

蛋白质结构与相互作用研究新方法 ——交联质谱技术*

樊盛博^{1,2)} 吴妍洁^{1,2)} 杨兵³⁾ 迟浩^{1,2)} 孟佳明^{1,2)}
 卢珊³⁾ 张昆^{1,2)} 邬龙^{1,2)} 孙瑞祥¹⁾ 董梦秋^{3)**} 贺思敏^{1)**}

¹⁾中国科学院计算技术研究所, 中国科学院智能信息处理重点实验室, 北京 100190;

²⁾中国科学院大学, 北京 100049; ³⁾北京生命科学研究所, 北京 102206)

摘要 蛋白质的空间结构信息以及蛋白质间的相互作用信息对于研究蛋白质的功能有重要意义。研究蛋白质结构与相互作用的传统技术, 如核磁共振技术、X射线晶体衍射技术等, 对于蛋白质的纯度、结晶性和绝对量均有比较高的要求, 限制了其广泛应用。交联质谱技术是近十多年来发展起来的新技术, 它将质谱技术与交联技术相结合, 在研究蛋白质结构与相互作用方面具有速度快、成本小、蛋白质各方面性状要求低等优势。本文就交联质谱技术各个环节的技术方法加以综述, 包括交联质谱实验分离富集技术、常见交联剂特性、交联质谱数据库搜索算法、结果验证研究和交联质谱技术的应用等方面, 并展望了该研究方向未来的发展。

关键词 交联质谱技术, 蛋白质结构, 蛋白质相互作用, 蛋白质组学, 质谱技术

学科分类号 Q51, TP39

DOI: 10.3724/SP.J.1206.2013.00484

蛋白质以其自身的结构和与其他蛋白质之间的相互作用为基础发挥功能, 因此, 研究蛋白质的结构和相互作用一直是生命科学的重要方向^[1-2]。

蛋白质结构研究的传统技术, 如核磁共振技术(nuclear magnetic resonance, NMR)、X射线晶体衍射(X-ray crystallography)技术, 有精度高、获得信息多等优点, 但对蛋白质样品本身有比较苛刻的要求, 如蛋白质的绝对量、纯度、浓度、分子大小和可结晶性等都必须达到一定的标准, 才能利用这类技术研究其结构。同时, 这类方法往往耗时较长、成本较高, 进一步限制了它们的应用^[3-5]。

检测蛋白质相互作用的传统方法, 如酵母双杂交(yeast two-hybrid)、亲和色谱(affinity chromatography)和免疫共沉淀(co-immunoprecipitation)等, 也有各自的局限性。酵母双杂交技术可以揭示蛋白质间的直接相互作用, 甚至通过大规模筛选发现未知的相互作用, 但酵母细胞未必能为异源表达的其他物种蛋白质提供合适的相互作用的环境条件。亲和色谱技术和免疫共沉淀技术其通量比较低, 背景结合蛋白质与特异性结合蛋白质有时难以区分, 直接与间

接相互作用通常也难以区分。另外, 这三种方法对于瞬时、微弱的相互作用, 比如信号转导过程中的松散变化的蛋白质复合物, 都很难获得有效信息。最后, 这类技术还有一个共同缺点是均无法提供相互作用的界面信息^[6-8]。

化学交联结合质谱技术(chemical cross-linking coupled with mass spectrometry), 简称交联质谱技术或CXMS, 是近年发展起来的新方法。它利用化学交联剂(chemical cross-linker)处理蛋白质样品, 将空间距离足够接近、可以与交联剂反应的两个氨基酸以共价键连接起来, 然后利用基于质谱技术的蛋白质组学手段分析交联产物。交联质谱技术在研究蛋白质结构和蛋白质相互作用上有着自己独特的优势: a. 可以高通量地进行多种蛋白质的结构和

* 国家自然科学基金面上项目(21475141)和国家重点基础研究发展计划(973)资助项目(2010CB912701, 2012CB910602, 2013CB911203)。

** 通讯联系人。Tel: 010-62600822

董梦秋。E-mail: dongmengqiu@nibs.ac.cn

贺思敏。E-mail: smhe@ict.ac.cn

收稿日期: 2014-02-10, 接受日期: 2014-05-26

相互作用分析, 比传统方法速度快、信息产出规模大; b. 通过交联剂的共价交联作用, 可以固定原本不稳定的蛋白质相互作用, 从而可以研究一大类容易解离的、结构松散的蛋白质复合物; c. 相对于核磁共振、X 射线晶体衍射, 交联质谱技术灵敏度高、对蛋白质性状要求低、需要样品绝对量小; d. 可以进行体内交联, 有助于研究体内蛋白质结构和相互作用^[9-11]; e. 由于交联后的蛋白质先经过酶切之后才进行质谱分析, 所以不受蛋白质本身的长度影响; f. 使用不同臂长、不同反应基团的交联剂可以得到更多的蛋白质空间结构和相互作用信息^[12].

将化学交联与质谱技术结合起来分析蛋白质结构的尝试始于 20 世纪末、21 世纪初^[13], 到 2012 年, 有关 CXMS 的研究文章发表数目已经从 2000 年的 124 篇稳步增长到 260 篇, 引用数目从 2000 年的 2 057 篇增加到 8 864 篇 (数据来自于 Web of Knowledge, <http://apps.webofknowledge.com/>, 2013-09-01). 交联质谱技术从实验技术环节到数据分析环节都有长足的进步, 获得了学术界的广泛关注, 已成为持续增长的新的科研热点^[14-21].

本文将按照交联质谱技术各环节的先后顺序进行综述. 首先介绍因酶切和交联情况的不同而产生的三种交联肽段形式, 以方便下文中理解与使用; 然后依次介绍交联质谱分离富集技术、常见交联剂特性、数据库搜索算法、结果验证研究和技术应用; 最后对该研究方向未来的发展进行了展望.

1 交联产生的肽段形式

交联蛋白质在不同位置发生酶切, 可以产生三种交联产物 (图 1): a. 被交联剂修饰的单肽 (交联

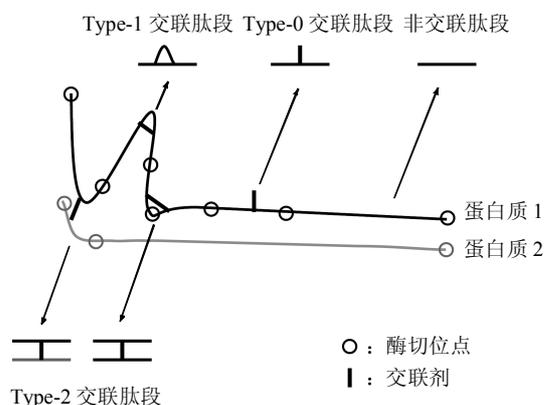


Fig. 1 Multiple cross-linked peptide forms

图 1 交联蛋白质酶切后产生的交联肽段的各种形式

剂只一端交联一条肽段, mono-linked peptide), Type-0 类型的交联肽段; b. 肽段内部交联 (交联剂两端交联同一条肽段, loop-linked peptide 或 intra-linked peptide), Type-1 类型的交联肽段; c. 肽段间交联 (交联剂交联两条肽段, Inter-linked peptides), Type-2 类型的交联肽段.

三种类型的交联肽段中, Type-0 交联肽段实际上是被交联试剂修饰的普通肽段. 这三种类型的交联肽段都是单个交联剂分子的反应产物. 也有更为复杂的共价连接两个或更多交联剂分子的交联肽段, 比如 Type-0 和 Type-1 并存的 Type-0, 1 形式的交联肽段等, 但是由于这些形式不常出现, 通常不予考虑^[22].

2 实验分离与富集

交联质谱技术, 相较于常规质谱技术, 由于目标交联肽段在整个样品中相对含量较少、样品成分更加复杂, 一定的分离与富集处理可以有效地提高最终鉴定效果. 在这方面, 多年来已做出许多探索工作, 本节仅对一些主要方法加以简述.

交联样品的分离可以在蛋白质水平或肽段水平进行. 前者最常用的是聚丙烯酰胺凝胶电泳方法 (sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE); 后者一般使用反相高效液相色谱技术 (reverse phase high performance liquid chromatography, RP-HPLC). 与之配合的, 常见的肽段水平预分离方法包括亲和色谱技术、强阳离子交换技术 (strong cation exchange, SCX)^[23]和分子筛色谱技术 (size exclusion chromatography, SEC)^[24]等.

基于 SDS-PAGE 的策略, 通常做法是切下交联蛋白质的条带, 酶切后再用基质辅助激光解吸电离与飞行时间质谱联用 (matrix-assisted laser desorption/ionization time of flight mass spectrometry, MALDI-TOF-MS) 或液相色谱与串联质谱联用 (liquid chromatography-tandem mass spectrometry, LC-MS/MS) 分析. 这种方法不能用于太过复杂的样品, 通常针对单个蛋白质或蛋白质复合物.

更加复杂的样品, 在进入质谱分析之前, 除了一些非常小的蛋白外, 需要先酶切, 再用 RP-HPLC 分离^[23]. 这种方法能够有效分离复杂样品中的交联肽段, 但不能对交联肽段进行富集.

要对交联肽段进行富集, 可以使用亲和色谱技术. 这种方法需要在交联剂上设计相应的亲和标示

物. 亲和色谱技术能够利用交联肽段中含有的亲和和标示物来富集含有交联剂的肽段. 不过这种方法不能区分 Type-0、Type-1 和 Type-2 三种交联形式, 因为它们均含有交联剂.

如果要进一步分离这三种交联形式的肽段, 可以尝试强阳离子交换色谱分离^[24], 利用 Type-0、Type-1 和 Type-2 三种交联形式在电荷量上的差别来进行分离. 但如果在 Type-0、Type-1 类型的交联肽段中含有组氨酸或漏切的胰酶位点 (精氨酸或未交联的赖氨酸) 时, 电荷会与交联肽段相近, 这种方法便无法分离. 此外, 由于交联肽段通常质量大于常规肽段, 所以也可以利用分子筛色谱技术^[25]将两者分开. 但如果常规肽段较长, 便无法与相近质量的交联肽段分离.

