Skyline 处理已有定量实验数据

本教程介绍了如何处理不是基于Skyline创建的离子对列表进行的选择反应监测实验（简称 SRM；亦称为多重反应监测，简称 MRM）数据。本教程还介绍了如何使用 Skyline 处理加入了同位素标记内标肽段的定量实验数据。

本教程将处理工具MRMer1 发表时采用的数据，还将处理由 Addona 及其合作者2开展的跨实验室研究（以下简称“study 7”）所获取的数据。Study 7 是美国国家癌症研究所 (NCI) 支持的癌症临床蛋白质组学技术评估(CPTAC) 的一部分。在Study 7计划中，CPTAC验证工作组优先采用 Skyline 作为实验室间大规模研究的工具。因此，研究计划累积的数据为Skyline提供了完善的测试，确保Skyline能支持验证工作组正在进行的实验类型。

即使缺少非Skyline创建的离子对列表，您仍然可以了解Skyline处理同位素标记肽段作为内标的LC-MRM 实验的数据的特点。特别是，如果您想以极高的可信度鉴定肽段的色谱峰，并匹配参照肽段，Skyline提供了良好的支持，能够帮助您实现这些目标。

# 入门指南

如要开始本教程，请下载下面的 ZIP 文件：

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

将其中的文件解压到您的电脑上合适的文件夹中，如：C:\Users\brendanx\Documents

该操作将创建一个新文件夹：C:\Users\brendanx\Documents\ExistingQuant

现在启动 Skyline，将自动创建一个新的空白文档。

导入现有离子对列表之前，您应当尽可能多地向 Skyline 提供关于离子对列表设计和数据采集实验方面的背景信息。要做到这一点，您可以在导入离子对列表之前，尝试调整文档的设定值。

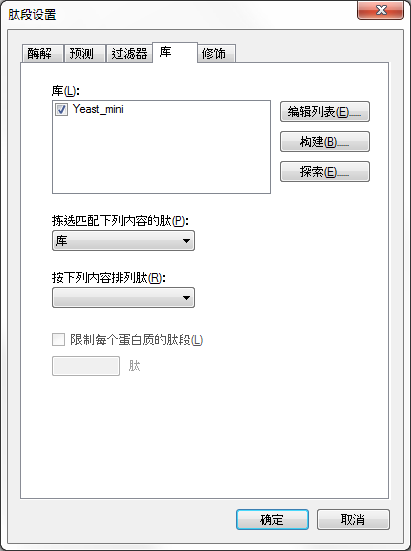
# 准备接受离子对列表的文档

本教程开始时，首先要看看名为 MRMer1（其发音如英文单词的“murmur”）的 MRM 分析软件工具提供的数据集。在查看和积分整合MRM色谱图方面，MRMer是Skyline的前身。MRMer发表时采用的数据集产生于2008年，可以从其官方网页(http://proteomics.fhcrc.org/CPL/MRMer.html) 下载。在 MRMer 数据集中，所有肽段均源自于酵母菌，并且对应有国家标准与技术研究所 (NIST) 谱图库中的 MS/MS 谱图。这就意味着可以通过谱图库和背景蛋白质组，非常容易地向Skyline提供实验中所监测肽段的相关信息。“[靶向方法编辑](https://skyline.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=tutorial_method_edit)”教程详细介绍了创建谱图库和背景蛋白质组文件的方法。在本教程中，您将使用已经准备好的文件，这些文件已经被压缩至完成本教程所需的最小信息量，以便下载的 ZIP 文件大小能控制在合理范围之内。

设置 MRMer文档的谱图库请按以下步骤操作：

* 在“设置”菜单上，单击“肽段设置”。
* 单击“库”标签。
* 单击“编辑列表”按钮。
* 单击“编辑库”表单中的“添加”按钮。
* 在“编辑库”表单的“名称”字段内输入“Yeast\_mini”。
* 单击“浏览”按钮。
* 导航至先前创建的ExistingQuant文件夹下的MRMer子文件夹。
* 选择仅包含本研究中使用的酵母肽谱图的“Yeast\_MRMer\_mini.blib”文件。
* 单击“打开”按钮。
* 单击“编辑库”表单中的“确定”按钮。
* 单击“编辑库”表单中的“确定”按钮。
* 选中“库”列表中新的“Yeast\_mini”库。

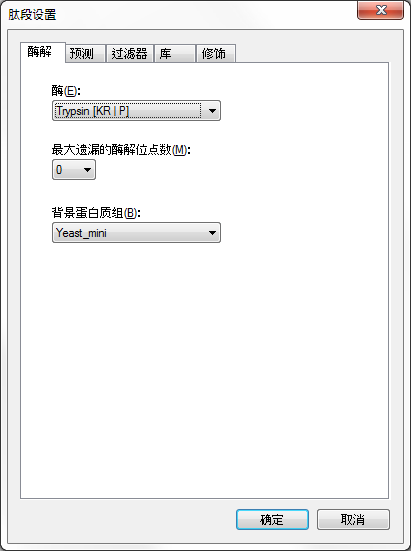
“肽段设置”表单将显示如下：



现在设置MRMer文档的背景蛋白质组，请按以下步骤操作：

* 在“肽段设置”表单中，单击“酶解”标签。
* 在“背景蛋白质组”下拉列表中，选择“<添加……>”。
* 在“编辑背景蛋白质组”表单的“名称”字段中输入“Yeast\_mini”。
* 单击“浏览”按钮。
* 导航至先前创建的ExistingQuant文件夹下的MRMer子文件夹。
* 选择“Yeast\_MRMer\_mini.protdb”文件。
* 单击“打开”按钮。
* 在“编辑背景蛋白质组”表单中单击“确定”按钮。

“肽段设置”表单将显示如下：

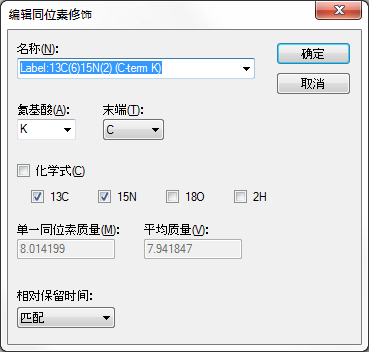


最后，在将MRMer实验的离子对列表导入当前文档之前，您还需要定义肽段轻重同位素修饰。MRMer实验包括未标记的轻肽段和对应的重肽段，后者使用细胞培养的稳定同位素氨基酸标记 (SILAC)方法对赖氨酸和精氨酸进行了标记。如果事先没有指定正确的同位素修饰， Skyline 将无法识别出离子对列表中重链肽的质荷比值。

要在Skyline文档设置中指定SILAC 标记，请执行下列步骤：

* 单击“修饰”标签。
* 单击“同位素修饰”列表旁边的“编辑列表”按钮。
* 单击“编辑同位素修饰”表单中的“添加”按钮。
* 在“编辑同位素修饰”表单的“名称”字段内输入“Label:13C(6)15N(2) (C-term K)”。
* 在“氨基酸”字段中输入“K”。
* 在“末端”下拉列表上，选择“C”。
* 选中“13C”和“15N”复选框，告诉Skyline对赖氨酸分子中的所有碳原子使用13C，所有氮原子使用 15N，总质量偏移为 8 Da (6x(13C-12C)+ 2x (15N-14N))。

此时“编辑同位素修饰”表单将显示如下：

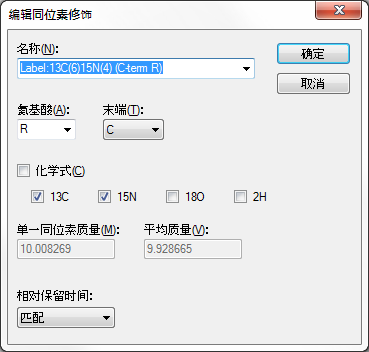


单击“确定”按钮，再通过以下步骤添加第二个同位素修饰：

* 单击“编辑同位素修饰”表单中的“添加”按钮。
* 在“编辑同位素修饰”表单的“名称”下拉列表中选择“Label:13C(6)15N(4) (C-term R)”。

