置**分类应看起来像：



- 确保在**挑选匹配下列内容的肽**下拉列表中选中“库”。
- 单击**确定**按钮来确认新设置和库。

现在，所有匹配 DDA 运行中 MS/MS 谱图和其保留时间的肽段均可以加入您的 Skyline 文档中（您可以在查看·谱图库中进行查看）。在本教程的后面部分中，Skyline 将使用这些信息从DIA数据中提取最丰富的片段离子，以及使用经测量的片段离子相对丰度值和 MS/MS 谱图保留时间来匹配DIA数据中的色谱图峰。

### 限制色谱图提取保留时间范围

在导入 DIA 数据之前，您仍需要告诉 Skyline 在什么保留时间（RT）范围内提取每一个色谱图。在时间预定的 SRM 实验中，仪器仅会在预期的 LC 洗脱时间左右的受限时间窗口记录每一离子对。与其不同的是，在 DIA MS/MS 谱图覆盖中，所有可能的母离子片段均在整个运行过程中被记录。因此，我们有可能在全梯度中提取每一目标离子对的色谱图。

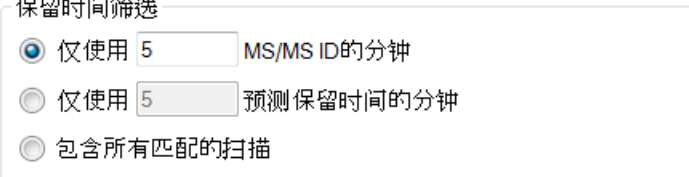
但是在实际操作中，从全梯度中提取目标离子对通常会导至生成含有许多混乱干扰峰的让人极度困惑的色谱图，这会为操作者和自动化软件找到色谱图中的正确峰值造成困难。使用全梯度色谱图也会使您的文件更大、数据导入时间会更久。出于这些原因，尽管 Skyline 允许全梯度提取，但 Skyline 团队仍强烈建议根据 RT 预测器限制色谱图提取范围。也就是说，虽然在 SRM 中色谱图时间范围通常受仪器采集方法中的时间设置的限制，但在 DIA 中，我们建议在色谱图提取过程中限制色谱图时间范围。

限制 RT 范围需要预测每一肽段的 RT，而预测 RT 则有多种方法。本教程集中讲述最简单、最方便的方法，即仅仅使用在 DDA 运行中相对应的肽段 ID 的保留时间（您将在已构建的谱图库中找到）。这种方法没有对 DDA 和 DIA 运行之间的任何色谱变化做更正（例如均匀偏移），尽管如此，许多用户亦认为这对他们要达到的目的来说已足够精确了。我们也可以通过其他方法来预测保留时间，例如：根据多肽的氨基酸序列来预测保留时间的方法 (SSRCalc<sup>4</sup>) 或基于已被标准化和保存的实际测量的 RT 的方法（术语为“iRT”<sup>5</sup>，若需更多信息，请见于 [iRT 保留时间预测教程](#)）。

按以下指导可在您的谱图库中限制基于肽段 ID 时间的 RT 范围：

- 在设置菜单中单击**离子对设置**。
- 单击**全扫描**标签。
- 在**全扫描**标签底部旁边的**保留时间筛选**窗格中，选择**仅使用在 MS/MS ID [5] 分钟内的扫描**。

此窗格应看起来像：



保留时间筛选

仅使用 5 MS/MS ID的分钟

仅使用 5 预测保留时间的分钟

包含所有匹配的扫描

这样 Skyline 会知道，它应该仅在您的库中相对应的肽段谱图的 5 分钟内从 DIA 谱图中提取相关信息。对于单个肽段谱图匹配，总提取时间窗口是 10 分钟。如果一个已知的肽段有不止一个 ID，那么将最小 ID 时间减去 5 分钟、将最大 ID 时间加 5 分钟后会得出一个范围，Skyline 将从在这个范围内采集的谱图进行提取。（请注意，所有 ID 保留时间亦会保存在您已构建的非冗余库中，您将在稍后看到。）第二个选择是，**仅使用在预测 RT [5] 分钟内的扫描**，使用一个 RT 预测器（例如 SSRCalc 或一个 iRT 计算器）来确定提取时间范围。在本教程中您不会使用到 RT 预测器，但它们有在其他几个教程和 [导入化验物库](#) 贴士中讨论（在 Skyline 网页中，[Tips/小贴士](#) > [Working with Other Quantitative Tools/使用其他定量工具](#)）。

选择**包含所有匹配的扫描**将使 Skyline 在全梯度中提取色谱图（在**仪器**标签中指定的任何**最少时间**和**最多时间**之间）。这种方法只偶尔有用，但一般来说不建议按如上所述操作，且您不会在这里用到。

- 单击**离子对设置**分类中的**确定**按钮。

您的 Skyline 文档中不会立即发生任何变化，但当您稍后将 DIA 数据导入本教程时，色谱图时间范围将受到限制。

## 从 FASTA 文件导入目标

到现在为止，您还没有定义您的 DIA 分析的目标蛋白质、肽段或离子对。在 DIA 中，您可以在收集了质谱仪数据后、数据被导入前才这样做。实际上，当前的 Skyline 文档可被当作一份可以重复使用的模板。该模板可以用来从 DIA 原始文件中提取任何与根据 DDA 数据所创建的库相关的离子对组。

虽然您可以在一次 DIA 运行中检测成百上千的肽段（或许更多），但本教程仅讨论 6 个蛋白质，它们是本教程的实验设计的特别对象。教程文件夹包括含有指定了这些蛋白质的 FASTA 文件。在导入前，首先应设置 Skyline，使其知道应导入这些蛋白质的哪些肽段和离子对：

- 在**设置**菜单上单击**肽设置**。
- 单击**过滤器**标签。
- 勾选**自动选择所有匹配的肽**复选框。

这样 Skyline 会在您导入了 FASTA 文件后自动使用符合资格的肽段填满**目标**列表。

在此实验中，您可以保持**过滤器**标签中的其他设置不变，它们会被忽略以支持在您已构建的库中某一谱图匹配的任何肽段。如果您想应用这些过滤器设置，那么您需要更改**拣选匹配下列内容的肽**选项，选择除“库”之外的任何选择。使用“库和过滤器”将只允许生成同时在库中找到并符合**过滤器**标签标准的肽段。

在此特定的实验中，肽段中的半胱氨酸被脲甲基化修饰，所以您需要使 Skyline 将它添加为固定修饰：

- 单击**修饰**标签。
- 如果您在**结构修饰**列表看到修饰“脲甲基化 (C)”，请确保它是被选中的。

否则：

- 单击**结构修饰**中的**编辑列表**按钮。
- 单击**编辑结构修饰**分类中的**添加**按钮。
- 在**名称**字段单击右边的下拉箭头，可查看 Unimod 修饰。
- 在列表中找到“脲甲基化 (C)”并单击它。

编辑结构修饰分类现在应看起来像：

编辑结构修饰

名称(N):  
Carbamidomethyl (C)

