

Proteinanalytik mittels Crosslinking-Massenspektrometrie

Synergistische Kombination von kovalenter Bindungsknüpfung und LC-MS

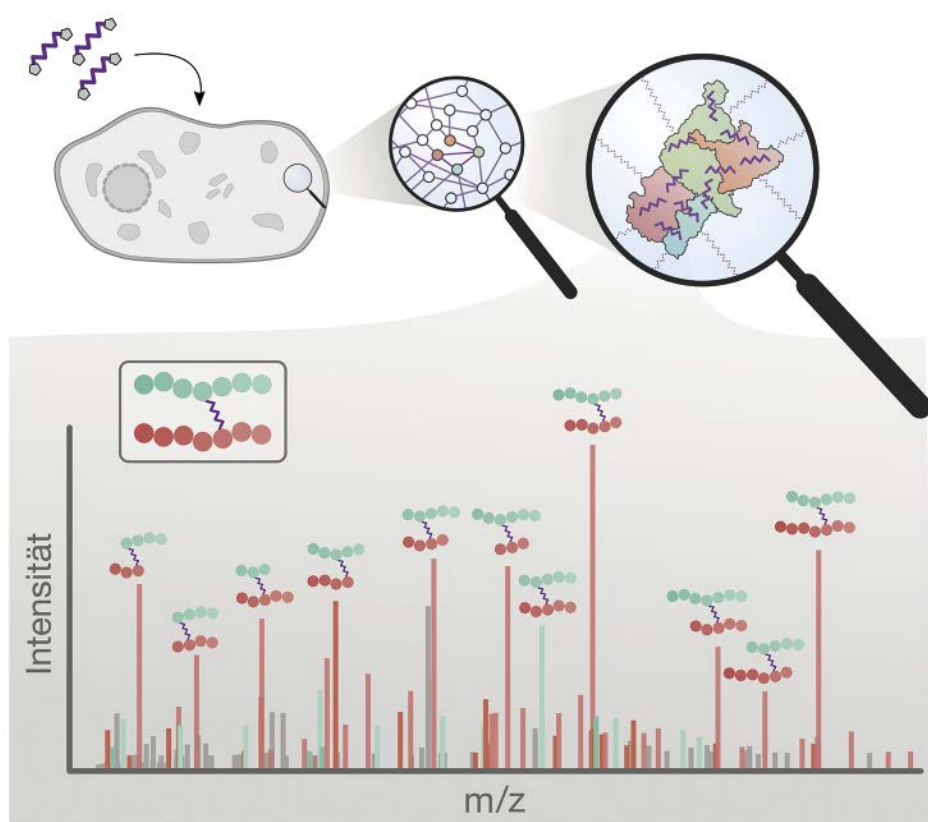
Ludwig R. Sinn¹ und Juri Rappsilber^{1,2}

Crosslinking-Massenspektrometrie (CLMS) ist eine Methodik zur Untersuchung von Struktur und Interaktionen von Proteinen. Dazu werden räumlich nahe Aminosäuren in und zwischen Proteinen erst quervernetzt und nach einem Proteinverdau mittels Massenspektrometrie identifiziert. Durchbrüche in Probenaufbereitung, analytischem Messaufbau und Datenprozessierung erweitern die Anwendbarkeit von CLMS als vielseitige Brückentechnologie zwischen Strukturbiologie und Systembiologie.

Erforschen von Proteinstrukturen auf zellulärer Ebene

Das Leben bedarf des Zusammenspiels zahlreicher Biomoleküle, von denen Proteine eine zentrale Rolle einnehmen, z.B. als Enzyme des Stoffwechsels. Die Funktion eines Proteins ist eng mit dessen individueller molekularer Struktur verbunden, weswegen die Erforschung von Proteinstrukturen elementar zum mechanistischen Verständnis vom Wirken der Proteine beiträgt. Dieses Vorhaben stellt jedoch hohe Ansprüche an Technologie und Arbeitszeit. Darüber hinaus agieren Proteine meist nicht isoliert, sondern im Verbund als Protein-Komplexe und auch größere Gruppen, die z.B. auf Störung des zellulären Gleichgewichts reagieren. Das Proteom (die Gesamtheit der Proteine einer Zelle/eines Gewebes) und dessen Dynamik wird im Forschungsfeld der Systembiologie untersucht.