自动选中“13C”和“15N”复选框，告诉Skyline对精氨酸分子中的所有碳原子使用13C，所有氮原子使用 15N，总质量偏移为 10 Da ((6x(13C-12C)+ 4x (15N-14N)))。

此时“编辑同位素修饰”表单将显示如下：



Skyline 自动计算单一同位素质量和平均质量偏移，如图，赖氨酸(K)质量偏移 约为 8 Da，精氨酸(R) 约为10 Da，这是由于在这些氨基酸残基中使用了13C和15N。

MRMer离子对列表导入当前文档的准备工作如下：

* 单击“编辑同位素修饰”表单中的“确定”按钮。
* 单击“编辑同位素修饰”表单中的“确定”按钮。
* 选中“同位素修饰”列表中您刚创建的“Label:13C(6)15N(2) (C-term K)”和“Label:13C(6)15N(4) (C-term R)”修饰对应的复选框。
* 还要确保选中“结构修饰”列表中的“脲甲基半胱氨酸”（烷基化）复选框，注意在新安装的 Skyline下此项已为默认设置。
* 在“肽段设置”表单中，单击“确定”按钮。

您将看到Skyline文档区域出现的空白谱图。现在您已经为导入MRMer离子对列表做好了准备。

# 导入具有对应蛋白质的离子对列表

将离子对列表导入Skyline的方法有两种：

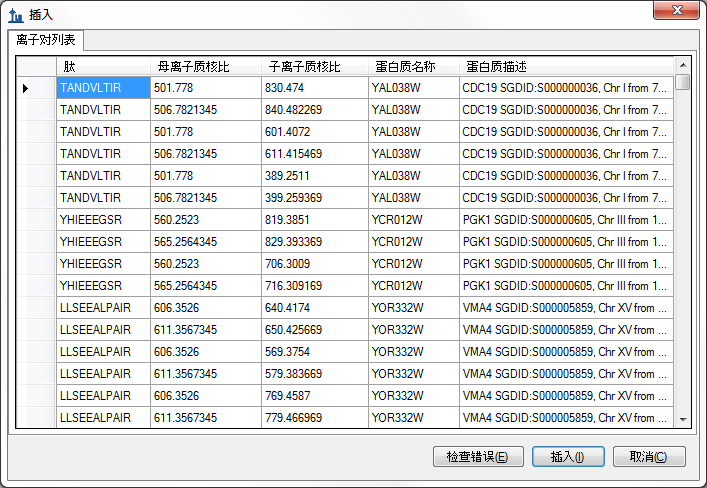
1. 单击“编辑”>“插入”>“离子对列表”，插入到所显示的表单。
2. 单击“编辑”>“粘贴”，直接插入到文档。

对于MRMer数据集，可使用第一种方法。对于Study 7数据集，可使用第二种方法。注意，如果您的文档包括背景蛋白质组，应采用“插入”表单的第一种方法，其优势在于可以将肽段与包含这些肽段的蛋白质自动关联。Skyline当前仅支持肽段在背景蛋白质组中对应单一蛋白质的情况，未来版本将能够处理一个肽段对应多个蛋白质的情况。MRMer离子对列表里面有两条肽段对应到了多个蛋白质，在本教程中已被删除。

如要向当前文档添加离子对列表肽，请执行下列步骤：

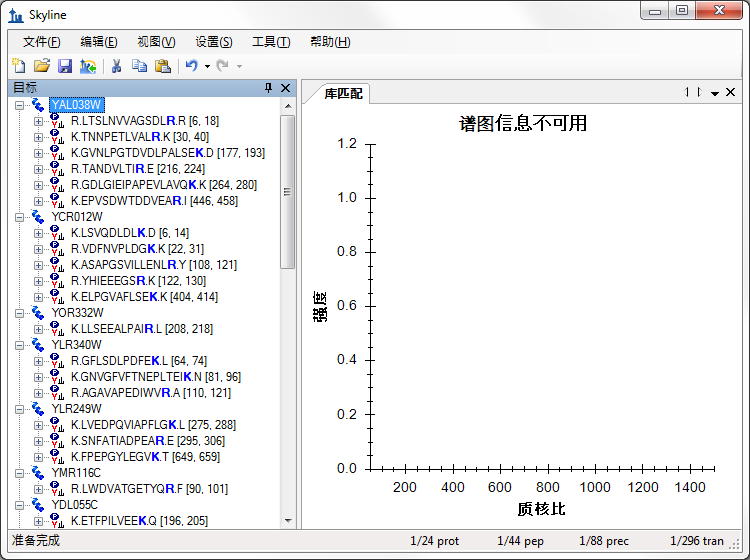
* 使用Microsoft Excel打开为本教程创建的ExistingQuant文件夹下的MRMer子文件夹中的“silac\_1\_to\_4.xls”文件。此Excel文件创建自MRMer下载的“silac\_1\_to\_4.transition.tsv”（制表符分隔值）文件，包含了轻重标按照1:4进行混合的实验。
* 确保在水平滚动条左侧标记有“固定”的电子表格中的页面处于活动状态。
* 选择所有 296 行的前 3 列单元格：肽段序列（A 列）、母离子质荷比（B 列）、子离子/片段离子质荷比（C 列）。
* 复制单元格 (ctrl-C)。
* 切换回 Skyline。
* 在“编辑”菜单上，选择“插入”，并单击“离子对列表”。
* 在键盘上按下 ctrl-V，粘贴复制的单元格。

“插入”表单现在将显示如下：



已添加单元格，以及相关蛋白质的名称和描述。Skyline通过将肽段序列与背景蛋白质组文件中的蛋白质序列进行匹配，自动添加蛋白质信息。如要向Skyline文档中插入这些肽段，请单击“插入”按钮。

Skyline主窗口将显示若干按照从属蛋白质分组的肽段。肽段图标应包含3条垂直线和一条基线 ()，就好像在其右下角有一个非常小的MS/MS谱图 ()。出现此图标表明肽具有相关的MS/MS 库谱图。标记肽段C-末端赖氨酸或精氨酸以蓝色突出显示，标明重链标记形式中稳定同位素标记的氨基酸：

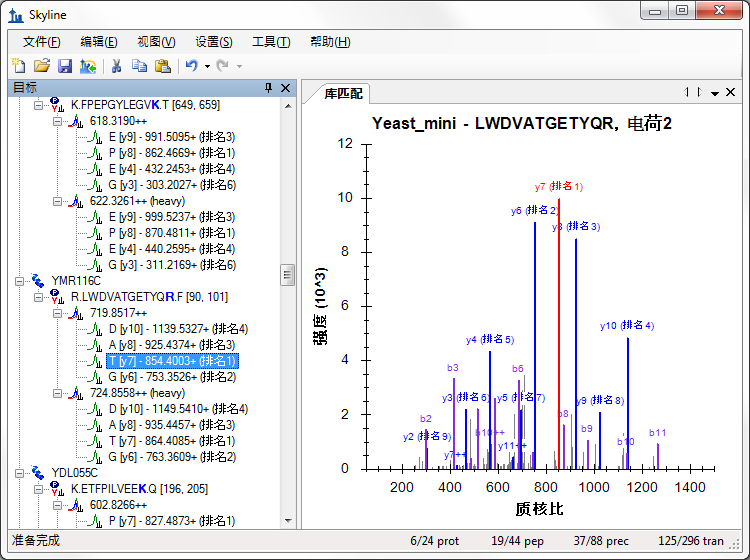


您还可以在窗口的右下角状态栏中看到指示符“296 tran”，表明MRMer离子对列表中的所有296 个离子对已添加到此文档中。在状态栏的左侧，您可以看到此文档包含24个蛋白质、44个肽段和88个母离子，其中每个肽段对应2个母离子。