氨基酸(A): C 末端(T): 变量

化学式:  
H3C2NO

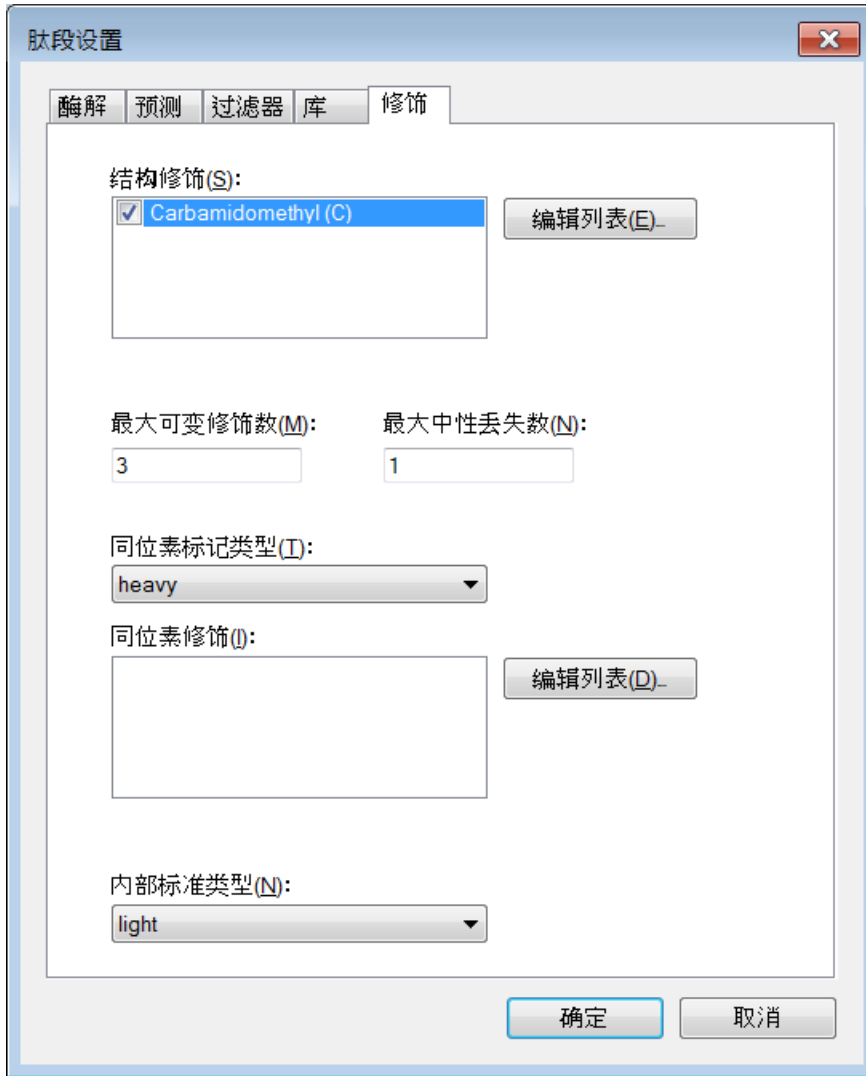
单一同位素质量(M): 57.021464 平均质量(V): 57.05162

确定 取消 丢失(L) >>

- 单击编辑结构修饰分类中的确认按钮。
- 单击编辑结构修饰分类中的确认按钮。
- 在结构修饰列表中勾选“脲甲基化 (C)”复选框。

肽设置分类应看起来像：





- 单击**确定**按钮。

最后，您需要让 Skyline 知道从每个肽段中提取哪些离子对。由于目标蛋白质非常少，所以您可以开始指定范围非常广的离子对，再稍后缩小范围：

- 在**设置**菜单中单击**离子对设置**。
- 单击**过滤器**标签。
- 在**母离子电荷**字段中，输入“1、2、3、4”。
- 在**离子电荷**字段中，输入“1、2”。
- 在**离子类型**字段中，输入“y、b、p”。（p 将添加从 MS1 谱图中提取的母离子。）
- 在**从下拉列表**中，选择“离子 3”。
- 在**至下拉列表**中，选择“最后一个离子 - 1”。
- 勾选**自动选择所有匹配的离子对**。

- 勾选使用 DIA 母离子窗口进行排除。

离子对设置分类应看起来像：

离子对设置

预测 过滤器 库 仪器 全扫描

母离子电荷(P): 1, 2, 3, 4      离子电荷(I): 1, 2      离子类型(T): y, b, p

子离子

从(E): 离子 3      至(O): 最后一个离子 - 1

特殊离子(S):

- N-terminal to Proline
- C-terminal to Glu or Asp
- iTRAQ-114
- iTRAQ-115
- iTRAQ-116
- iTRAQ-117
- TMT-126
- TMT-127L

使用 DIA 母离子窗口进行排除(U)

自动选择所有匹配的离子对(A)

编辑列表(E)

确定 取消

这些设置会使得 Skyline 包括一系列广泛的离子对，包括 MS1 和 MS/MS 扫描。使用 DIA 母离子窗口进行排除使得 Skyline 不会包括在同一 DIA 隔离窗口内的离子对（例如，如果母离子处在 500 至 520 质荷比的 DIA 窗口，那么 513 质荷比的子离子将不会被包括在内）。在一些情况下这可能是理想的排除方法，因为没有被打碎的母离子可能在原始隔离范围内的 MS/MS 谱图内造成噪声和潜在干扰，这样会使基于此范围碎片离子的定量的可靠性降低。

- 单击**确定**按钮。

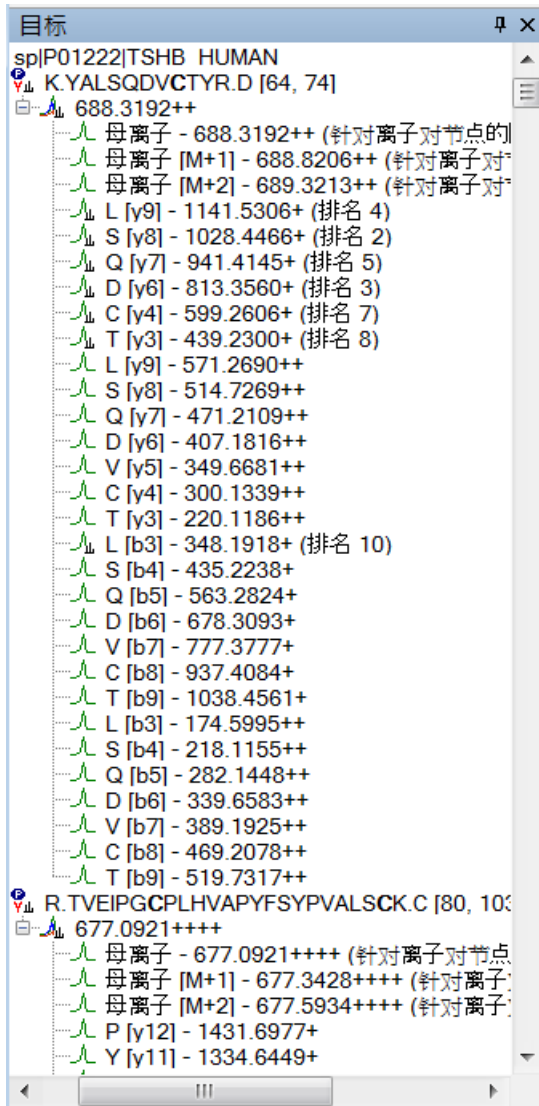
现在您已准备好可以导入 FASTA 文件了：

- 在**文件**菜单上选择**导入**，然后单击 **FASTA**。
- 在**导入 FASTA** 分类中，导航至您一直使用的教程文件夹，双击名称为“pituitary\_database.fasta”的文件。

在**目标**窗格中，将出现一份列表，包括属于 FASTA 文件中 6 个蛋白质的 26 个肽段。在每一个蛋白质中，只有存在于谱图库的肽段被包括在此份列表中。

- 在**编辑**菜单选择**展开全部**并单击**母离子**。

这样 Skyline 已包括的完整离子对集将显示出来。请注意，每一个肽段均包括在 500 至 900 质荷比之间的所有合格的母离子，且每一个母离子均包括所有可能的 b 和 y 离子对以及 3 个母离子（M、M+1 和 M+2）。另外请注意，每一个母离子均有一个相应的库谱图。**目标**视图看起来应该像：



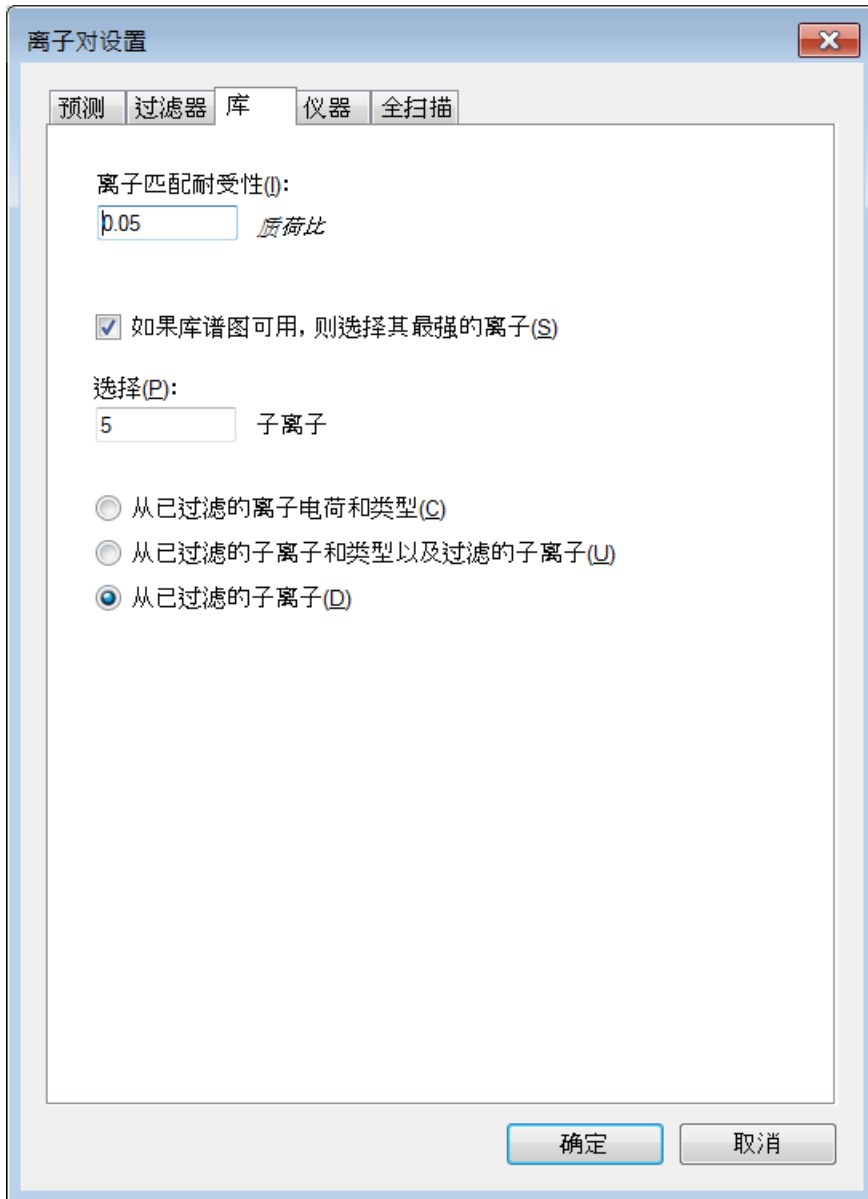
在 DIA 中，我们可能提取所有这些子离子，因为仪器采集了包括处于 500 至 900 质荷比范围的每一组可能的母离子—子离子对（至少是那些处于 MS/MS 谱图已测量范围内的）。但在实际操作中，提取每一种可能的子离子通常是不必要的，而且实际上会由于添加了一些色谱图而阻碍了目标肽段的检测，因为这些色谱图背景强，信号低。在大多数情况下，我们建议选择强度可能较高的几个离子对。您可以使用早先通过 DDA 搜索结果构建的谱图库来得到这些信息。