Heutzutage können Tausende von Proteinen, mittels Flüssigchromatographie-Massenspektrometrie (LC-MS) identifiziert und quantifiziert werden. Während der vorgeschalteten Spaltung von Proteinen in leichter analysierbare Bruchstücke (Peptide) geht jedoch die dreidimensionale Proteinstruktur verloren. Um dennoch Information zu Proteinstrukturen zu erhalten, werden in der CLMS zuerst chemische Quervernetzungen in Proteine eingeführt, welche die anschließende enzymatische Proteinspaltung in Peptide überdauern und schließlich per LC-MS als verknüpfte Peptide beobachtet werden (Abb.1). Die so erlangte Information über Distanzbeziehungen zwischen zwei Punkten (hier Aminosäureresten) kann dann zur



Modellierung der originären Proteinstruktur bzw. von Protein-Protein-Interaktionsnetzwerken (PPI) verwendet werden.

Methodische Herausforderungen und deren schrittweise optimierte Lösung

Die Erfassung des Proteoms gelingt zwar zunehmend besser, wird aber durch dessen großen Dynamikumfang (bis zu sieben Größenordnungen in einer *H. sapiens*-Zelle) [1] erschwert. Proteinmodifikationen vergrößern diese Komplexität weiter – sowohl natürlich vorkommende, wie beispielsweise regulative Acetylierungen, aber auch als Teil der Analyse eingefügte, wie beispielsweise Quervernetzungen. Zudem liegen die einzelnen Quervernetzungen nur substöchiometrisch vor. Weder ist die Reaktion vollständig, noch eindeutig, weil jede beteiligte Aminosäure in der Regel mehr als eine reaktive Aminosäure als möglichen Partner hat,

noch würde man eine komplette Vernetzung der Proteine wünschen, da dies die Analyse erschweren würde. Die Folge ist, dass gerade seltene Proteine, bzw. seltene Proteinkonformationen in den Daten unterrepräsentiert sind, oder sogar unterhalb des Detektionslimits liegen. Um diese aus dem dynamischen Bereich der Probe folgenden Herausforderungen zu bewältigen, wird daran gearbeitet, die vielen einzelnen Prozessschritte weiter zu verbessern (Abb.1).

Im Nass-Labor

Die Wahl eines Crosslinker richtet sich nach gewünschter chemischer Reaktionsspezifität (z.B. N-Hydroxysuccinimidyl-Ester für NH_2 - und OH - bzw. Maleimide für SH -), aber auch weiteren Charakteristika, die etwa einer nachgestellten spezifischen Anreicherung oder auch verbesserten massenspektrometrischen Detektion (z.B. isotoopen-markiert oder

spaltbar) dienlich sind [2]. Nach Quervernetzung und enzymatischer Proteinspaltung müssen die modifizierten Peptide in der Regel aus dem von linearen Peptiden dominierten Gemisch angereichert werden [3]. Verschiedene molekulare Volumina oder auch elektrische Ladungszustände werden etwa während Größenausschluss- und Kationenaustauscher-Chromatographie ausgenutzt und auch ein spezifischer Affinitäts-tag im Crosslinker selbst, wie z.B. Biotin, hat sich hierfür als geeignet erwiesen. Spezifische Gasphasen-Ionenmobilitäten können ebenso zur Anreicherung genutzt werden [4] und für die Untersuchung sehr anspruchsvoller Stoffgemische werden oftmals auch mehrere Anreicherungs- und Fraktionierungstechnologien miteinander gekoppelt [5].