如要进一步查看在此文档中的母离子和离子对，请执行以下操作：

* 在“编辑”菜单上，选择“全部展开”并单击“母离子”。

您可以花一些时间，选择肽视图中的各个肽和离子对，以浏览这些离子对，以及它们的碎片离子峰 在MS/MS库谱图中是如何按照强度大小排名。在肽段树形视图选择不同的肽段，则MS/MS谱图也会同步更新，显示出与当前肽段匹配的谱图。同时，与已选离子对相匹配的谱峰还会以红色突出显示：



您会注意到，并非所有的肽段都选择了丰度最高的碎片离子作为离子对，正如上图显示的那样。同时，并非所有的谱图都有良好的碎片离子匹配。

如未看到 b-离子或者突出显示的双电荷离子，您可以通过进行以下菜单选择使 Skyline 将其显示出来：

* 在“视图”菜单上，选择“离子类型”并单击“B”。
* 在“视图”菜单上，选择“电荷”并单击“2”。

您可能已经注意到Skyline对于轻母离子和重母离子都显示相同的谱图。此谱图库可能仅包含轻母离子的谱图，然而，即使此库包含轻和重母离子的匹配谱图，Skyline也仅使用其中一个（默认是轻母离子的），以避免两个 MS/MS 谱图之间碎片离子出现信号强度排序的差异。如果此库中仅包含匹配肽段的重型、已标记形式的谱图，那么 Skyline 将使用该图谱对轻和重母离子的离子对排名。

# 导入数据

当然，从现有离子对列表创建Skyline文档，一个重要的原因是我们想利用Skyline查看和验证在三重四级杆 MS 仪器上采集的数据，因此需要Skyline能够导入已有的原始离子对列表。

如要将 MRMer公布的数据导入到您已创建好的文档中，请执行下列步骤：

* 在“文件”菜单上，单击“另存为”。
* 导航至先前创建的 ExistingQuant 文件夹下的 MRMer 子文件夹。
* 在“文件名”字段内输入“MRMer”。
* 单击“保存”按钮。
* 在“文件”菜单上，选择“导入”并单击“结果”。
* 单击“导入结果”表单中的“确定”按钮。
* 选择文件列表中的“silac\_1\_to\_4.mzXML”文件。是从Waters Quattro Premier原始文件（无法使用）转换成的 mzXML，是因为 MRMer 无法读取仪器原始数据文件格式。
* 单击“打开”按钮。

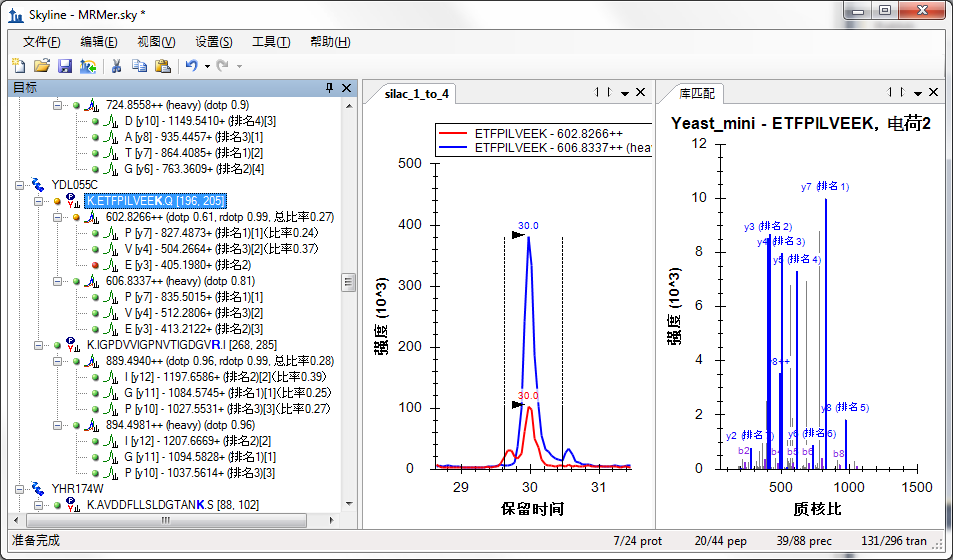
Skyline开始把文件导入其高速缓冲器中，这需要占用少量的磁盘空间，但是能够让Skyline 从中非常快地检索到所需信息。Skyline窗口底部的状态栏将显示进度。

导入完成后，在积分边界之间（以黑色虚线显示）测出信号的离子对，将会有绿色圆圈添加到离子对图标的左侧。那些在已选色谱峰组中不能找到的峰的离子对，会被显示出为红色圆圈。如果母离子仅含有绿色圆圈的离子对，此母离子和肽段还将显示出绿色圆圈。示例数据质量很好，所以您将主要看到绿色圆圈。

如要看到显示为红色圆圈的离子对，请执行以下步骤：

* 在“编辑”菜单上，单击“查找肽” (ctrl-F)。
* 在“序列片段”字段内输入“ETFP”。
* 单击“查找下一个”按钮。

此操作将使 Skyline 显示如下：

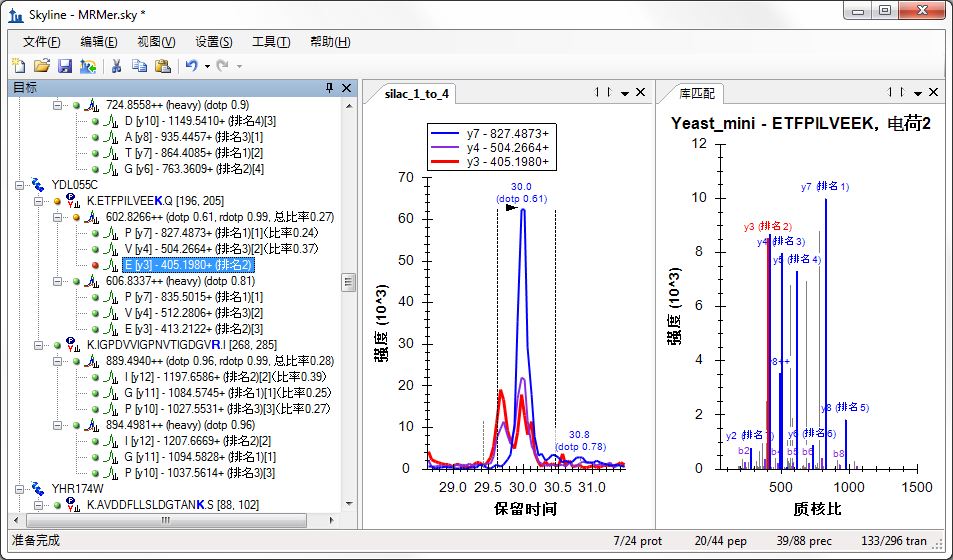


假如 Skyline窗口中图表窗口相互之间有重叠，您可以通过单击其中一个图表标签并拖动至新的位置上，重新安排这些窗口的布局。您还可以通过单击并拖动窗格之间的分割控制条来改变分配到各自的比例。

如果不想让色谱图视图像上图中那样放大显示，请执行以下步骤：

* 在“视图”菜单上，选择“自动缩放”并单击“最佳峰值”(F11)。

如要进一步了解此肽段的y3离子未包含在选取谱峰组之内的原因，请在肽段视图中选择该离子。Skyline 将发生如下变化：



如果未看到上图所示的三个离子对的色谱图，请执行以下步骤：

* 在“视图”菜单上，选择“离子对”并单击“全部” (shift-F10)。

现在您可能会看到y3和y4离子似乎存在一些信号干扰，以至于它们显示出两个未完全分离的谱峰。这些很显然不是由于单个肽段造成的，因为y7上并不存在重叠，而且这两个离子对之间的相对强度在两个峰中是不同的。

如果您仍然想优化该方法，则您可以下次尝试测定y5和y8，因为MS/MS库谱图表明这两个离子具有较好的可测性。然而，如果您确实想从这些数据中得到一个初始定量结果，则您有两种选择：