- 在设置菜单中单击离子对设置。
- 单击库标签。
- 选择如果库谱图可用，则选择其最强的离子。
- 在拣选字段，输入“5”来对每个母离子使用 5 个子离子。
- 选择从已过滤的子离子单选按钮。

这表示，从库中选择的 5 个子离子也必须满足在**离子对设置**的**过滤器**标签中定义的过滤器设置。

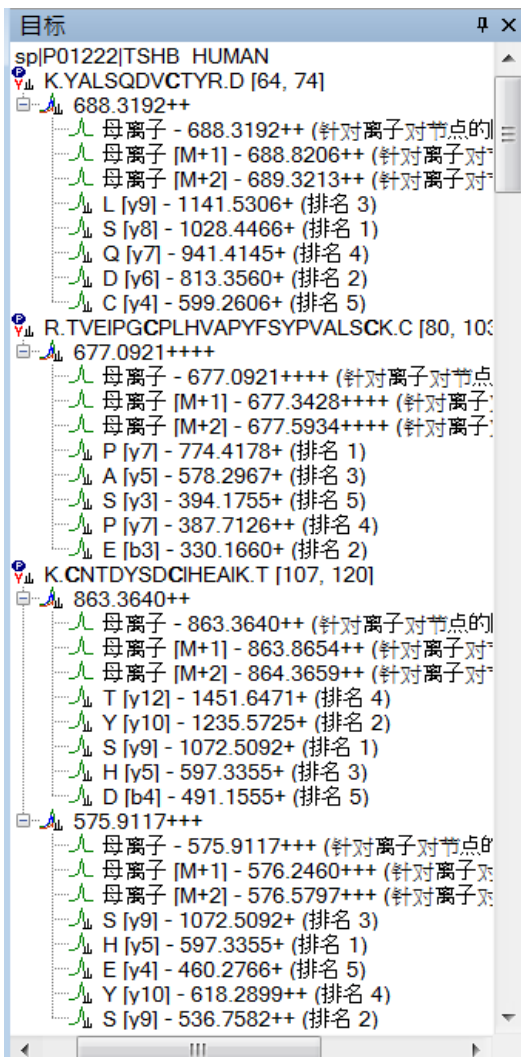
- 另外，将**离子匹配耐受性**设置为 0.05 或在 1000 *质荷比*下的 50 ppm（百万分率），由于 DDA 数据也是在高质量准确度下收集的，所以这已经相当够用了。

库标签现在应看起来像：



- 单击**确定**按钮。

再看一次**目标**窗格。现在此窗格应该只包括针对每个母离子的 5 个最好子离子，外加 3 个母离子：



除了通过 FASTA 文件，还有许多其他方法可以在 DIA 实验中把肽段、母离子和离子对导入至 Skyline 中使用。若需更多详情，请参考[靶向方法编辑](#)教程。如果您已使用其他工具创建了离子对列表，则您可以参考[现有和定量实验](#)教程。其他 DIA 工具可能创建一份称为“化验物库”的加长离子对列表格式，添加了离子对片段离子相对丰度值和保留时间预测信息。化验物库未在本教程中使用，但在[导入化验物库](#)小贴士中有详细说明。

在大型的 DIA 实验中，如果使用诱饵肽<sup>6</sup>来训练高级峰拣选（肽段检测）模型会很有帮助，这样可以提升 Skyline 自动化峰拣选能力，允许 Skyline 产生检测概率分数（q 值）。诱饵肽在 DIA 工作流程中不是非有不可的，且只在您选择应用自定义峰值得分方法时才需要。在许多情况下，Skyline 的默认峰拣选可以很好地完成工作，且并不需要自定义峰值得分。而甚至当使用自定义峰值得分时，也存在诱饵的替代方案（见于[高级峰拣选模型](#)教程）。所以在本教程中不需要向文档中添加诱饵肽。

在导入数据前，仔细考虑您的目标选择是很重要的。在您导入 DIA 数据后，如目标有任何变化，都将要求重新导入数据以用于更新的色谱图提取和自动化峰拣选。

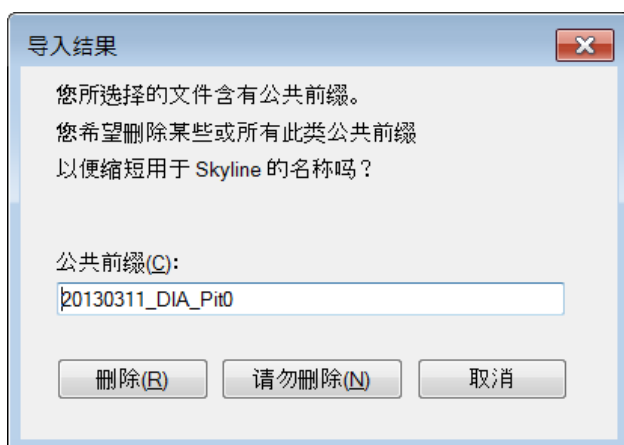
## 导入质谱数据

现在，您已经准备好可以导入本教程完整版本中包括的 DIA 质谱数据了。这些文件非常大（大约 5 GB），且仅在本教程的完整版本中提供。如果您下载的是小版本，则您应跳过本节的剩余部分并打开“DIAImported.sky”文件，这是含有完全导入 DIA 运行的 Skyline 文档。

如果您下载的是本教程的完整版本，那么您可以自己导入 DIA 数据。导入 DIA 数据非常简单，与导入 SRM 数据的步骤完全一样：

- 首先，在**文件**菜单上单击**另存为**。
- 在**文件名称**字段，将名称更改为“DIATutorial.sky”。
- 单击**保存**按钮。
- 下一步，在**文件**菜单上选择**导入**，然后单击**结果**。
- 单击**确定**按钮，在文件中添加**单次注射重复测定**。
- 选择以下的文件，使用 Shift+鼠标单击进行重选择：
  - 20130311\_DIA\_Pit01.raw
  - 20130311\_DIA\_Pit02.raw
- 单击**打开**按钮。

Skyline 将显示一个对话框，说明它可以删除文件名中的公共前缀：



- 删除文本“DIA\_Pit0”，使“20130311\_”仅成为待删除的公共前缀。
- 单击**删除**按钮。

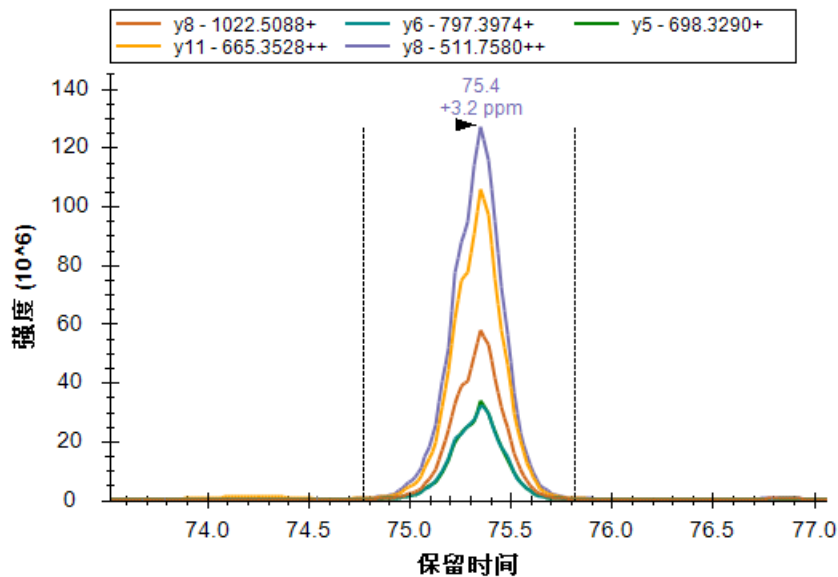
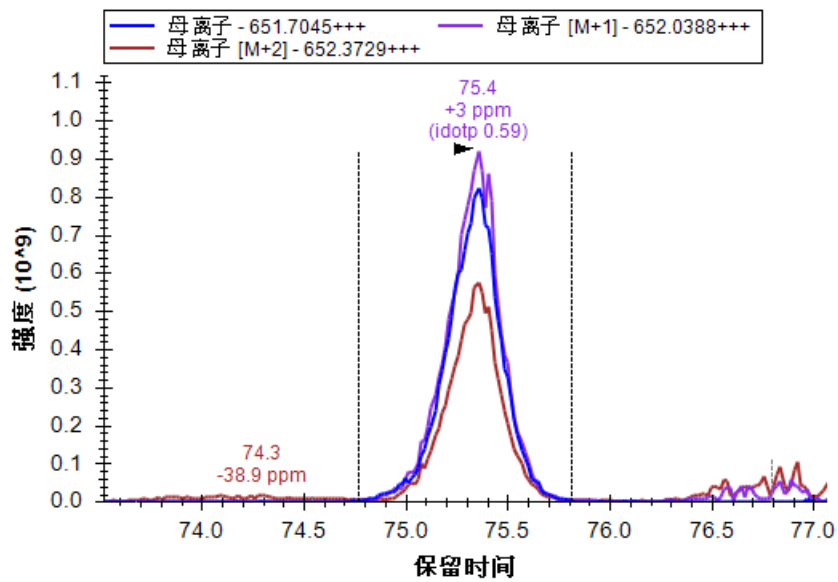
在**目标**窗格中的所有母离子和子离子的色谱图现在可以从 DIA 数据中提取。这个步骤需要几分钟时间。加载完成后，目标和诱饵肽的色谱图会显示出来。