Die Messung & Datenverarbeitung

Während der eigentlichen Messung bieten z.B. eine für die Probe zugeschnittene chromatographische Trennung und auch die Verwendung hochauflösender Massenspektrometrie mit intelligenter Messlogik weiteres Optimierungspotenzial [6]. Die sich anschließende Datenprozessierung berücksichtigt die chimäre Natur eines quervernetzten Peptids, welches aus zwei unabhängig voneinander zu identifizierenden linearen Peptiden besteht [7]. Da die Anzahl an Möglichkeiten aus der Kombination zweier Peptide ungefähr mit quadratischer Abhängigkeit zur Datenbankgröße ansteigt, stellt dieser Schritt ein anspruchsvolles algorithmisches Problem dar, für welches mehrere Lösungen vorgeschlagen wurden. Eine hohe Massengenauigkeit und Sensitivität des LC-MS-Systems, sowie die Ausnutzung charakteristischer Fragment-Ionen im MS₂-

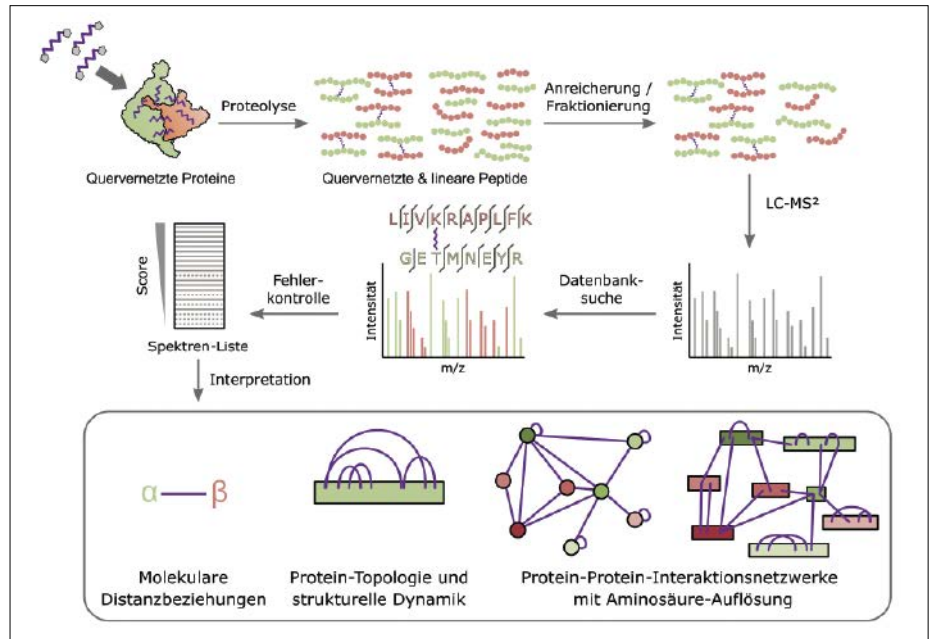


Abb. 1: Klassischer Workflow der Quervernetzungs-Massenspektrometrie. Gezeigt ist der typische Arbeitsablauf eines Quervernetzungs-Massenspektrometrie Experiments (siehe Text für Details). Die erhaltene Information in Form von molekularen Distanzbeziehungen kann sowohl zum Modellieren einzelner Proteine, als auch von Protein-Netzwerken zellulärer Komplexität verwendet werden.

Spektrum helfen hierbei. Außerdem bieten Proben-spezifische Informationen, wie etwa chromatographisches Verhalten der Analyten während der Anreicherung bzw. Fraktionierung, die Möglichkeit, weitere Konfidenz-Filter beispielsweise mittels maschinellen Lernens einzubauen [8].

Ein ganz zentrales Thema jeder Methode ist die Fehlerkontrolle. Diese wurde hier aufbauend auf etablierte Verfahren aus der Proteomik entwickelt und ergänzt, um Anforderungen vernetzter Peptide gerecht zu werden [5, 7]. Die Gültigkeit der Erwei-

terungen wurde kürzlich mittels theoretischer und alternativer experimenteller Kontrollen bewiesen, und so können nunmehr verlässlich PPIs mittels CLMS bestimmt werden.