3 交联剂

交联质谱技术流程中, 使用何种交联剂, 交联剂如何设计, 对于结果和后续处理有很大的影响.

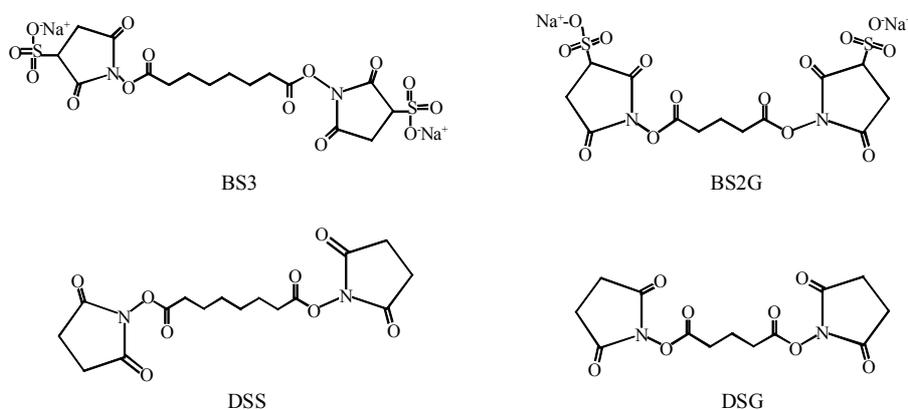


Fig. 2 Structure of BS3, BS2G, DSS, DSG

图 2 BS3, BS2G, DSS, DSG 的结构

琥珀酰亚胺酯类交联剂有以下优点. 首先, K 在肽段中含量较多, 使得交联更加容易. 其次, 反应效率高, 在控制好反应条件的前提下, 交联特异性也非常高, 很少与除 K 和肽段 N 端之外的其他氨基酸发生反应^[26-27]. 相比起来, 用非氨基特异性的交联剂^[28-29]得到的交联肽段数量较少.

除此之外, 部分交联剂可以进行非对称交联, 因其两端的化学基团性质不同, 可以与不同种类的氨基酸反应^[30-31]; 部分交联剂如甲醛只含有一个基团, 但也可以与多种氨基酸反应^[10-11, 32].

交联剂大致可分为简单交联剂和特色交联剂两大类. 本节给出了部分常见交联剂的特性.

3.1 简单交联剂

简单的交联剂通常包含两个反应基团和一个连接区域. 其中, 最常用的交联剂是与氨基反应的琥珀酰亚胺酯类 (N-hydroxy succinimidyl or sulfosuccinimidyl esters, NHS ester), 包括双琥珀酰亚胺辛二酸酯 (disuccinimidyl suberate, DSS)、双琥珀酰亚胺戊二酸酯 (disuccinimidyl glutarate, DSG)、二(磺基琥珀酰亚胺)辛二酸酯 (bissulfosuccinimidyl suberate, BS3) 和二(磺基琥珀酰亚胺)戊二酸酯 (bissulfosuccinimidyl glutarate, BS2G)(图 2). BS3 和 BS2G 相对 DSS 和 DSG 增加了助溶的基团, 而 BS3 与 BS2G 之间、DSS 与 DSG 之间的区别则在于交联剂的长度. 它们可与赖氨酸 (K) 侧链的氨基或者蛋白质 N 端的氨基发生交联反应.

3.2 特色交联剂

特色交联剂, 按功能可分为三类: a. 交联剂含有亲和基团, 进一步可以利用磁珠富集^[33-42]; b. 交联剂能够在二级质谱图中产生特殊碎裂信号离子^[34-41, 43-47]; c. 交联剂可以使用轻重同位素标记^[33, 39, 41, 48-49]. 目前主流的交联剂, 一般均含有其中的一种或几种功能(表 1).

3.2.1 具有亲和基团的交联剂

亲和基团 (affinity tag) 富集技术最早用于蛋白质的定量和修饰鉴定^[50]. 让交联剂带上特定的亲和

基团, 与固定相中磁珠相结合, 就可以富集含有交联剂的肽段, 而排除不含交联剂的肽段的干扰. 沿着这条线路, 目前已发展出多种带亲和基团的交联剂. 这种方法的局限在于, 交联剂需要特别设计合成, 并且由于加入特定基团的缘故, 一般臂长要比简单的交联剂更长, 这使得交联效率有所下降, 得到的交联位点之间的距离限制作用也会变弱.

3.2.2 具有特征离子的交联剂

这类交联剂的交联肽段产物在质谱碎裂产生的二级谱图中, 会产生特征性的谱峰. 其原理是在交联剂设计时加入了某种特殊的、易在质谱仪碎裂条件下脱落的基团, 这样交联肽段在碎裂生成二级谱图的同时也会发生基团的脱落, 形成谱峰. 这类交联剂的优点是, 能够利用二级谱图中信号谱峰的存在性来判断谱图是否含有交联剂, 从数据层面进行富集, 进而提升鉴定效果.

3.2.3 同位素标记的交联剂

一般在合成交联剂时, 可以选择加入同位素标记, 从而产生轻重两种交联剂^[51-54]. 使用时可以混合两种交联剂, 进行后续实验. 这种可标记交联剂的优点是存在轻重两种交联剂, 在谱图中会形成轻重同位素标记谱峰对, 进而可以利用这一点来判断含有交联剂的谱图, 二级谱图上的碎片离子也会有轻重关系的对应. 缺点是由于信号被分到轻重

两个通道, 相同量样品产生的谱图信号不及非同位素标记得到的信号强, 信噪比有所下降.

特色交联剂一般均会发生交联剂断裂. 按照断裂机理的不同, 特色交联剂又可分为化学断裂 (chemically cleavable)、气相/质谱断裂 (gas phase/mass spectrometry cleavable) 和光断裂 (photocleavable) 三类. 化学断裂是在送入质谱仪之前即用化学手段使交联剂断裂, 因而无法获得母离子和交联剂断裂形成的单肽之间的对应关系, 而且需要额外的清洗步骤, 可能造成样品的损失; 气相/质谱断裂克服了化学断裂的这些缺点, 它在样品进入气相/质谱仪器后才发生碎裂, 可以保持对应关系, 缺点是其断裂后形成的单肽往往电荷较低, 因而不适用于质谱鉴定; 光断裂则是对气相/质谱断裂的一种弥补手段, 它通过额外的紫外线或激光等来使交联剂发生断裂, 提高了单肽段所带的电荷, 使之更适合于质谱鉴定.

简单交联剂与特色交联剂这两类交联剂的界限并不明显, 如 DSS 和 BS3 等交联剂均可以同时进行同位素标记. 一般而言, 特色交联剂虽然有各自的优势, 但由于交联效率、碎裂效率、可溶性、成本、鉴定效率等问题, 广泛使用的交联剂仍以简单交联剂为主. 不过, 近年来随着技术的成熟与发展, 特色交联剂的鉴定数目也在逐渐上升^[55].

Table 1 Characteristics of some common linkers (Sorted by first publication time in ascending order)

表 1 部分常见交联剂特性 (按首次发表时间升序排列)

交联剂	发表年份[参考文献]	富集作用	信号离子	同位素标记	交联剂断裂类型	臂长	额外特性
PIR	2005 [34], 2006 [35], 2009 [36]	√	√		气相/质谱断裂	43Å	可用于体内交联
SuDP (CID-CXL-MS/MS)	2006 [56], 2007 [57], 2012 [58]				气相/质谱断裂	11.2Å	
BAMJ	2007 [59]				化学断裂	7.5Å	
IRCX	2008 [60]				光断裂	16Å	
BiPS	2009 [48]			√	光断裂	10Å	交联剂带有荧光
SBC	2009 [49]			√	光断裂	10.2Å	可与多种氨基酸 发生交联反应
Azide-DSG	2010 [37]	√	√		气相/质谱断裂	7.7Å	
DEB	2010 [38]	√	√		气相/质谱断裂	7.3Å	支持 ETD 碎裂
BuUrBu	2010 [43]		√		气相/质谱断裂	12.5Å	
pcPIR	2010 [45]		√		光断裂	40Å	
BAH	2010 [46]		√		气相/质谱断裂	12.3Å	支持 ETD 碎裂
CBDPS	2011 [39]	√	√	√	气相/质谱断裂	14Å	
SSIN	2011 [42]	√			光断裂	14Å	
DSSO	2011 [47]		√		气相/质谱断裂	10.1Å	
BDRG	2012 [40]	√	√		气相/质谱断裂	27.0Å	
CXLs	2012 [41]	√	√	√	气相/质谱断裂	20Å	
DC4	2012 [44]		√		气相/质谱断裂	18Å	

4 交联质谱谱图鉴定

4.1 交联质谱谱图鉴定难点

蛋白质交联质谱数据, 与常规单肽段质谱数据相比, 在鉴定上更具有挑战性, 主要表现在下述几点:

4.1.1 丰度低

由于每个位点的实际交联效率往往远低于100%, 所以交联肽段丰度偏低, 交联谱图数目少、信号差, 给交联谱图鉴定带来困难(图3)。

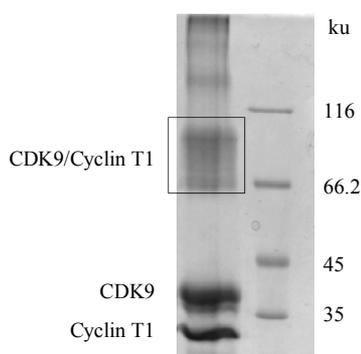


Fig. 3 Low abundance of inter-linked peptides

图3 充分交联情况下交联肽段含量仍然较低

4.1.2 交联形式多样

肽段交联后会形成多种组合形式(图4)。相较于普通肽段来讲, 更加复杂的组合形式, 给鉴定算法的设计和软件架构带来了一定的困难^[61-62]。

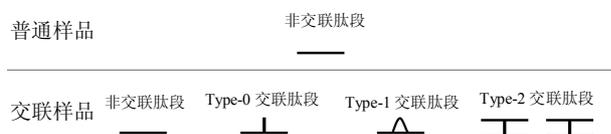


Fig. 4 Complex peptide forms after cross-linking

图4 肽段交联后会形成多种组合形式

4.1.3 谱图更复杂

由于存在着两条肽段上的多处断裂形成的离子, 以及因交联而产生的交联特异离子, 交联肽段二级谱图的碎片离子比起常规谱图在种类和数目上都更多(图5)^[57, 60, 62-65]。