## 研究 DIA 结果

现在 DIA 数据已导入，您可以查看一下结果：

- 确保文件 DIA\_Pit01 在色谱图窗格中已被选中。
- 在编辑菜单选择**折叠全部**并单击肽段。
- 单击目标列表中的肽段 R.VLQAVLPLPQVVCTYR.D。（应该是向下数第 7 个肽段。）
- 在视图菜单选择**离子对**并单击**分割图**。
- 在视图菜单选择**自动缩放**并单击**最佳峰值**（F11）。

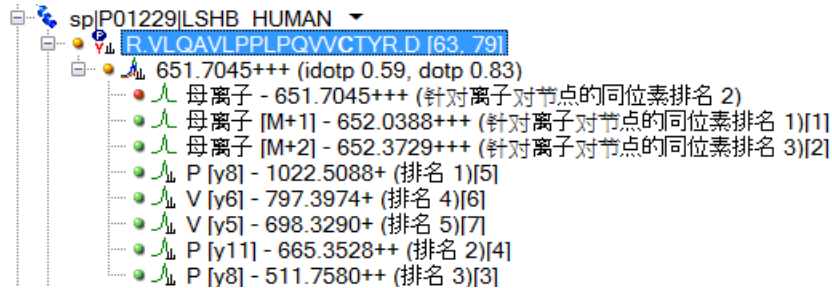
色谱图应看起来像：





注意包括母离子的图显示了“同位素点积 0.59”。它测量了母离子同位素峰面积与预期同位素分布是怎样关联的。母离子来自隔行的 MS1 扫描，这些扫描系包括在 DIA MS/MS 采集循环之间。0.59 的点积对于预期的同位素分布来说是非常低的值，它应该比来自 MS/MS 扫描的片段离子分布拥有更低的方差。

如果您展开**目标**视图中的肽段，您将看到单一同位素峰有一个红点，说明它没有峰面积。



若要使 Skyline 包括所有整合边界之间的所有已整合面积，则无论峰形状的匹配度如何，均按以下操作：

- 在**设置**菜单上单击**整合所有**。

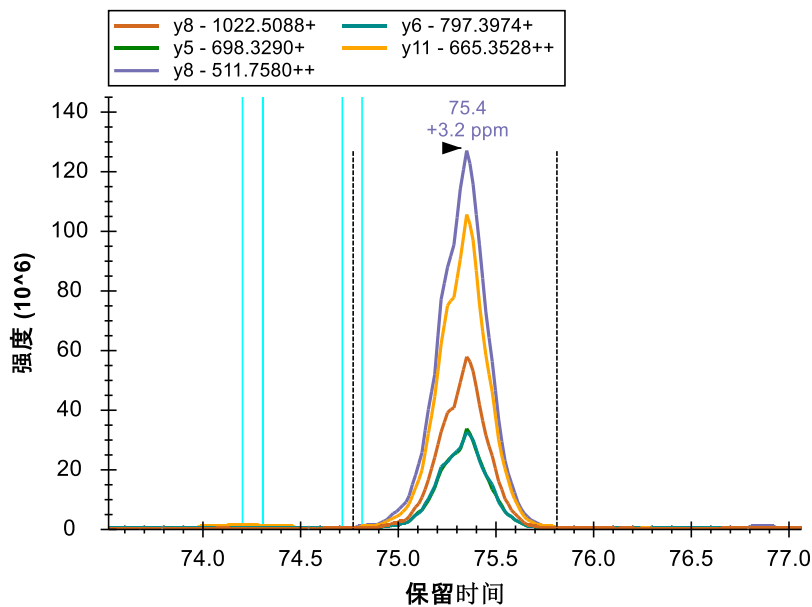
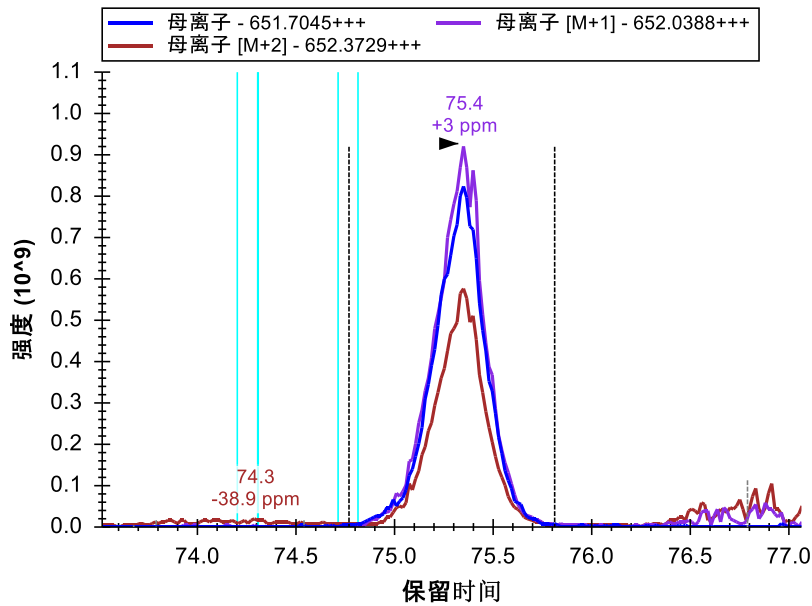
红点将变绿，**目标**视图中的同位素点积值将变为“同位素点积 0.99”。由于此同位素点积值是现在所有在已提取色谱图中检测到峰组的最佳匹配分布，所以它也将从色谱图视图中消失。

“点积 0.83”在**目标**视图中位于同位素点积值的后面，它是用来衡量 5 个已选的子离子峰面积与在匹配库谱图中发现的强度的关联程度。两个值现在相对来说都很高，并可帮助确认这个峰组和目标肽段的关系。这两个高点积的组合、共同洗脱的离子对和峰形以及大约 3 ppm 的质量误差，这一切可以使您确定是目标肽段的洗脱形成了这个峰值。

现在按以下步骤可显示用于预测保留时间的肽段 ID 时间：

- 右键单击色谱图图形，选择**肽段 ID 时间**并单击**来自其他运行**。

在色谱图窗口中应显示一组淡蓝色垂直线，如下：



淡蓝色线表示 MS/MS 谱图匹配的肽段的保留时间，这个肽段是您在曾使用的生成谱图库的 DDA 运行时采集得到的。它们都在整合范围的 1 分钟内，这更加确定了整合的峰对应了相关的肽段。（请注意，要查看所有匹配的 MS/MS 谱图的保留时间不一定需要一个冗余库。所有保留时间均保存在非冗余库，但每个肽母离子只有一个谱图。）

Skyline 峰的自动拣选在决定拣选哪个峰时会同时考虑到峰面积点积和 ID 保留时间。如您所看到的，在本例中它已选择了正确的峰。

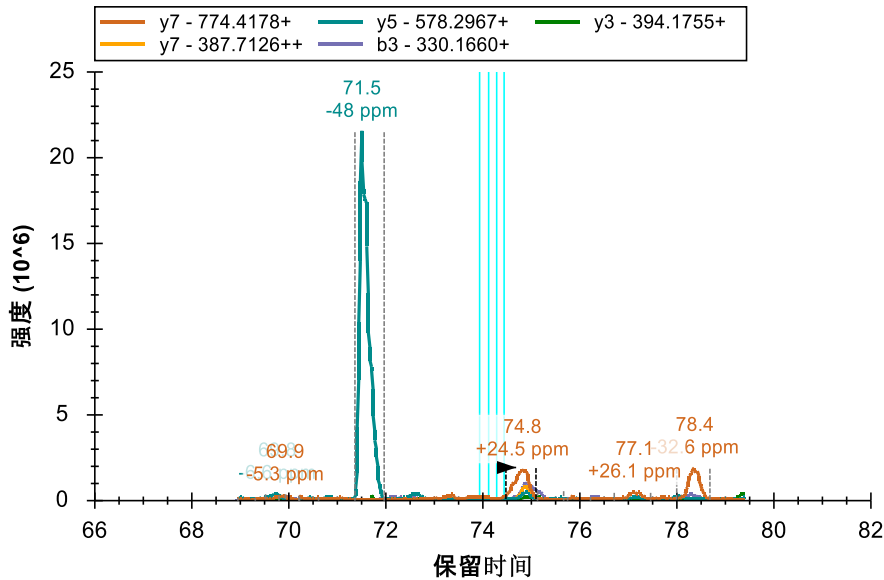
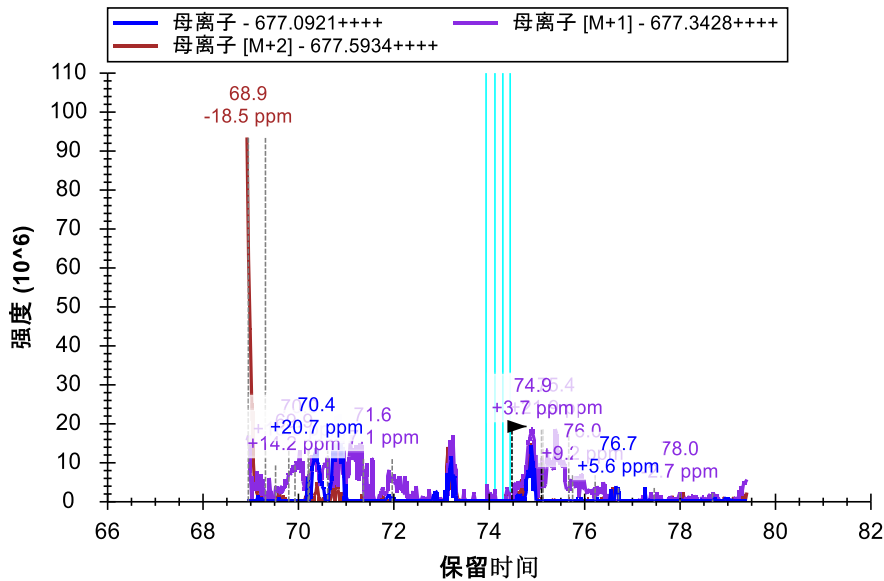
按以下步骤可查看已提取色谱图的完整范围：