1. 删除受到干扰影响的峰值。
2. 调整峰值边界，以排除产生干扰的区域。

在 MacCoss实验室，我们倾向于第一种选择，是因为第二种选择需要让人们来确定未知干扰的信号界限，这种不确定因素会引入未知的误差。在本教程中，您可以尝试这两种选择。

## 移除存在干扰的离子对峰

如要从轻峰值组中除去y4离子对，请执行以下步骤：

* 右键单击色谱图。
* 在右键菜单上，选择“删除峰值”并单击“y4 - 504.2664+”。

在肽段视图中，您会看到文本“（比率为 0.38）”从轻离子对的末端消失，同时母离子比值“（总比率为 0.27）”变为“（总比率为 0.24）”。

母离子总面积比值是加权平均值，以内标峰面积作为权重，数学公式表示为：

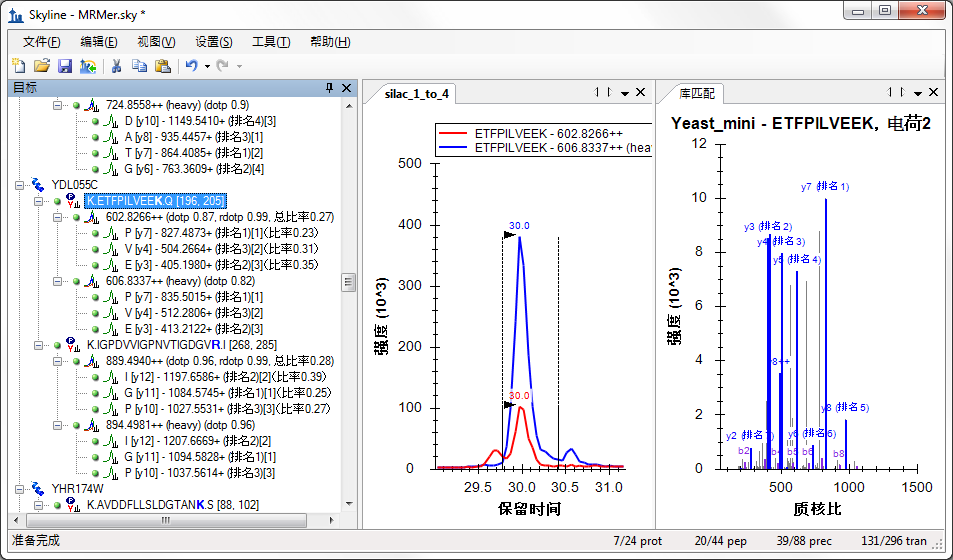
移除轻峰或重峰将移除与其相匹配的离子对的色谱峰。因此，仅有一对离子对保留时，正如本研究中的y7，离子对轻峰和重峰之间的比值（此处为 0.24）将成为母离子的总面积比值。

## 调整峰边界以排除干扰

如要通过调整峰的积分边界，以消除定量检测中的干扰，请执行下列步骤：

* 在“编辑”菜单上，单击“撤销” (ctrl-Z)，以撤销最后删除部分的峰值。
* 选择肽段“ETFPILVEEK”。
* 在 X-轴下面单击鼠标，从轻（红色）色谱图在预期峰与干扰峰（约 29.8 分钟）之间最小值的右侧，拖动至重（蓝色）色谱图右边的第二个峰的边界 （约 30.4 分钟）。

此操作将使 Skyline 显示如下：



此时，y4和y3之间的比值显示为 0.31，y7的显示为 0.23，加权平均产生的母离子比值为 0.26。

采用两种不同处理干扰的方法，获取的两个总比值分别为 0.24 和 0.26，与 1 比 4（轻和重的比例）SILAC 混合的真实值 0.25 同样接近。然而，显示存在干扰作用（EFP 肽段的y3和y4）的峰面积比值更接近于1:3，而不是1:4。由此看出，人为调节实际上操作（第一种选择）的可能出现问题，也强调了我们更倾向于移除干扰离子对（第二种选择）的原因。

对本文档中的数据的进一步分析将显示大多数肽段的比值非常接近于预期值0.25。此外，对于具有4个或更多个离子对的母离子，Skyline将显示峰面积和匹配MS/MS峰强度之间的点积 (dotp)。精确匹配时这些数值的大部分都非常接近于 1.0。

最后，在转向本教程的第二个文档之前，您可能会注意到2个肽段（K.YVDPNVLPETESLALVID**R**.L和K.FPEPGYLEGV**K**.T）的离子对为空白，而其它都具有绿色或红色圆圈。这就意味着之前导入的mzXML文件未包含这些离子对的数据。您可以在文本编辑器中打开 mzXML文件，查找母离子质荷比，以亲自确认母离子在原始离子对列表中缺失。对于处理手动创建的离子对列表，或由那些比Skyline接受的测试与使用较少的工具创建出的离子对列表，这些异常现象是非常典型的。

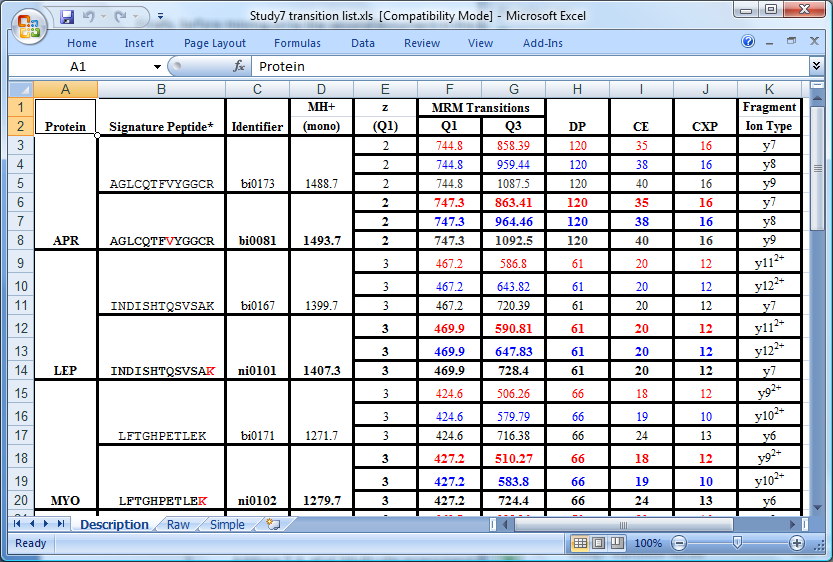
# 为 CPTAC Study 7 准备文档

在下一个章节中，您将处理跨实验室CPTAC Study 72的数据。该研究在Skyline 0.1 版首次发行之前已经由验证工作组完成。当时的实验方法是利用电子表格创建的，而产生的数据则由供应商提供的软件进行分析。

准备接受Study 7 离子对列表的文档

同样，首先要完成的第一项任务是从现有的离子对列表中创建Skyline文档。向Skyline中插入离子对列表的第一步是检查离子对列表，了解 Skyline 要求完成哪些设置才能识别出列表中的质荷比。如要开始这项检查，使用 Microsoft Excel 打开之前在本教程中创建的 ExistingQuant 文件夹下Study 7 子文件夹中的“Study7 transition list.xls”文件。

您将看到如下所示的电子表格：



您可以判定这是一个由某人手动创建的电子表格。创建时此人尽了很大努力，使用表格边界、合并单元格和突出显示，让其他人也能够清晰明了地查看这些离子对，而现在Skyline可以自动处理这项工作。

这个列表中的每一肽段均具有轻、重两种形式。每一个重肽的“Signature peptide”（署名肽／代表蛋白质的特定肽段）列中，单个稳定同位素标记的氨基酸残基以红色突出显示。向下滚动整个列表，您将看到有4种标记方案：

1. 6 个肽段具有标记的 C-末端赖氨酸，增加了 8 Da。
2. 2 个肽段具有标记的 C-末端精氨酸，增加了 6 Da。
3. 2 个肽段具有标记的内部缬氨酸，增加了 5 Da。
4. 1 个肽段具有标记的内部亮氨酸，增加了 6 Da。