4.1.4 搜索空间巨大

交联肽段是由两条肽段进行组合而形成, 所以

理论上讲, 其面对的数据库规模是普通数据库搜索的平方级规模(表2)。

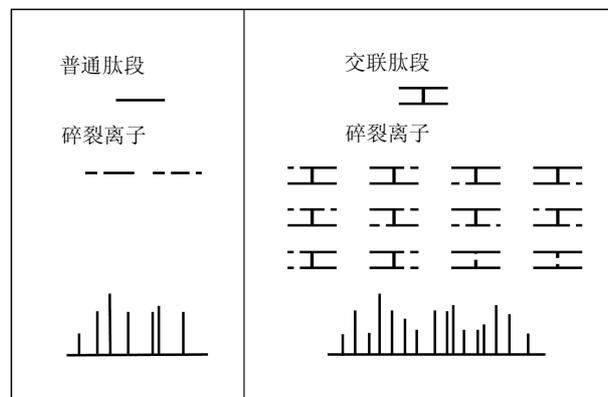


Fig. 5 More ion types in cross-linked MS/MS spectra

图5 交联肽段二级谱图的离子种类更多

Table 2 The search space difference between traditional database and cross-linked database (Trypsin)

表2 常规数据库与交联数据库搜索空间上的差别 (Trypsin 酶切)

Database	Proteins	Peptides	Peptide pairs
<i>E. coli</i> (MG1655)	6 126	3.35×10^5	5.63×10^{10}
<i>C. elegans</i> (WS232)	24 652	1.18×10^6	6.96×10^{11}
<i>H. sapiens</i> (ipi.HUMAN.v3.87)	91 491	3.72×10^6	2.77×10^{13}

4.1.5 缺乏评价方法

由于交联肽段由两条肽段组成, 无法直接利用传统统计方法进行结果的正确性评价。依靠人工、阈值等方法, 在准确度、可行性等方面都存在问题。所以在交联质谱谱图鉴定的错误率控制上, 必须要发展出新的一套方法来判断鉴定结果的可信度。

4.2 交联质谱谱图鉴定算法与软件

交联质谱谱图数据早期主要依靠手工解析。手工解析方法存在着效率低、标准不统一和易出错等问题, 并且由于交联肽段谱图本身的复杂性, 手工解析也非常困难。在这种情况下, 交联质谱数据的自动鉴定算法研究与软件开发, 就显得尤为重要。

4.2.1 一级谱图鉴定

2000~2005年, 交联质谱鉴定一般是基于一二级质谱谱图进行。一级质谱鉴定的原理是, 首先计算目标交联肽段的母离子理论质量, 再在一级谱图中寻找对应于这个质量的候选谱峰。当谱峰存在时, 即认为交联肽段存在, 进一步可利用二级谱图

进行人工标注. 一级谱图鉴定往往也多与 MALDI-TOF 仪器合用. 这一时期比较著名的软件包括 Links^[13]、MS2Links^[22]、ProteinXXX^[66] 和 CLPM^[67] 等^[54, 68-72].

Links (又名 ASAP) 是最早的交联肽段一级谱图鉴定软件, 发表于 2000 年. 文中利用 Links 软件得到交联肽段鉴定结果后, 进一步利用此信息对计算预测的 FGF-2 结构模型进行约束, 并验证其与已知结构一致. MS2Links (又名 MS2Assign) 是 Links 的后续软件, 它可与 Links 联用, 在得到可能的交联肽段鉴定结果后, 用 MS2Links 标注二级谱图, 再人工判定鉴定结果可靠性.

值得一提的是, 利用同位素标记来辅助交联鉴定的方法在这段时期内出现^[53, 73-74]. 同位素标记法对于交联质谱数据的鉴定有一定的特殊作用, 关于其具体细节, 请参见特色交联剂部分.

一级谱图鉴定法, 由于其鉴定方法的限制, 只能用于纯化的单个蛋白质的研究. 此外, 基于一级谱图的鉴定方法, 也存在着错误率较高的问题. 高精度一级谱图理论上能够缩减候选肽的规模, 但实际上, 由于噪音、遗漏酶切、非特异酶切和修饰等因素的影响, 往往无法确定肽段的身份. 另外, 仅用是否存在与交联肽段质量一致的谱峰进行鉴定, 也是不可靠的. 因而, 这种方法一直有较高的错误率, 需要进一步利用如 MS2Links (MS2Assign)^[22] 这样的工具来进行人工标注. 这些缺陷导致该方法逐渐被淘汰, 二级谱图鉴定成为了主流方法.

4.2.2 二级谱图鉴定

交联肽段质谱数据的二级谱图鉴定方法是随着质谱仪的发展而出现的, 主要分为两大类: 一类是将交联肽段的二级谱图鉴定问题设法转化为非交联肽段的常规肽段二级谱图鉴定问题, 再利用常规肽段的数据库搜索引擎进行鉴定^[34, 71, 75-79]; 另一类是直接针对交联肽段的二级谱图发掘规律, 设计和实现专用于交联肽段鉴定的二级谱图数据库搜索引擎^[30, 54, 73-74, 80-91]. 第一类是早年二级谱图鉴定的主要方法, 第二类在人们对交联肽段二级谱图的理解逐渐加深后, 得到了广泛的发展.

a. 将交联肽段鉴定问题转化为非交联肽段鉴定问题

这一类方法可利用常规数据库搜索引擎进行搜索. 具体又可分为三大类技术路线: I. 处理数据库, 使之可搜索得到交联肽段候选肽; II. 设置较大的母离子误差窗口, 进行开放式搜索; III. 利用

特殊交联剂, 使交联肽段在二级谱图中断裂, 形成普通单肽. 分述如下.

I. 数据库处理类

数据库处理类是通过将蛋白质数据库的拼合处理, 将交联肽段候选肽转化为线性肽段候选肽, 从而能够进行常规数据库搜索的一种方法. 其原理是, 通过线性组合两条单肽而得到的长肽段, 与真实的交联肽段对相比, 有一半的碎片离子质量是完全一致的. 利用这部分离子进行匹配, 就可以在一定程度上鉴定到部分交联结果.

2007 年, 由 Rappsilber 等设计提出的 XDB 方法, 是首个数据库处理类方法的尝试^[75], 后续 XComb 方法对 XDB 方法进行了进一步的改进^[76]. XDB 方法利用预搜索得到的可靠单肽段结果进行全组合式首尾拼合, 构建连接数据库, 进一步利用连接数据库和遗漏酶切设置, 就可以使用常规肽段搜索引擎鉴定得到交联肽段结果. XComb 方法将全组合式长蛋白质连接方法改进为短蛋白连接方法, 去除冗余肽段, 提高了鉴定效率.

交联肽段的数据库处理方法, 可以利用并不复杂的数据库处理过程将交联肽段对的搜索空间转化为常规线性肽段的搜索空间, 从而可以使用成熟的常规搜索引擎进行搜索, 是专用工具出现之前的变通方法. 但是, 这种路线有其本身的缺陷: 其一, 通过线性组合得到的长肽段的理论碎片离子与真实的交联肽段对相比, 有一半的碎片离子不相符, 这使得在鉴定中这部分离子信息不能得到利用(图 6); 其二, 这种方法由于其全组合特性, 无法针对较大的数据库进行, 一般以 100 条蛋白质序列为上限, 从根本上限制了这种方法在复杂生物样品上的应用.



Fig. 6 Only parts of the ions are used (black solid line)
图 6 数据库处理方法与开放式搜索方法只利用部分碎片离子 (黑色实线部分)

II. 开放式搜索类

设置较大的母离子误差, 用常规数据库搜索引擎进行开放式搜索的方法, 最早是应用于普通质谱领域的修饰发现问题. 由于母离子误差窗口大至几

百 Da 范围, 发生了意外修饰的肽段能够进入到打分候选中, 并且因其有部分碎片离子与未发生修饰的肽段完全相同, 往往可以得到鉴定. 对于鉴定得到的谱图, 再进行人工核对, 即可发现意外修饰.

交联肽段可被看作是一条肽段在交联位点处发生了等于另一条肽段质量的修饰, 因而也可以使用开放式搜索方法进行鉴定. 2008 年, 由 Singh 等^[78]发表的 Popitam 方法, 首次利用了开放式搜索方法进行交联肽段的鉴定. Popitam 方法包括两步: 先用普通单肽数据库对数据进行常规搜索, 得到常规肽段结果, 并将高可信度结果对应的单肽谱图去除; 然后利用开放式搜索方法进行交联肽段的鉴定. 2010 年, Chu 等^[79]同样用开放式搜索软件 ProteinProspector 对已知结构的样品进行鉴定, 并通过已知的结构信息确认了交联鉴定结果的正确性.

开放式搜索, 也是将交联肽段鉴定问题转化为常规肽段鉴定问题的一种方法. 其核心思想是将交联肽段对中的一条肽段及交联剂部分视为一个大质量的修饰, 利用另一条肽段的碎片离子来进行常规鉴定. 这种方法与前面提到的数据库处理方法有同样的缺陷, 它只能利用一条肽段碎裂产生的碎片离子, 另一条肽段的碎裂则没有进行考虑. 对于某些单条肽段碎裂并不好的交联肽段谱图来讲, 用这种方法难以得到鉴定 (图 6).

III. 交联剂可断裂类

另外一种方法是利用特别设计的交联剂, 让两条交联肽段在二级质谱阶段分离开变为单肽, 再分别采集单肽的三级质谱图进行鉴定, 从而将交联肽段谱图鉴定问题转化为传统单肽段谱图鉴定问题. 这一类技术中的代表技术为 PIR (protein interaction reporter) 技术, 最早于 2005 年开始应用^[84, 71].

PIR 交联试剂一般由四部分组成, 氨基酸反应基团、可断裂部分、标记离子和富集用亲和基团. 可断裂部分在质谱仪中断裂, 将交联肽段断裂为两条单肽, 同时释放标记离子, 形成特殊的碎片信号峰 (reporter ion).

