Top-down and middle-down mass spectrometry structural characterization of monoclonal antibodies

THÈSE N^O 6052 (2014)

PRÉSENTÉE LE 24 JANVIER 2014
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE SPECTROMÉTRIE DE MASSE DE BIOMOLÉCULES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Luca FORNELLI

acceptée sur proposition du jury:

Prof. H. Girault, président du jury
Prof. Y. Tsybin, directeur de thèse
Dr M. Kussmann, rapporteur
Dr C. Masselon, rapporteur
Prof. R. Zubarev, rapporteur

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2014

Table of contents

Abstract.

Top-down (TD) mass spectrometry (MS) is a protein structural analysis technique consisting in mass determination of intact proteins followed by their gas-phase fragmentation and mass analysis of structure-specific product ions. Applied to large scale studies, i.e., proteomics, this is the only approach that can provide identification of proteoforms, i.e., genetic products with specific chemical or genetic modifications. From this comes the great impulse to advance TD MS. High instrumental requirements in terms of mass accuracy, mass resolution and signal-to-noise ratio, in addition to difficulties in achieving efficient fragmentation of intact proteins and interpreting the resulting convoluted tandem MS (MS/MS) spectra have so far limited TD MS analysis primarily to small, ~20 kDa, proteins. The present work represents the attempt to structurally investigate progressively more complex peptides, proteins, and their mixtures, finally reaching detailed characterization of intact monoclonal antibodies. These, also known as immunoglobulins G (IgGs), are ~150 kDa homoheterodimers composed of two identical light (~25 kDa) and two identical heavy (~50 kDa) chains. We first examined the potential of the state-of-the-art radical-driven MS/MS techniques, namely electron capture dissociation (ECD) and electron transfer dissociation (ETD), as well as of high mass accuracy and resolution measurements at the peptide level. Particularly, ECD demonstrated its utility in combination with Fourier transform MS (FTMS) to decipher the sophisticated pool of modified products derived from the catalytic activity of the enzyme tissue transglutaminase on its peptidic substrate. High mass accuracy was also at the base of an elegant label-free quantitation approach applied to the analysis of glycopeptides derived from stably and transiently produced IgGs. Afterwards, an ETD-based top-down MS investigation of proteins from 30 kDa (carbonic anhydrase) to ~79 kDa (serotransferrin) was performed. For the first time a protein of this size was characterized with ETD MS/MS. The capability of this ion activation technique to cleave disulfide bridges was confirmed for large proteins, after having being reported for peptides and smaller proteins. Finally, top-down analysis of monoclonal IgGs was accomplished using two distinct high-resolution, ETD-enabled MS platforms, equipped with either time-of-flight (TOF) or Orbitrap mass analyzers. Remarkably, proteins of this size have been previously analyzed in a top-down fashion only in ion cyclotron resonance mass analyzers. ETD led to reach up to 32% sequence coverage for an immunoglobulin G1, and several IgGs belonging to different organisms or subclasses were analyzed. Notably, we also identified some conformational and structural constrains that are at the base of the specific fragmentation patterns observed for the IgGs. The sequence coverage of IgGs was then improved by the application of a dedicated middle-down MS approach, taking advantage of a robust and quick novel proteolytic protocol, and of all the technical advances developed for TD MS, including fine-tuning of ETD parameters, product ion transmission, and data analysis. The demonstration of capabilities of TOF and Orbitrap mass analyzers for TD MS of large proteins has opened the new horizons for development and application of the top-down approach. Particularly, the achieved results should be useful to advance in the near future the possibility of studying immunoglobulin mixtures from different biological sources in a proteomic fashion, potentially improving the efficiency of the drug discovery process.

Keywords: mass spectrometry, MS; tandem mass spectrometry, MS/MS; top-down MS, TD MS; middle-down MS, MD MS; electron capture dissociation, ECD; electron transfer dissociation, ETD; time-of-flight, TOF; Fourier transform mass spectrometry, FTMS; Orbitrap; ion cyclotron resonance, ICR; protein; peptide; monoclonal antibody, mAb; immunoglobulin G, IgG; proteolysis.

Abstract in italiano.

La spettrometria di massa (MS) top-down (TD) è una technica consistente nella determinazione della massa di proteine intatte seguita dalla loro frammentazione in fase gassosa e analisi di massa di ioni frammento struttura-specifici. Applicato a studi su ampia scala, ossia proteomica, questo è l'unico approccio che puó fornire l'identificazione di proteoforme, ovvero prodotti genici con specifiche modifiche chimiche o genetiche. Da ció deriva il grande impulso per migliorare la TD MS. Elevati requisiti strumentali in termini di accuratezza di massa, risoluzione di massa e rapporto segnale-rumore, in aggiunta a difficoltà nell'ottenere una frammentazione efficiente di proteine intere e nell'interpretazione dei complessi spettri di MS tandem (MS/MS) risultanti hanno finora limitato l'analisi TD MS prevalentemente a proteine piccole, di ~20 kDa. Il presente lavoro rappresenta il tentativo di investigare strutturalmente peptidi e proteine progressivamente piú complessi, fino a giungere alla caratterizzazione dettagliata di anticorpi monoclonali interi. Questi, anche noti come immunoglobuline G (IgG), sono omoeterodimeri composti di due identiche catene leggere (~25 kDa)e due identiche pesanti (~50 kDa). Pertanto, abbiamo da prima esaminato il potenziale dello stato dell'arte di tecniche MS tandem guidate da radicali, ovvero la dissociazione per cattura elettronica (ECD) e la dissociazione per trasferimento elettronico (ETD), nonchè di misure ad alta accuratezza di massa e risoluzione al livello dei peptidi. Specificamente, l'ECD ha dimostrato la sua utilità in combinazione con la spettrometria di mass a trasformata di Fourier (FTMS) nel decifrare il sofisticato gruppo di prodotti modificati derivati dall'azione catalitica dell'enzima transglutaminasi tissutale sul suo substrato peptidico. L'alta accuratezza di massa è stata invece alla base di un elegante approccio di quantificazione label-free applicato all'analisi di glicopeptidi derivati da IgG. Di seguito è stata svolta un'investigazione top-down MS basata su ETD di proteine da 30 kDa (anidrasi carbonica) ai ~79 kDa della serotransferrina. Per la prima volta una proteina di queste dimensioni è stata caratterizzata tramite ETD, e in questa occasione è stata provata in modo definitivo la capacità di questa tecnicadi attivazione ionica, prima riportata per peptidi corti, di rompere ponti disolfuro. Infine, l'analisi top-down di IgG monoclonali è stata compiuta usando due distinte piattaforme MS ad alta risoluzione e con funzione ETD, equipaggiate o con analizzatore di massa a tempo di volo od Orbitrap. E' da sottolineare che proteine di queste dimensioni erano state in precedenza analizzate in top-down solo attraverso analizzatori di massa a risonanza ionica ciclotronica. L'ETD ha portato a raggiungere fino al 32% di copertura di sequenza per un'immunoglobulina G1, e diverse IgG appartenenti a diversi organismi o sottoclassi sono state analizzate. Da segnalare come siano stati anche evidenziate alcune limitazioni conformazionali e strutturali alla base del peculiare pattern di frammentazione che si osserva per le IgG. In seguito la copertura di sequenza delle IgG è stata estesa grazie all'adozione di un un approccio middle-down dedicato, che si avvantaggia di un robusto e veloce protocollo di proteolisi, e degli avanzamenti tecnici sviluppati per la TD MS, includenti l'ottimizzazione fine dei parametri di ETD, trasmissione degli ioni prodotto e analisi dei dati. La dimostrazione delle capacità di analizzatori di massa TOF e Orbitrap per TD MS di grandi proteine ha aperto nuovi orizzonti per lo sviluppo e l'applicazione dell'approccio top-down. In particolare, i risultati raggiunti dovrebbero essere utili ad aprire nel futuro prossimo la possibilità di studiare mix di immunoglobuline provenienti da differenti fonti biologiche in uno stile proteomico, migliorando potenzialmente l'efficienza del processo di scoperta di farmaci.

Parole chiave: spettrometria di massa, MS; spettrometria di massa tandem, MS/MS; top-down MS, TD MS; middle-down MS, MD MS; dissociazione per cattura elettronica, ECD; dissociazione per trasferimento elettronico, ETD; tempo di volo, TOF; spettrometria di massa a trasformata di Fourier,

FTMS; Orbitrap; risonanza ionica ciclotronica, ICR; proteina; peptide; anticorpo monoclonale, mAb; immunoglobulina G, IgG; proteolisi.

List of papers.

The present Thesis is based on the following research articles (* indicates equal contribution):

I. **Fornelli L***, Schmid AW*, Grasso L, Vogel H, Tsybin YO. Deamidation and transamidation of substance P by tissue transglutaminase revealed by electron-capture dissociation Fourier transform mass spectrometry. *Chemistry – A European Journal* (2011) 17(2), 486-497

II. Nallet S*, **Fornelli L***, Schmitt S, Parra J, Baldi L, Tsybin YO, Wurm FM. Glycan variability on a recombinant IgG antibody tra nsiently produced in HEK-293E cells. *New Biotechnology* (2012) 29(4), 471-47

III. **Fornelli L**, Parra J, Hartmer R, Stoermer C, Luebeck M, Tsybin YO. Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry, *Analytical Bioanalytical Chemistry*, (2013) 405(26), 8505-8514

IV. Tsybin YO, **Fornelli L**, Stoermer C, Luebeck M, Parra J, Nallet S, Wurm FM, Hartmer R. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Analytical Chemistry* (2011) 83(23), 8919-8927

V. **Fornelli L**, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Molecular and Cellular Proteomics* (2012) 11(12), 1758-1767

VI. **Fornelli L**, Ayoub D, Aizikov K, Liu X, Damoc E, Pevzner PA, Makarov A, Beck A, Tsybin YO. Top-down analysis of immunoglobulins G with electron transfer dissociation high-field Orbitrap FTMS, *manuscript in preparation*

VII. **Fornelli L***, Ayoub D*, Aizikov K, Beck A, Makarov A, Tsybin YO. Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap FTMS, *submitted*

My personal contribution to these papers is the following: I have performed the entire FTMS experiments for papers I, II, VI and VII. Data acquisition for the remaining papers was performed with the assistance of Prof. Yury O. Tsybin at either Thermo Scientific or Bruker Daltonics GmbH, Bremen, Germany. I personally carried out most of sample preparation and data analysis for all the papers, except for sample preparation of papers II and III and data analysis of papers IV and V for which we received important support from collaborators (listed as co-authors). Finally, I was involved in the discussion of all the research projects and actively contributed to the writing and correction of all manuscripts.

Papers (research articles and reviews) not included in the discussion:

1.   Tsybin YO, **Fornelli L**, Kozhinov AN, Vorobyev A, Miladinovic SM. High-resolution and tandem mass spectrometry--the indispensable tools of the XXI century. *Chimia* (2011) 65(9), 641-645

2.   Miladinović SM, **Fornelli L**, Lu Y, Piech KM, Girault HH, Tsybin YO. In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Analytical Chemistry* (2012) 84(11), 4647-4651

3.   Gorshkov MV, **Fornelli L**, Tsybin YO. Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. *Rapid Communications in Mass Spectrometry* (2012) 26(15), 1711-1717

4.   Zhurov KO, **Fornelli L**, Wodrich MD, Laskay ÜA, Tsybin YO. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews* (2013) 42(12), 5014-5030

5.   Laskay ÜA, Srzentić K, **Fornelli L**, Upir O, Kozhinov AN, Monod M, Tsybin YO. Practical considerations for improving the productivity of mass spectrometry-based proteomics. *Chimia,* (2013) 67(4), 244-249

6.   Zhurov KO, Kozhinov AN, **Fornelli L**, Tsybin YO. Distinguishing analyte from noise components in mass spectra of complex samples: where to cut the noise?, *submitted*

List of abbreviations

BU	Bottom-up
CID	Collision-induced dissociation
ECD	Electron capture dissociation
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FT	Fourier transform
FTMS	Fourier transform mass spectrometry
FWHM	Full width at half maximum
ICR	Ion cyclotron resonance
Ig	Immunoglobulin
IgG	Immunoglobulin G
IR	Infrared
IRMPD	Infrared multiphoton dissociation
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
LTQ	Linear trap quadrupole
MALDI	Matrix-assisted laser desorption ionization
MD	Middle-down
mAb	monoclonal antibody
MS	Mass spectrometry
MS/MS or MS^2	Tandem mass spectrometry
PIA	Product ion abundance
PTM	Post-translational modification
RF	Radio frequency
SNR	Signal-to-noise ratio
TD	Top-down
TOF	Time-of-flight
UPLC	Ultra-high pressure liquid chromatography

Chapter 1: Introduction

The term "intelligent design" (ID) was introduced in its modern acceptation in 1987 by the Supreme Court of the United States, which sentenced that the teaching of this specific form of creationism is unconstitutional in American public schools. Although absolutely anti-scientific, ID has the merit of efficaciously illustrating the high complexity of biological systems that scientists have to investigate. The ID movement tried indeed to give explanation of this complexity as the necessary product of a superior mind (possibly, of God). Scientists, to whom this Thesis is addressed and will probably not be interested in discussing theological topics, will be convinced of the absolute level of sophistication reached by cells of superior organisms by numbers: in April 2003, with the completion of the sequencing of human genome, the Human Genome Project officially declared that ~20'500 genes are encoded in our DNA.[1] Genes are not, though, the effectors in molecular mechanisms governing cells. These are represented by gene products known as *proteins*.[2] From the chemical point of view, proteins are large polymers of *amino acids*, whose structure is differentiated by the so-called side chain. Although amino acid could exist in a variety of different stereo configurations and could potentially carry any chemical group at the side chain, insomuch that more than 500 have been classified so far, only 22 of these molecules are known as proteinogenic amino acids (i.e., effective building blocks of proteins), 21 of which are encoded in eukaryotes whilst only 20 are directly encoded by the genes. Each amino acid is represented in the DNA by a *codon*, which is a sequence of three nucleotides.[3] The molecular steps required in a cell for gene expression are essentially two (Figure 1.1): *transcription* and *translation*. During transcription, the double-stranded DNA helix is temporarily opened so that one of the strands is accessible to the molecular machinery responsible for the synthesis of a molecule of RNA, named messenger RNA or mRNA, complementary to the DNA filament used as template. The central enzyme involved in this process is the RNA polymerase, which copies with high fidelity the gene for its whole length, including both non-coding DNA regions, or introns, as well as coding ones, or exons. The product of RNA polymerase catalytic activity is a pre-mRNA which has to undergo maturation, a multistep process including splicing, which serves to remove introns and leave in the RNA uniquely the coding portion of the gene sequence.[4]

After maturation, the codons present in mRNAs are converted by ribosomes, large ribonucleic protein assemblies, into specific sequences of amino acids.
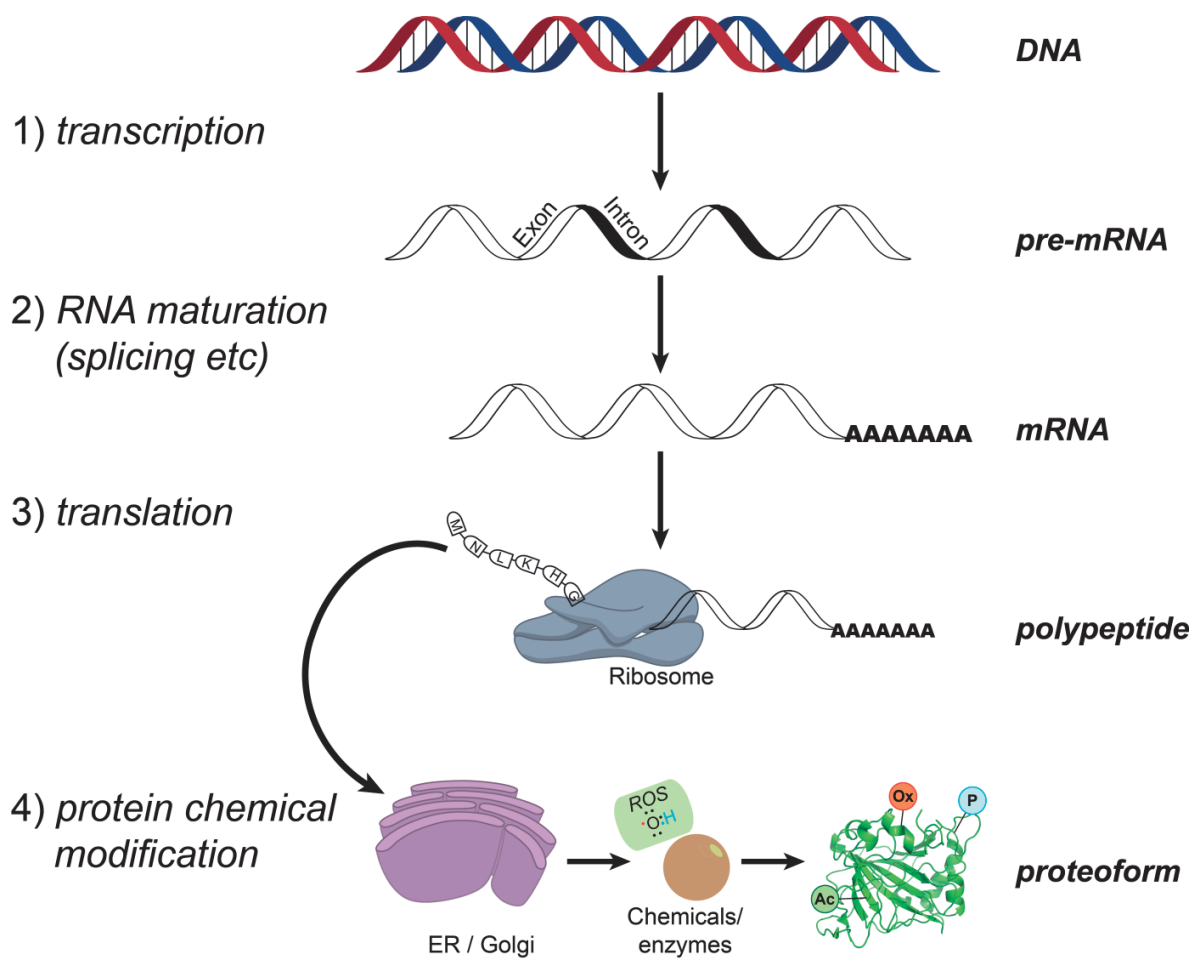
Figure 1.1. Gene expression pathway. Several processes are required for passing from DNA, protein blueprint, to the final proteoform. Modifications to the gene product can occur either at the DNA level or as post-transcriptional events.

Described in these terms, the complexity of genes and their products would seem identical. This assumption was also at the base of the famous hypothesis, formulated in 1941 by Beadle and Tatum, known as "one gene – one protein". The two scientists speculated that each step in a metabolic pathway had to be regulated by an enzyme codified by a specific gene. Implicit in this hypothesis was the concept that one gene could produce only a single enzyme. The inaccuracy of Beadle and Tatum's hypothesis, though, became rapidly apparent. Currently, the consensus view states that a single gene encoding for a polypeptide can produce more than a product. Variation sources are classified, according to IUPAC, at two different levels: DNA mutation and post-

transcriptional events. The former level includes genetic variations as polymorphisms (e.g., *single nucleotide polymorphism*, SNP), which can induce the substitution of the original codon with a different one, possibly encoding for a different amino acid or causing protein truncation (as specific sequences of three nucleotides, called stop-codons, determine the termination of polypeptide synthesis). The latter is a more heterogeneous class of modifications, comprising alternative mRNA splicing events as well as chemical modifications (which can be induced by enzymes or chemical agents) called post-translational modifications (PTMs). Gene products subjected to the first class of modifications are named *isoforms*, whereas they are referred to as *protein species* when including modifications of the second kind. With a newly introduced terminology, each combination of modifications on a gene product, independently from the origin of the modifications themselves, can be classified generally as a *proteoform*.[5]

The importance of proteoforms can be deduced taking into consideration, for instance, that mutations in protein sequence are often responsible for abnormal folding (i.e., the process through which the protein reaches its three-dimensional conformation) and are therefore implicated in the onset of diseases, and PTMs are often used by cellular mechanisms to activate or de-activate specific enzymes.

The set of modifications which can be applied to a gene product cannot be inferred by the analysis of neither the genome, which includes all the genes present in a certain organism, nor the transcriptome, which is a dynamic pool of active genes transcribed into RNA molecules. Hence, the need for the direct analysis of proteins seems apparent and inevitable. The first methods for protein sequencing, such as Edman degradation,[6] were laborious and could not easily complete the sequencing of very long polypeptides, and furthermore were limited to the analysis of single, purified proteins. Similar issues are common to all the biochemical protein characterization techniques.

To improve qualitatively as well as quantitatively the information about the proteins to investigate, analytical techniques were applied to protein characterization. The technique of choice for the analysis of gene products and, particularly, of proteoforms is currently represented by *mass spectrometry*.

1.1 Mass spectrometry for proteins: what approach?

Mass spectrometry (MS) is defined as an analytical method aimed at determining the mass-to-charge ratio (*m/z*) of charged particles, or ions. A typical MS experiment, performed using a *mass spectrometer*, consists of three parts: ionization, separation and detection. The ionization process ensures the indispensable acquisition of a charge by the analyte and, if this was originally in the solid or liquid phase, its passage to the gas phase. Subsequently, analyte ions can be manipulated by the application of electric and/or magnetic fields, in order to separate them from ions with other *m/z* values or temporarily confine them in specific areas of the spectrometer for several purposes. Finally, ions are detected, so that ion abundances are associated to ions' mass-to-charge ratios.[7]

The above described procedure determines the *m/z* of analyte ions and, therefore, leads to the calculation of the mass of the analyte as a neutral molecule. According to the contemporary terminology, the sequence of these events is known as *MS^1* or *survey scan*. For the characterization of biomolecules, and particularly of those, such as nucleic acids and polypeptides, which are polymers built from multiple monomers with different chemical formula, the accurately measured mass is theoretically sufficient to infer the monomer composition, but not their sequence. Therefore, after MS^1 an additional step is performed, consisting in the gas phase fragmentation of a selected ion type (*precursor ion*). This methodology, which produces charged fragments of the precursor called *product ions*, is referred to as *tandem mass spectrometry* (*MS^2* or *MS/MS*). Further selection and fragmentation of product ions are known as MS^n: for instance, the fragmentation of a product ion obtained by MS^2 is termed MS^3.

Through the years, mass spectrometry has developed so that from the very first application by J.J. Thomson, which revealed the existence of isotopes for non-radioactive elements, it could be applied to the analysis of small organic molecules and, currently, to large protein complexes. Nowadays, biomolecular mass spectrometry is used to address very sophisticated biological questions.

When describing the different approaches which have been developed for mass spectrometry-based protein analysis, one should always consider that for satisfying the requests coming from the world of life sciences, the goal rapidly passed from the characterization of a single protein (i.e., *targeted* analysis) to that of a protein mixture of various complexity, up to the *proteome*

level. According to its original definition, a proteome is defined as the set of all the proteins expressed by a cell in a defined moment. Let us assume, using arbitrary underestimation, that a human cell expresses only one fifth of its genes. It then follows that, with the products on average differentiated into two proteoforms, the proteome will be composed of at least 8'000 proteoforms. On the other hand, if we consider the *transcriptome* (i.e., the equivalent of the proteome for RNAs) of human cell lines, a significantly larger proteome would be expected, with roughly one order of magnitude more proteoforms.[8]

To reduce this extreme degree of complexity, ensuring efficient protein analysis, MS is often coupled on-line or off-line with protein purification, fractionation and separation techniques, so that in the unit of time only a limited number of analytes is ionized and subjected to MS processing. The most popular separation technique is liquid chromatography (LC) and, specifically, reversed phase LC which can be performed in a fully MS-compatible fashion.

Two main philosophies, ultimately ending in three possible approaches, can be identified in MS-based protein analysis, Figure 1.2. The first one relies on proteolysis, a chemical or enzymatic cleavage of proteins into shorter polypeptidic chains known as *peptides*. Depending of the method applied for the generation of peptides, these can have different lengths. When proteolytic peptides are smaller than 3 kDa on average, like in the case of the widely employed digestion with the protease trypsin, the analysis is named *bottom-up* (BU) MS (or bottom-up proteomics, whether applied to an entire proteome).[9, 10] If peptides are produced by the cleavage of rarely expressed amino acids or combinations of amino acid residues, like for the chemical cleavage at methionine with cyanogen bromide, their average mass exceeds the 3 kDa limit and can normally vary from ~5 to 15 kDa. This approach is called *middle-down* (MD) MS.[11, 12] Proteolytic peptides are subjected to MS and MS/MS analysis, with resulting mass spectra of reduced complexity.

Conversely, the opposite philosophy is based on the analysis of intact gene products. This methodology, known as *top-down* (TD) MS, includes the same MS and MS/MS steps used for smaller peptides.[13-15]
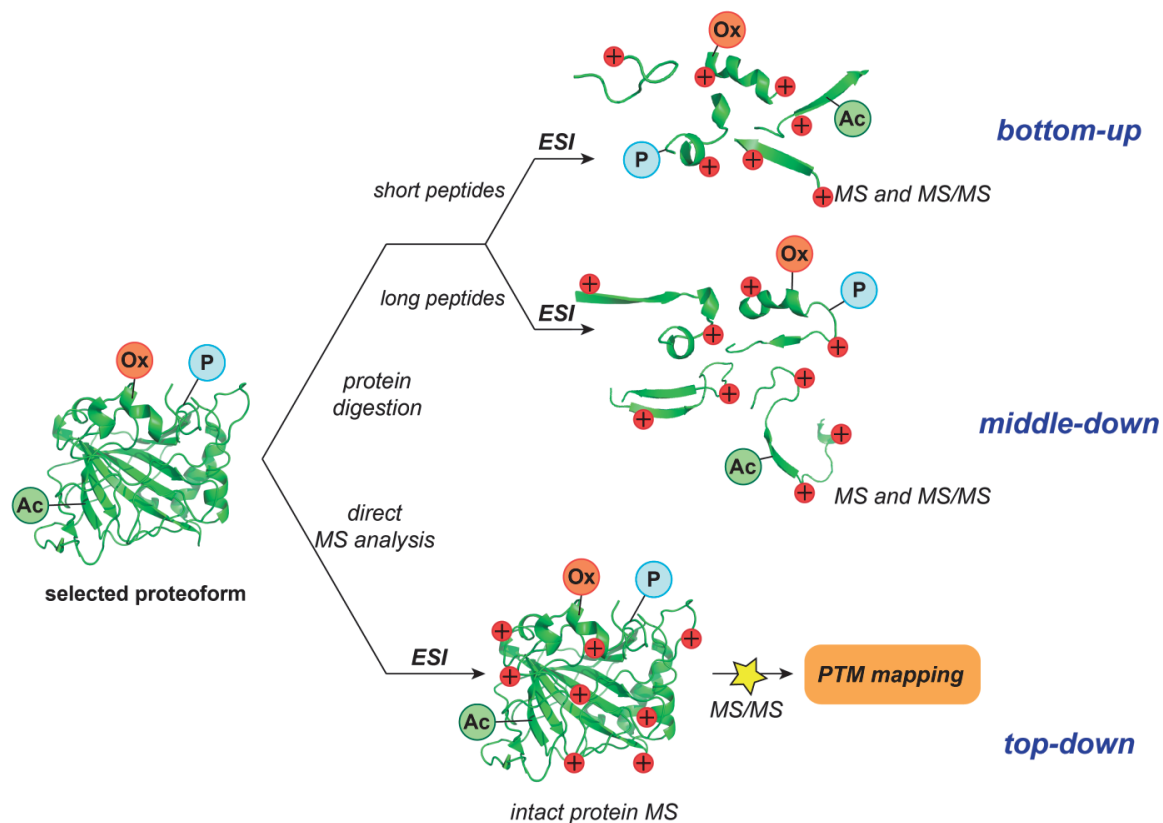
Figure 1.2. Schematics of proteolysis-based approaches for protein analysis, namely bottom-up and middle-down MS, and intact protein analysis or top-down MS.

Operations prior to ionization, parameters of MS analysis and data interpretation are significantly different among the three methods. Proteolytic peptides produced by bottom-up and middle-down create complex mixtures but, on the other hand, these relatively short polypeptides are easy to separate, ionize, and produce relatively simple mass spectra dominated by a limited number of ions with similar charge state (i.e., the number of charges of a single ion); moreover, their MS/MS results in relatively easily interpretable mass spectra composed of a reduced population of product ions. Conversely, proteins are more difficult to separate, ionize and both their MS[1] and MS/MS spectra are complicated by the larger size of these analytes. In MS[1], proteins show a wide charge state distribution; their MS/MS spectra are often convoluted, and product ion assignment is complicated not only by their high number but also by their high charge state and potential overlaps of ions with very similar *m/z* values.

Besides the technical differences, which will be discussed more in detail in the following chapters, the most relevant discriminant between the two approaches is conceptual, and precisely connected to the final outcome of the analysis. Most of the currently used protein databases are gene-centered, as they include protein entries obtained by *in silico* translation of genes sequenced and annotated for a certain organism of choice. Bottom-up and middle-down MS can directly determine the presence of specific peptides, which can be assigned to a related protein (or, eventually, to all those which share that polypeptidic sequence) which is considered as positively identified through statistical validation and, possibly, assignment of multiple peptides. Despite the fact that several kinds of modifications can be also attributed to single peptides, their identification always points to a *protein family*, which includes all the possible proteoforms generated by the translation of a given gene. The ways these modifications are actually combined within a single intact protein are, in fact, impossible to infer. On the contrary, top-down MS can determine the mass of the intact protein analyzed in MS^1, which already accounts for all the modification present in that specific analyte, whereas MS/MS should confirm possible modifications and precisely assign their position. In this way, top-down MS should theoretically lead to the identification of a specific proteoform.

Despite the theoretical advantages of top-down MS over proteolysis-based approaches, protein identification throughput allowed by middle-down and bottom-up MS is largely superior due to the above mentioned ease in peptide analysis. Furthermore, top-down requires a specific, sophisticated pipeline to be implemented in a mass spectrometry laboratory, which includes dedicated high-end instrumentation and data analysis software that can cope with the increased complexity of this type of analysis.

Therefore, the current scenario in MS-based protein analysis is variegated, and the strategy to implement for a specific study is chosen on the basis of available instrumentation, scale of proposed investigation and its specific aims, with the awareness that, oftentimes, more than one solution would be potentially applicable to the case.

1.2 Aim of the thesis.

In a single sentence, the common trait of the hereinafter presented research can be defined as *the development of novel analytical methods for the characterization of complex proteins through direct application to targeted studies of state-of-the-art mass spectrometry technology.*

Going from general to particular according to the deductive method, it is apparent that the development of new ways of studying complex objects requires a *targeted* investigation. Large scale studies would indeed introduce issues in the interpretation of results. Moreover, the *complexity* we aimed to challenge is represented by the objects under investigation *per se*, not by their number. Specifically, this thesis is focused on proteins whose complexity arises from the large size, the presence of multiple polypeptidic chains interconnected by solid bonds (e.g., by covalent bonds such as disulfide bridges between side chains of cysteins), multiple chemical modifications, or combinations of these elements. The beauty of the designed methods relies on the fact that these proteins were mostly studied in intact form, without over-manipulating them.

The term mass spectrometry *technology* refers in this context to the synergy of instrumentation (i.e., hardware), ion fragmentation techniques applied to structurally explore protein analytes as well as data treatment and analysis dedicated and adapted to a specific study. Although hardware modifications on mass spectrometers are beyond the scope of this thesis, the final outcome is that for pursuing the goals of this research we introduced new ways in using commercially available instruments and re-modelled the commonly employed instrument control and data collection strategies.

As described in detail in the following Chapter, several mass analyzers and ion fragmentation methods are currently available. The term *state-of-the-art* is not referred simply to hardware/software sophistication, but rather to the application of advanced instruments to extreme cases, finalized to obtain a proof-of-concept and to push the limits of technology underlining its actual limitations. Notably, we prefer to deeply exploit the information condensed in traditional MS^1 and MS^2, for instance by using not only the sequence information derivable by tandem mass spectrometry but also analyzing the abundance of identified product ions, rather than to combine them with a number of additional MS-related techniques.

In order to provide a logical and, possibly, comprehensive explanation of results, this Thesis is structured as follows: Chapter 2 is aimed at clarifying in detail the techniques used in the single studies, and the rationale supporting their choice. Chapters 3 to 5 report an introduction and summary of results achieved: for the elucidation of the effects of enzymatically-induced modifications on a peptide (Chapter 3); in the analysis and quantification of modifications on large proteins by the application of intact mass measurement and BU MS (Chapter 4); in the TD MS characterization of large proteins (Chapter 5), Finally, Chapter 6 includes the research articles related to the works described in the previous Chapters, whilst Chapter 7 summarizes the obtained results and outlines future research directions.

Chapter 2: Experimental methods

2.1 Mass spectrometry basics for the analysis of biological molecules

As expected from Chapter 1, the analytical technique employed in the research studies described in this Thesis is mass spectrometry. Therefore, before proceeding with the description of the experimental procedures employed for characterizing complex biomolecules, it is appropriate to define some basic concepts, which will also help in clarifying the peculiarities and sophistication of the research projects described hereinafter.

2.1.1 Molecules, elements and isotopes in mass spectrometry.

The only particles that can be detected in MS are the charged ones, and in almost all of MS applications these are ions. Generally, most of the studied ions are cations, or positively charged species, and this is the case for polypeptides. Although in principle a cation can be obtained by removal of one or more electrons from a neutral molecule, in the case of biopolymers the ionization in positive mode is carried out through the adduction to the molecule of a positively charged particle. Commonly, the charged adduct is a proton, but also metals such as sodium or potassium can be bound to polypeptides. An ion formed by charging the analyte during the ionization process is known as the *molecular ion* (more strictly speaking, *"pseudomolecular ion"* when protonation occurs). If we assume protons as charging adducts, we can conclude that the elemental composition of the analyte cation is the same as of the neutral analyte, plus the adducted protons. Indeed, each organic molecule is formed by a number of different elements and these are normally present in multiple isotopes, which differ from each other by the number of neutrons present in the nucleus, and hence have different masses.[16]

The practical consequence of the existence of isotopes is that when ionizing a pure analyte we deal with a mixture of molecules with the same elemental formula but different masses, and, consequently, different *m/z* values even when the number of added protons is fixed. In other words, given that we almost always analyze a group of ions at a time, and practically never a single ion, for each *charge state* of our analyte cation, defined by the number of positive charges it owns, we will detect not a single signal but multiple ones, arising from the different *combinations* of isotopes. The groups of signals detected for a selected charge state of the analyte of interest is collectively called an *isotopic envelope* or an *isotopic distribution*. Each combination of isotopes for the same molecule, to which a specific *m/z* signal should correspond,

is defined as an *isotopologue*.[17] The relative abundance of each isotopologue within a given isotopic distribution can be calculated theoretically by considering the elemental formula and the relative abundances of each element included in it. Amino acids and, thus, polypeptides are composed of a limited number of different elements, specifically: carbon, hydrogen, nitrogen, oxygen and sulfur. It is important to remind that these five elements are not equally represented in amino acids (sulfur, for instance, is present only in two of the proteinogenic amino acids, cysteine and methionine); moreover, the relative abundance of the respective isotopes changes substantially among these elements, and it is primarily carbon that presents, aside of the most abundant isotope (^{12}C, constituting 98.9% of overall carbon content), a fairly abundant heavier isotope, ^{13}C, accounting for ~1.1% of overall carbon content. The presence of these specific five elements in natural polypeptides explains why the so-called *monoisotopic peak*, which by definition corresponds to the peak of the isotopologue composed exclusively of the most abundant of the isotopes of each element, corresponds also to the peak of the lightest isotopologue. Figure 2.1 summarizes the concepts introduced above and shows the effect of mass variation on the relative abundance of the monoisotopic peak within a certain isotopic distribution. With the increase in the mass of the analyte, the monoisotopic peak intensity is progressively reduced. Importantly, the (spectral) *dynamic range* provided by modern mass analyzers, expressed as the ratio between the intensities of the most abundant over the least abundant peak within the mass spectrum typically span three to four orders of magnitudes. As a consequence, the monoisotopic peak may not be detectable for polypeptides of considerable mass.[7]

2.1.2 Mass accuracy, mass resolution and their relevance in proteomics.

It is apparent that, with over 20'000 possible gene products, to confidently identify a polypeptide sequence, not only MS^2 experiments have to be applied, but the measured mass (actually, *m/z*) of the analyte has to be determined with high accuracy. *Mass accuracy* is defined as the difference between experimental and theoretical *m/z* value of an analyte relative to the theoretical *m/z* value, and it is commonly measured in parts-per-million (ppm). In MS-based proteomics, two definitions for mass are usually considered: *monoisotopic mass*, or the mass of the monoisotopic isotopologue, and *average mass*, calculated on the base of the molecular formula considering the

average atomic masses of each element. Experimentally, the average mass is determined as the centroid of the isotopic distribution of the biomolecule of interest; unfortunately, due mainly to statistical limitations (i.e., the number of ions considered in a MS analysis is generally too low to reflect precisely the real relative ratios of different isotopes of a certain element), it has been proven that the average mass determination has an inherent uncertainty, such that the mass accuracy is limited to 10 ppm for biomolecules larger than 10 kDa and ±0.1 Da for smaller ones, as estimated by Zubarev *et al*.[18] This consideration immediately emphasizes the importance of measuring directly, whether possible, the mass of the monoisotopic peak with high accuracy, as in practice the monoisotopic mass is much less affected by poor statistics than the average mass. High mass accuracy measurements are highly beneficial in proteomics for confident protein identification: for instance, with mass accuracy of ±1 ppm it has been calculated that it is possible to filter out 99% of peptides having the same *nominal mass* (i.e., the mass of a molecule or ion calculated considering the integer mass of the most abundant isotope of each atom) but different *elemental composition*.[19]

Another fundamental parameter in MS analysis of polypeptides is *mass resolution* (or, simply, resolution).[20] This is defined as the capability of the mass analyzer to distinguish two ion signals in a mass spectrum. Historically, this parameter has been based on the resolution of two ion signals above 10% of their intensity, and calculated as $R=M/\Delta M=M_2/(M_1-M_2)$, where M_1 and M_2 (with $M_1 > M_2$) are the *m/z* values of the signals and ΔM, known as *resolving power*, is the *m/z* difference of the two measured signals. A more recent definition is based on the width of the ion signal at its half height. In this case the resolution is known as full width at half maximum, or FWHM. The relevance of *high resolution* in MS-based proteomics and in protein characterization is twofold. First, it allows the discrimination between isotopologues even for large, multiply charged polypeptides or proteins, which is necessary if we intend to determine the ion charge state in a complex mixture and to use the monoisotopic mass, and also the differentiation between (partially) overlapping isotopic distributions. In fact, we have to consider that in the most commonly employed ionization technique for proteomics, the number of proton adducts, and therefore the cation charge state, is generally increased with the polypeptide mass (as it will be explained in the section 2.3.1). If we approximate the distance between isotopologues in a distribution to the mass of a neutron, which for the sake of simplicity we can now consider as equal to 1.003 u (for the most common case of ^{13}C), this then implies that the *m/z* distance

between two consecutive isotopologues for a singly charge peptide cations is of 1.003/1~1 *m/z* units, whereas for a quadruply charged one this corresponds to only 1.003/4~0.25 *m/z* units. In other words, the resolution required for discriminating consecutive isotopologues increases with the mass of the analyte. Note, that this aspect becomes more relevant, thus, passing from BU to MD and TD proteomics. Secondly, the achieved mass accuracy is a function of the determination of the ion peak position in the mass spectrum; the higher the resolution, the lower the possibility that two close peaks can interfere with each other with subsequent mass shift which would ultimately lower the mass accuracy.
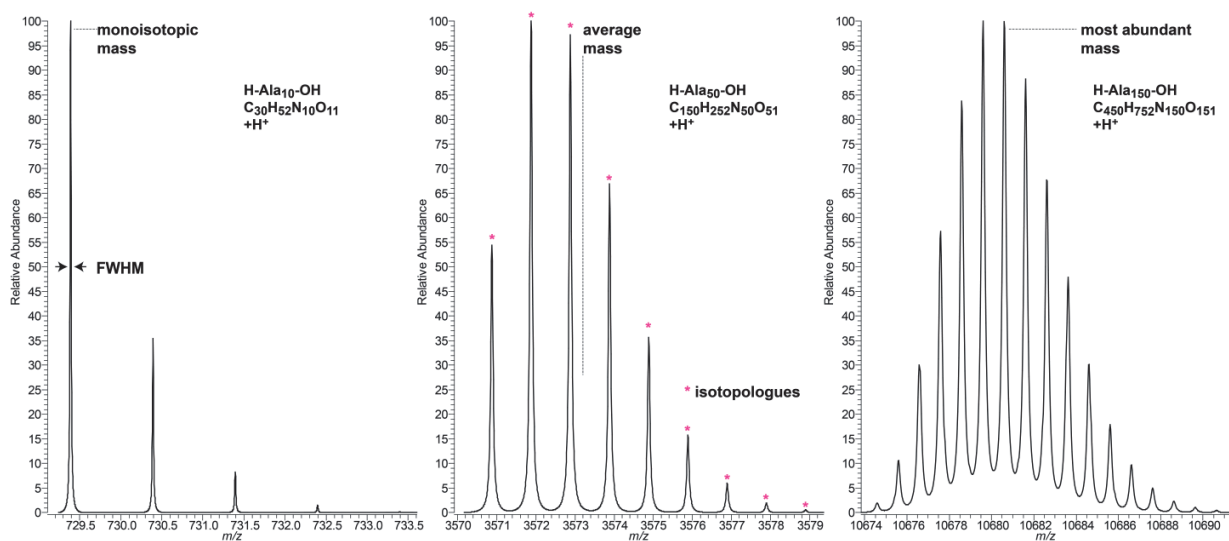
Figure 2.1. Variation of the isotopic distribution according to the analyte mass. Three examples are reported for (from left to right) polypeptides composed of 10, 50 and 150 alanine amino acid residues. Monoisotopic, average and most abundant masses are indicated. Point at which the FWHM resolution is calculated is marked for a selected peak.

2.1.3 Signal-to-noise ratio and high resolution mass spectrometry.

As previously mentioned, the dynamic range is the parameter indicating what components of a complex mixture can be detected in a single mass spectrum. In the current generation of high resolution mass spectrometers, it is limited to three-four orders of magnitude, effectively preventing the possibility of observing the monoisotopic peak of proteins larger than ~25-30 kDa (as beyond this threshold the ratio between the most abundant isotopologue in the distribution

and the monoisotopic peak exceeds this range). The lower limit of abundances in a mass spectrum is populated by non-analyte signals, i.e., spectral noise (and specifically, for most mass analyzers, either chemical or electrical noise). Increasing the *signal-to-noise ratio* (SNR) beneficially contributes to analysis of polypeptides. First, it is intuitively obvious that high SNR prevents the possibility that lowest components of an isotopic distribution, including for large polypeptide ions even the monoisotopic peak, can be confused with noise. Secondly, high SNR measurements have been demonstrated to lead to mass estimations with higher mass accuracy.

2.2 Protein structure and intrinsic complexity

In Chapter 1, we described the complexity characterizing proteomes in terms of high number of proteoforms present. Hereinafter we will introduce basic concepts about *intrinsic* structural complexity of proteins, underlining the importance of those aspects which are generally relevant from a mass spectrometry-centered perspective, and more specifically for high resolution MS.

2.2.1 Amino acids: functional groups and possible interactions in a polypeptidic chain.

As mentioned in Chapter 1, 20 proteinogenic amino acids are encoded in genomes of eukaryotes. As suggested by the name, an amino acid invariably includes two main functional groups: an amino group, and a carboxylic acid. The central carbon, termed alpha carbon or C_α, interconnects the amino and the acid group, and is linked to a hydrogen and to the so-called side chain, a variable group that discriminates the different amino acids. These are classified according to the physical-chemical properties of their side chains as displayed in Figure 2.2. Importantly, the protonation process mentioned in the previous section (and further explained in the following one) involves the N-terminal amino group (when unmodified) and the amine groups on the side chains of lysine, arginine and histidine.
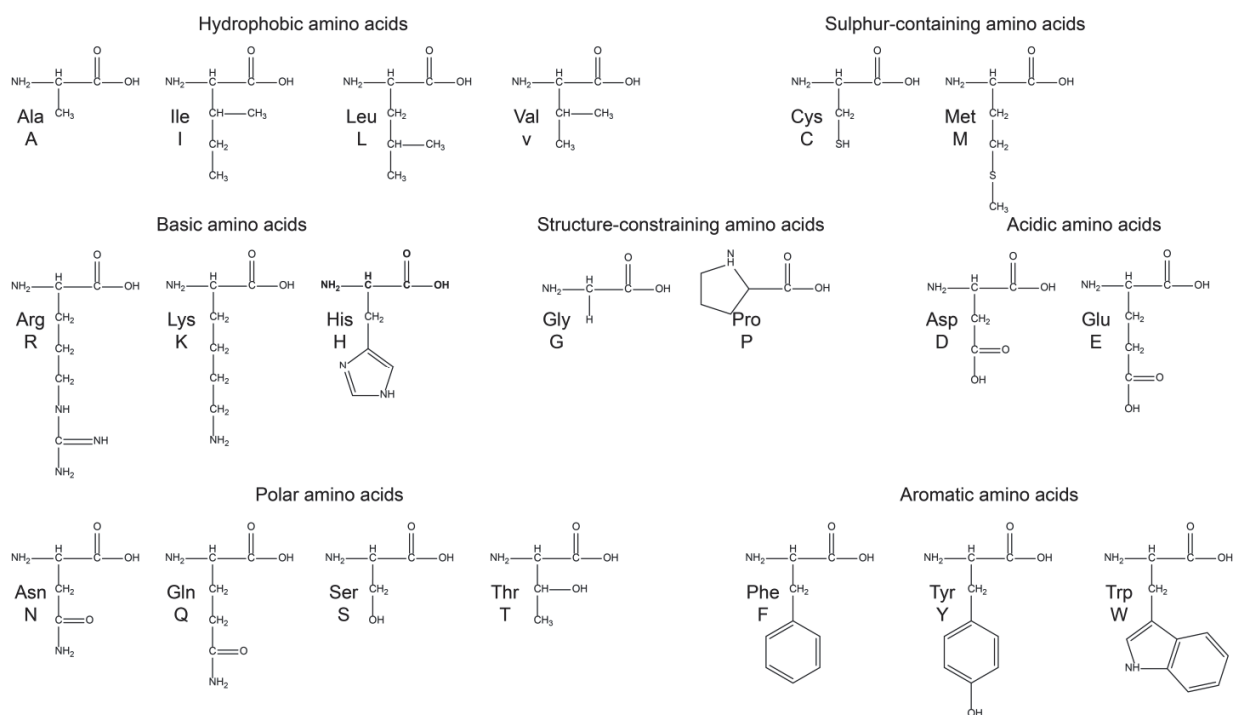
Figure 2.2. The 20 proteinogenic amino acids grouped by physical-chemical properties of their side chains. Amino acids are indicated with both three and single letter codes.[21]

Polypeptides are polymers of amino acids.[22] A condensation reaction between the amino and the carboxylic acid group of two distinct amino acids, with the elimination of a water molecule, serves to covalently bind two monomers, which are called *residues*, through the formation of an amide bond also known as the *peptide bond*, Figure 2.3. As highlighted in the Figure, the repeated sequence N- C_α-C is known as the *peptide backbone*. The geometry of each polypeptide can be described as a function of the angles between these covalent bonds. Particular importance has to be attributed to the dihedral angles Φ, Ψ and Ω between backbone planes. Given that side chains can introduce steric hindrance, an interesting use of dihedral angles, specifically of Φ and Ψ, is provided by the Ramachandran plot, which indicates the allowed combinations of values for these two angles for the amino acid residues constituting a polypeptide. Notably, different structural elements are characterized by specific (Φ, Ψ) pairs.[23, 24]
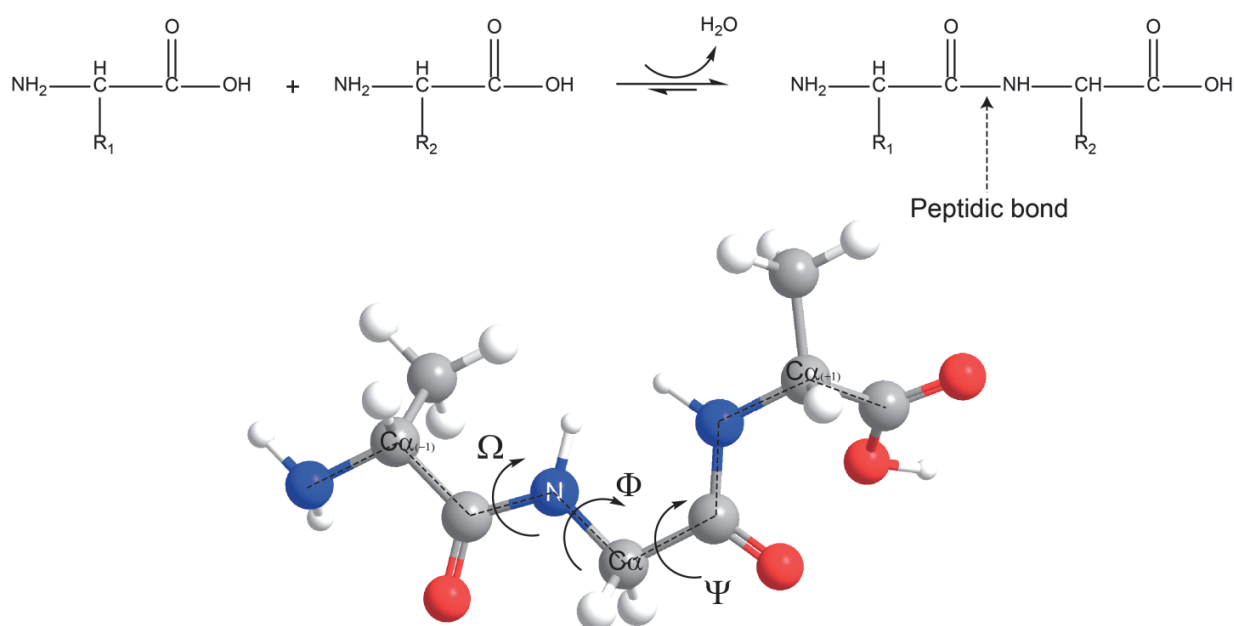
Figure 2.3. Top panel: schematic representation of the formation of a peptidic bond through a condensation reaction and water elimination. Bottom panel: positioning of the dihedral angles on the peptide backbone (marked with a dashed line) of the tripeptide H-Ala-Gly-Ala-OH.

Main interactions between amino acid residues in a polypeptide include: electrostatic interactions between positively and negatively charged amino acids; hydrogen bonds; dipole-dipole interactions; Van der Waals interactions; and hydrophobic interactions. Notably, interacting residues can be placed at variable distances, depending on the nature of the interaction. For instance, electrostatic interactions generally require amino acid side chains to be closer than 400 pm, with interaction force decreasing with the square of the distance.[25] Hydrogen bonds are generally formed within 200-300 pm of distance between donor and acceptor.[26]

2.2.2 Protein structure: a category-based nomenclature.

Spatial organization of polypeptides is fundamental to any of their functions. If, for the sake of simplicity, we grossly divide proteins into two main groups, *enzymes* and *structural proteins*, we can immediately pinpoint how overall geometry and three-dimensional reciprocal positioning of amino acids are governing protein activities. Enzymes are biological catalyzers of chemical reactions, and generally they present an active site with highly specific spatial compatibility for reaction substrates. Structural proteins, on the other hand, stabilize cell structures by interacting spatially with each other, creating supramolecular entities based on the interaction and binding of several protein units. It is intuitive that strength of binding and geometry of the final structure of a supramolecular entity depends on the spatial conformation and shape of the subunits.

The process through which a polypeptidic chain reaches its native three-dimensional conformation, capable of executing a specific biological function, is known as *folding*.[27] This process starts before the polypeptidic chain is completely translated by ribosomes, and is believed to start at a local level, with single portions of the polypeptide that start being structured. These nucleation sites are then driving the rest of the chain to a more complex and ordered three-dimensional structure. The number of interactions involved in folding and in keeping stable the native conformation is high (*vide supra*), so that a classification for describing protein structure according to a reductionist principle was necessary.[28] Currently, four levels of structural organization are considered for proteins:

1. The ordered sequence of amino acids within a polypeptide is called *primary structure*. This level includes all the information required for the determination of higher order structure. From the MS point of view, identifying the primary sequence of a protein is the first step towards the assignment of a selected protein family and, then, of a specific proteoform. This operation is conducted by combining information from MS^1, which returns the amino acid composition, and MS^2, which is used for establishing the amino acid order. Primary structure determination can be performed in a database-driven fashion, as in proteomics (i.e., by matching MS^1 and related MS^2 mass spectra to a certain number of entries listed in a protein database, or, in the case of middle-down and bottom-up approaches, in a database composed of peptides obtained by *in silico* digestion of the same proteins). Conversely, *de novo* sequencing relies more extensively on MS^2, as no

sequence information stored in a database can be used and all the amino acid residue positions have to be determined by gas phase fragmentation. Further fragmentation of a product ion (i.e., MS^3) may be necessary for distinguishing between structural isomers of amino acids such as α- and β-aspartate (i.e., "normal" aspartate and its variant with a carbon atom – C_β – moved from the side chain to the backbone). Note that, by convention, the amino acid sequence is ordered from the amino acid residue with the free amino group (N-terminus) to that with the free carboxylic acid group (C-terminus).

2. Simple structural elements composed by a limited number of amino acids, normally consecutively positioned within a polypeptidic sequence, are known on the whole as *secondary structure*. In general, the secondary structure elements present in a protein can be classified according to two different, but correlated, features: the hydrogen bond connectivity or the dihedral angles. The former classification relies on the observation that secondary structure is based on hydrogen bonds connecting amide and carboxyl groups of polypeptidic backbone. The latter refers to the aforementioned Ramachandran plot, as each secondary structure element is characterized by a specific range of possible dihedral angular values for the amino acid residues involved. Among the most common elements we can mention the α-helix, the β-sheet and the β-turn. Portion of proteins which do not present any specific classified secondary structure elements are normally defined as random coils.

3. The achievement of the final, three-dimensional conformation of a protein, or *tertiary structure*, can be chemically described in terms of a hydrophobicity-driven process. Indeed, as proteins are immersed in a water-based environment, hydrophobic amino acids will naturally tend to exclude water molecules and aggregate. Energetically, the native conformation is positioned at the bottom of an energy funnel, Figure 2.4. If hydrophobicity is fundamental for determining what will be the amino acid residues exposed to the interface with a solvent, other interactions are important for maintaining native folding. Particularly, non-covalent Coulombic interactions between charged amino acids (also known as "salt bridges") and dipolar interactions, as well as covalent disulfide bridges formed between side chains of cysteine residues (see Section 2.2.3) all contribute to tertiary structure stability.

4. Although most proteins are biologically active when they reach a stable tertiary structure, some need to form complexes with others. The association of multiple polypeptides/proteins is called a *quaternary structure*. It is apparent how this structural level can be properly achieved only in presence of a correct tertiary structure. Generally, quaternary structure is reached through non-covalent interactions among the different polypeptides. For example, proteins which are composed of multiple chains linked also by inter-molecular covalent bonds, such as antibodies, are considered as having a quaternary structure.
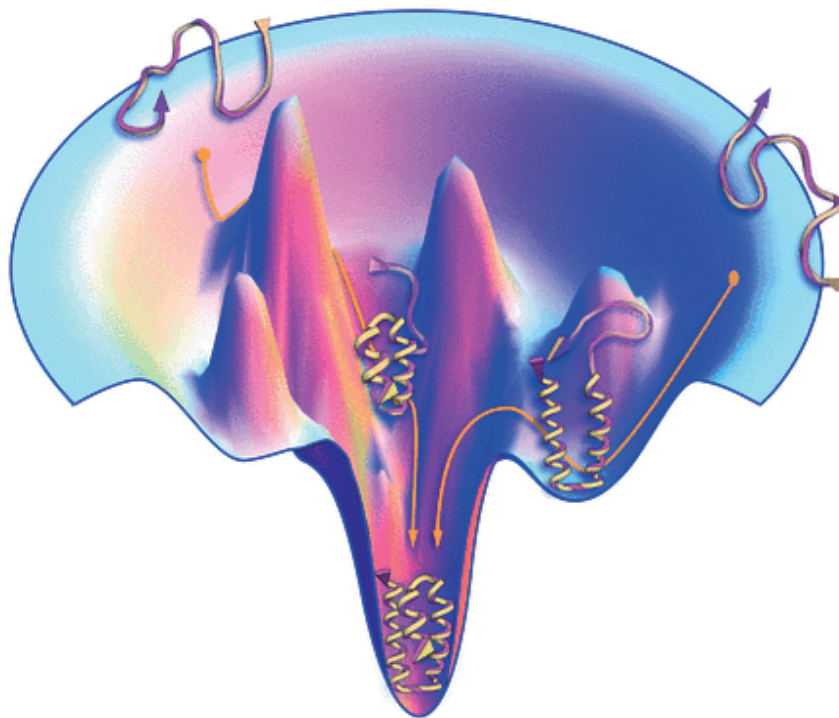

Figure 2.4. Graphic representation of the energy funnel theory for protein folding. Proteins minimize their potential energy by reaching the native conformation, whereas partially-folded conformations are energetically unfavorable. Figure adapted from [27].