Anwendungsfelder für die CLMS-Methodik

Der Einsatz von CLMS hat sich stetig weiterentwickelt (Abb. 2). In der Vergangenheit war deren Anwendung limitiert auf das Studium einzelner Proteinstrukturen

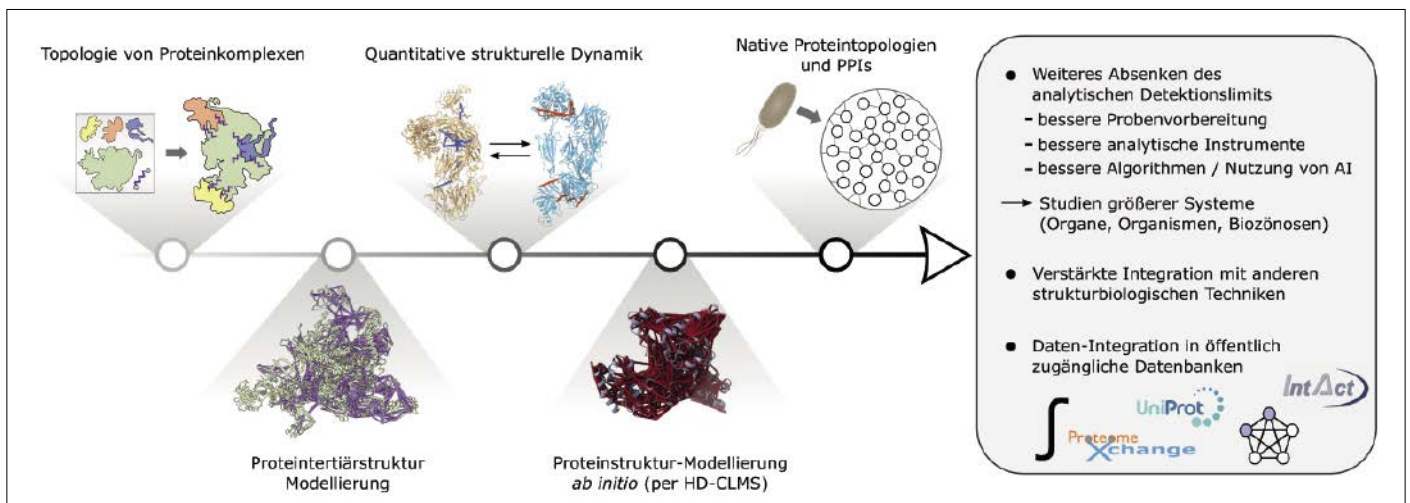


Abb. 2: Anwendungsbeispiele und Perspektive der Quervernetzungs-Massenspektrometrie. Beginnend mit der Topologie-Analyse gereinigter Proteinkomplexe entwickelte sich CLMS zu einer Methodik zum systematischen Studium von Struktur und Topologie des Proteoms mit weitreichendem Potenzial hin zu einer Plattformtechnologie.

und Proteinkomplextopologien, wenn gleich auflösungslimitiert durch die maximale Distanz zwischen reaktiven Gruppen des Crosslinkers (ca. 10–20 Ångström). Ein Durchbruch gelang mit der Verwendung von Licht-aktivierbarer Crosslink-Chemie, wie z.B. Diazirine, welche eine hohe Reaktionsgeschwindigkeit mit breiter chemischer Spezifität verbinden, um in der Folge mehr Datenpunkte zu erzeugen. Damit gelang es in einer Machbarkeitsstudie sogar eine Proteinstruktur vollständig basierend auf CLMS-Daten mit nur geringer Abweichung zur bekannten Kristallstruktur zu erstellen [9]. Weitere Forschungsanstrengungen widmen sich außerdem dem Einbau ähnlicher Crosslinker direkt in Proteine als Licht-aktivierbare Aminosäurederivate [10].

