



Doctoral Thesis

Structural characterization of macromolecular protein assemblies by protein cross-linking and mass spectrometry

Author(s):

Walzthöni, Thomas

Publication Date:

2013

Permanent Link:

<https://doi.org/10.3929/ethz-a-009759654> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 20901

**STRUCTURAL CHARACTERIZATION OF MACROMOLECULAR
PROTEIN ASSEMBLIES BY PROTEIN CROSS-LINKING AND MASS
SPECTROMETRY**

A dissertation submitted to the

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

THOMAS WALZTHÖNI

Mag. Mag. Biol.

Leopold-Franzens-Universität Innsbruck

born February 9, 1983

Austrian citizen

accepted on the recommendation of

Prof. Dr. Ruedi Aebersold, examiner

Prof. Dr. Nenad Ban, co-examiner

Dr. Martin Beck, co-examiner

2013

Abstract

Virtually all biological processes are controlled and catalyzed by proteins which are, in many cases, in complexes with other proteins and form macromolecular protein assemblies. Therefore, understanding the architecture and structure of protein complexes is critical to understanding their biological role and function.

Traditionally, high-resolution data for structural analysis of proteins or protein complexes has been generated by the powerful methods of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. While these methods are enormously powerful and allow to generate models at atomic resolution, macromolecular assemblies are often refractory to be studied by these methods. More recently, mass spectrometry (MS) based methods have been developed that provide low-resolution structural information of proteins or protein complexes. One of these methods is chemical cross-linking in combination with mass spectrometry (CX-MS). Thereby protein complexes are chemically cross-linked with a cross-linking reagent, subsequently digested to peptides and analyzed by tandem mass spectrometry. Sequencing of the cross-linked peptides allows to identify the residues which are indicative for their spatial proximity in the native state of the protein complex. This structural information can then be used by integrative modeling approaches to generate structural models of the underlying protein complexes.

This thesis describes the development of algorithms for the automated analysis of MS data obtained from CX-MS experiments and the application of the technology to large protein complexes and a modular protein interaction network.

The first part of the thesis describes the development of the algorithms xQuest and xProphet. xQuest was developed to identify cross-linked peptides from fragment ion spectra of CX-MS experiments. The algorithm makes use of isotopically labeled cross-linkers, which greatly facilitates the identification process and allows the identification of cross-linked peptides from large sequence databases. Subsequently the development of xProphet is described, a software to estimate false discovery rates (FDRs) for cross-linked peptides. The method is based on a target-decoy strategy, which takes the specific properties of data from CX-MS experiments into account. Together the development of xQuest and xProphet represents a fully automated software pipeline for the routine analysis of cross-linked peptides identified by MS.

The second part of the thesis represents the application of the CX-MS technology to study large protein complexes. These include the TRiC/CCT chaperonin, which is an essential chaperone in eukaryotes that assists folding of up to 10% of the proteome, for which the definitive topology of the complex could not be deduced so far. The application of CX-MS and the combination with combinatorial structural modeling approaches led to the unambiguous assignment of the complex

topology. A second large protein complex that was studied is the 26S proteasome. Parts of the structure of the complex were elusive to high resolution methods so far, especially the 19S regulatory particle. The application of CX-MS and the combination with additional structural data allowed the generation of a structural model of the 26S proteasome by the use of an integrative modeling approach.

The third part describes the application of the CX-MS technology to affinity purified protein complexes that are part of a modular protein network. Thereby the method was applied to systematically probe the structure of protein complexes of the human protein phosphatase 2 A network.

Finally, the fourth part of the thesis describes an integrated step-by-step protocol of the CX-MS technology for the structural analysis of protein complexes.

In summary, during my PhD I developed the algorithms that are necessary for the automated identification and statistical validation of data from CX-MS experiments. The potential of the technology has been shown by revealing the molecular architectures of large protein complexes and the structural characterization of a modular protein network. Therefore, I expect the CX-MS technology to become an integral part for hybrid structural biology approaches.

Zusammenfassung

Nahezu alle biologischen Prozesse werden von Proteinen kontrolliert und katalysiert. Oftmals führen Proteine ihre Funktion als Multiproteinkomplexe aus. Aus diesem Grund ist es wichtig, die Architektur und Struktur von Multiproteinkomplexen zu bestimmen, um deren biologische Funktion zu verstehen.