2.2.3 Post-translational modifications and protein structure.

Most of currently known PTMs involve the addition of small molecules to selected amino acid side chains. This is the case, for instance, with the widely studied acetylation, methylation and phosphorylation processes. Other PTMs involve the addition of larger molecules, such as peptides (e.g., ubiquitination[29] and SUMOylation[30]) or oligosaccharides (e.g., N-linked glycosylation[31]). Large scale MS-based studies of these PTMs benefit from the relatively large mass difference between unmodified and modified amino acid residues, which, combined with the fact that specific modifications can be present only on selected amino acid side chains, allows in some cases for the localization of the modified sites even with low mass accuracy and low resolution instruments, although high mass resolution and accuracy surely increase the identification confidence.

Despite the fact that some of the above-mentioned PTMs might induce non-negligible changes in the physical-chemical properties of the involved proteins, as is the case of phosphorylation which accounts for the introduction of a negatively charged chemical group, their overall impact on the protein structure is limited compared to that of structural PTMs. Among them, the most important is surely represented by the formation of a disulfide bridge (i.e., a covalent bond) between thiol groups of two cysteine residues. The role played by disulfide bridges in stabilizing protein structures is achieved by placing in vicinity two non-consecutive portions of the polypeptidic sequence, potentially enhancing hydrophobic interactions and preventing the possibility that water molecules, which will have no access to the disulfide-protected area, will disrupt secondary structure elements by creating alternative hydrogen bonds with backbone amide groups. It is noteworthy that disulfide bridges can limit the efficiency of polypeptide sequencing in MS^2, as their presence requires the cleavage of two bonds for obtaining a product ion. Interestingly, other cross-linkages between side chains are possible, like enzymatically-induced transamidation which involves the ε-amino group of lysine and the amide group of glutamine. Finally, a limited number of PTMs, induced either chemically or enzymatically, provokes the variation substitution of an amino acid in the sequence with another one. For instance, deamidation of glutamine or asparagine results in the formation of glutamate or aspartate,

respectively. The mass variation caused by deamidation is lower than 1 Da, therefore high mass accuracy and mass resolution are required especially when the overall mass of the polypeptide of interest is large.

2.3 Instrumentation

This section is dedicated to the description of mass spectrometers employed for the research of this Thesis, and will particularly underline, in an order that allows ideally to follow to the logics of a real experiment (i.e., ionization, ion separation, *m/z* isolation, ion activation and fragmentation followed by product ion detection, and finally analysis of the acquired data), the following aspects:

- The type of ionization employed, namely electrospray.
- High resolution hybrid mass spectrometers.
- Tandem mass spectrometry techniques for complex polypeptides.
- Data analysis.

2.3.1 Electrospray ionization.

Analyte ionization represents the *sine qua non* operation for mass spectrometry analysis. Biomolecules, and polypeptides in particular, are composed of a number of relatively weak covalent bonds (e.g., most of those connecting PTMs to amino acid side chains). From this derived the need of *soft ionization methods*, which could induce the analyte charging phenomenon avoiding destructive effects.

Two techniques suitable for biomolecule ionization were introduced between the 80s and 90s, for which John B. Fenn and Koichi Tanaka were awarded the Nobel Prize in Chemistry in 2002:[32, 33] *electrospray ionization* (ESI) and ionization techniques that finally led to development of *matrix-assisted laser desorption/ionization* (MALDI) by Franz Hillenkamp and Michael Karas.[34, 35] The studies presented in this Thesis are entirely realized applying ESI in its different variants. The success of ESI is largely due to its ability to form multiply-charged ions (*vide infra*) and possibility of coupling on-line this *continuous* ionization method with widespread separation techniques such as *chromatography* (and, particularly, *reversed phase* chromatography) or capillary electrophoresis, which entail a constant liquid flow.

From the theoretical point of view, an electrospray source resembles a controlled current electrolytic cell. Indeed, a liquid pushed through a capillary is put under a high voltage (in the kV range), with the electric field being comprised between the tip of the capillary and the inlet of the mass spectrometer, Figure 2.5. Essentially, talking about the generation of analyte cations, the capillary tip acts as an anode and the mass spectrometer inlet as a cathode.

With regard to the sequential stages required for ion formation, we distinguish between formation of spray, desolvation of droplets, emergence of charged particles from droplets and sampling of ions from the atmospheric pressure environment inside the mass spectrometer. The formation of the spray is ensured by the application of the electric field. In presence of low field, a droplet eluting from the capillary appears spherical. Its shape changes into an elongated form by ramping up the voltage until, at onset field, the spray is formed as the pressure is higher than the surface tension. What is formed is the so-called Taylor cone, from which droplets are released. Depending on the flow of the liquid through the capillary we discriminate between ESI (also known as microESI, as it is characterized by a flow in the order of µl/min) and nanoESI (with a lower flow in the order of hundreds of nl/min). The formation of droplets and their desolvation is caused by the accumulation at the liquid surface of particles charged by the combination of the applied field and the presence in the ESI solution of proton donor species (generally, acids such as trifluoroacetic acid or formic acid). The droplet desolvation/evaporation process, favored also by the proximity of the instrument inlet, which is usually heated, and by the eventual use of nebulizing gas, increases the charge per unit of volume, so that, because of the excessive Coulombic repulsion, higher than the surface tension, larger droplets divide into smaller ones and finally release ions. In other words, these charged droplets reach a radius value lower than the stability one, also known as Rayleigh radius, due to their progressive evaporation, with the result of undergoing a fission.[36] To better understand this phenomenon, we can mention that in a typical microESI source the first generation droplets released from the Taylor cone have a diameter of ~1.5 µm, whilst offspring droplets have a reduced diameter of ~0.1 µm; this implies for the second generation a volume equal to ~2% of the parent droplet but a charge content estimated in ~15% of the initial one.[7, 37] Considering the mechanism of this process, it is apparent why organic solvents such as methanol or acetonitrile are often used to help the spray stability, as they have a lower surface tension than water. The nanoESI allows for a higher sensitivity as the parent droplets formed at the Taylor cone are already smaller than in the

microESI counterpart by about one order of magnitude. NanoESI is also more tolerant towards salts. Finally, ions formed at atmospheric pressure have to be introduced in the mass spectrometer, whose analyzer generally operates in medium to high vacuum conditions. Besides the already mentioned heated ion inlet (generally, a glass or metallic capillary), the solution for creating a pressure gradient is offered by the so-called *differential pumping system*, consisting in the definition of compartments, on which vacuum pumps operate, connected by small apertures so that a large difference in pressure between consecutive chambers can be maintained. An ion optics system, placed in the region at intermediate vacuum and composed of focusing lenses and/or multipoles, ensures efficient ion transmission towards the internal part of the mass spectrometer.
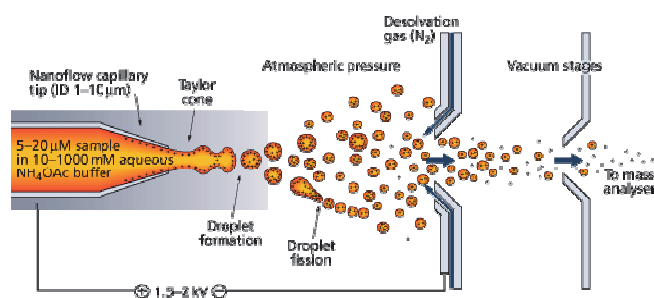
Figure 2.5. Graphic representation of an electrospray ion source. Image adapted from the Royal Society of Chemistry.

Main features of ESI of polypeptides (in positive mode) include: formation of *multiply-charged ions* (in contrast with MALDI, which produces almost uniquely singly-charged ions) by protonation (or, eventually, complexation of other charged species such as metal ions); generation of a *charge state distribution* (or charge state envelope), composed of analyte ions with different charge states (i.e., with different number of proton adducts) and, hence, different *m/z*; concentration of polypeptides of interest that, depending on the mass analyzer employed for their analysis, finds a lower limit in the pmol or even fmol range. The advantages offered by the here described ESI peculiarities will be illustrated in the following sections of this Chapter.

2.3.2 High resolution hybrid mass spectrometers

An important characteristic shared by all the employed instruments mentioned in this Thesis is their *hybrid design*.[38, 39] This means that two mass analyzers, generally a low resolution and a high resolution one, are positioned in series. This technology has a high impact on the general characteristics of the instrument as the low resolution mass analyzer is chosen to be also capable of working as a collision or reaction cell, i.e., to enable tandem mass spectrometry. In this way, the capabilities of the mass spectrometer are extended, as often (but not always) MS^2 parameters are not compatible with operational conditions of high resolution mass analyzer (*vide infra*). In extreme cases, the first mass analyzer actually operates exclusively for selection of ions for MS^2. Furthermore, a sophisticated design allows for the ion manipulation in the low resolution mass analyzer while either precursor or eventually product ions are analyzed with high mass resolution and accuracy in the high resolution mass analyzer, improving the duty cycle of a typical sequence composed by one MS^1 scan followed by a series of MS^2 scans of selected precursors.

Research studies presented in the following Chapters of this Thesis were performed with three different high resolution mass analyzers: time-of-flight (TOF) and Fourier transform – ion cyclotron resonance (FT-ICR) and Orbitrap.[7] These will be hereinafter described according to their ion detection principle, that is *single ion counting* for TOF and *induced current detection* for FT-ICR and Orbitrap. For reason that will be explained in the dedicated paragraph of this Chapter, MS analysis relying on mass analyzers of the latter group is also known as Fourier transform mass spectrometry (FTMS).

2.3.2.1 Fourier transform mass spectrometry.

High resolution mass spectrometry can rely on a non-destructive ion detection principle, the *induced current detection*. Specifically, ions can be trapped in a confined space (i.e., an ion trap), where their periodic motion produces an induced current in the detection electrodes circuit. This current is digitized as a time-domain signal containing all the motion frequencies of measured ions, with intensities proportional to the abundance and charge of the ions in the trap. Effectively, the time-domain signal, whose length can last up to some seconds (with a record duration of 50 sec on the employed FT-ICR MS), contains a convoluted waveform resulting from the sum of all the waveforms produced by the motion of different ions. The time-domain signal, also known as transient, can be converted to a frequency-domain signal by applying the mathematical operation

of Fourier transform. The resulting frequency spectrum is a complex one, i.e., it contains both a real and an imaginary part. Importantly, the real part returns the so-called *absorption spectrum*, whereas the imaginary part gives a second spectrum, the *dispersion* one. Mathematically, it is fundamental to know the phase of ion at zero time (i.e., the phase at the beginning of signal recording). In fact, at zero degree phase the absorption spectrum only contains positive values and peaks have the highest possible resolution, whereas the dispersion spectrum includes both positive and negative values at a slightly lower resolution. This means that at zero degree phase the absorption mode spectrum can be used alone to determine the frequencies of trapped ions. For any other phase value different than zero, though, both components, real and imaginary, are composed of linear combinations of absorption and dispersion modes, i.e., they both contain asymmetric peaks described by positive and also negative values (with a reduced resolution). The final conclusion is that, whether it is not possible to establish the ion phase at time zero (which is necessary to eventually apply a phase correction), a third spectral representation is used: the *magnitude spectrum*, which allows the representation of symmetric peaks independently from the initial phase value. It is obtained by combining both real and imaginary parts according to the following equation:

$$magnitude = \sqrt{Re^2 + Im^2} \qquad\qquad \text{(eq. 2.1)}$$

The drawback of a magnitude mode spectrum is that its final resolution is about one half of that achieved by the corresponding absorption spectrum at zero degree phase.

In summary, in FTMS *m/z* values are represented by specific frequencies and it is the possibility of measuring frequencies with high accuracy that ensures, ultimately, high mass accuracy. The conversion of frequencies into masses is realized by calibrating the instruments for the *m/z* range of use with a limited number of ions with accurately known *m/z*. Since averaging the current period improves the accuracy in induced current detection, elongating the time of signal recording, i.e., increasing the length of the transient signal, allows for proportional increase in final mass resolution (and, thus, accuracy).[40]

Hereinafter we will briefly summarize the working principles and illustrate the technical features of the two instruments present in our laboratory, LSMB.

- Fourier transform – ion cyclotron resonance-based hybrid mass spectrometer.

The first FTMS instrument in use in the lab, installed in 2007, is a hybrid LTQ FT Ultra (Thermo Scientific, Bremen, Germany). As shown in the schematics, Figure 2.6, its architecture consists of (from left, ion inlet, to right): an ion optics system interfaced with the atmospheric pressure inlet; a linear ion trap (linear trap quadrupole, LTQ); transfer multipole (octopole); and an ICR cell immersed in a homogeneous high magnetic field. The instrument present at LSMB is a customized version currently equipped with a 10 T superconductive magnet.
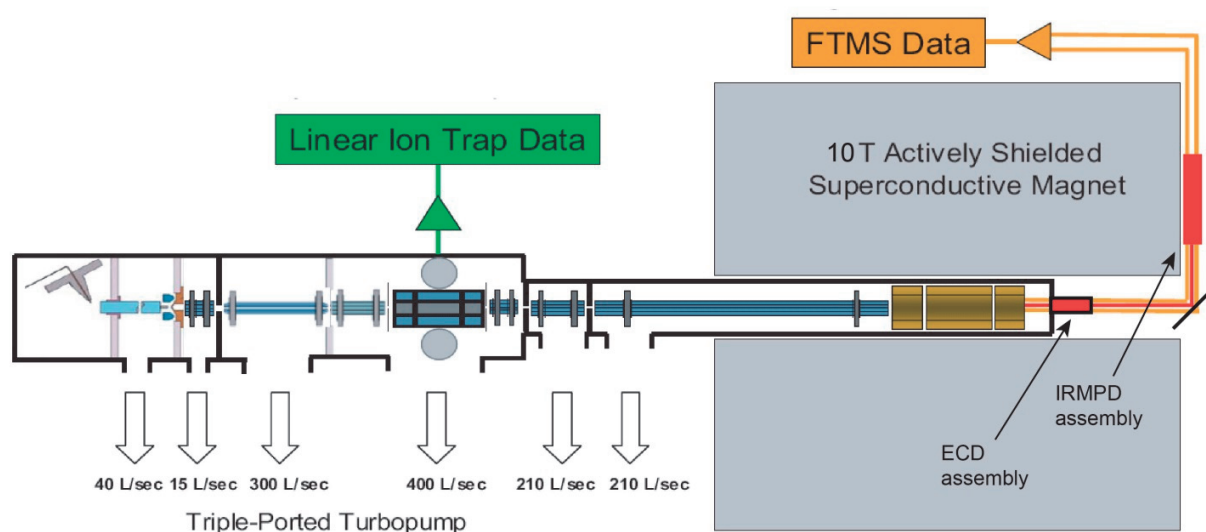
Figure 2.6. Schematics of the LTQ FT Ultra mass spectrometer (courtesy of Thermo Scientific).

Going further into detail, the mass spectrometer works with different spray-based sources (micro and nasoESI; atmospheric pressure photoionization (APPI) source). Importantly, as depicted in Figure 2.6, the vacuum is increased by a differential pumping system from the inlet towards the ICR cell, so that ions formed at atmospheric pressure are in a ~1E-3 torr environment when they reach the central multipole, called MP0, of the ion optics placed before the linear ion trap. The pressure drops further in the LTQ, which has an operational pressure of ~5E-5 torr, arriving finally to the so-called ultrahigh vacuum (UHV) region located at the level of the ICR cell (~1-2E-10 torr).

The front end of this mass spectrometer belongs to the LTQ XL series of Thermo Scientific instruments. The ion inlet consists of a heated stainless steel capillary (no orthogonal ion

injection implemented), and the first part of the ion optics is of the *tube lens – skimmer* type, specifically with a configuration that includes two consecutive skimmers followed by a quadrupole and a transfer octopole, spaced out to the LTQ. In the ion trap, ions are confined and manipulated (including ion ejection for detection as well as ion selection) through the application of a radiofrequency (RF)-modulated electric fields. The linear ion trap, derived from the Paul or 3D ion trap, consists of four rods with parabolic section and exploits the quadrupole principle, such that the radial ion confinement is performed through a quadrupolar alternate electric field, whereas the axial confinement is provided by additional static or RF-based electric fields applied at both ends of the trap.[41, 42] Note, that for trapping both positively and negatively charged ions simultaneously, RF trapping potentials are used. Neutral gas molecules (He) present in the trap are used to lower ion kinetic energy, and eventually for ion activation/fragmentation (*vide infra*). Main features of the LTQ are higher trapping efficiency (~20-fold) and increased ion capacity (~50-fold) relative to a 3D ion trap. The given LTQ can scan up to *m/z* 4000.[43]

Ion *m/z* measurement in the LTQ is performed by ramping the RF voltage and ejecting the ions orthogonally towards ion detectors. Particularly, in the LTQ the radial ion ejection is followed by detection operated by a dual dynode system for ion-electron conversion. Signal gain is obtained through a secondary electron cascade through electron multipliers. Despite the high scan speed allowed by the LTQ (~15'000 Th/s) and its high sensitivity, the linear ion trap has been primarily used for ion isolation and, eventually, activation. Ion isolation window can have a width up to 100 *m/z* units. The function named *data dependent acquisition* allows for the automatic isolation (and, whether desired, activation) of the most intense peaks in a spectrum.

Finally, a noteworthy feature of the LTQ is the possibility of controlling the number of trapped ions (more precisely, of overall injected charges) by a function named *automatic gain control* (AGC). Briefly, a short prescan (of a defined time, 50 ms) is used to determine the ion current within the selected mass range. This current value is associated to a corresponding number of charges (through a specific routine performed during the instrument calibration), therefore the instrument control software can calculate a specific ion injection time to fill the LTQ according to a user-defined desired number of charges (AGC target value). The number of charges (and, thus, number of ions) injected and stored in the trap is important to determine the operation duty cycle, instrument sensitivity, spectral SNR etc. Furthermore, as ions confined in a limited space interact

with each other, an excessively high number of trapped ions can cause phenomena collectively known as *space-charge effects* that might negatively affect ion measurements. This is valid for both low and high resolution ion trap mass analyzers.

From the LTQ, ions can be transferred to the ICR cell, immersed in a static magnetic field. A charged particle with a velocity vector perpendicular to the magnetic field experiences a force, the "Lorentz force", perpendicular to the plane defined by the velocity and the magnetic field vectors. The ion velocity vector is thus changed in its direction (but not in its magnitude), so that the ion starts a circular trajectory being finally trapped by the magnetic field in the radial direction.[44] This low amplitude motion is called cyclotron motion and its frequency can be described as follows:

$$\omega_c = \frac{qB}{m}$$
(eq. 2.2)

Where ω_c represents the cyclotron frequency, q the charge of the ion, B the applied magnetic field and m the mass of the ion. Note, that ω_c is an angular frequency and it thus depends on the velocity v and the radius r. Importantly, for measuring the cyclotron frequency, from which the *m/z* value of ions is derived, ions are excited by the application of an RF electric field on two opposite plates of the ICR cell. This amplifies the cyclotron motion radius and increases the induced current detected by two other plates on the cell, connected to a preamplifier (Figure 2.7, left panel).
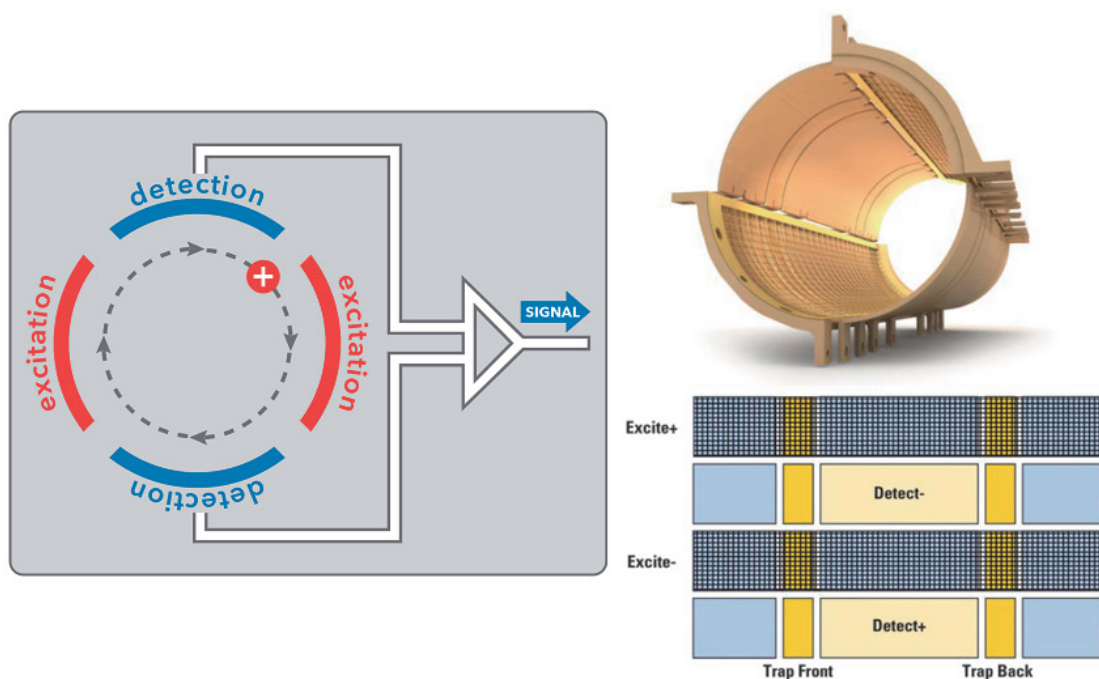
Figure 2.7. FT-ICR cell working principle and schematics. Left panel, the schematics of an ICR cell section illustrates the ion cyclotron motion as well as detection and excitation plates. Right panel, picture and schematics of the Ultra ICR cell equipping the LTQ FT Ultra mass spectrometer. Figure adapted from reference [40].

As shown in Figure 2.7, right panel, which displays a schematic of the Ultra ICR cell equipping the LTQ FT Ultra mass spectrometer, trapping electrodes are used to axially confine ions within the cell. Other two characteristics of ion motions arise from the application of trapping voltages in the ICR cell: the *axial motion* and the *magnetron motion*.

The FTMS mass spectra obtained with the LTQ FT Ultra mass spectrometer in this work are represented in magnitude mode. To extend the MS^2 capabilities of the LTQ, the back end of the instrument is equipped with a CO_2 laser and a cathode assembly, with which it is possible to activate ions trapped in the ICR cell with photons and electrons, respectively (*vide infra*).

- Orbitrap-based hybrid mass spectrometer.

Since 2012 the LSMB is equipped with a second FTMS instrument, an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific).[45] The scheme presented in Figure 2.8 clearly illustrates that

the logics behind the design of this instrument represents a direct evolution of that employed in the LTQ FT Ultra.
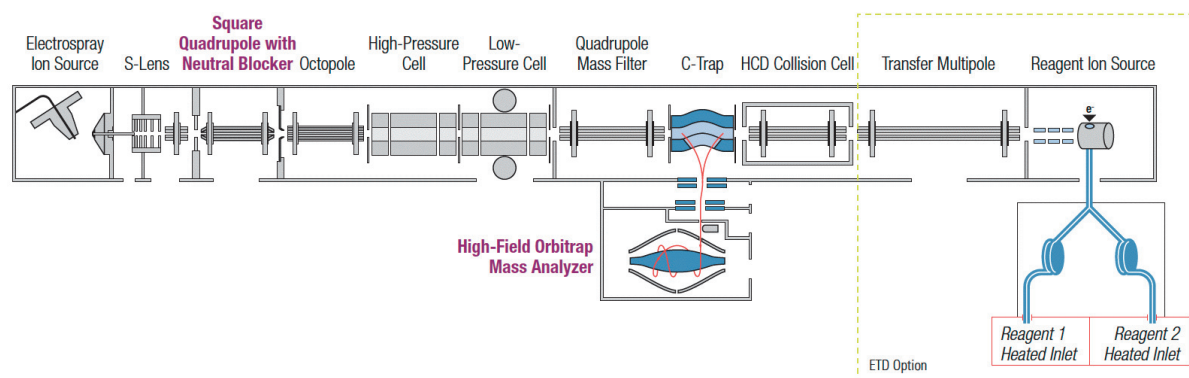

Figure 2.8. Schematics of the hybrid LTQ Orbitrap Elite mass spectrometer. Note, that the only relevant difference, in terms of hardware, with the LTQ Orbitrap Velos Pro is the type of employed Orbitrap (standard for Velos Pro, high-field in the case of Elite). Figure adapted from http:\\ planetorbitrap.com.

The instrument front end corresponds to the LTQ Velos Pro series of high performance ion traps. Improvements relative to the XL series involve both the ion optics placed immediately after the ion inlet and the LTQ itself. In regard to the inlet region, the tube lens – skimmer ion optics equipping the XL series has been replaced with a stacked ion ring guide (S-lens), composed of six progressively spaced circular ion guides. The S-lens allows for a ~10-fold increase in sensitivity compared with the old tube lens – skimmer design. Furthermore, after the S-lens the transfer quadrupole owns a curved design with a neutral blocker, limiting the transmission of neutrals to the LTQ and maximizing the number of transmitted ions. Finally, the LTQ has a completely renewed design, and it is composed by a high pressure and a low pressure cell, where He is present, in standard operating conditions, at 5E-3 and 3.5E-4 torr, respectively. The high pressure cell is dedicated to the ion storage, activation and isolation, whereas the low pressure cell serves for fast and efficient ion ejection and, subsequently, detection. The scan speed of the low pressure cell is increased in comparison to the old generation of LTQ XL to ~33'300 Th/s when operating in "normal mode".[46]

Core of the LTQ Orbitrap Elite is the high resolution mass analyzer, a high-field compact Orbitrap, Figure 2.9. This mass analyzer consists essentially of two electrodes: a larger outer electrode, split into two parts for induced current detection, and a central electrode of elongated spindle shape around which trapped ions move. A negative potential is applied to the central electrode when cations are analyzed. Ions stabilized within the Orbitrap by the applied quadro-logarithmic electric field start oscillating along the central electrodes with trajectories resembling spirals. It is possible to distinguish two motions of the ions, which are considered as independent in the first-order approximation: a *rotational motion*, highly dependent on ion initial energy, and an *axial motion*, independent from ion initial energy and which is therefore used for determining the *m/z* of trapped ions.[47, 48] Specifically, the frequency of harmonic oscillation of ions along the *z*-axis (i.e., along the central electrode) is defined as:

$$\omega_z = \sqrt{\frac{k}{(m/q)}}$$ (eq. 2.3)

Where ω_z represents the axial oscillation frequency, q the ion total charge and k is a constant proportional to the potential difference between central and outer electrode. Increasing the electric field is fundamental for increasing ions ω_z, with beneficial effects on the final mass resolution (*vide infra*). As previously mentioned, the LTQ Orbitrap Elite is equipped with an evolution of the original Orbitrap presented in 2005, whose compact geometry allows for a higher electric field (Figure 2.9).[49, 50] The compact Orbitrap works under UHV conditions improved in respect to those typical of the Ultra ICR cell described above: the standard operating pressure of our instrument at LSMB is 0.5E-10 torr. It is noteworthy that this version of the Orbitrap mass analyzer is coupled with a signal processing algorithm, commercially named "eFT", that enables mass spectra representation in absorption mode.[51]

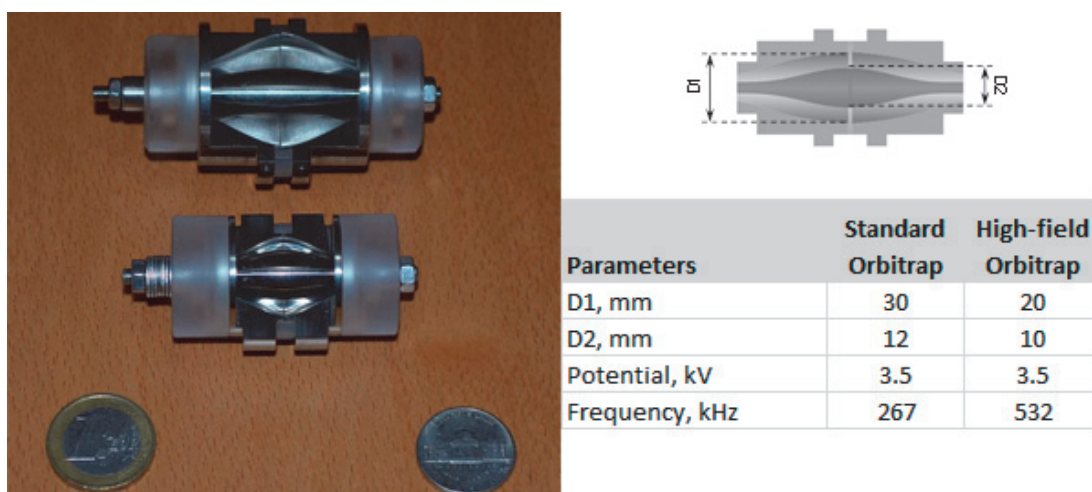| Parameters	Standard Orbitrap	High-field Orbitrap
D1, mm	30	20
D2, mm	12	10
Potential, kV	3.5	3.5
Frequency, kHz	267	532

Figure 2.9. Comparison between standard and high-field Orbitrap. Left panel, picture of both the mass analyzers, indicating the size reduction occurred passing from the old to the new generation of mass analyzer. Right panel, some dimensions and other features distinguishing the two Orbitrap generations.

Another principal component of the instrument is the *C-trap*. The name of this RF-only, gas-filled multipole derives from its shape, resembling that of the letter "C". The C-trap is used to store a packet of ions prior to their injection in the Orbitrap, which takes place from a tiny hole present in one of the outer electrodes.

Furthermore, the LTQ Orbitrap Elite is equipped with a N_2-filled HCD cell. This multipole is generally used for ion activation (*vide infra*), but can also be exploited for improving the efficiency of large ion injection to the Orbitrap through mild ion collisional cooling with neutral N_2 gas molecules.

Finally, a chemical ionization-type ion source is placed at the back end of the instrument, and it is used for the production of radical anions used for a specific ion activation technique (*vide infra*).

Note, that part of the measurements reported in this Thesis was performed on a LTQ Orbitrap Velos Pro mass spectrometer.[52] The main differences of this instrument relative to the above described LTQ Orbitrap Elite consist in the lack of eFT signal processing (i.e., magnitude mode representation is given for the mass spectra) and the presence of a standard Orbitrap in place of the high-field one.

2.3.2.2 Time-of-flight mass spectrometry.

Differently from the previously described mass analyzer, where ions where trapped either by applying a magnetic (FT-ICR) or an electric (Orbitrap) field, time-of-flight mass analyzers rely on ion separation by their *m/z* values performed in a *field-free* area.[53] Briefly, ions are first accelerated by a source potential V_s, and continue moving further into a field-free space under ultra high vacuum conditions until they reach a detector, typically a microchannel plate electron multiplier or an array detector. Therefore, for an ion of mass *m* and total charge *q=ze*, moving with velocity *v*, the initial kinetic energy E_k is expressed as follows:

$$E_k = \frac{mv^2}{2} = qV_s \qquad \text{(eq. 2.4)}$$

The velocity of the ion is considered constant within the field-free region of the flight tube with length *d*. Therefore, with the time needed for reaching the detector, *t*, being equal to

$$t = \frac{d}{v} \qquad \text{(eq 2.5)}$$

We finally obtain that:

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2V_s e}\right) \qquad \text{(eq. 2.6)}$$

With *t* that can be measured accurately by a time-to-digital converter (TDC), thus enabling the calculation of the ion *m/z*.

In practice, one of the limitations of this design is that ions entering the mass spectrometer upon ionization are not perfectly focused when they reach the acceleration zone. This limited focusing is translated in a dispersion of E_k which can be corrected by the implementation of an *electrostatic reflector*, also known as *ion mirror* or *reflectron*, in the flight tube. This device consists of a series of ring electrodes to which a potential of the same polarity as analyzed ions is applied. The reflectron creates an electric field that reflects the ions and sends them back through

the flight tube. The E_k dispersion is corrected as ions with higher kinetic energy penetrate the ion mirror more deeply before being reflected.

Among the figures of merit of modern TOF mass analyzers there are high sensitivity (e.g., attomolar for peptides, when couple to MALDI)[54] as well as elevated scan speed (e.g., modern qTOFs can perform MS^2 at 50 Hz rate) and dynamic range (determined by signal detection and digitization, with modern 10 bit signal converters capable of exceeding 5 orders of magnitude), generally higher than in FTMS mass analyzers.

- High resolution quadrupole – time-of-flight mass spectrometer

The instrument chosen for intact mass measurement and top-down characterization of large proteins in a research project described in the following Chapters is a hybrid quadrupole – time-of-flight maXis UHR qTOF (Bruker Daltonics GmbH, Bremen, Germany). The schematics of this instrument, designed with a quadrupole and a TOF mass analyzer in series, is shown in Figure 2.10. This configuration is typically used for TOF analyzers coupled to ESI sources.
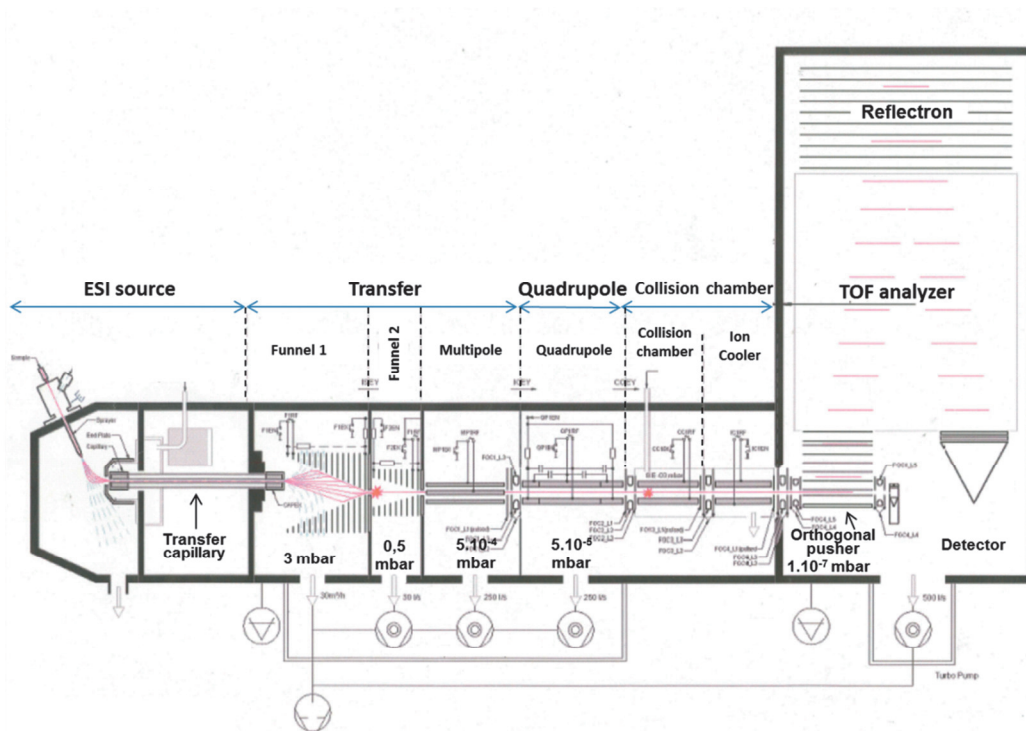

Figure 2. 10. Schematics of the maXis UHR qTOF mass spectrometer (Bruker Daltonics GmBH).

The off-axis ion inlet (solution applied to limit the introduction of neutral molecules in the rest of the mass spectrometer) is followed by two ion funnels and a multipole. The central portion of the instrument is dominated by the presence of a low resolution mass analyzer, the quadrupole, used as an ion filter and for ion activation. Precursor ions selected in the quadrupole are subjected to ion activation in the *collision chamber*.[55] Product ions resulting from precursor ion fragmentation are transferred to a thermalizing chamber (named Ion Cooler) for collisional cooling (with N_2) prior to orthogonal acceleration. This design is often referred to as QqTOF, where "Q" indicates a mass resolving quadrupole (that used also for precursor ion selection), and "q" an RF-only quadrupole or multipole dedicated to ion activation.[56]

Once left the thermalizing chamber, ions are focused by an optics device (skimmer) and enter the modulator present at the base of the TOF mass analyzer. Initially, the modulator region is field-free, so that the ions continue moving along their original trajectory. Then, a pulsed electric field is applied along the modulator gap so that ions are accelerated orthogonally to their initial direction. From here, they can be analyzed by the TOF mass analyzer working under UHV conditions (~1E-7 torr). This is equipped with a multichannel plate detector coupled with an analog-to-digital converter (ADC) operating with a digitization frequency up to 4 GHz. The pulsed ion current from the MCP is thus digitized at 250E-12 s intervals.

The employed maXis UHR qTOF was additionally equipped with a chemical ionization source for generation of radical anions to be introduced from the front end into the mass spectrometer and trapped in the collision chamber.

2.3.2.3 Concluding remarks on mass resolution and accuracy of the employed instruments.

As previously mentioned, the mass resolution achieved by FTMS mass analyzers is a function of the transient length. This implies that when time constrains are applied to the analysis, this cannot be conducted with the highest resolution possible for the employed mass analyzer. Importantly, due to the non-linear relationship between ion frequencies and corresponding *m/z* values, the mass resolution of ion peaks in FTMS changes along the *m/z* axis. In detail, whilst in FT-ICR MS the ion cyclotron motion frequency is inversely proportional to the ions' mass-over-charge, in Orbitrap FTMS ion axial motion frequency is inversely proportional to the square root of *m/z*. As

a final result, even when a high magnetic field is used to increase the mass resolution, the advantage of FT-ICR MS over Orbitrap FTMS can be limited to high frequency ions (i.e., those with low *m/z*) whereas high-field Orbitrap mass analyzers can overcome the mass resolution obtained by ICR-based instruments for lower frequency ions (typically, Orbitrap starts outperforming the mass resolution of ICR above 1000-2000 *m/z*, depending on the ICR and Orbitrap configurations used in the comparison).

With regard to the instruments that equip LSMB, the Ultra ICR cell in the 10 T magnet configuration can substantially exceed 1'000'000 resolution FWHM at 400 *m/z* with a transient length of 12.288 s. The transient length in our customized instrument has been successfully extended to 24.576 s, but the resolution has been observed to increase linearly with transient length up to 1.536 s, value after which the mass resolution increase is non-linear. For achieving resolutions that allow top-down MS analysis, for instance ~50'000 or ~100'000 FWHM at 400 *m/z*, transients of 384 and 768 ms are used, respectively.

The compact, high-field Orbitrap has been proven to reach ~1'000'000 resolution FWHM at 400 *m/z* with transient length of 3.072 s but, most importantly for routine operations, it provides 120'000 resolution with 384 ms transients (and enabled spectral representation in absorption mode with eFT).

The maXis UHR qTOF of the second generation is accredited of a maximal mass resolution of ~50'000 FWHM (recently improved to >60'000 in the newest version) which is substantially equal throughout the entire mass range considered. Nevertheless, due to the characteristics of the signal detector, the mass resolution decreases when particularly complex mixtures (i.e., in both the number of ions and their dynamic range) are analyzed, like in the case of top-down MS. Under these operational conditions, the resolution is attested at ~20-30'000 FWHM.

Moreover, an additional consideration has to be done for what concerns the usable mass range. In a hybrid configuration, the first mass analyzer may restrict the upper *m/z* limit reachable by the second mass analyzer. In fact, despite the undeniable benefits in terms of ion selection and extended ion activation capabilities granted by linear or 3D traps, their *m/z* ranges do not generally exceed 4-5000 *m/z* (specifically, both LTQ XL and Velos Pro have a fixed upper limit at 4000 *m/z*). Notably, ICR and Orbitrap alone are capable of a much more extended *m/z* range.

Usually less stringent *m/z* limitation is present for quadrupoles when operating as mass filters. The maXis is accredited of an upper *m/z* limit of 20000, but practically, similarly to ion traps, for the quadrupole an increase in the allowed upper limit is generally accompanied by a corresponding increase in the minimal *m/z* detected with good sensitivity.

The mass accuracy of all the three high resolution mass analyzers described above is at the state-of-the-art. Particularly, high field Orbitrap and Ultra ICR cell reach ~1-2 ppm mass accuracy with external calibration, and can further increase the accuracy to parts-per-billion with internal calibrants. Similarly, the mass accuracy achieved by the maXis UHR qTOF is between 1 and 5 ppm with external calibration, and goes to sub-ppm values by using a mass-lock.

2.3.3 Tandem mass spectrometry

After mass-to-charge and, therefore, accurate mass determination of the polypeptide of interest in MS^1, tandem mass spectrometry (MS^2 or MS/MS) adds a second dimension to the analysis by fragmenting selected precursor ions. Originally, the term "tandem" was used as in the early days of mass spectrometry the (i) precursor selection, (ii) fragmentation and (iii) product ion detection steps were performed in different mass analyzers/collision cells; the most typical example in this sense is represented by triple quadrupole instruments. In this case, MS^2 is named *in space*, as the different operations take place in different regions of the mass spectrometer. When all the steps occur sequentially in the same mass analyzer, as for instance in the case of Paul or linear ion traps, the MS^2 process is defined as *in time*.[7]

Sequencing a polypeptide implies the cleavage of the bonds forming the backbone. More in detail, for facilitating the identification of the backbone cleavage site a single cleavage should occur, such that each polypeptide ion is split into two fragments, one containing C- and another the N-terminal portion of its sequence. Additional backbone cleavages might lead to the production not only of smaller fragments containing N- and C-termini, but also internal fragments which are difficult to characterize and complicate the tandem mass spectrum. Other cleavages such as neutral losses (e.g., the loss of H_2O, CO, CO_2 etc.) or losses of side chain groups can be useful for confirming the putative amino acid sequence or eventual PTM position at the cost of more convoluted MS^2 spectra. The nomenclature used for defining the product ions

generated in MS^2 of polypeptides, introduced in the currently accepted version by Biemann, takes into account the possibility of cleavage at all the three backbone bonds (i.e., N-C_α, C_α-C and C-N), Figure 2.11.[57]
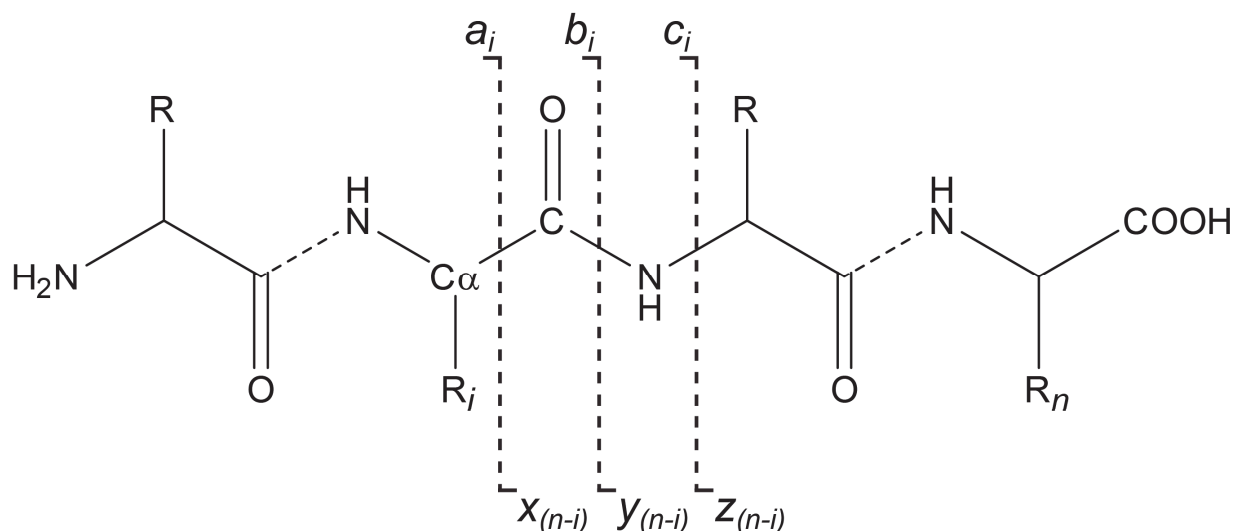
Figure 2.11. Scheme of product ions produced by MS^2 for a polypeptide parent ion with n residues, fragmented at the C-terminal side of the ith amino acid residue. Product ions are the N-terminal a-, b- and c-type ions, and complementary C-terminal x-, y- and z-ions.

Besides the practical differences that can be observed in experiments with MS^2 performed in time or in space, and which will be discussed in a separate section at the end of this Chapter and specifically referred to the mass spectrometers used for this Thesis, the different fragmentation techniques applied to polypeptides can be classified on the basis of the applied mechanism for backbone cleavage. Two main groups of ion activation techniques can be defined: that based on vibrational bond activation, and the radical-driven one. Hereinafter some of the most widely employed ion fragmentation techniques are briefly described.

2.3.3.1 Vibrational activation-based tandem MS

The MS^2 techniques belonging to this group exploit interaction of ions with either IR photons or neutral molecules to increase ion internal energy until the fragmentation threshold is reached.

Generally these fragmentation techniques lead to the cleavage of the weakest bonds in the polypeptide, eventually including even those present at side chains of amino acid residues.

- Collision-induced fragmentation

Ion trap devices (namely 3D or linear ion traps and quadrupoles/multipoles) filled with neutral gas molecules, such as N_2 or He, can be used for ion activation by inducing collision of the ions themselves with neutrals. In *collision-induced dissociation* (CID), the most widely employed ion activation technique in MS-based molecular structure analysis, including proteomics, ions' kinetic energy is first increased by the application of ramping RF potentials, with the consequence that the ions' velocity is increased and with that also the frequency of inelastic collisions with neutrals. As these collision are progressively more energetic with the increase in ions' kinetic energy, the fragmentation threshold is finally reached.[58]   In *higher-energy collision dissociation* (HCD; formerly called *higher-energy C-trap dissociation*, from the bent multipole where it was originally performed) the fragmentation mechanism is similar, but the RF potentials are replaced with a high axial potential applied along the gas-filled multipole that acts as collision cell.[59] Both CID and HCD induce mainly the cleavage of the peptidic bond (C-N), resulting in the formation of complementary *b*- and *y*-type ions. In HCD, due to the higher energy of the process, not only the used ion activation time is generally much lower than in CID (typically, 0.1-1 ms against the 10-30 ms used in CID), but also the obtained product ions differ from CID, as *b*-ions, less stable than *y*-ions, are a minority and ultimately HCD returns mostly *y*-ions.

- Infrared multiphoton dissociation.

Ions' internal energy can be increased also by absorption of infrared (IR) photons. As the single IR photon absorption leads to a small increment in ion internal energy, multiple photons have to be absorbed to increase the vibrational energy of an ion to the bond dissociation threshold. Therefore this dissociation technique is termed *infrared multiphoton dissociation* (IRMPD).[60] Specifically, the FT-ICR mass spectrometer is equipped with a 10.6 µm CO_2 laser, such that each photon absorbed accounts for a 0.117 eV increment in ions' internal energy. This energy is typically lower than that gained through the collision of an ion with a neutral, with the consequence that the activation time in IRMPD is generally longer than in CID/HCD.

Nevertheless, IRMPD can lead to the peptide bond fragmentation (i.e., yielding *b*- and *y*-type ions) under both UHV conditions, like in the case of an ICR cell, and in the higher pressure linear ion trap. The internal energy relaxation process due to slow ion collisions with neutral gas molecules are to be overcome in the latter case. Furthermore, the IR laser can be used for ion pre- or post-activation, by irradiating ions with a proper combination of irradiation time-laser power that induces an increase in ion internal energy without reaching the dissociation limit.

2.3.3.2 Radical-driven tandem MS

In contrast with activation methods based on the increase of ions' vibrational energy, other activation techniques consist in the formation of a radical species within the poplypeptide backbone followed by backbone rupture. Several techniques can be used for the generation of radical ions, however the most commonly employed ones are based on the acquisition by a biomolecular multiply-charged cation of an electron. Particularly, the polypeptide fragmentation following the capture of a low energy (~1eV) electron is named *electron capture dissociation* (ECD),[61] whereas the fragmentation occurring after the transfer of an electron from a radical anion to a polypeptide cation is termed *electron transfer dissociation* (ETD).[62]

ECD was discovered in 1998, whereas ETD was introduced six years later. The electron-cation reaction on which ECD is based can be expressed as follows:

$$[M + nH]^{n+} + e^- \rightarrow [M + nH]^{(n-1)+\bullet} \rightarrow fragments \qquad \text{(eq. 2.7)}$$
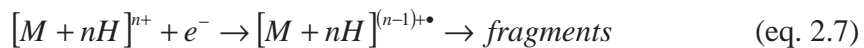
Similarly, the ion-ion reaction characteristic of ETD is written as:

$$[M + nH]^{n+} + A^- \rightarrow [M + nH]^{(n-1)+\bullet} + A \rightarrow fragments \qquad \text{(eq. 2.8)}$$
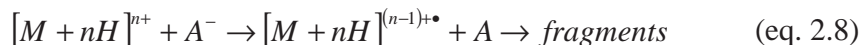
Where "A" represents the electron donor molecule.

The mechanism underlying ECD is still under debate. The so-called "Cornell mechanism" postulates the electron attachment to occur at one protonated nitrogen of a basic amino acid residue, followed by electron relaxation with subsequent hydrogen transfer to the amide group, forming a radical aminoketyl group. The formation of this group finally leads to the fragmentation of the backbone. To explain phenomena observed in ECD and not accounted by this mechanism, the "Utah-Washington mechanism" was then introduced. This mechanism

hypothesizes the capture of the electron directly on the π* orbital of the amide group, with formation of the radical aminoketyl species followed by fragmentation. Both the mechanisms, as well as others suggested recently, agree on the non-ergodic nature of the process, that implies that the fragmentation occurs rapidly after the electron capture event, so that the energy acquired by the electron capture itself is not redistributed among all the vibrational levels of the polypeptide cation. A similar scenario is suggested for ETD, which energetically is an even softer process than ECD.

It is apparent how the charge annihilation subsequent to the electron capture/transfer event prevents the possibility of using singly charged precursors for electron mediated tandem MS. From an analytical chemistry point of view, the fragmentation efficiency of ECD and ETD (expressed as ratio between overall abundance normalized to the charge state of product ions over abundance normalized to the charge state of the selected precursor ion) is a function of the charge state of the precursor ion. Conversely, the duration of the process (i.e., the electron-ion interaction period in ECD or the ion-ion interaction period in ETD) is progressively reduced by increasing the precursor charge state. Both the fragmentation techniques lead to the fragmentation of mainly N-C$_\alpha$ bond, with formation of even-electron c'-ions and odd-electron radical $z^\bullet$-ions. Depending on the lifetime of the $[c'+ z^\bullet]^{(n-1)+\bullet}$ complex, hydrogen rearrangement can take place resulting in the presence in the final spectrum of $c^\bullet$ and z'-ions. For the sake of simplicity, N- and C-terminal product ions obtained by N-C$_\alpha$ bond cleavage will be referred to as c- and z-type ions. Note, that the cleavage at the N-terminal side of proline cannot lead to product ion formation due to the structure of this amino acid.

Experimentally, ECD and ETD are performed under very different conditions. The former is commercially enabled uniquely in FT-ICR instruments, where cations interact with the electron beam produced by a cathode positioned close to the ICR cell under UHV conditions. Implementations of ECD in other types of mass spectrometers, specifically ion traps and TOF,[63, 64] although successful, have never been adopted on commercial instruments, and are replaced by more efficient ETD implementations. ETD, which requires the simultaneous trapping of precursor cations and radical anions, is performed in ion trap mass analyzers. The previously described maXis UHR qTOF and the LTQ Orbitrap of both Velos Pro and Elite series are ETD-enabled mass spectrometers. The former uses the collision cell for trapping cations and anions,

both injected from the front end of the mass spectrometer, whereas the LTQ Orbitrap instruments are designed with radical anion injection from the back end followed by ETD in the linear ion trap.

ECD and ETD are classically the ion activation techniques of choice for analysis of complex polypeptides or protein top-down analysis as their figures of merit include: (i) the capability of cleaving disulfide bridges;[65] (ii) the retention of labile PTMs, which are normally lost using collision-based ion activation of peptide ions;[66, 67] (iii) an extended sequence coverage with cleavage sites that are not limited to the N- and C-terminal ends of the protein, but are rather the function of charge location.[68]

2.3.3.3 Further considerations on MS^2 in a top-down approach.

In the tandem mass spectrometry of a protein, i.e., a polypeptide containing more than 50 amino acid residues, a large number of fragmentation channels are open, not only due to the length of the peptide chain, but also because electrospray-generated ions are multiply charged and, therefore, a single cleavage site can be identified by complementary b- and y-type, or c'- and $z^{\bullet}$-type ions present in different charge states. As a consequence, the population of each product ion is relatively limited in top-down MS. Therefore, differently from bottom-up MS, in TD MS the application of strategies aimed at maximizing the SNR of the measurements is inalienable. As previously mentioned, this parameter is important as it allows the determination of the monoisotopic peak and, moreover, high SNR peaks are characterized by improved mass accuracy. The typical strategy used to increase SNR consists in averaging a number of tandem mass spectra. For TOF mass analyzers, spectra can be averaged directly, eventually requiring recalibration whether they were obtained from different experiments. Differently, in FTMS what is averaged, following sophisticated procedures, are time-domain transients. This is necessary in the case of data acquired in separate experiments, but the same strategy is available in the commercial instrument software control, which can set a number of so-called microscans, corresponding to transients that are averaged before FT conversion. At a practical level, although no significant difference in terms of duty cycle or data acquisition rate is apparent between summing an x number of already FT-processed mass spectra and using the same x number of microscans in a single scan, the latter strategy results in a slightly higher SNR improvement than

the former.[69] It is useful to recall that the theoretical SNR improvement in the case of transient summation prior to Fourier transformation is proportional to the square root of the number of summed transients.[70]

Another aspect that is important to consider is the role played by ion transmission in tandem mass spectrometry in space. As mentioned before, to this category belongs ETD, which is the activation technique on which most of the studies of this Thesis are based. While in TOF-based instruments the transmission of ions occurs without a substantial preference for low mass or high mass ions, in hybrid Orbitrap-based instrument there are instrumental parameters capable of affecting deeply the results of transmission from the first mass analyzer to the Orbitrap. As exemplified in Figure 2.12, which shows ETD mass spectra of a ~23 kDa protein, myoglobin, recorded in the Orbitrap, the gas pressure at the HCD cell, which is proportional to the gas pressure in the C-trap (as the gas is leaking from the HCD cell to the C-trap), and here is measured as the difference in pressure between gas switched on and off (i.e., the "Δp"), determines the quality of transmission of heavy, highly charged ions. It is apparent how, by reducing the gas pressure, the intensity of remaining precursor and of charge-reduced species is increased. Contemporary, as shown in the insets of each panel, the SNR of large, highly charged product ions such as c_{67}^{10+}, which has a mass of ~7.4 kDa, becomes progressively higher. On the contrary, the SNR of light product ions can be reduced. For a protein of the size of myoglobin the pressure reduction is generally beneficial, but for larger proteins, whose fragmentation generates in general a larger number of product ions, and specifically a high number of heavy fragments, favoring too strongly the transmission of large ions can be counterbalanced by the loss of information on the edges of the polypeptide chain.
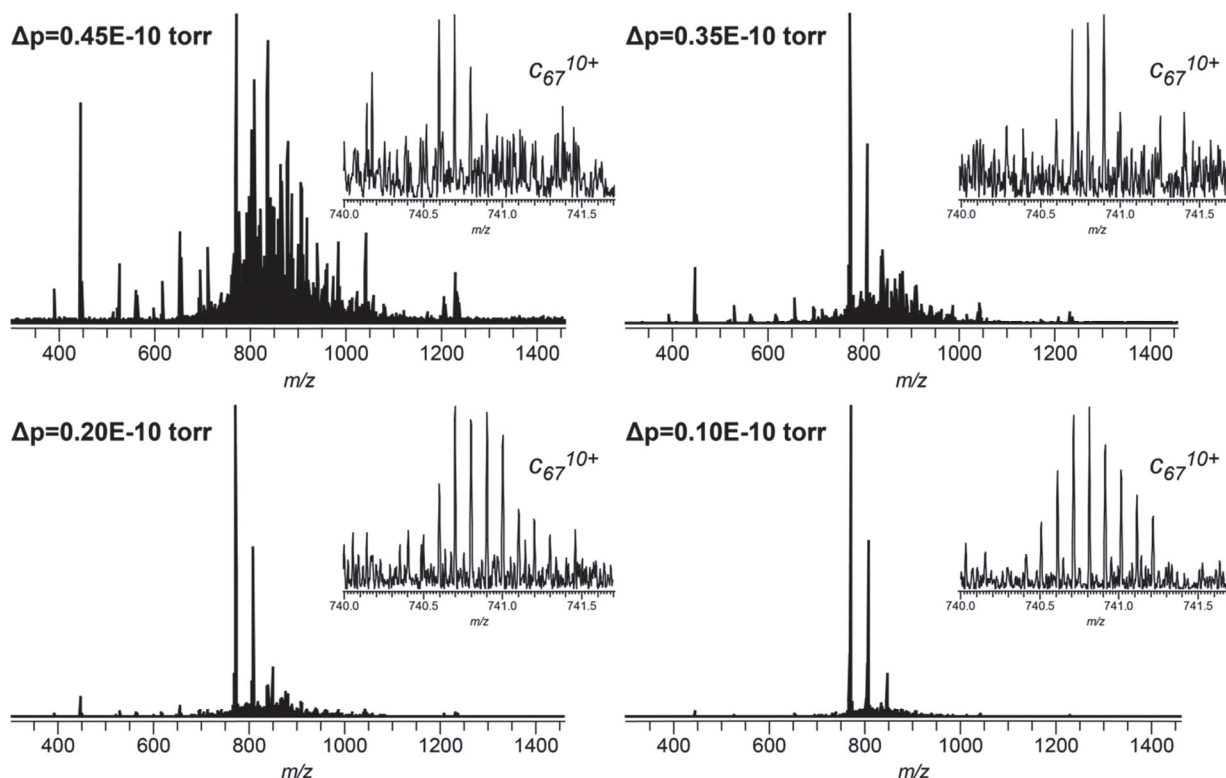
Figure 2.12. Orbitrap ETD MS^2 of horse myoglobin, precursor 22+. The four panels represent the broadband mass spectra obtained with different pressure of collision gas (N_2) in the HCD chamber. Mass spectra are obtained by averaging 2 scans recorded at 120'000 resolution at 400 m/z, each scan being composed of 10 microscans. Insets show the c_{67}^{10+} ion in each mass spectrum. Differences in SNR are visually apparent.