2007 年发表的 X-links 软件是与 PIR 相配合的软件^[77]. 软件首先依据碎片信号峰的存在性判断谱图是否含交联剂; 对于含交联剂谱图, 利用其母离子质量和交联剂断裂产生的单肽之间的质量关系来确定谱图交联形式; 之后质谱仪自动选择单肽段谱峰对应离子进行再次碎裂, 产生三级谱图; 最后再利用常规数据库搜索工具进行搜索. 2010 年发表的 Blinks 软件进一步作了改进^[92], 它比较了交联肽

段与断裂产生的单肽之间的色谱相似性, 减少误判; 同时集成了单肽段搜索引擎 Mascot^[93]的鉴定结果, 使引擎更加完整.

PIR 策略是用二级质谱将交联肽段对拆开, 生成普通单肽段谱图, 从而得以用常规数据库搜索工具进行搜索的方法. 这种方法同时可以利用碎片信号峰来进行谱图中交联剂存在性的判定; 后续这部分工作发展到可以进行体内交联, 进一步提高了其应用范围^[55, 94-96]. 这种方法的缺点是, 一般均需要与仪器控制软件进行联用, 同时必须使用特殊制作的交联剂, 门槛较高. 此外, 这类交联剂一般臂长较长, 获得的空间距离对于推算空间结构的约束作用弱一些.

b. 针对交联肽段二级谱图本身特点而设计的数据库搜索软件

专用于交联肽段二级谱图鉴定的数据库搜索引擎是随着质谱仪和实验技术的发展而出现的. 这一类软件同转化类软件的最大区别在于其利用了两条交联肽段主干碎裂产生的全部理论离子, 而不是仅其中部分离子. 这一点无法用任何现有常规肽段鉴定软件实现. 利用全部理论离子的好处在于, 谱图信息可得到完整利用, 鉴定正确交联肽段的能力更强.

由于交联肽段搜索空间随肽段数据库规模平方级增长, 这一类软件一度只能用于纯化样品或简单蛋白质复合物, 要求数据库不能太大. 如软件 ProCrossLink^[89-90]、XLink^[54]、X!Links^[83-84]、CRUX^[73]、Xlink-Identifier^[85]、FINDX^[86]和 StavroX^[88]等, 数据库规模上的限制使这类软件不能得到更广泛的应用. 后续软件设法提高了可支持数据库的规模^[30, 74, 81-82, 87, 91]. 除了鉴定软件之外, 用于交联数据存储和交联数据展示的软件也有一定的发展^[97].

目前支持大规模数据库的交联肽段搜索引擎有三个: xQuest, MassMatrix 和 pLink. 下文中将分别对这三个软件进行介绍.

I. xQuest

Rinner 等^[82]于 2008 年发表的 xQuest 软件是第一个尝试对一般意义上的大规模数据库 (大肠杆菌, *E. coli*) 进行交联肽段鉴定的数据库搜索引擎. 大数据库交联肽段搜索的主要困难在于两点: 第一, 非交联肽段丰度远超过交联肽段丰度, 生成的二级谱图数目更多, 对交联肽段谱图的鉴定产生干扰; 第二, 搜索空间大, 其候选肽规模是普通数据库中候选肽规模的平方数量级大小.

xQuest 利用轻重同位素标记的特殊交联剂, 解决了这两个问题. xQuest 所利用的轻重交联剂质量相差 12 Da, 等量混合后进行交联反应, 其交联产物在一级谱图上会形成相差为 12 Da 的两簇谱峰. 由于这两簇谱峰的质荷比距离一般大于碎裂窗口宽度, 轻重标记的交联肽段各自生成一张二级谱图. 对两张二级谱图的碎片离子进行对比, 含有交联剂的离子有 12 Da 的质荷比差距, 不含交联剂的则没有. 在此基础上构建离子索引, 根据区分后的碎片离子得到候选肽列表, 再组合候选肽、打分、过滤, 得到最终结果.

xQuest 类型的鉴定算法通过轻重同位素标记交联剂的方法, 一方面能够区分谱图对应的肽段是否含有交联剂, 一方面能够确定二级谱图中部分碎片离子是否含有交联剂, 并通过离子索引技术有效缩减了候选肽规模, 使得复杂大数据库、复杂样品的交联鉴定成为可能. xQuest 的不足之处在于: 第一, 要取得比较好的结果, 需使用特殊的轻重同位素标记的交联剂来进行实验, 轻重交联剂的反应效率对于结果也有非常大的影响; 第二, 轻重标记之间的质量差必须足够大, 使母离子落入不同的碎裂窗口; 第三, xQuest 无法一次得到常规单肽、Type-0 和 Type-1 类型的结果, 需要进行额外的搜索; 第四, 在 xQuest 于 2012 年发展出配套的结果评价软件 xProphet^[9]之前, 其假发现率控制也是一个问题, 鉴定错误可能较多.

II. MassMatrix

MassMatrix 也能够通过缩减数据库的方法来支持大规模交联数据库的搜索. 最早发表于 2007 年的 MassMatrix 软件, 起初用于常规数据库搜索, 进行普通肽段的数据库鉴定^[9]; 后于 2008 年发展成为可鉴定二硫键的交联质谱数据库搜索引擎, 利用两个统计模型和一个描述性模型进行打分, 得到二硫键数据的鉴定结果^[9]; 发展到 2010 年, MassMatrix 实现了大数据库 (*E. coli*) 搜索以鉴定交联质谱谱图, 并且不再局限于二硫键的交联类型, 有效地支持了多数交联质谱谱图鉴定任务^[9].

针对复杂样品, MassMatrix 设计了两步搜索的缩减数据库策略: 第一步, 在大规模数据库中, 预搜索普通单肽肽段, 得到高可信肽段列表; 第二步, 使用第一步中得到的比较显著的肽段建立小规模数据库, 取代原大规模数据库, 进行第二步的交联搜索. 这种方法的依据是, 交联肽段一般会与非交联肽段共存, 从而可以预先搜索得到普通单肽肽

段列表, 再构造数据库进行交联搜索. MassMatrix 方法工作完整有效, 但存在的一个问题是并没有使用有效的假发现率控制手段, 报告结果的标准通常是对三个引擎打分进行经验阈值限定, 导致结果中存在一定的不可估计的错误鉴定.

III. pLink

pLink 是于 2012 年发表的专用于交联肽段鉴定的数据库搜索软件^{[8], [10]}. 它同样支持大规模交联数据库的搜索. 除此之外, 它在交联质谱技术的各个环节上进行了优化, 克服了过去的交联鉴定软件存在的一些缺点, 在大规模 *E. coli* 数据库上利用简单交联剂取得了到目前为止最多的交联位点鉴定数目.

pLink 相较于以往的交联肽段数据库搜索软件, 除了同样具有充分利用全部碎片离子的特点外, 还具有自己的优点. 这主要体现在:

针对新一代高精度质谱仪器进行深度优化. 研究组设计合成了大规模标准交联肽段数据集, 挖掘其中的碎片离子特点, 确认内部离子、交联剂断裂离子等特殊离子对于交联肽段谱图鉴定的作用, 并且引入到搜索引擎的设计中. 这是之前的工作所没有的, 实际也证明, 在 HCD 碎裂高精度谱图中, 这些离子对于搜索引擎鉴定能发挥相当大的作用.

针对大数据库在交联情况下组合爆炸的问题, 设计了粗细两步打分的策略, 有效缩减候选肽规模, 提高鉴定速度. 与 xQuest 的缩减搜索空间方法相比, 这种方法不需要对交联剂进行同位素标记, 从而能得到更广泛的引用; 与 MassMatrix 相比, 如果一部分交联肽段对应的单肽肽段没有得到鉴定, 用这种方法也可以搜索得到.

在肽谱匹配打分设计上, 对业已成熟的核谱向量点积打分进行进一步的改进^[10], 引入离子共存性加权概念, 有效利用内部离子、交联剂断裂离子等新增离子, 提高了交联谱图鉴定率.

在鉴定结果的假发现率控制方面, 将传统目标-诱饵库策略加以推广, 首次提出了在交联情况下的假发现率控制公式, 从而使利用目标-诱饵库策略控制交联鉴定的假发现率成为可能, 具体见下文结果验证部分.

在应用范围上更广. 由于 pLink 算法本身具有通用性, 样品制备不需要进行同位素标记, 也不需要利用特色交联剂, 对交联剂的臂长、质量、对称性等性质均无要求, 给生物实验的灵活性带来了很大的便利.

pLink 交联肽段搜索引擎在前人工作的基础上, 针对高精度数据, 深入挖掘交联质谱谱图离子特征, 设计实现了整套高效的算法, 是目前交联鉴定搜索引擎中的佼佼者。就交联鉴定效果而言,

pLink 在 *E. coli* 全裂解数据中利用简单交联剂鉴定到 394 对交联肽段对, 创下了当时的最高纪录。

表 3 总结了常见交联软件及其特性。

Table 3 Characteristics of some common CXMS software tools (Sorted by first publication time in ascending order)
表 3 部分常见交联软件特性 (按首次发表时间升序排列)