Die traditionellen Methoden zur Strukturanalyse von Proteinen und Proteinkomplexen sind die Röntgenstrukturanalyse und Kernspinresonanzspektroskopie (NMR-Spektroskopie). Während diese Methoden sehr leistungsfähig sind und die Ermittlung der atomaren Struktur erlauben, stellt sich die Untersuchung von Multiproteinkomplexen mit diesen Methoden oft als sehr schwierig heraus. Vor kurzem wurden neue, auf Massenspektrometrie basierende Methoden entwickelt, die zur Untersuchung der Struktur von Multiproteinkomplexen eingesetzt werden können.

Eine dieser Methoden ist die Kombination der chemischen Quervernetzung von Proteinen und Massenspektrometrie (CX-MS). Dabei werden Proteine oder Multiproteinkomplexe durch ein chemisches Reagens (Crosslinker) quervernetzt, anschliessend enzymatisch verdaut und die daraus resultierenden Peptide massenspektrometrisch analysiert. Durch die Sequenzierung der vernetzten Peptide können diejenigen Aminosäuren identifiziert werden, die sich in der nativen Proteinstruktur in räumlicher Nähe befinden. Diese Strukturinformation kann anschliessend mittels integrativen Strukturmodellierungsmethoden zur Generation von Strukturmodellen der untersuchten Multiproteinkomplexe verwendet werden.

Der erste Teil der Arbeit stellt die Entwicklung der Algorithmen xQuest und xProphet vor. xQuest wurde entwickelt, um quervernetzte (gecrosslinkte) Peptide anhand von Fragmentationsspektren von CX-MS Experimenten zu identifizieren. Der Algorithmus verwendet die Information von isotoptenmarkierten Crosslinkern, die den Identifizierungsprozess erleichtern. Dies ermöglichte die Identifizierung von gecrosslinkten Peptiden aus komplexen Proben und die Verwendung von umfangreichen Sequenzdatenbanken. Anschliessend wird die Entwicklung von xProphet beschrieben, eine Methode zur Abschätzung von False Discovery Raten von gecrosslinkten Peptiden. Diese Arbeit stellt einen „Target-Decoy“ Ansatz vor, der die speziellen Anforderungen der Daten von CX-MS Experimenten berücksichtigt. Insgesamt stellt die Entwicklung der Algorithmen xQuest und xProphet und deren Implementierung eine voll automatisierte Softwareplattform zur Routineanalyse von gecrosslinkten Peptiden dar.

Der zweite Teil der Arbeit beschreibt die Anwendung der CX-MS Technologie zur Untersuchung von grossen Proteinkomplexen. Dies umfasst den Chaperonkomplex TRiC/CCT, ein essentieller Multiproteinkomplex in Eukaryonten, der die Faltung von bis zu 10% aller Proteine vorantreibt. Die Topologie dieses Multiproteinkomplexes konnte bisher nicht eindeutig bestimmt werden. Mithilfe

der CX-MS Technologie in Kombination mit kombinatorischen Modellierungsansätzen konnte die Topologie des Proteinkomplexes nun eindeutig bestimmt werden.

Des Weiteren wurde die Methode auf das 26S Proteasom angewendet. Teile der Struktur waren bisher für die klassischen Methoden der Strukturbiologie, im Speziellen das 19S regulatorische Partikel, nicht zugänglich. Die Anwendung der CX-MS Methode in Kombination mit zusätzlicher struktureller Information erlaubte die Generierung eines Strukturmodelles des 26S Proteasoms mithilfe eines integrativen Modellierungsansatzes.

Der dritte Teil der Arbeit beschreibt die Anwendung der CX-MS Technologie in Kombination mit der Affinitätsaufreinigung von Proteinkomplexen. Dabei wurde die CX-MS Methode angewendet, um die Proteinkomplexe des modularen Proteinnetzwerkes der humanen Protein-Phosphatase 2A strukturell zu charakterisieren.

Abschließend wird im vierten Teil der Arbeit ein durchgängiges Schritt-für-Schritt Protokoll der CX-MS Technologie zur strukturellen Analyse von Proteinkomplexen beschrieben.

Im Rahmen dieser Forschungsarbeit wurden Algorithmen und Softwareprogramme entwickelt, welche die automatisierte Identifizierung und statistische Validierung der Daten von CX-MS Experimenten ermöglichen. Das technologische Potenzial der CX-MS Methode wurde aufgezeigt, indem es gelang, grosse Multiproteinkomplexe und ein modulares Proteinnetzwerk strukturell zu charakterisieren. Daher bin ich davon überzeugt, dass die CX-MS-Technologie künftig einen wesentlichen Bestandteil für hybride strukturbiologische Ansätze darstellen wird.