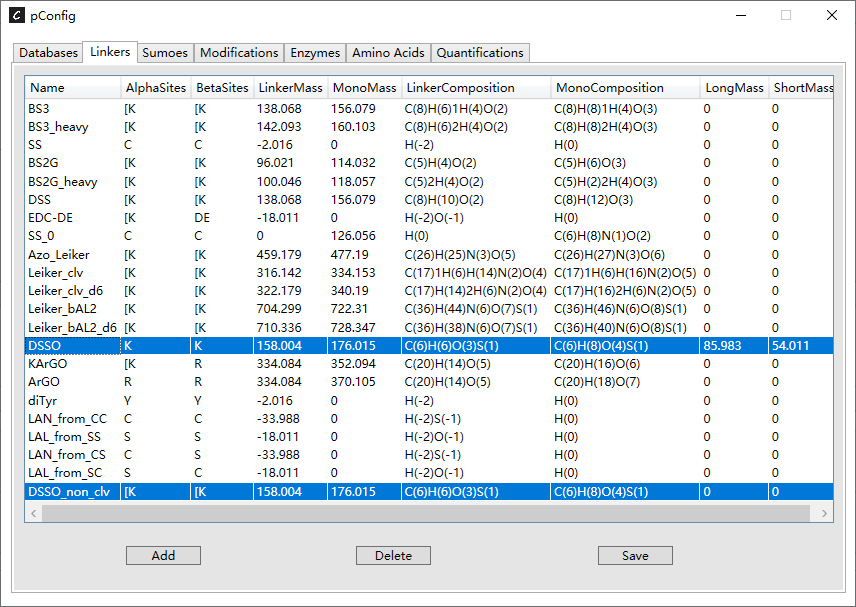
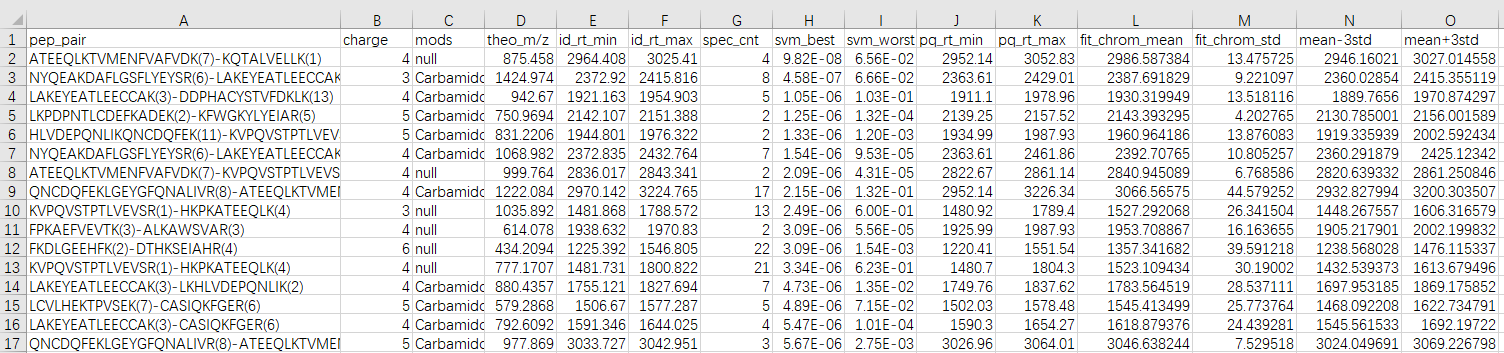
构建肽段inclusion list，在优化SteppedHCD能量时，仅对inclusion list中的肽段打MS2。

1. 使用pLink2界面搜索，设置DSSO\_non\_clv，然后使用Conventional crosslinking流程DSSO\_non\_clv交联剂搜索DSSO数据，即用不可碎裂流程搜索DSSO数据。



1. 搜索完毕之后，使用打开自动生成的pQuant参数文件pQuant\_cfg.txt，修改参数NUMBER\_SCANS\_HALF\_CMTG=1000;，即pQuant在重构色谱曲线时往前后查看的MS1数目。
2. cmd进入pLink2安装文件夹，使用命令行运行pQuant.exe pQuant\_cfg.txt，在pLink2结果文件夹中会新增一个pQuant文件夹，里面存放了pQuant定量结果。
3. 打开filter.py文件，设置pLink2\_id\_folder、inclusion\_list和inclusion\_list\_detail的路径，分别表示鉴定结果文件夹路径、期望存放的inclusion\_list路径和期望存放的inclusion\_list\_detail路径
4. 运行filter.py
5. 打开在第2步设置的inclusion\_list文件，对每个交联肽段，按（序列，电荷，修饰）汇总为一行，每一列的含义如下：
   1. 交联肽段对
   2. 电荷
   3. 修饰
   4. 理论m/z
   5. 所有鉴定到这个肽段的ms2 scan的rt的最小值
   6. 所有鉴定到这个肽段的ms2 scan的rt的最大值
   7. 鉴定到这个肽段的谱图数目
   8. 鉴定到这个肽段的所有谱图的svm打分最好值
   9. 鉴定到这个肽段的所有谱图的svm打分最差值
   10. pQuant会对每个PSM输出其母离子在ms1中的最小rt和最大rt，这一列表示所有鉴定到这个肽段的PSM的最小rt的最小值
   11. 和第J列类似，表示所有鉴定到这个肽段的PSM的最大rt的最大值
   12. pQuant在重构色谱曲线时，可能会陷入局部极小值，我把pQuant输出的该肽段的所有色谱曲线拼接起来，对全局色谱曲线进行拟合，这是高斯拟合的均值
   13. 和第L列类似，是高斯拟合的标准差
   14. 根据高斯函数的3σ准则，这可以认为是rt的最小值
   15. 和第N列类似，这可以认为是rt的最大值。

可以用Excel对spec\_cnt排序，把比较可靠的肽段信息输入质谱仪的inclusion list。



inclusion\_list\_detail.csv在inclusion\_list.csv的基础上，对每个肽段，额外输出了三行信息，分别是ms1 scans、ms1\_rt、ms1\_mono\_intensity，表示重构全局色谱曲线时，拼接起来的ms1 scan、rt和mono峰强度。举个例子，比如inclusion\_list\_detail.csv的第一个肽段（如下图所示），第6行表示出现了该肽段母离子mono峰的所有ms1 scan列表，第7行对应第6行ms1 scan的rt，第8行对应mono峰的强度。如果用excel画出第7、8行，则可以看到重构的色谱曲线。高斯拟合出来的rt均值为2986.587384，标准差为13.475725，根据3σ准则，可以认为这个肽段的洗脱时间区间在[2946.16021，3027.014558]。对比由鉴定结果导出的rt区间（第E、F列）和由pQuant直接输出的rt区间（第J、K列），这三个好像差不多，个人觉得用（第N、O列）和（第J、K列）应该都可以。

另外，需要注意的是，由下图可知，重构出来的色谱曲线可以看到有两个高斯曲线，这是因为数据来源2个raw，可能rt没有对齐，建议只对单个raw跑这个脚本。另外，我手工看过几个重构出来的色谱曲线，并不是所有色谱曲线的高斯曲线都像下图那样漂亮，所以重构出来的rt（第N、O列）不一定很准，建议直接用pQuant输出的rt区间（第J、K列）。