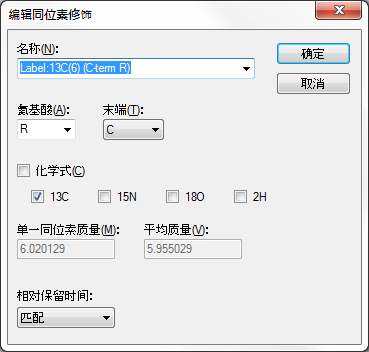
这个标记方案不可能像在MRMer文档中那样使用全局修饰，因为部分赖氨酸和精氨酸标记的肽段还包括内部缬氨酸和亮氨酸标记。处理这种现象最简单的方式就是对C-末端赖氨酸和精氨酸指定全局修饰，正如您在MRMer文档中做的那样，然后再手动应用缬氨酸和亮氨酸修饰。

如要为Study 7 离子对列表准备具有合适修饰的新文档，请执行下列步骤：

* 在 Skyline工具栏上，单击“保存”按钮，以保存对MRMer文档所做的改变。
* 在 Skyline工具栏上，单击最左边的“新文档”按钮。
* 在“设置”菜单上，单击“肽段设置”。
* 单击“修饰”标签（若有必要）。
* 单击“同位素修饰”列表旁边的“编辑列表”按钮。
* 选择“Label:13C(6)15N(4) (C-term R)”修饰，以便在下方添加将您的新修饰。
* 单击“编辑同位素修饰”表单中的“添加”按钮。
* 从“编辑同位素修饰”表单中的“名称”下拉列表中选择“Label:13C(6) (C-term R)”。

自动选中“13C”复选框，告诉 Skyline 对赖氨酸分子中的所有碳原子使用 13C，总质量偏移为 6 Da (6x 13C)。

此时“编辑同位素修饰”表单将显示如下：



如要完成Study 7 离子对列表文档准备，请执行以下步骤：

* 单击“编辑同位素修饰”表单中的“确定”按钮。
* 单击“编辑同位素修饰”表单中的“确定”按钮。
* 选中刚才创建的“Label:13C(6) (C-term R)”修饰。
* 取消选中为MRMer文档创建的“Label:13C(6)15N(4) (C-term R)”修饰，如果未取消选中的话。
* 确保仍然选中为 MRMer 文档创建的“Label:13C(6)15N(2) (C-term K)”修饰的复选框。
* 还要确保已选中“脲甲基半胱氨酸”的“结构修饰”列表中的复选框。
* 单击“库”标签。
* 取消选中在 MRMer 文档中使用的“Yeast\_mini”库，如果未取消选中的话。
* 单击“酶解”标签。
* 如果尚未选择“无”，则在“背景蛋白质组”中选择“无”。
* 在“肽段设置”表单中，单击“确定”按钮。

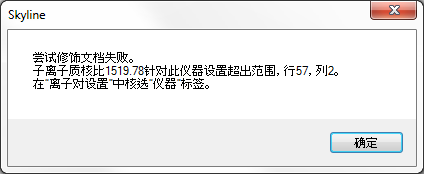
## 将离子对列表粘贴到文档

现在切换回到 Excel，并单击Study 7 电子表格中的“raw”（原始）标签。在本页面，您将看到66 个离子对已经从原始离子对列表中导入至本研究所采用的4000 QTRAP仪器中。此时，由于Skyline仍然缺少处理缬氨酸和亮氨酸标记的肽段中重离子对的设置，单击底部水平滚动条左侧的“Simple”（简单）电子表格标签，您将发现移除了9 个这种标记的离子对的表格。

如要把这些离子对添加到新的空白Skyline文档，请执行下列步骤：

* 选择Study 7 电子表格“Simple”页面上所有的6列57行数据。
* 复制单元格 (ctrl-C)。
* 切换回 Skyline。
* 在“编辑”菜单上，单击“粘贴” (ctrl-V)。

Skyline 将呈现以下错误消息：



在处理Skyline之外创建的离子对列表时，此类消息并不少见。在本主题的Skyline教学视频中还列出了其他消息（[视频 3：现有实验](https://skyline.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=video_0-5b)）。出现这种类型的错误最常见的原因有以下几个：

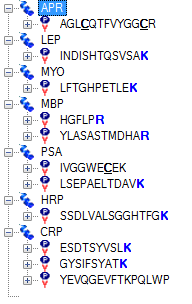
1. 忘记指定必要的修饰
2. 在手动或利用其他工具创建的质荷比数值时出现错误
3. 包含了超出当前仪器设置的范围的质荷比

最后一个可能的原因是由于当前的离子对列表与仪器类型不对应，您可以看到检查“离子对设置”表单中的“仪器”标签的消息。

如要纠正这一问题，请执行下列步骤：

* 在错误消息中单击“确定”按钮。
* 在“设置”菜单上，单击“离子对设置”。
* 单击“仪器”标签。
* 在“最大质荷比”字段中输入“1800”。
* 单击“确定”按钮。
* 在“编辑”菜单上，再次单击“粘贴” (ctrl-V)。
* 在“编辑”菜单上，选择“全部折叠”，并单击“肽” (ctrl-shift-D)。

此操作将使Skyline肽段视图显示如下：



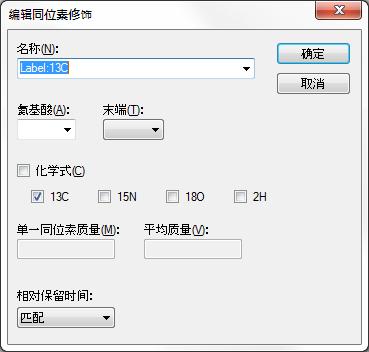
## 手动调整修饰

8个肽段显示加粗、蓝色的 C-末端 K 或 R，但其他 3 个缺少标记的氨基酸。这些都是需要手动指定 V 或 L 标记的肽段，因为通过文档设置无法实现这些修饰。

如要指定第一个肽段中V上的同位素修饰，请执行下列步骤：

* 选择“AGL**C**QTFVYGG**C**R”肽段。
* 在“编辑”菜单上，单击“修改肽段”。
* 从 C-末端 R 的“重同位素”下拉列表上选择空白条目，移除该精氨酸中稳定同位素标记。
* 在该肽段中 V 残基的“重同位素”下拉列表中选择“<添加…>”。
* 从“编辑同位素修饰”表单的“名称”下拉列表中选择“Label:13C”。

“编辑同位素修饰”界面如下：



通过修改同位素碳原子数量，这种修饰将能够在任意氨基酸上引入不同的质量偏移。单击“确定”按钮返回“编辑修饰”表单，表单现在显示如下：

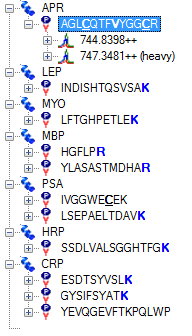


单击“编辑修饰”表单中的“确定”按钮，返回文档。

选中肽段中的缬氨酸将不会加粗，因为此肽段尚未包含重母离子。如要添加重母离子，请执行以下步骤：

* 将鼠标悬停在此肽段上，并单击出现在其右侧的箭头。
* 选中弹出选择列表中的“747.3481++（重）”复选框。
* 单击左上角的绿色选中按钮（或者按 Enter 键）。

肽段视图现在将如下所示：



单击轻母离子和重母离子左侧的“+”符号，展开“744.8398++”母离子和“747.3481++（重）”母离子，确认其包含匹配离子对，且如预期的那样它们之间相差 5 Da。