- 在视图菜单选择**自动缩放**并单击**无** (Shift-F11)。

这样您会看到，Skyline 仅在大约 DDA ID 时间左右的大约 5 分钟窗口中提取了色谱图 — ID 时间发生在将近 75 分钟，所以 Skyline 是在 70 至 80 分钟之间进行的提取。

在 DIA 中会有许多干扰，因为母离子隔离窗口非常宽（例如 10-25 质荷比）。但尽管会存在许多干扰，Skyline 自动化峰拣选也通常能够拣选正确的峰。若要看实际运行情况：

- 单击**目标**列表中的第 2 个肽段 R.TVEIPGCPLHVAPYFYSYPVALSCK.C。



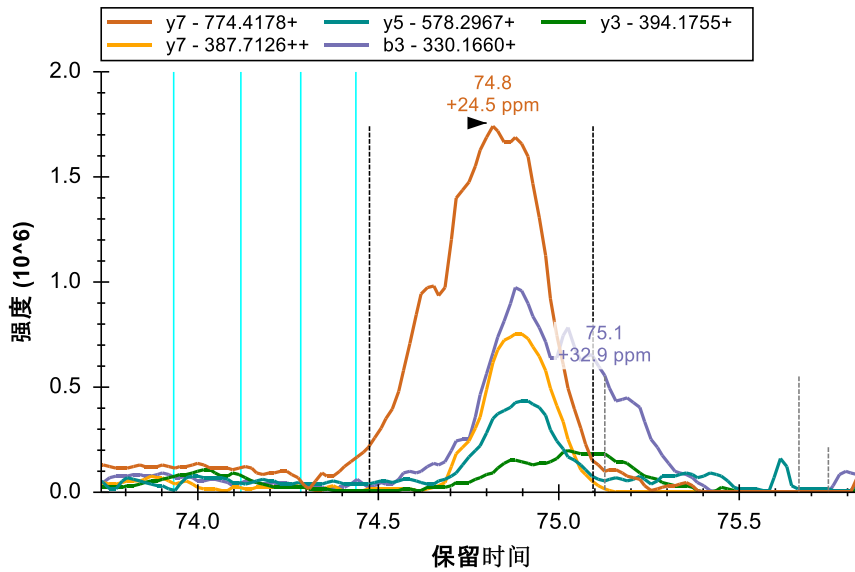
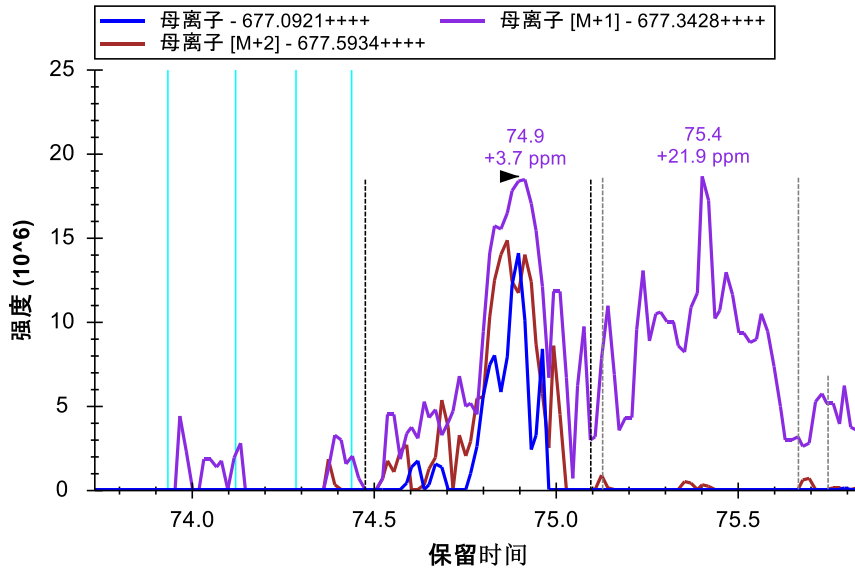
如您所见，有许多峰组可以从 MS1 色谱图中选择。在 MS/MS 色谱图中，在 71.5 分钟时有一个离子对出现高峰，在 75 分钟时出现一组强度低很多的共洗脱峰，更接近了预测的 RT 且点积很高。Skyline 不会被大峰值分散重心（它很可能成为干扰），并会选择小型的共洗脱峰（这的确是最有可能的选项）。

## 理解提取的色谱图

子离子峰仍显示了超过 24.5 ppm 的质量误差，而这对于 Orbitrap 质量分析仪来说，几乎不会发生。按以下步骤可更好地理解造成问题的原因：

- 在**视图**菜单选择**自动缩放**并单击**最佳峰值** (F11)。

在更近距离查看整合的峰时，您会看到像这样的内容：



看到它可不会像看到 R.VLQAVLPPLPQVVCTYR.D 的峰时那样令人安心。尽管在 75 分钟 RT 附近的相关母离子左右很明显地看到在 MS1 扫描上有很多活动，但整合的峰却只有 3.7 ppm 的质量误差。如果您展开目标视图中的肽段，您会看到它有一个 0.91 的同位素点积。如果您将鼠标光标悬浮在母离子图第 74.5 分钟的整合边界，直至看到一个分割光标 (⊕)，向右拖至大约 74.7 分钟，接近于主母离子峰的左边，那么您会看到同位素点积值上升到 0.96、质量误差降低到 2.3 ppm。

**注释：**只有当您下载的是完整教程数据集时，下方的全扫描视图才可以全部使用。否则 S kyline 将显示一条消息，说明原始数据文件缺失。

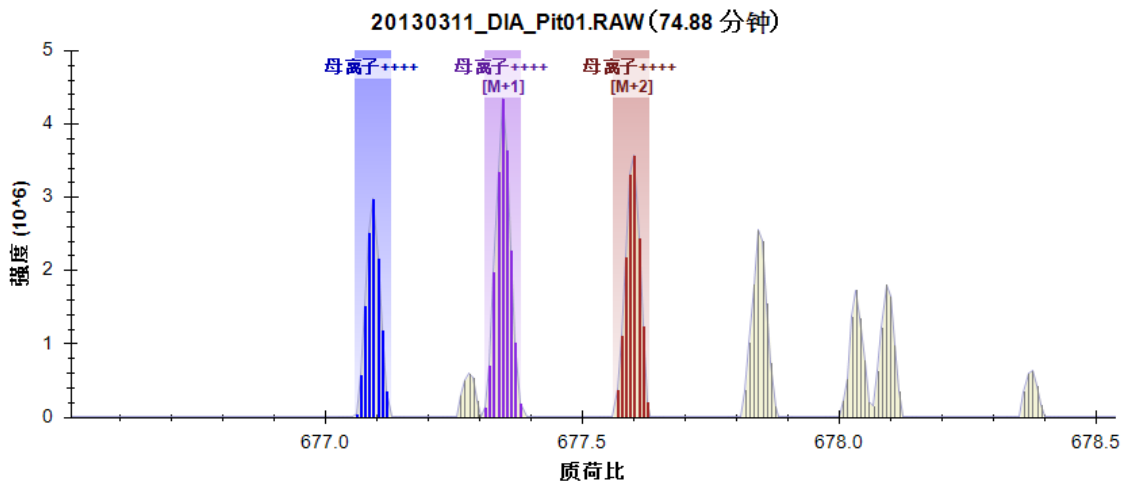
按以下步骤可更好地理解在 MS1 扫描中发生了什么情况（色谱图即从这些扫描中提取）：

- 将鼠标光标悬浮在“母离子 [M+1]” 峰的顶点，直到看到紫圈出现且光标变为指向手。
- 单击此圈。

Skyline 将显示**全扫描**视图，其中显示了一系列的 MS1 扫描，而母离子色谱图点即是从中提取的。单击视图右上角的放大镜加号 (🔍) 按钮，您可以看到整个 MS1 扫描。或者按以下操作：

- 单击并拖拽标记为 M、M+1 和 M+2 峰周围的矩形来放大。
- 确保您查看的是正好在 74.88 分钟采集的谱图（在标题有指示），或者使用右上角的箭头导航至此谱图。

**全扫描**图现在看起来应像这样：



个体强度在分布图模式峰中呈长条状，有一条线和阴影将其连接起来。汇总到一起在色谱图上创建点的个体强度用色谱图颜色标注了出来，提取范围以阴影区域表示。您可以看到 M+3 和 M+4 峰，它们在阴影峰的右面，没有被提取。这看起来应该就是我们预期的目标肽段，但您也能看到两个峰（677.275 和 678.025）已非常有可能干扰肽段的 M+1 和 M+4 峰。