软件	发表年份 [参考文献]	一级 / 二级 谱图鉴定	离子 利用	交联剂限制	支持 免费(单机 / 网页)	评价 方法	谱图 显示	获取方法 (2013.9.1 检查)
GPMAW	2000[66]	一级	-	无限制	免费单机	A	√	http://gpmaw.com/
Links,	2000[13],	一级	-	无限制	免费单机	A	√	http://ms3d.ca.sandia.gov (不能访问)
MS2Links	2003[22]							
CLPM	2005[67]	一级	-	无限制	网页	无		http://bioinformatics.ualr.edu/mbc/services/CLPM.html
X-links	2005[34], 2007[77]	二级 / 三级	A	PIR	√ 免费单机	C		http://www.wsu.edu/proteomics/ (不能访问)
virtualmslab	2006[72]	一级	-	无限制	-	无		Send email to ldk@science.uva.nl
SearchXLinks	2006[102]	一级	-	二硫键交联	免费单机	无		http://www.searchxlinks.de (不能访问)
Pro-CrossLink	2006[90], 2008[89]	二级	B	¹⁸ O 标记	免费单机	A	√	http://depts.washington.edu/medchem/faculty/NelsonS.htm
XLink	2006[54]	二级	B	¹⁸ O 标记	免费单机	A	√	http://tools.proteomecenter.org/XLink.php
X!Links	2007[84], 2009[83]	二级	B	无限制	-	B		-
XDB	2007[75]	二级	A	无限制	免费单机	C		-
Popitam	2008[78]	二级	A	无限制	-	A		http://www.expasy.ch/tools/popitam/ (不能访问)
MassMatrix	2008[91], 2010[80]	二级	B	无限制	√ 免费单机和 网页	B		http://www.massmatrix.net
CrossSearch	2008[30]	二级	B	无限制	-	A		-
xQuest	2008[81], 2012[98]	二级	B	大质量差同 位素标记	√ 网页 / 免费 单机	D	√	http://www.xquest.org
xComb	2010[76]	二级	A	无限制	-	无		-
Protein prospector	2010[79]	二级	A	无限制	免费单机	C		http://prospector.ucsf.edu
CRUX	2010[73]	二级	B	无限制	免费单机	B		http://noble.gs.washington.edu/proj/crux
DBond	2010[74]	二级	C	二硫键交联	-	B		Send email to paek@uos.ac.kr
USTags	2010[87]	二级	B	二硫键交联	√ -	无		-
BLinks	2010[92]	二级	A	PIR	√ 免费单机	C		http://brucelab.gs.washington.edu/BLinks.php
ICC-CLASS	2010[103]	二级	A	同位素标 记可断裂	√ 免费单机	无		http://www.creativemolecules.com/CM_Software.htm
Xlink-Identifier	2011[85]	二级	B	无限制	-	C	√	Send email to Xiuxia.Du@uncc.edu
pLink	2012[81]	二级	C	无限制	√ 免费单机	D	√	http://pfind.ict.ac.cn/software/pLink/index.html
FINDX	2012[86]	二级	B	同位素标记	-	无		Send email to findxlink@gmail.com
StavroX	2012[88]	二级	B	无限制	-	B	√	Send email to michael.goetze@biochemtech.uni-halle.de

-: 表示不适用此条目或未查询到相关信息。离子利用: A, 部分二级谱图主干碎片离子 (转化为非交联肽段鉴定问题); B, 全部二级谱图主干碎片离子 (针对交联肽段二级谱图设计); C, 全部二级谱图主干碎片离子及其他交联特征离子 (针对交联肽段二级谱图进行碎片特征挖掘)。评价方法: A, 人工读图判定; B, 打分经验阈值判定; C, 传统目标-诱饵库假发现率控制方法; D, 交联谱图假发现率控制方法。

4.3 交联质谱谱图鉴定结果验证

基于数据库搜索方法的质谱谱图鉴定，在一定程度上会发生随机匹配，使鉴定结果中存在着不正确的结果；交联质谱谱图鉴定，也有着同样的问题。而且，对于交联质谱数据，由于数据库中候选肽规模庞大，随机匹配更容易发生，一定要有比较有效的结果评价方法对结果中的假发现率 (false discovery rate, FDR) 进行控制，才能得到可信的结果。

在常规数据库搜索引擎中，目前使用最普遍的假发现率控制方法，是有统计理论基础的目标 - 诱饵库方法^[104-105]。这种方法是将数据库进行处理，生成含有标记的随机诱饵数据库，合并目标库和诱饵库之后数据库搜索；在鉴定结果中，利用鉴定得到的随机匹配到诱饵数据库的谱图数量来估计目标数据库中随机匹配的数目。可惜的是，由于在交联质谱中同时存在两条肽段，该方法不能简单使用，这就需要对结果的评价方法作出改进。

早年的一级谱图鉴定方法中，对结果的判定通过后期人工读图来进行，如 GPMW^[66]、与 ASAP 联用的 MS2Assign^[22]等。人工判定方法处理数据的通量较低，且由于不同人的判定标准不一致，主观性大，不易推广。

利用二级谱图肽谱匹配时，由于交联质谱数据库搜索引擎一般会对二级谱图的肽谱匹配给出一定的打分，因而打分经验阈值法也是一种常见方法。其做法是针对引擎的一个或几个打分，利用标注集等经验判定方法来给出一定的可能分数阈值，对于分数阈值以下的鉴定判定为不可靠，分数阈值以上的鉴定作为结果进行输出。这种方法操作简便，使用容易；不足之处是经验阈值可靠性差，在不同数据集上往往不具备推广意义，无法控制结果中错误鉴定的数目。

数据库处理、开放式搜索和特殊可碎裂交联剂等方法，由于是将交联质谱谱图鉴定问题转化为普通肽段谱图鉴定问题，所以跟常规数据库搜索引擎一样，可以沿用传统的目标 - 诱饵库方法来估计结果中的假发现率。在交联质谱谱图数据的随机库构建方法上，针对交联数据的特点，可以利用随机交联剂质量或是随机交联位点氨基酸来配置^[75]。不过，这些构建诱饵库的方法也有各自的缺陷：用随机交联剂质量的方法，随机质量不易把握；随机交联位点氨基酸的方法，则容易因氨基酸含量上的差别而造成偏差。

部分引擎直接使用传统的目标 - 诱饵库策略进行假发现率的控制，即用匹配到随机肽段的交联谱图数目来估计目标肽段的假发现率。由于交联谱图中存在着全对、半对和全错三种情况 (图 7)，这使其与常规肽段匹配情况不同。在这种情况下，忽略半对情况，依靠传统策略用目标库和诱饵库的匹配数目来进行交联谱图鉴定的假发现率估计，往往是不准确的。

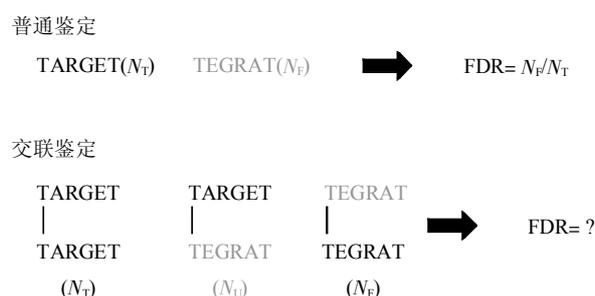


Fig. 7 Target-decoy FDR control not directly suitable for CXMS

图 7 交联质谱技术难以用传统方法直接估计 FDR

2012 年，中国科学院计算技术研究所 pFind 团队吴妍洁等^[81, 100]和瑞士分子系统生物学研究所 Walzthoeni 等^[98]，同时提出了在交联谱图鉴定情况下，利用目标 - 诱饵库进行假发现率估计的方法。在交联谱图鉴定情况下，存在着全对、半对和全错三种情况，通过一定的统计方法进行推断和估计，双方提出了同样的交联谱图估计公式：

$$FDR = N_{T+}(N_U - 2*N_F)/N_T = (N_U - N_F)/N_T$$

其中，N_T 为鉴定结果中交联的两条肽段均匹配到目标库的结果谱图数目，N_U 为鉴定结果中一条肽段匹配到目标库，一条肽段匹配到诱饵库的结果，N_F 为鉴定结果中两条肽段均匹配到诱饵库的结果谱图数目。

利用推广至交联谱图情况下的目标 - 诱饵库交联公式，可以有效地进行交联谱图数据库搜索的假发现率控制，而且操作简单，是目前交联谱图数据库搜索鉴定结果评估的最有效方法。

5 交联质谱技术的应用

交联质谱技术，主要用于辅助进行蛋白质的结构分析和相互作用研究。在结构研究中，交联质谱技术能够提供低分辨的结构信息，与蛋白质三维结

构预测软件相结合能够有效确定蛋白质的空间构象。在蛋白质相互作用研究中,一方面,化学交联能够固定弱的瞬时的相互作用,另一方面,化学交联能够帮助区分蛋白质之间的直接相互作用与间接相互作用,确定蛋白质相互作用的界面。

5.1 交联质谱技术在蛋白质结构研究中的应用

化学交联剂的臂长决定了能够被交联的两个氨基酸的空间距离上限。通过使用相同类型、不同碳链长度的交联剂,理论上能够更进一步确定被交联氨基酸的空间距离。与蛋白质三维结构预测软件相结合,有了化学交联肽段提供的空间距离限制,能够减少蛋白质结构预测的运算量。当软件预测出多种蛋白质构象,无法进一步区分哪一种更合理时,化学交联质谱数据可以辅助排除一些不合理的构象。对于一些大型的蛋白质复合物,通过不同亚基之间的交联信息,能够了解这些蛋白质复合物的组装方式,知道不同亚基之间的相互作用界面。

2000年,Young等^[13]尝试寻找成纤维细胞生长因子 FGF-2 蛋白的折叠模型 (protein fold)。仅使用穿线法 (sequence threading),两个正确答案 (与 FGF-2 同属于 β -trefoil fold family 的蛋白) 排名第 5 和第 12 位;引入化学交联质谱所提供的空间距离信息,排名上升至第 1 和第 3 位。

晶体结构未知的蛋白质及其复合物,例如 nebulin-calmodulin^[69]、calmodulin-melittin^[106]、BSA^[102]、AbrB^[107]和 bacteriophage P22 procapsids^[108]等也曾使用交联质谱方法辅助进行结构判定。除此之外,交联质谱技术也被应用于晶体结构已知的蛋白质及其复合物,做为结构信息的辅助验证,如 CaM/M13^[109-110]、CYP2E1-b5^[78, 111]和 CYTOCHROME C^[84, 112]等;甚至用于全 *E. coli* 库的交联蛋白质鉴定问题中^[81-82, 94]。

5.2 交联质谱技术在蛋白质相互作用研究中的应用

5.2.1 固定弱的或者瞬时的蛋白质-蛋白质相互作用

交联技术与普通的免疫共沉淀实验方法相结合,能够通过共价键固定弱的或者瞬时的相互作用。普通的免疫共沉淀实验中使用比较温和的漂洗条件,大量特异相互作用的蛋白质会被保留下来,这对于鉴定与目标蛋白质相互作用的蛋白质有利,不过很多非特异结合的污染蛋白质也会被保留下来。如果使用比较严格的漂洗条件,虽然能更好地去除污染蛋白质,但是一些弱的相互作用蛋白质会丢掉。引入化学交联,可以在目标蛋白质与相互作用的蛋白质之间形成共价连接,从而保护弱的或者