如要对剩下两个肽段创建标记的母离子，请执行下列步骤：

* 右键单击肽段“IVGGWE**C**EK”，并单击“修饰”。
* 从 C-末端 K 的“重同位素”下拉列表上选择空白条目，从该赖氨酸中删除稳定同位素标记。
* 在该肽段中 V 残基的“重同位素”下拉列表中选择“Label:13C”修饰。
* 单击“确定”按钮。
* 将鼠标悬停在此肽段上，并单击出现在其右侧的箭头。
* 选中弹出选择列表中的“541.7637++（重）”复选框。
* 单击左上角的绿色选中按钮（或者按 Enter 键）。
* 右键单击肽段“YEVQGEVFTKPQLWP”，然后单击“修饰”。
* 在该肽段中 L 残基的“重同位素”下拉列表中选择“Label:13C”修饰。
* 单击“确定”按钮。
* 将鼠标悬停在此肽段上，并单击出现在其右侧的箭头。
* 选中弹出选择列表中的“913.9746++（重）”复选框。
* 单击左上角的绿色选中按钮（或者按 Enter 键）。

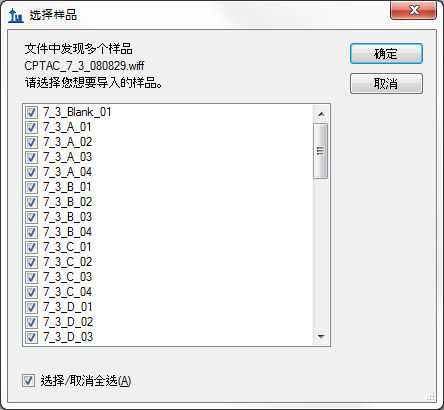
现在已创建好Skyline文档，准确地反映出原始Study 7 离子对列表中的信息。将此文档保存至 ExistingQuant 文件夹下的 Study 7 子文件夹中的 “Study 7.sky”。

## 从多样品 WIFF 文件导入数据

由于4个主要三重四级杆质谱仪厂商的支持，现在Skyline不需要任何格式转换，就能够完全支持导入所有仪器原始数据格式的文件。这意味着您可以直接导入4000 QTRAP 检测这些离子的原始数据，步骤如下：

* 在“文件”菜单上，选择“导入”并单击“结果”。
* 单击“导入结果”表单中的“确定”按钮，导入单次注射重复测定。
* 在Study 7 文件夹中选择“CPTAC\_7\_3\_080829.wiff”文件。
* 在“导入结果文件”表单中单击“打开”按钮。

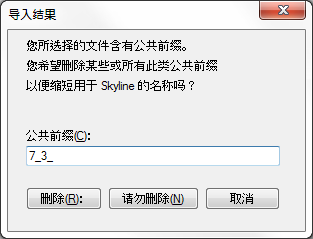
Skyline可能会花一两秒的时间读取多样品WIFF文件中的样品名称列表，然而它将呈现如下表单：



对于本教程，为了缩短导入所需的时间：

* 取消选中包含文本“空白”（1 个样品）、“QC”（4 个样品）和“梯度洗脱”（4 个样品）的条目。
* 同时还取消最后的“A2”（4 个样品）和“A3”（4 个样品）条目。
* 单击“确定”按钮。

Skyline将弹出移除重复前缀“7\_3\_”的对话框，这个重复前缀被Skyline认定和标示为重复实验：



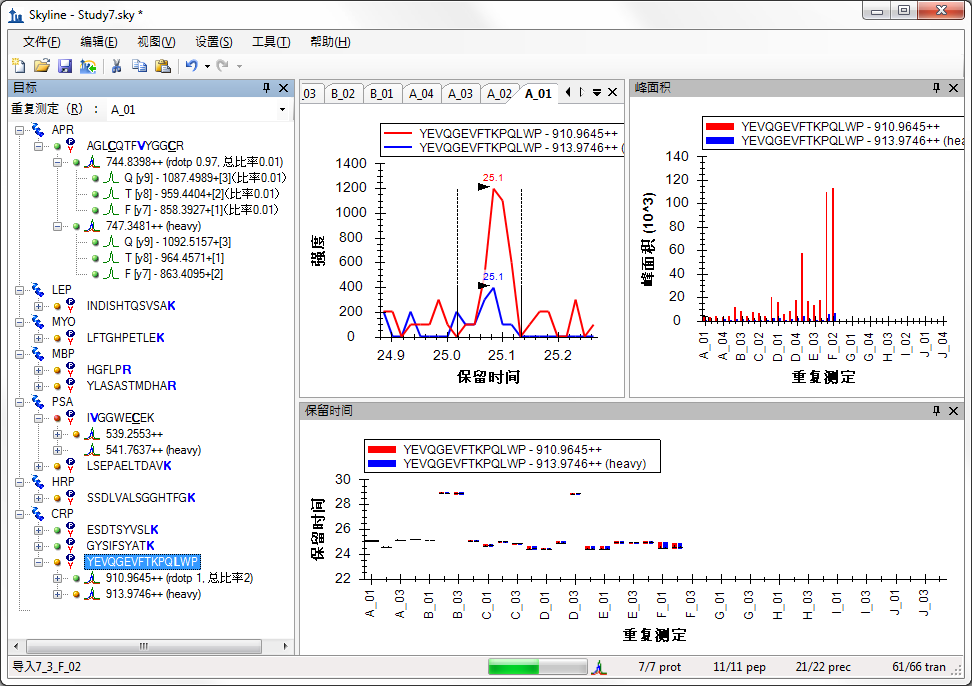
单击“删除”按钮，接受这项操作。

Skyline开始将此WIFF文件中的数据导入其高速数据缓存文件 (Study 7.skyd) 中，以便数据存取更快，并将文档中导入的所有数据压缩成单一文件以便共享。

## 检查和调整色谱峰积分

Skyline导入原始数据需要花费几分钟时间，在此期间您可以开始重新调整图表视图，以达到最佳效果。如要显示并重新定位峰面积和保留时间重复测定比较视图，请执行以下操作：

* 在“视图”菜单上，选择“保留时间”并单击“重复测定比较” (F8)。
* 单击并拖动窗口至 Skyline 窗口的底部，将其停放在那里。
* 在“视图”菜单上，选择“峰面积”并单击“重复测定比较”(F7)。
* 单击并拖动窗口至 Skyline 窗口的右边，将其停放在那里。
* 调整视图比例，使之如下图所示。



目前选择的肽段是“YEVQGEVFTKPQ**L**WP”（此为 CRP 蛋白质的C-末端肽段），事实上Study 7 中的验证工作组的前期分析可能有些问题。在导入数据时，您可以看到Skyline并不总能在重复检测中得到合理的色谱峰积分。仍然存在一些保留时间异常值，甚至有一些保留时间非常接近于25分钟的色谱峰好像并不是同一个肽段，因为大多数色谱峰峰值在 24.7 分钟附近。一个清晰的问题是，作为内标加入的重标肽段，总是比内源轻标肽段信号强度低不少。

需要说明的是，您可以将鼠标悬停在任何一个重复实验的图表上，直到在手形光标出现，然后单击导航至相应重复检测的色谱图。您可以使用这种方法导航至每一个可能错误识别的峰值所对应的色谱图，并利用在 MRMer 文档中使用的方法（即单击并向下拖动 X 轴）对其纠正。但是对于本教程而言，您可以采用此验证工作组的后期处理方式，即删除该肽段。

您将看到Skyline对此文档中其它肽段做了更好的色谱峰积分。比之前发行的Skyline版本，目前的 Skyline 0.7 版在更好的进行自动峰值积分方面得到了大大的改进。然而，在真实的数据集中，此文档还包含一些需要手动调整积分边界的离子对。

首先，在上图中，您可以看到肽段“I**V**GGWE**C**EK”的母离子“541.7637++（重）”似乎缺失数据。这是因为提供给质谱仪的离子对列表仅精确到母离子质荷比的一位小数点，将其错误地四舍五入为“541.7”。您可以在 Excel 电子表格中“Raw”标签上检查这项内容。