2.3.4 Data analysis

The analysis of large polypeptides is generally not performed manually, although manual validation of results is often advised. Essentially, for large polypeptides like those typical of MD and, particularly, TD proteomic experiments, two operations are required, in the following order: (i) spectral *deconvolution* and (ii) *product ion assignment*.

Deconvolution of a mass spectrum (either MS^1 or MS^2) is the operation through which ion clusters having different charge states are all transformed in their neutral or singly charged counterparts. Sometimes, depending on the software in use, this operation includes also peak centroiding. Software used for this Thesis include Maximum Entropy for processing Bruker BAF

files and Xtract,[71] THRASH[72] and MS-Deconv for Thermo Scientific RAW files. Deconvoluted spectra were processed for product ion assignment with either ProSightPC (Thermo Scientific) or MS-Align+.[73]

Deconvolution is a fundamental operation as the assignment of product ions, i.e. the determination of cleavage sites, depends on its accuracy. Therefore, the efficacy in determining the charge state of an ion and the possibility of including a selected peak into an isotopic envelope distinguish the different available deconvolution algorithms. The problem is not trivial. As an example of the solutions introduced to remedy its complexity, the "*averagine*" concept used in THRASH can be cited. Averagine is defined as an "amino acid with weighted average mass" and it was firstly used for charge state determination in high resolution FTMS spectra.[74] THRASH applies the average amino acid concept for simulating a charge state distribution of a theoretical ion that resembles in mass the one under analysis. The match between this theoretically simulated distribution and the experimentally obtained one will be measured, and a threshold value on the similarity score can be applied to filter out spectra peaks prior to the subsequent product ion assignment process. Obviously, the results obtained by this method are heavily dependent on ion statistics, as a poor ion population will generally result in distorted isotopic distributions.[18] Finally, hydrogen transfer between c- and z-type ions (i.e., resulting in $c^{\bullet}$- and z'-type ions) in ETD experiments, or simple misassignment of the monoisotopic peak (usually mismatched with the first ^{13}C peak of the isotopic distribution) by the deconvolution algorithm should be taken into account during validation of identified protein cleavage sites. In targeted top-down analysis, a partial solution for the problem is to search for ions which differ in mass from the theoretical ones ±1.003 Da (^{13}C misassignment) or 1.007 Da (hydrogen transfer). The latter event is not often observed in top-down spectra due to the high charge states of product ions involved and the subsequent high coulombic repulsion (i.e., short lifetime of the $[c'+ z^{\bullet}]^{(n-1)+\bullet}$ complex). Unfortunately, manual validation of top-down spectra is not used when this methodology is used for large scale studies (e.g., proteomics). Description of the strategies used for reducing the misassignment of monoisotopic peaks in top-down spectra can be found in Chapter 6, Papers V and VI.

Chapter 3: ECD high resolution MS in the study of peptides

This Chapter is focused on the investigation of two post-translational modifications, namely deamidation and transamidation, by electron capture dissociation and Fourier transform - ion cyclotron resonance MS. Hereinafter reported considerations are based mainly on this published research article, enclosed in Chapter 6:

- Deamidation and transamidation of substance P by tissue transglutaminase revealed by electron-capture dissociation Fourier transform mass spectrometry (*paper I*).

3.1 Relevance of non-widely studied PTMs in contemporary biology and biotechnology.

For reasons related to both their undisputable biological relevance and also the relative ease in identification via MS-based analysis, a restricted number of PTMs has historically been the center of proteomic studies. To this list belong phosphorylation,[75] methylation[76] and acetylation.[77-79] Biologically, reversible phosphorylation represents the main way to modulate enzymatic activity, protein-protein interactions and other signal transduction processes; methylation and acetylation play important roles in determining the functionality of relevant classes of proteins such as histones and tumor suppressor proteins. From an analytical chemistry point of view, these three PTMs induce a pronounced mass shift in the modified residues, and although proteomic research is mainly based on collision-based ion activation techniques, which do not always preserve labile PTMs, neutral losses can be used for localization of these modifications. On the other hand, confident identification of modifications introducing small mass variations in an amino acid residue requires high mass resolution and accuracy MS. This is the case, for instance, of arginine citrullination[80] and the hereinafter described glutamine and asparagine deamidation. Deamidation converts Gln and Asn in glutamate and aspartate, respectively, with a mass shift of +0.984 Da. This implies that deamidation replaces a polar amino acid with a charged one, with potential structural consequences on the involved protein. Importantly, deamidation occurs via the formation of a cyclic intermediate, so that the final product can be a structural isomer of Glu and Asp, like γ-glutamic acid or β-aspartic acid (also known as iso-Asp), respectively (Figure 3.1).[81]

Notably, deamidation is a spontaneous phenomenon believed to target proteins for degradation,[82] but it can also be enzymatically induced. Tissue transglutaminase (tTGase) is an

enzyme that can deamidate glutamine residues. It has been reported to play a major role in celiac disease and it is suspected to contribute to the onset of neurodegenerative pathologies such as Alzheimer's disease and Parkison's disease.[83, 84] However, as its name indicates, tTGase main catalytic activity is transamidation, a reaction through which this enzyme can form a new covalent bond between the side chains of Gln and Lys, introducing protein *cross-linking*. Therefore, an *in vitro* study based on the activity of tTGase on model peptides can serve to investigate the potential of high resolution MS not only in elucidating a specific enzymatic activity but also in future application to biotherapeutics quality control.

3.2 Radical-driven ion activation as a tool for deep characterization of deamidation and transamidation.

Although the evidences of deamidation and subsequent amino acid rearrangements leading to the formation of structural isomers were first detected in MS-base studies applying collision-induced ion activation, electron capture dissociation and electron transfer dissociation introduced new possibilities for detailed analysis of deamidation products. Specifically, ECD has been used for the quantification of deamidation in O_2 labeled peptides,[85] discrimination of α-glutamic acid and γ-glutamic acid[86] and finally the differentiation of α-aspartic acid from β-aspartic acid, even at a proteome-wide scale.[87, 88] ETD was instead applied to the differentiation of α-Asp from β-Asp.[89] The distinction between amino acid isomers is achieved through the observation of specific reporter ions that can be identified in the ECD/ETD tandem mass spectrum only in presence of non-α amino acids.
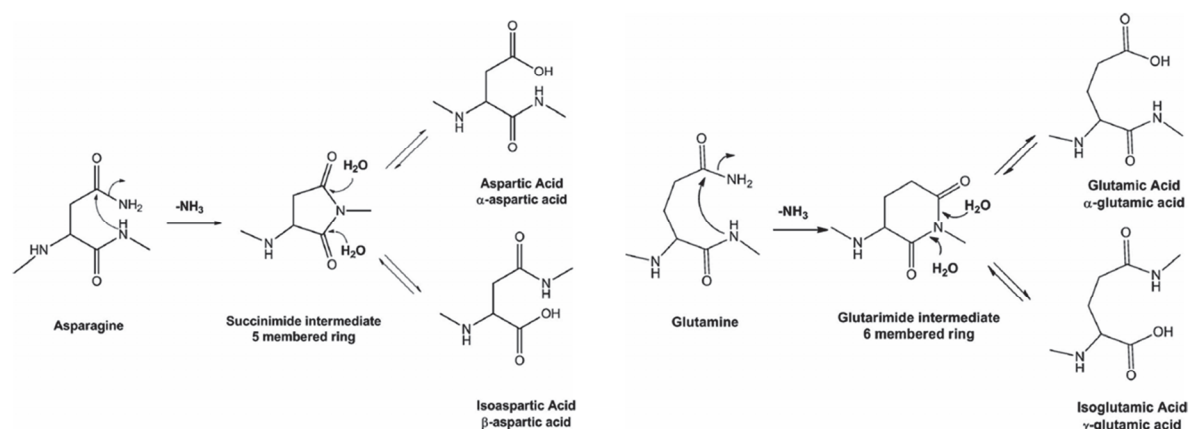
Figure 3.1. Pathways leading to the formation of structural isomers of Asp and Glu as a consequence of deamidation of Asn and Gln, respectively. Left panel shows the formation of a succinimide intermediate with possible production of either α-Asp or β-Asp. Right panel illustrates the formation of α-Glu or γ-Glu from a glutarimide intermediate. Schemes derived from [87] and [86]

Despite the potential demonstrated by ETD and ECD in resolving structural amino acid isomers, their application has been so far limited uniquely to peptides, with the notable exception of the top-down study on β2-microglobulin by Li *et al.* Even in this case, however, ECD was used as MS^3 on an isolated CID-produced fragment of β2-microglobulin, and not directly on the intact protein, despite its relatively small molecular weight of ~11.7 kDa.[90] This methodology, in other words, relies on low abundant ions that are difficult to detect in convoluted top-down tandem mass spectra. Extending the mass range of precursor ions for which the isomerization is detectable by ECD/ETD MS is one of the future goals for applying routinely this methodology to the analysis of biotherapeuticals.

Results published in the research article about deamidation and transamidation of the undecapeptide substance P show no evidence of isomeric rearrangement of Gln. Nevertheless, in this study the intriguing and at the same time challenging aspect is represented by the *competition* of two mechanisms, deamidation and transamidation, both possible for this peptide (which contains within its sequence both Gln and Lys residues) and that might use as a substrate two Gln residues, consecutively placed along the peptide sequence. Tandem mass spectra generated by collision-induced dissociation generally present side chain losses, and furthermore CID generally

cleaves the amide bond at the backbone (C-N bond) and potentially other labile bonds such as those of several PTMs. As illustrated in Figure 3.2, transamidation leads to the formation of an amide bond interconnecting the side chains of glutamine and lysine.
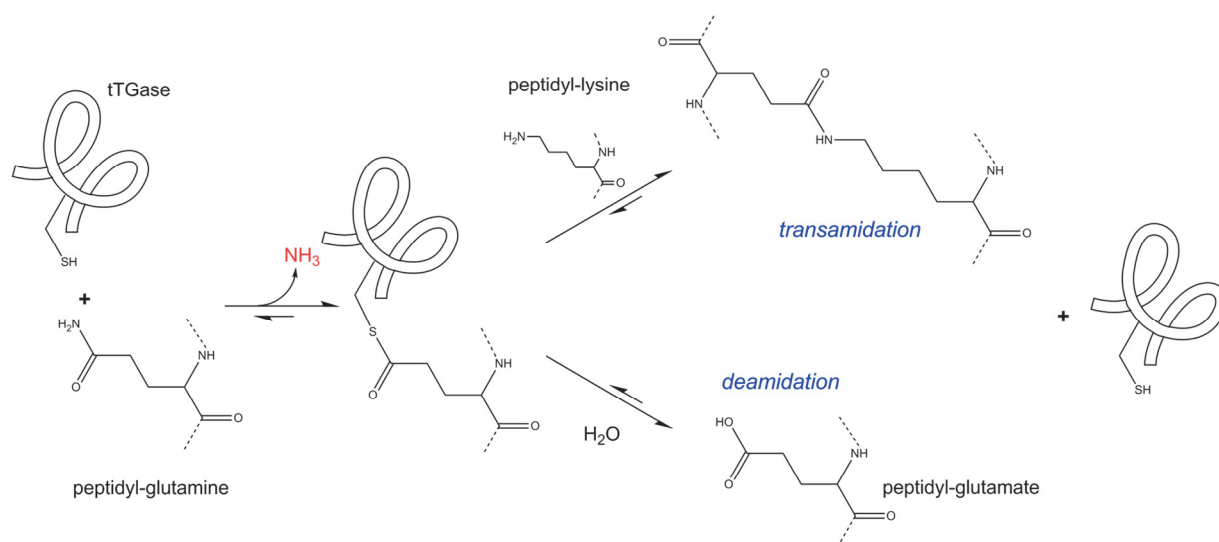
Figure 3.2. Schematics of deamidation and transamidation catalyzed by tissue transglutaminase.

The combination of high resolution MS and ECD is fundamental for elucidating the order in consecutive deamidation of both the Gln residues present in substance P, as well as for determining the position of the Gln residue involved in transamidation without rupture of the newly formed bond.

The experience reached through this proof-of-principle study will be important for future applications on biopharmaceuticals and, particularly, organic molecule-labeled biotherapeutics.

Chapter 4: High mass accuracy MS for biotechnological applications

After the description of results achieved by the investigation of post-translational modifications at the peptide level (Chapter 3), and before introducing top-down and middle-down MS approaches for structural characterization of large proteins, a logical interlude is represented by the application of different MS-based approaches, from intact mass measurement to bottom-up MS, for the *qualitative* and *quantitative* study of a typical biotherapeutic, an IgG1, an its major PTM, N-glycosylation. The information and experimental results on which the present Chapter is based derive from the following peer-reviewed research article:

- Glycan variability on a recombinant IgG antibody transiently produced in HEK-293E cells

4.1 N-linked glycosylation in biotechnology and mass spectrometry.

Glycosylation in proteins is differentiated between N- and O-linked glycosylation, the former referring to oligosaccharidic chains linked to Asn residues, the latter indicating short sugar chains (generally mono-, di- or tri-saccharides) covalently bond to Ser and Thr residues.[91-94] For most biotherapeutics, N-linked glycosylation is the most relevant as it usually deeply affects the biological activity and function of the protein. Erythropoyetin, follicle-stimulating hormones and antibodies are some examples of glycosylated biopharmaceutical proteins.[95, 96] Furthermore, N-glycosylation also conditions the half-life of antibodies in serum (*vide infra*).[97]

From a structural point of view, N-linked oligosaccharides are relatively long and branched heteropolymers, with multiple linkage possibilities among single monomers; therefore, their MS-based analysis is performed by limiting the size of the analyte, either by releasing the glycans from polypeptides or by proteolytically digesting the glycoprotein into peptides. Enrichment or purification procedures are hence needed prior to MS experiments. In principle, a complete structural analysis of glycans, released or associated to short peptides, is composed of MS^1 followed by MS^2. Product ions produced by fragmentation of released sugars have a specific nomenclature that accounts for all the possible bond ruptures observed in MS^2.

Interestingly, different ion activation methods lead to complementary information: in the case of released glycans, vibrational-based (e.g., CID or IRMPD) and electron-based (e.g., ECD or ETD) fragmentation techniques are known to cleave different bonds in the oligosaccharidic chains.

Similarly, the analysis of glycopeptides by CID or IRMPD generates primarily fragmentation of the sugar moiety, whereas ETD and ECD mainly conserve the PTM intact and produce extended cleavage of the peptide backbone.

In summary, it is apparent how the biological and biotechnological relevance of N-glycosylation is matched by the complexity of MS-based glycan analysis.

4.2 Conceptual and practical relevance of the obtained results

In principle, although chemistry alone would allow an extremely large number of different combinations from a reduced number of possible initial monomers, actual biological processes are normally limited to a reduced fraction of the theoretical possibilities. This is the case for N-linked glycans, which exist in a restricted number of combinations.

This consideration enables creation of searchable databases, e.g., GlycoPep DB,[98] which group known glycopeptides according to the protein source, i.e., their expression system. N-linked glycosylation is, in fact, a multistage process which requires different enzymes to be accomplished.[31] Type and stoichiometry of these enzymes depend on the source organism.[99] As a consequence, when high mass accuracy MS is used, within a small mass tolerance (i.e., of a few ppm) in general only one possible glycan structure with biological relevance can be associated to the monoisotopic mass of a glycopeptide of interest. Obviously, only tandem mass spectrometry can confirm linkage and branching between monomers, as GlycoPep DB returns only a list of monomers.

On these bases, high mass accuracy MS was therefore applied, using an FT-ICR mass spectrometer, to perform identification and subsequent label-free quantitation of tryptic glycopeptides obtained from ten batches of an IgG transiently expressed in HEK293E cells, the final goal being the batch-to-batch comparison of IgG glycosylation. On-line LC-MS of tryptic peptide was employed, with separation of modified peptides carried out with a special variant of traditional reversed phase chromatography. Traditional aliphatic chain-based stationary phases (e.g., the so-called C4, C8 or C18) poorly bind glycopeptides, whereas porous graphitized carbon columns efficiently retain peptides modified with a large hydrophilic moiety such as glycopeptides.

Although the composition of single oligosaccharidic chains derived from tryptic proteolysis of the IgG of the ten batches already provided a good estimation of the similarity of overall IgG glycosylation, a final confirmation in this sense could be inferred only from glycoforms, that is by IgG intact mass measurements. High resolution qTOF MS was therefore employed for obtaining a qualitative estimation of the similarity in glycoform abundances for the recombinant IgG expressed in the ten batches.

In conclusion, high mass accuracy and resolution MS, proposed in its different possible uses, elegantly served the purpose of answering an intriguing applied biology question, leading to the validation that transient expression of IgG with a robust protocol can effectively result in the production of batch-to-batch consistent IgG glycoforms.

Chapter 5: ETD-based top-down and middle-down MS of large proteins

This Chapter is dedicated to the ETD top-down and middle-down MS analysis of large proteins. Specifically, the focus is on the structural characterization of ~150 kDa monoclonal antibodies. The present summary is based on the experimental results reported in the following research studies included in Chapter 6:

- Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry (*paper III*).
- Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry (*paper IV*).
- Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS (*paper V*).
- Top-down analysis of immunoglobulins G with electron transfer dissociation high-field Orbitrap FTMS (*paper VI*).
- Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap FTMS (*paper VII*).

The reasons for thinking of the structural analysis of intact antibodies as the acme of a mass spectrometry-based research aimed at characterizing large proteins in a top-down fashion are intrinsically linked to the complexity of these biomolecules, as described hereinafter (Section 5.1). Preliminary steps evaluating the feasibility and efficiency of the analysis of large proteins by electron transfer dissociation MS are also illustrated (Section 5.2). Finally, results of different TD MS analyses of antibodies are collectively analyzed (Section 5.3), for outlining current achievements and limitations, and proposing therefore possible alternatives or complementary approaches that will not completely sacrifice the theoretical advantages of TD MS, such as middle-down MS (Section 5.4).

5.1 Monoclonal antibodies.

Antibodies or *immunoglobulins* (Igs) are proteins playing an important role in the immune system. Produced by B cells, they are characterized by a quaternary Y-shaped structure composed by two identical light chains, and two identical, N-glycosylated heavy chains.[100]

Igs primary function is the recognition and binding of a specific molecule, known as an antigen. Each region formed by the combination of the N-termini of light and heavy chains (i.e., the superior tips of the "Y") is called a paratope which selectively binds a specific portion of the antigen, or epitope. *Monoclonal antibodies* are produced by a single B cell clone and, being all identical, bind exclusively a single epitope on the antigen. *Polyclonal antibodies*, in contrast, represent a mixture of Igs produced by different clones and are generally active against different epitopes on the same antigen.

Among immunoglobulins we distinguish different classes or isotypes, differentiated by the kind of heavy chain they present. In placental mammals we recognize 5 isotypes: IgA, IgD, IgE, IgG and IgM. IgGs are the most common in human blood and the most studied. Immunoglobulins G are further distinguished in 4 subclasses, named with numbers in decreasing order of abundance in human blood.

The IgG subclass is the object of research projects presented in this Thesis also due to their increasing relevance as a biotherapeutic, i.e., as a biotechnological drug. Particularly, most of the IgGs already approved by the Food and Drug Administration and available on the market belong to the IgG1 subclass, and only the IgG3 subclass is not targeted as potential biotherapeutic molecule at the moment.[101]

If we take the structure of an IgG1 as reference (structural differences among the subclasses are limited), two characteristics are immediately clear, Figure 5.1. First, the presence of so called *"constant" and "variable" domains* in both light and heavy chains (indicated by the letters "V" and "C", respectively, followed by the initial of the chain where they are located and, eventually, a progressive number). Secondly, a high number of *disulfide bonds* (sixteen for IgG1 and IgG4, higher number for IgG2 and IgG3). These bonds are either *inter-molecular*, connecting each light chain with one of the heavy and the two heavy chains between themselves, or *intra-molecular*. As mentioned above, the regions responsible for the high specific and selective binding of an IgG to its antigen are the tips of the "Y" structure, where the variable regions are located. On each variable region of each chain there are three short sequences (i.e., generally not longer than 15 amino acids, and often much shorter) known as *complementarity determining regions* (CDRs). The folding of the IgG, with heavy and light chain bound together and exposing the corresponding CDRs, forms the antigen binding domain. From a genetic point of view, the hyper

variability of CDRs, required for producing an extremely high number of different antibodies against all possible antigens, finds its origin in a phenomenon called "V(D)J recombination", which is a DNA rearrangement process that randomly combines *V*ariable, *D*iverse and *J*oining genes, originally separated in the genome, in a unique one before its translation in mRNA.[102]

Another common way for classifying regions of an IgG is based on the proteolytic action of specific enzymes such as papain. This protease cleaves the IgG above the junction between the two heavy chains (so-called *hinge region*), producing two identical "fragments antigen binding" (Fab), each composed of one light chain and the N-terminal portion of one heavy chain (also known as Fd), and a "fragment crystallizable" (Fc), which biologically is important for binding the IgG to its receptor and structurally is a dimer of two identical C-terminal portions of the heavy chain (called Fc/2).
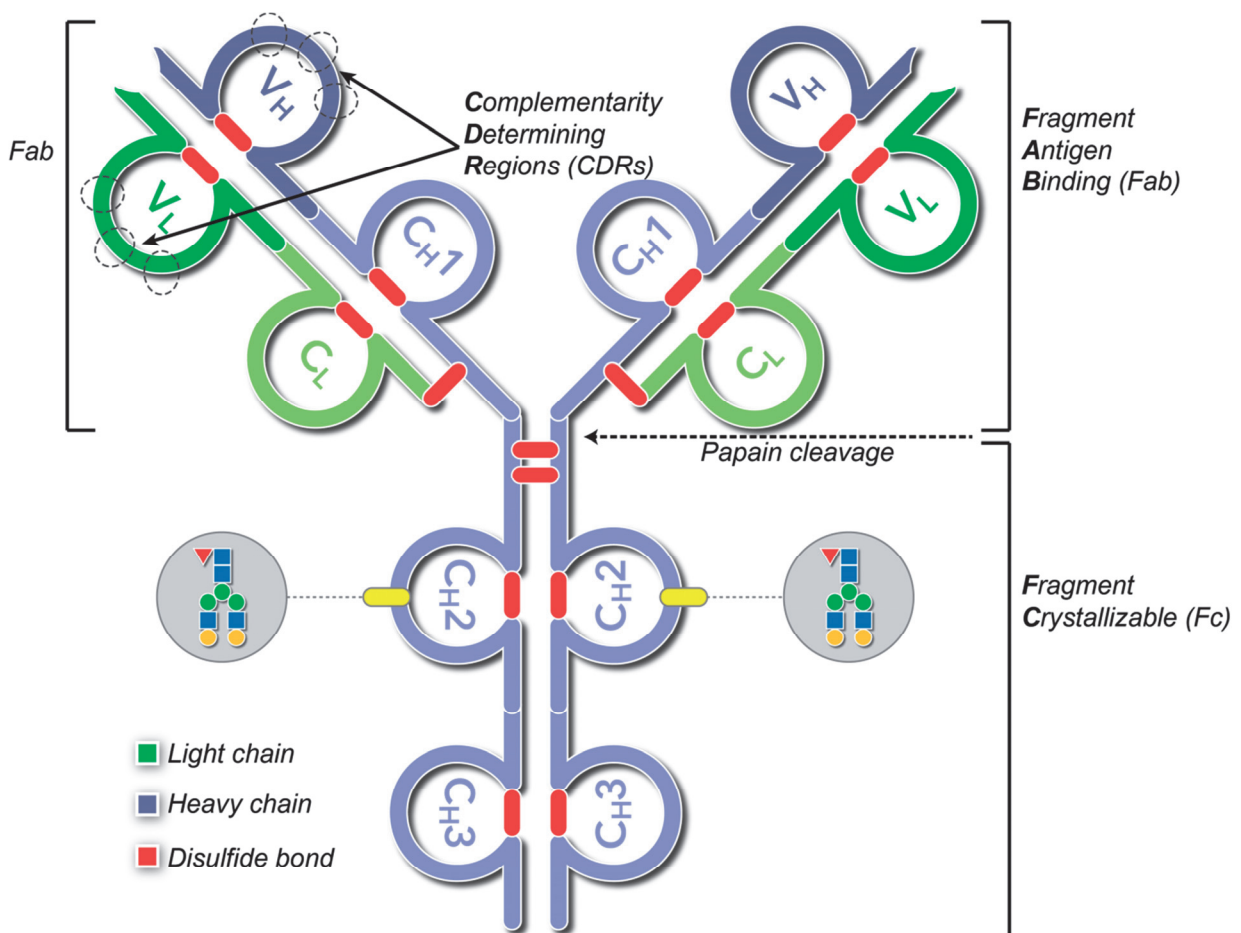

Figure 5.1. Schematic representation of an immunoglobulin of the G class (isotype: IgG1).

The half-life of immunoglobulins G in blood is very high, reaching or even exceeding 30 days. This is mainly due to their extremely robust and compact structure. Besides the already mentioned hinge region, the only highly flexible areas in an IgG correspond to the solvent exposed loops interconnecting consecutive variable or constant domains, characterized by the presence of polar and basic amino acid residues. Each domain is, in fact, highly structured and constitutes a compact configuration known as *immunoglobulin domain*.[103] Observed also in other classes of proteins than antibodies, these domains are characterized by the presence of seven to nine beta strands organized in two antiparallel beta sheets. In IgGs, the rigidity of the immunoglobulin domain is enhanced by the presence of an intra-molecular disulfide bridge, Figure 5.2.
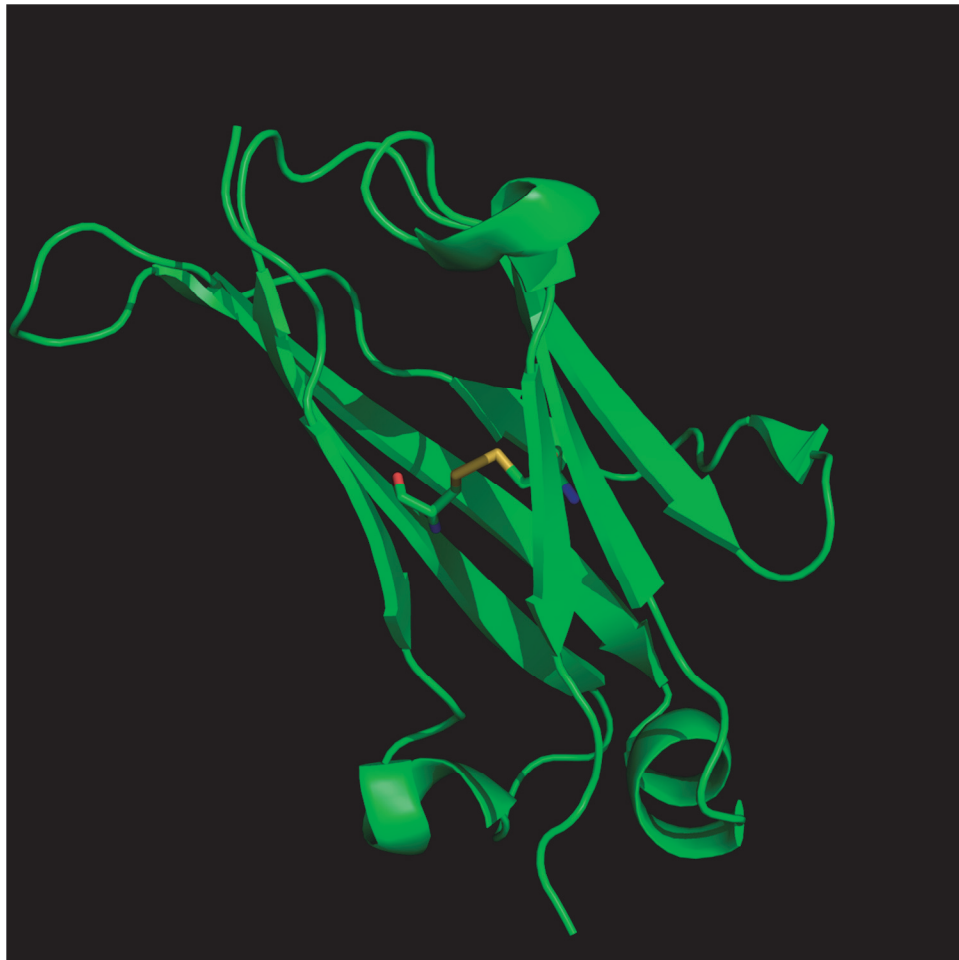

Figure 5.2. Three-dimensional structure of an immunoglobulin domain. The disulfide bond is highlighted. Image from crystallographic data of the IgG1 Trastuzumab (PDB entry: 3D6G).

The role of mass spectrometry in the characterization of immunoglobulins is fundamental for different reasons. First, the size of these proteins complicates the analysis with other traditionally employed techniques such as gel electrophoresis. Secondly, these antibodies can carry several different post-translational modifications, whose masses are always very small compared to the overall size of the intact protein. Among the most common PTMs we find: N-glycosylation (+>1000 Da), which plays a role in antibody-dependent cell-mediated cytotoxicity and in complement-dependent cytotoxicity;[104] clipping of C-terminal Lys;[105] and block of the N-terminus due to conversion of Gln or Glu to pyroGlu.[106] Aside of these PTMs involving the chain termini or the Fc region, others can be found in the CDRs where they can effectively affect or even impair the binding capability of the IgG. Among these we list: deamidation of Asn and, less commonly, Gln (+0.984 Da); isomerization of Asp to iso-Asp; racemization of Asp; as well as oxidation of Met and Trp (+16 Da).

Finally, it is absolutely important to confirm the Ig sequence obtained by genomic data, particularly for the CDR domains, given the high variability of these sequences.

Traditionally, detailed characterization of Igs has been carried out primarily by bottom-up MS.[107] The combination of data derived from the digestion of the antibody by different proteases can effectively result in high sequence coverage (up to 100%) and facilitates the identification of both large (e.g., glycosylation) and small (e.g., deamidation) PTMs.[108] Nevertheless, top-down as well as middle-down mass spectrometry own theoretical and practical advantages over bottom-up MS. As previously mentioned, the measurement of intact mass of the antibody or of some large fragments (e.g., the Fc) is fundamental for a first assessment of the main glycoforms, as BU MS can provide detailed information about single glycans, but not about their combinations. Furthermore, the intact mass can be used for determining the presence of possible truncation or unexpected modifications. This is of particular importance for bioengineered Igs, which are often expressed in host system which differ metabolically and genetically from the source system of the Ig.

Considering the increasingly growing interest in IgGs as biopharmaceuticals, the challenges that an MS-based analysis has to face are numerous. Aside from the traditional primary sequence confirmation or, in the case of *de novo* sequencing, determination, important goals are represented by detection of disulfide bond scrambling, localization of covalently bound organic

molecules in the so-called antibody-drug conjugates, precise determination of binding sites. Finally, if currently these analyses are performed on selected, isolated antibodies, a future goal is surely the extension of the throughput of analysis, to be able to accurately investigate IgGs present in mixtures.

The research on which this Thesis is based does not pretend to focus on all the above mentioned tasks. It is, on the contrary, specialized essentially on the structural analysis of monoclonal antibodies by a single activation technique, electron transfer dissociation. Conceptually, the starting point is represented by the analysis of a model protein to evaluate ETD capabilities in TD MS.

5.2 ETD analysis of intact transferrin: importance of disulfide bridges and protein gas phase conformation.

Experimental results contained in paper III are useful for describing in detail the workflow required for a complete protein analysis by ETD top-down MS, from parameters affecting experimental results to, notably, data analysis. Although the introductory part is based on ~29 kDa carbonic anhydrase, the experimental part is primarily focused on serotransferrin.

This is a ~79 kDa protein commonly employed as a standard in bottom-up MS experiments. Besides the large number of tryptic peptides it can generate, one of the main features of this protein is the presence of nineteen disulfide bridges, which can hinder efficient fragmentation in TD MS.

The ETD-based analysis of oxidized transferrin described in *paper III* was concluded with the assignment of mainly *c*-type product ions derived from a disulfide-free loop interconnecting two disulfide protected areas. Nevertheless, four of the identified N-terminal containing (i.e., *c*-type) product ions arose from disulfide-protected regions. After chemical reduction of disulfide bridges prior to MS analysis, ETD generated a different fragmentation pattern, with most of the assigned *c*- and *z*-type ions deriving from areas, generally not far from the termini, previously protected by disulfide bonds. Effectively, the central portion of the protein sequence remained unaffected by fragmentation.

These two data sets provide good evidences for a number of important conclusions. First, although ETD is capable of cleaving disulfide bonds the probability of this occurrence seems relatively low in this large, structured protein. Secondly, the imbalance towards N-terminal assigned product ions characteristic of the oxidized proteins, which disappears after reduction (for the reduced protein 34 and 39 c- and z-type ions were identified, respectively), might suggest that also protein tertiary structure is playing a non-negligible role in determining the final protein fragmentation, most likely by limiting the residues susceptible to protonation and by preventing an easy dissociation of fragments that are tightly bound together by non-covalent interaction. Finally, as general consideration, the fragmentation maps for the two investigated protein states (i.e., oxidized and reduced) suggest that disulfide bonds *direct* the protein fragmentation, determining the areas susceptible of extended sequencing.

5.3 Top-down MS of immunoglobulins G.

The here summarized results should be interpreted both as an attempt of improving the characterization of intact immunoglobulins G as a specific class of biomolecules, and also, more generally, as an exploration of the current limits in ETD-based top-down FTMS analysis of large proteins.

5.3.1 Intact mass measurement and glycoforms

High resolution time-of-flight mass spectrometry and Orbitrap FTMS demonstrated similar performances in terms of mass accuracy for the determination of IgG glycoforms. Both the employed mass analyzers were used on-line coupled to reverse phase LC, which was solely applied for ensuring efficient desalting and partial denaturation of the antibody. Upon averaging of all the spectra recorded during elution (typical elution time of some minutes, due to to LC column overloading), the final mass accuracy for the most abundant glycoforms in a single LC run was lower than 1 Da for both the mass analyzers. Although the use of different IgGs prevent a direct comparison between the two setups, time-of-flight mass spectrometry seems to present an advantage over Orbitrap FTMS in detecting low-abundant proteoforms (e.g, essentially glycoforms), probably due to the higher dynamic range of this technology over FT-based

instruments. It is important to underline, however, that in both cases no isotopically resolved information was obtained and all the values are referred to protein average mass. A more detailed discussion about glycoforms is reported in Chapter 4.

5.3.2 Sequence coverage and disulfide bridge cleavage of IgGs by ETD MS/MS

The capabilities of electron transfer dissociation of producing extensive backbone cleavage and also induce rupture of disulfide bonds were tested on intact human IgGs of subclass 1 and 2 as well as on murine IgG1. Independently from the employed MS platform, the results we obtained overtook the previous state-of-the-art in MS/MS of IgGs. The main reference in this sense was represented by report of Bondarenko *et al.* who performed CID on reduced and alkylated isolated chains of a murine IgG2 on a hybrid Orbitrap XL mass spectrometer.[109] The final sequence coverage for that IgG was attested around 10%. In our first study, conducted by ETD qTOF mass spectrometry, we reached ~15% for a human IgG1 and ~21% for the murine IgG1. It is noteworthy that the above mentioned difference in sequence coverage derived mostly from the sequencing of the N-terminus of both heavy and light chains: in the case of the murine IgG, we achieved almost complete sequencing of the region preceding the Cys residue involved in the intra-molecular disulfide bond of the variable domains of both chains; on the contrary, poor sequencing for the same area was reached for the human IgG1. This might reflect differences in the (gas phase) conformation of the variable domains of the two IgGs. Following studies carried out on two different generations of ETD-enabled hybrid LTQ-Orbitrap mass spectrometers pushed the sequence coverage up to ~32% for a human IgG1. The increase in identified unique backbone cleavages can be attributed to both the higher resolution of the mass analyzer and, particularly, to improved ETD efficiency. The latter aspect derived from: (i) a refined experimental design that took advantage of the acquired experience in fragmentation of large proteins, and substantiated in the selection of multiple highly charged precursors for ETD (allowed by a large isolation window, up to 600 Th) and the variation of ETD duration time in different experiments; (ii) geometry of the ETD reaction cell (linear ion trap) that, presumably, offers superior trapping of both multiply-charged cations and radical anions, as proved by reduced ion-ion reaction time required for maximum sequence coverage in comparison with the qTOF platform (ETD duration of 10-25 ms versus ~60 ms).

A common feature of all the here presented studies is the application of strategies aimed at improving the spectral signal-to-noise. Depending on the ion detection principle used by the specific mass analyzer the operations on data recorded in single LC-MS/MS experiments were conducted directly on *m/z* spectra (qTOF) or on time-domain (transient) signals (Orbitrap). The creation of high SNR spectra from data recorded in consecutive LC runs allowed the reduction of sample preparation to the minimum, differently from what typically happens when proteins have to be highly purified and desalted off-line for direct infusion. The main limitation of this approach consisted in its time-consuming nature, which we tried to balance in our most recent work by optimizing both ETD experimental parameters (i.e., selection of highly charged precursors) and the transfer efficiency of product ions from the linear ion trap to the high resolution mass analyzer (Orbitrap).

Finally, these ETD experiments showed the presence of backbone cleavages in disulfide-protected areas. Most of those belong to the C_H3 domain, which probably has a relatively loose and/or flexible structure, particularly towards the C-terminus. Several basic residues are also located within this C-terminal portion of this globular domain. The presence of these residues could be important not only as protonation might induce unfolding as a consequence of Coulombic repulsion, which also favors the gas phase dissociation of the non-covalently bound fragments after fragmentation, but also because the sequencing of disulfide protected areas requires a double ETD event to occur, which implies double charge reduction.

5.3.3 Preferential ETD cleavage sites in intact IgGs: a product ion abundance analysis

Independently from the MS platform in use, electron transfer dissociation on IgG resulted in the fragmentation of the same portions of sequence within a specific IgG subclass. Furthermore, if we exclude from the analysis the above mentioned complete sequencing of the N-terminal ends of light and heavy chains achieved for murine IgG1 only, we can extend the similarity of fragmentation pattern to all the analyzed IgGs, independently from subclass they belong to. More in detail, as previously described we found only one disulfide-protected domain which was highly sequenced (i.e., the C_H3 domain), whereas the remaining fragmented areas corresponded to *disulfide-free loops* interconnecting consecutive immunoglobulin domains. This is the case for both heavy and light chain. This common fragmentation pattern might find an explanation in the

retention of a relatively compact tertiary/quaternary structure in the gas phase, caused primarily by the difficulty in unfolding disulfide-protected immunoglobulin domains. A similar scenario would also give reason of the apparently impassable limit of ~30% sequence coverage that we reached by ETD and corresponds also to state-of-the-art results achieved by ECD. Technical expedients such as the reduction of the HCD gas pressure in the hybrid Orbitrap Elite mass spectrometer (see Chapter 2 and *paper VI*) helped in gaining higher SNR mass spectra but did not substantially changed the final sequence coverage.

Passing from a general view to a detailed observation only of the sequenced areas, it appeared manifest that specific cleavage sites were identified by several highly intense product ions of different charge states, whilst other cleavage sites had to be manually verified due to the presence of a single product ion of medium or low intensity. The phenomenon was validated when, on side of the assignment of typical ETD product ions, namely *c*- and *z*-type ions, also *y*-type ions were searched (for the light chain and, importantly, the C-terminal portion of the heavy chain). Most of the identified *y*-ions arose from cleavage sites for which related *z*-ions in multiple charge states had been discovered.

A *product ion abundance* (PIA) analysis on C-terminal containing product ions (i.e., *z*-ions) on different IgGs revealed that, as expected, a large discrepancy exists between product ions derived from disulfide-free and disulfide-protected regions, the formers being substantially more abundant than the latters. Furthermore, specific hot spots of fragmentation, luckily conserved between different IgGs due to their positioning in the constant domains and hence pinpointed in experiments involving different IgGs, were identified. Notably, they are located in proximity of basic residues. Increase of ETD duration caused reduction in the average charge state of assigned product ions, presumably due to secondary electron transfer events (and possible re-fragmentation), but did not dramatically modify the general fragmentation landscape which remained dominated by the presence of product ions from disulfide-free areas. These facts come out in favor of the retention of a relatively compact structure by IgGs from ionization to the gas phase. Notably, this hypothesis would account for the peculiar IgG sequencing pattern which alternates highly sequenced regions with areas totally unaffected by ETD. To a limited extent, PIA analysis of top-down spectra of highly-complex, large proteins might be useful for

determining or hypothesizing disulfide bonds connectivity and, possibly, discriminate between solvent exposed (i.e., highly protonated) regions from those inaccessible to the solvent.

5.3.4 What do we learn from the ETD-based top-down MS analysis of IgGs?

Electron transfer dissociation is an ion activation technique whose efficiency is increased with the charge state of the precursor ion. Therefore, one would expect that an extremely short ion-ion reaction time is required for a ~150 kDa protein fragmentation. Nevertheless, despite the use of a high number of electron donor radical anions, in large excess relative to the number of protein cations, activation times needed to reach good sequence coverage are in the order of tens of milliseconds. The results of the PIA analysis combined with a look at the localization of basic amino acid residues and disulfide bridges, suggest that structural constraints present in the gas phase are the major limitation for obtaining a more extended sequence coverage than the achieved ~30%. This hypothesis would find additional support in the observation that only a few identified fragments are larger than 20 kDa, which can be the consequence of a difficulty in cleaving all the inter-molecular disulfide bonds: generally only the intra-molecular ones are considered, but it must be underlined that currently the data analysis software cannot deal with internal fragments (i.e., non-canonical product ions where both the termini has been cleaved) nor with branched product ions composed by more chains (or parts of them). Effectively, the presence of these specific ions is suggested not only by the high number of unassigned peaks in IgG top-down spectra, but also by the results achieved through the PIA analysis, which shows evidences of secondary fragmentation of z-type ions by increasing ETD duration.

Moreover, structural constraints are not only reducing the finally achieved sequence coverage, but are impeding the analysis even of large PTMs (e.g., the portion of the heavy chain where the glycosylation is located is not sequenced). The only confidently assigned PTMs so far were located at the chain termini.

In conclusion, while improvements for TD MS of IgGs can be suggested on the base of the achieved results, approaches alternative to TD should also be pursued, ideally provided that they are based on the removal of major aforementioned structural constraints without affecting the sample quality (e.g., by introducing artifacts) like it can happen in BU MS.

5.4 Reducing IgGs to large polypeptides: a middle-down MS approach.

Processing an IgG in solution prior to its gas phase analysis represents the most logical approach to overcome the limitations in TD MS sequencing caused by retention in the gas phase of tertiary and quaternary structure of the IgG. Nevertheless, in order to avoid problems typical of BU MS the final size of the polypeptides produced in the sample preparation step must be adequately large, and the sample processing itself has to be fast, robust and occur under mild conditions.

The immunoglobulin-G degrading enzyme of *Streptococcus pyogens* (Ides) shows a papain-like activity, cleaving the IgG at the hinge region and releasing two Fab fragments and an Fc.[110] Differently from papain, the peculiarity of this enzyme is its high specificity and, most importantly, high efficiency: the proteolytic reaction can be performed in only ~30 minutes. The proteolysis is followed by denaturation and chemical reduction of disulfide bridges to yield three fragments of ~25 kDa each: light chain (Lc), C-terminal portion of the heavy chain (Fd) and the monomeric Fc (Fc/2). Notably, the entire protocol is carried out under optimal storage conditions for the IgGs (i.e., in IgG formulation buffer of slightly acidic phosphate buffer), and finally in slightly acidic environment which decreases the possibility of reduced disulfide bonds to be re-oxidized, and thus eliminates *de facto* the need of alkylating free thiols.

Polypeptides of 25 kDa require analysis conditions, from the point of view of the employed mass spectrometry parameters, typical of TD MS. Therefore, in Orbitrap FTMS (the only technique employed for this analysis in the Thesis) 10 microscans were used for both MS^1 and MS^2 spectra. Moreover, other settings concerning ion transmission (e.g., reduced HCD gas pressure) were borrowed from the analysis of intact IgGs. In addition, reversed phase liquid chromatography was used not for desalting but for polypeptide separation, which was fully achieved (i.e., with baseline resolution of chromatographic peaks) under ultrahigh pressure conditions (UPLC). The beneficial consequences of the application of typical TD MS instrumental parameters coupled with UPLC are apparent when results of single ETD LC-MS/MS runs are analyzed: almost 50% sequence coverage is obtained for Lc and Fc (including the localization of the N-glycosylation site) with the average of 7 mass spectra (corresponding to 70 microscans). Similarly to TD MS, upon averaging of transients derived from distinct ETD-based LC runs and variation of ETD activation time, for the same two fragments almost 70% sequence coverage was reached.

Importantly, the method was demonstrated to be sensitive enough to localize up to two oxidized methionine residues in the Fc.

Considering the similarity in size of the three IgG fragments and the sharpness of elution peaks under UPLC conditions, the results obtained in targeted studies, both in terms of chromatographic separation and also of MS-based protein sequencing, were encouraging enough to suggest the application of the same method to the analysis of IgG mixtures, which represents an important future goal. Preliminary results showed that reversed phase chromatography can partially separate a simple mixture of these polypeptides. This is an important achievement given that the analysis of mixtures of intact IgGs cannot rely on reversed phase chromatography and would require the use of an alternative chromatographic separation such as ion exchange chromatography (IEC). However, IEC is highly sensitive to the presence of specific PTMs such as charged glycans or deamidated amino acids and therefore can produce the elution of different proteoforms of the same IgG in multiple peaks, increasing the chromatogram complexity.

Chapter 6: Papers

Paper I

Deamidation and Transamidation of Substance P by Tissue Transglutaminase Revealed by Electron-Capture Dissociation Fourier Transform Mass Spectrometry

Luca Fornelli,[a] Adrien W. Schmid,[a, c] Luigino Grasso,[b] Horst Vogel,[b] and Yury O. Tsybin*[a]

Abstract: Tissue transglutaminase (tTGase) catalyzes both deamidation and transamidation of peptides and proteins by using a peptidyl glutamine as primary substrate. A precise consensus sequence for the enzyme is unknown and the ratio between deamidated and transamidated (or cross-linked) reaction products is highly substrate-dependent. Due to its overlapping body distribution with tTGase and ease of manipulation with tandem mass spectrometry, we used the neuropeptide substance P as a model to investigate the associated enzymatic kinetics and reaction products. Online liquid-chromatography Fourier-transform ion-cyclotron-resonance mass spectrometry (FT-ICR MS) combined with electron-capture dissociation (ECD) was employed to study the tTGase-induced modifications of substance P. A particular strength of ECD for peptide-enzyme reaction product monitoring is its ability to distinguish isomeric amino acids, for example, Glu and iso-Glu, by signature product ions. Our studies show that the primary reaction observed is deamidation, with the two consecutive glutamine residues converted sequentially into glutamate: first Gln_5, and subsequently Gln_6. We then applied ECD FT-ICR MS to identify the transamidation site on an enzymatically cross-linked peptide, which turned out to correspond to Gln_5. Three populations of substance-P dimers were detected that differed by the number of deamidated Gln residues. The higher reactivity of Gln_5 over Gln_6 was further confirmed by cross-linking SP with monodansylcadaverine (MDC). Overall, our approach described herein is of a general importance for mapping both enzymatically induced post-translational protein modifications and cross-linking. Finally, in vitro Ca-signaling assays revealed that the main tTGase reaction product, the singly deamidated SP $(RPKPEQFFGLM-NH_2)$, has increased agonist potency towards its natural receptor, thus confirming the biologically relevant role of deamidation.

Keywords: enzymes · fragmentation · peptides · radical ions · reaction dynamics

[a] L. Fornelli,+ Dr. A. W. Schmid,+ Dr. Y. O. Tsybin
Biomolecular Mass Spectrometry Laboratory
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne (Switzerland)
Fax: (+41) 121-693-9700
E-mail: yury.tsybin@epfl.ch

[b] L. Grasso, Dr. H. Vogel
Laboratory of Physical Chemistry of Polymers and Membranes
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne (Switzerland)

[c] Dr. A. W. Schmid+
Current address: Proteomics Core Facility
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne (Switzerland)

[+] Contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201002483.

Introduction

Tissue transglutaminase (type 2 transglutaminase; tTGase) is a ubiquitously expressed enzyme that is found extra- and intracellularly.[1] tTGase is the only member of its protein family to be expressed in human brain.[2] It is involved in several biological processes[1] and acts as a GTPase[3] and a kinase.[4] The most studied biological function of tTGase enzyme is the formation of peptidase-resistant isopeptidic bonds in a Ca^{2+}-dependent acyl-transfer reaction.[1,5–7] Under physiological conditions, the acyl-donor group can be the γ-carboxamide of a peptidyl-glutamine (Gln) residue, whereas the acyl-acceptor group can be either a low-molecular-weight primary amine or the ε-amino group of a lysine (Lys). In the latter case, the transamidation reaction (also known as a cross-linking) yields an ε-(γ-glutamyl)−lysine

bond. In the absence of a good acyl acceptor, tTGase can also deamidate Gln to glutamic acid (Glu)[8] by using a water molecule to react in place of the primary amine with the Gln–enzyme intermediate (Scheme 1). The tTGase acyl-transfer activity, which can result in both inter- and intramolecular protein cross-linking, has been widely investigated in relation to the onset of several neurodegenerative pathologies, characterized by the presence of protein aggregates.[9] Specifically, tTGase has been proven to induce the formation of protease-resistant aggregates, which are thought to play a crucial role in the initiation and development of Alzheimer disease, using both Aβ-40 and Aβ-42 at physiologically relevant concentrations in vitro.[10] Moreover, tTGase-mediated intramolecular cross-linking of α-synuclein has been shown to prevent the formation of typical amyloid aggregates.[11]
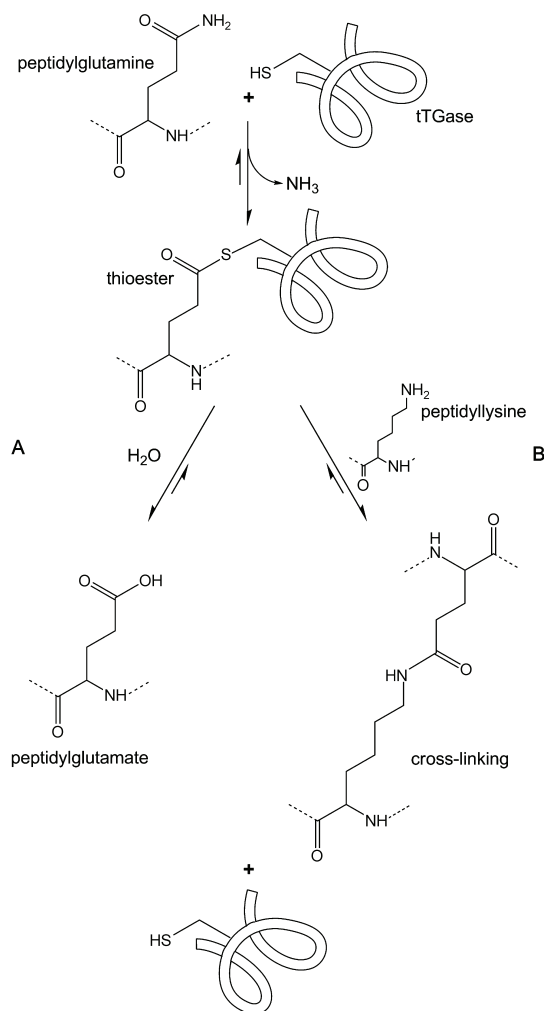
Scheme 1. Representation of two calcium-ion-dependent peptide-modifying reactions catalyzed by tissue transglutaminase (tTGase): A) deamidation and B) transamidation using the γ-acyl group of Gln as a first substrate, and water or ε-amino group of Lys as a second substrate.

On the other hand, glutamine deamidation is a post-translational modification (PTM) that introduces the charged amino acid Glu in place of the uncharged Gln within a polypeptide chain. Gln-to-Glu substitution can strongly influence protein–protein or protein–receptor interactions by altering signaling pathways and inducing pathological insurgence. An example of the latter case is tTGase-mediated deamidation of gliadin, a protein found in wheat gluten.[12,13] Deamidation of gliadin protein was reported to be responsible for the inflammatory responses associated with celiac disease.[14] Deamidation-induced formation of new disease-associated protein complexes might be recognized as neo-epitopes by T-cell receptors and lead to autoimmune response.

The substrate specificity of tTGase has been widely investigated recently.[15–17] Although a number of generic requirements have been reported, no conclusive amino acid sequence pattern for substrate–enzyme recognition has been found yet, neither for the acyl-donor nor for the acyl-acceptor group.[18,19] The lack of a defined consensus amino acid sequence that indicates the glutamines and lysines that can serve as substrates for tTGase, together with the fact that not all the actual substrates show the same affinity for the enzyme, requires the development and use of specific tools for determining the residues that are deamidated or transamidated by transglutaminase on peptides and proteins of interest, including pathology-related neuropeptides.[20,21] Traditional strategies for the identification utilize cross-validation approaches, as they comprise both the use of mutant peptides or proteins and bottom-up experiments, which consist primarily of mass-spectrometry-based analysis of protein digests sometimes coupled with collision-induced dissociation (CID) for the biomolecular ion fragmentation.[11]

Taking into account the current understanding of the central biological and pathological role of the tTGase-induced modification of peptides and proteins, we selected tTGase-catalyzed transamidation and deamidation of the neuropeptide substance P (SP)[22] to demonstrate the applicability of the tandem mass spectrometry for the characterization of tTGase reaction products. The undecapeptide SP belongs to the tachykinin neurotransmitter family.[23,24] It is distributed both in the peripheral nervous system,[25] where it mediates different physiological responses, including smooth muscle contraction,[26] vasodilation,[27] and increase in vascular permeability,[28] as well as in the central nervous system,[29] where it is involved mainly in inflammatory response[30] and pain perception.[31] SP action is regulated through its binding to the specific neurokinin-1 receptor (NK-1R), a G-protein-coupled transmembrane protein.[32–34] Binding to NK-1R results in phospholipase C activation and to the subsequent increase in the intracellular Ca^{2+} concentration.[35] Moreover, stimulated neurokinin-1 receptor can affect different neuronal pathways.[23] For example, analgesic response of the opiates can be influenced by NK-1R activation, mediated by its primary agonist SP, to modulate the reward (addition) mechanism of the opiates.[36,37] The use of SP as a substrate for tTGase is widely described in the literature. Particularly,

the presence of two consecutive Gln residues in the SP sequence (positions 5 and 6) has been exploited to induce the formation of cross-linked species with low-molecular-weight fluorescent amines, for example, spermidine, putrescine, and monodansylcadaverine.[38–40] First, this method allows to detect SP–amine complexes in different assays using the fluorescent properties of the amines.[38] Secondly, peptides modified with low molecular weight compounds can be efficiently analyzed by low-resolution mass spectrometry.[39] Although the overlapping body distribution with tTGase suggests that SP could be a natural substrate for this enzyme, no conclusive analysis of tTGase-induced SP deamidation and transamidation have been conducted. Importantly, as the C-terminal part of SP (residues 5–11) is involved in the NK-1R binding process,[41,42] SP deamidation could change the properties of the peptide sufficiently to affect its biological function.

The current study is thus aimed to provide in-depth analysis of tTGase-mediated deamidation of SP by using liquid chromatography (LC) coupled with high-resolution tandem Fourier-transform ion-cyclotron-resonance mass spectrometry (FT-ICR MS).[43] Identification of residue-specific modifications was carried out by electron-capture dissociation (ECD).[44–46] Electron addition to the multiply charged ions in the gas phase during the ECD process leads to formation of intermediate charge-reduced radical species. These metastable species readily dissociate by the cleavage of the $N-C_\alpha$ backbone bonds of the peptides, whereas the lateral chains of the amino acids and also labile post-translational modifications (PTMs) are typically preserved. Importantly, ECD FT-ICR MS has been recently shown to be able to differentiate aspartic acid from iso-aspartic acid[47] and glutamic from γ-glutamic acid.[48] The differentiation is based on the specific side-chain loss that is characteristic for the structure of the residues.