是瞬时的相互作用,随后使用非常严格的漂洗条件可以有效去除非特异的蛋白质,同时保留特异相互作用的蛋白质。

除了能够固定弱的或者瞬时的蛋白质相互作用外,从免疫共沉淀样品中鉴定得到交联肽段,还能有效地确定蛋白质之间的相互作用是直接相互作用还是间接相互作用(免疫共沉淀技术本身并不能确定这一点)。以往的尝试中通常使用甲醛交联免疫共沉淀样品。由于至今无法有效地从甲醛交联样品中直接鉴定到交联多肽,所以直到最近,化学交联技术都仅用于固定相互作用,而无法区分蛋白质之间是直接相互作用还是间接相互作用。

2012年,Yang等^[10]成功地使用 BS3 交联剂处理 FIB-1 免疫共沉淀样品,再利用 pLink 鉴定得到了交联肽段,从而在固定相互作用的基础上得以进一步鉴定直接相互作用蛋白质,最终鉴定结果与结构模型吻合。与以往相比,之所以这种方法能够成功,一方面是由于在实验步骤上进行了改革和优化,另一方面则在于软件鉴定算法的进步。同年,Herzog等^[113]亲和纯化蛋白磷酸酶 2A (protein phosphatase 2A, PP2A) 或其结合蛋白的蛋白复合体,也通过化学交联质谱的方法,揭示了 PP2A 通过不同的调节亚基或桥梁蛋白与不同底物蛋白相互作用的网络。

5.2.2 鉴定蛋白质相互作用界面

常用的交联剂臂长都很短,从而保证被交联的氨基酸必定是处在一定的空间距离内。这样,不同蛋白质之间的交联一般只能发生在蛋白质相互作用的界面上。通过鉴定交联肽段,能够很好地找到蛋白质相互作用的界面,帮助解释蛋白质之间相互作用的机制。

举例来讲,获取蛋白质复合物三维空间形状的电镜技术和用于研究结构功能关系的点突变实验,如果结合交联质谱技术,能够使实验更加高效,更具针对性。Jennebach等^[114]使用交联质谱技术,从 670 ku、含 15 个组成亚基、难以共结晶的 RNA 聚合酶 Pol II 转录起始复合物中,鉴定到上百个交联赖氨酸对,确定了 Pol II 与 TFIIF 结合的区域。这项研究工作揭示了 TFIIF 如何抑制 Pol II 非特异性结合 DNA 以及如何协助 Pol II 定位到启动子 DNA、确定转录起始位点。类似的方法还被用于解析 RNA 聚合酶 Pol I^[115]和 RNA 聚合酶 Pol III 蛋白复合体的结构^[116]。Ciferri等^[117]采用电镜结合交联质谱解析出了 polycomb repressive complex 2 (PRC2)

蛋白复合物的结构和亚基之间的组装关系。

5.2.3 全蛋白质组水平研究蛋白质相互作用

对细胞裂解液做化学交联处理，然后使用质谱鉴定化学交联多肽，能够在整体水平上鉴定蛋白质之间的相互作用。从大肠杆菌细胞裂解液的交联样品中，Xu 等^[80]鉴定到了 20 对交联多肽；Rinner 等^[82]鉴定到了 71 对交联多肽；Yang 等^[81]在大肠杆菌全蛋白质组样品的交联实验中鉴定得到了 394 对交联多肽；Weisbrod 等^[94]利用特殊交联剂鉴定得到了 708 对交联多肽。这其中，Xu、Rinner 和 Yang 等利用简单交联剂 BS3 或 DSS 进行交联肽段搜索鉴定，Weisbrod 等利用特殊合成的可断裂交联剂 PIR 将交联肽段断裂成单肽进行单肽段搜索鉴定；在鉴定结果的假发现率控制上，Xu 和 Rinner 等利用分数阈值法和部分已知结构进行验证，Yang 等利用假发现率估计公式按用户设定值进行自动控制，Weisbrod 等则是使用人工核实的方法进行结果评估。

这些实验的设计是为了证明交联质谱技术处理复杂样品的能力。从目前的实验水平来看，利用全蛋白质组的交联实验鉴定所有相互作用的蛋白质，不是十分现实，还有很长的一段路要走。但是，我们不能忽略的是，交联质谱技术各个相关技术环节都在不断进步，例如，具有富集功能交联剂的不断优化、质谱仪越来越灵敏、扫描速度越来越快、算法越来越先进，等等，全蛋白质组水平研究蛋白质相互作用的研究也许不会很遥远。

6 总结与展望

交联质谱技术结合了传统化学交联方法与现代质谱方法，作为一种近十几年来新兴的技术，正日益显示出其在蛋白质结构和相互作用研究中的价值和意义。高通量、低价格、高精度，结合目前逐渐完备成熟的软件体系，交联质谱技术在蛋白质结构和相互作用研究中已逐步成为常规的组学分析技术手段，在越来越多的实验室得到应用。

目前，在大肠杆菌全细胞裂解液的交联实验中，用简单交联剂 BS3，最高可以鉴定到 394 对交联肽段；利用特殊设计的交联剂可以鉴定到 708 对。随着交联剂的发展和鉴定算法的发展，这个数字有望进一步提升。而对于数据库较小的蛋白复合物，如牛血清白蛋白、核糖核酸酶 A (ribonuclease A, RNase A) 等，在当前的技术水平下，通过一定的实验和算法优化，对高丰度的交联形式如二硫键

等，全部的交联位点通常均能得到鉴定。

交联质谱技术的发展，除了与质谱仪的发展、实验技术的进步密切相关之外，主要依托于两条主线的发展：交联剂的发展和交联质谱鉴定方法的发展。交联剂上，从传统的化学交联剂，到特别开发的针对质谱而设计的多种交联剂，各有特点，从而使研究者在利用交联质谱技术进行研究时，有了更多选择。交联质谱鉴定方法上，从早年的一级谱图母离子匹配鉴定，到同位素交联方法的引入、二级谱图鉴定、二级交联谱图的离子特点挖掘，利用信息越来越全面，鉴定精度越来越高；算法上，肽谱匹配打分的算法设计越来越精巧，区分性越来越强，从只考虑母离子误差大小，到考虑二级谱图的离子匹配个数，到基于概率的打分模型；鉴定能力上，从早年只能支持单一纯化样品，到略复杂样品的小规模鉴定，到复杂样品的大规模鉴定，交联质谱技术的适用性越来越广。

交联质谱方法上，随着信息学上障碍的逐渐清除，交联质谱技术有望成为质谱实验室的常规技术^[14]。但是，目前仍存在着众多有挑战性的问题。

举例来讲，由于酶切效率、酶切位点分布等因素的影响，在常规交联实验中，除了本文综述的两条肽段交联在一起而形成的二肽交联形式之外，三条肽段交联产生的三肽交联形式，以及更多肽段彼此交联产生的多肽交联形式，也是广泛存在的(图 8)。部分蛋白质复合物由于交联位点距离近，在实验中甚至完全无法得到二肽交联形式的肽段，而只有通过多肽交联鉴定才能确定其交联产物。但是，由于多肽交联鉴定问题相对二肽交联鉴定来讲，数据库规模更大、谱图离子更多、样品复杂度更高，目前国际国内尚没有太多研究。

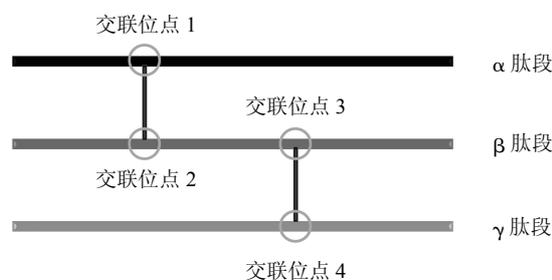


Fig. 8 Triple-linked peptide

图 8 三肽交联形式肽段

除此之外，即使在二肽交联鉴定方面，由于交联质谱技术本身的数据库规模限制，目前的搜索引

擎对于超大规模数据库的处理, 也依然存在着鉴定速度低、灵敏度有待提高等问题。

交联质谱领域是一个富有潜力的潜在技术领域。可以预见, 随着前端实验技术的成熟, 和后端算法的完善, 交联蛋白质鉴定有望同普通蛋白质鉴定一样简单有效, 得到广泛应用, 在科学研究中占有一席之地。