如要正确地计算此文档中的母离子，与测定的数据相匹配，请执行以下步骤：

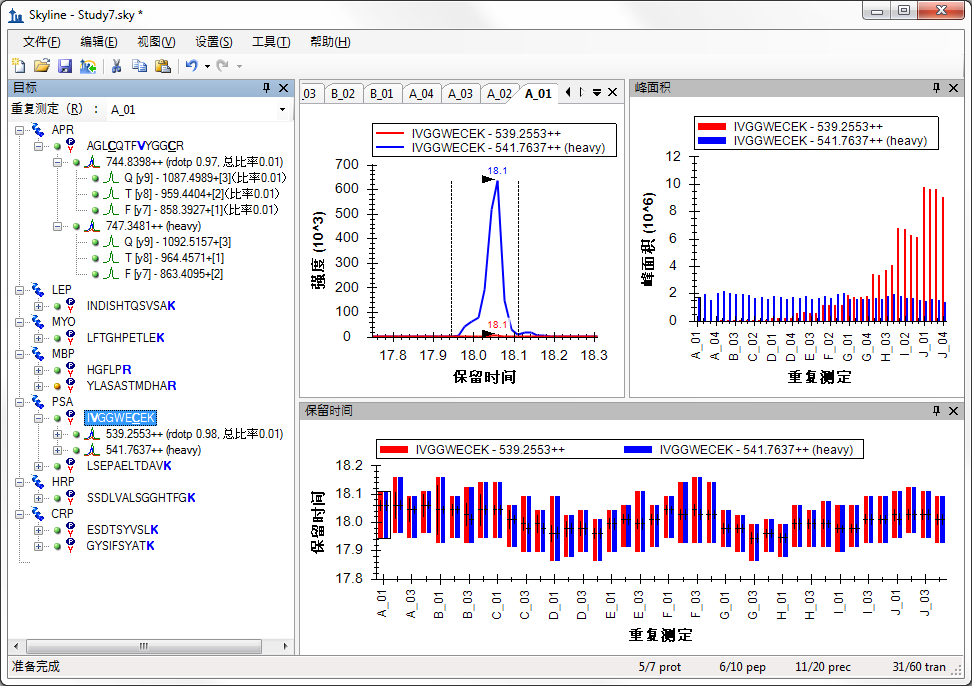
* 在“设置”菜单上，单击“离子对设置”。
* 在“仪器”标签的“质荷比匹配误差容限”字段内输入“0.065”
* 单击“确定”按钮。

这将导致在“541.7637++（重）”母离子旁边出现一个绿色圆圈。

另一个问题是，在肽段视图中有许多元素显示出橙色和红色圆圈，这表明离子对没有积分面积。正如在“靶向方法编辑”教程中讨论的那样，在优化方法时，标出这些肽段，然后进行优化是非常有用的。但是对于像本教程中已经高度优化方法，这将产生不太精确的定量结果。因此，Skyline引入了一个新的选项，即要求所有离子对的峰面积在最强峰值的界限范围之间进行整合。如要打开这项功能，请执行以下操作：

* 在“设置”菜单上，单击“全部合并”。

此时 Skyline 将显示如下：



## 用峰面积视图检查数据

上述图像中的“峰面积”视图给出了Study 7清晰的全局图景。它是四次技术重复检测各个浓度的校正曲线。您会看到，重同位素标记的内标是恒定浓度的，但经过 50 多次进样后，其峰面积有所减小。在“保留时间”视图，您可以看到肽保留时间非常稳定，非常适合应用于预定的运行中。

在本节中，您将专注于“峰面积”视图，以及其提供的许多用于检查多个重复检测数据集的选项。在屏幕上为“峰面积”视图提供更多的显示空间：

* 双击标记有“峰面积”的标题栏。

“峰面积”视图将脱离 Skyline 主窗口，并在其上浮动。重新调整其位置，以免遮挡肽段视图。

开始逐个选择此文档中剩下的 10 个肽段，您将看到，第1个肽段和最后 5 个肽段将会出现与上述图表类似的峰面积图表，而中间的 4 个肽段显得很不精确。

同时进样稳定同位素标记参照肽段的一个重要原因是，它们可用于对内源性未标记肽段的峰面积进行归一化，消除一些不同测定之间的变异（系统误差）。如要在“峰面积”视图中形象地看到这些，请执行以下操作：

* 右键单击“峰面积”图表，选择“归一化”并单击“重”。

再次检查之前表现正常的 6 个肽段。选择肽段 SSDLVALSGGHTFG**K**，您将看到如下所示的归一化图表：



确实，您将发现每一种浓度下的重复检测所获得结果之间的精确度得到了提高。

再看下其他 4 个肽段，您将发现它们仍然未显示出预期的模式。

查看归一化图表的另一种有趣的方式是分别查看每一个离子对。在行为正常的肽段中，每一个轻离子和对应的重离子对之间的比值较为相似。如要分别复查离子对的比值，请执行以下操作：

* 在“编辑”菜单上，选择“全部展开”，并单击“肽”(ctrl-D)。
* 选择每一个肽段的轻母离子。

对于 6 个行为正常的肽段，您会看到如下所示的图表（ESDTSYVSL**K** 肽段轻母离子 564.7746++ 被选中）：



不出所料，这些比值非常相似。第二个和第三个肽段 (INDISHTQSVSA**K**和LFTGHPETLE**K**)的图表并不是非常干净，但离子对比值中没什么问题。然而，对于第四个肽段 (HGFLP**R**)，选择其轻 363.7059++ 母离子时，会出现如下所示的峰面积图表：



看起来 y3 的离子对中存在干扰，因为其低浓度下比值与其它肽段差异较大。

“峰面积”视图提供了另一种检查母离子中离子对相对强度的方式。现在您可以通过执行以下操作来再次查看 HGFLP**R** 肽段：

* 右键单击“峰面积”图表，选择“归一化”，并单击“总数”。

图表将发生如下变化：



此图表再次清晰地显示出 y3（棕色）的离子对存在干扰，随着内源性肽的浓度增加（超出 E组重复测定时的浓度），其影响变小。移动峰面积视图，以便看到色谱图，您这时可以单击各个条以查看存在干扰的峰值。本例中可以非常清楚的看到y3的色谱峰存在干扰（重复检测E\_03）：



与 MRMer 文档中的情况一样，您可能为了排除干扰峰而尽力调整积分界线，但删除 y3 离子对很可能是一种更好的办法，尤其是因为该离子对对肽段的总的峰面积贡献不大。如要深入了解这些测定结果的精确性，请执行以下操作：

* 在“视图”菜单上，选择“峰面积”，并单击“肽段比较” (ctrl-F7)。
* 在“视图”菜单上，选择“离子对”，并单击“总数” (ctrl-F10)。
* 右键单击“峰面积”图表，并单击“CV 值”。

“峰值面积”图表将发生变化：



如果该肽段未按照这个顺序出现，请执行以下操作：

* 右键单击“峰面积”图表，选择“优化”并单击“文档”。

当然，当前显示的肽段（以红色显示）有很大的变异系数 (CV) ，因为数据集包含了被分析蛋白质及其肽段的 10 个不同浓度水平的数据。然而，重肽（以蓝色显示）的 CV 值提供了大量的信息，因为在所有样品中其是以恒定浓度进样的。基于目前为止您所看到的，对于6个结果清晰的肽段来说，变异系数约为 10% 或更低，这应该说十分正常。然而其他 4 个肽段的变异系数更接近于 40%，甚至50%。

在Skyline中仅使用几个简单的操作，您就可以了解到一个数据集的很多信息。而在此前，这些信息需要验证工作组的统计和程序人员花数周时间分析和提取。如果您已经完成了[靶向方法优化](https://brendanx-uw1.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=tutorial_method_refine)教程，现在您可能明白了为什么在分析预定重复实验数据时要进行优化环节，这样做是为了在分析珍贵样品之前能更好地理解目标肽段的信号响应特性。