This study first confirms the previously reported higher reactivity of Gln_5 compared to Gln_6.[38] Second, we report the detection and characterization of SP dimers.[39] Both results are accompanied by the kinetics studies of tTGase deamidation and also transamidation. The latter point is of particular importance, considering that the mechanism of the highly substrate-dependent tTGase acyl-donor recognition is not completely understood.[15,49] Finally, to determine a putative biological role of tTGase-mediated deamidation of SP, a singly deamidated variant of SP was used to reveal changes in agonist potency upon binding to neurokinin-1 receptors.

Results

Kinetics of tTGase-induced deamidation of substance P: Figure 1 shows the kinetics of the reaction between substance P and tTGase monitored by LC–MS. More specifically, the time dependence of the population of singly deamidated substance P versus the unmodified peptide is shown. According to the literature, tTGase exhibits high specificity toward acyl-donor group recognition.[15] Theoretically, both
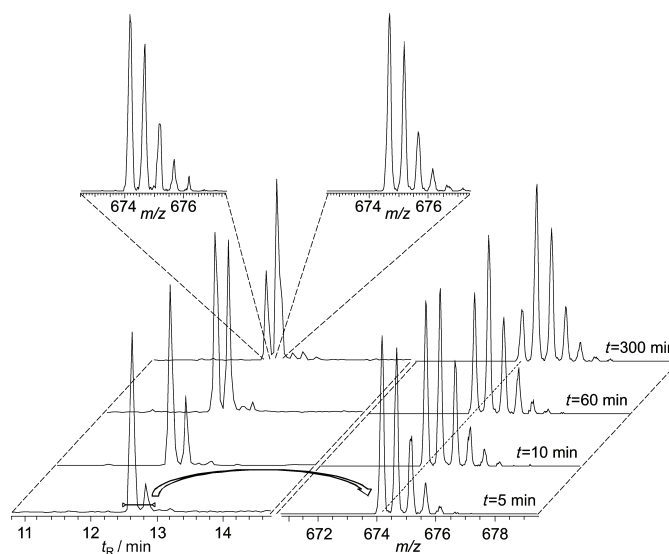

Figure 1. Time-course LC–MS analysis of substance-P deamidation catalyzed by tissue transglutaminase (tTGase). Four timepoints are considered, namely, 5, 10, 30, and 60 min. Left: Total ion current chromatograms exhibit a first peak that corresponds to unmodified (or control) substance P, and a second one that corresponds to the singly deamidated substance P. Right: Combined isotopic distribution of both peaks (the two isotopic distributions are partially overlapping) demonstrates the relative abundance of the unmodified and deamidated species. The inset shows separate isotopic distributions for the two species that correspond to the LC–MS of the 300 min sample. The m/z scale of the LC ESI LTQ MS data shown was only externally calibrated (for doubly charged SP, m/z: calcd: 674.37; found: 674.18).

glutamine residues in SP, namely, Gln_5 and Gln_6, could be deamidated. To monitor the deamidation order and rate, a reaction was performed with relatively low concentrations for both substrate (74 μM) and enzyme (0.002 U).

As depicted in Scheme 1, the deamidation reaction occurs when a water molecule enters the active site of tTGase after the formation of an intermediate between enzyme and Gln residue of the substrate. A low concentration of SP should favor the deamidation process over transamidation, whereas low enzyme concentration was chosen to allow the study of a fast process. Figure 1 shows that the singly deamidated substance P is the main reaction product under the employed reaction conditions (both SP and tTGase concentrations are low) at 300 min from the reaction onset. Figure 1, left, displays two baseline-resolved peaks from a total-ion chromatogram related to substance P control (substrate retention time ≈ 12.6 min) and singly deamidated substance P (product retention time ≈ 12.9 min). The combined isotopic distribution of the two peaks that differ in molecular weight by 0.98 Da demonstrates their relative abundance change as a function of time (Figure 1, right).

Competition between deamidation and transamidation of substance P: The results presented in Figure 1 demonstrate that substance P is a good substrate for tTGase, but do not clarify which one of the two Gln residues is deamidated.

The reaction pathway that leads to peptide transamidation with subsequent formation of dimers or polymers is also not considered in Figure 1. Previous findings by Ferrándiz et al. have shown that Gln_5 serves as an indispensable tTGase substrate and is more reactive as an acyl donor.[38] Thus, we used the LC–MS/MS to investigate if both Gln residues can be tTGase substrates and elucidate the possible reaction sequence as a function of time. Electron-capture dissociation (ECD) was chosen as a tandem mass spectrometry method mainly due to the efficient and easy-to-interpret fragmentation of substance P it induces. Furthermore, the specific character of ECD allows one to distinguish amino acid isomerization that may occur during the enzymatic reaction (vide supra). According to a study by Marino's group, substance P dimers should not be formed.[39] However, this finding contrasts with the general understanding of tTGase catalytic activity, which demonstrates a relatively good specificity for the peptidic sequence that contains the acyl donor (Gln), but not for the acyl acceptor (Lys).[18]

We assume that transamidation, which can be thought of as a competitive reaction with respect to deamidation, can be promoted by increasing the probability of two substance P monomers to be recognized in rapid succession by the enzyme. This result was achieved by using a high substrate concentration. To accelerate the reaction and increase the enzyme activity, we used higher substrate and tTGase concentrations and increased the temperature. Figure 2 shows
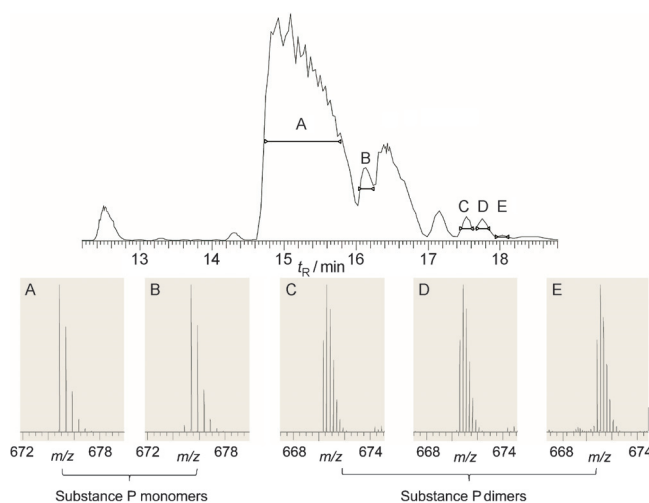

Figure 2. Reverse-phase HPLC separation of the tTGase-catalyzed substance-P deamidation and transamidation reaction products.

the total ion current (TIC) chromatogram of reaction products obtained after 5 h incubation of a solution of 222 μM SP with 0.02 U of tTGase at 37 °C. The species of interest were detected by LC FT-ICR MS (Figure 2, insets). The singly deamidated SP elutes at retention time, $t_R = 15$ min, followed by doubly deamidated SP at $t_R = 16.2$ min. The two peaks at $t_R = 16.4$ and 17.2 min correspond to the truncated forms, which lack the first two residues at the N terminus, of

singly and doubly deamidated SP, respectively. Degradation of SP after the second residue is observed normally after incubation for several hours at 37 °C and represents a cleavage at the C-terminal side of proline residue. These species are followed by three distinct peaks of lower intensity that correspond to the three different populations of SP dimers ($t_R = 17.55$, 17.75, and 18.1 min). Under these experimental conditions, the reaction was completed without detection of unreacted SP.

The substantially higher intensity of the peaks assigned to deamidated SP compared to the cross-linked SP suggests that the Lys residue of SP is not a preferred substrate for the enzyme. A similar conclusion can be made by comparing the reaction described here and the one between SP and the primary amine monodansylcadaverine (MDC) (vide infra).

Selective deamidation of Glu_5 and Glu_6: ECD mass spectra collected for singly and doubly deamidated SP in the data-dependent LC FT-ICR MS/MS experiment are shown in Figure 3. The expected and almost complete sequence coverage is achieved in both cases. The absence of the c_1^+/z_{10}^+ or
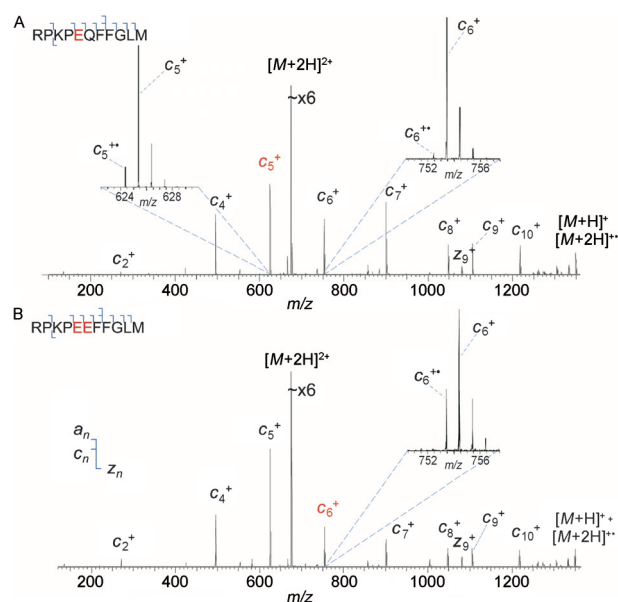

Figure 3. Tandem mass spectra of tTGase-catalyzed (top) singly and (bottom) doubly deamidated substance P, acquired with LC ECD FT-ICR MS.

c_3^+/z_8^+ ions is due to the ring structure of a proline residue, which prohibits the formation of the corresponding ECD product ions. The fragmentation patterns are similar to those of nondeamidated SP.[50] The predominance of c-type ions over z-type ions is, presumably, due to the position of the charged residues close to the N terminus of the peptide.

ECD reveals that the singly deamidated SP has a single Gln-to-Glu substitution that occurs exclusively at position 5 (Figure 3A). The inset in Figure 3A shows the presence of a

radical species, $c_5^{+\cdot}$, at the amount typically obtained in ECD of the unmodified SP for c_5^{+} ions. Moreover, all the heavier c-type ions, from c_6^{+} to c_{10}^{+}, show a 0.98 Da mass shift in comparison to the corresponding ions of unmodified SP, as demonstrated by c_6^{+} isotopic distribution (Figure 3A, inset). The ECD mass spectrum of doubly deamidated SP shown in Figure 3B correlates with the ECD pattern for the singly deamidated SP species for c_2^{+}, c_4^{+}, and c_5^{+} ions. Consistently, an additional 0.98 Da shift in mass is observed for the rest of the c_n^{+} ion series, for $n > 5$.

Importantly, no population of SP with a single deamidation at position 6 is observed. Therefore, ECD FT-ICR MS results, coupled with the intensity ratio of the TIC chromatographic peaks for the two modified SP species, suggest that either the deamidation of Gln_5 is necessary for the subsequent deamidation of Gln_6, or that Gln_6 is a much less probable substrate for tTGase than the Gln_5 residue.

Substance-P transamidation products: As displayed in Figure 4, the three populations of SP dimers revealed by LC–MS differ by a 0.98 Da mass shift, which corresponds to a single deamidation event.

Calculation of the expected molecular weight of crosslinked dimers from transamidation (twice molecular weight of SP minus ammonia; 17.024 Da) indicates the three dimers to be singly, doubly, and triply deamidated species. Proposed sequences for each population inferred by the measured m/z values are reported between the two isotopic distributions for the two charge states, 4+ and 5+, ob-

served in the broadband mass spectra (Figure 4). Note that the intensities of the first two dimer populations are similar and higher than the one of the third population, Figure 2.

ECD FT-ICR MS was performed on every isolated SP dimer population by using the $[M+4\,H]^{4+}$ ions as precursors (Figure 5). The FT-MS was set to operate in a targeted
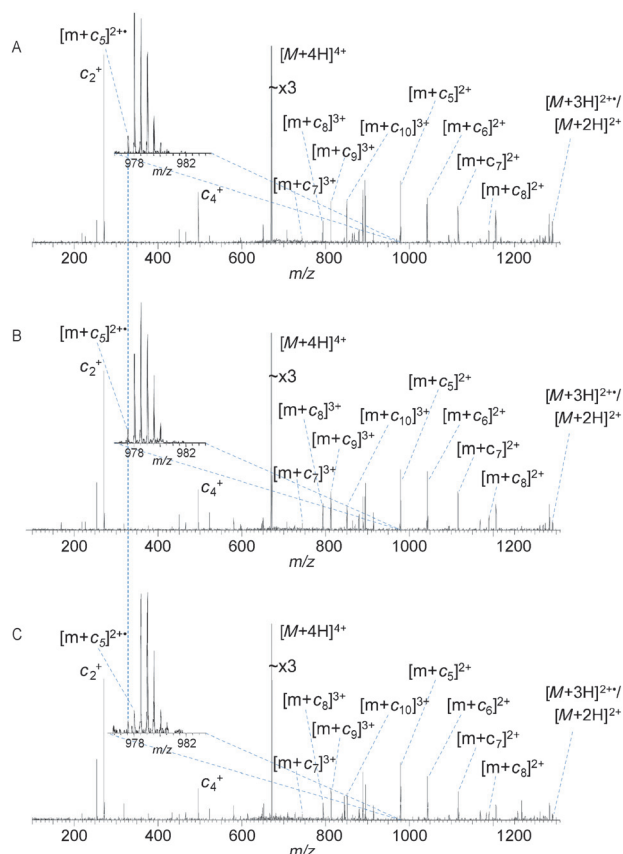
Figure 5. ECD FT-ICR MS of (top) singly, (middle) doubly, and (bottom) triply deamidated dimers of substance P formed by tTGase-induced transamidation ("m" indicates the monomeric form of SP). Tandem mass spectra were acquired under standard operating conditions in the LC–MS/MS experiment.

mode to isolate (isolation window of 5 Th) and fragment species at m/z 670.10. In all three ECD mass spectra reported in Figure 5, a total absence of c_5^{+} ion, that could originate from an SP monomer if cross-linked at position 6, is observed. Therefore, only Gln_5 is involved in the transamidation reaction. The low reactivity of Gln_6 revealed by the study of ECD mass spectra of the two deamidated forms of monomeric SP (Figure 3) supports this conclusion. The remarkably high abundance of c_2^{+} ions in comparison to other product ions, specifically c_4^{+} ions, reflects the unique possibility for both monomeric units that form the dimers to contribute.

ECD product ions that contain the transamidation-reaction-produced covalent bonding between an intact mono-
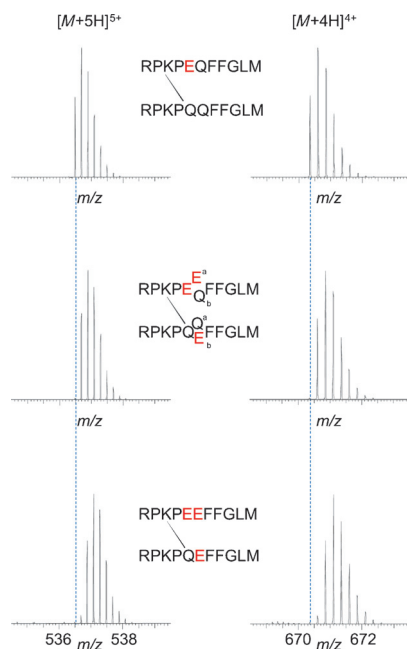

Figure 4. Isotopic distributions of the (left) quintuply and (right) quadruply charged populations of (top) singly, (middle) doubly, and (bottom) triply deamidated SP dimers formed by transamidation and observed by LC FT-ICR MS. Insets show the corresponding suggested structures of the dimers.

meric SP and a *c*-type fragment of the second SP monomer are assigned in Figure 5 as $[m+c_n]^{k+}$. The analysis of the series of $[m+c_n]^{k+}$ ions gives an important indication about the position of deamidated residues. The ECD fragmentation pattern obtained from the singly deamidated SP dimer population contains a characteristic $[m+c_5]^{2+}$ ion that indicates the deamidation of Glu_5 (Figure 5A, inset). The first deamidation of the Gln_5 residue is further confirmed by the fact that one of the two Gln_5 residues in the dimer is involved in the formation of the cross-linking bond, and a population of $[m+c_5]^{2+}$ ions without a corresponding mass shift is not observed.

ECD mass spectra for the SP dimer population with two deamidated residues is shown in Figure 5B. Depending on both the position of the second Glu residue and the fragmentation site, two possible $[m+c_5]^{2+}$ fragment ions can be produced by ECD: one with the same *m/z* as observed for the singly deamidated dimers, and a heavier one (Scheme 2).
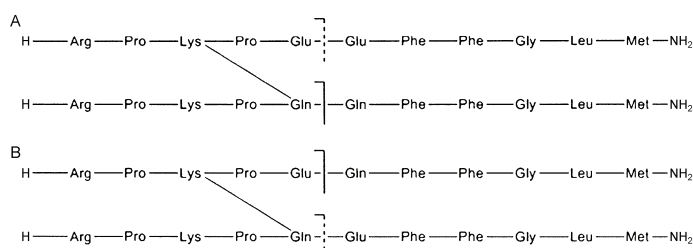
Scheme 2. Possible deamidation and fragmentation sites of doubly deamidated SP dimers for obtaining $[m+c_5]^{2+}$ ions ("m" indicates the monomeric form of SP); dashed lines indicate the lighter product ions, solid lines the heavier ones. A) Dimer with both the deamidation on the same monomer; B) dimer with deamidation on both SP monomers.

A partial overlap between the two product ion isotopic distributions can be observed (Figure 5B, inset). Finally, as a consequence of deamidation of all the three originally available Gln residues in the third population of SP dimers, a further mass shift in the $[m+c_5]^{2+}$ isotopic distribution is apparent (Figure 5C and inset).

Substance-P cross-linking with MDC: The above-illustrated results demonstrate that the tTGase-induced deamidation of Glu_6 can occur not only after the deamidation of Glu_5, but also after a cross-linking between Glu_5 and a peptydil lysine. To further investigate this process, a tTGase reaction with SP was performed in the presence of a primary amine, monodansylcadaverine (MDC).[38] The acyl-acceptor MDC was used in a 10-fold excess amount to promote the transamidation between SP and MDC in a competition with SP deamidation. Kinetics studies were performed by using the same algorithm as for the SP-tTGase reaction described above. We were able therefore to monitor both the formation of transamidated SP-MDC complexes and the results of the competition between deamidation and transamidation reac-

tions. In contrast to the previous reports on the kinetics of SP interaction with tTGase, the reaction product analysis described here did not take advantage of MDC fluorescent properties but was obtained through LC–MS analysis.[38]

The advantage of the LC–MS method is in the detailed characterization of peptide modifications, specifically, Gln deamidation, on transamidated SP-MDC complexes. The time-course analysis clearly indicates that the main reaction product is the cross-linked MDC-SP ($t_R = 11$ min, *m/z* 833.439 for the doubly charged species, as displayed on the left inset of Figure 6), in presence of both higher (right) and lower (left) concentration of substrates. After approximately one hour from the reaction start, the abundant peak of MDC-SP complex is followed by a less abundant peak ($t_R = 11.2$ min, *m/z* 833.931, doubly charged), which corresponds to an MDC-SP with deamidated Gln_6 (Figure 6, central inset). Although the intensity of the latter peak increases with time, it remains low. Nevertheless, the presence of this peak demonstrates that the second glutamine of SP can be deamidated also in presence of an isopeptidic bond on Gln_5 side chain.
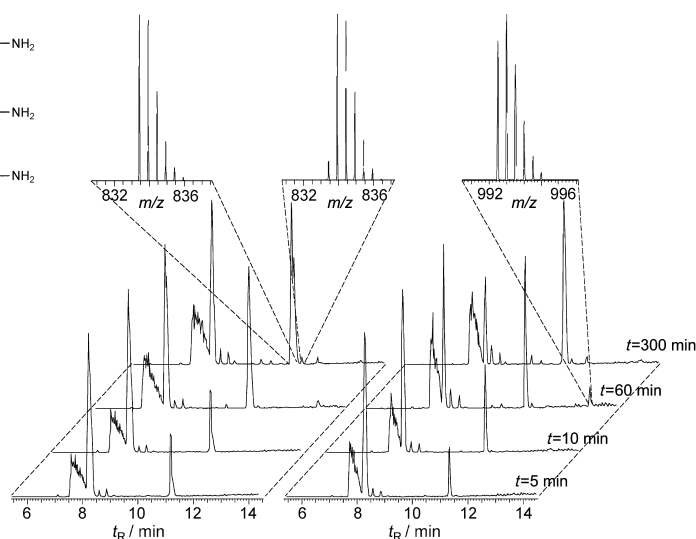
Figure 6. LC FT-ICR MS of a tTGase-catalyzed reaction between substance P and monodansylcadaverine (MDC) used in a 10-fold excess amount. Left: 74 μM SP; right: 148 μM SP.

Abad and co-workers investigated the reactivity of Gln_6 in SP during its tTGase-catalyzed reaction with MDC detecting the reaction products by reverse-phase HPLC combined with fluorescence spectroscopy.[38] They observed a high propensity of SP to be cross-linked to both Gln_5 and Gln_6 when the reaction is performed at 37 °C with an approximately 100-fold excess amount of MDC over SP. On the contrary, we show that under our reaction conditions, the formation of a doubly cross-linked SP (($MDC)_2$-SP) is not favored and that this reaction product is probably unstable. After 60 min from the beginning of the reaction, a peak related to ($MDC)_2$-SP is detected for both reactions ($t_R =$

13.3 min, m/z 992.508 for the doubly charged species; spectrum on Figure 6, right inset).

The reasons for the low abundance of $(MDC)_2$-SP can be found in the competition between deamidation and transamidation and in a possible hydrolytic activity of the enzyme towards isopeptidic bonds, as reported by Sollid et al.[51] This tTGase reaction could explain the absence of $(MDC)_2$-SP peak at the later reaction times.

Biochemical confirmation of substance-P deamidation: To verify the data collected with LC–MS/MS, and to determine possible modifications that occurred at SP as an effect of side reactions of tTGase, we performed enzymatic digestion of SP with endoproteinase Glu-C from *Staphylococcus aureus* V8 (Figure 7).[52] Glu-C can selectively cleave the
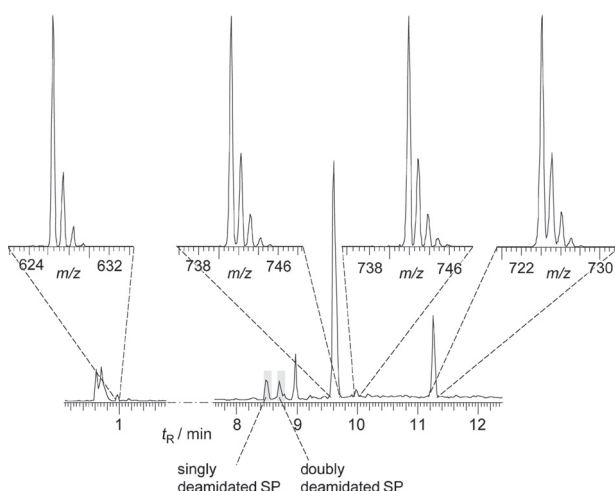

Figure 7. LC–MS of Glu-C digestion of reaction products of substance P with tTGase.

peptidic bond on the C-terminal side of a glutamic acid residue when used in ammonium bicarbonate 0.1 M, pH 8.0. A solid-phase-synthesized Glu_5-SP variant was used to compare its Glu-C digestion products with the ones obtained by using the tTGase-deamidated SP as a substrate for the endoproteinase. In summary, the following substrates were compared by the proteolitic digestion: SP after reaction with tTGase, Glu_5-SP after tTGase deamidation, and Glu_5-SP alone, without previous reaction with tTGase. All the samples were tested in the presence (positive control) and absence (negative control) of Ca^{2+}. Our enzymatic digestion described before was carried out using the products of up-scaled tTGase reactions (performed in 1 h with 0.05 units of enzyme to prevent possible aggregation of the products that could occur after many hours from the start) without intermediate purification. The substrates for Glu-C were dissolved in a buffer that contained Tris-HCl to obtain the final concentration in the new reaction mixture of around 110 μM. The synthetic Glu_5-SP was used for the reaction in presence

of the same concentration and composition of the buffer. The detection of the digested peptides was performed by LC–MS by using the linear trap quadrupole (LTQ) mass spectrometer.

All the screened samples provided the expected fragments: no cleavage for SP negative control, 2 main fragments, SP [1–5] and SP [6–11], for the singly deamidated SP (both enzymatically obtained from unmodified SP and a synthetic modified peptide), and, more interestingly, the same two fragments for the doubly deamidated SP, with a shift in mass of 0.98 Da in the SP [6–11] fragment in comparison with the one obtained from singly deamidated SP. This result, combined with the lack of detection of the SP [1–6] or SP [7–11] fragment, reveals that Glu-C cleaves specifically only after the Glu_5 residue (see Figure S1 in the Supporting Information). Figure 7 illustrates the products detected for the upscaled tTGase reaction of SP after 24 h. The highly hydrophilic SP [1–5] peptide elutes very early, and most probably it is partially undetected, whereas the above-mentioned variants of the SP [6–11] fragment elute after more than 9 min. Note that the second most abundant species in the TIC is at m/z 724.12. We suggest that it corresponds to the SP [6–11] fragment in which the N-terminal Gln or Glu residue is converted into pyroglutamic acid (pGlu) through a nucleophilic attack reaction of the free N-terminal amine to the γ-carboxylic group. The mass difference between this peak and the most abundant SP [6–11] peptide is 17.1 Da. The formation of pGlu after proteolitic cleavage of peptides and proteins has been previously reported, but only for tryptic digestion.[53] The LC–MS data alone is not sufficient to determine if only the N-terminal Gln residue is converted into the pGlu or if the N-terminal Glu of the second and less abundant SP [6–11] fragment also contributes. Literature reports suggest that the pGlu conversion is more favorable in the presence of an N-terminal glutamine, which correlates here with the observation of the more abundant of the two species.[54,55]

Conformational studies on substance P and singly deamidated substance P: To determine potential differences in the conformation of unmodified SP and its singly deamidated Glu_5-SP variant in solution, circular dichroism (CD) spectra were recorded for both peptides (Figure 8).

Both peptides in aqueous solution show CD spectra typical for a random coil conformation (Figure 8A) as reported previously.[56–58] Apparently, a single amino acid substitution as in the Glu_5-SP variant does not induce the formation of a stable secondary structure. However, it is known that SP can acquire an ordered secondary structure when dissolved in organic solvents or bound to detergent micelles. Figure 8B and C represent SP (solid line) CD spectra in 50% 2,2,2-trifluoroethanol (TFE) and 30 mM sodium dodecyl sulfate (SDS), respectively. In the former case, the spectra collected in 50% TFE are characterized by a minimum in ellipticity at 205 nm and a maximum at 195 nm, thus indicating a β-turn or mixed β-turn/α-helix structure.[57] Upon binding to negatively charged SDS micelles, SP forms predominantly
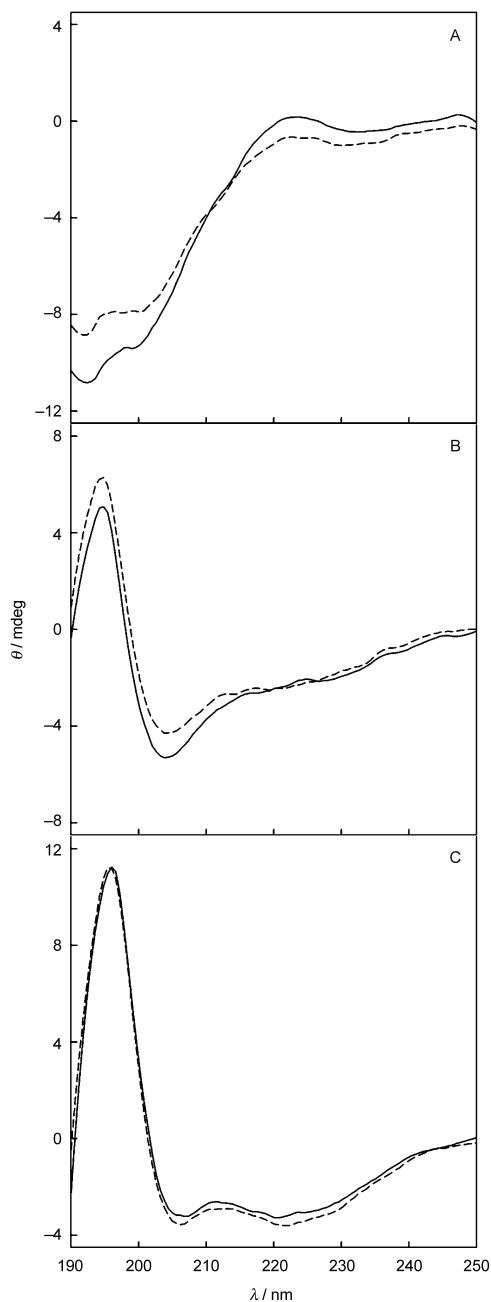
Figure 8. Circular dichroism spectra of SP (solid line) and Glu$_5$-SP (dashed line) in different solutions: A) water, B) 50% TFE–water, C) 30 mм SDS in water.

Although peptide structure assignment based on the CD spectra in SDS and TFE solution is not unambiguous, the high similarity of the spectra trends for SP and Glu$_5$-SP suggests that the substitution at the fifth residue does not alter the peptide structure in solution. Nevertheless, conclusive indications with regards to Glu$_5$-SP receptor binding characteristics could not be derived from our CD data.

Receptor activation studies: To demonstrate the biological relevance of our structural investigations further, we investigated the capability of Glu$_5$-SP to induce a cellular response through binding to the neurokinin-1 receptor (NK1-R). For these studies, HEK293 cells that stably expressed the NK1-R were used. We employed the fluorescence assay based on the signal transduction pathway, which is activated after the binding of an agonist to NK1-R. This process leads to an increase in the free cytosolic concentration of Ca^{2+} and subsequently to a fluorescent response mediated by Fluo-4 Ca^{2+} indicator dye.

First, a confocal microscopy-based analysis was carried out to compare the cellular response to the two peptides and the morphological changes induced by the potential stimulation (Figure 9).[60]
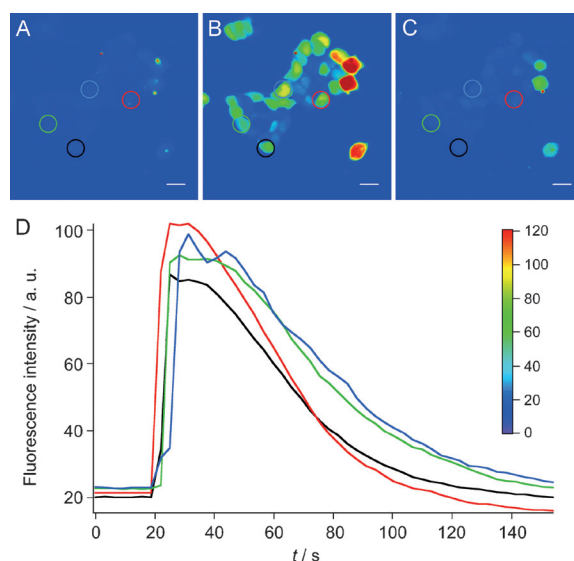
Figure 9. Ca-ion concentration mobilization in HEK293 after the addition of Glu$_5$-SP. The confocal images show the fluorescence of the Ca-ion indicator Fluo-4 inside of the HEK cells at A) 0 s, B) 30 s, and C) 110 s after the addition of Glu$_5$-SP to a final bulk concentration of 20 nм. D) Time dependence of fluorescence intensity of Fluo-4.

an α-helical conformation in agreement with published data.[57,59] The synthetic singly deamidated SP variant shows a similar behavior (Figure 8C). The CD spectra of the two peptides in the presence of 30 mм SDS are almost identical, whereas in 50% TFE, Glu$_5$-SP shows minor structure variations in comparison to the unmodified SP, in particular in the 215–230 nm region.

As shown in Figure 9, the addition of Glu$_5$-SP produces a quick increase of fluorescence with the cellular response duration, approximately 100 s, comparable with the reported values[61,62] and those observed for SP (data not shown). Moreover, the binding of SP to NK1-R eventually results in the formation of blebs on the cell membrane.[63] The same phenomenon is observed for Glu$_5$-SP.
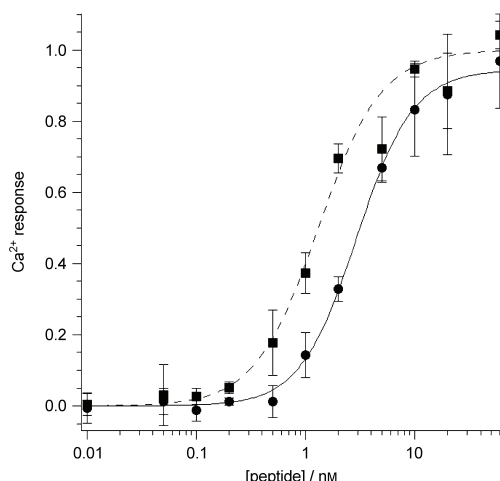
Figure 10. Change of the intracellular Ca-ion concentration at different bulk concentrations of SP (●) and Glu$_5$-SP (■) obtained as described in Figure 9. Lines are fits of the experimental data to the Hill equation. Experimental data are expressed as mean (± standard deviation) of four measurements.

We subsequently compared the cellular Ca-ion responses in a dose-dependent stimulation for the two agonists.[64] Figure 10 shows the concentration-response curves that indicate that Glu$_5$-SP induces the intracellular response at slightly lower concentrations than the unmodified counterpart. The calculated EC$_{50}$ values for the wild-type and deamidated SP are (2.9±0.5) and (1.3±0.2) nM, respectively. The repetition of the experiment resulted in a substantial confirmation of the ratio between the EC$_{50}$ values.

Discussion

The first biology-related question we addressed here by LC FT-ICR MS/MS concerns tTGase activity on SP. Although it is known that tTGase can catalyze two competitive reactions, one that involves both glutamine and lysine (transamidation), the other glutamine and a water molecule (deamidation), the reactivity of lysine residues in SP is a source of controversy in the literature. Whereas some reports claim that no dimers could be detected[39] and that lysines preceded by a proline are not good acyl acceptors, Pastor et al. mentioned the detection of intermolecular cross-linking of SP analogues, however, without delivering further information.[49] Here we have shown that the reactivity of Lys$_3$ is low. We confirmed it first by the kinetic study on SP only, which shows that the deamidation is much favored on this substrate over transamidation, and second by the analysis of SP–monodansylcadaverine reaction products, which on the contrary highlight the strong propensity of tTGase for cross-linking this primary amine to the neuropeptide.

However, we demonstrated that at the specific reaction conditions, for instance, at an increased substrate concentration, tTGase efficiently produces SP dimers, although dea-

midated SP remains the main reaction product. Moreover, tTGase can modify the structure of the monomers and create different populations of intermolecularly cross-linked SP. ECD FT-ICR MS analysis demonstrates the presence of three different subpopulations of dimers that are distinguished by the number of deamidated glutamines they contain, from 1 to 3 (three is the maximum number of Gln residues in SP dimer that may be modified after the formation of one isopeptidic bond).

As indicated by previous studies[38] and confirmed here by tandem mass spectrometry, in the case of SP, the two glutamine residues can both act as substrates for tTGase, but strictly sequentially. The Gln$_6$ residue can be either deamidated or transamidated, but only after the modification of the Gln$_5$ residue. This conclusion is based on the observation that, as displayed in Figure 2, in the attempt to force the tTGase-mediated production of dimers, the main reaction products are singly and doubly deamidated SP, and the former of these species is deamidated at position 5 only. This is revealed by the corresponding ECD mass spectra, which present a c_5^+ ion that corresponds to a Gln$_5$-to-Glu$_5$ substitution. The corresponding isotopic distribution is the one expected for a c^+ ion that terminates with Glu (RPKPE) and not the one terminated with Gln (RPKPQ), despite the presence of a radical species. The characterization of the SP dimers by ECD reaffirms this conclusion. ECD mass spectra show that the transamidation takes place only at position 5 of the first SP monomer, and the first of the three detected deamidations proceeds at Gln$_5$ of the second SP monomer. The Gln$_6$ residue is modified only afterwards. Consistently, in the MDC–SP reaction, the Gln$_6$ residue is deamidated only in doubly deamidated SP or in a small population of MDC–SP dimers in which the acyl donor for cross-linking remains Gln$_5$.

The fact that the modification that occurs at Gln$_5$ can be either a deamidation or a transamidation suggests that the probable reason for Gln$_6$ to start acting as a substrate is not the deamidation-induced local charge modulation (increase of a negative charge upon conversion from glutamine to glutamate). Although in the presence of two consecutive glutamines in the primary structure the one closer to the C-terminal end is normally less reactive than the other, this cannot be considered as a general rule.[15] Presumably, Gln$_6$ starts to be reactive after a modification of Gln$_5$ as a consequence of a conformational change in the peptide that leads to improved steric accessibility to the active site of tTGase. Therefore, a modification at position 5 might also affect SP properties in other binding processes. Generally, it is believed that SP assumes a partial α-helical conformation when interacting with membranes and proteins such as receptors. The attainment of an α-helical secondary structure has been previously indicated by CD experiments.[56,57] Recent NMR spectroscopic studies suggested that in a membrane-bound form SP is organized with a central α-helix and a flexible N-terminal end.[65] We confirmed the fact SP rearranges itself in an α-helix by CD performed in 30 mM SDS, a membrane-mimicking environment (see Figure 8). Not

surprisingly, the same behavior is retained also by Glu$_5$-SP. The activity of Glu$_5$-SP was a subject of the studies reported here due to the high rate of its formation by tTGase from SP. The capability of SP analogues to reach an α-helical conformation despite single and even double sequence mutations has been demonstrated by NMR spectroscopy.[65,66]

To reveal the capability of Glu$_5$-SP to interact with a protein, we chose a calcium-ion signaling assay in live HEK293 cells that stably express the neurokinin-1 receptor (NK1-R), the natural agonist of which is SP. We demonstrated that Glu$_5$-SP acts as an agonist for NK1-R, with an activity that is even higher than that of unmodified SP. NMR spectroscopic studies conducted on polypeptides that represent the predicted second and third extracellular loops (named EC2 and EC3, respectively) of NK1-R showed that both peptides are involved in SP binding; specifically Gln$_5$ is among those SP residues that interact with EC3 and, presumably, that maintains a solvent-accessible position.[67,68] This is in agreement with the case of the NK2 receptor, which is very closely related to the NK1-R, in which the binding site for agonist and antagonist peptides has been localized with high precision in living cells also at the extracellular side of the receptor.[69,70] On the other hand, Perrine et al. demonstrated that in cells that express the rat NK1-R, the Gln$_5$-to-Ala$_5$ substitution only slightly reduces the capability of SP to induce intracellular calcium-ion responses, thus suggesting that the residue at the fifth position is involved in the binding to the receptor but is not essential for the activation of the downstream signal transduction.[66] Nevertheless, the latter study considers a nonconservative substitution, whereas the tTGase-induced deamidation results in a less dramatic modification. In the Gln-to-Glu substitution, the main variation is associated with the introduction of a negative charge into SP, whereas the steric hindrance of the two side chains is similar. The results obtained here suggest that the absence of the amide group in the lateral chain of glutamate slightly alters the interaction of the fifth residue with both the solvent and, more notably, the third extracellular loop of the receptor without compromising at the same time the general structure of the peptide. These findings offer the prospect for further analysis directed toward the elucidation in detail of the differences in the interaction of substance SP and Glu$_5$-SP with EC3.

Conclusion

By coupling reverse-phase liquid chromatography, for the separation of tissue transglutaminase (tTGase) reaction products, with online electron capture dissociation high-resolution mass spectrometry, we were able to unambiguously determine the deamidation and transamidation sites on the neuropeptide substance P (SP). The identification of these sites on our model peptide was unambiguous even without using traditionally employed strategies such as testing modified peptides that carry point mutations. In fact, ECD offered a complete fragmentation pattern of both deamidated

and transamidated SP (without, in the latter case, cleavage of the newly formed isopeptidic bond). Therefore our proposed method has the potential of being used in the elucidation of structural features of larger peptides and proteins, for example, amyloid-beta peptide,[71] subjected to both enzymatic "zero-length" cross-linking and post-translational modifications.

Taken together, this study shows that under different conditions the major reaction products are always singly and doubly deamidated SP, tereby highlighting the low propensity of Lys$_3$ of SP to act as a substrate in a cross-linking reaction. Moreover, ECD spectra have demonstrated that both the two consecutive glutamine residues of the substrate (Gln$_5$ and Gln$_6$) can be deamidated, but strictly sequentially, and with Gln$_5$ as a preferred substrate for tTGase. In fact, when the reaction conditions allow the formation of SP dimers, ECD reveals that the cross-linking site is on Gln$_5$ only, whereas Gln$_6$ can react the same but, again, only after Gln$_5$ modification (deamidation or transamidation). This results in the formation of three different dimers populations with one, two, or three deamidated Gln residues, respectively. With regards to the identification of the isomeric residues, for example, glutamate to γ-glutamic acid isomerization, the results reported here did not demonstrate the presence of characteristic product ions in ECD FT-ICR MS specified by O'Connor and co-workers.[48] Therefore, the reaction catalyzed by tTGase indeed does not induce isomerization and deamidated SP species do not undergo spontaneous Glu isomerization after 5 h in solution at 37 °C. Finally, we discovered that the cellular response of SP upon binding to the natural receptor NK1-R is not decreased but, on the contrary, slightly increased by a single deamidation event that occurs at residue 5, as proven by an in vitro Ca^{2+} assay with HEK293 cells that express the receptor.

Experimental Section

Samples and reagents: Water, acetonitrile (HPLC MS grade), NaCl, and CaCl$_2$ were purchased from Fluka analytical (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). Tris-HCl buffer was purchased from Calbiochem (Nottingham, United Kingdom). Guinea pig liver transglutaminase, endoprotease Glu-C from *Staphylococcus aureus* V8, and substance P were purchased from Sigma–Aldrich (Buchs, Switzerland). Dulbecco's modified Eagle medium (DMEM), GlutaMAX medium, fetal calf serum (FCS), and Dulbecco phosphate-buffered saline (D-PBS) were purchased from Invitrogen (Breda, Netherlands). Fluo-4 NW calcium assay kit was obtained from Invitrogen (Carlsbad, CA, USA). Substance P variant modified at position 5 (E$_5$-SP) with the substitution of the glutamine residue by glutamic acid, RPKPEQFFGLM-NH$_2$, was produced by solid-phase synthesis (Protein and Peptide Chemistry Facility, University of Lausanne, Switzerland). E$_5$-SP crude peptide was purified by reverse-phase HPLC chromatography, whereas other reagents and samples were used without further purification.

Transglutaminase-catalyzed reactions: In a typical reaction catalyzed by Guinea pig liver transglutaminase (tTGase), the reaction mixture contained 40 mM Tris-HCl (pH 7), 70 mM NaCl, and 2 mM CaCl$_2$. For the kinetics studies of substance-P deamidation following interaction with tTGase, tTGase (0.002 units) was mixed with a solution of substance P

(74 μм) at an incubation temperature of 30 °C. In total, 7 fractions were collected at the following time points: 5, 10, 15, 30, 60, 120, and 300 min. For substance-P deamidation and transamidation, the peptide concentration was increased to 222 μм. For this reaction and those described below, tTGase (0.02 units) was used at 37 °C unless otherwise stated. The reaction was stopped after 24 h with fraction collection performed after 1, 2, and 5 h. For the monodansylcadaverine (MDC)–substance P cross-linking reaction, substance P at two concentrations, 74 and 148 μм, was used. A 10-fold excess amount of MDC was employed to yield the corresponding final MDC concentrations of 740 and 1480 μм, respectively. Several fractions were collected following the scheme used for substance-P deamidation and transamidation. The enzymatic catalysis was always quenched by immediately storing the liquid samples at −80 °C.

Liquid chromatography mass spectrometry (LC–MS): Analysis of the reaction products was performed by liquid chromatography combined online with high-resolution tandem mass spectrometry (LC–MS/MS). For the reverse-phase liquid chromatography, a Surveyor LC pump (Thermo Fisher Scientific, Bremen, Germany) was used. After a 1:2 dilution in water, samples (10 μL) were loaded onto a silica-based Hypersil Gold (100×1 mm, 1.9 μm particle size) C_{18} resolving column (Thermo Fisher Scientific, Bremen, Germany). Eluents consisted of 0.1 % (v/v) formic acid in water (mobile phase A) and 0.1 % (v/v) formic acid in acetonitrile (mobile phase B). The employed flow rate was 50 μL min^{-1} with an initial flow of 5 % B maintained for 2 min followed by a linear gradient up to 80 % B in 20 min.

Tandem mass spectrometry analysis was performed using a linear-ion-trap Fourier-transform ion-cyclotron-resonance mass spectrometers (LTQ FT-ICR MS) equipped with a 7 and 12 T superconducting magnet (Thermo Fisher Scientific). Eluting analytes were electrosprayed by using a standard electrospray ion source (ESI). The LTQ FT-ICR MS were operated in the data-dependent mode to automatically switch between broadband MS and MS/MS acquisitions. Peptide cations of interest were isolated in the LTQ (isolation window of 5 Th, target charge counts 5× 10⁵) according to a predefined mass list and then transferred to the ICR ion trap for subsequent tandem mass spectrometry. Electron-capture dissociation (ECD) was performed under standard experimental parameters: selected precursor ions were irradiated with low-energy electrons for approximately 50 ms before product ion excitation and detection. Ion magnetron motion correction was taken into account by introducing an appropriate delay prior to electron injection.[72] The minimum signal-to-noise (S/N) ratio for selection of precursor ions was set to 500. Broadband MS spectra (m/z 500–800) were collected using transient recording over 768 ms (a resolution of 100 000 at m/z 400 for 7 T LTQ FT-ICR MS configuration), whereas ECD was performed with transient recording over 192 ms (a resolution of 25 000 at m/z 400 for 7 T LTQ FTMS configuration) across an m/z window of 100–2000 Th. Reported ECD mass spectra normally average 10 scans. Data analysis was performed using Xcalibur software, version 2.0.7 (Thermo Fisher Scientific).

Proteolytic digestion with Glu-C: Endoproteinase Glu-C digestion was carried out at 37 °C in 0.1 м ammonium bicarbonate at pH 8.0 by using an enzyme-to-substrate ratio of 3:100 (w/w). Reaction mixture aliquots were collected after 6, 10, and 24 h. Different substrates were tested at 110 μм concentration, including substance P with and without modifications: wild-type substance P; reaction products of tTGase catalytic activity (same conditions previously described except for the enzyme concentration, 0.05 U instead of 0.02) on substance P with and without Ca^{2+}; reaction products of tTGase on synthetic Glu₅-SP (same conditions mentioned above) with or without Ca^{2+}; synthetic Glu₅-SP in the same Tris-HCl final concentration of the previous sample, with and without Ca^{2+}. The GPMAW tool (version 8.0, Lighthouse data, Odense, Denmark) was used for generating a list of expected peptides from the enzymatic digestion. Reactions were always quenched by immediate sample storing at −80 °C.

CD measurements: Far-UV circular dichroic spectra were collected using a Jasco J-810 spectropolarimeter (Jasco Corporation, Japan) at 37 °C with a cuvette with a path length of 0.1 cm. CD spectra were recorded between 190 and 250 nm, with 0.2 nm wavelength steps and a bandwidth of 1 nm. Samples (SP and Glu₅-SP) were dissolved to a final concentration

of 50 μм in water, 50 % 2,2,2-trifluoroethanol (TFE) and 30 mм sodium dodecyl sulfate (SDS). For every reported spectrum, six scans were averaged to improve signal-to-noise ratio. All spectra were corrected for the buffer baseline.

Confocal imaging: Adherent HEK293 cells that stably expressed NK1 receptors[60] were grown in DMEM/F-12+GlutaMAX medium that contained 10 % FCS and Hygromycin B (100 μg mL^{-1}) in a humidified atmosphere with 5 % CO_2 at 37 °C. For confocal imaging, cells were seeded on glass coverslips in six-well dishes in DMEM/FCS (2 mL). After 24 h, cells were incubated at 37 °C for 30 min with 5 mм Fluo-4 AM (Invitrogen, Carlsbad, CA, USA) in serum-free medium, followed by 30 min incubation with fresh medium that contained 10 % FCS. Agonist-containing solution was added to the selected well during the experiments. Cells were imaged using confocal fluorescence microscopy with a Zeiss LSM510 confocal microscope equipped with a C-apochromat ×63/1.2 water-immersion objective (Zeiss, Germany). Excitation was at 488 nm (Ar laser) and the Fluo-4 fluorescence was monitored using a 505 nm longpass emission filter at a rate of 1 image every 2 s.

Calcium-signaling assay: For monitoring the receptor activation, HEK293 cells that stably expressed NK1 receptors were seeded in DMEM/FCS into a clear-bottom 96-well plate for 24 h at 37 °C and 5 % CO_2. Growth medium was removed and replaced with fluorescent Ca^{2+}-indicator solution (Invitrogen, Carlsbad, CA, USA). Cells were subsequently incubated for 30 min at 37 °C followed by another 30 min at room temperature before the measurements, without removing the dye-loading solution. The increase of fluorescence intensity upon addition of SP or Glu₅-SP in 11 different concentrations (60, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, 0.05, and 0.01 nм) in assay buffer (1× Hank's balanced salt solution (HBS), 20 mм 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes)) was measured on a fluorescence plate reader (FLEX Station, Molecular Devices) at room temperature. The fluorescence emission was recorded at (525 ± 4.5) nm upon excitation at (485 ± 4.5) nm, and final values of Ca^{2+} response were the result of averaging the measurements of 4 wells for each peptide concentration. The EC_{50} for SP and Glu₅-SP were obtained by fitting a sigmoidal dose–response equation to the Hill equation [Eq. (1)]:

$$E_L = E_0 + (E_{max} - E_0)/(1 + (EC_{50}/[L])^{n_H}) \tag{1}$$

in which E_L is the fluorescence intensity for a given ligand concentration [L]; EC_{50} is the concentration that produces 50 % of the maximal response E_{max}; E_0 is the baseline intensity in the absence of ligand; n_H is the Hill coefficient. Experimental data were fitted to the Hill equation using Igor Pro software (WaveMetrics Inc., Lake Oswego, OR). Experiments were performed in duplicate. To assess a possible statistical difference among the two dose–response curves, an F-test was performed, whereas a t-test was used for the comparison of EC_{50} values.

Acknowledgements

[1] S. Y. Kim, L. Marekov, P. Bubber, S. E. Browne, I. Stavrovskaya, J. Lee, P. M. Steinert, J. P. Blass, M. F. Beal, G. E. Gibson, A. J. Cooper, *Neurochem. Res.* **2005**, *30*, 1245.

[2] M. Griffin, R. Casadio, C. M. Bergamini, *Biochem. J.* **2002**, *368*, 377.

[3] K. E. Achyuthan, C. S. Greenberg, *J. Biol. Chem.* **1987**, *262*, 1901.

[4] T. S. Lai, T. F. Slaughter, C. M. Koropchak, Z. A. Haroon, C. S. Greenberg, *J. Biol. Chem.* **1996**, *271*, 31 191.

[5] J. S. Chen, K. Mehta, *Int. J. Biochem. Cell Biol.* **1999**, *31*, 817.

[6] S. Beninati, M. Piacentini, *Amino Acids* **2004**, *26*, 367.

[7] J. E. Folk, P. W. Cole, *J. Biol. Chem.* **1966**, *241*, 5518.

[8] L. Lorand, R. M. Graham, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 140.

[9] M. M. Wilhelmus, A. M. van Dam, B. Drukarch, *Eur. J. Pharmacol.* **2008**, *585*, 464.

[10] D. M. Hartley, C. Zhao, A. C. Speier, G. A. Woodard, S. Li, Z. Li, T. Walz, *J. Biol. Chem.* **2008**, *283*, 16790.

[11] A. W. Schmid, D. Chiappe, V. Pignat, V. Grimminger, I. Hang, M. Moniatte, H. A. Lashuel, *J. Biol. Chem.* **2009**, *284*, 13128.

[12] W. Dieterich, T. Ehnis, M. Bauer, P. Donner, U. Volta, E. O. Rieck-en, D. Schuppan, *Nat. Med.* **1997**, *3*, 797.

[13] O. Molberg, S. N. McAdam, R. Korner, H. Quarsten, C. Kristiansen, L. Madsen, L. Fugger, H. Scott, O. Noren, P. Roepstorff, K. E. Lundin, H. Sjostrom, L. M. Sollid, *Nat. Med.* **1998**, *4*, 713.

[14] S. Dorum, S. W. Qiao, L. M. Sollid, B. Fleckenstein, *J. Proteome Res.* **2009**, *8*, 1748.

[15] Y. Sugimura, M. Hosono, F. Wada, T. Yoshimura, M. Maki, K. Hitomi, *J. Biol. Chem.* **2006**, *281*, 17699.

[16] S. T. Khew, P. P. Panengad, M. Raghunath, Y. W. Tong, *Biomaterials* **2010**, *31*, 4600.

[17] E. Csosz, P. Bagossi, Z. Nagy, Z. Dosztanyi, I. Simon, L. Fesus, *J. Mol. Biol.* **2008**, *383*, 390.

[18] J. J. Grootjans, P. J. T. A. Groenen, W. W. de Jong, *J. Biol. Chem.* **1995**, *270*, 22855.

[19] T. Ohtsuka, M. Ota, N. Nio, M. Motoki, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 2608.

[20] J. Zhang, R. P. Guttmann, G. V. Johnson, *J. Neurochem.* **1998**, *71*, 240.

[21] G. J. Ho, E. J. Gregory, I. V. Smirnova, M. N. Zoubine, B. W. Festoff, *FEBS Lett.* **1994**, *349*, 151.

[22] M. M. Chang, S. E. Leeman, *J. Biol. Chem.* **1970**, *245*, 4784.

[23] C. Severini, G. Improta, G. Falconieri-Erspamer, S. Salvadori, V. Erspamer, *Pharmacol. Rev.* **2002**, *54*, 285.

[24] A. M. Khawaja, D. F. Rogers, *Int. J Biochem Cell Biol* **1996**, *28*, 721.

[25] D. M. White, *J. Peripher. Nerv. Syst.* **1997**, *2*, 191.

[26] J. Mizrahi, S. Dion, P. D'Orleans-Juste, E. Escher, G. Drapeau, D. Regoli, *Eur J Pharmacol* **1985**, *118*, 25.

[27] C. Bossaller, K. Reither, C. Hehlert-Friedrich, W. Auch-Schwelk, K. Graf, M. Grafe, E. Fleck, *Herz* **1992**, *17*, 284.

[28] A. Saria, Z. Yan, G. Wolf, D. Loidolt, C. R. Martling, J. M. Lundberg, *Acta Oto-Laryngol. Suppl.* **1989**, *457*, 25.

[29] A. Ribeiro-da-Silva, T. Hokfelt, *Neuropeptides* **2000**, *34*, 256.

[30] I. Marriott, *Front. Biosci.* **2004**, *9*, 2153.

[31] R. C. Frederickson, V. Burgis, C. E. Harrell, J. D. Edwards, *Science* **1978**, *199*, 1359.

[32] A. Saria, *Eur. J. Pharmacol.* **1999**, *375*, 51.

[33] L. Quartara, C. A. Maggi, *Neuropeptides* **1997**, *31*, 537.

[34] L. Quartara, C. A. Maggi, *Neuropeptides* **1998**, *32*, 1.

[35] P. C. Sternweis, A. V. Smrcka, *Trends Biochem. Sci.* **1992**, *17*, 502.

[36] P. Murtra, A. M. Sheasby, S. P. Hunt, C. De Felipe, *Nature* **2000**, *405*, 180.

[37] K. G. Commons, *Brain Res.* **2010**, *1314*, 175.

[38] C. Ferrandiz, E. Perez-Paya, L. Braco, C. Abad, *Biochem. Biophys. Res. Commun.* **1994**, *203*, 359.

[39] R. Porta, C. Esposito, S. Metafora, P. Pucci, A. Malorni, G. Marino, *Anal. Biochem.* **1988**, *172*, 499.

[40] C. Esposito, F. Mancuso, A. Calignano, P. Di Pierro, P. Pucci, R. Porta, *J. Neurochem.* **1995**, *65*, 420.

[41] S. Lavielle, G. Chassaing, S. Julien, J. Besseyre, A. Marquet, *Neuropeptides* **1986**, *7*, 191.

[42] H. Duplaa, G. Chassaing, S. Lavielle, J. C. Beaujouan, Y. Torrens, M. Saffroy, J. Glowinski, P. D'Orleans Juste, D. Regoli, A. Carruette et al., *Neuropeptides* **1991**, *19*, 251.

[43] A. G. Marshall, C. L. Hendrickson, M. R. Emmett, R. P. Rodgers, G. T. Blakney, C. L. Nilsson, *Eur. J. Mass Spectrom.* **2007**, *13*, 57.

[44] R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter, F. W. McLafferty, *Anal. Chem.* **2000**, *72*, 563.

[45] H. J. Cooper, K. Hakansson, A. G. Marshall, *Mass Spectrom. Rev.* **2005**, *24*, 201.