Insbesondere in Verbindung mit anderen strukturbiologischen Techniken, wie beispielsweise kryo-Elektronenmikroskopie, zeigt CLMS ihre Stärken. So gelang es, die Struktur des humanen Thyroglobulins zu lösen und damit den Mechanismus der Bildung lebensnotwendiger Schilddrüsenhormone aufzuklären [11]. Darüber hinaus ist CLMS geeignet, die Wissenslücken zwischen dem Agieren einzelner Proteine und deren Zusammenspiel im zellulären Kontext zu schließen [5, 12, 13]. So kann die Funktion eines bislang unerforschten Proteins via dessen Interaktion zu Proteinen mit bekannten Funktionen deduziert werden. Weiterhin können systemische Reaktionen des zellulären Proteoms auf externe Stimuli wie z.B. auf Medikamentengabe erforscht werden [14]. Ein eindrucksvolles Beispiel stellt die seit langem theoretisch angenommene Kopplung von Transkription und Translation in Bakterien dar, welche unter der Verwendung von kryo-Elektronentomographie und CLMS bestätigt werden konnte [15].

Die Zukunft der Quervernetzungs-Massenspektrometrie

Man kann die kontinuierliche Verbesserung der Methodik in all ihren Aspekten, der Crosslinking-Chemie, dem LC-MS-Instrumentarium und der Datenverarbeitung erwarten (Abb. 2).

Bei der Datenverarbeitung wird die sich stetig verbessernde Anwendung künstlicher Intelligenz eine optimierte Einbeziehung von proben- und messungsspezifischen Effekten, wie beispielsweise Wissen über die Probenvorbereitung und den biologischen Hintergrund, ermöglichen. Mittelfristig wird das gesamte Proteom, einschließlich seltener Proteine, in deren zellulärer Umgebung in *in situ* quervernetzter Form beobachtbar werden.

Parallel dazu ermöglicht die Licht-induzierbare Quervernetzungschemie mehr individuelle Datenpunkten zu Änderungen in Zusammensetzung und Struktur des Proteoms zu erhalten, sodass dann auch quantitative CLMS-Analysen in Proben zellulärer Komplexität [16] mit größerer analytischer Tiefe machbar werden.

Gleichzeitig muss aber auch an der bestmöglichen Verwendung der Daten gearbeitet werden. Hinsichtlich der Proteinstrukturmodellierung muss etwa das Vorkommen teils sehr verschiedener Konformere quantitativ berücksichtigt werden. Lediglich vom Crosslinker-modifizierte und nicht quervernetzte Peptide indizieren beispielsweise hierbei, inwieweit Aminosäurereste sterisch bzw. für Solvens zugänglich vorlagen.

Selbstverständlich sollten PPI aus CLMS-Experimenten zusätzlich auch anderweitig validiert werden. Hierbei eignen sich Antikörper-basierte Technologien, spezialisierte Assays und PPI-Datenbanken, insbesondere aber auch computergestützte Vorhersagen [17]. Diesen Trend begleitend bemüht sich das CLMS-Feld zuletzt auch um eine transparente und zugängliche Datenverwaltung, hin zu einer Nutzung durch das globale Forschungskollektiv [18]. All diese Entwicklungen werden schließlich die Zusammenführung von Strukturbiologie und Systembiologie bewirken, hin zu einem vollständigeren Bild über die molekulare Organisation des Lebens.

Zugehörigkeiten

¹Technische Universität Berlin, Institut für Biotechnologie, Fachgebiet Bioanalytik, Deutschland

²Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, United Kingdom

● KONTAKT |

Prof. Dr. Juri Rappsilber

Institut für Biotechnologie, Fachgebiet Bioanalytik
Technische Universität Berlin
Berlin, Deutschland
juri.rappsilber@tu-berlin.de



Weitere Beiträge zum Thema:
<https://bit.ly/WAS-Proteomik>



Literatur:
<https://bit.ly/GIT-Sinn>