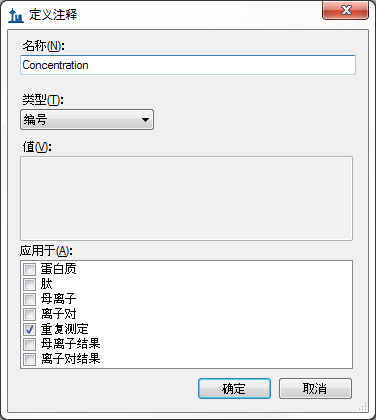
## 用浓度值注释重复检测

如果 X-轴显示的是样品浓度，而不是重复检测的名称，则将能更加直观的理解图表数据。

Skyline允许您通过定义“注释”的方式，关联重复检测的其他信息。

* 在“设置”菜单上，单击“注释…”。
* 单击“注释设置”表单中的“编辑列表”按钮。
* 单击“定义注释”表单中的“添加”按钮。
* 在“定义注释”表单中的“名称”字段内输入“Concentration”（浓度）。
* 在“类型”字段内选择“编号”。
* 选中“应用于”列表中“重复测定”左侧的复选框。

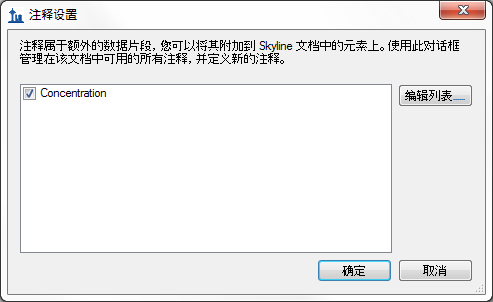
“定义注释”表单现在将如下所示：



现在完成了注释的定义，可以通过以下步骤将其添加到您的文档中：

* 单击“定义注释”表单中的“确定”按钮。
* 单击“定义注释”表单中的“确定”按钮。
* 选中新的“Concentration”注释的复选框。

“注释设置”表单现在将显示如下：



* 单击“确定”按钮。

可使用“结果网格”编辑这些注释的数值。如要显示结果网格视图，请执行以下操作：

* 在“视图”菜单上，单击“结果网格” (Alt-2)。

“结果网格”将显示峰面积以及其他测定的值。为了确保“Concentration”列可见，请执行以下操作：

* 在肽树形视图中选择蛋白质“APR”。

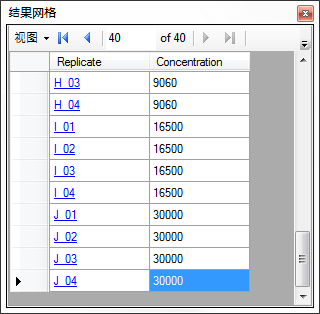
现在“结果网格”将仅显示出重复检测名称和浓度注释。配制样品的浓度为：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样品 | A | B | C | D | E | F | G | H | I | J |
| 浓度 (fmol/µL) | 0 | 60 | 175 | 513 | 1500 | 2760 | 4980 | 9060 | 16500 | 30000 |

利用上述表格，填写各个重复检测的“浓度”值。如果您在向“结果网格”中输入内容时看到 Skyline 主窗口在闪动，您可能需要通过完成以下操作，以允许 Skyline 在不激活重复检测色谱图的情况下改变网格中的行：

* 右键单击“结果网格”，并单击“同步选择”。

在你正在输入这些数值时，“结果网格”的底部将如下所示：



其中一个可以使用这个新注释的地方是“峰面积重复测定比较”图表。

* 在肽树形视图中选择肽段“SSDLVALSGGHTFGK”。
* 在“视图”菜单上，选择“峰面积”并单击“重复测定比较”(F7)。
* 右键单击“峰面积”图表，选择“分组依据”并单击“浓度”。
* 右键单击“峰面积”图表，选择“归一化”并单击“重”。

“峰面积”图表将显示如下：



该图显示出直到最低浓度CV值都非常小。通过执行以下操作，您可以轻易切换到平均比值和标准偏差图表：

* 右键单击“峰面积”图表，单击“CV 值”。

视图将发生如下变化：



## 进一步探索

8个不同实验室参与了Study 7，并且每一个实验室均产生了与本教程中您检查的相类似的数据集。然而，并不是所有的实验室都以相同的方式遇到了相同的问题。本教程还包含此研究中来自另一实验室及其他子项目的数据。本教程中，文档已创建好，并且数据集已导入。

如要打开Study 7-II 来自编号为52的实验室的数据，请执行下列步骤：

* 在“文件”菜单，单击“打开” (ctrl-O)。
* 导航至本教程开始时创建的 ExistingQuant 文件夹下的 Study 7 子文件夹中的 Study II 子文件夹。
* 选择“Study 7ii (site 52).sky”文件。
* 单击“打开”按钮。

此文件会快速打开，并且您会看到重同位素标记参考肽段的CV值不尽相同，如下图所示：



INDISHTQSVSA**K**是来自第一组数据的4个肽段中的一个，此数据集的变异系数仍然约为40%，而现在LSEPAELTDAV**K** 的变异系数约为 25%。然而，如果您在“保留时间”视图中看到“重复检测比较”图表，您将会看到 Skyline 从其中 3 次重复检测中选择了错误的色谱峰（这很可能会出现在将来的skyline版本得到完善）。



如果您通过单击并向右拖动 X-轴的方式纠正这些错误，LSEPAELTDAV**K** 的重母离子变异系数可以下降至类似其它大多数肽段的10% 左右。

如要查看此数据集中的轻:重比值，请执行下列步骤：

* 在“视图”菜单上，选择“峰面积”并单击“重复检测比较”(F7)。
* 右键单击“峰面积”图表，选择“归一化”并单击“重”。

您将发现所有肽段的图表，甚至会发现肽段INDISHTQSVSA**K** 在此数据集中结果非常好：



这似乎显示出内标对此肽段检测结果的变异起到了补偿作用。查看归一化前的数据：

* 右键单击“肽面积”图表，选择“归一化”并单击“无”。

如果不是这张图表，我们很难想象到归一化会如此有效：



最后，回到我们在第一组数据集中发现的存在干扰的肽段上来：

* 选择肽段 HGFLP**R** 的轻母离子 363.7059++。
* 右键单击“峰面积”图表，选择“离子对”并单击“全部”。
* 右键单击“峰面积”图表，选择“归一化”并单击“重”。

在这里您会看到 y3 离子对中存在干扰的明显证据：



然而，和第一组数据集不同，在这组数据里检查此肽段色谱图来发现干扰作用显得更困难。在这种情况下，由于稍微改变了色谱条件，几乎不可能在y3的积分范围中找到此肽段两个峰，如下面所示（重复检测E\_03）：



NCI CPTAC Study 7 提供了更多的数据，现在这些数据均已公开。Skyline可以帮助您迅速批量了解这些数据，即使参与此研究的人员并不知道 Skyline的存在。

# 结论

通过本教程的学习，您了解到了，对那些事先并没有针对使用skyline而设计和产生的实验数据， Skyline仍可以同样地大大简化数据分析处理的过程；这些数据可能来自于您开始使用 Skyline 之前自己进行的实验，抑或是来自于他人曾经进行的而您希望重复或回顾的实验。对于MRMer和NCI CPTAC Study 7 公布的数据集，您很快从离子对列表创建好了Skyline文档，即使这些列表包含着相当复杂的修饰情况，还包括同位素标记内标。

通过学习您还了解到了一些 Skyline 的特征，即 Skyline 能够处理引入了同位素标记的靶向蛋白质组实验数据。从同位素标记修饰的定义到设计，Skyline简化了为这些实验创建仪器方法的工作。从精确的轻:重峰面积比值到强大的图表显示选项，Skyline 让您可以深入地了解并分析这些实验采集的数据。

# 参考文献列表

1. Martin,D.B. *et al.*MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments.*Mol.Cell Proteomics.***7**, 2270-2278 (2008).

2. Addona,T.A. *et al.*Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma.*Nat.Biotechnol.***27**, 633-641 (2009).