[46] Y. O. Tsybin, J. P. Quinn, O. Y. Tsybin, C. L. Hendrickson, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 762.

[47] J. J. Cournoyer, J. L. Pittman, V. B. Ivleva, E. Fallows, L. Waskell, C. E. Costello, P. B. O'Connor, *Protein Sci.* **2005**, *14*, 452.

[48] X. Li, C. Lin, P. B. O'Connor, *Anal. Chem.* **2010**, *82*, 3606.

[49] M. T. Pastor, A. Diez, E. Perez-Paya, C. Abad, *FEBS Lett.* **1999**, *451*, 231.

[50] H. B. Hamidane, D. Chiappe, R. Hartmer, A. Vorobyev, M. Moniatte, Y. O. Tsybin, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 567.

[51] J. Stamnaes, B. Fleckenstein, L. M. Sollid, *Biochim. Biophys. Acta* **2008**, *1784*, 1804.

[52] G. R. Drapeau, Y. Boily, J. Houmard, *J. Biol. Chem.* **1972**, *247*, 6720.

[53] B. Thiede, S. Lamer, J. Mattow, F. Siejak, C. Dimmler, T. Rudel, P. R. Jungblut, *Rapid Commun. Mass Spectrom.* **2000**, *14*, 496.

[54] D. Chelius, K. Jing, A. Lueras, D. S. Rehder, T. M. Dillon, A. Vizel, R. S. Rajan, T. Li, M. J. Treuheit, P. V. Bondarenko, *Anal. Chem.* **2006**, *78*, 2370.

[55] G. N. Abraham, D. N. Podell, *Mol. Cell Biochem.* **1981**, *38 Spec No*, 181.

[56] G. Chassaing, O. Convert, S. Lavielle, *Eur. J. Biochem.* **1986**, *154*, 77.

[57] R. W. Williams, J. L. Weaver, *J. Biol. Chem.* **1990**, *265*, 2505.

[58] L. P. Choo, M. Jackson, H. H. Mantsch, *Biochem. J.* **1994**, *301*, 667.

[59] D. A. Keire, T. G. Fletcher, *Biophys. J.* **1996**, *70*, 1716.

[60] B. H. Meyer, K. L. Martinez, J. M. Segura, P. Pascoal, R. Hovius, N. George, K. Johnsson, H. Vogel, *FEBS Lett.* **2006**, *580*, 1654.

[61] H. Li, S. E. Leeman, B. E. Slack, G. Hauser, W. S. Saltsman, J. E. Krause, J. K. Blusztajn, N. D. Boyd, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9475.

[62] K. McConalogue, O. Dery, M. Lovett, H. Wong, J. H. Walsh, E. F. Grady, N. W. Bunnett, *J. Biol. Chem.* **1999**, *274*, 16257.

[63] J. Meshki, S. D. Douglas, J. P. Lai, L. Schwartz, L. E. Kilpatrick, F. Tuluc, *J Biol Chem.* **2009**, *284*, 9280.

[64] Y. Lill, K. L. Martinez, M. A. Lill, B. H. Meyer, H. Vogel, B. Hecht, *Chemphyschem* **2005**, *6*, 1633.

[65] D. J. Beard, S. A. Perrine, E. Phillips, S. Hoque, S. Conerly, C. Tichenor, M. A. Simmons, J. K. Young, *J. Med. Chem.* **2007**, *50*, 6501.

[66] S. A. Perrine, D. J. Beard, J. K. Young, M. A. Simmons, *Eur. J. Pharmacol.* **2008**, *592*, 1.

[67] M. Pellegrini, A. A. Bremer, A. L. Ulfers, N. D. Boyd, D. F. Mierke, *J. Biol. Chem.* **2001**, *276*, 22862.

[68] A. L. Ulfers, A. Piserchio, D. F. Mierke, *Biopolymers* **2002**, *66*, 339.

[69] G. Turcatti, H. Vogel, A. Chollet, *Biochemistry* **1995**, *34*, 3972.

[70] G. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel, A. Chollet, *J. Biol. Chem.* **1996**, *271*, 19991.

[71] A. W. Schmid, E. Condemi, H. Vogel, G. Tuchscherer, M. Mutter, R. Hamelin, D. Chiappe, M. Moniatte, Y. O. Tsybin, unpublished results.

[72] Y. O. Tsybin, C. L. Hendrickson, S. C. Beu, A. G. Marshall, *Int. J. Mass Spectrom.* **2006**, *255*, 144.

Paper II

Glycan variability on a recombinant IgG antibody transiently produced in HEK-293E cells

Sophie Nallet[1,3], Luca Fornelli[2,3], Simone Schmitt[2,4], Julien Parra[2,5], Lucia Baldi[1], Yury O. Tsybin[2] and Florian M. Wurm[1,]

[1] École Polytechnique Fédérale de Lausanne (EPFL), School of Life Sciences, Laboratory of Cellular Biotechnology, Station 6, CH-1015 Lausanne, Switzerland
[2] École Polytechnique Fédérale de Lausanne (EPFL), School of Basic Sciences, Biomolecular Mass Spectrometry Laboratory, CH-1015 Lausanne, Switzerland

In this study, a recombinant monoclonal IgG antibody was produced by transient gene expression (TGE) in suspension-adapted HEK-293E cells. The objective of the study was to determine the variation in recombinant IgG yield and glycosylation in ten independent transfections. In a ten-day batch process, the variation in transient IgG yield in the ten batches was less than 30% with the specific productivity averaging 20.2 ± 2.6 pg/cell/day. We characterized the N-glycosylation profile of each batch of affinity-purified IgG by intact protein and bottom-up mass spectrometry. Four major glycans were identified at Asn^{297} in the ten batches with the maximum relative deviation for a single glycoform being 2.5%. In addition, within any single transfection there was little variation in glycoforms over the ten-day culture. Our experimental data indicate that with TGE, the production of recombinant IgG with little batch-to-batch variation in volumetric yield and protein glycosylation is feasible, even in a non-instrumented cultivation system as described here.

Introduction

Two main strategies are available for the production of recombinant therapeutic proteins in mammalian cells, stable (SGE) and transient gene expression (TGE). The former is based on the constitutive expression of the protein from a clonal recombinant cell line having one or more integrated copies of the transgene and is the method of choice for the large-scale production of recombinant therapeutic proteins for the commercial market [1,2]. TGE is based on protein production from a pool of transiently transfected cells and is mainly used to generate small amounts of protein for research and/or for pre-clinical studies [3–5]. In contrast to SGE, TGE has not yet been used for the commercial production of therapeutic proteins. One of the major concerns

with bioprocesses based on TGE is the batch-to-batch variability in protein yield and quality, especially with regard to protein glycosylation. The purpose of this study was to determine the batch-to-batch variability in these two parameters for a transiently produced recombinant IgG antibody. The protein was produced in ten independent batches of transiently transfected HEK-293E cells. Affinity-purified protein was analyzed by SDS–PAGE and size exclusion chromatography (SEC). Mass spectrometry (MS) was used for the identification and relative quantitation of glycans.

Materials and methods

Cell culture and transfection

Four days before each transfection, one vial of frozen suspension-adapted HEK-293E cells was thawed for two 150-mL transfections. Cells were rinsed once with Ex-cell293 medium (SAFC, St. Louis, MO) and resuspended at 0.4×10^6 cells/mL in Ex-cell293 medium supplemented with 4 mM glutamine in an orbitally shaken 250-mL glass bottle (SciLabware Ltd., Stone, United Kingdom). Cells were incubated at $37°C$ in the presence of 5% CO_2 with orbital shaking at 110 rpm (ISF-4-X incubator, Kühner AG, Birsfelden, Switzer-

land) [6]. After three days, the cells were resuspended at 1×10^6 cells/mL in a 1-L glass bottle in Ex-cell293 medium with glutamine. Transfections were performed on the following day as previously described with some modifications [7]. Briefly, cells were centrifuged before transfection and resuspended at 20×10^6 cells/mL in 7.5 mL RPMI 1640 medium (Lonza, Verviers, Belgium) supplemented with 25 mM HEPES and 0.1% Pluronic F68 (Applichem GmbH, Darmstadt, Germany) and placed in 50-mL glass bottles equipped with home-made vented caps with 1.5 cm polyether sulfone membranes. To each bottle, 187.5 μg of plasmid DNA (2% pEGFP-N1 (Clontech, Palo Alto, CA) and 98% pXLG^{CHO}-A2 by weight) and 600 μg of linear 25 kDa polyethylenimine (PEI; 1 mg/mL in H_2O) (Polysciences, Eppenheim, Germany) were added separately. The vector pEGFP-N1 carries the enhanced green fluorescent protein (EGFP) gene. Plasmid pXLG^{CHO}-A2, a dual expression vector for the production of a human anti-Rhesus D IgG1 antibody, was constructed as described for pXLG^{CHO}-A3 [8], except that the expression cassettes for the light and heavy chain genes were in a head-to-tail rather than a head-to-head orientation. After DNA and PEI addition, the cultures were incubated at 37°C with agitation as before. After three hours the cells were transferred to a 500-mL glass bottle with a vented cap and diluted to a final cell density of 1×10^6 cells/mL by addition of Ex-cell293 medium to a total volume of 150 mL. Valproic acid (VPA) (Sigma–Aldrich, St. Louis, MO) was added to a final concentration of 3.75 mM, and the cultures were incubated as before. The cell density and viability were determined with the Trypan blue exclusion method using a Neubauer hemocytometer. The recombinant IgG concentration in the cell culture medium was determined by ELISA as previously described [9].

IgG purification and analysis by SDS–PAGE and size exclusion chromatography

Recombinant IgG was purified from the cell culture medium using Streamline Protein A according to the manufacturer's instructions (GE Healthcare, Uppsala, Sweden). Purified IgG was analyzed by reducing SDS–PAGE either before or after treatment of the protein with PNGase F for three hours at 37°C (New England Biolabs, Ipswich, MA). Analysis of purified IgG by SEC was performed on a Superdex 200 10/30 GL column connected to an Äkta system (GE Healthcare). Before injection 50 μg of purified IgG was diluted in 100 μL of running buffer (50 mM K_2HPO_4 and 150 mM NaCl, pH 8.0). The protein was eluted with running buffer at a flow rate of 0.4 mL/min and detected by UV absorbance. The IgG produced by TGE was compared to a control IgG produced for clinical trials in a stable Chinese hamster ovary (CHO) cell line that was kindly provided by the authors of the study [10].

IgG sample preparation for mass spectrometry

For the bottom-up MS analysis, IgG samples were dried using a SpeedVac system and resuspended (200 μL, 50 μg of protein) in Tris–HCl buffer, pH 8.8 (Calbiochem-Merck, Darmstadt, Germany). Protein aliquots were denatured and reduced in the presence of 4 M urea (Acros Organics, Thermo-Fischer Scientific, Geel, Belgium) and 10 mM DTT (SAFC) at 37°C for one hour with shaking and subsequently alkylated by incubation with 40 mM iodoacetamide (SAFC). Reduced and alkylated IgG was subjected to proteolysis with trypsin (0.5 μg, substrate-to-enzyme ratio of 100:1, w/w)

(Promega AG, Dübendorf, Switzerland) at 37°C for 18 hours. After denaturation of trypsin by heating for 5 min to 90°C, the samples were aliquoted in four fractions containing 12.5 μg of peptides each, dried, and stored at −80°C.

Mass spectrometry of IgG

Measurements of the intact IgG were performed on a electrospray ionization (ESI) – quadrupole time-of-flight mass spectrometer (MaXis qTOF MS, Bruker Daltonics, Bremen, Germany) operated in positive ion mode at a resolution of ∼50,000 and on-line coupled with a reverse phase liquid chromatography (RPLC) system (HPLC 1200, Agilent, Santa Cara, CA) equipped with a C$_8$ column for sample desalting. Mass spectra were smoothed and deconvoluted with BioTools and Max Entropy (Bruker Daltonics). Bottom-up analysis of the IgG tryptic digestions was performed according to the following protocol. After resuspension in 35 μL of water, 10 μL of a digested sample aliquot was loaded onto a porous graphitized carbon column (Hypercarb column, Thermo Scientific, Bremen, Germany) heated at 45°C. Peptides were separated over a binary gradient (from 100% to 55% mobile phase A in 36 min, followed by column wash and re-equilibration; mobile phase A: 95% H_2O, 4.9% acetonitrile (ACN, Fluka, Sigma–Aldrich) and 0.1% formic acid; mobile phase B: 95% ACN, 4.9% H_2O and 0.1% formic acid) at a flow rate of 150 μL/min. Peptide cations eluted from the column were ionized with a standard ESI source and subsequently detected on a linear ion trap Fourier transform ion cyclotron resonance (LTQ FT-ICR) mass spectrometer (Thermo Scientific) equipped with a 12 T superconducting magnet (Oxford Nanosciences, Oxon, UK). Broadband mass spectra were collected in positive ion mode at a resolution of 25,000 at 400 m/z over a m/z window of 700–2000. For the data analysis the Xcalibur software version 2.0.7 (Thermo Scientific) was used.

Label-free quantitation of glycopeptides

Label-free quantitation of glycopeptides was performed by applying a previously described method [11] adapted here for liquid chromatography (LC)–MS analysis. Briefly, a list of tryptic glycopeptide ions was created by considering the IgG N-glycosylation site (Asn^{297}, Kabat numbering [12]) and a maximum of two trypsin missed cleavages. Only biologically relevant oligosaccharidic structures were chosen by browsing a glycopeptide database (GlycoPep DB software, http://hexose.chem.ku.edu/sugar.php) [13] and other sources [14–17]. Glycopeptide masses were calculated by adding to the peptide mass the different glycan masses. The obtained masses were converted into protonated glycopeptide ion m/z values (from 2+ to 4+; Table 1). These theoretical values were used to validate the presence of glycopeptides in the sample mixture by using a threshold of theoretical-to-experimental mass accuracy of 3 ppm. For the relative quantitation of the detected glycoforms, the intensity of the peaks corresponding to glycopeptides were extracted from a specific area of the chromatogram (typically from 24 to 29 min) and transferred to Microsoft Excel. The intensities of the first four isotopic peaks were summed for obtaining a single value of intensity for each glycopeptide ion. The relative intensity of each ion was obtained by normalizing over the total glycopeptide intensity, and the relative abundance of each glycoform was calculated by summing the relative intensities of all the related glycopeptide ions.

TABLE 1

List of glycopeptide ions detected by LC–FT-ICR MS

Glycoform[a]	Peptide aa sequence[b]	Theoretical m/z[c] (monoisotopic)	Charge state
G1F	EEQYNSTYR	1398.5528	2+
	TKPREEQYNSTYR	1639.7018	2+
	TKPREEQYNSTYR	1093.4686	3+
	TKPREEQYNSTYR	820.3537	4+
G2F	EEQYNSTYR	1479.5794	2+
	TKPREEQYNSTYR	1147.4875	3+
	TKPREEQYNSTYR	860.8674	4+
G0F	EEQYNSTYR	1317.5265	2+
	TKPREEQYNSTYR	1558.6748	2+
	TKPREEQYNSTYR	1039.4523	3+
	TKPREEQYNSTYR	779.841	4+
	EEQYNSTYR	1378.0397	2+
	TKPREEQYNSTYR	810.0976	4+
	EEQYNSTYR	1203.4710	2+
	EEQYNSTYR	1386.0371	2+
	TKPREEQYNSTYR	1133.8120	3+
	EEQYNSTYR	1446.5503	2+
	EEQYNSTYR	1365.5239	2+
	EEQYNSTYR	1215.9869	2+
	EEQYNSTYR	1297.0133	2+

[a] (■) N-Acetylglucosamine; (●) mannose; (○) galactose; (▲) fucose.
[b] Glycosylation site located on Asn[297] (Kabat numbering [12]).
[c] Theoretical m/z values were calculated according to the procedure described in Materials and methods.

Results and discussion

The variability of the TGE bioprocess was assessed in ten independent transfections of 150 mL performed under the same conditions. For the ten batches, the cell density and viability varied by 16–26% (Fig. 1a) and the recombinant antibody concentrations varied by 24–30% over the ten-day culture (Fig. 1b). The calculated cell specific productivity in the ten batches averaged 20.2 ± 2.6 pg/cell/day. The recombinant IgG was purified from cell-free supernatants by affinity chromatography with Protein A and analyzed by SEC and SDS–PAGE. For each analysis, the IgG produced by TGE was compared to a control IgG produced in a CHO cell line [10]. On the chromatograms, only one sharp protein peak was observed at the same elution volume for all the samples, suggesting that the ten batches produced a protein of the same size (Fig. 2a). No other protein peak was observed, which indicated the IgG was pure and did not contain degradation products. The two other peaks observed at a later elution time represent salt peaks. For all samples resolved by reducing SDS–PAGE, two bands were observed at ~25 kDa and ~55 kDa, corresponding to the IgG light and heavy chains, respectively (Fig. 2b). Each IgG heavy chain had about the same extent of N-linked glycosylation as determined by treatment with PNGase F (Fig. 2b). No difference in the size was observed between the IgG produced by TGE in HEK-293E cells and by SGE in

CHO cells (Fig. 2a,b). The higher peak intensities for the CHO-produced antibody (Fig. 2a) can be explained by the fact that it was stored at a different concentration in a different buffer than the IgGs produced in HEK-293E cells.

To further analyze the variability of glycosylation among the ten batches of transiently produced antibody, we employed high resolution MS. First, analysis of intact IgG was performed by on-line LC–electrospray quadrupole time-of-flight (LC–ESI-qTOF) MS at a resolution of 50,000. The deconvoluted mass spectra showed that each IgG was composed of a complex mixture of several glycoforms with MWs between ~149,700 and 150,500 Da (Fig. 3a), which is in agreement with the theoretical MW calculated on the basis of the IgG light and heavy chain cDNA sequences. Moreover, the qualitative analysis of three selected batches clearly indicated that their glycosylation pattern was conserved (Fig. 3a). We then applied a label-free bottom-up methodology using FT-ICR MS coupled on-line with LC. This method, adapted from Rebecchi et al. [11], consisted in the LC–MS-based identification of the glycopeptides generated upon trypsin digestion of the intact IgG and their subsequent quantitation after normalization over the total glycopeptide intensity (see Materials and methods for additional details). This approach allowed us to identify both high mannose and complex oligosaccharides. However, no sialic acid-containing
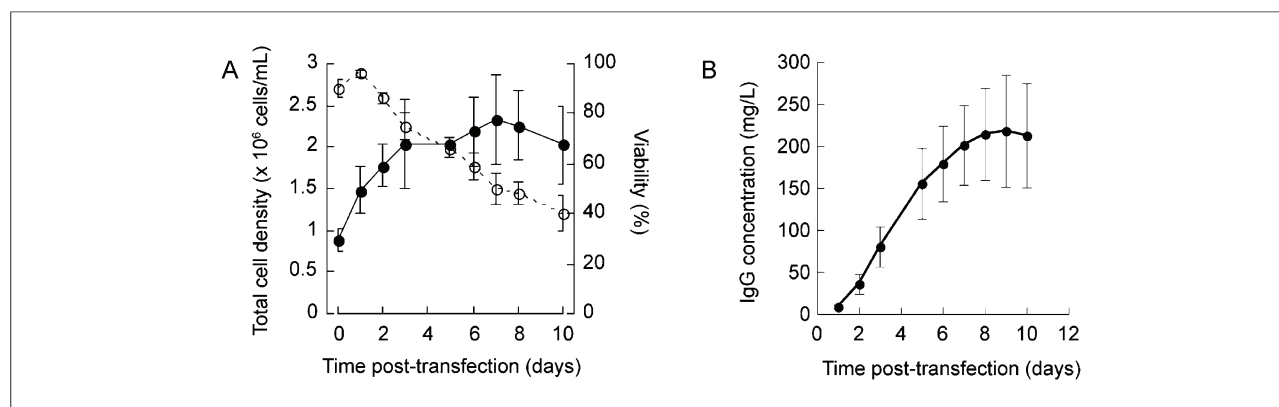
FIGURE 1

Cell growth and IgG production in ten independent TGE batches. HEK-293E cells were transfected and cultivated for ten days as described in Materials and methods. The data are expressed as the average of ten batches ± s.d. At the times indicated the cell density (closed circles) and viability (open circles) were measured by manual counting using the Trypan blue exclusion method **(a)** and the IgG concentration in the medium was measured by ELISA **(b)**.

glycopeptides were detected. This was probably due to their low ionization efficiency in positive mode. Eleven different glycans were detected, but four of these contributed 98 ± 1% of the total glycan amount (Table 1). These four species, represented in Fig. 3b with their relative abundances averaged over the ten batches [order of relative abundance: G1F (58.5%), G2F (20.8%), G0F (14.3%) and the mixed complex form (4.1%)], were fucosylated (the nomenclature

used for the glycoforms is described in [18]). Fucosylation is important for the activity of therapeutic antibodies as it can reduce their antibody-dependent cytotoxicity [19]. It is noteworthy that the relative ratio order of the four main glycans was maintained over nine of the ten batches with the only exception being batch 8, where G0F (21.3%) was slightly more abundant than G2F (17%) (Fig. 3c). Despite the generally low ionization efficiency of glycopeptides, the
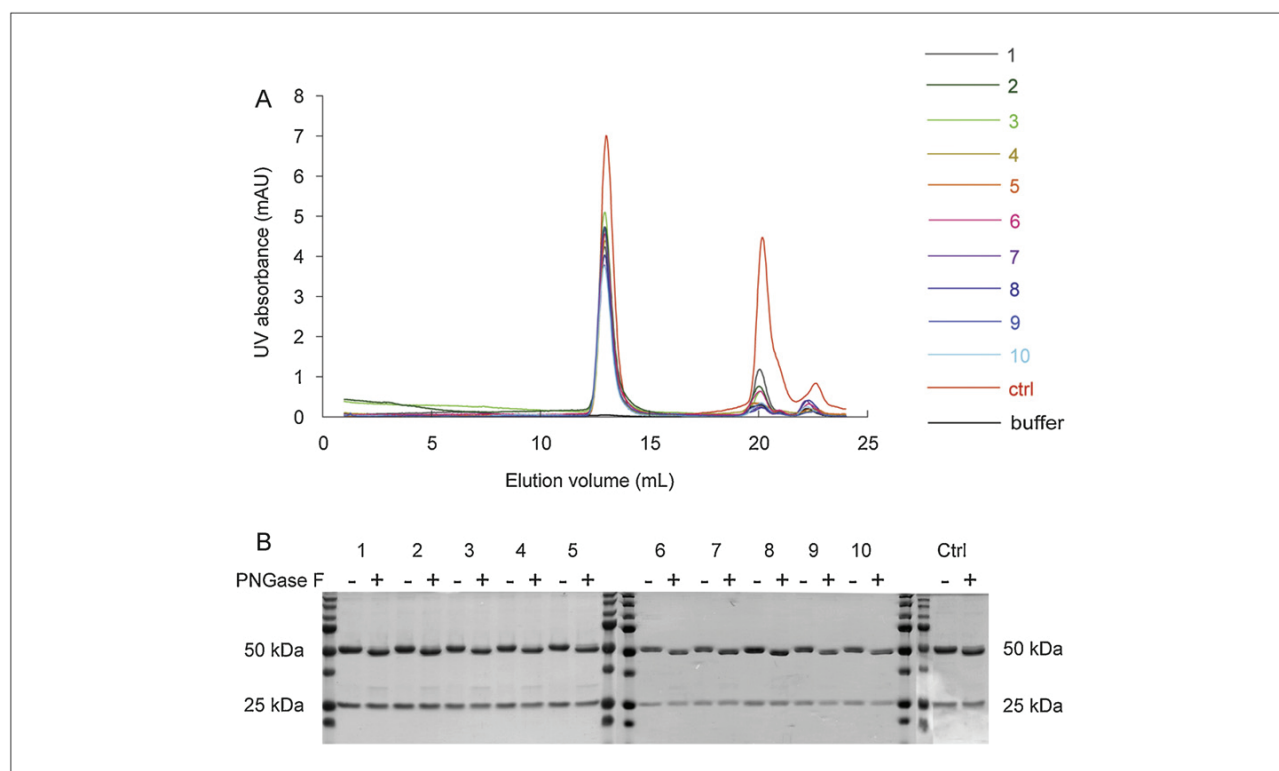
FIGURE 2

Analysis of the IgG produced in ten TGE batches and purified by affinity to Protein A. The control refers to the same IgG produced for clinical trials in a stable CHO cell line [10]. **(a)** SEC of purified IgGs. The buffer alone was used as a negative control. **(b)** Reducing SDS–PAGE of IgG either deglycosylated with PNGase F (+) or not (−).
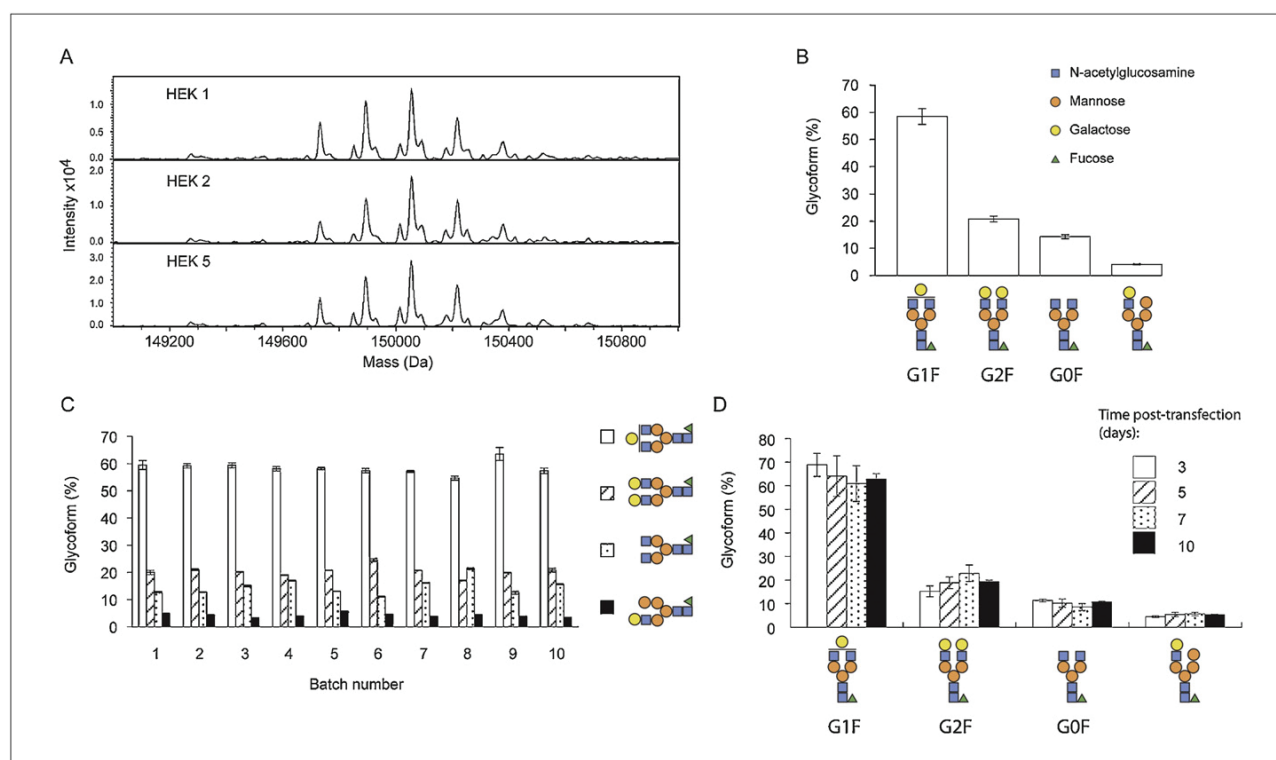
FIGURE 3

Mass spectrometric analysis of recombinant IgGs. **(a)** LC–ESI-qTOF MS-produced smoothed and deconvoluted mass spectra of TGE batches nos. 1, 2 and 5. **(b)** Relative abundance of the four main glycans detected in the IgG by FT-ICR MS applying label-free quantitation. The data are expressed as the average of ten batches ± s.d. **(c)** Relative abundances of the same four glycans among the ten batches expressed as the mean of three technical replicates ± s.d. **(d)** Comparison of the glycan relative composition in cell batches harvested at the times indicated. The data are expressed as mean of three technical replicates ± s.d.

maximum standard deviation over three technical replicates for these glycoforms (with samples analyzed according to a random sequence) was 9.1% and normally it remained below 6%. The maximum relative deviation for a single glycoform among the ten batches was only 2.5%.

Finally, to determine if the type and distribution of the glycoforms varied with time in cell culture, we compared the glycosylation profile of one batch harvested at three, five, seven and ten days after transfection. The results indicated that the ranking of the four main glycoforms was constant from three days post-transfection (Fig. 3d). This suggested that the *N*-linked glycans were stable in cell culture despite the decrease in cell viability over time (*cf*. Fig. 2a). The same observation has been reported for an IgG produced by TGE in CHO cells [20].

In conclusion, the results presented here show that a TGE process can be reproducible even in a non-instrumented cultivation system. As measured by LC–MS, the IgG produced by TGE in

HEK-293E cells had a glycosylation profile which was reproducible over ten batches and similar to glycosylation profiles reported in the literature for human IgGs [21,22]. Therefore, this study demonstrated that TGE processes, even when performed in simple, non-instrumented bioreactors, can yield a protein with little batch-to-batch variation in yield or quality. The relevance of sialic acids in glycosilation patterns has not been overlooked in this study; however, we focused here on the overall glycans detectable by the positive mode of MS. Our future work will also address the sialic acid content of recombinant proteins produced by TGE.

References

1 Walsh, G. (2010) Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* 28, 917–924
2 Wurm, F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* 22, 1393–1398
3 Baldi, L. *et al.* (2007) Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnol. Lett.* 29, 677–684
4 Geisse, S. and Fux, C. (2009) Recombinant protein production by transient gene transfer into mammalian cells. *Methods Enzymol.* 463, 223–238
5 Pham, P.L. *et al.* (2006) Large-scale transfection of mammalian cells for the fast production of recombinant protein. *Mol. Biotechnol.* 34, 225–237
6 Muller, N. *et al.* (2005) Orbital shaker technology for the cultivation of mammalian cells in suspension. *Biotechnol. Bioeng.* 89, 400–406

7 Backliwal, G. *et al.* (2008) High-density transfection with HEK-293 cells allows doubling of transient titers and removes need for a priori DNA complex formation with PEI. *Biotechnol. Bioeng.* 99, 721–727

8 Rajendra, Y. *et al.* (2011) A simple high-yielding process for transient gene expression in CHO cells. *J. Biotechnol.* 153, 22–26

9 Meissner, P. *et al.* (2001) Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnol. Bioeng.* 75, 197–203

10 Miescher, S. *et al.* (2004) A single recombinant anti-RhD IgG prevents RhD immunization: association of RhD-positive red blood cell clearance rate with polymorphisms in the FcgammaRIIA and FcgammaIIIA genes. *Blood* 103, 4028–4035

11 Rebecchi, K.R. *et al.* (2009) Label-free quantitation: a new glycoproteomics approach. *J. Am. Soc. Mass Spectrom.* 20, 1048–1059

12 Kabat, E.A. *et al.* (1991) *Sequences of Proteins of Immunological Interest* (5th edn), NIH Publication n. 91-3242

13 Go, E.P. *et al.* (2007) GlycoPep DB: a tool for glycopeptide analysis using a Smart Search. *Anal. Chem.* 79, 1708–1713

14 Jefferis, R. (2009) Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev. Drug Discov.* 8, 226–234

15 Raju, T.S. *et al.* (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* 10, 477–486

16 Stadlmann, J. *et al.* (2008) Analysis of immunoglobulin glycosylation by LC–ESI-MS of glycopeptides and oligosaccharides. *Proteomics* 8, 2858–2871

17 Zhang, Z. *et al.* (2009) Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom. Rev.* 28, 147–176

18 Krapp, S. *et al.* (2003) Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.* 325, 979–989

19 Ferrara, C. *et al.* (2006) Modulation of therapeutic antibody effector functions by glycosylation engineering: influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-*N*-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnol. Bioeng.* 93, 851–861

20 Galbraith, D.J. *et al.* (2006) Control of culture environment for improved polyethylenimine-mediated transient production of recombinant monoclonal antibodies by CHO cells. *Biotechnol. Prog.* 22, 753–762

21 Flynn, G.C. *et al.* (2010) Naturally occurring glycan forms of human immunoglobulins G1 and G2. *Mol. Immunol.* 47, 2074–2082

22 Scherer, H.U. *et al.* (2010) Glycan profiling of anti-citrullinated protein antibodies isolated from human serum and synovial fluid. *Arthritis Rheum.* 62, 1620–1629

Research Paper

Paper III

RESEARCH PAPER

Top-down analysis of 30–80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry

Luca Fornelli · Julien Parra · Ralf Hartmer · Carsten Stoermer · Markus Lubeck · Yury O. Tsybin

Abstract Electron transfer dissociation (ETD)-based top-down mass spectrometry (MS) is the method of choice for in-depth structure characterization of large peptides, small- and medium-sized proteins, and non-covalent protein complexes. Here, we describe the performance of this approach for structural analysis of intact proteins as large as the 80 kDa serotransferrin. Current time-of-flight (TOF) MS technologies ensure adequate resolution and mass accuracy to simultaneously analyze intact 30–80 kDa protein ions and the complex mixture of their ETD product ions. Here, we show that ETD TOF MS is efficient and may provide extensive sequence information for unfolded and highly charged (around 1 charge/kDa) proteins of ~30 kDa and structural motifs embedded in larger proteins. Sequence regions protected by disulfide bonds within intact non-reduced proteins oftentimes remain uncharacterized due to the low efficiency of their fragmentation by ETD. For serotransferrin, reduction of S–S bonds leads to significantly varied ETD fragmentation pattern with higher sequence coverage of N- and C-terminal regions, providing a complementary structural information to top-down analysis of its oxidized form.

L. Fornelli · J. Parra · Y. O. Tsybin (✉)
Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
e-mail: yury.tsybin@epfl.ch

R. Hartmer · C. Stoermer · M. Lubeck
Bruker Daltonics GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany

Present Address:
J. Parra
Laboratory for Analysis and Modelling for Biology and Environment, CNRS UMR 8587, Université d'Evry Val d'Esonne, 91025 Evry Cedex, France

Introduction

The term top-down mass spectrometry (TD MS) refers to the structure analysis of intact protein ions by their fragmentation in the gas phase [1]. When applied to a mixture of proteins from different protein families or proteoforms of a single protein family, TD MS translates into a top-down proteomics (TDP). TDP presents several theoretical advantages over bottom-up mass spectrometry-based proteomics (BUP), which consists in the analysis of peptide mixtures derived from proteolysis [2, 3]. In both TDP and BUP, the survey scan (MS[1] or simply MS) yields information about the total mass of biomolecules, whereas the tandem MS (MS/MS) step confirms the sequence of the selected precursor ion. TDP thus identifies a specific proteoform, whereas BUP is typically employed in identification of protein families [4].

TD MS and TDP analyses are technically more challenging since proteins are generally more difficult to separate, ionize, and fragment than peptides. In addition, the resulting MS/MS spectra are convoluted and data interpretation is not trivial. Recent introduction of novel in-solution protein fractionation techniques, e.g., gel-eluted liquid fraction entrapment electrophoresis [5], or the adoption of new stationary phases for liquid chromatography (LC) [6] have partially filled the gap between peptide and protein fractionation and separation. However, for MS data acquisition and analysis, new developments and improvements are needed. In terms of hardware, only high-resolution instruments are capable of resolving and distinguishing complex product ion clusters in MS/MS spectra of proteins. In terms of software, specific data analysis workflows have to be applied to efficiently reconstruct and

combine the information from MS^1 and, most importantly, MS/MS.

Although all the common ion activation methods can be applied to intact proteins, electron-based fragmentations, specifically electron capture dissociation (ECD) [7, 8] and electron transfer dissociation (ETD) [9], have been shown to produce a higher sequence coverage than "slow-heating" methods (such as collision-induced dissociation [10] or infrared multiphoton dissociation [11]) and also have the capability of cleaving disulfide bridges [12]. Importantly, ECD and ETD perform most efficiently (i.e., with low ion activation time resulting in an extensive backbone fragmentation) with highly charged precursors, which is the typical condition present during TD MS or TDP where highly charged protein cations are analyzed. Recently, developments in ultraviolet photodissociation (UVPD) and surface-induced dissociation MS/MS demonstrated intriguing capabilities specifically for intact protein analysis [13].

Originally, radical-driven TD MS of large proteins was introduced using MS/MS "in time" because the required high resolution was only available for Fourier transform ion cyclotron resonance mass spectrometers (FT-ICR MS) and the ability to perform radical-induced fragmentation was only available in ion traps, e.g., ICR ion trap or Paul ion trap [14, 15]. Beginning with comprehensive analysis of ~10-kDa model proteins, such as ubiquitin [16], the application range of ECD-based TD MS was significantly expanded, delivering a single-amino acid resolved structure of a 29 kDa carbonic anhydrase [17, 18] and further analysis of proteins in the 50–70 kDa range [19]. TD MS of proteins in these mass ranges can be now performed on the time scale of on-line LC.

The quest for radical-driven fragmentation in the widespread, low resolution quadrupole (q) mass analyzers and product ion analysis in the high-resolution mass analyzers led to the development of radical-driven top-down mass spectrometry "in space". After numerous attempts, ECD was efficiently implemented in a quadrupole mass analyzer that could be coupled to a higher resolution time-of-flight mass spectrometer (TOF MS) [20, 21]. However, the real revolution in TD MS "in space" was made by implementing ETD first in a linear ion trap and then in other quadrupole mass analyzers [22]. Currently, the most successful implementations of ETD specifically for protein analysis include ETD on hybrid qTOF MS [23, 24], LTQ Orbitrap FTMS [25] and qFT-ICR MS platforms [26]. These recent technologies have already shown strong potential applications, including the structural analysis of ~150 kDa intact immunoglobulins G (IgG) on different platforms [27, 28].

The ETD fragmentation and subsequent mass analysis using a qTOF MS of bovine carbonic anhydrase and human serotransferrin are presented herein. Carbonic anhydrase was used as a model for medium-size proteins, which can be analyzed without specifically designed sample preparation or instrumental set-up. Serotransferrin was used to challenge the instrument capabilities with a large protein and investigate the performance of ion manipulation, fragmentation, and detection. Structural study of this protein is complicated in part because its large MW and the presence of 19 disulfide bridges hinders ion–ion interactions in ETD. This structural complexity manifested itself by significant differences in the ETD fragmentation patterns depending on serotransferrin oxidation state. To characterize these proteins in detail we applied dedicated data analysis workflow based on the SNAP2 algorithm (Bruker Daltonics, Bremen, Germany). Our results thus demonstrate the current capabilities of high-resolution qTOF MS in sequence analysis of large proteins by TD MS.

Experimental

Sample preparation Bovine carbonic anhydrase and human serum transferrin (serotransferrin) were obtained commercially (Sigma, Taufkirchen, Germany). For direct infusion experiment of carbonic anhydrase and non-reduced serotransferrin, proteins were dissolved in water/acetonitrile/formic acid (50:49.9:0.1, *v/v/v*) to a final concentration of ~1 μM and used without further purification. Transferrin reduction was obtained by dissolving the lyophilized protein in 50 mM sodium carbonate buffer in presence of 5 mM *tris*(2-carboxyethyl)phosphine (TCEP, Sigma). Prior to infusion, the reduced protein was desalted by centrifugation with microcentrifuge 10 kDa MW cut-off filters (Millipore, Zug, Switzerland).

Time-of-flight tandem mass spectrometry Mass measurements of intact protein ions and ETD MS/MS were performed on a high-resolution, high-mass accuracy quadrupole time-of-flight (qTOF) mass spectrometer (maXis UHR qTOF MS, Bruker Daltonics GmbH, Bremen, Germany) equipped with electrospray ion source. ETD MS/MS was performed as previously described [24]. The main steps in top-down ETD TOF MS are schematically shown in Fig. 1. The differences in ETD of high-mass proteins compared to smaller analytes are: (1) prolonged cation accumulation and radical anion injection times to ensure the accumulation of a sufficient number of analyte ions and, at the same time, the possibility of efficiently inducing ETD; (2) decreased ion–ion reaction time as ETD proceeds faster when highly charged ions are used as precursors. After optimization (see Supplementary Information), precursor ion accumulation time was increased up to 800 ms, anion injection time was set to 100 ms, and additional ion–ion interaction time to 40–50 ms. Product ions were transferred to the ion cooler cell from the reaction cell and orthogonally pushed to the TOF mass analyzer at the rate of 10 kHz. The orthogonal acceleration potential was at 10 kV. Acquisition in a broad *m/z* window (100–5,000 *m/z*) was performed with signal digitization at 2 GHz, whereas higher resolution MS and MS/
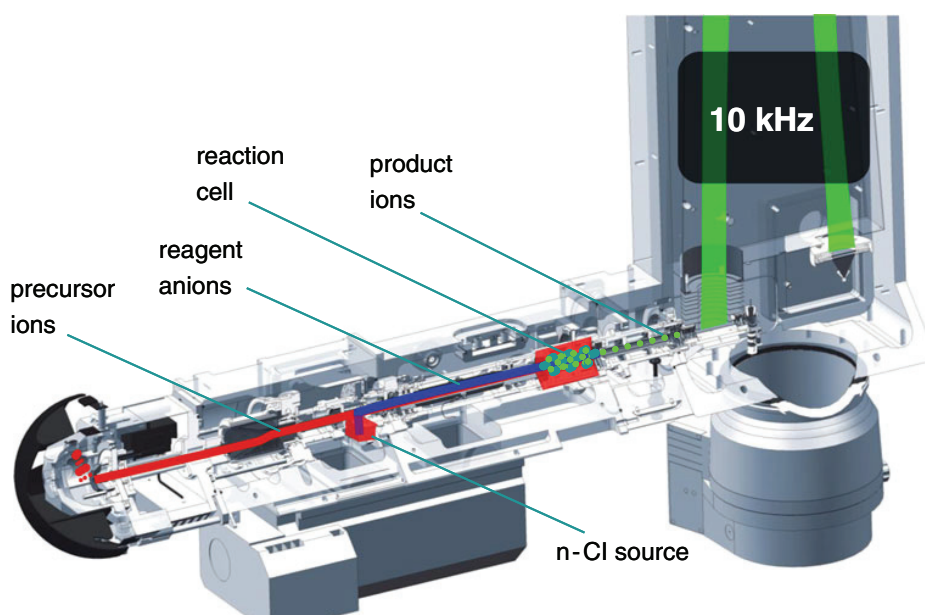
Fig. 1 Sequence of top-down electron transfer dissociation tandem mass spectrometry on a qTOF MS. Step 1: electrospray ionization of the samples in the ion source and precursor ion accumulation in the reaction cell (800 ms). Step 2: ETD reagent injection from the chemical ionization source to the reaction cell (20–100 ms). Step 3: ion–ion reaction in the reaction cell (20–100 ms). Step 4: ion–ion reaction product ions transfer into the ion cooler. Step 5: ion extraction from the ion cooler into the orthogonal acceleration region. Step 6: orthogonal ion extraction into the time-of-flight tube with an iron mirror (10 kHz, 10 kV). Step 7: ion detection at the multichannel plate with time to digital converter (2 or 4 GHz sampling rate)

MS data were acquired at a faster digitization rate of 4 GHz for narrower m/z range of 500–3,000 m/z. Intact protein and product ion mass spectra were acquired in separate direct infusion experiments. For the larger protein, data acquisition time of up to 30 min was employed to average up to 1,000 MS/MS spectra.

Data analysis The MS and MS/MS spectra were analyzed using the dedicated top-down data analysis procedure. Briefly, precursor and product ion mass spectra were summed over the infusion time in each MS and MS/MS experiment with Data Analysis software version 4.0 (Bruker Daltonics). Intact protein mass spectra were submitted to the MaxEntropy deconvolution procedure to yield heterogeneous distribution of neutral (deprotonated) glycosylated proteoforms.

Before ETD data analysis, the SNAP 2 algorithm was used to simulate MS/MS spectra with theoretical isotopic distributions of the peaks that matched the experimental mass spectra with a correlation coefficient of 0.4 or better. Charge states and monoisotopic masses were assigned automatically for the product ion isotopic distributions in both the experimental and theoretical mass spectra. Manual analysis considering both experimental and theoretical mass spectra was performed to assign ETD-specific *c*-type and *z*-type product ions. Alternatively, automatic data analysis was performed by deisotoping the simulated isotopic clusters, keeping only the monoisotopic masses. Following charge state deconvolution, the obtained mass list of product ions was searched against the known sequence of the proteins to determine the product ion identity and the sequence coverage. The results of both manual and automatic data analysis were compared and pooled to generate the verified protein sequence coverage map. Internal product ions, formed by secondary fragmentation of the primary product ions, have not been considered in the current work.

Results and discussion

Evaluation of ETD top-down mass spectrometry workflow The complete workflow for TD MS includes: (1) sample preparation, (2) intact protein mass measurement in MS mode, (3) sequence analysis in MS/MS mode, and (4) data analysis. The present work employed a state-of-the-art, ETD-enabled high-resolution qTOF mass spectrometer (Fig. 1). Some of its figures of merits for intact protein analysis are: resolution up to 50,000 over a wide m/z range in MS[1], and up to 30,000–40,000 for MS/MS (the resolution is influenced by the complexity and the abundance of the product ion population); a scan rate up to 20 Hz; mass accuracy of ~1 ppm (for both MS[1] and MS/MS). Remarkably, TOF mass analyzers generally have a higher dynamic range than FT-based instruments, and this characteristic is particularly important for intact protein analysis: for the ~9-kDa ubiquitin the ratio between the monoisotopic and the most abundant peaks is ~10, but the same

ratio increases to ~6E4 for ~29 kDa carbonic anhydrase, ~E15 for ~79 kDa transferrin, and up to ~E32 for a ~150-kDa IgG.

Figure 2 shows the ETD mass spectrum of 35^+ precursor ion of carbonic anhydrase. Most intense product ions are shown in the expanded segments. Generally, the envelope of the undissociated charge-reduced species (also referred to as "ETnoD products") in ETD TD MS of small to medium proteins is positioned in the low m/z range (~500–1,500 m/z) therefore overlapping with typical product ions of ~300–1,400 m/z. Here, three different intense ETnoD products (with charge states 34^+, 33^+, and 32^+) were detected, impeding the product ion observation between ~850 and 920 m/z.

The complexity of the product ion mixture for ETD TD MS of carbonic anhydrase is illustrated in Fig. 3, top panel. Several partially overlapping clusters of multiply charged ions can be observed in an expanded 675–684 m/z window. Due to the high density of peaks, the first data analysis step is to filter the ion clusters from noise or spurious peaks by matching the experimentally obtained clusters with theoretical peaks (Fig. 3, middle panel). For a successful noise filtration, high-mass accuracy and

resolution are needed. In addition, high signal-to-noise signal is required to confidently identify the charge state and possibly the monoisotopic peak of each ion cluster. For carbonic anhydrase, the qTOF TD MS allowed the assignment of product ions of up to 14^+ charge state. Software parameters such as signal-to-noise cut-off can be manually selected to enhance peak identifications. The SNAP2 algorithm matches the experimental and theoretical distributions applying a user-defined correlation coefficient. Due to the complexity of product ion population for a 29 kDa protein, this coefficient is usually set to 0.5 or higher (equivalent to 50 % correlation), but this value can be reduced for larger proteins, where generally the complexity of product ion mixture is higher and the average signal-to-noise ratio is lower. Finally, as illustrated in Fig. 3, bottom panel, the identified clusters are deconvoluted and deisotoped, as only the monoisotopic masses are used for the database search or, alternatively, for de novo sequencing. In all our experiments, we imported the sequence of the employed proteins in the software BioTools (Bruker Daltonics) to reduce the time needed for a complete database search.
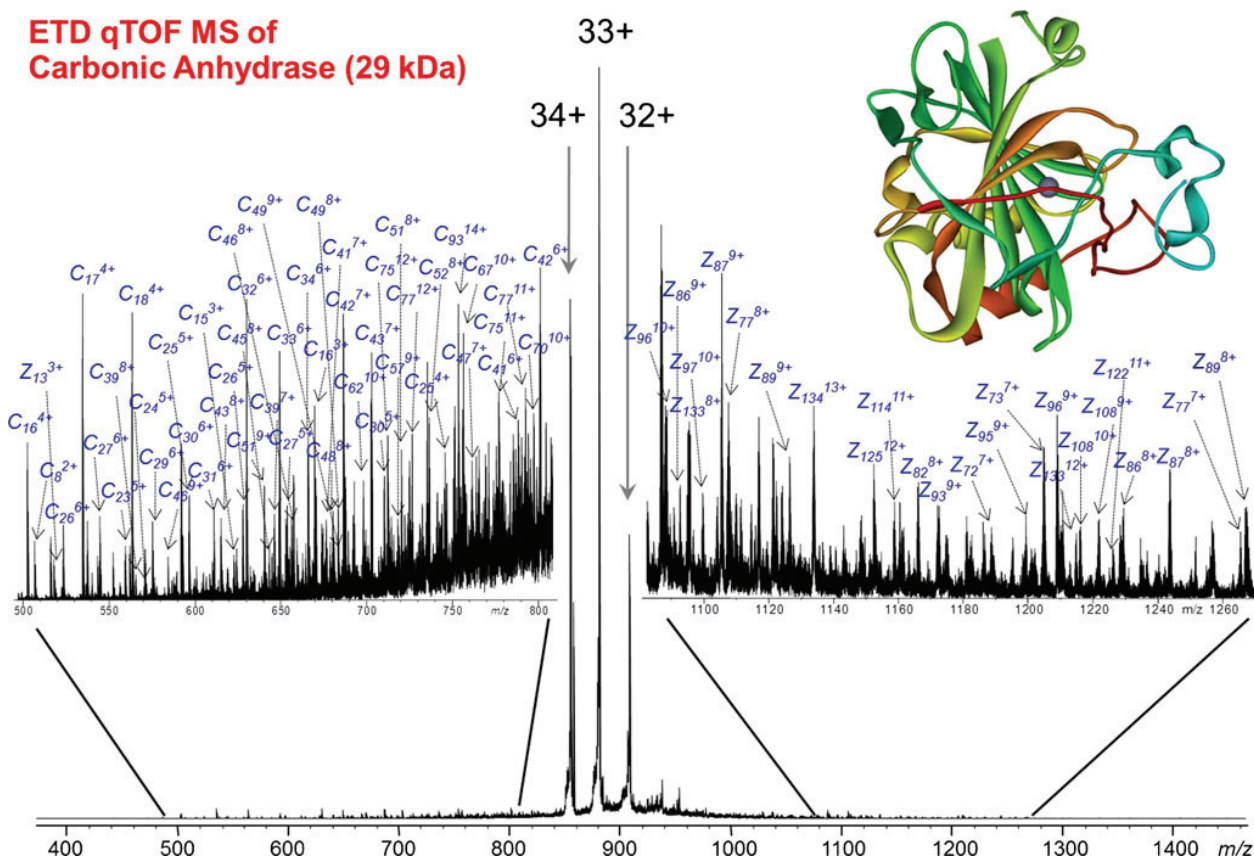
Fig. 2 Top-down electron transfer dissociation (ETD)-based ESI qTOF MS/MS of carbonic anhydrase (29 kDa). Prior to ETD, a single charge state, $[M+35H]^{35+}$, of carbonic anhydrase was selected with an isolation window of 30 m/z. Efficient charge reduction upon electron transfer resulted in electron transfer products without dissociation (ETnoD products), specifically $[M+35H]^{34+\bullet}$, $[M+35H]^{33+\bullet\bullet}$, and $[M+35H]^{32+\bullet\bullet\bullet}$ ions. The *insets* show the expanded regions of the MS/MS mass spectra

Fig. 3 Top-down MS/MS data analysis procedure. Example is made for ETD qTOF MS/MS of 29 kDa carbonic anhydrase protein, m/z=675–684 (Fig. 2). Step 1: raw data file analysis and matching of theoretical and experimental isotopic clusters. Step 2: deconvolution and deisotoping of the selected theoretical isotopic clusters. Step 3: protein theoretical product ion database search and/or de novo sequencing based on the monoisotopic masses and charge states data (product ion assignment)
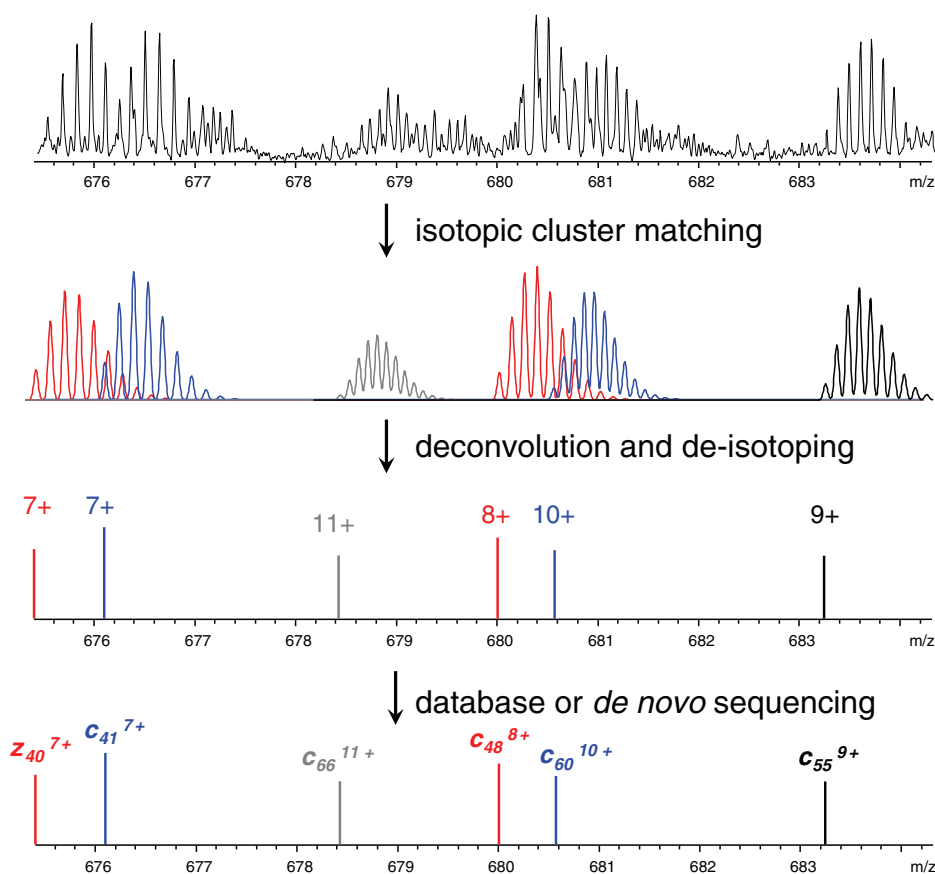
Top-down mass spectrometry of oxidized transferrin Mature human serotransferrin (UniProt entry: P02787) is a 679 amino acid-long protein, Scheme 1. First, the intact mass of non-

reduced transferrin was determined (Fig. 4, insets). The observed charge state distribution was centered at ~2,200 m/z and included charge states 22–50$^+$ with the most abundant

Scheme 1 Fragmentation map of intact transferrin after ETD-based ESI qTOF MS/MS. Backbone cleavage sites identified uniquely for the oxidized form of serotransferrin are indicated in *black*, whereas cleavages characteristic of the reduced form are indicated in *red*. Common cleavage sites are shown in *green*. Disulfide bridges are indicated by *purple dotted lines*
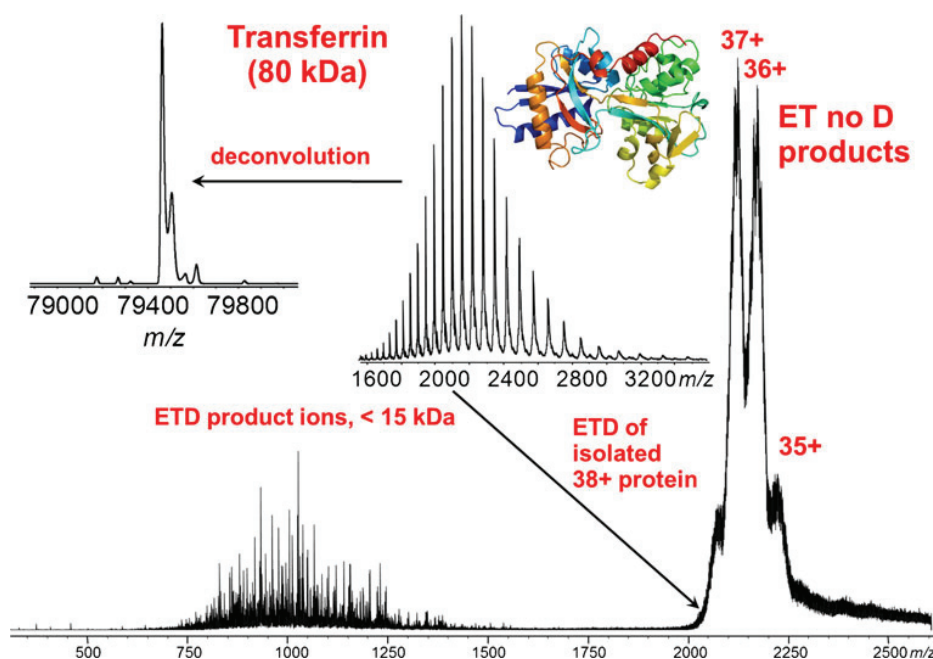
Fig. 4 Top-down electron transfer dissociation (ETD)-based ESI qTOF MS/MS of transferrin (79 kDa). The *insets* show ESI qTOF MS of intact transferrin with and without deconvolution. Prior to ETD, a single charge state, $[M+38H]^{38+}$, of transferrin was selected with an isolation window of 30 m/z. Efficient charge reduction upon electron transfer resulted in electron transfer products without dissociation (ETnoD products), specifically $[M+38H]^{37+\bullet}$, $[M+38H]^{36+\bullet\bullet}$, and $[M+38H]^{35+\bullet\bullet\bullet}$ ions. Product ions are observed as a group primarily in the $m/z=700–1,600$ region. Due to their high charge state and lower abundance compared to the ETnoD charge-reduced precursor ions these ions are not resolved. There are no product ions detected above 2,500 m/z

charge states at 35–38$^+$. This corresponded to a charge density of only 1 proton/2.1 kDa. The relatively low protonation level can be presumably explained by the compact conformation retained by the protein in the liquid as well as in the gas phase due to the constraints from S–S bridges, which limit the accessibility of potential protonation sites. The deconvoluted mass spectrum showed a main peak at ~79.5 kDa (Fig. 4), left inset. Intact protein mass measurements showed the presence of additional, less abundant, proteoforms, presumably due to protein glycosylation. Although the proteoform characterization is beyond the scope of this work, it might be useful to recall that the capabilities of qTOF MS in determining with high mass accuracy protein glycoforms have already been demonstrated for 150 kDa intact IgGs (see Electronic Supplementary Material, Fig. S1 of reference [27]).