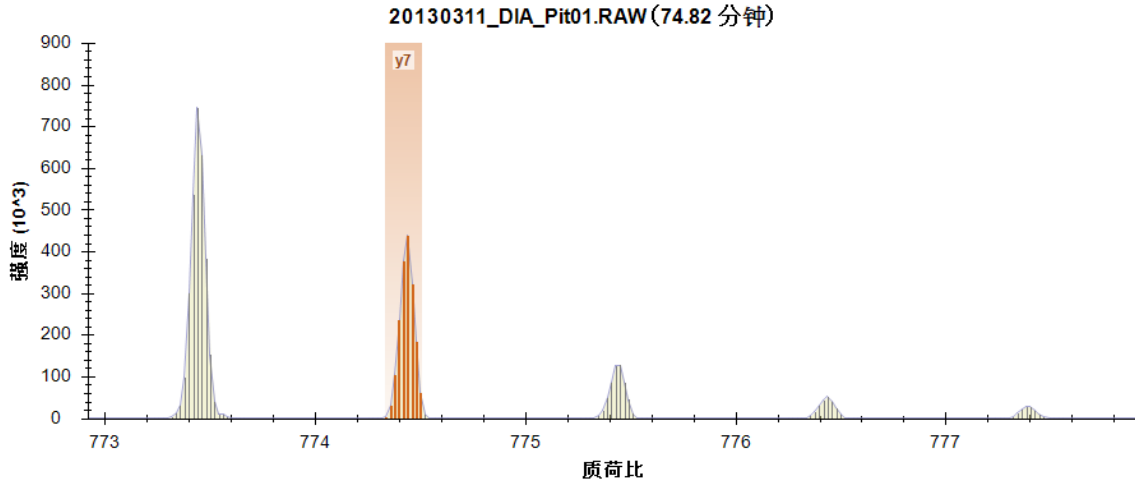
- 现在使用**全扫描**视图工具栏中的箭头按钮 (← →) 可查看色谱图峰洗脱分布图的扫描。

您可以看到，从大约 74.77 分钟至 74.92 分钟内的 MS1 信号与预期同位素分布非常吻合，没有任何明显干扰到提取范围的因素。在 74.92 分钟之后，信号则变得不太清晰。如果您缩小一点（使用鼠标滚轮），那么您会看到，最终其他清晰的肽段信号开始占据此质荷比范围。

若要对子离子 MS/MS 扫描进行相似的分析：

- 将鼠标光标悬浮在“y7+” 峰的顶点，直到看到紫圈出现且光标变为指向手。
- 单击此圈。

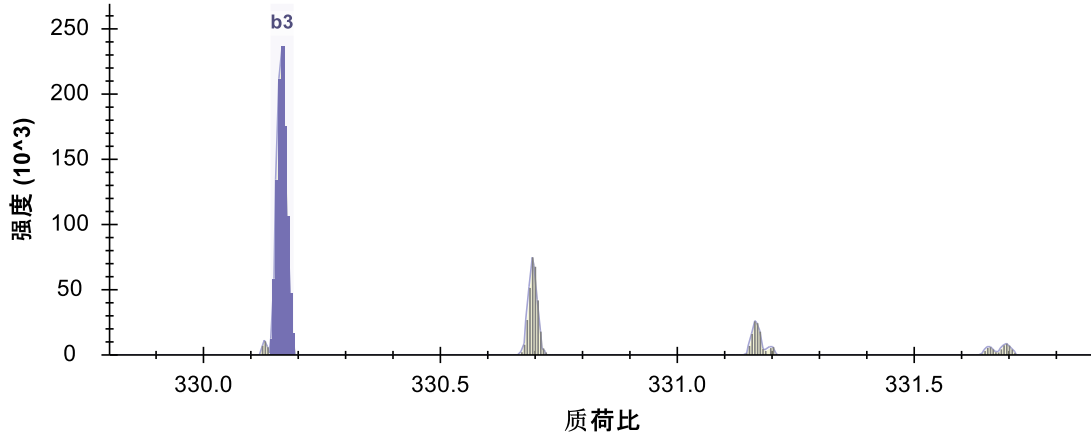
您会在**全扫描**视图中看到一个图形，如下所示：



显而易见，y7 色谱图中的信号不是来自于目标片段，因为 y7 提取范围并不在单一同位素峰上。相反，您会看到 y7 提取范围在第二个同位素峰上，而单一同位素峰则轻了 1 Da（道尔顿），再加上 y7 色谱图与其他子离子和母离子色谱图并没有很好地共洗脱，导致向左偏离太多，所以您也许想抛弃这个离子对（在**目标**窗格中右键单击 y7 并单击**删除**）。而事实上，它是预期库 MS/MS 谱图中强度最高的片段，而点积与库谱图的关联当前是 0.88，这当然会让人有些困惑。

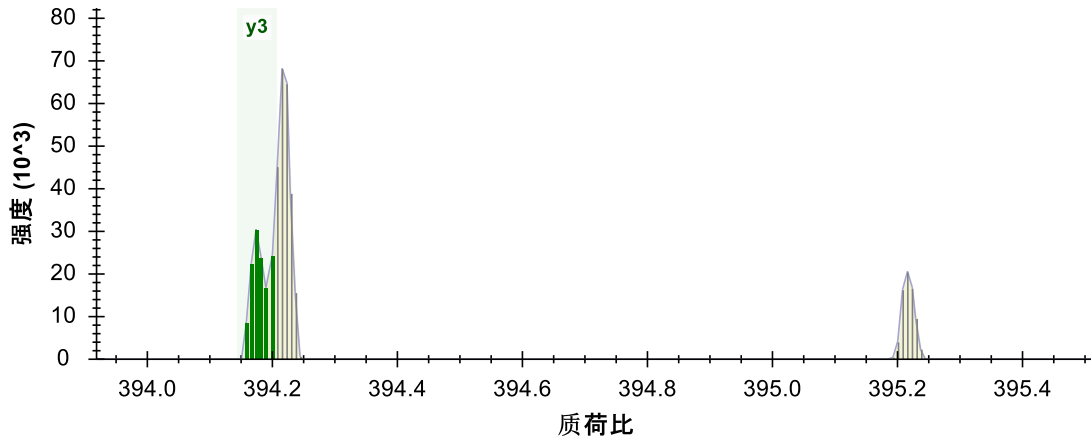
接下来，如果您单击 b3+ 离子的色谱图并缩放，则您会看到它的信号来自一个双电荷离子（同位素峰之间 0.5 质荷比）：

20130311\_DIA\_Pit01.RAW (74.89 分钟)



y7<sup>++</sup> 和 y5<sup>+</sup> 的峰看起来很合理，但 y3<sup>+</sup> 的色谱图很明显在洗脱过程中受到了干扰，而最终在大约 74.92 分钟时被干扰控制。

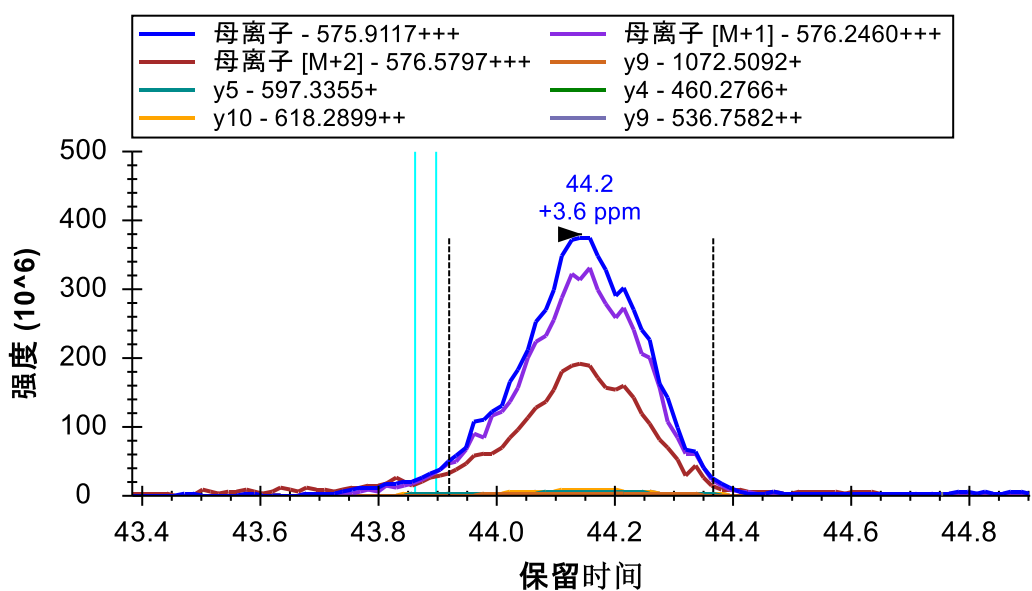
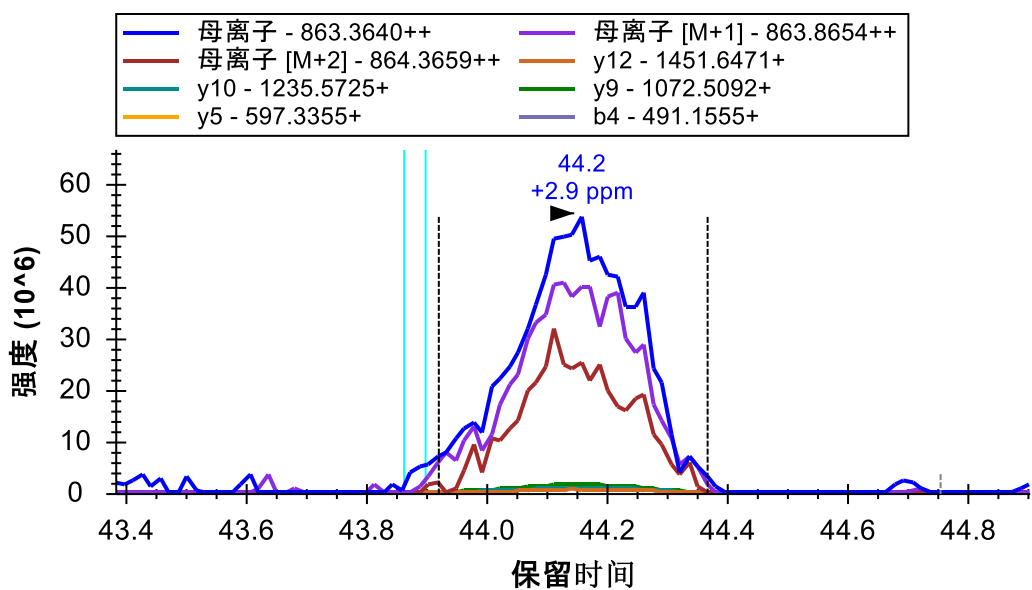
20130311\_DIA\_Pit01.RAW (74.92 分钟)