参 考 文 献

- [1] Berggard T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics*, 2007, **7**(16): 2833–2842
- [2] Phizicky E M, Fields S. Protein-protein interactions: methods for detection and analysis. *Microbiol Rev*, 1995, **59**(1): 94–123
- [3] Wilderman P R, Shah M B, Liu T, *et al.* Plasticity of cytochrome P450 2B4 as investigated by hydrogen-deuterium exchange mass spectrometry and X-ray crystallography. *J Biol Chem*, 2010, **285**(49): 38602–38611
- [4] Zuiderweg E R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, 2002, **41**(1): 1–7
- [5] Marion D. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics*, 2013, **12**(11): 3006–3025
- [6] Shen Z, Cloud K G, Chen D J, *et al.* Specific interactions between the human RAD51 and RAD52 proteins. *J Biol Chem*, 1996, **271**(1): 148–152
- [7] Howell J M, Winstone T L, Coorssen J R, *et al.* An evaluation of *in vitro* protein-protein interaction techniques: assessing contaminating background proteins. *Proteomics*, 2006, **6**(7): 2050–2069
- [8] Ethier M, Lambert J P, Vasilescu J, *et al.* Analysis of protein interaction networks using mass spectrometry compatible techniques. *Anal Chim Acta*, 2006, **564**(1): 10–18
- [9] Guerrero C, Tagwerker C, Kaiser P, *et al.* An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified *in vivo* cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics*, 2006, **5**(2): 366–378
- [10] Klockenbusch C, Kast J. Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin beta1. *J Biomed Biotechnol*, 2010, **2010**: 927585
- [11] Klockenbusch C, O'Hara J E, Kast J. Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal Bioanal Chem*, 2012, **404**(4): 1057–1067
- [12] Petrotchenko E V, Olkhovik V K, Borchers C H. Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol Cell Proteomics*, 2005, **4**(8): 1167–1179
- [13] Young M M, Tang N, Hempel J C, *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA*, 2000, **97**(11): 5802–5806
- [14] Trakselis M A, Alley S C, Ishmael F T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug Chem*, 2005, **16**(4): 741–750
- [15] Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*, 2006, **25**(4): 663–682
- [16] Singh P, Panchaud A, Goodlett D R. Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique. *Anal Chem*, 2010, **82**(7): 2636–2642
- [17] Tang X, Bruce J E. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. *Mol Biosyst*, 2010, **6**(6): 939–947
- [18] Leitner A, Walzthoeni T, Kahraman A, *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics*, 2010, **9**(8): 1634–1649
- [19] Sinz A. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal Bioanal Chem*, 2010, **397**(8): 3433–3440
- [20] Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol*, 2011, **173**(3): 530–540
- [21] Stengel F, Aebersold R, Robinson C V. Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Mol Cell Proteomics*, 2012, **11**(3): R111 014027
- [22] Schilling B, Row R H, Gibson B W, *et al.* MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom*, 2003, **14**(8): 834–850
- [23] Kruppa G H, Schoeniger J, Young M M. A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*, 2003, **17**(2): 155–162
- [24] Fritzsche R, Ihling C H, Gotze M, *et al.* Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun Mass Spectrom*, 2012, **26**(6): 653–658
- [25] Leitner A, Reischl R, Walzthoeni T, *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics*, 2012, **11**(3): M111 014126
- [26] Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem*, 2008, **392**(1–2): 305–312
- [27] Madler S, Bich C, Touboul D, *et al.* Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J Mass Spectrom*, 2009, **44**(5): 694–706
- [28] Zhang Q, Crosland E, Fabris D. Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies. *Anal Chim Acta*, 2008, **627**(1): 117–128
- [29] Novak P, Kruppa G H. Intra-molecular cross-linking of acidic residues for protein structure studies. *Eur J Mass Spectrom (Chichester, Eng)*, 2008, **14**(6): 355–365
- [30] Nadeau O W, Wyckoff G J, Paschall J E, *et al.* CrossSearch, a user-friendly search engine for detecting chemically cross-linked peptides in conjugated proteins. *Mol Cell Proteomics*, 2008, **7**(4):

- 739–749
- [31] Rozbesky D, Man P, Kavan D, *et al.* Chemical cross-linking and H/D exchange for fast refinement of protein crystal structure. *Anal Chem*, 2012, **84**(2): 867–870
- [32] Toews J, Rogalski J C, Kast J. Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal Chim Acta*, 2010, **676**(1–2): 60–67
- [33] Chu F, Mahrus S, Craik C S, *et al.* Isotope-coded and affinity-tagged cross-linking (ICATXL): an efficient strategy to probe protein interaction surfaces. *J Am Chem Soc*, 2006, **128**(32): 10362–10363
- [34] Tang X, Munske G R, Siems W F, *et al.* Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem*, 2005, **77**(1): 311–318
- [35] Chowdhury S M, Munske G R, Tang X, *et al.* Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers. *Anal Chem*, 2006, **78**(24): 8183–8193
- [36] Zhang H, Tang X, Munske G R, *et al.* Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol Cell Proteomics*, 2009, **8**(3): 409–420
- [37] Vellucci D, Kao A, Kaake R M, *et al.* Selective enrichment and identification of azide-tagged cross-linked peptides using chemical ligation and mass spectrometry. *J Am Soc Mass Spectrom*, 2010, **21**(8): 1432–1445
- [38] Trnka M J, Burlingame A L. Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene: the power of high resolution electron transfer dissociation for determination of both peptide sequences and their attachment sites. *Mol Cell Proteomics*, 2010, **9**(10): 2306–2317
- [39] Petrotchenko E V, Serpa J J, Borchers C H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Mol Cell Proteomics*, 2011, **10**(2): M110 001420
- [40] Luo J, Fishburn J, Hahn S, *et al.* An integrated chemical cross-linking and mass spectrometry approach to study protein complex architecture and function. *Mol Cell Proteomics*, 2012, **11**(2): M111 008318
- [41] Sohn C H, Agnew H D, Lee J E, *et al.* Designer reagents for mass spectrometry-based proteomics: clickable cross-linkers for elucidation of protein structures and interactions. *Anal Chem*, 2012, **84**(6): 2662–2669
- [42] Yan F, Che F Y, Nieves E, *et al.* Photo-assisted peptide enrichment in protein complex cross-linking analysis of a model homodimeric protein using mass spectrometry. *Proteomics*, 2011, **11**(20): 4109–4115
- [43] Muller M Q, Dreiocker F, Ihling C H, *et al.* Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem*, 2010, **82**(16): 6958–6968
- [44] Clifford-Nunn B, Showalter H D, Andrews P C. Quaternary diamines as mass spectrometry cleavable crosslinkers for protein interactions. *J Am Soc Mass Spectrom*, 2012, **23**(2): 201–212
- [45] Yang L, Tang X, Weisbrod C R, *et al.* A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. *Anal Chem*, 2010, **82**(9): 3556–3566
- [46] Gardner M W, Brodbelt J S. Preferential cleavage of N-N hydrazone bonds for sequencing bis-arylhydrazone conjugated peptides by electron transfer dissociation. *Anal Chem*, 2010, **82**(13): 5751–5759
- [47] Kao A, Chiu C L, Vellucci D, *et al.* Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics*, 2011, **10**(1): M110 002212
- [48] Petrotchenko E V, Xiao K, Cable J, *et al.* BiPS, a photocleavable, isotopically coded, fluorescent cross-linker for structural proteomics. *Mol Cell Proteomics*, 2009, **8**(2): 273–286
- [49] Krauth F, Ihling C H, Ruttinger H H, *et al.* Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and electrospray ionization LTQ-Orbitrap mass spectrometry. *Rapid Commun Mass Spectrom*, 2009, **23**(17): 2811–2818
- [50] Gygi S P, Rist B, Gerber S A, *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, 1999, **17**(10): 994–999
- [51] Petrotchenko E V, Serpa J J, Borchers C H. Use of a combination of isotopically coded cross-linkers and isotopically coded N-terminal modification reagents for selective identification of inter-peptide crosslinks. *Anal Chem*, 2010, **82**(3): 817–823
- [52] Muller D R, Schindler P, Towbin H, *et al.* Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem*, 2001, **73**(9): 1927–1934
- [53] Collins C J, Schilling B, Young M, *et al.* Isotopically labeled crosslinking reagents: resolution of mass degeneracy in the identification of crosslinked peptides. *Bioorg Med Chem Lett*, 2003, **13**(22): 4023–4026
- [54] Seebacher J, Mallick P, Zhang N, *et al.* Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J Proteome Res*, 2006, **5**(9): 2270–2282
- [55] Yang L, Zheng C, Weisbrod C R, *et al.* *In vivo* application of photocleavable protein interaction reporter technology. *J Proteome Res*, 2012, **11**(2): 1027–1041
- [56] Soderblom E J, Goshe M B. Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. *Anal Chem*, 2006, **78**(23): 8059–8068
- [57] Soderblom E J, Bobay B G, Cavanagh J, *et al.* Tandem mass spectrometry acquisition approaches to enhance identification of protein-protein interactions using low-energy collision-induced dissociative chemical crosslinking reagents. *Rapid Commun Mass Spectrom*, 2007, **21**(21): 3395–3408
- [58] Liu F, Wu C, Sweedler J V, *et al.* An enhanced protein crosslink