One of the most intense precursor ions, the $[M+38H]^{38+}$ ion, was selected for ETD TOF MS/MS. Fragmentation parameters, specifically (1) cation accumulation time, (2) ion–ion interaction time and (3) radical anion injection time were optimized to maximize the protein fragmentation efficiency. The optimum transferrin cation accumulation time was found to be at ~800 ms, which ensured a sufficient precursor ion population. A similar value was previously applied also to the ETD of ~150 kDa IgG [27]. The ion–ion interaction time was found to be optimal for ETD of transferrin between 40 and 50 ms. At these interaction times, ETD process resulted in the

formation of a significant product ion population without excessive internal fragment production, and incomplete dissociation of the charge-reduced species (Fig. S1, Electronic Supplementary Material). The radical anion reaction time can deeply influence the ETD efficiency as it determines the final reagent anion number involved in ETD. For a fixed ion–ion interaction time, a low anion injection time impedes ETD, and conversely at too long injection times the occurrence of multiple, consecutive electron transfer events is highly enhanced, so that many charge-reduced species, low-charged product ions and internal fragments are produced. The final applied value was 100 ms (Fig. S2, Electronic Supplementary Material). The ETD mass spectrum of oxidized transferrin is shown in Fig. 4, bottom. Charge-reduced species (particularly, $[M+38H]^{37+\bullet}$, $[M+38H]^{36+\bullet\bullet}$, and $[M+38H]^{35+\bullet\bullet\bullet}$ ions) dominated the mass spectrum in the high m/z region, whereas the product ions were grouped and centered at ~1,000 m/z. Notably, in contrast to what is typically observed for a smaller protein, e.g., carbonic anhydrase, the highly abundant ETnoD products were positioned above 2,000 m/z, thus they did not interfere with the product ion detection and analysis. Most product ions were located below ~1,700 m/z.

An expanded view of the ETD mass spectrum of transferrin containing most of the product ions is presented in Fig. 5. The top panel illustrates the capabilities of the qTOF mass spectrometer to isotopically resolve multiply charged product ions

Fig. 5 Expanded views on ETD product ions of transferrin obtained on a qTOF MS. (*Bottom*) expanded segment, $m/z=700–1,500$, shows ETD product ions that are assigned to be mainly c-type product ions of transferrin in charge states from 5+ to 14+. Product ions c_{69} and c_{56} are particularly abundant and present in a number of charge states. (*Top*) expanded segment, $m/z=975–1,015$, shows efficient separation and baseline resolution of multiply charged ETD product ions. For example, the inset shows a well-resolved isotopic distribution of c_{55}^{6+} product ion at $\sim1,000\,m/z$. The annotated here MS/MS peaks are included into the protein sequence coverage summary presented in Scheme 1
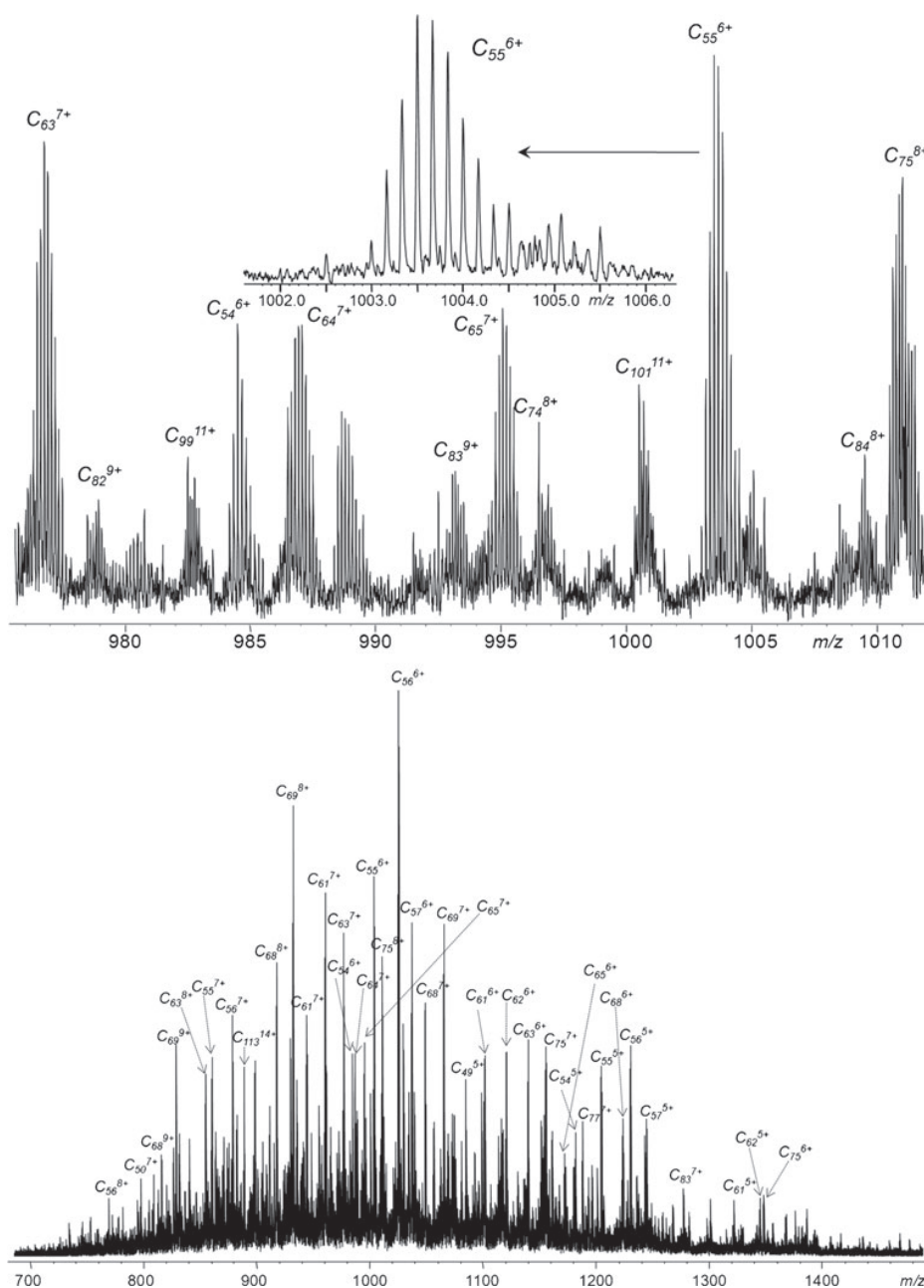
present in complex mixture. The assignment of the most abundant product ions shown in the bottom panel indicates that at the applied ETD conditions product ions up to 14^+ in charge state were detected, with charge states primarily between 5^+ and 8^+. Almost all identified product ions were c-type, containing information about the N-terminus of the protein. The relative position of the disulfide bridges may explain the fragmentation pattern observed. Close to the C-terminus, two Cys residues (Cys_{665} and Cys_{674}) are located. These are believed to be involved in the formation of distal S–S bridges with Cys_{474} and Cys_{402}, respectively. The

entire C-terminal portion of serotransferrin is folded in a tight conformation by a number of additional S–S bonds present between Cys_{402} and Cys_{674}. The extent of disulfide-protected sequence is therefore most likely responsible for the formation of only few z-type ions. Particularly, most z-ions identified derived from backbone cleavages occurring at the C-terminal side of Cys_{665} (i.e., with no need of cleaving a disulfide bond, but only the peptide bond for product ion formation), with the notable exceptions of z_{18}, z_{75} and z_{99}.

Conversely, at the N-terminus there was a 69-residue loop completely disulfide-free (between Ile_{49} and Ser_{117}), which

interconnects the first disulfide-protected area (from Cys_9 to Cys_{48}, with two S–S bonds) to the following one (starting from Cys_{118}, which is linked to Cys_{194}). As expected, most of the identified backbone cleavages arose from the unprotected linker, which was almost fully sequenced, and the small N-terminal sequence preceding Cys_9. Only four cleavages in the first disulfide-protected area were identified (i.e., ions c_{28}, c_{37}, c_{43}, and c_{44}), and no product ions derived from the second area were observed. Overall, we identified 62 c-ions and 11 z-ions, for a total sequence coverage of ~11 % (Table 1).

The capability of ETD (and, more generally, radical-driven ion activation methods) to cleave disulfide bonds has been confirmed to some extent in these experiments with transferrin, in addition to those reported earlier for IgGs. Of 73 backbone cleavages reported for the oxidized serotransferrin, 13 derived from disulfide-protected regions, and were mainly located towards the C-terminus. However, recent results by Ganisl and Breuker suggest that, at least when applied to proteins, ECD does not preferentially nor frequently cleaves S–S bonds rather than peptide backbone ones [29], as reported in the past for peptides [12, 30]. Nevertheless, the controversy may be limited to the term "preferential", as backbone cleavages from S–S bond-protected regions for many of the investigated proteins were reported. The authors explain these cleavages as resulting from a contribution from either vibrational fragmentation prior to ECD, or secondary radical fragmentation in ECD. Due to the differences in ETD and ECD processes, especially for what concerns the pressure of the ion traps where ion activation and fragmentation take place, further studies are needed to address this interesting question in relation specifically to ETD.

Top-down mass spectrometry of reduced transferrin Top-down experiment on disulfide-reduced serotransferrin was performed by direct infusion of a protein sample treated overnight with TCEP (Fig. 6). As expected, the reduction of S–S bonds induced the exposure of protonation sites not accessible to the acidic solvent in the oxidized protein. As a result, the charge state envelope of the reduced intact protein is significantly shifted towards lower m/z in respect to that of the oxidized counterpart. A large number of high charge state ions

Table 1 Summary of the assigned unique backbone cleavages for oxidized or reduced transferrin, and for the combination of the two datasets.

Serotransferrin form	N-terminal fragments	C-terminal fragments	Total sequence coverage (%)
Oxidized	62 (4)[a]	11 (9)[a]	10.7
Reduced	34	39	10.7
Combined	82	41	18.1

[a] For the oxidized form, the number of backbone cleavages from disulfide-protected regions is indicated within parenthesis

were detected (Fig. 6, right inset). Specifically, the charge state distribution included charge states from 40^+ to 100^+, with the most intense transferrin ions located around 900 m/z and carrying 83–88 protons.

To maximize the fragmentation efficiency in radical-driven mass spectrometry, high charge state precursors are usually preferred. Nevertheless, positioning the precursor isolation window around 1,000 m/z resulted in a partial overlapping between the ETnoD products and the heavy product ions. An intensity difference was observed between the charge-reduced species and dissociation products. This impeded the identification of fragments above 1,000 m/z, therefore we opted for a 300 m/z isolation window centered at 1,285 m/z, to include a number of charge states of the precursor (from 56^+ to 70^+) (Fig. S3, Electronic Supplementary Material). These precursor ions ensured improved fragmentation efficiency and the presence of ETnoD products in a region of the mass spectrum not occupied by product ions, beyond 1,400 m/z. Different from the oxidized transferrin, the ETD mass spectrum of the reduced protein showed a large m/z distribution of the product ions, which were not grouped in a specific, limited region but instead spread between ~200 and 1,300 m/z, with a variation of intensity that generally is not higher than two- to three-folds over the whole m/z range (Fig. 6, left insets). Since the same ETD fragmentation parameters (injection and interaction times) were applied for both the lower charge states oxidized, and higher charge state reduced forms, it is likely that for reduced transferrin analysis multiple electron transfer events occurred on already formed product ions, reducing their charge state and increasing their final mass-to-charge ratio. This is supported by the identification and assignment of several $z+1$ ions, generally not detected for the oxidized protein, and presumably obtained through a charge reduction event by secondary electron transfer on a primary product ion.

As summarized in Scheme 1, the distribution of the assigned product ions for ETD of reduced serotransferrin was more balanced between c- and z-type ions. Regarding the N-terminal fragments, the backbone cleavages observed were complementary to those previously identified for the oxidized transferrin. Specifically, the protein motif comprised between Cys_9 and Cys_{48} was highly sequenced (18/30 backbone cleavages assigned). On the contrary, the loop between Cys_{48} and Cys_{118} was poorly sequenced. The number of z-type ions was increased, with good sequence coverage in the region between Cys_{620} and the C-terminus. This motif included also Cys_{637}, which is linked to Cys_{418} in the oxidized protein; a non-complete reduction of this bond might explain the poor or absent sequence coverage for the sequence preceding Cys_{637}. In general, for the reduced protein the average size of assigned product ions was smaller than for the oxidized counterpart, likely because of a modified conformation and charge location. Here, the only "high-mass" product ions assigned were c_{101}, c_{102} and c_{104} (which were still lighter than

Fig. 6 Top-down electron transfer dissociation (ETD)-based ESI qTOF MS/MS of disulfide bond reduced serotransferrin (79 kDa). The *right inset* shows ESI qTOF MS data of intact S–S bond reduced transferrin without deconvolution demonstrating a significant shift in precursor ion charge states toward increased protein charging. Prior to ETD, a number of transferrin precursor ions in different charge states, from about $[M+56H]^{56+}$ to $[M+70H]^{70+}$, were selected with an isolation window of $\pm 150\,m/z$ around m/z 1,285. Efficient charge reduction upon electron transfer resulted in electron transfer products without dissociation (ETnoD products). The observed product ions occupy the m/z region from charge-reduced precursor ions to the lowest m/z detected

other c-ions assigned for the oxidized transferrin). In total, for the reduced form we report 34 c-type ions and 39 z-type ions (Table 1). Similarly to the previous experiments on IgG fragmentation, no complementary fragments for the same backbone cleavage have been identified.

Conclusions

Our previous results reported for the analysis of immunoglobulin G (IgG) had already shown light on the new possibilities open by high-resolution TOF mass spectrometry for the top-down analysis of large proteins. Here, we focused our attention on electron transfer dissociation-based TD MS characterization, for stressing the current possibilities offered by this experimental set-up and further investigating the dependence of protein sequence coverage on the protein structure/conformation and its disulfide bond connectivity. The comparison between oxidized and reduced transferrin clarifies, despite the use of a different precursor charge state for the ETD reaction, that the presence or absence of S–S bonds redistribute entirely the fragmentation channels (see Scheme 1, where the backbone cleavages in common between oxidized and reduced precursors, signed in green, are a minority). Not only the disulfide-protected regions need two bond cleavages to be sequenced, but also the different charge localization and local conformation

can play an important role in this scenario. Moreover, the comparison between the fragmentation maps of transferrin and IgG represents incontrovertible evidence that ETD is primarily directed to the disulfide-free loops interconnecting consecutive disulfide-protected domains when the proteins are analyzed in non-reduced forms [27]. On the other hand, for these two protein forms the product ion populations are distributed very differently in their ETD mass spectra (a single-maximum distribution for transferrin versus a double-maximum distribution for the IgG), indicating that also the primary structure of the protein influences the final outcome (see Fig. 1 and Schemes 1 and 2 in Ref. [27]). Serotransferrin presents a much more complicated S–S bond connectivity in comparison to the IgG, and this might explain the limited sequence coverage for the C-terminus of its oxidized form, as well as the smaller size of assigned fragments when compared to the IgG. The analysis of the reduced transferrin, indeed, complemented the results obtained for the oxidized form very well: although the extent of fragmentation obtained in each single experiment (i.e., for the oxidized or reduced protein) is the same, the combination of the two increases the final sequence coverage by ~70 % (from 10.7 to 18.1 %, Table 1).

Achieved ETD TD MS performance shall extend the application range of protein–drug binding characterization to larger proteins, compared to the current level of ubiquitin (8.6 kDa)-size proteins [31, 32]. However, further enhancement in (1)

sensitivity of detection of multiply charged product ions (to reach high signal-to-noise by averaging a lower number of scans); (2) dissociation of charge-reduced species, which are constantly dominant over the product ions in all the reported mass spectra; (3) detection of complementary heavy product ions as well as (4) internal product ions from secondary fragmentation will be needed to increase the overall capabilities of ETD-based TD MS and TDP on the qTOF and alternative, e.g., Orbitrap FTMS, platforms.

References

1. Cui W, Rohrs HW, Gross ML (2011) Analyst 136:3854–3864
2. Tran JC et al (2011) Nature 480:254–258
3. Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Kelleher NL (2011) J Biol Chem 286:25451–25458
4. Smith LM, Kelleher NL, Consortium for Top Down Proteomics (2013) Nat Methods 10:186–187
5. Tran JC, Doucette AA (2008) Anal Chem 80:1568–1573
6. Zheng S, Yoo C, Delmotte N, Miller FR, Huber CG, Lubman DM (2006) Anal Chem 78:5198–5204
7. Zubarev RA, Kelleher NL, McLafferty FW (1998) J Am Chem Soc 120:3265–3266
8. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW (2000) Anal Chem 72:563–573
9. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) P Natl Acad Sci USA 101:9528–9533
10. Wells JM, McLuckey SA (2005) Methods Enzymol 402:148–185
11. Little DP, Speir JP, Senko MW, O'Connor PB, McLafferty FW (1994) Anal Chem 66:2809–2815
12. Zubarev RA, Kruger NA, Fridriksson EK, Lewis MA, Horn DM, Carpenter BK, McLafferty FW (1999) J Am Chem Soc 121:2857–2862
13. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS (2013) J Am Chem Soc doi:10.1021/ja4029654
14. Zubarev RA (2003) Mass Spectrom Rev 22:57–77
15. Silivra OA, Kjeldsen F, Ivonin IA, Zubarev RA (2005) J Am Soc Mass Spectrom 16:22–27
16. Cooper HJ, Hakansson K, Marshall AG (2005) Mass Spectrom Rev 24:201–222
17. Sze SK, Ge Y, Oh H, McLafferty FW (2002) Proc Natl Acad Sci USA 99:1774–1779
18. Tsybin YO, Ramstrom M, Witt M, Baykut G, Hakansson P (2004) J Mass Spectrom 39:719–729
19. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP, McLafferty FW (2002) J Am Chem Soc 124:672–678
20. Baba T, Hashimoto Y, Hasegawa H, Hirabayashi A, Waki I (2004) Anal Chem 76:4263–4266
21. Satake H, Hasegawa H, Hirabayashi A, Hashimoto Y, Baba T, Masuda K (2007) Anal Chem 79:8755–8761
22. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Proc Natl Acad Sci USA 101:9528–9533
23. Huang TY, McLuckey SA (2010) Proteomics 10:3577–3588
24. Hartmer RG, Kaplan DA, Stoermer C, Lubeck M, Park MA (2009) Rapid Commun Mass Spectrom 23:2273–2282
25. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ (2007) Anal Chem 79:3525–3534
26. Kaplan DA, Hartmer R, Speir JP, Stoermer C, Gumerov D, Easterling ML, Brekenfeld A, Kim T, Laukien F, Park MA (2008) Rapid Commun Mass Spectrom 22:271–278
27. Tsybin YO, Fornelli L, Stoermer C, Luebeck M, Parra J, Nallet S, Wurm FM, Hartmer R (2011) Anal Chem 83:8919–8927
28. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO (2012) Mol Cell Proteomics 11:1758–1767
29. Ganisl B, Breuker K (2012) ChemistryOpen 1:260–268
30. Gunawardena HP, Gorenstein L, Erickson DE, Xia Y, McLuckey SA (2007) Int J Mass Spectrom 265:130–138
31. Hartinger CG, Tsybin YO, Fuchser J, Dyson PJ (2008) Inorg Chem 47:17–19
32. Meier SM, Tsybin YO, Dyson PJ, Keppler BK, Hartinger CG (2012) Anal Bioanal Chem 402:2655–2662

Paper IV

Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry

Yury O. Tsybin,*,[†] Luca Fornelli,[†] Carsten Stoermer,[‡] Markus Luebeck,[‡] Julien Parra,[†,○] Sophie Nallet,[§,▽] Florian M. Wurm,[§] and Ralf Hartmer[‡]

[†]Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[‡]Bruker Daltonics GmbH, Bremen, Germany

[§]Cellular Biotechnology Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

S *Supporting Information*

ABSTRACT: Improving qualitative and quantitative characterization of monoclonal antibodies is essential, because of their increasing popularity as therapeutic drug targets. Electron transfer dissociation (ETD)-based top-down mass spectrometry (MS) is the method of choice for in-depth characterization of post-translationally modified large peptides, small- and medium-sized proteins, and noncovalent protein complexes. Here, we describe the performance of ETD-based top-down mass spectrometry for structural analysis of intact 150 kDa monoclonal antibodies, immunoglobulins G (IgGs). Simultaneous mass analysis of intact IgGs as well as a complex mixture of ETD product ions at sufficiently high resolution and mass accuracy in a wide m/z range became possible because of recent advances in state-of-the-art time-of-flight (TOF) mass spectrometry. High-resolution ETD TOF MS performed on IgG1-*kappa* from murine myeloma cells and human anti-Rhesus D IgG1 resulted in extensive sequence coverage of both light and heavy chains of IgGs and revealed information on their variable domains. Results are superior and complementary to those previously generated by collision-induced dissociation. However, numerous disulfide bonds drastically reduce the efficiency of top-down ETD fragmentation within the protected sequence regions, leaving glycosylation uncharacterized. Further increases in the experiment sensitivity and improvement of ion activation before and after ETD reaction are needed to target S—S bond-protected sequence regions and post-translational modifications.
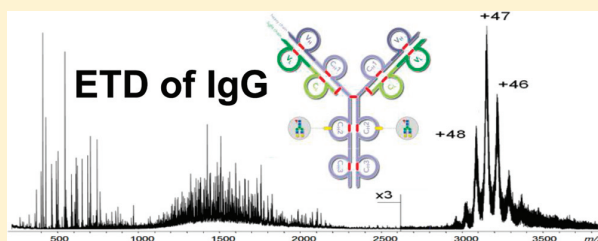
With the inception of new protein engineering technologies, monoclonal antibodies, once deemed "magic bullets," are again gaining popularity as an effective therapeutic tool against various diseases, including cancer. As a result, monoclonal antibodies in general and immunoglobulins in particular currently constitute an important number of therapeutic drugs (also called biologicals) under development in the biopharmaceutical industry.[1] Immunoglobulins play an essential role in the immune systems and can be found in the blood of vertebrates. Among the different classes of immunoglobulins, secreted immunoglobulins G (IgGs) are the most abundant.[2] A schematic representation of the Y-shaped structure of IgGs is shown in Scheme S1, which is given in the Supporting Information. All IgG antibodies have two identical light chains and two identical heavy chains. Each chain consists of constant and variable regions, the latter being responsible for the recognition and binding of antigens. The identification of the intact mass and sequencing of the constant and variable domains of the IgGs is important for the quality control of recombinant antibodies to be used as drugs.[3] Therefore, mass spectrometry (MS) and tandem mass spectrometry (MS/MS) play an important role in the qualitative and quantitative characterization of IgGs.[4] Specifically, mass spectrometric analysis is used to confirm the stability of IgGs in processing and storage, to identify and quantify their often highly heterogeneous post-translational modification (PTM) profiles, and to provide in-depth information on PTMs location and structure. MS/MS techniques are primarily used in the so-called bottom-up MS approach, in which IgGs are enzymatically cleaved in solution prior to ionization and MS/MS analysis. Similarly, specific PTM analysis may involve an enzymatic digestion that, in turn, may lead to the loss of PTMs (e.g., carbohydrates) prior to MS and MS/MS analysis.[5] Prolonged digestion at relatively high (37 °C) temperature, which is required for complete enzymatic action, can produce the artifacts and, particularly for IgGs, lead to an increased rate of deamidation and oxidation, preventing the correct identification of certain PTMs, such as Asp isomerization.[2]

Top-down mass spectrometry refers to mass spectrometric monitoring of the gas-phase fragmentation of biomolecular ions of any size.[6−8] A growing number of protein and RNA characterization experiments performed by top-down mass spectrometry in both direct infusion and coupled with online protein mixture purification and separation has been reported.[9−18]

Direct mass spectrometric analysis of intact IgGs has several advantages over the bottom-up approach, including quantitative differentiation of heterogeneously modified proteins and direct indication of N-terminal and C-terminal truncation.[19] Top-down MS/MS characterization of intact IgGs can potentially allow direct assignment of protein sequence fragments with multiple amino acid variations, even if the mutations occur within a distance of several amino acids. Furthermore, top-down mass spectrometry has found an important application in detecting the differences in protein splicing variants (isoforms), as highlighted in a recent work by Ge and co-workers on extensive troponin T protein characterization.[20] Finally, bottom-up analysis generally requires a higher amount of sample than the top-down approach.[2]

To date, the fragmentation of intact proteins of the size of intact monoclonal antibodies in the gas phase has been attempted with collision-induced dissociation (CID), either in the dedicated CID cell or directly in the ion source region.[21,22] However, CID-generated sequence coverage, particularly from IgGs, has been limited. Electron capture dissociation (ECD)[23−25] and electron transfer dissociation (ETD)[26] mass spectrometry of large proteins occupies a significant niche in top-down MS analysis, because of the specific advantages of these radical-driven MS/MS techniques, compared to slow-heating fragmentation methods, such as CID and infrared multiphoton dissociation (IRMPD). Advantages include more extensive sequence coverage obtained on large proteins and efficient rupture of disulfide bonds.[27,28] However, the significant advantage of ECD/ETD over CID/IRMPD in the characterization of labile PTMs on peptides[29] is less pronounced on proteins, where vibrational energy distribution over a large number of degrees of freedom increases the chance for labile PTMs to remain intact during the rupture of protein backbone bonds in CID/IRMPD.[30,31] As the current state-of-the-art in ECD-based top-down mass spectrometry "in time", Ge and co-workers identified 36 c-type ions and 23 z-type ions (59 cleavages) of 142 kDa cardiac myosin binding protein C (cMyBP-C).[32] Further increases in protein molecular weight for top-down mass spectrometry "in time" await the implementation of ECD on Fourier transform ion cyclotron resonance (FT-ICR) MS platforms equipped with superconducting magnets with strengths of 15−21 T.[33]

The reported here work became possible because of the newest addition to the family of tandem mass spectrometry "in space" instruments, particularly ETD-enabled state-of-the-art high-resolution quadrupole time-of-flight (qTOF) MS.[34] Specific features of a particular importance for the top-down mass spectrometry of intact proteins on the high resolution qTOF MS platform include the following: a resolution of up to 50 000 over a wide m/z range is achieved for intact protein analysis (see Figure S1 in the Supporting Information), a resolution of up to 30 000−40 000 over a wide m/z range is achieved for product ion analysis (resolution depends on the complexity of the product ion mixture and on the abundances of the product ions), a mass accuracy of ∼1 ppm exists in both MS and MS/MS modes, and a mass spectra acquisition rate of up to 20 per second is observed. The tandem mass spectra acquisition rate is dependent on the time required for the fragmentation reaction and will be discussed below. Particular advantages of top-down mass spectrometry of TOF MS platforms include larger dynamic range than the high-resolution Orbitrap and FT-ICR MS instruments and single-ion counting versus the ion package accumulation required for Fourier transform mass spectroscopy (FTMS) platforms.[35]

Therefore, we present the top-down characterization of two intact recombinant monoclonal antibodies, murine and human IgGs, performed with the above ETD-enabled state-of-the-art qTOF MS platform.

■ METHODS

Sample Preparation. Glycosylated monoclonal antibodies, specifically immunoglobulins G, were obtained commercially or produced in house. Proteins were dissolved in water to a final concentration of ∼1 μM. Purified IgG1-*kappa* from murine myeloma (clone number MOPC 21) was purchased from Sigma (Taufkirchen, Germany) and used without further purification.[36] In-house produced human anti-Rhesus D IgG1 was transiently expressed in HEK293 cells. Transfection of HEK293 cells in suspension for the expression of human anti-Rhesus D IgG was conducted as previously reported (see the Supporting Information).[37,38] Protein purification was performed following standard protocols to yield a nondetectable level of impurities (e.g., IgG fragments).[39]

Liquid Chromatography. Online protein purification and concentration for intact protein measurements was performed using an Agilent 1200 HPLC system with reversed phase cartridges (Zorbax SB C8 3.5 μm, 2.1 mm i.d. × 30 mm, Agilent, Waldbronn, Germany) at 70 °C and a flow rate of 300 μL/min. The applied gradient was 100% phase A (0.1% formic acid in water) for 3 min, followed by a linear gradient to 100% phase B (0.1% formic acid in acetonitrile) in 7 min. For ETD experiments, a capillary liquid chromatography (LC) setup (Dionex Ultimate 3000) with monolithic columns heated to 60 °C (PepSwift PSDVB, 200 μm i.d. × 50 mm, 5 mm trapping column; Dionex, Germering, Germany) was used. To obtain protein elution time from the LC column of ∼5−10 min, a long and shallow gradient (3%-80% phase B in 30 min at 2 μL/min) in combination with intentionally overloading the column was applied (see Figure S2, top inset, in the Supporting Information). Protein injection volume was 1 μL per each injection to provide ∼1 pmol of protein per injection.

Time-of-Flight Tandem Mass Spectrometry. Intact protein mass measurements and ETD MS/MS were performed on a high-resolution and high-mass-accuracy quadrupole time-of-flight (qTOF) mass spectrometer (maXis UHR qTOF MS, Bruker Daltonics GmbH, Bremen, Germany). ETD MS was performed as previously described.[34] The specific difference in ETD applied to high mass proteins is in the increased precursor ion accumulation time (up to 800 ms) prior to radical anion (azuline) injection for electron transfer reaction. Typical anion injection time was 40 ms, and additional ion−ion interaction time was normally set to 60 ms. Product ions were transferred to the ion cooler cell from the reaction cell and orthogonally pushed to the TOF mass analyzer at the rate of 10 kHz. The orthogonal acceleration potential was at 10 kV. Acquisition in a broad m/z window (100−5000 m/z) was performed with signal digitization at 2 GHz, whereas higher-resolution MS and MS/MS data were acquired at a faster digitization rate of 4 GHz for narrower m/z range of 500−3000 m/z. Intact protein MS data and product ion MS/MS data were acquired in separate LC runs to allow sufficient time for data acquisition during protein elution from the chromatographic column.

Data Analysis. The LC-MS and LC-MS/MS data were analyzed using the standard top-down data analysis procedure (Bruker Daltonics).[40] (See the Supporting Information for more details.)

Figure 1. Top-down electron transfer dissociation (ETD)-based LC electrospray ionization (ESI) quadrupole time-of-flight (qTOF) MS/MS of MOPC 21 IgG. Inset: expanded segment, $m/z = 250-950$, shows ETD product ions that are assigned to be mainly z-type product ions containing the C-terminal part of the heavy chain of MOPC 21 IgG. The inset shows a typical isotopic distribution of doubly charged product ion demonstrating high resolving power and true isotope relative abundances achieved. To increase the confidence in ETD product ion assignment, the demonstrated here tandem mass spectrum was obtained by summation of the data from 10 consecutive LC MS/MS runs.

Briefly, precursor and product ion mass spectra were summed over the elution time of a protein in each LC-MS and LC-MS/MS run with Data Analysis software version 4.0 (Bruker Daltonics). Intact protein MS data were directly submitted to the MaxEntropy deconvolution procedure to yield a heterogeneous distribution of neutral (deprotonated) glycosylated protein isoforms (see Figure S1 in the Supporting Information). Protein glycosylated forms were assigned using the built-in post-translational modification calculator. Internal product ions, formed by secondary fragmentation of the primary product ions, have not been considered in the current work and will be analyzed in a separate study. To increase the signal-to-noise level of the product ions, up to 10 LC-MS/MS runs were summed.

■ RESULTS

Intact Protein Molecular Weight and Distribution of Glycoforms. Electrospray ionization time-of-flight (TOF) mass spectra of both MOPC 21 and HEK293 cell derived monoclonal antibodies ("HEK IgG") show broad charge-state distributions from approximately +30 to +60 (see Figure S1, panels A and B, respectively, in the Supporting Information). Deconvolution of the charge states toward the neutral protein mass distribution allows the estimation of the intact protein molecular weight (MW) and the differentiation of a number of glycosylated protein isoforms (see Figure S1, panels C and D, respectively, in the Supporting Information). Knowledge of the intact protein mass indicates any potential modification of the protein,

including N-terminal and C-terminal truncations, which are common for monoclonal antibodies. Importantly, information on the terminal truncations may be lost when performing bottom-up mass spectrometry. Figure S1 in the Supporting Information suggests protein molecular weights (MWs), without modifications, of ∼147 kDa for MOPC 21 IgG and ∼149 kDa for HEK IgG (see the Supporting Information for more details). Finally, intact protein mass spectrometry results presented in Figure S1 in the Supporting Information allow direct quantification of the relative abundances of the different glycoforms, whereas relative quantification of protein glycoforms by bottom-up mass spectrometry is typically performed on one or several enzymatically derived peptides that contain these glycoforms.[39,41]

Top-Down Mass Spectrometry of Murine IgG. To increase the performance of protein fragmentation by ETD in a single LC-MS experiment we explored the possibility that the LC-MS system might provide a substantially long protein elution time. Therefore, we have not attempted to optimize the separation efficiency. Total ion chromatogram showing the elution profile of the MOPC 21 IgG indicates protein elution time of ∼5 min for the experimental setup employed (see Figure S2, top inset, in the Supporting Information).

Although top-down analysis of large peptides and small proteins typically relies on the isolation of a single charge state, we isolated three charge states simultaneously for the online LC-purified proteins of 150 kDa. Indeed, online purified and separated proteins of such high m/z values, as in electrospray ionization mass spectrometry (EI-MS) of IgGs, dominate the
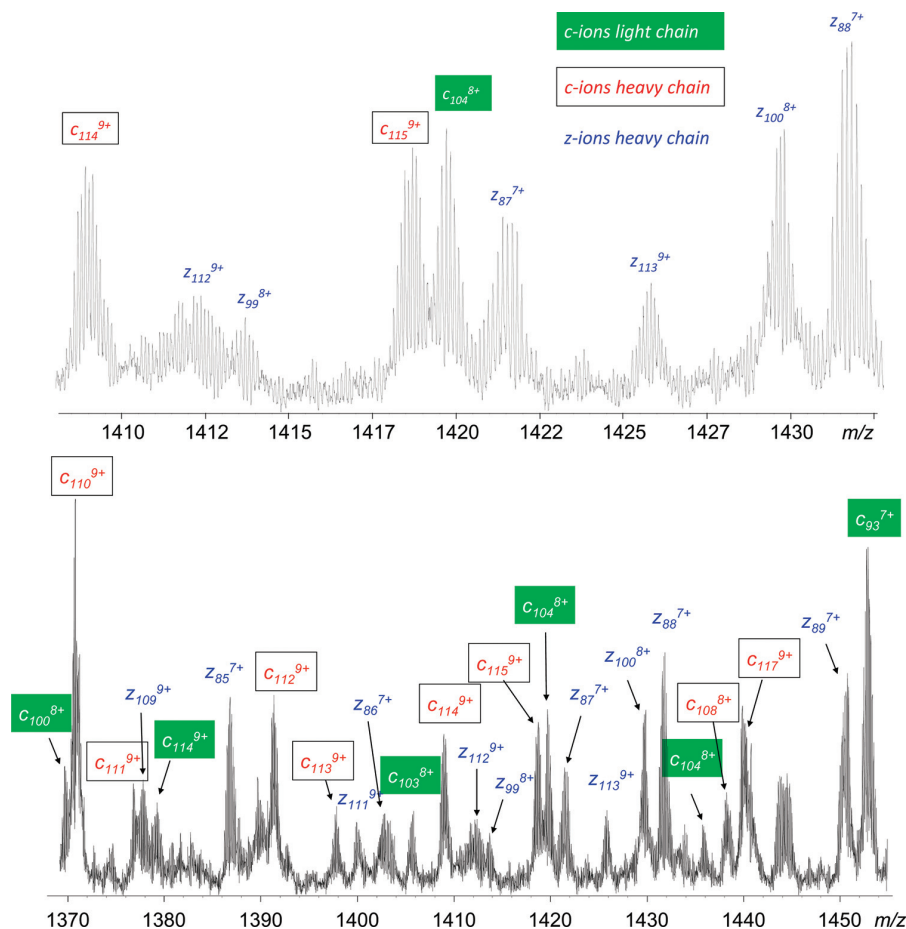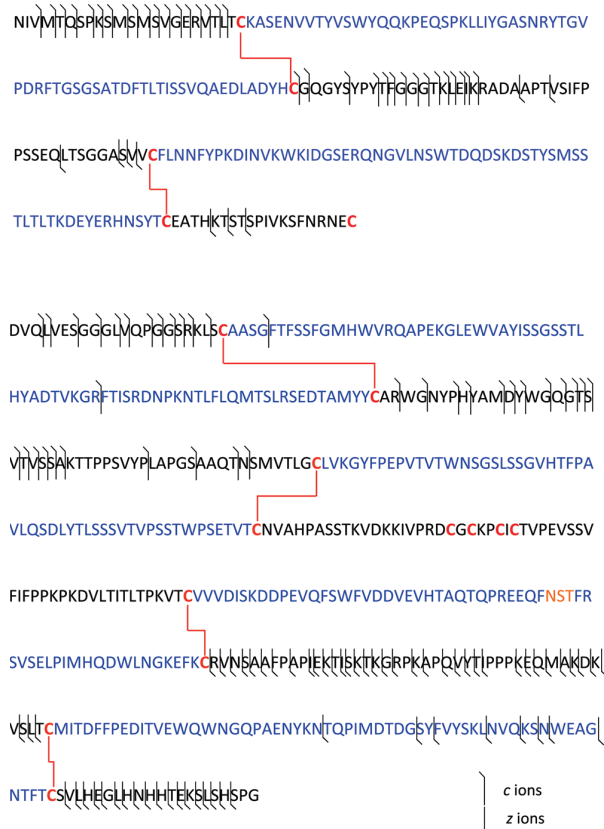
Figure 2. Medium-size product ions from a tandem mass spectrum of MOPC 21 IgG acquired with ETD-based LC ESI qTOF MS/MS, as shown in Figure 1. In the bottom image, the expanded segment, $m/z = 1370-1460$, shows multiply charged product ions from the center part of the heavy chain of MOPC 21 IgG, variable domain of the light chain, and other parts of the IgG. In the top image, the expanded segment, $m/z = 1408-1443$, shows the typically obtained resolved isotopic clusters of ETD product ions. To increase the confidence in ETD product ion assignment, the tandem mass spectrum demonstrated here was obtained by summation of the data from 10 consecutive LC MS/MS runs.

precursor ion distribution in the isolation window for the samples studied in the current work. Figure 1 and Figure S2 in the Supporting Information show the result of the top-down ETD qTOF MS approach applied to the online LC-purified MOPC 21 IgG (1 pmol on a column per injection). The isolation window was m/z 100 and covered precursor ion charge states +49, +50, and +51. Electron transfer from the reagent radical anions resulted in an efficient charge state neutralization of the precursor protein ions, as evidenced by the abundant protein peaks for charge states +48, +47, and +46. The extent of the reaction, shown by the intensity of the charge-reduced peaks and the value of the remaining charge, is a function of the injection time of radical anions and ion–ion reaction time. Sufficiently long injection and interaction times can reduce the precursor and charge-reduced ion amounts to undetectable levels. More efficient charge state neutralization and reduced species fragmentation can be achieved by increasing protein ion internal energy, similar to the supplemental activation.[42,43] For the experimental setup here, the supplemental activation can be provided by more energetic product ion transfer from the reaction cell to the ion cooler cell (by setting higher potential drop).

ETD MS/MS of MOPC 21 IgG shows abundant product ions grouped around m/z 500 and m/z 1500, Figure 1. Interestingly, there are no product ions in the large m/z area preceding precursor and charge-neutralized ion region. Closer analysis of product ions in the region $m/z \sim 500$ demonstrates primarily light ($\sim$1–2 kDa) product ions with few charges (Figure 1, top inset). The data demonstrate the sufficient capacity of the employed high-resolution qTOF MS to resolve the isotopic distributions and the overlapping product ions in this m/z range for the given MS/MS spectrum complexity. Expanded segments of the ETD mass spectrum at $m/z \approx 1500$ show the presence of heavier (up to 10–15 kDa) product ions that are more highly charged (see Figure 2). Despite a much higher complexity of the MS/MS spectrum, the employed mass spectrometer allows efficient separation of the isotopic and overlapping peaks. The results of ETD product ion identification are summarized in Scheme 1 for both light (Scheme 1, top) and heavy (Scheme 1, bottom) chains of murine IgG. In total, we report 69 c-ions (33 fragments for light chain and 36 for heavy chain) and 73 z-ions (9 fragments for light chain and 64 for heavy chain) identified from the ETD mass spectra

8922

dx.doi.org/10.1021/ac201293m |*Anal. Chem.* 2011, 83, 8919–8927

Scheme 1. MOPC 21 IgG Sequence Coverage Obtained by ETD-Based LC ESI qTOF MS/MS for (Top) Light Chain and (Bottom) Heavy Chain of the Protein[a]

NIVMTQSPKSMSMSVGERVTLTCKASENVVTYVSWYQQKPEQSPKLLIYGASNRYTGV

PDRFTGSGSATDFTLTISSVQAEDLADYHCGQGYSYPYTFGGGTKLEIKRADAAPTVSIFP

PSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNSWTDQDSKDSTYSMSS

TLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

DVQLVESGGGLVQPGGSRKLSCAASGFTFSSFGMHWVRQAPEKGLEWVAYISSGSSTL

HYADTVKGRFTISRDNPKNTLFLQMTSLRSEDTAMYYCARWGNYPHYAMDYWGQGTS

VTVSSAKTTPPSVYPLAPGSAAQTNSMVTLGCLVKGYFPEPVTVTWNSGSLSSGVHTFPA

VLQSDLYTLSSSVTVPSSTWPSETVTCNVAHPASSTKVDKKIVPRDCGCKPCICTVPEVSSV

FIFPPKPKDVLTITLTPKVTCVVVDISKDDPEVQFSWFVDDVEVHTAQTQPREEQFNSTFR

SVSELPIMHQDWLNGKEFKCRVNSAAFPAPIEKTISKTKGRPKAPQVYTIPPPKEQMAKDK

VSLTCMITDFFPEDITVEWQWNGQPAENYKNTQPIMDTDGSYFVYSKLNVQKSNWEAG

NTFTCSVLHEGLHNHHTEKSLSHSPG

c ions
z ions

[a] The anticipated intra-chain disulfide bonds are shown with the red lines that connect Cys residues. Glycosylation is not shown.

of murine MOPC 21 IgG (Table 1). All 142 product ions correspond to unique backbone cleavages out of 657 possible cleavage sites (see Scheme 1 for details). Therefore, the overall sequence coverage obtained is about 21%. The overall sequence coverage calculated for fragmentation of S—S unprotected regions of the sequence, shown in Scheme 1 in black, increases to ~48%. Sequence coverage of the constant domain C_H3 of the heavy chain (see Scheme S1 for details) is the highest and is ~85% (only S—S unprotected regions are taken into consideration).

Top-Down Mass Spectrometry of Human IgG. Application of ETD qTOF MS/MS to another monoclonal antibody, human anti-Rhesus D IgG, was similar to the MOPC 21 IgG efficiency of precursor ion neutralization and fragmentation. However, the observed product ion distribution indicates some differences (see Figure 3). Note that only the product ion region of the MS/MS data is shown in Figure 3, following data acquisition with a 4 GHz sampling rate. Interestingly, the distinction between the light and heavy product ion groups is less pronounced than for ETD of MOPC 21 IgG, whereas the region between the heavy product ion group and the charge-neutralized precursor ions is still empty (data not shown). ETD mass spectra assignment results show a similar to murine IgG sequence coverage formed by product ions belonging to both light and heavy chains

of human IgG (see Scheme 2). In total, we report 32 c-ions (21 fragments for light chain and 11 for heavy chain) and 77 z-ions (15 fragments for light chain and 62 for heavy chain) identified from ETD mass spectra of human anti-Rhesus D IgG (see Table 1). The 109 product ions correspond to 103 unique backbone cleavages out of 669 possible cleavage sites (see Scheme 2 for details). Therefore, the overall sequence coverage obtained is ~15%. Compared to ETD of murine IgG, fewer c-type product ions in ETD of human IgG were identified, most likely due to the structural differences between these two antibodies and lower purity of the sample employed. Sequence coverage distribution over the S—S protected and unprotected regions, as well as increased sequence coverage of the constant domain C_H3, are similar to the ETD of murine IgG.

■ DISCUSSION

Typically, ECD/ETD-based top-down mass spectrometry has been applied to 10—50 kDa proteins. The resulting ECD/ETD patterns demonstrate almost symmetrical distributions of product ions around the isolated precursor protein charge state, which is the $m/z \approx 1000$ region for proteins of that size.[28,44,45] In contrast, here, the ETD mass spectra of 150 kDa proteins in charge states below 50 exhibit a significant product-ion-free region between charge-reduced precursor ions (centered at $m/z \approx 3000$) and abundant product ions (centered at $m/z \approx 1000$). Interestingly, the product ions occupy similar m/z space as in the ECD/ETD of 10—50 kDa proteins, presumably demonstrating protein structure related preference toward the formation of product ions of a given size and charge state, as well as the product ion charge-state dependent efficiency of the secondary fragmentation and charge neutralization reactions in ETD. Furthermore, differences in protein sequence and high-order structure organization between human and murine IgGs (e.g., the number of disulfide bonds between the heavy chains) are also responsible for the substantially higher ETD efficiency for the fragmentation of the N-terminal regions of both heavy and light chains in the murine IgG (see Scheme 1, compared to Scheme 2), as manifested also by the more-abundant and numerous light product ions in Figure 1, compared to Figure 3.

For intact IgGs, the formation of several distinct product ion clusters can be explained by the specific structural organization of these covalently bound protein complexes (see Scheme S1 in the Supporting Information for details). IgGs employed in the current work are formed by two heavy chains of ~50 kDa each (443 amino acids for murine MOPC 21 IgG and 455 amino acids for human anti-Rhesus D IgG) and two light chains of ~25 kDa each (214 amino acids for both murine MOPC 21 IgG and human anti-Rhesus D IgG). These are covalently linked by interchain disulfide bonds between each heavy and light chain, as well as between the heavy chains. Observation of C-terminal product ions from light chains (see Schemes 1 and 2) indicates a possible rupture of interchain disulfide bonds upon electron absorption by precursor ions, which leads to the separation of heavy and light chains (Scheme S1 in the Supporting Information). As a result, fragmentation patterns are formed following the ETD fragmentation rules for 25 kDa and 50 kDa proteins, corresponding to isolated light and heavy chains. If the interchain S—S bonds remain intact, extensive fragmentation of N-terminal and C-terminal regions of heavy chains of IgG and N-terminal regions of light chains can occur. We note that the complementary product ions to either light or heavy product

Table 1. Sequence Information Obtained by Tandem Mass Spectrometry from Different Immunoglobulins G, IgGs[a]

	Sequence Information Containing N-terminus			Sequence Information Containing C-terminus		
	muIgG1[1]	huIgG1[2]	huIgG2[3]	muIgG1[1]	huIgG1[2]	huIgG2[3]
	# c-ions	# c-ions	# b-ions	# z-ions	# z-ions	# y-ions
light chain	33	21	24	9	15	32
heavy chain	36	11	26	64	62	20
total	69	32	50	73	77	52

[a] Results are presented for ETD MS/MS performed on (superscript 1) murine IgG1 (muIgG1) used in this work, in its non-reduced form and (superscript 2) human IgG1 (huIgG1) used in this work, in its non-reduced form; and for CID MS/MS performed on (superscript 3) human IgG2 (huIgG2), in its reduced and alkylated form, as reported by Bondarenko et al.[4]
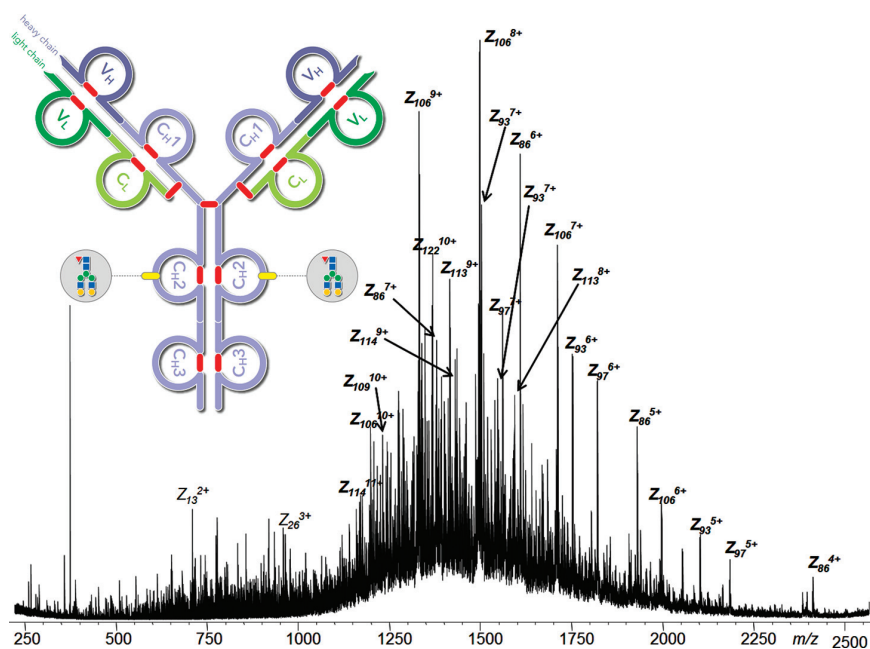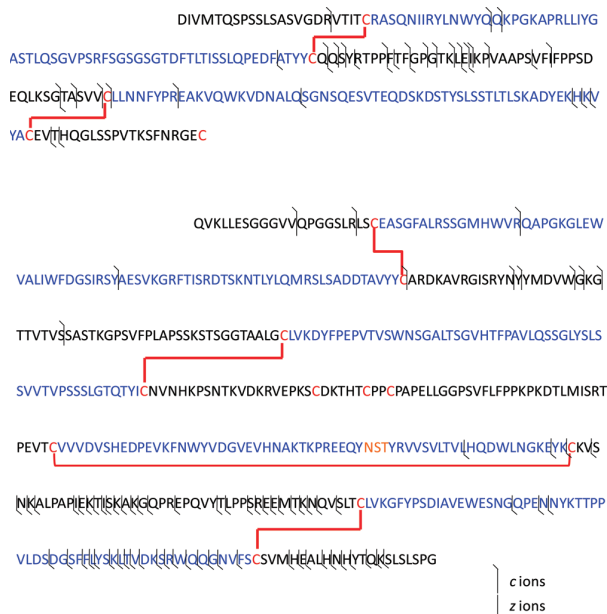

Figure 3. Expanded segment, m/z = 250–2500, of ETD-based LC ESI qTOF tandem mass spectrum of human anti-Rhesus D IgG transiently expressed in HEK293 cells. Only product ions due to the backbone cleavages in the IgG heavy chain are shown. Note the abundant product ions, e.g., z_{86}, z_{93}, z_{97}, and z_{106}, present in many charge states and presumably indicate a structure-related facile backbone rupture within the IgG heavy chain. To increase the confidence in ETD product ion assignment, the tandem mass spectrum demonstrated here was obtained by summation of the data from 10 consecutive LC MS/MS runs. The inset shows schematic representation of a generic IgG monoclonal antibody (see Scheme S1 in the Supporting Information for a detailed description).

ions would be positioned in the region around the charge-reduced precursor ions in the $m/z \approx 3000$ region. However, these complementary product ions are not shown with sufficient abundance and resolving power to be identified. Similarly, ion signals corresponding to the entire heavy or light chains following the possible interchain disulfide bond rupture have not been detected.

Fragmentation of monoclonal antibodies using CID of reduced and alkylated proteins previously yielded 26 b-ions and 20 y-ions (43 unique cleavages) for the isolated heavy chain and 24 b-ions and 32 y-ions (54 unique cleavages) for the isolated light chain for IgG2.[4] Collision-induced in-source fragmentation of intact IgG2 has produced efficient fragmentation of the variable domains of both heavy and light chains, leaving the constant domains almost intact.[3] Therefore, the top-down ETD data presented here for fragmentation of intact antibodies is superior

to the CID performed on the separated heavy and light chains and in-source CID performed on the entire IgG. A more-detailed comparison of the CID fragmentation patterns reported by Bondarenko and co-workers in Figure 6 of ref 4, with the ETD fragmentation patterns reported in Schemes 1 and 2 of the current work (Table 1), demonstrates a significant difference in the number of cleavages specifically in the C-terminal part of the protein (i.e., C_H2 and C_H3 domains) between the two activation/dissociation techniques, with ETD being more efficient than CID. The number of ETD-generated z-ions produced in C_H2 and C_H3 domains is twice that of the CID-generated y-ions, despite the fact that ETD experiments were performed on the nonreduced IgGs, whereas CID was done on the reduced and alkylated proteins and it is known that C_H2 and C_H3 domains are highly stable in the native folding. On the other hand, the sequence coverage obtained at the N-terminal parts of all three

Scheme 2. Transiently Expressed in HEK293 Cells Human Anti-Rhesus D IgG Sequence Coverage Obtained by ETD-Based LC ESI qTOF MS/MS for (Top) Light Chain and (Bottom) Heavy Chain of the Protein[a]

DIVMTQSPSSLSASVGDRVTITCRASQNIIRYLNWYQQKPGKAPRLLIYG

ASTLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSYRTPPFTFGPGTKLEIKPVAAPSVFIFPPSD

EQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKV

YACEVTHQGLSSPVTKSFNRGEC

QVKLLESGGGVVQPGGSLRLSCEASGFALRSSGMHWVRQAPGKGLEW

VALIWFDGSIRSYAESVKGRFTISRDTSKNTLYLQMRSLSADDTAVYYCARDKAVRGISRYNYYMDVWGKG

TTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLS

SVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRT

PEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVS

NKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPP

VLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG

c ions
z ions

[a] The anticipated intra-chain disulfide bonds are shown with the red lines that connect Cys residues. Glycosylation is not shown.

proteins studied is substantially less extensive. Neither ETD no CID sufficiently cover the biologically significant complementary determining regions (CDRs), especially those identified as CDR1 and CDR2, which are located in the variable domains of both light and heavy chains. Finally, a difference in the number of N-terminal fragments can be noted for the murine and the human IgGs used in this work, which reflects a variation in ETD efficiency, which is possibly due to either primary, secondary, and tertiary structures between IgGs of different species. The different states of the proteins used (reduced or not reduced), combined with their different sequence and structure, make definitive conclusions difficult; however, we can speculate that the sequence coverage achievable by ETD MS could be further improved by reducing and alkylating the IgGs, gaining an additional advantage over CID in terms of total sequence coverage. Nevertheless, due to the natural differences between ETD and CID, e.g., different method behavior toward cleaving at the N-terminal side of Pro, complementary information can be still obtained from the combined data.