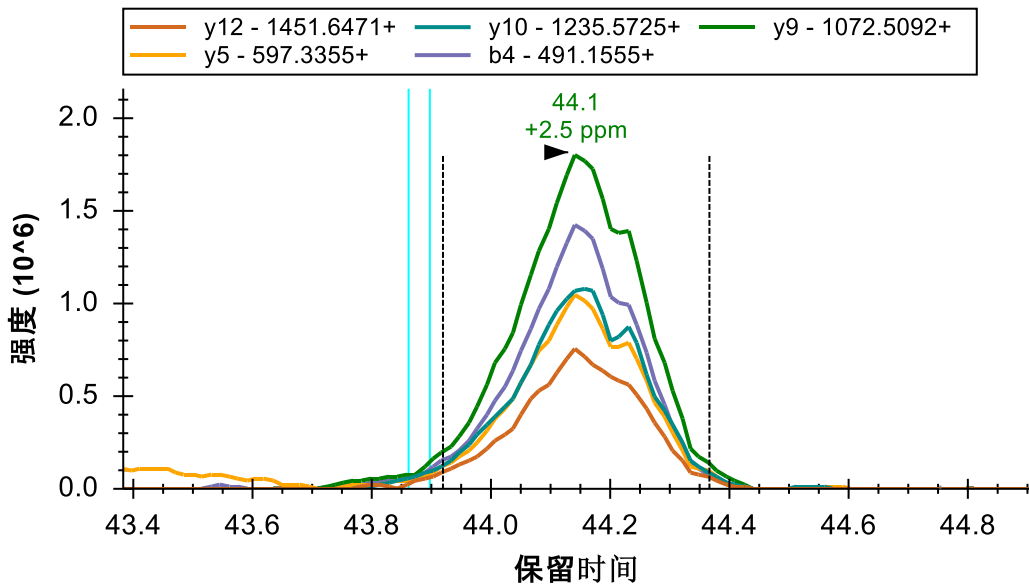
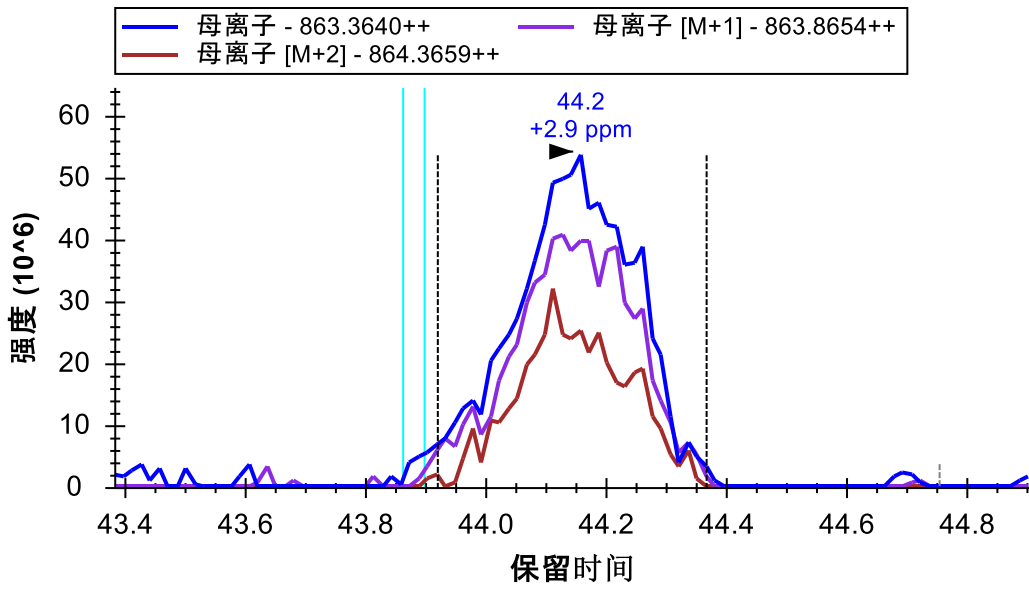
到这里，您可能在这样的环境下从可计量的角度考虑想抛弃这个肽段。虽然我们有可能对此肽段的新片段离子重新运行色谱图提取，但您不应在本教程中这样做，因为此操作将要求您首先在**目标**窗格中选择新离子对，然后再重新导入 DIA 文件。

继续前进到肽段 K.CNTDYSDCIHEAIK.T，您会看到当有多个电荷态时，离子对分割视图会按电荷态而非离子类型分割，并且从 MS1 扫描提取出的母离子强度通常会超过从 MS/MS 扫描中提取的子离子强度：



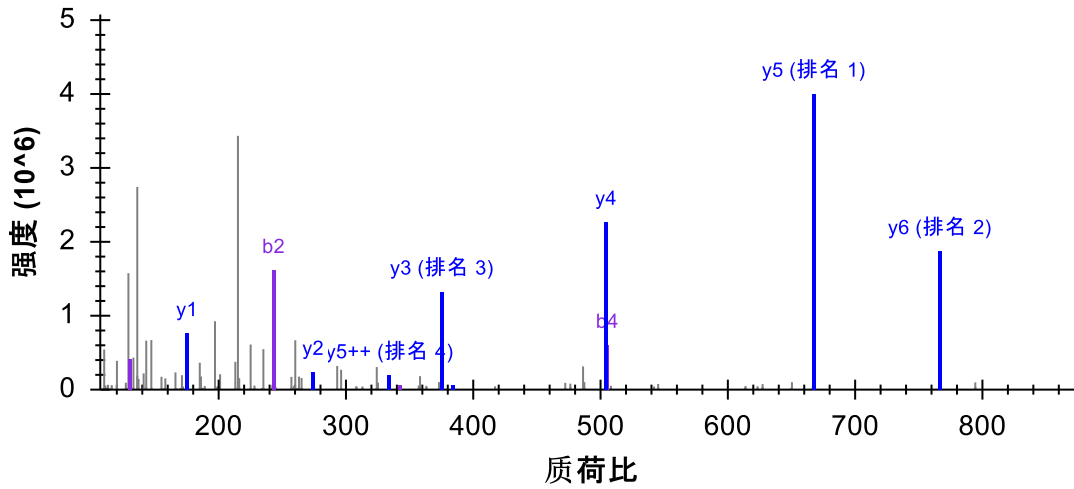


如果您展开目标视图中的肽段再选择每个母离子，则您会看到两个均有高质量的峰，而母离子和子离子离子对都有低噪声的良好信号、质量误差在 5 ppm 以下、同位素点积值接近 1 且点积值大于 0.8:



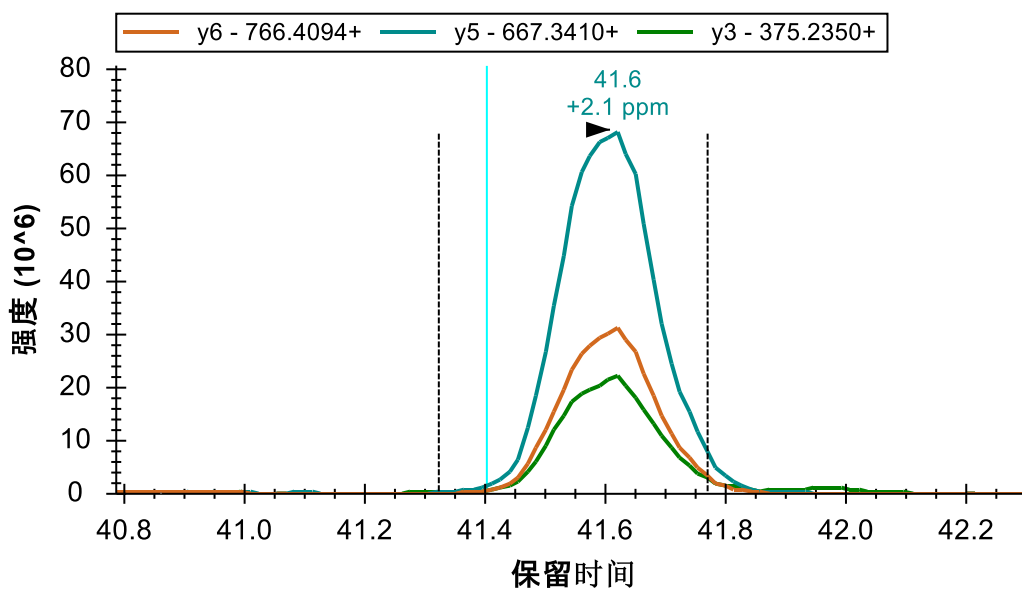
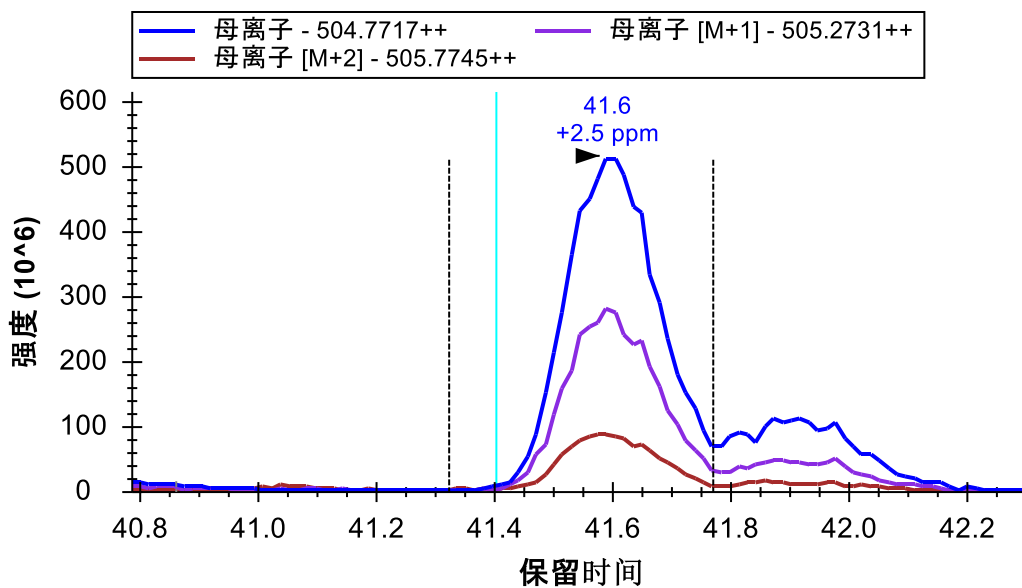
在本教程中最后一个值得研究的肽段即是下一个肽段 K. ELVYETVR. V。在目标视图中选择它并展开，您会看到 0.99 的同位素点积和点积值，但只有 3 个子离子离子对。您可能会奇怪为什么没包括 y4 离子对，在库匹配视图中它貌似是显示在谱图中的第二高选项：