- identification strategy using CID-cleavable chemical crosslinkers and LC/MS(n) analysis. *Proteomics*, 2012, **12**(3): 401-405
- [59] Kasper P T, Back J W, Vitale M, *et al.* An aptly positioned azido group in the spacer of a protein cross-linker for facile mapping of lysines in close proximity. *Chembiochem*, 2007, **8**(11): 1281-1292
- [60] Gardner M W, Vasicek L A, Shabbir S, *et al.* Chromogenic cross-linker for the characterization of protein structure by infrared multiphoton dissociation mass spectrometry. *Anal Chem*, 2008, **80**(13): 4807-4819
- [61] Swaim C L, Smith J B, Smith D L. Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis (sulfosuccinimidyl propionate), DTSSP with peptides. *J Am Soc Mass Spectrom*, 2004, **15**(5): 736-749
- [62] Iglesias A H, Santos L F, Gozzo F C. Collision-induced dissociation of Lys-Lys intramolecular crosslinked peptides. *J Am Soc Mass Spectrom*, 2009, **20**(4): 557-566
- [63] Back J W, Hartog A F, Dekker H L, *et al.* A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes. *J Am Soc Mass Spectrom*, 2001, **12**(2): 222-227
- [64] Santos L F, Iglesias A H, Gozzo F C. Fragmentation features of intermolecular cross-linked peptides using N-hydroxy-succinimide esters by MALDI- and ESI-MS/MS for use in structural proteomics. *J Mass Spectrom*, 2011, **46**(8): 742-750
- [65] Gaucher S P, Hadi M Z, Young M M. Influence of crosslinker identity and position on gas-phase dissociation of Lys-Lys crosslinked peptides. *J Am Soc Mass Spectrom*, 2006, **17**(3): 395-405
- [66] Bennett K L, Kussmann M, Bjork P, *et al.* Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping—a novel approach to assess intermolecular protein contacts. *Protein Sci*, 2000, **9**(8): 1503-1518
- [67] Tang Y, Chen Y, Licht C F, *et al.* CLPM: a cross-linked peptide mapping algorithm for mass spectrometric analysis. *BMC Bioinformatics*, 2005, **6**(Suppl 2): S9
- [68] Back J W, Notenboom V, de Koning L J, *et al.* Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling. *Anal Chem*, 2002, **74**(17): 4417-4422
- [69] Sinz A, Wang K. Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin-calmodulin complexes. *Biochemistry*, 2001, **40**(26): 7903-7913
- [70] Chu F, Shan S O, Moustakas D T, *et al.* Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. *Proc Natl Acad Sci USA*, 2004, **101**(47): 16454-16459
- [71] Taverner T, Hall N E, O'Hair R A, *et al.* Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem*, 2002, **277**(48): 46487-46492
- [72] de Koning L J, Kasper P T, Back J W, *et al.* Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. *FEBS J*, 2006, **273**(2): 281-291
- [73] McIlwain S, Draghicescu P, Singh P, *et al.* Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J Proteome Res*, 2010, **9**(5): 2488-2495
- [74] Choi S, Jeong J, Na S, *et al.* New algorithm for the identification of intact disulfide linkages based on fragmentation characteristics in tandem mass spectra. *J Proteome Res*, 2010, **9**(1): 626-635
- [75] Maiolica A, Cittaro D, Borsotti D, *et al.* Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol Cell Proteomics*, 2007, **6**(12): 2200-2211
- [76] Panchaud A, Singh P, Shaffer S A, *et al.* xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J Proteome Res*, 2010, **9**(5): 2508-2515
- [77] Anderson G A, Tolic N, Tang X, *et al.* Informatics strategies for large-scale novel cross-linking analysis. *J Proteome Res*, 2007, **6**(9): 3412-3421
- [78] Singh P, Shaffer S A, Scherl A, *et al.* Characterization of protein cross-links *via* mass spectrometry and an open-modification search strategy. *Anal Chem*, 2008, **80**(22): 8799-8806
- [79] Chu F, Baker P R, Burlingame A L, *et al.* Finding chimeras: a bioinformatics strategy for identification of cross-linked peptides. *Mol Cell Proteomics*, 2010, **9**(1): 25-31
- [80] Xu H, Hsu P H, Zhang L, *et al.* Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. *J Proteome Res*, 2010, **9**(7): 3384-3393
- [81] Yang B, Wu Y J, Zhu M, *et al.* Identification of cross-linked peptides from complex samples. *Nat Methods*, 2012, **9**(9): 904-906
- [82] Rinner O, Seebacher J, Walzthoeni T, *et al.* Identification of cross-linked peptides from large sequence databases. *Nat Methods*, 2008, **5**(4): 315-318
- [83] Lee Y J. Probability-based shotgun cross-linking sites analysis. *J Am Soc Mass Spectrom*, 2009, **20**(10): 1896-1899
- [84] Lee Y J, Lackner L L, Nunnari J M, *et al.* Shotgun cross-linking analysis for studying quaternary and tertiary protein structures. *J Proteome Res*, 2007, **6**(10): 3908-3917
- [85] Du X, Chowdhury S M, Manes N P, *et al.* Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. *J Proteome Res*, 2011, **10**(3): 923-931
- [86] Soderberg C A, Lambert W, Kjellstrom S, *et al.* Detection of crosslinks within and between proteins by LC-MALDI-TOF/TOF and the software FINDX to reduce the MSMS-data to acquire for validation. *PLoS One*, 2012, **7**(6): e38927
- [87] Shen Y, Tolic N, Purvine S O, *et al.* Identification of disulfide bonds in protein proteolytic degradation products using *de novo*-protein unique sequence tags approach. *J Proteome Res*, 2010, **9**(8): 4053-4060
- [88] Gotze M, Pettelkau J, Schaks S, *et al.* StavroX—a software for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom*, 2012, **23**(1): 76-87
- [89] Gao Q, Xue S, Shaffer S A, *et al.* Minimize the detection of false positives by the software program DetectShift for 18O-labeled

- cross-linked peptide analysis. *Eur J Mass Spectrom* (Chichester, Eng), 2008, **14**(5): 275–280
- [90] Gao Q, Xue S, Doneanu C E, *et al.* Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. *Anal Chem*, 2006, **78**(7): 2145–2149
- [91] Xu H, Zhang L, Freitas M A. Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. *J Proteome Res*, 2008, **7**(1): 138–144
- [92] Hoopmann M R, Weisbrod C R, Bruce J E. Improved strategies for rapid identification of chemically cross-linked peptides using protein interaction reporter technology. *J Proteome Res*, 2010, **9**(12): 6323–6333
- [93] Perkins D N, Pappin D J, Creasy D M, *et al.* Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999, **20**(18): 3551–3567
- [94] Weisbrod C R, Chavez J D, Eng J K, *et al.* *In vivo* protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J Proteome Res*, 2013, **12**(4): 1569–1579
- [95] Bruce J E. *In vivo* protein complex topologies: sights through a cross-linking lens. *Proteomics*, 2012, **12**(10): 1565–1575
- [96] Zheng C, Yang L, Hoopmann M R, *et al.* Cross-linking measurements of *in vivo* protein complex topologies. *Mol Cell Proteomics*, 2011, **10**(10): M110 006841
- [97] Zheng C, Weisbrod C R, Chavez J D, *et al.* XLink-DB: database and software tools for storing and visualizing protein interaction topology data. *J Proteome Res*, 2013, **12**(4): 1989–1995
- [98] Walzthoeni T, Claassen M, Leitner A, *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Methods*, 2012, **9**(9): 901–903
- [99] Xu H, Freitas M A. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinformatics*, 2007, **8**: 133
- [100] 吴妍洁. 交联蛋白质谱鉴定的算法研究及软件开发. 北京: 中国科学院研究生院, 2011
- Wu Y J. Research on Cross-linked Protein Identification Algorithm and Software Development. Beijing: Graduate School of The Chinese Academy of Sciences, 2011
- [101] Fu Y, Yang Q, Sun R, *et al.* Exploiting the kernel trick to correlate fragment ions for peptide identification *via* tandem mass spectrometry. *Bioinformatics*, 2004, **20**(12): 1948–1954
- [102] Wefing S, Schnaible V, Hoffmann D. SearchXLinks. A program for the identification of disulfide bonds in proteins from mass spectra. *Anal Chem*, 2006, **78**(4): 1235–1241
- [103] Petrotchenko E V, Borchers C H. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC Bioinformatics*, 2010, **11**: 64
- [104] Elias J E, Haas W, Faherty B K, *et al.* Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods*, 2005, **2**(9): 667–675
- [105] Elias J E, Gygi S P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, 2007, **4**(3): 207–214
- [106] Schulz D M, Ihling C, Clore G M, *et al.* Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes. *Biochemistry*, 2004, **43**(16): 4703–4715
- [107] Olson A L, Liu F, Tucker A T, *et al.* Chemical crosslinking and LC/MS analysis to determine protein domain orientation: Application to AbrB. *Biochem Biophys Res Commun*, 2013, **431**(2): 253–257
- [108] Kang S, Hawkridge A M, Johnson K L, *et al.* Identification of subunit-subunit interactions in bacteriophage P22 procapsids by chemical cross-linking and mass spectrometry. *J Proteome Res*, 2006, **5**(2): 370–377
- [109] Kalkhof S, Ihling C, Mechtler K, *et al.* Chemical cross-linking and high-performance Fourier transform ion cyclotron resonance mass spectrometry for protein interaction analysis: application to a calmodulin/target peptide complex. *Anal Chem*, 2005, **77**(2): 495–503
- [110] Dimova K, Kalkhof S, Pottratz I, *et al.* Structural insights into the calmodulin-Munc13 interaction obtained by cross-linking and mass spectrometry. *Biochemistry*, 2009, **48**(25): 5908–5921
- [111] Gao Q, Doneanu C E, Shaffer S A, *et al.* Identification of the interactions between cytochrome P450 2E1 and cytochrome b5 by mass spectrometry and site-directed mutagenesis. *J Biol Chem*, 2006, **281**(29): 20404–20417
- [112] Dihazi G H, Sinz A. Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom*, 2003, **17**(17): 2005–2014
- [113] Herzog F, Kahraman A, Boehringer D, *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science*, 2012, **337**(6100): 1348–1352
- [114] Tabb D L. Evaluating protein interactions through cross-linking mass spectrometry. *Nat Methods*, 2012, **9**(9): 879–881
- [115] Jennebach S, Herzog F, Aebersold R, *et al.* Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res*, 2012, **40**(12): 5591–5601
- [116] Wu C C, Herzog F, Jennebach S, *et al.* RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proc Natl Acad Sci USA*, 2012, **109**(47): 19232–19237
- [117] Ciferri C, Lander G C, Maiolica A, *et al.* Molecular architecture of human polycomb repressive complex 2. *Elife*, 2012, **1**: e00005

A New Approach to Protein Structure and Interaction Research: Chemical Cross-linking in Combination With Mass Spectrometry*

FAN Sheng-Bo^{1,2}, WU Yan-Jie^{1,2}, YANG Bing³, CHI Hao^{1,2}, MENG Jia-Ming^{1,2}, LU Shan³,
ZHANG Kun^{1,2}, WU Long^{1,2}, SUN Rui-Xiang¹, DONG Meng-Qiu³**, HE Si-Min¹**

¹Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences,
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

²University of Chinese Academy of Sciences, Beijing 100049, China;

³National Institute of Biological Sciences, Beijing 102206, China)

Abstract For any protein, its 3-D structure and interaction with other proteins are vital information for the understanding of its function. Traditionally, NMR and X-ray crystallography are employed to acquire such information at high resolution. However, application of these methods is limited because they are extremely demanding for protein purity and quantity, or protein size and crystalizability. Chemical cross-linking in combination with mass spectrometry (CXMS) is a novel technology developed over the past decade that complements the traditional methods with its many advantages such as high throughput, low cost and relaxed sample requirements. Here we review the various aspects of CXMS, including separation and enrichment techniques, characteristics of common cross-linkers, development of database-search tools for CXMS data, algorithms for validation of results, and application. The future development of CXMS is also discussed.

Key words cross-linking mass spectrometry, protein structure, protein interaction, proteomics, mass spectrometry

DOI: 10.3724/SP.J.1206.2013.00484

* This work was supported by grants from The National Natural Science Foundation of China(21475141), and The National Basic Research Program of China (2010CB912701, 2012CB910602, 2013CB911203).

**Corresponding author. Tel: 86-10-62600822

DONG Meng-Qiu. E-mail: dongmengqiu@nibs.ac.cn

HE Si-Min. E-mail: smhe@ict.ac.cn

Received: February 10, 2014 Accepted: May 26, 2014