The particular power and attraction of top-down mass spectrometry is that post-translational modifications can be characterized without the use of enzymatic interaction with the product.[6] Whereas intact protein mass spectrometry clearly demonstrated the heterogeneous glycosylation patterns of the employed IgGs (see Figure S1 in the Supporting Information), ETD-based top-down mass spectrometry is not yet capable of locating these modifications. The reasons for this failure in top-down mass spectrometry, as presented here, are the specific structural organization of IgGs, where the single glycosylation site is located in a middle of 50 kDa heavy chains and is protected by disulfide bonds. In addition to the overall MS/MS

performance increase, top-down mass spectrometry performance for post-translational modification analysis could be further improved through increases in the signal-to-noise ratio of product ions through summation of data from more than 10 LC-MS/MS runs and extensive use of ion activation before and after ETD reaction.[42,43] Further increase in sequence coverage can be envisioned from an increased rate of MS/MS data acquisition, which would translate to a higher number of spectra per LC MS/MS run.[46] Therefore, the use of CO_2 laser-based infrared multiphoton-induced ion activation should be considered.[47]

In addition to peptide and protein sequencing, ECD and ETD have been suggested to reflect the secondary and tertiary structures of peptides and proteins. It is believed that hydrogen bonding networks in the gaseous biomolecular ions modulate the site-specific efficiency of ECD/ETD peptide backbone cleavages.[48−50] Therefore, the structural organization of amino acids into α-helical, random coil, or beta strands may find specific signatures in the fragmentation pattern through distinct product ion abundance (PIA) distributions.[51] For example, ECD of amphipathic peptides produces sequence-specific periodic distribution of PIAs that may reflect a strong preference by amphipathic peptides for α-helical organization.[52] Similar observations of specific ECD PIA distributions have been reported for larger systems, including proteins consisting of a bundle of 3 α-helix segments.[53] Interestingly, Figures 1−3 also reveal very distinct preferential ETD fragmentation of intact IgGs.[46] For example, Figure 3 shows abundant z_{86}, z_{93}, z_{97}, z_{106}, and other product ions that are present in many charge states and clearly dominate the ETD PIA distribution. Presumably, these product ions indicate a structure-related facile backbone rupture within the IgG heavy chain. Indeed, these preferential cleavage sites are located between two highly structured, disulfide-protected domains of the heavy chain.[46]

■ CONCLUSIONS

Recent advances in high-resolution time-of-flight mass spectrometry (TOF-MS) made it possible to extend the upper mass limit of electron transfer dissociation (ETD)-based top-down mass spectrometry to 150 kDa intact monoclonal antibodies, specifically immunoglobulins G (IgGs). The number of identified product ions and the corresponding unique cleavages in ETD of IgGs reported here is double that of the electron capture dissociation (ECD) of a similarly sized (142 kDa) protein previously reported. Compatible with online purification and separation by high-performance liquid chromatography (HPLC), ETD-based top-down mass spectroscopy (MS) can now be efficiently applied to large intact proteins and protein complexes. The extent of protein sequence coverage is protein-structure-dependent and governed by disulfide bond connectivity and complex organization. The sequence coverage of IgGs achieved in this study is superior and complementary to the results of the previous attempts made with slow-heating fragmentation techniques such as collision-induced dissociation (CID) in a dedicated collisional cell or ion source region of a mass spectrometer.[3,4] Importantly, ETD-generated sequence tags provide valuable structural information on variable domains of both light and heavy chains. On the other hand, failure of the current implementation of ETD to locate and characterize glycosylation on the IgGs should be addressed by increasing the method sensitivity and adding efficient ion activation techniques. Despite the assignment of the most peaks in the ETD mass spectra, a number of the

unassigned isotopic clusters are still present. We estimate this number to be at least 20% of the total number of the isotopic clusters. Considering internal product ions and other peptide backbone cleavage products (e.g., *y*-ions and *a*-ions) should further enhance the sequence coverage and confidence in product ion assignment. The unambiguous assignment of all the unknown peaks is especially important for the characterization of antibodies in the biopharmaceutical industry and, thus, must be considered in the future data analysis work. Taking into account all product ions will allow the total efficiency of ETD to be estimated in this experimental setup. However, it is expected that the ETD efficiency of highly charged large protein ions shall be above 50% if all the neutral loss-produced product ions, as well as both primary and secondary (internal) product ions, are considered. Sufficiently long ion−ion interaction periods allow one to deplete the precursor and charge-neutralized ion populations completely, especially when assisted by supplemental activation (data not shown). The high complexity of the fragmentation pattern also demands further (at least 2−3 fold) resolution increase in a wide *m/z* range. Further improvement in the confidence and speed of complex top-down data analysis is also needed to account for the vast number of heterogeneously charged product ions formed and to increase the efficiency of pooling data from different charge states of precursor ions and experimental conditions (e.g., ion−ion interaction time).[54,55] Finally, further investigation of the abundant ions present in the product ion region is needed to confirm the assignment of these ions to the preferential ETD products or an eventual CID-based fragmentation. However, recent investigations suggest the ETD-based nature of these ions;[46] therefore, a study into the preferential fragmentation in ETD is needed to relate the observed preferential dissociations channels with the high-order structural features of these large proteins and to contribute to the general understanding of ECD/ETD mechanisms.

ASSOCIATED CONTENT

ⓢ Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*Address: Prof. Yury O. Tsybin, EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland. E-mail: yury.tsybin@epfl.ch.

Present Addresses
○Laboratory for Analysis and Modeling for Biology and Environment, CNRS UMR 8587, Université d'Evry Val d'Esonne, 91025 Evry Cedex, France.
▽Crucell Switzerland AG, 3018 Bern, Switzerland.

ACKNOWLEDGMENT

REFERENCES

(1) Sheridan, C. *Nat. Biotechnol.* **2010**, *28*, 307–310.
(2) Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom. Rev.* **2009**, *28* 147–176.
(3) Zhang, Z. Q.; Shah, B. *Anal. Chem.* **2007**, *79*, 5723–5729.
(4) Bondarenko, P. V.; Second, T. P.; Zabrouskov, V.; Makarov, A. A.; Zhang, Z. Q. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1415–1424.
(5) Hansen, R.; Dickson, A. J.; Goodacre, R.; Stephens, G. M.; Sellick, C. A. *Biotechnol. Bioeng.* **2010**, *107*, 902–908.
(6) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 196A–203A.
(7) Garcia, B. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 193–202.
(8) Reid, G. E.; McLuckey, S. A. *J. Mass Spectrom.* **2002**, *37*, 663–675.
(9) Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne, M. T.; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. *Anal. Chem.* **2007**, *79*, 7984–7991.
(10) Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L. *Anal. Chem.* **2010**, *82*, 1234–1244.
(11) Zhang, J. A.; Dong, X. T.; Hacker, T. A.; Ge, Y. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 940–948.
(12) Ayaz-Guner, S.; Zhang, J.; Li, L.; Walker, J. W.; Ge, Y. *Biochemistry* **2009**, *48*, 8161–8170.
(13) Zabrouskov, V.; Whitelegge, J. P. *J. Proteome Res.* **2007**, *6* 2205–2210.
(14) Xie, Y. M.; Zhang, J.; Yin, S.; Loo, J. A. *J. Am. Chem. Soc.* **2006**, *128*, 14432–14433.
(15) Ferguson, J. T.; Wenger, C. D.; Metcalf, W. W.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1743–1750.
(16) Hartinger, C. G.; Tsybin, Y. O.; Fuchser, J.; Dyson, P. J. *Inorg. Chem.* **2008**, *47*, 17–19.
(17) Taucher, M.; Breuker, K. *J. Am. Soc. Mass Spectrom.* **2010**, *21* 918–929.
(18) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24*, 201–222.
(19) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.
(20) Zhang, J.; Zhang, H.; Ayaz-Guner, S.; Chen, Y.-C.; Dong, X.; Xu, Q.; Ge, Y. *Biochemistry* **2011**, *50*, 6081–6092.
(21) Karabacak, N. M.; Li, L.; Tiwari, A.; Hayward, L. J.; Hong, P. Y.; Easterling, M. L.; Agar, J. N. *Mol. Cell. Proteomics* **2009**, *8*, 846–856.
(22) Ryan, C. M.; Souda, P.; Bassilian, S.; Ujwal, R.; Zhang, J.; Abramson, J.; Ping, P.; Durazo, A.; Bowie, J. U.; Hasan, S. S.; Baniulis, D.; Cramer, W. A.; Faull, K. F.; Whitelegge, J. P. *Mol. Cell. Proteomics 9*, 791-803.
(23) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
(24) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15*, 12–16.
(25) Zubarev, R. A. *Mass Spectrom. Rev.* **2003**, *22*, 57–77.
(26) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
(27) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 2857–2862.
(28) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563–573.
(29) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Anal. Chem.* **1999**, *71*, 4431–4436.
(30) Siuti, N.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 817–821.
(31) Mikhailov, V. A.; Iniesta, J.; Cooper, H. J. *Anal. Chem.* **2010**, *82* 7283–7292.
(32) Ge, Y.; Rybakova, I. N.; Xu, Q. G.; Moss, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12658–12663.
(33) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 3985–3990.
(34) Hartmer, R. G.; Kaplan, D. A.; Stoermer, C.; Lubeck, M.; Park, M. A. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 2273–2282.

(35) Kaplan, D. A.; Hartmer, R.; Speir, J. P.; Stoermer, C.; Gumerov, D.; Easterling, M. L.; Brekenfeld, A.; Kim, T.; Laukien, F.; Park, M. A. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 271–278.

(36) Strathearn, M. D.; Strathearn, G. E.; Akopiantz, P.; Liu, A. Y.; Paddock, G. V.; Salser, W. *Nucleic Acids Res.* **1978**, *5*, 3101–3112.

(37) Baldi, L.; Hacker, D. L.; Adam, M.; Wurm, F. M. *Biotechnol. Lett.* **2007**, *29*, 677–684.

(38) Baldi, L.; Jacquet, R.; Picasso, S.; Tromba, P.; Derow, E.; Girard, P.; Hacker, D.; Wurm, F. M. *Animal Cell Technology Meets Genomics*, Proceedings of the 18th ESACT Meeting, Grenada, Spain, May 11−14, 2003.

(39) Nallet, S.; Fornelli, L.; Schmitt, S.; Parra, J.; Baldi, L.; Wurm, F. M.; Tsybin, Y. O. Submitted to *New Biotechnology*.

(40) Hartmer, R.; Stoermer, C.; Decker, J. In *Proceedings of the 58th Conference on Mass Spectrometry and Allied Topics*, Salt Lake City, UT, May 23−27, 2010.

(41) Rebecchi, K. R.; Wenke, J. L.; Go, E. P.; Desaire, H. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1048–1059.

(42) Ben Hamidane, H.; Chiappe, D.; Hartmer, R.; Vorobyev, A.; Moniatte, M.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2009**, *20* 567–575.

(43) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 477–485.

(44) Tsybin, Y. O.; Ramstrom, M.; Witt, M.; Baykut, G.; Hakansson, P. *J. Mass Spectrom.* **2004**, *39*, 719–729.

(45) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774–1779.

(46) Tsybin, Y. O.; Damoc, E.; Fornelli, L.; Miladinovic, S.; Nolting, D.; Zeller, M.; Grouzmann, E.; Makarov, A. A. In *Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO, June 5−9, 2011.

(47) Ledvina, A. R.; Beauchene, N. A.; McAlister, G. C.; Syka, J. E. P.; Schwartz, J. C.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2010**, *82*, 10068–10074.

(48) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15863–15868.

(49) Breuker, K.; Oh, H. B.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 6407–6420.

(50) Breuker, K.; Oh, H. B.; Lin, C.; Carpenter, B. K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14011–14016.

(51) Vorobyev, A.; Hamidane, H. B.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2273–2283.

(52) Ben Hamidane, H.; He, H.; Tsybin, O. Y.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1182–1192.

(53) Breuker, K.; Bruschweiler, S.; Tollinger, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 873–877.

(54) Easterling, M. L.; Sander, P.; Roemer, T.; DeNoyer, L.; Berg, C. B.; Decker, J.; Witt, M. *Mol. Cell. Proteomics* **2006**, *5*, 1061.

(55) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. *Nucleic Acids Res.* **2007**, *35*, W701–W706.

Paper V

Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS*⑤

Luca Fornelli‡, Eugen Damoc§, Paul M. Thomas¶, Neil L. Kelleher¶, Konstantin Aizikov§, Eduard Denisov§, Alexander Makarov§, and Yury O. Tsybin‡‖

The primary structural information of proteins employed as biotherapeutics is essential if one wishes to understand their structure–function relationship, as well as in the rational design of new therapeutics and for quality control. Given both the large size (around 150 kDa) and the structural complexity of intact immunoglobulin G (IgG), which includes a variable number of disulfide bridges, its extensive fragmentation and subsequent sequence determination by means of tandem mass spectrometry (MS) are challenging. Here, we applied electron transfer dissociation (ETD), implemented on a hybrid Orbitrap Fourier transform mass spectrometer (FTMS), to analyze a commercial recombinant IgG in a liquid chromatography (LC)-tandem mass spectrometry (MS/MS) top-down experiment. The lack of sensitivity typically observed during the top-down MS of large proteins was addressed by averaging time-domain transients recorded in different LC-MS/MS experiments before performing Fourier transform signal processing. The results demonstrate that an improved signal-to-noise ratio, along with the higher resolution and mass accuracy provided by Orbitrap FTMS (relative to previous applications of top-down ETD-based proteomics on IgG), is essential for comprehensive analysis. Specifically, ETD on Orbitrap FTMS produced about 33% sequence coverage of an intact IgG, signifying an almost 2-fold increase in IgG sequence coverage relative to prior ETD-based analysis of intact monoclonal antibodies of a similar subclass. These results suggest the potential application of the developed methodology to other classes of large proteins and biomolecules. *Molecular & Cellular Proteomics 11: 10.1074/mcp.M112.019620, 1758–1767, 2012.*

Top-down mass spectrometry (MS)[1] (1–3) has continued to demonstrate its particular advantages over traditionally employed bottom-up MS strategies (4). Specifically, top-down MS allows the characterization of specific protein isoforms originating from the alternative splicing of mRNA that code single nucleotide polymorphisms and/or post-translational modifications (PTMs) of protein species (5). Intact protein molecular weight (MW) determination and subsequent gas-phase fragmentation of selected multiply charged protein ions (referred to as tandem MS or MS/MS) theoretically might result in complete protein sequence coverage and precise assignment of the type and position of PTMs, amino acid substitutions, and C- or N-terminal truncations (6), whereas the bottom-up MS approach allows only the identification of a certain protein family when few or redundant peptides are found for a particular protein isoform. At a practical level, however, top-down MS-based proteomics struggles not only with the single- or multi-dimensional separation of undigested proteins, which demonstrates lower reproducibility and repeatability than for peptides, but also with technical limitations present in even state-of-the-art mass spectrometers. The outcome of a top-down MS experiment depends indeed on the balance between the applied resolution of the mass spectrometer and its sensitivity. The former is required for unambiguous assignment of ion isotopic clusters in both survey and MS/MS scans, whereas the latter is ultimately dependent on the scan speed of the mass analyzer, which determines the number of scans that can be accumulated for a given analyte ion on the liquid chromatography (LC) timescale to enhance the resulting signal-to-noise ratio (SNR). Until recently, the instrument of choice for top-down MS has been the Fourier transform ion cyclotron resonance (FT-ICR) mass

[1] The abbreviations used are: ETD, electron transfer dissociation; ECD, electron capture dissociation; CID, collision-induced dissociation; FT, Fourier transform; FTMS, Fourier transform mass spectrometry; IgG, immunoglobulin G; MS, mass spectrometry; MS/MS, tandem mass spectrometry; LC, liquid chromatography; PTM, post-translational modification; SNR, signal-to-noise ratio; FT-ICR MS, Fourier transform ion cyclotron resonance mass spectrometry; MW, molecular weight; qTOF MS, quadrupole time-of-flight mass spectrometry; LTQ, linear trap quadrupole (linear ion trap); pyroGlu, pyroglutamic acid.

spectrometer, primarily because of its superior resolving power and the availability of electron capture dissociation for the efficient MS/MS of large biomolecules (7, 8). However, this solution has been shown to have some limitations in the analysis of large proteins (9). The main issue, as described by Compton *et al.* (10), is that the SNR in Fourier transform mass spectrometry (FTMS) is inversely proportional to the width of the isotopic and charge state distributions (11), which both increase as a function of MW. Particularly, the SNR dramatically decreases with MW under standard on-line LC-MS/MS operating conditions if isotopic resolution is required. It is noteworthy that such SNR reduction can affect not only intact mass measurements, but also the subsequent MS/MS performance.

The most widely employed solution for improving top-down analysis is thus a substantial reduction of the protein mixture complexity, for example, through off-line sample prefractionation (12). Furthermore, when the MW exceeds 100 kDa, proteins are often analyzed *via* direct infusion after off-line purification of the single isoform or species of interest (13). Overall, these strategies aim to improve the quality of mass spectra, specifically their SNR, by increasing the number of scans dedicated to each selected isoform or species. However, off-line intact protein analysis has limitations, including sample degradation and modification (*e.g.*, oxidation during long off-line measurements and sample storage). The time required for multistep LC-based protein purification can also be substantial.

Electron capture dissociation (ECD) (14, 15) and electron transfer dissociation (ETD) (16) are ion activation techniques that allow polypeptide fragmentation with reduced PTM losses (17, 18). Nevertheless, ECD and ETD generally provide larger sequence coverage for intact proteins than slow-heating activation methods such as collision induced dissociation (CID) and infrared multiple photon dissociation (19, 20). Furthermore, ECD and ETD are known to cleave disulfide bonds, a fundamental feature for the analysis of proteins in their native state (*i.e.*, without cysteine reduction and alkylation) (21–23).

The structural analysis of high MW intact proteins with MS has garnered much recent attention in the literature (24, 25), mainly because of the improved capabilities offered by rapidly developing sample preparation, protein separation, and mass spectrometric methods and techniques. Immunoglobulin G (IgG) proteins are antibodies with an MW of about 150 kDa that are composed of two identical sets of light and glycosylated heavy chains with both intra- and intermolecular disulfide bridges (Fig. 1) (26). IgGs represent an attractive target for structural analysis method development, given their high importance as biotherapeutics (27). A unit-mass resolution mass spectrum demonstrating an isotopic distribution of an isolated charge state of a 148 kDa IgG1 has been recently achieved with FT-ICR MS equipped with 9.4 T superconducting magnet and a statically harmonized ICR cell (24). However, further
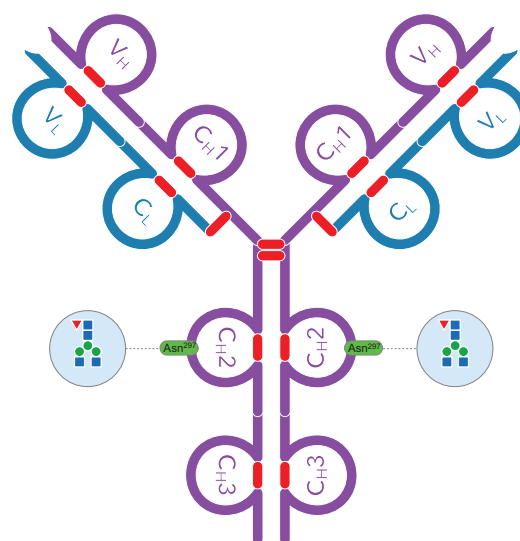

FIG. 1. **Schematic representation of IgG1.** Two identical light (*blue*) and two identical heavy (*fucsia*) chains form the intact IgG. The light chain is composed of a variable domain (V_L) and a constant domain (C_L), whereas the heavy chain comprises one variable domain (V_H) and three constant domains ($C_H 1$–3). Each domain contains an intramolecular disulfide bridge (in *red*); intermolecular disulfide bridges link the heavy chains to each other (two bonds) and each heavy chain to one light chain (one bond). Each heavy chain includes an N-glycosylation site (located at Asn^{297}; here, a G0F/G0F glycosylation is shown).

analytical improvements are needed to achieve routine and reproducible MS operation at the required level of resolution and sensitivity.

Fragmentation of intact antibodies in the gas phase following the top-down MS approach has been previously attempted without precursor ion charge state isolation by means of nozzle-skimmer CID on a linear trap quadrupole (LTQ)-Orbitrap™ (28, 29) and with precursor ion isolation *via* ETD on a high resolution quadrupole time-of-flight (qTOF) mass spectrometer (25). Relative to the results previously obtained with slow-heating MS/MS methods, the ETD qTOF MS/MS demonstrated substantially higher sequence coverage, reaching 15% for human and 21% for murine IgGs. Important for future top-down proteomics development for complex protein mixtures, the ETD qTOF MS/MS results were obtained on the LC timescale. To increase the sequence coverage and confidence in product ion assignment, a substantial increase in SNR was achieved by averaging MS/MS data from up to 10 identical LC-MS/MS experiments. The high complexity of the product ion population reduced the effective resolution to about 30,000, presumably limiting the assignment of overlapping high charge state product ions in the 1000–2000 *m/z* range. Even higher peak complexity was observed in the region of charge reduced species and complementary heavy product ions, above 3000 *m/z*. Finally, numerous disulfide bonds drastically reduced MS/MS efficiency in the disulfide bond-protected regions.

Here we demonstrate that ETD-enabled hybrid linear ion trap Orbitrap FTMS allows us to further improve the top-down ETD-based LC-MS/MS of monoclonal antibodies, introduced earlier for TOF-based MS. To fully take advantage of the high resolving power of Orbitrap MS/MS for increasing both the number of assigned product ions and the confidence of the assignments, maintaining an LC-MS/MS setup useful in a general proteomics workflow for protein desalting and separation, we averaged time-domain transients (derived from separated LC-MS/MS runs) before Fourier transform signal processing.

EXPERIMENTAL PROCEDURES

Sample Preparation—Humira monoclonal IgG1, *kappa*, was kindly provided by Abbott Laboratories (Abbott Park, IL, USA). The protein, originally stored in a buffer solution (sodium phosphate/sodium citrate/polysorbate 80, with the addition of NaCl), was dissolved in water at a concentration of 10 μM and used without further purification.

Liquid Chromatography—Protein desalting and preconcentration for subsequent on-line MS/MS analysis was performed with reverse phase high performance LC (Surveyor MS Pump Plus with Micro Autosampler, Thermo Fisher Scientific, San Jose, CA, USA). For each chromatographic run, 20 pmol (2 uL) of IgG was loaded onto a BioBasic-4 column of 1 mm i.d. and 100 mm length with a 5 μm particle size (Thermo Fisher Scientific, Runkorn, UK). A linear gradient from 20% to 80% of acetonitrile with 0.1% formic acid at a 100 μl/min flow rate was applied to ensure an IgG elution time of ~5 min. The total run time, including column wash and re-equilibration, was about 15 min.

MS and Tandem MS—An ETD-enabled hybrid linear ion trap Orbitrap FTMS (Orbitrap Velos Pro, Thermo Scientific, Bremen, Germany) with an IonMAX ion source was employed for both IgG intact mass measurement and fragmentation, in separate experiments. For ETD MS/MS, precursor ions were isolated at either 2750 $\pm$ 50 *m/z* (hereinafter referred to as the "narrow isolation window") or 2900 $\pm$ 300 *m/z* (hereinafter referred to as the "wide isolation window") in the high pressure LTQ. Fluoranthene radical anions were introduced into the LTQ over 40 to 50 ms and were allowed to interact with multiply charged cations of IgG for 10 or 25 ms. Product ions were transferred to the Orbitrap FTMS, which was operated with a charge target value (automatic gain control) set to 1 million. Precursor and product ion detection were performed in the Orbitrap mass analyzer over a 200–4000 *m/z* range. The resolving power was set at 15,000 (at 400 *m/z*) for MS experiments and at 100,000 (at 400 *m/z*) for ETD MS/MS. The precursor ion target was set to 1 $\times$ 10^6 for all experiments, and the target for fluoranthene anions was set to 2 $\times$ 10^6. To optimize the SNR improvement as a function of single scan number, 10 microscans were averaged for each scan in all the experiments. The S-lens rf level was set to the maximum (70), and the transfer tube temperature was set to 350 °C. Sheath gas was set to 20, and auxiliary to 10. The Orbitrap FTMS was calibrated for the high (2000–4000 *m/z*) mass range, keeping mass accuracy for the low (50–2000 *m/z*) mass range at the acceptable level.

Tandem MS Data Processing and Analysis—For each ETD LC-MS/MS experiment, Orbitrap FTMS time-domain transients were recorded in MIDAS .dat format (30) for additional signal processing. A dedicated routine was developed in-house for recording and adding the transients. Briefly, transient recording can be performed either directly by the Orbitrap FTMS on-board computer or through an additional oscilloscope connected to the preamplifier outputs, as described elsewhere (31). Transient processing was performed following standard methods developed for FT-ICR MS, taking advantage of the Orbitrap FTMS transient MIDAS format's being fully compatible with the available FT-ICR MS data analysis freeware (32). Importantly, all transient signals originating from separate LC-MS/MS experiments were summed in the time domain prior to Fourier transform (FT) signal processing. In the current work, the following numbers of summed transients were realized: (i) 18 scans (corresponding to 180 microscans) from a single LC-MS/MS run with an ETD duration of 10 or 25 ms; (ii) 180 scans (1800 microscans) from 10 LC-MS/MS runs with either 10 or 25 ms ETD; (iii) 360 scans (3600 microscans) from 20 LC-MS/MS runs, with equal numbers of 10 and 25 ms ETD experiments; and (iv) 1000 scans (10,000 microscans) from 40 LC-MS/MS runs, with 10 and 25 ms ETD combined.

For the SNR estimation, the noise in the ETD mass spectra was first calculated point by point (in a 45 *m/z* unit-large portion of the mass spectra, always the same for all of the ones considered), and after that the abundance of the peak at 772 *m/z* (taken as a reference for the light product ion population) was divided over the obtained noise value. Mass spectral deconvolution was performed with ProMass software (Thermo Scientific), and comprehensive analysis of fragmentation patterns was performed with the dedicated top-down MS analysis software ProSight PC 2.0 (Thermo Scientific) (33). In the ProSightPC analysis, searches were performed against a custom protein warehouse incorporating the known sequences of both light and heavy chains of Humira in both oxidized and reduced forms (both chains) and with both cyclized and uncyclized isoforms of the heavy chain. Data were searched in two ways: first at $\pm$15 ppm fragment tolerance, and finally at $\pm$4.2 Da fragment tolerance. Fragments matching the first search were considered valid without further inspection. Data acquired in the second search were all manually validated. The large search tolerance required for these analyses arises from two separate phenomena: first, peak picking algorithms can mismatch the isotopic distribution, and second, electron-based fragmentations are known to undergo hydrogen atom migration (34, 35). Both of these processes can shift the observed mass by 1 or more Da (1.0033 Da in the former, 1.0078 Da in the latter). Spurious matches were removed as false positives in this analysis.

RESULTS

Intact Protein MS—Fig. 2 shows the results of a single LC-MS experiment, without fragmentation, on the human, recombinant Humira IgG1. Over the selected gradient, the elution time of the IgG from the column was ~5 min (Fig. 2A). The gradient, in combination with the relatively high quantity of IgG injected, maximized the elution time and allowed for efficient desalting. The charge state distribution is centered on the 52+ protein cation, and charge states observed ranged from 38+ to 80+ (Fig. 2B). After the deconvolution of broadband MS data, several potential glycoforms were identified (Fig. 2C). The peak at 148,080.3 Da can be attributed to two G0F oligosaccharidic chains (often referred to as G0F/G0F glycoform (36)), whereas the remaining two glycoforms can be considered, respectively, as a G0F/G1F glycoform (considered the mass difference of 162 Da, corresponding to the addition of a hexose to one of the oligosaccharides) and G1F/G1F (or possibly, but less likely, G0F/G2F). Considering the known sequence of the IgG, which corresponds to an average mass of 145,465.5 Da, the mass accuracy error for the first glycan combination is ~0.3 Da, corresponding to ~2 ppm. The wide charge state envelope (from 38 to 80+) ensured good statistics for the deconvolution operation. The
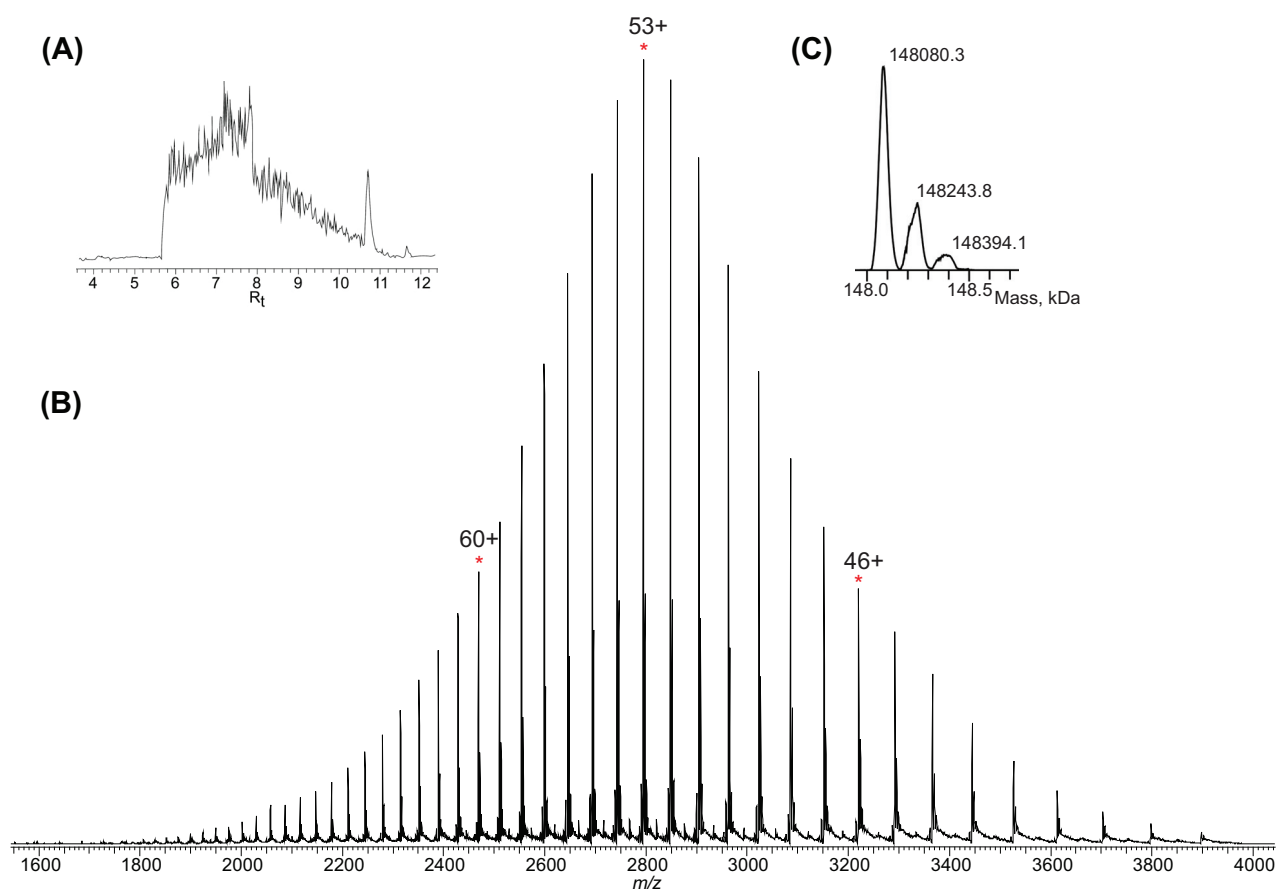
FIG. 2. **Results of intact IgG LC-MS analysis in the Orbitrap FTMS acquired at 15,000 resolution (at *m/z* 400).** *A*, base peak chromatogram displaying the ~5 min elution of the analyzed IgG1. *B*, charge state distribution of the Humira IgG, from 38+ to 80+. *C*, Deconvoluted mass spectrum showing distinct IgG glycoforms; the first peak represents a G0F/G0F glycoform, and the second and third peaks are G0F/G1F and G1F/G1F, respectively.

inferred glycosylation pattern for the Humira IgG1 produced by intact protein mass measurements is consistent with what reported in the literature for similarly produced antibodies (36, 37). Moreover, the intact mass measurements confirmed that no truncation was observed at either the N- or the C-terminus of the analyte.

Typically, qualitative and quantitative analysis of the glycosylation profile are performed in a bottom-up fashion—for instance, by quantifying tryptic glycopeptides (38). However, and more commonly, in order to reduce the high complexity of all possible combinations of different glycans that can be found on recombinant IgGs, glycans are released *via* enzymatic reaction with PNGase F (39), with or without combination with physical methods such as microwave-assisted digestion (40), for both qualitative and quantitative analysis.

Top-down MS Method Optimization—As previously mentioned, the main performance limitation in the LC-based top-down MS and MS/MS of large proteins is the relatively low SNR provided by averaging a limited number of scans acquired during rapid protein elution in a single LC-MS/MS

experiment. To increase the SNR of the MS/MS data and to maximize the number of confidently identified product ions, summing of Orbitrap MS/MS transients coming from a number of distinct LC-MS/MS experiments could be implemented before performing FT for time-to-frequency conversion. Fig. 3 displays expanded segments of ETD mass spectra resulting from analysis of a different numbers of transients acquired for two ion–ion interaction periods (10 ms and 25 ms) and two isolation windows (narrow and wide). Figs. 3*A*–3*E* show the results of experiments performed with a 2700–2800 *m/z* precursor ion isolation window in the LTQ, the narrow isolation window, comprising precursor ions from 53+ to 55+ charge states. The ETD mass spectrum shown in Fig. 3*F* was obtained for the 2600–3200 *m/z* precursor ion isolation window, the wide isolation window, comprising precursor ions from 47+ to 57+ charge states. For a 10 ms ion–ion interaction period (ETD duration), the main product ion population is centered at about 1600 *m/z*, and a second group of less abundant, lighter product ions is located below 1000 *m/z* (Fig. 3*A*). The experiment with a longer (25 ms) ETD duration shows
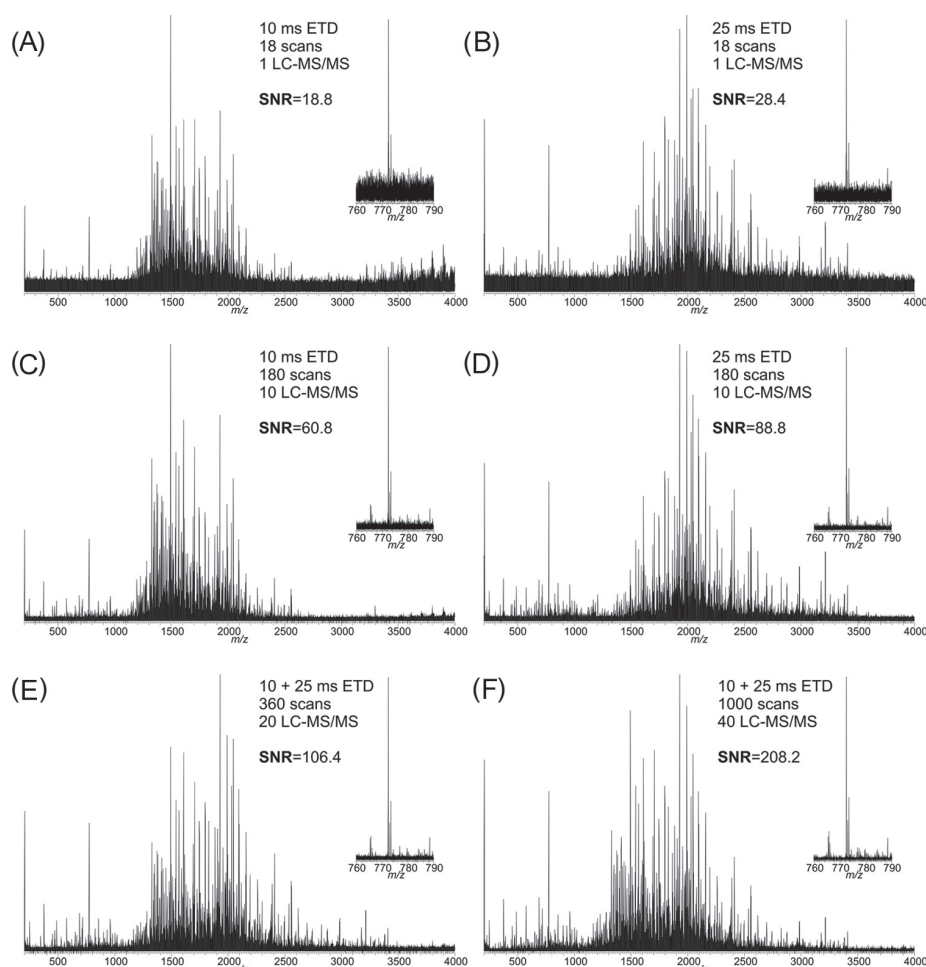
Fig. 3. **Comparison of ETD mass spectra resulting from the addition of different numbers of transients and acquired under different ETD conditions.** Precursor ion isolation window for panels *A–E* was 2750 ± 50 *m/z*, whereas for panel *F* a wider isolation window of 600 *m/z* units centered at 2900 *m/z* was employed. *A*, ETD mass spectra obtained by summing 18 scans (corresponding to 180 microscans) acquired in a typical single ETD LC-MS/MS experiment with 10 ms ETD duration. *B*, single LC-MS/MS with 25 ms ETD duration. *C*, ETD mass spectra obtained by summing 180 transients (corresponding to 1800 microscans) collected in 10 LC-MS/MS experiments with 10 ms ETD duration. *D*, 10 LC-MS/MS runs with 25 ms ETD duration. *E*, ETD mass spectrum resulting from the sum of 360 transients, coming equally from 10 ms and 25 ms ETD duration experiments, equivalent to 20 LC-MS/MS runs. *F*, ETD mass spectrum resulting from the sum of 1000 transients (10,000 microscans in total), coming equally from 10 ms and 25 ms ETD duration experiments.

a clear mass shift toward higher *m/z* values, with the final result being that the main product ion population is centered at ~2000 *m/z* (Fig. 3*B*). A longer ETD duration presumably causes secondary electron transfer and product ion fragmentation. A greater number of different product ions with low masses is present in Fig. 3*B* than in Fig. 3*A*. Under both experimental conditions, the precursor ions and the corresponding charge-reduced species were barely detectable. This contrasts with our previous ETD MS/MS analysis of a human IgG conducted on a qTOF MS (for a comparison, see Fig. 1 of Ref. (25)). Figs. 3*C* and 3*D* show ETD mass spectra obtained by summing 180 scans from 10 and 25 ms ETD LC-MS/MS experiments, respectively. Here, the lack of detected IgG charge-reduced species is even more apparent.

Finally, Figs. 3*E* and 3*F* display mass spectra combining scans from both 10 and 25 ms ETD MS/MS, with the former resulting from the sum of 360 scans and the latter 1000 scans coming from ETD experiments in which a wide isolation window was used. These two ETD mass spectra present a bimodal product ion distribution character. Note that in addition to the increase in SNR for mass spectra gained by summing experiments with ETD interaction times of 10 ms and 25 ms, it is believed that higher sequence coverage may be obtained by adding potentially different product ions produced at a different degree of secondary fragmentation.

To estimate the sensitivity (or SNR) gain achieved by summing time domain transients before FT, we calculated SNR values for each of the ETD mass spectra presented in Fig. 3

using the abundance of the 772 *m/z* ion peak as a reference (see insets in Fig. 3). The reference peak was selected for its presence in both the 10 ms and 25 ms ETD experiments and its location in a spectral region with low interference from other peaks. Although the abundance of the 772 *m/z* ion peak relative to that of the base peak changes between the 10 ms and 25 ms datasets, it is clear that the measured SNR is roughly tripled when the number of scans increases from 18 to 180, as expected (the SNR goes from 18.8 to 60.8 and from 28.4 to 88.8 in 10 ms and 25 ms experiments, respectively). For ETD mass spectra combining an equal number of scans from the two types of ETD experiments, the gain in SNR is more difficult to determine, but by considering an averaged value for the initial SNR at 18 scans (that is, $(18.8 + 28.4)/2 = 23.6$) we can assume an increment of more than 4 times for the ETD mass spectrum in Fig. 3*E* relative to the result of a single LC-MS/MS experiment. The same calculation returns an almost 9-fold increment for the 1000-scans averaged ETD mass spectrum, which results from the use of different ETD precursor ions due to a wider isolation window and thus is not fully comparable with the preceding examples. A further increase in SNR obtained by adding more LC-MS/MS runs is feasible, although the envisioned SNR gain might not be justified by the increased sample consumption and experiment duration. Therefore, other ways of improving SNR in a single LC-MS/MS experiment are to be explored, including (i) activated ion-ETD (35, 41), (ii) protein ion supercharging (42–45), (iii) an increased number of precursor ions obtained through multiple fills of the ETD reaction cell, (iv) improved product ion transmission to the Orbitrap mass analyzer from the ETD reaction cell, and (v) an increased number of scans acquired per LC-MS/MS experiment *via* super-resolution-method-based signal processing applied complementary to the FT signal processing (46).

The complexity of high SNR ETD mass spectra can be appreciated in Fig. 4, which emphasizes the importance of the employed high resolution acquisition (100,000 at 400 *m/z*) in the MS/MS mode. Particularly, the overlapping isotopic distributions of multiply charged product ion clusters are resolved, allowing one to determine their charge state with high confidence. Although the high resolution employed in this work profoundly contributed to the identification of IgG backbone cleavage sites (*vide infra*), it must be stressed that the increase in SNR achieved with the addition of transients taken from different LC-MS/MS experiments (see supplemental Fig. S1) is currently the most important factor for the successful analysis of populations of large, multiply charged product ions *via* FTMS.

ETD-based Top-down MS: Narrow Isolation Window—The mass spectra obtained by summing 360 and 1000 ETD MS/MS scans were subjected to data analysis with ProSight PC (see "Experimental" section for more details). The obtained fragmentation map is represented in Fig. 5. A summary of identified N- and C-terminal end-containing ions for both
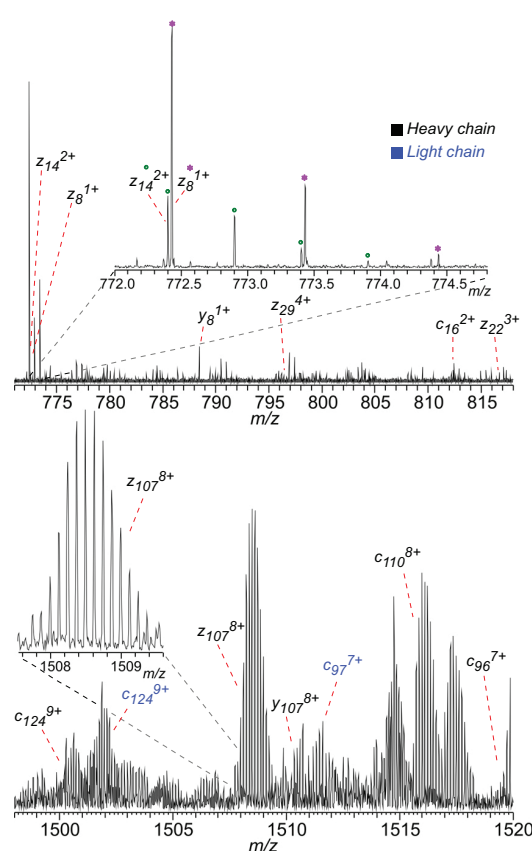

Fig. 4. **Expanded segments of ETD mass spectrum obtained by averaging 1000 scans (mass spectrum shown in Fig. 3*F*).** *Upper panel*, low *m/z* region; the importance of the high resolution in discriminating between z_{14}^{2+} and z_8^{1+} (heavy chain) is confirmed in the inset. *Bottom panel*, expanded segment of an ETD mass spectrum with highly charged and partially overlapping product ion clusters; the isotopic distribution of z_{107}^{8+} is shown in the inset.

light and heavy chains is presented in Table I. The ETD mass spectrum obtained from 360 scans with simultaneous fragmentation of three different IgG precursor ions (charge states 53+, 54+, and 55+) shows 46 *c*-ions, 31 $z^{\bullet}$-ions, and 23 *y*-ions for the light chain, corresponding to 72 unique backbone cleavages, and 48 *c*-ions, 35 $z^{\bullet}$-ions, and 21 *y*-ions, equivalent to 90 unique backbone cleavages, for the heavy chain (Fig. 5, *black bars*). In the heavy chain, the Glu[1] residue was observed to be also in pyroglutamic acid (pyroGlu) form, as highlighted in purple in Fig. 5. The total number of 162 identified unique backbone cleavages obtained *via* ETD of the three charge states of IgG precursor ions, taking into account the 214 residues of the light and 451 residues of the heavy chain (altogether resulting in 663 potential backbone cleavage sites), corresponds to 24.4% sequence coverage. Note the cleavages at both sides of proline residues (whose N–C$_\alpha$ bond cleavage in ETD would not result in separated product ions). Furthermore, although most of the assigned cleavage sites are located in the disulfide-bond-free regions, 22 backbone
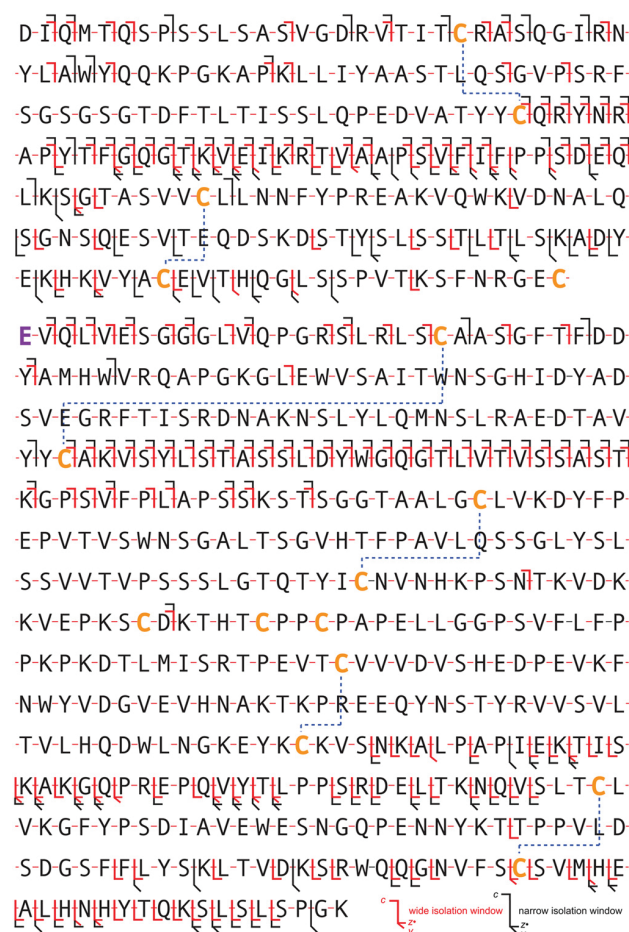
FIG. 5. **Fragmentation map of light (*top*) and heavy (*bottom*) chains of Humira IgG.** Cleavage sites identified for experiments conducted with isolation windows of 100 *m/z* (narrow window, centered at 2750 *m/z*) or 600 *m/z* (wide window, centered at 2900 *m/z*) are indicated in *black* and *red*, respectively. Cysteine residues are represented in *orange*, and intermolecular disulfide bonds are indicated by the *blue dashed lines*. On the heavy chain, the detected pyroGlu[1] mutation is highlighted in *purple*.

cleavages for the light chain and 10 for the heavy chain are notably located in disulfide-protected portions of the polypeptide sequence. Finally, it is noteworthy that the central portion of the light chain, comprising parts of both variable and constant domains, has been extensively sequenced, with most of the backbone cleavages confirmed by the identification of fragments containing information about the N terminus (*c*-ions) and the C terminus ($z^\bullet$- and/or *y*-ions). However, the central portion of the heavy chain (including the glycosylation site) was not sequenced at all.

ETD-based Top-down MS: Wide Isolation Window—The wider ion isolation window (600 *m/z* units, including precursor ions from 47+ to 57+ charge states), which might open new dissociation channels, was employed in order to further improve the ETD sequence coverage. Indeed, ETD of precursor

TABLE I
Summary of identified product ions from top-down MS of IgGs

	Light chain				Heavy chain				
	N-Term. product ions	C-Term. product ions	Number of unique cleavages	Chain sequence coverage	N-Term. product ions	C-Term. product ions	Number of unique cleavages	Chain sequence coverage	Total sequence coverage
Orbitrap MS ETD, 360 scans, narrow isolation window[a]	46	54	72	33.8%	48	56	90	20%	24.4%
Orbitrap MS ETD, 1000 scans, wide isolation window[a]	45	43	63	29.6%	57	71	117	26%	27.1%
Orbitrap MS ETD, combined[b]	54	67	92	43.2%	60	85	125	27.8%	**32.7%**
qTOF MS ETD[c]	21	15	31	14.4%	11	62	73	16.1%	15.4%
Orbitrap MS CID[d]	24	32	53	24.8%	26	20	42	9.5%	14.5%

[a] Measurements performed in the present study on intact human IgG1, *kappa*, with *c*-ions considered for the N-terminus and $z^\bullet$- and *y*-ions considered for the C-terminus.
[b] Values calculated by combining the backbone cleavage sites of the two Orbitrap FTMS experiments.
[c] Measurements performed in a previous work (Ref. 25) on an intact IgG1, *kappa*, with *c*-ions considered for the N-terminus and $z^\bullet$-ions considered for the C-terminus.
[d] Data derived from Ref. (29); measurements performed on reduced and alkylated human IgG2, with *b*-ions considered for the N-terminus and *y*-ions considered for the C-terminus.

ions from the wide isolation window generated 63 and 117 unique backbone cleavages from the light and heavy chains, respectively (Fig. 5, *red bars*). The total number of thus-obtained unique cleavage sites (180) corresponds to 27.1% sequence coverage. Relative to the ETD results obtained with the narrow isolation window, the C-terminal portion of the light chain presents a lower number of assigned backbone cleavages, whereas the sequence coverage of the region between the second and third constant domains (C_H2 and C_H3) and the C-terminal part of the heavy chain is substantially improved. The outlined difference in fragmentation patterns is apparent in the branching ratios of product ions formed. Indeed, the total number of product ions containing information on the C terminus, namely, $z^{\bullet}$- and *y*-type ions, was 56 in ETD with the narrow isolation window, whereas it reached 71 for ETD with the wide isolation window for the heavy chain (Table I). Conversely, for the light chain, the same number decreased from 54 for the former to 43 for the latter ion selection conditions. Considering the good SNR characterizing both analyzed mass spectra, the observed differences in sequence coverage at the C terminus are most likely due to the use of different ETD precursor ions for the two experimental sequences.

To adequately evaluate the progress made by employing Orbitrap ETD MS/MS for the structural analysis of intact and S–S bond reduced IgG, a comparison with previous work employing ETD MS/MS on a qTOF MS (see Scheme 2 in Ref. (25)) and CID MS/MS on an Orbitrap FTMS (29) can be done (Table I). Note that CID experiments were performed on reduced and alkylated IgG of a different subclass (IgG2, not IgG1). For ETD experiments, the analyzed human IgGs were different only in the variable domains, as they belonged to the same subclass and had the same *kappa*-type light chain. An appreciable increase in sequence coverage provided by ETD on Orbitrap FTMS is apparent (Table I). After we combined the results of narrow and wide isolation window ETD Orbitrap FTMS experiments, the final number of unique backbone cleavage sites rose to 217, representing 32.7% sequence coverage. A detailed investigation of the dependence of sequence coverage on the number of scans has yet to be addressed.

DISCUSSION

Given their multimeric structure comprising two heavy and two light chains, complex intra- and intermolecular disulfide connectivity, and heterogeneous glycosylation on heavy chains and their high molecular weight (~150 kDa), the full characterization of intact IgGs represents a difficult challenge for any MS platform. The state-of-the-art high resolution Orbitrap FTMS allows LC-timescale-compatible intact IgG ionization, charge state selection in a range up to 4000 *m/z*, and fragmentation by both slow-heating methods (*e.g.*, CID and higher energy collisional dissociation) and radical chemistry-based dissociation (*e.g.*, ETD). It also supports the detection

of isotopically resolved, multiply charged product ions present in a complex mixture, enabling a top-down MS approach for the structural analysis of intact IgGs. Here, we have demonstrated that time-domain transient signal averaging from a number of consecutive LC-MS/MS experiments before FT signal processing and peak picking significantly improves the performance of Orbitrap FTMS-based top-down MS. We note that although state-of-the-art top-down proteomics of 10 to 50 kDa proteins is already performed with on-line nanoLCMS, the analysis of intact proteins heavier than 50 kDa is often performed after off-line LC-based protein purification followed by off-line nanospray-MS. We opted for an on-line LC-MS/MS approach here because it is more universal than the off-line approach in terms of the types of samples to be addressed, as well as its suitability for automated quality control workflows. Indeed, the near-future goal of the top-down proteomics is to perform analyses of complex protein mixtures that contain heavy proteins as large as 150 to 200 kDa. With an increase in MS sensitivity and speed of high resolution data acquisition, substantially fewer LC-MS/MS runs will be required for protein characterization.

In the current work, the observed increase in SNR rose as expected with the square root of the number of scans; specifically, we estimated a roughly 4-fold to 8-fold gain in SNR from a single LC-MS/MS experiment to the final mass spectra. Furthermore, to maximize the number of assigned unique cleavage sites, we not only averaged 1000 scans, but we also used a large (600 *m/z*) precursor ion isolation window. As a result, the data presented here demonstrate a 2-fold higher sequence coverage than was obtained on a similar human IgG1 in a prior ETD implementation for top-down MS analysis of IgGs. In addition, we report the partial cyclization of Glu[1] to pyroGlu[1] on the IgG heavy chain. We note that significant structural differences between IgG variants considered in Table I certainly influence ETD MS/MS performance. Further work should be dedicated to comparing similar IgG subclasses. Also, information about the ETD preference for a particular IgG sequence or higher order structure provides an additional source of information on the ETD fundamentals.

The novelty of the results presented here for ETD-based top-down MS is primarily in the substantially improved analytical characteristics of the employed experimental set-up that increase the overall capabilities of the method and allow an important step to be taken toward its routine application. Although previous work demonstrated the proof-of-principle implementation, it also demonstrated the corresponding limitations, specifically in terms of the mass resolving power. The large number of laboratories equipped nowadays with high-resolution Orbitrap FTMS instrumentation suggests a high impact of the results described here. Nevertheless, similar to previously reported results, we were not able to sequence long portions of the constant domains of light chains or the entire central portion of the heavy chain, presumably because of a combination of disulfide-bond networks and the partial

retention of the secondary and tertiary structures of the IgG in the gas phase, which would explain the almost complete sequence coverage obtained for the flexible, unstructured loops connecting consecutive domains in the light and heavy chains (see Fig. 5). We are currently addressing these issues by implementing ion activation before and after ETD reaction, unfolding IgG *via* the partial reduction of disulfide bonds, and increasing the number of protonated sites *via* precursor ion supercharging. Future improvements in instrumentation—for example, the use of a compact high-field Orbitrap mass analyzer or Orbitrap Elite FTMS instrument or improved ion transfer and trapping—are expected to increase substantially the speed of analysis and reduce the number of scans required for top-down analysis.