## Dia\_Tutorial\_Library - ELVYETVR, 电荷 2

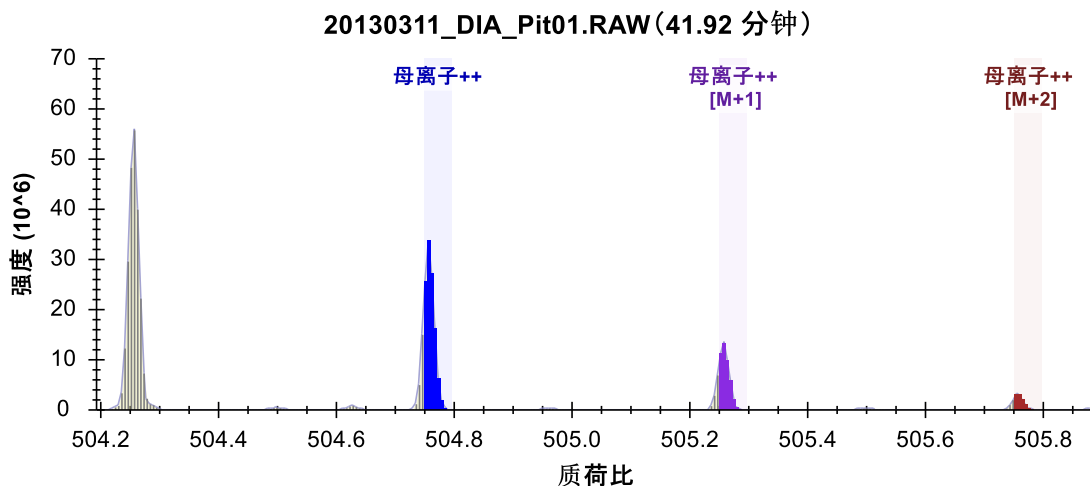


这个问题的答案就在于，由于您已在**离子对设置** — **过滤器**标签中勾选了**使用 DIA 母离子窗口进行排除**的复选框，而母离子**质荷比** 504.77 和子离子**质荷比** 504.28 均处于 500 至 520 隔离窗口中，所以它被排除了。这样，对于此肽段，您只有 y3、y5 和 y6 这些不太理想的选择了。您更应该倾向于对每个肽段使用 5 个或甚至 6 个离子对。

而色谱图看起来却很不错。现在位于熟悉的峰左边的共洗脱母离子和附近的蓝绿色 ID 线会让您对峰值分配更有把握。



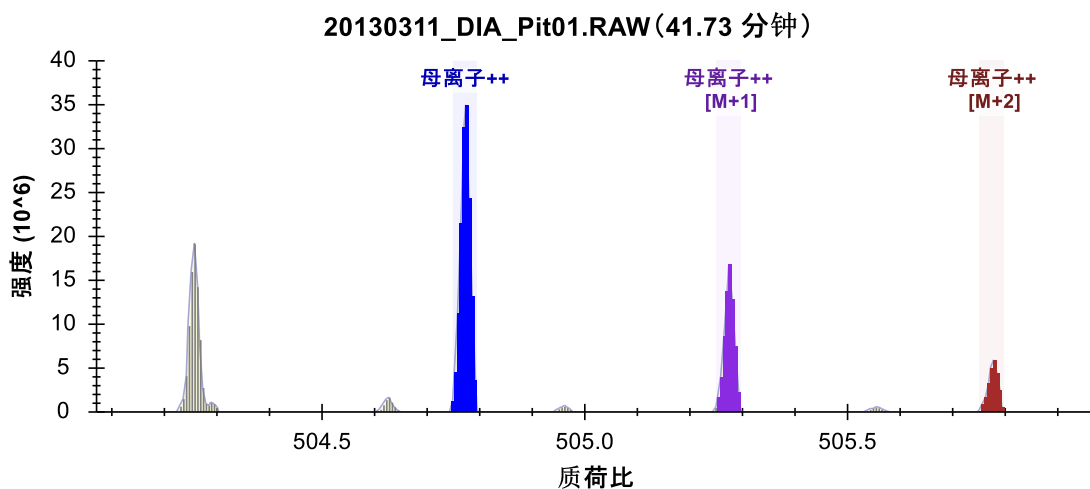
在母离子色谱图中，从主峰延伸出的第二个峰是一个干扰，而在子离子色谱图中却完全没有这个峰，这使得这个事实更明显。您可以单击它使其在**全扫描**视图中显示 MS1 扫描，从而更加有把握。您会看到信号来自比目标肽轻 1 Da 的肽段，质量误差明显高于 5 ppm:



您可能会想（而且想得相当正确），如果您已使用了带有质心质谱（它会将每个分布图峰转变为一个单独的条形）的文件和一个非常窄的提取窗口，那么这个峰就会消失。您可以向 Skyline 提供质心数据并使用非常高的分辨力从而在 Skyline 中完成此操作。但您应注意，由于数据实际上是由仪器采集的，所以数据的可视性会降低。另一个需记住的重点是，无论您是否已使用了质心算法，在质荷比维度中的真正分布图峰分辨率决定着干扰何时会出现。若要通过视觉理解：

- 单击**全扫描**视图中的向左箭头 (←)，移动 MS1 扫描直至到达在干扰和真正肽段峰之间的离子对。

您会看到来自目标肽段的信号开始扩张，而峰中心会早在左边的峰消失前向提取范围的中心移动，这使得您清晰地看到，两个肽段的信号在重叠处是不可分离的。



在色谱图提取前使用质心法可以减少在目标肽段峰外看到的噪声，但却无法移除相关峰内干扰的信号因素。

## 结论

在本教程中，您定义了一项用于生成 DIA 仪器方法或仅仅数据分析的 DIA 采集方案。您根据 DDA 搜索结果构建了一个谱图库、根据库中的已测量 RT 设置了保留时间限制、根据相关蛋白质和来自 DDA 运行匹配谱图定义了一组要被测量的离子对、导入了 DIA 运行并分析了结果数据质量。最终与所有 Skyline 文档一样，您可以生成相关肽段的峰面积和统计信息。在本教程中亦提及了其他几个教程以用于进阶学习（[靶向方法编辑](#)、[现有和定量实验](#)、[iRT 保留时间预测](#)、[高级峰拣选模型](#)、[导入化验物库](#)和 [Panorama 色谱图库](#)）。

如果在 DIA 运行前，运行间，或运行后采集相应的 DDA 数据，则这个方法可使您在 Skyline 中分析任何 DIA 数据集。目前这是 DIA 数据分析的最简单方法。但是，如果您或其他人已为相关的蛋白质构建了含有片段离子相对丰度值信息和标准化保留时间的库，那么您便可以使用这些库而不用从特定实验的 DDA 运行中采集和提取信息。为 DIA 实验构建和使用此类库是更高阶段的内容，需要一份单独的教程学习，或者您需要将其他 Skyline 和 Panorama 教程的信息汇总到一起，这些教程已在这里提及。但在完成了本教程后，您应该掌握了使用 DIA 和 Skyline 用于定量蛋白质组分析的至少一组完整 workflow。

## 参考文献

1. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39 - 45 (2004).
2. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics MCP* **11**, 0111.016717 (2012).
3. Egertson, J. D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* **10**, 744 - 746 (2013).
4. Krokhin, O. V. *et al.* An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics MCP* **3**, 908 - 919 (2004).
5. Escher, C. *et al.* Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics Accept.* (2012).
6. Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8**, 430 - 435 (2011).