REFERENCES

1. Cui, W., Rohrs, H. W., and Gross, M. L. (2011) Top-down mass spectrometry: recent developments, applications and perspectives. *Analyst* **136,** 3854–3864
2. Kellie, J. F., Tran, J. C., Lee, J. E., Ahlf, D. R., Thomas, H. M., Ntai, I., Catherman, A. D., Durbin, K. R., Zamdborg, L., Vellaichamy, A., Thomas, P. M., and Kelleher, N. L. (2010) The emerging process of top down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol. Biosyst.* **6,** 1532–1539
3. Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R., and Kelleher, N. L. (2011) Analysis of intact protein isoforms by mass spectrometry. *J. Biol. Chem.* **286,** 25451–25458
4. Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., Imhaus, A. F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X., and Dumenil, G. (2011) Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science* **331,** 778–782
5. Schluter, H., Apweiler, R., Holzhutter, H. G., and Jungblut, P. R. (2009) Finding one's way in proteomics: a protein species nomenclature. *Chem. Cent. J.* **3,** 11
6. Zhang, J., Zhang, H., Ayaz-Guner, S., Chen, Y. C., Dong, X., Xu, Q., and Ge, Y. (2011) Phosphorylation, but not alternative splicing or proteolytic degradation, is conserved in human and mouse cardiac troponin T. *Biochemistry* **50,** 6081–6092
7. Nikolaev, E. N., Boldin, I. A., Jertz, R., and Baykut, G. (2011) Initial experimental characterization of a new ultra-high resolution FTICR cell with dynamic harmonization. *J. Am. Soc. Mass Spectr.* **22,** 1125–1133
8. Tsybin, Y. O., Ramstrom, M., Witt, M., Baykut, G., and Hakansson, P. (2004) Peptide and protein characterization by high-rate electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **39,** 719–729
9. Patrie, S. M., Ferguson, J. T., Robinson, D. E., Whipple, D., Rother, M., Metcalf, W. W., and Kelleher, N. L. (2006) Top down mass spectrometry of < 60-kDa proteins from Methanosarcina acetivorans using quadrupole FTMS with automated octopole collisionally activated dissociation. *Mol. Cell. Proteomics* **5,** 14–25
10. Compton, P. D., Zamdborg, L., Thomas, P. M., and Kelleher, N. L. (2011) On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **83,** 6868–6874
11. Scigelova, M., Hornshaw, M., Giannakopulos, A., and Makarov, A. (2011) Fourier transform mass spectrometry. *Mol. Cell. Proteomics* **10,** O111.009431
12. Tran, J. C., and Doucette, A. A. (2009) Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **81,** 6201–6209
13. Ge, Y., Rybakova, I. N., Xu, Q., and Moss, R. L. (2009) Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 12658–12663
14. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120,** 3265–3266
15. Zubarev, R. A. (2003) Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom. Rev.* **22,** 57–77
16. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 9528–9533
17. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **71,** 4431–4436
18. Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., and Zubarev, R. A. (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spec.* **14,** 1793–1800
19. Molina, H., Matthiesen, R., Kandasamy, K., and Pandey, A. (2008) Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Anal. Chem.* **80,** 4825–4835
20. Santos, L. F. A., Eberlin, M. N., and Gozzo, F. C. (2011) IRMPD and ECD fragmentation of intermolecular cross-linked peptides. *J. Mass Spectrom.* **46,** 262–268
21. Zubarev, R. A., Kruger, N. A., Fridriksson, E. K., Lewis, M. A., Horn, D. M., Carpenter, B. K., and McLafferty, F. W. (1999) Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *J. Am. Chem. Soc.* **121,** 2857–2862
22. Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K., and McLafferty, F. W. (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72,** 563–573
23. Wu, S. L., Jiang, H. T., Hancock, W. S., and Karger, B. L. (2010) Identification of the unpaired cysteine status and complete mapping of the 17 disulfides of recombinant tissue plasminogen activator using LC-MS with electron transfer dissociation/collision induced dissociation. *Anal. Chem.* **82,** 5296–5303
24. Valeja, S. G., Kaiser, N. K., Xian, F., Hendrickson, C. L., Rouse, J. C., and Marshall, A. G. (2011) Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **83,** 8391–8395
25. Tsybin, Y. O., Fornelli, L., Stoermer, C., Luebeck, M., Parra, J., Nallet, S., Wurm, F. M., and Hartmer, R. (2011) Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal. Chem.* **83,** 8919–8927
26. Zhang, Z. Q., Pan, H., and Chen, X. Y. (2009) Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom. Rev.* **28,** 147–176
27. Sheridan, C. (2010) Fresh from the biologic pipeline-2009. *Nat. Biotechnol.* **28,** 307–310
28. Zhang, Z., and Shah, B. (2007) Characterization of variable regions of monoclonal antibodies by top-down mass spectrometry. *Anal. Chem.* **79,** 5723–5729
29. Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., and Zhang, Z. Q. (2009) Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **20,** 1415–1424
30. Senko, M. W., Canterbury, J. D., Guan, S., and Marshall, A. G. (1996) A high-performance modular data system for Fourier transform ion cyclo-

tron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **10,** 1839–1844

31. Perelman, D. H., Moscovets, E., Shwe, H., and Kyin, S. (2012) Analysis of transient data from an Orbitrap mass spectrometer with filter diagonalization method. *Proceedings of the 60th American Society for Mass Spectrometry (Santa Fe, NM) Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada,* May 20–24, 2012

32. Blakney, G. T., Hendrickson, C. L., and Marshall, A. G. (2011) Predator data station: a fast data acquisition system for advanced FT-ICR MS experiments. *Int. J. Mass Spectrom.* **306,** 246–252

33. Zamdborg, L., LeDuc, R. D., Glowacz, K. J., Kim, Y. B., Viswanathan, V., Spaulding, I. T., Early, B. P., Bluhm, E. J., Babai, S., and Kelleher, N. L. (2007) ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **35,** W701–W706

34. Savitski, M. M., Kjeldsen, F., Nielsen, M. L., and Zubarev, R. A. (2007) Hydrogen rearrangement to and from radical z fragments in electron capture dissociation of peptides. *J. Am. Soc. Mass Spectrom.* **18,** 113–120

35. Ledvina, A. R., McAlister, G. C., Gardner, M. W., Smith, S. I., Madsen, J. A., Schwartz, J. C., Stafford, G. C., Jr., Syka, J. E., Brodbelt, J. S., and Coon, J. J. (2009) Infrared photoactivation reduces peptide folding and hydrogen-atom migration following ETD tandem mass spectrometry. *Angew Chem. Int. Ed. Engl.* **48,** 8526–8528

36. Jefferis, R. (2005) Glycosylation of recombinant antibody therapeutics. *Biotechnol. Progr.* **21,** 11–16

37. Kim, W. D., Tokunaga, M., Ozaki, H., Ishibashi, T., Honda, K., Kajiura, H., Fujiyama, K., Asano, R., Kumagai, I., Omasa, T., and Ohtake, H. (2010) Glycosylation pattern of humanized IgG-like bispecific antibody produced by recombinant CHO cells. *Appl. Microbiol. Biot.* **85,** 535–542

38. Nallet, S., Fornelli, L., Schmitt, S., Parra, J., Baldi, L., Tsybin, Y. O., and Wurm, F. M. (2012) Glycan variability on a recombinant IgG antibody transiently produced in HEK-293E cells. *N. Biotechnol.* **29,** 471–476

39. Hansen, R., Dickson, A. J., Goodacre, R., Stephens, G. M., and Sellick, C. A. (2010) Rapid characterization of N-linked glycans from secreted and gel-purified monoclonal antibodies using MALDI-ToF mass spectrometry. *Biotechnol. Bioeng.* **107,** 902–908

40. Tzeng, Y. K., Chang, C. C., Huang, C. N., Wu, C. C., Han, C. C., and Chang, H. C. (2008) Facile MALDI-MS analysis of neutral glycans in NaOH-doped matrixes: microwave-assisted deglycosylation and one-step purification with diamond nanoparticles. *Anal. Chem.* **80,** 6809–6814

41. Ledvina, A. R., Beauchene, N. A., McAlister, G. C., Syka, J. E., Schwartz, J. C., Griep-Raming, J., Westphall, M. S., and Coon, J. J. (2010) Activated-ion electron transfer dissociation improves the ability of electron transfer dissociation to identify peptides in a complex mixture. *Anal. Chem.* **82,** 10068–10074

42. Sterling, H. J., and Williams, E. R. (2009) Origin of supercharging in electrospray ionization of noncovalent complexes from aqueous solution. *J. Am. Soc. Mass Spectrom.* **20,** 1933–1943

43. Valeja, S. G., Tipton, J. D., Emmett, M. R., and Marshall, A. G. (2010) New reagents for enhanced liquid chromatographic separation and charging of intact protein ions for electrospray ionization mass spectrometry. *Anal. Chem.* **82,** 7515–7519

44. Yin, S., and Loo, J. A. (2011) Top-down mass spectrometry of supercharged native protein-ligand complexes. *Int. J. Mass Spectrom.* **300,** 118–122

45. Miladinovic, S. M., Fornelli, L., Piech, K. M., Lu, Y., Girault, H. H., and Tsybin, Y. O. (2012) In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Anal. Chem.* **84,** 4647–4651

46. Kozhinov, A. N., and Tsybin, Y. O. (2012) Filter diagonalization method-based mass spectrometry for molecular and macromolecular structure analysis. *Anal. Chem.* **84,** 2850–2856

Paper VI

Top-down analysis of immunoglobulins G with electron transfer dissociation high-field Orbitrap FTMS

Luca Fornelli[1], Daniel Ayoub[1], Konstantin Aizikov[2], Xiaowen Liu[3,4], Eugen Damoc[2], Pavel A. Pevzner[5], Alexander Makarov[2], Alain Beck[6], and Yury O. Tsybin[1*]

[1] Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[2] Thermo Fisher Scientific GmbH, 28199 Bremen, Germany

[3] Department of BioHealth Informatics, Indiana University-Purdue University Indianapolis, 46202 Indianapolis, IN, USA

[4] Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 46202 Indianapolis, IN, USA

[5] Department of Computer Science and Engineering, University of California in San Diego, 92093 San Diego, CA, USA

[6] Centre d'Immunologie Pierre Fabre, 74160 St Julien-en-Genevois, France

*Correspondence should be addressed to Prof. Yury O. Tsybin, EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland. E-mail: yury.tsybin@epfl.ch

Running title: Top-down ETD Orbitrap Elite FTMS of IgG

To be submitted to: *Proteomics*

List of abbreviation:

Automatic gain control, AGC

Electron transfer dissociation, ETD

Electron transfer dissociation – higher energy collisional dissociation, EThcD

Fourier transform mass spectrometry/spectrometer, FTMS

Immunoglobulin G, IgG

Mass spectrometry, MS

Tandem mass spectrometry, MS/MS

Keywords: IgG, ETD, Orbitrap, EThcD

(Abstract)

The increasing importance of immunoglobulins G (IgGs) as biotherapeutics calls for improved analytical methods of in-depth structural characterization of these large (~150 kDa) macromolecules. Required methods have to be rapid, robust, and with minimal sample preparation. In a previous work we showed the potential of Orbitrap Fourier transform mass spectrometry (FTMS) combined with electron transfer dissociation (ETD) for the top-down investigation of an intact IgG1, resulting in ~30% sequence coverage. Here, we describe top-down analysis of IgGs1 (Adalimumab, Trastuzumab) and IgG2 (Panitumumab) performed with ETD on the higher field Orbitrap mass analyzer. For two different IgGs1, sequence coverage comparable to the previous results was achieved in a two-fold reduced number of averaged transients, which corresponds, taken together with the increased acquisition rate of the high-field Orbitrap, to ~six-fold improvement in analysis time. Furthermore, we compared the ETD process performed on the same IgG with different ion-ion interaction times and the top-down fragmentation patterns obtained for IgGs1 with that of an IgG2 (Panitumumab), which presents structural differences. These results reinforce the hypothesis that portions of the IgG chains remaining unsequenced are those retaining structural constraints in the gas phase.

1. Introduction

Monoclonal antibodies (mAbs) are an effective therapeutic tool in a variety of indications such as cancer and inflammatory disease.[1] These large glycoproteins constitute, with presently more than 40 approved products and 30 others in advanced clinical investigations, the fastest growing class of therapeutics.[2] The main class of antibodies employed as biotherapeutics is immunoglobulin G (IgG). IgGs are large (~150 kDa) biomolecules with a quaternary structure made of two ~25 kDa light and two ~50 kDa heavy chains, identical among themselves, with the latter being N-glycosylated, **Fig. 1**. The four chains are interconnected and their tertiary/quaternary structure is reinforced by inter- and intra-molecular disulfide bridges. Heavy chains can be translated from different genes, and their sequence determines the so-called IgG isotype (or subclass). In the pharmaceutical industry, three of the four IgG isotypes are of interest: IgG1, IgG2, and IgG4. Although all isotypes share ~90% of homology in their amino acid sequences, the structure of their hinge regions and the disulfide bond connectivity are significantly different. IgG1 and IgG4 contain 16 disulfide bridges, two in the hinge linking the heavy chains, one disulfide bridge linking each light chain to the heavy chain and 12 intramolecular bridges (2 and 4 for each light and heavy chain, respectively). IgG2 contains 18 disulfide bonds with four in the hinge region linking the two heavy chains together.[3] Figure 1 compares the structures of an IgG1 and an IgG2. Note that the C-terminal cysteine of the light chain is linked to the third

cysteine of the heavy chain in IgG2 (and also IgG4) whereas in IgG1 it is linked to the fifth cysteine.

Due to their bioproduct nature, high-throughput extensive characterization of IgGs structure is needed.[4] Mass spectrometry (MS) plays a pivotal role in structure analysis of IgGs. MS-based methods have been developed and implemented across all stages of mAb production.[3] The basic workflow may include intact IgG mass measurement, IgG fragment (reduced light and heavy chains or papain-produced fragments) mass measurement (referred to as middle-up MS approach),[5] middle-down[6] and bottom-up MS.[2, 7] Although bottom-up approaches provide the most structural information today, they suffer a number of drawbacks such as artifact introduction and lengthy sample preparation.[8-10] Top-down mass spectrometry (TD MS) refers to the gas-phase dissociation of intact biomolecules and subsequent mass measurement of the reaction products.[11] In this perspective, TD MS of intact IgGs is a convenient way to obtain primary structure information while keeping the connectivity between light and heavy chains. Furthermore, this approach generally allows access to sequence confirmation of terminal regions as well as variable domains. It also allows the identification and localization of major modifications and the characterization of major glycoforms. It has the advantage of implicating limited sample preparation and therefore minimal artifact introduction due to sample processing. Despite these advantages, TD MS is not yet widely used for the characterization of mAbs, primarily due to the large number of highly charged

products yielded by IgG fragmentation which requires high-resolution MS and substantial acquisition time for improved sensitivity.[12]

Electron capture dissociation (ECD)[13, 14] and electron transfer dissociation (ETD)[15] of proteins usually generate larger sequence coverage than slow-heating activation methods (such as collision-induced dissociation, CID).[16] ECD and ETD MS/MS are also known as powerful methods for characterization of labile post-translational modifications, e.g., glycosylation.[17, 18] Fourier transform ion cyclotron resonance (FT-ICR) MS provides the highest resolution and might be hyphenated with ECD making it, until recently, the technique of choice for TD MS experiments.[19-21] Notably, Mao *et al.* achieved 25% sequence coverage of IgG1 using qFT-ICR MS and ECD when isolating a single charge state in a direct infusion ionization mode. They increased the sequence coverage to ~33% with a broadband tandem mass spectrometry (MS/MS), e.g., using all the IgG1 ions present in the charge state envelope as precursors for ECD.[22] However, the general use of FT-ICR MS instruments for IgG characterization is rare,[3] presumably due to their limited presence in biopharmaceutical laboratories owing to the tedious maintenance and operational costs.

With the advent of ETD efficiently performed in an external ion trap with product ion detection in a high-resolution mass analyzer, e.g., Orbitrap FTMS[23] or time-of-flight (TOF) MS,[24, 25] the performance of TD MS has improved. Particularly, TD MS of proteins on-line separated with reversed phase liquid chromatography (LC) prior to electrospray ionization (ESI) became

possible routinely.[26, 27] We have previously described the use of ETD with a high-resolution qTOF MS reporting the overall sequence coverage of 21% for mouse IgG1 and 15% for human IgG1 which provided structural information on variable domains of both the heavy and the light chain.[28] As a result of a more extensive investigation, we achieved 33% sequence coverage of a human IgG1 using ETD with a hybrid linear ion trap (LTQ) Orbitrap FTMS (considering c-, z- and y-type product ions). The results obtained with narrow (i.e., 100 Th) and large (i.e., 600 Th) isolation windows for precursor ions were combined to increase the final number of identified product ions.[29] In all the aforementioned TD MS studies, over one thousand scans (or transients, in the case of FTMS) were averaged to achieve maximum sequence coverage.[22, 28, 29] Naturally, a number of LC MS/MS runs were acquired followed by time-domain data averaging to provide the required number of scans in the LC-MS/MS approach of TD MS. The current state of the art in TD MS, however, does not enable the sequencing of the disulfide bond protected domains of both light and heavy chains, neither it allows the direct localization of the glycosylation site in the C_H2 domain of the heavy chain. This is due to the compact protein structure and its disulfide bond network (see Fig. 1), which constitute the major limitation in IgG TD MS characterization. Overall, the performance of all three approaches, using ETD on a qTOF MS or on LTQ Orbitrap FTMS, as well as ECD on FT-ICR MS, provide comparable results.

Here we investigated the potential of ETD implemented on a high-field Orbitrap FTMS[30] for faster top-down characterization of three therapeutic

mAbs: Adalimumab and Trastuzumab (both of the IgG1 subclass), and the IgG2 Panitumumab. The high-field Orbitrap mass analyzer is characterized by a higher frequency of ion oscillations along the central electrode compared to the prior generation of Orbitrap mass analyzers.[31] Increased frequency of ion oscillations leads to the corresponding enhancement of the resolving power over a fixed acquisition time. Moreover the implementation of FT absorption mode of spectral representation (known as "eFT" algorithm) allows a further gain (of up to two-fold) in acquisition rate at a constant resolution.[32] In addition to the advantages of the employed mass analyzer, we optimized precursor ion isolation and product ion transfer between the LTQ and the Orbitrap to increase the efficiency of ETD-based TD MS. Finally, IgG TD MS was performed under the optimized conditions with a two-stage ion activation method, a combination of ETD and higher-energy collision dissociation (HCD)[33] termed EThcD, which previously was shown to be beneficial for extending peptide sequence coverage.[34, 35]

2. Materials and methods

2.1 Chemicals. Water, acetonitrile (ACN), methanol (MeOH), formic acid (FA) and trifluoroethanol (TFE) were obtained in LC-MS purity grade. Water, ACN and MeOH were purchased from Fluka Analytical (Buchs, Switzerland). FA was obtained from Merck (Zug, Switzerland), and TFE from Acros Organics (Geel, Belgium).

2.2 Sample preparation. Therapeutic monoclonal antibodies of the IgG1 class, namely Adalimumab (Humira, Abbot Laboratories) and Trastuzumab (Herceptin, Genentech), as well as the IgG2 Panitumumab (Vectibix, Amgen) are the European Medicines Agency approved versions and formulations, available commercially to a general public. Prior to direct infusion analysis, all antibodies were buffer exchanged against a 150 mM ammonium acetate solution pH 6.7 using microspin desalting columns (Zeba 75 µL, Thermo Scientific, Zug, Switzerland).

2.3 Sample delivery and ionization. The antibodies were injected into the mass spectrometer either by on-line coupled LC or by direct infusion in a positive ion mode of ESI. In the former case, we employed an Ultimate 3000 LC system (Thermo Scientific, Amsterdam, The Netherlands) equipped with a reversed phase C4 column (Hypersil gold, 1x100 mm, 5 µm particle size, Thermo Scientific, Runkorn, UK), applying a gradient of ACN from 5 to 80% in 20 min, in presence of 0.1% FA, and a flow rate of 100 µl/min generated by the loading pump. 1.5 µg of antibody (~10 pmol), diluted in 20% ACN and 0.1% FA, were typically loaded in each injection, resulting in ~2 min elution of the IgG. The microESI source (IonMax ion source, Thermo Scientific, Bremen, Germany) was operated at 3.7 kV, with sheath gas set to 20 and auxiliary gas to 10 arbitrary units. In the latter case, the antibody, buffer exchanged as described above, was diluted in 50% ACN and 0.1% FA to a final concentration of ~10 µM; 6 µl of this IgG solution were loaded in a 20 µl loop connected to the injection

valve of the LC system and infused through a silica capillary to the nanoESI source (Nanospray Flex ion source, Thermo Scientific) equipped with a metallic emitter to which a 2.4 kV voltage was applied. The composition of the solution used to push the IgG from the loop to the source was 39.9% water, 30% ACN, 20% MeOH, 10% TFE, and 0.1% FA. The flow rate generated by the nanopump of the LC system was 0.6 μl/min.

2.4 Mass spectrometry. All experiments were performed on a hybrid dual linear ion trap high-field Orbitrap FT mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific, Bremen, Germany) equipped with ETD and HCD. Similarly to our previous report,[29] the S-lens RF level was set to 70% and the temperature of heated transfer capillary was 350° C. Broadband mass spectra of intact IgGs and tandem mass spectra of the same molecules were recorded in separate experiments using ion detection in the Orbitrap FTMS (m/z 200-4000). All Orbitrap FTMS scans were recorded averaging 10 microscans to improve the SNR. Orbitrap FTMS was calibrated in the high mass range (m/z 2000-4000) using ion at 2021.93954 m/z of commercial LTQ Velos Calmix (Thermo Scientific). Intact mass measurements of IgGs were performed at 15'000 resolution at 400 m/z with a target value of charges (automatic gain control, AGC) of 1 million. For ETD experiments, precursor ions were isolated between m/z 2400 and 2900 in the high pressure LTQ and subsequently subjected to ETD reaction. The AGC value for fluoranthene radical anions was set to 2 million charges, anion maximum injection time was set to 50 ms and

ETD duration (i.e., ion-ion interaction time) was set to 10 or 25 ms. Product ion detection in the Orbitrap mass analyzer was performed with 120'000 resolution (at 400 m/z, eFT enabled) and an AGC of 1 million charges.

For both intact mass measurements and tandem mass spectrometry, the gas (N_2) "delta pressure" in the Orbitrap mass analyzer region, calculated as the difference between the pressure in presence of gas flow into the HCD cell and when the gas is completely shut down, was of 0.1E-10 torr. The typical value of this pressure difference for bottom-up proteomics (smaller size ions) is ~0.35E-10 torr. To further improve ion transmission to the high-resolution mass analyzer from the LTQ, ions ejected from the LTQ were first trapped and stored in the HCD cell (known as "HCD trapping"), then transferred backward to the C-trap, and finally injected into the Orbitrap as previously described.[36]

EThcD experiments were performed as follows. After ETD with 10 ms duration, all ETD products, including both charge-reduced species and product ions (precursors are generally completely consumed in ETD under the applied conditions), were transferred from the LTQ to the HCD cell. Here the collision energy was set to 25 or 40 eV, so that instead of ion capture and relaxation a second fragmentation event could occur. Finally, ions were transferred back to the C-trap and further into the Orbitrap for detection.

2.5 FTMS data processing and analysis. The data processing routine for obtaining high SNR tandem mass spectra was performed as previously described.[29] Briefly, in both LC-MS/MS and direct infusion MS/MS

experiments, Orbitrap FTMS time-domain transients were recorded in MIDAS .dat format[37] and underwent averaging before being processed with the standard FT procedure for time-to-frequency conversion. In this way it was possible to average scans derived from separate LC runs or distinct direct infusion experiments. The eFT signal processing was not applied to the averaged transients, whereas it should be feasible in principle. The averaged transients were then converted into standard Thermo .RAW files for further viewing and processing with XCalibur software (Thermo Scientific), as well as data analysis (*vide infra*). Final ETD mass spectra were derived from summed transients recorded for 10 and 25 ms ETD duration experiments. For an in-depth evaluation of the effect of ETD duration on IgG fragmentation, 10 and 25 ms ETD datasets were also analyzed separately.

Previously, the combination of deconvolution software Xtract and product ion assignment software ProSight PC 3.0 (both from Thermo Scientific) was successfully employed for IgG TD MS data analysis.[30] Here, we evaluated the performance of a recently developed workflow combining a deconvolution software MS-Deconv and product ion assignment software MS-Align+ (both developed at University of California San Diego, San Diego, CA, USA).[38] A comparison between results produced by MS-Deconv/MS-Align+ and by Xtract/ProSight PC is presented in Fig. S1, Supporting Information. Both the data analysis packages led to comparable results, with MS-Deconv/MS-Align+ being faster to use (less manual intervention). Briefly, the MS-Deconv/MS-Align+ analyses were realized according to the following steps: (i) ReAdW was

used to convert .RAW files into centroided mzXML files; (ii) a customized version of MS Deconv was used to deconvolute the mzXML files using a SNR cutoff of 3. The custom version of MS Deconv was written to process the data sets containing MS/MS information only, whereas the standard version of MS-Deconv requires both MS and the directly corresponding to it MS/MS data sets; (iii) MS-Align+ was used for product ion identification, considering c- and $z\cdot$- type ions for ETD experiments and c-, $z\cdot$-, b- and y-type ions for EThcD data. Cysteine residues were considered as not modified and all the mass tolerances were set to 20 ppm. Manual validation was performed for cleavage sites identified from product ions present only in a single charge state.

2.6 Product ion abundance analysis (PIA). Abundances of identified product ions were derived from MS-Align+ analysis. For each cleavage site, all the related product ions were considered and grouped. Final abundances were obtained by dividing the abundance values reported for each product ion over the ion charge state, and summing all the resulting charge-normalized abundance values of product ions generated from a single cleavage site. For plotting together product ion abundance (PIA) analysis results of different sets of data, the abundances were expressed as relative percentage of the most abundant product ion cluster (i.e., the group of all the z-type ions, with different charge state, derived from a single cleavage site) in each data set.

3. Results and discussion

3.1 Optimization of ETD MS/MS of IgG. To maximize the ETD efficiency, we sought to optimize precursor cation selection and product ion transfer. In the former work based on a hybrid LTQ Velos Pro Orbitrap FTMS we isolated precursor ions around *m/z* 2750 (isolation window width: 100 Th, precursor ions from 53+ to 55+) or 2900 (isolation window width: 600 Th, precursor ions from 47+ to 57+). Considered the charge state envelope obtained with the current LC-MS setting on the LTQ Orbitrap Elite for the intact IgG1 Adalimumab, we centered the isolation window at *m/z* 2700, with a width of 600 Th, thus including precursors from 50+ to 61+ (Fig. S2, Supporting Information). Higher charge state precursor ions are supposed to increase the fragmentation efficiency in ETD and facilitate the separation of obtained product ions bound by non-covalent bonds. Furthermore, cleavage preferences within protein structure shift as a function of charge location. In regard to product ion transfer, we applied a combination of reduction of the gas pressure in the HCD cell and the "HCD trapping", *vide supra*. The gas pressure reduction improves SNR of large product ions by improving their transmission from LTQ into the HCD cell and further in the Orbitrap analyzer. Trapping of ions in the HCD cell (and not in the C-trap) favors the transmission also of lighter ions, counterbalancing the reduced damping effect in the C-trap resulting from the lower gas pressure (Fig. S3, Supporting Information).

Overall, the implementation of these settings allowed for SNR increase in the MS/MS mode for high *m/z* ions, which resulted in a reduced number of

transients needed to be averaged for obtaining high-quality mass spectra in comparison to our previous work, *vide infra*. Each ETD mass spectrum derived from a combination of transients was obtained applying two different ion-ion interaction times, 10 and 25 ms, **Fig. 2**. Increasing ETD duration enhanced the formation of product ions with a reduced charge state, likely because of the occurrence of multiple electron transfer and consequent charge reduction events. Nevertheless, the separate analysis of mass spectra obtained with 10 or 25 ms ETD duration revealed that most of the identified cleavage sites are shared between the two experiments (see product ion abundance analysis results below), suggesting that ETD is primarily limited by the retention of a compact conformation of the IgG in the gas phase. Summing transients from these two experimental sets, though, increased the confidence in the assignment of cleavage sites, as product ions obtained by the same backbone cleavage are identified in multiple charge states.

3.2 ETD and EThcD MS/MS of intact human IgG1. The analysis of product ions obtained by ETD MS/MS was conducted for two IgG1 mAbs, Adalimumab and Trastuzumab. These share more than 92% of their sequences, the differences being almost entirely concentrated in the complementarity determining regions (CDRs), involved in the antigen binding. **Fig. 3A** displays the ETD fragmentation map of Adalimumab, with the typical fragmentation pattern expected for this IgG isotype: extensive fragmentation is achieved on the heavy chain for the loop interconnecting V_H and C_H1, the loop between C_H2

and C_H3 and part of the C_H3 domain; on the light chain, the characterized region is represented by the loop between V_L and C_L. For both chains, therefore, the sequenced areas are mainly disulfide bond-free loops exposed to the solvent. The results of ETD MS/MS application to Trastuzumab are nearly identical to those for Adalimumab (Fig. S4, Supporting Information). For both the IgGs only the CDR3 domains of light and heavy chains are fully sequenced, and only a few product ions positioned on CDR2 of heavy and light chain have been assigned for Trastuzumab (which shows slightly higher sequence coverage). As summarized in **Table 1**, 310 ETD MS/MS scans (which corresponds to 3100 transients) were summed to result in 23.8% and 26.5% sequence coverage for Adalimumab and Trastuzumab, respectively. Notably, from the comparison of these results and the previously reported ones it is apparent how, although the total sequence coverage (i.e., sum of cleavages in light and heavy chains) is similar, the distribution of identified cleavage sites between the two chains is changed, with the heavy chain being more sequenced in the present report than in the previous one. For Trastuzumab we reached 29% sequence coverage which represents the highest value obtained so far for the heavy chain of an intact IgG1 by ETD TD MS. On the other hand, the light chain has been less sequenced with the new settings and data analysis routine than in the past. Manual inspection of MS/MS spectra reveals that the missing product ions are mainly those arising from cleavage sites located at the two termini of the light chain: these are light, lowly charged ions that presumably were not favorably transmitted with the newly employed

166

settings (Fig. S3 Supporting Information). The fully sequenced region in the light chain, corresponding to the loop between V_L and C_L, as mentioned above, remained the same between the two studies, and it is equal between Adalimumab and Trastuzumab (both of which often show complementary *c*- and *z*-type ions for the same cleavage site). Cleavages outside this area do not follow any clear pattern and seem randomly distributed.

To improve the sequence coverage of the IgG1 and, specifically, of the light chain, we applied EThcD on Adalimumab. We summed transients obtained with EThcD (i.e., both product ions and charge-reduced species) performed at both 25 and 40 eV collision energy. As shown in Figure 3B, EThcD produced fragmentation mainly in the same areas previously covered by ETD, but also induced new cleavages at the N-terminus of both light and heavy chain, and provided *b*- and *y*-type ions from the N-terminal side of proline that could not be previously identified from ETD data. The sequence coverage for the heavy chain of Adalimumab is lower than for ETD, but it passed from ~19% to ~28% for the light chain, as reported in Table 1. Overall, the combination of 310 ETD scans and 260 EThcD scans, for a total of 570 scans, produced 30.5% sequence coverage, which is the same result obtained by averaging 1360 scans in our previous work based on the Orbitrap Velos Pro. In addition to the reduced number of scans, a two-fold scan speed improvement due to the implementation of the high-field Orbitrap mass analyzer and a further two-fold improvement potentially granted by the application of eFT substantially decreased the required experimental time. In particular, at 120'000 resolution

(at 400 *m/z*) and under LC-MS/MS conditions, 120 transients (microscans) could be recorded per minute of IgG elution on the Orbitrap Elite FTMS versus the previous 40 transients (microscans) on the Orbitrap Velos Pro (note, that the ETD MS/MS experimental sequence, including cation accumulation, anion injection and subsequent cation/anion interaction largely contributes to the duty cycle, partially counterbalancing the time gain in FTMS data acquisition due to combination of high-field Orbitrap and eFT).

Finally, from the comparison of the ETD TD MS results obtained with different ion transfer parameters (especially the pressure conditions in the HCD cell), those reported here and in the previous work, it seems beneficial to acquire ETD TD MS data with both conditions and pooling the data together. Indeed, results reported in Fig. S5 (Supplementary Information) show that the sequence coverage for ETD MS/MS of Adalimumab with only *c* and *z*-ions is increased to 41.3% and 34.7% for light and heavy chain, respectively, with a total sequence coverage of 36.8%.

3.3 Product ion abundance (PIA) analysis of ETD MS/MS of IgGs1. A PIA analysis was conducted on ETD MS/MS data for the C-terminal portions of the almost completely sequenced heavy chains of the investigated IgGs1. **Fig. 4** shows a comparison between the relative abundances of *z*-ions from 10 ms ETD of Adalimumab, as well as 10 ms and 25 ms ETD of Trastuzumab. Note, that the two IgGs share identical sequence in the analyzed area. For all the three data sets, the average abundance of product ions generated by the

disulfide-free loop connecting C_H2 and C_H3 is ~40% of the maximum, and global maxima are all located there. Conversely, ions derived from the disulfide-protected area of C_H3, as well as those from the disulfide-free C-terminal end (which is, structurally speaking, a part of the C_H3 domain), have generally an abundance lower than 10% of the corresponding maximum. Furthermore, the global maxima for the two 10 ms ETD data sets are both located at the cleavage site between K_{105} and G_{106} (which results in z_{106} ions). Other local maxima correspond to the formation of z_{96}, z_{93}, and z_{86} ions. The 25 ms data set shows a similar trend, with z_{106} still abundant but surpassed by z_{93}, z_{91}, and z_{86}. Therefore, the ETD results are conditioned by the secondary and tertiary structures of the gaseous protein. Moreover, similarly to ECD MS/MS,[39] the ETD radical process seems to be affected by the electron transfer site, as the presence of basic residues at one side of the cleavage maxima would suggest. Finally, the shift in maxima occurring when passing from 10 to 25 ms ETD might be explained by the increased probability of secondary electron transfer events when the ETD duration is prolonged. The occurrence of secondary electron transfer and consequent charge neutralization events might be confirmed by the reduced average charge state of the product ions identified for 25 ms ETD in comparison to 10 ms ETD experiments (Table S1, Supporting Information). The data suggests that these events are most prominent at the flexible loop and with a C_H3 domain in a compact conformation. Therefore, despite the second fragmentation event the abundance of product ions originated within the sequence of the C_H3 domain is not higher than for the 10

ms experiments. Instead, we observe the shift between z_{106} and z_{86} as maxima in the disulfide-free region. Visually, this is represented in **Fig. 5** through a color-coded map based on crystal structure of Trastuzumab (image elaborated with The PyMol Molecular Graphics System, version 1.3, Schrodinger, LLC). The passage from 10 ms to 25 ms ETD duration induces the translation of highly frequent cleavage sites (represented in red) towards the C-terminus, but still in solvent accessible positions. On the contrary, the compact "immunoglobulin-like domain" corresponding to C_H3, which resemble the structure of a beta barrel, and contains an intra-molecular disulfide bridge, remains poorly sequenced (and thus represented in black-blue). These results reinforce the hypothesis that such large proteins as IgGs retain a relatively compact high-order structure in the gas phase, even under denaturing ionization conditions.

3.4 ETD MS/MS of intact human IgG2. ETD MS/MS was performed on Panitumumab, an IgG2, to provide a confirmation to our hypothesis that limitations in achieved sequence coverage for IgGs1 have mainly to be attributed to gas-phase retention of a structured, compact conformation by the protein. As displayed in **Fig. 6**, identified product ions for this IgG are localized in the areas sequenced in the experiments involving Adalimumab and Trastuzumab, e.g., the solvent accessible loops interconnecting consecutive disulfide-protected domains and the C_H3 domain. In-depth analysis reveals that the C-terminal portion of the heavy chain is highly sequenced, with 83

assigned z-type ions versus 82 z-ions identified for Trastuzumab, which has overall the highest sequence coverage for its heavy chain (Table 1). Nevertheless, given that, in accordance to what is observed for IgGs1, the hinge region is not sequenced also in the case of Panitumumab, differences between the two IgG isotypes are present in the loop interconnecting V_H and C_H1. In the case of IgGs1, this loop was fully sequenced, whereas for Panitumumab only partial sequencing was achieved. This can be explained by the presence in the loop of the IgG2 of Cys_{133}, involved in an inter-molecular disulfide bridge connecting the heavy with the light chain. Notably, most of the identified c-type product ions are located to the N-terminus of C_{133}. Further studies are needed to determine whether ETD cleavages were primarily impeded in the portion of the loop at the C-terminal side of C_{133} or, conversely, cleavages occurred but primarily produced product ions with portions of the light chain still present and therefore not identifiable by the present version of the data analysis approach. Finally, Panitumumab's light chain has been sequenced in its central area (i.e., disulfide-free loop between V_L and C_L), similarly to IgGs1, but its incomplete sequencing and, importantly, the limited number of identified z-ions (9, corresponding to only ~50-30% of those assigned in the other experiments), could indicate that the inter-molecular disulfide bridge with the heavy chain is less efficiently cleaved in IgG2 than in IgG1.

4. Concluding remarks

Recent developments in Orbitrap FTMS technology mark a substantial improvement in high-resolution MS data acquisition rate. Specifically, around three-fold faster acquisition of IgG ETD data under optimized conditions in the Orbitrap Elite FTMS (high-field compact Orbitrap) combined with a two-fold reduction in the number of transients to be averaged resulted in about six times reduced acquisition time in the present study relative to the analysis with the Orbitrap Velos Pro FTMS (standard field Orbitrap). Particularly, selection of higher charged precursors, optimization of product ion transfer from the ETD reaction cell to the high-resolution mass analyzer by reduction of HCD gas pressure, combination of experiments conducted with different ion-ion reaction times, and use of EThcD with two different collisional energy values all contributed to the reduction of number of scans needed for high quality TD MS analysis. Nevertheless, the previously determined limit in sequence coverage, ~30%, could not be overcome. The primary cause for the limited sequence coverage is the gas-phase retention of highly structured areas in the IgG, mainly in correspondence of the immunoglobulin-like domains.

Extending the sequence coverage in TD MS of IgGs may be provided from improved ETD reaction cell,[40] chemical ionization source for front-end ETD on Orbitrap FTMS,[41] and the use of ion activation methods alternative to ETD, e.g., ultraviolet photodissociation.[42] Furthermore, improved sequence coverage shall be achieved by combining ETD experimental data acquired with high HCD gas pressure (i.e., maximizing light product ion transmission), and

low gas pressure (i.e., maximizing heavy product ion transmission), as it results from the combination of present data with those acquired in standard operational conditions with the LTQ Orbitrap Velos Pro FTMS. Finally, the probable presence of internal products and product ions retaining parts of both light and heavy chains bonded by inter-molecular disulfide bridges will require developments in the data analysis tailoring this specific class of proteins.

5. Acknowledgements

6. References

[1] Beck, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal Chem* 2012, *84*, 4637-4646.

[2] Ayoub, D., Jabs, W., Resemann, A., Evers, W. et al., Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *MAbs* 2013, *5*, 699-710.

[3] Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsselaer, A., Sanglier-Cianferani, S., Characterization of therapeutic antibodies and related products. *Anal Chem* 2013, *85*, 715-736.

[4] Beck, A., Diemer, H., Ayoub, D., Debaene, F. et al., Analytical characterization of biosimilar antibodies and Fc-fusion proteins. *TrAC Trends in Analytical Chemistry* 2013, *48*, 81-95.

[5] Wang, B., Gucinski, A. C., Keire, D. A., Buhse, L. F., Boyne, M. T., Structural comparison of two anti-CD20 monoclonal antibody drug products using middle-down mass spectrometry. *Analyst* 2013, *138*, 3058-3065.

[6] Fornelli, L., Ayoub, D., Aizikov, K., Beck, A., Tsybin, Y. O., Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap FTMS. *submitted*

[7]   Zhang, Z., Pan, H., Chen, X., Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom Rev* 2009, *28*, 147-176.

[8]   Krokhin, O. V., Antonovici, M., Ens, W., Wilkins, J. A., Standing, K. G., Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* 2006, *78*, 6645-6650.

[9]   Dick, L. W., Kim, C., Qiu, D. F., Cheng, K. C., Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. *Biotechnology and Bioengineering* 2007, *97*, 544-553.

[10]   Diepold, K., Bomans, K., Wiedmann, M., Zimmermann, B. et al., Simultaneous assessment of Asp isomerization and Asn deamidation in recombinant antibodies by LC-MS following incubation at elevated temperatures. *PLoS One* 2012, *7*, e30295.

[11]   Zhang, Z., Shah, B., Characterization of variable regions of monoclonal antibodies by top-down mass spectrometry. *Anal Chem* 2007, *79*, 5723-5729.

[12]   Han, X., Jin, M., Breuker, K., McLafferty, F. W., Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science* 2006, *314*, 109-112.

[13]   Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 1998, *120*, 3265-3266.

[14]   Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L. et al., Electron capture dissociation for structural characterization of multiply charged protein cations. *Analytical Chemistry* 2000, *72*, 563-573.

[15]   Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2004, *101*, 9528-9533.

[16]   Wells, J. M., McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol* 2005, *402*, 148-185.

[17]   Meier, S., Tsybin, Y., Dyson, P., Keppler, B., Hartinger, C., Fragmentation methods on the balance: unambiguous top–down mass spectrometric characterization of oxaliplatin–ubiquitin binding sites. *Analytical and Bioanalytical Chemistry* 2012, *402*, 2655-2662.

[18]   Molina, H., Matthiesen, R., Kandasamy, K., Pandey, A., Comprehensive Comparison of Collision Induced Dissociation and Electron Transfer Dissociation. *Analytical Chemistry* 2008, *80*, 4825-4835.

[19]   Ge, Y., Rybakova, I. N., Xu, Q., Moss, R. L., Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc Natl Acad Sci U S A* 2009, *106*, 12658-12663.

[20]   Valeja, S. G., Kaiser, N. K., Xian, F., Hendrickson, C. L. et al., Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 2011, *83*, 8391-8395.

[21]   Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R. et al., Nano-LC FTICR tandem mass spectrometry for top-down proteomics: routine baseline unit mass resolution of whole cell lysate proteins up to 72 kDa. *Anal Chem* 2012, *84*, 2111-2117.

[22]   Mao, Y., Valeja, S. G., Rouse, J. C., Hendrickson, C. L., Marshall, A. G., Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-Fourier transform ion cyclotron resonance-mass spectrometry. *Anal Chem* 2013, *85*, 4239-4246.

[23]   McAlister, G. C., Berggren, W. T., Griep-Raming, J., Horning, S. et al., A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res* 2008, *7*, 3127-3136.

[24]   Hartmer, R. G., Kaplan, D. A., Stoermer, C., Lubeck, M., Park, M. A., Data-dependent electron transfer dissociation of large peptides and medium size proteins in a QTOF instrument on a liquid chromatography timescale. *Rapid Commun Mass Spectrom* 2009, *23*, 2273-2282.

[25]   Fornelli, L., Parra, J., Hartmer, R., Stoermer, C. et al., Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal Bioanal Chem* 2013, *405*, 8505-8514.

[26] Tian, Z., Tolic, N., Zhao, R., Moore, R. J. et al., Enhanced top-down characterization of histone post-translational modifications. *Genome Biol* 2012, *13*, R86.

[27] Ahlf, D. R., Compton, P. D., Tran, J. C., Early, B. P. et al., Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J Proteome Res* 2012, *11*, 4308-4314.

[28] Tsybin, Y. O., Fornelli, L., Stoermer, C., Luebeck, M. et al., Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Analytical Chemistry* 2011, *83*, 8919-8927.

[29] Fornelli, L., Damoc, E., Thomas, P. M., Kelleher, N. L. et al., Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics* 2012, *11*, 1758-1767.

[30] Michalski, A., Damoc, E., Lange, O., Denisov, E. et al., Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 2012, *11*, O111.013698.

[31] Makarov, A., Denisov, E., Lange, O., Performance evaluation of a high-field Orbitrap mass analyzer. *J Am Soc Mass Spectrom* 2009, *20*, 1391-1396.

[32] Lange, O., Damoc, E., Wieghaus, A., Makarov, A., Enhanced Fourier Transform for Orbitrap Mass Spectrometry. *Proc. 59th Conf. Amer. Soc. Mass Spectrom.,* Denver, June 5 9, 2011.

[33]   Olsen, J. V., Macek, B., Lange, O., Makarov, A. et al., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007, *4*, 709-712.

[34]   Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D. et al., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* 2012, *84*, 9668-9673.

[35]   Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F. et al., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* 2013, *12*, 1520-1525.

[36]   Rosati, S., Rose, R. J., Thompson, N. J., van Duijn, E. et al., Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew Chem Int Ed Engl* 2012, *51*, 12992-12996.

[37]   Senko, M. W., Canterbury, J. D., Guan, S., Marshall, A. G., A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 1996, *10*, 1839-1844.

[38]   Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C. et al., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics* 2010, *9*, 2772-2782.

[39]   Breuker, K., Oh, H., Lin, C., Carpenter, B. K., McLafferty, F. W., Nonergodic and conformational control of the electron capture

dissociation of protein cations. *Proc Natl Acad Sci U S A* 2004, *101*, 14011-14016.

[40] Rose, C. M., Russell, J. D., Ledvina, A. R., McAlister, G. C. et al., Multipurpose dissociation cell for enhanced ETD of intact protein species. *J Am Soc Mass Spectrom* 2013, *24*, 816-827.

[41] Earley, L., Anderson, L. C., Bai, D. L., Mullen, C. et al., Front-end electron transfer dissociation: a new ionization source. *Anal Chem* 2013, *85*, 8385-8390.

[42] Shaw, J. B., Li, W. Z., Holden, D. D., Zhang, Y. et al., Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *Journal of the American Chemical Society* 2013, *135*, 12646-12651.

Figure caption

Figure 1. Comparative schematics of the IgG1 (Adalimumab and Trastuzumab) and IgG2 (Panitumumab). The differences are related to a disposition of inter-molecular disulfide bridges (connecting either heavy and light chain, or the two heavy chains). IgG1 has the C-terminal Cys of light chain connected to the fifth Cys residue of the heavy chain, whereas the C-terminal Cys of light chain in IgG2 is connected to the third Cys of the heavy chain. The hinge region of IgG1 has 2 disulfide bridges between heavy chains, whereas IgG2 contains 4 S-S bonds there.

Figure 2. Top-down ETD Orbitrap FTMS of IgG1 (Adalimumab) with ion-ion interaction time of 10 and 25 ms. (A) Combined mass spectrum including transients from 10 and 25 ms, with product ion population arising from the former experiment highlighted in blue, and from the latter indicated in red. (B) and (C) Expanded views of the two product ion populations, with assigned product ions derived from the heavy chain. Note the difference in the average charge state of the product ions of the two populations.

Figure 3. Fragmentation map of Adalimumab by top-down Orbitrap FTMS. (A) ETD MS/MS. (B) EThcD MS/MS. The heavy chain is position on top, light chain at the bottom.

Figure 4. Product ion abundance analysis of ETD top-down MS of Adalimumab with 10 ms ETD duration and Trastuzumab with 10 and 25 ms duration.

Figure 5. Color-coded representation of ETD PIA results for the C-terminus of the heavy chain of Trastuzumab. (A) 10 ms ETD. (B) 25 ms ETD. The displayed image is based on X-ray crystallography data (PDB entry: 3D6G).

Figure 6. Fragmentation map of Panitumumab (IgG2) obtained with top-down ETD Orbitrap FTMS.

Table 1. Sequence coverage obtained for different IgGs1 and IgG2 by top-down ETD Orbitrap FTMS. Indicated scans consist of 10 microscans each. Note, only c- and z-type ions were considered for ETD MS (for both "high-field" and "standard" Orbitrap data), whereas also b- and y-type ions were included in EThcD MS analysis.
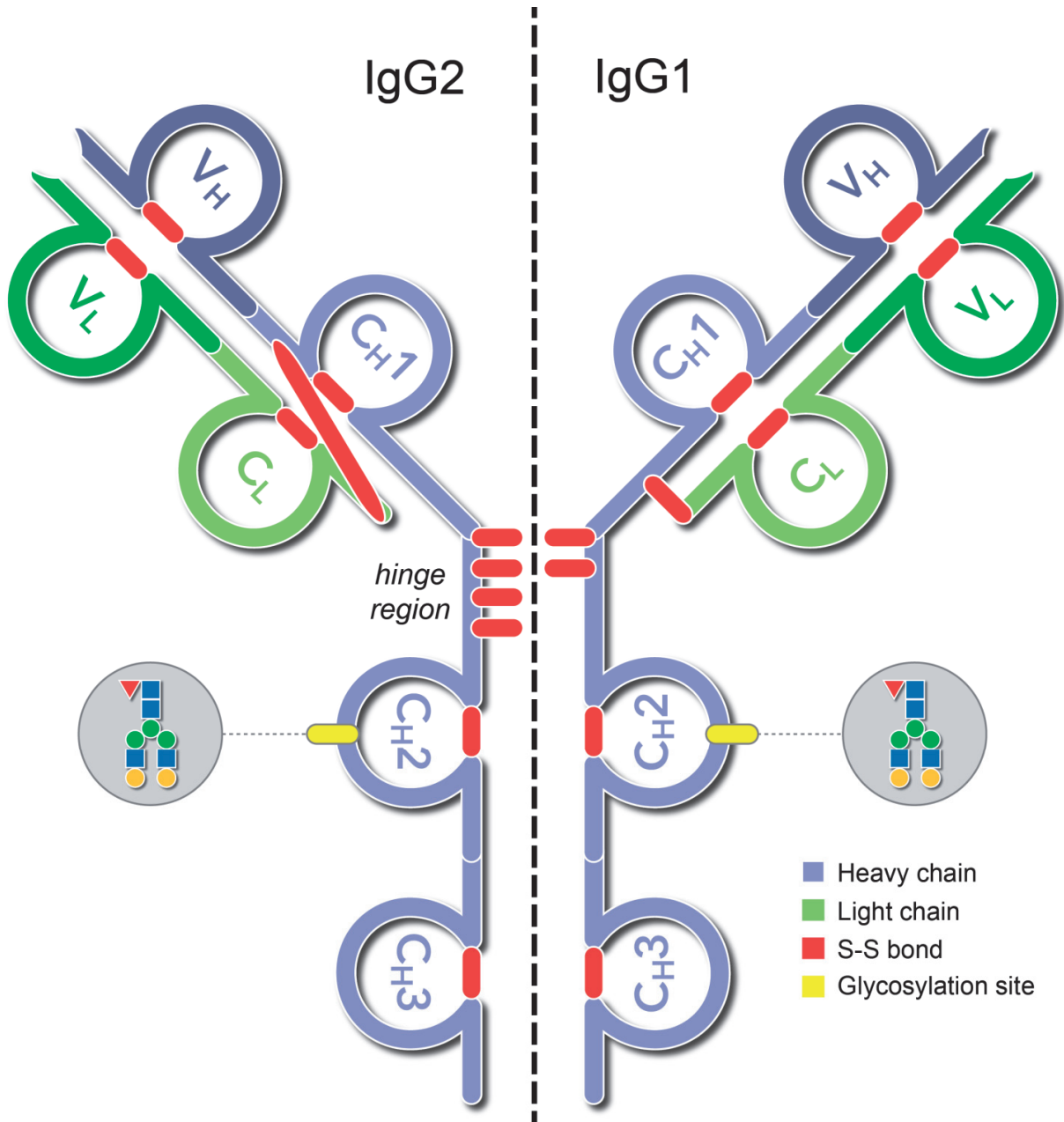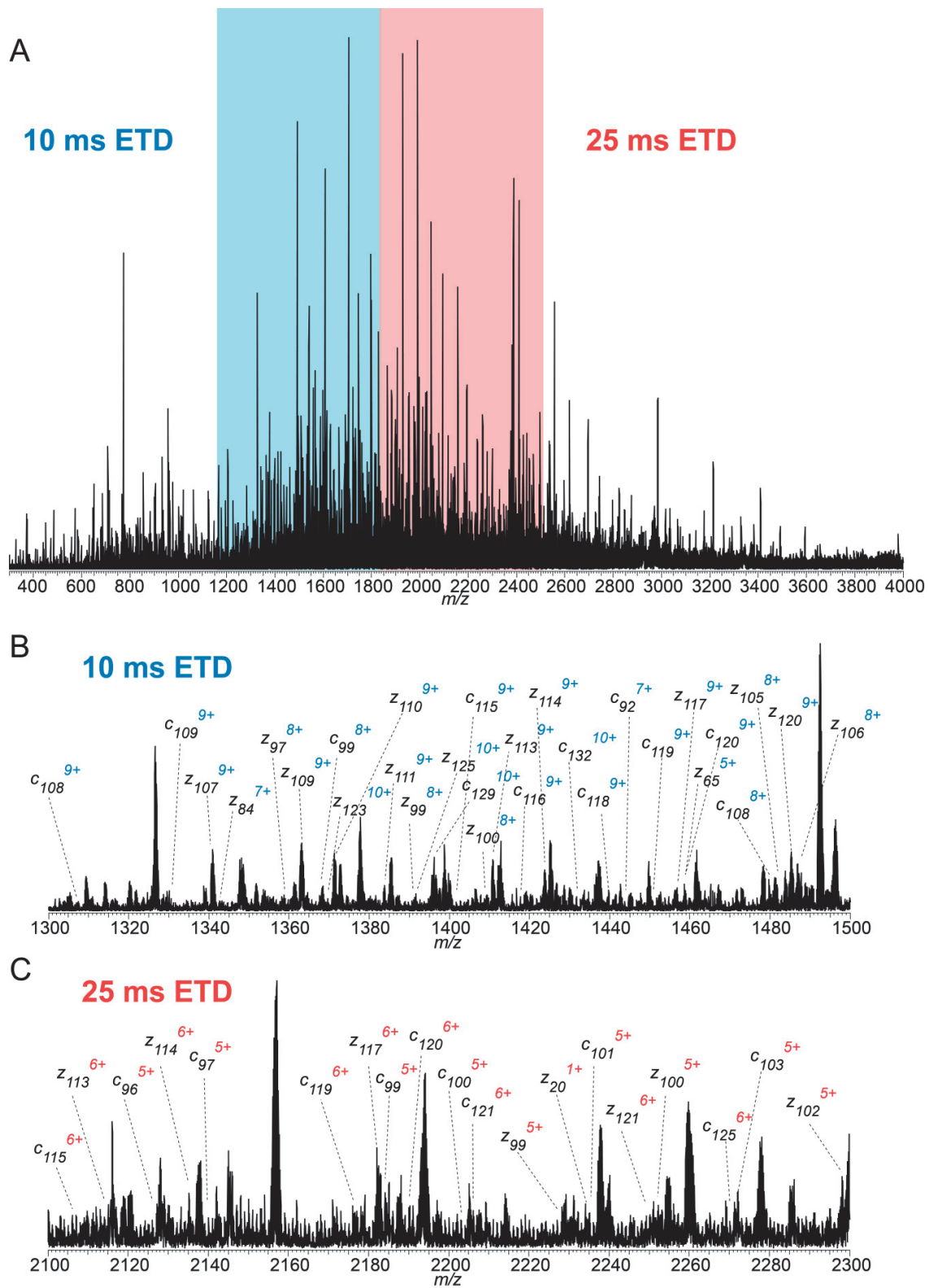
Figure 1.

Figure 2.

Figure 3.

A
ETD

B
EThcD

product ions:
b
c
z
y

Figure 4.

Figure 5.

Figure 6.

```
Q-V-Q-L-Q-E-S-G-P-G-L-V-K-P-S-E-T-L-S-L-T-C-T-V-S
G-G-S-V-S-S-G-D-Y-Y-W-T-W-I-R-Q-S-P-G-K-G-L-E-W-I
G-H-I-Y-Y-S-G-N-T-N-Y-N-P-S-L-K-S-R-L-T-I-S-I-D-T
S-K-T-Q-F-S-L-K-L-S-S-V-T-A-A-D-T-A-I-Y-Y-C-V-R-D
R-V-T-G-A-F-D-I-W-G-Q-G-T-M-V-T-V-S-S-A-S-T-K-G-P
S-V-F-P-L-A-P-C-S-R-S-T-S-E-S-T-A-A-L-G-C-L-V-K-D
Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P-A-V
L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-N-F-G-T-Q-T-Y
T-C-N-V-D-H-K-P-S-N-T-K-V-D-K-T-V-E-R-K-C-C-V-E-C
P-P-C-P-A-P-P-V-A-G-P-S-V-F-L-F-P-P-K-P-K-D-T-L-M
I-S-R-T-P-E-V-T-C-V-V-V-D-V-S-H-E-D-P-E-V-Q-F-N-W
Y-V-D-G-V-E-V-H-N-A-K-T-K-P-R-E-E-Q-F-N-S-T-F-R-V
V-S-V-L-T-V-V-H-Q-D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-G
L-P-A-P-I-E-K-T-I-S-K-T-K-G-Q-P-R-E-P-Q-V-Y-T-L-P
P-S-R-E-E-M-T-K-N-Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I
A-V-E-W-E-S-N-G-Q-P-E-N-N-Y-K-T-T-P-P-M-L-D-S-D-G
S-F-F-L-Y-S-K-L-T-V-D-K-S-R-W-Q-Q-G-N-V-F-S-C-S-V
M-H-E-A-L-H-N-H-Y-T-Q-K-S-L-S-L-S-P-G-K
```

```
D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-Q-A
S-Q-D-I-S-N-Y-L-N-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-D
A-S-N-L-E-T-G-V-P-S-R-F-S-G-S-G-S-G-T-D-F-T-F-T-I
S-S-L-Q-P-E-D-I-A-T-Y-F-C-Q-H-F-D-H-L-P-L-A-F-G-G
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L
K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V
D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L
S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G
L-S-S-P-V-T-K-S-F-N-R-G-E-C
```

Table 1.

| | Number of scans | Light chain | | | | Heavy chain | | | | Total sequence coverage |
		N-terminal product ions	C-terminal product ions	Unique cleavage sites	Sequence coverage	N-terminal product ions	C-terminal product ions	Unique cleavage sites	Sequence coverage	
high-field Orbitrap										
Adalimumab ETD	310	34	19	41	19.2%	39	78	117	26.0%	23.8%
Adalimumab EThcD	260	45	29	59	27.7%	22	58	80	17.8%	21.0%
Adalimumab combined	570	50	32	67	31.5%	53	82	135	30.0%	30.5%
Trastuzumab ETD	310	37	19	45	21.1%	55	75	130	29.0%	26.5%
Panitumumab ETD	300	34	9	34	15.9%	23	83	106	23.9%	21.3%
standard Orbitrap										
Adalimumab combined	1360	54	59	84	39.4%	60	79	119	26.4%	30.6%

Supporting Information.

Top-down analysis of immunoglobulins G with electron transfer dissociation high-field Orbitrap FTMS

Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, Xiowen Liu, Eugen Damoc, Pavel A. Pevzner, Alexander Makarov, Alain Beck, and Yury O. Tsybin

Figure S1. Fragmentation map obtained with MS-Deconv/MS-Align+ (right) and Xtract/ProSight PC (left) for ETD MS of Adalimumab performed on the LTQ Orbitrap Velos Pro FTMS (data set from Ref. [30], so-called "wide isolation window"). This data set represents the first test bed for assessing the capabilities of MS-Deconv/MS-Align+ for product ion assignment of TD MS of IgGs.

The combination of Xtract/ProSight PC used for the data analysis presented in Ref. [30] works efficiently, as shown in the left panel of the figure, with 60 and 112 unique cleavage sites for light and heavy chain respectively, identified by searching for c- and z-type ions. Results were achieved through a two-step search, with a first group of ions identified "automatically" under stringent mass tolerance settings (i.e., 20 ppm), and a second group manually validated after a 4.2 Da tolerance was set. The large tolerance allowed the assignment of product ions after hydrogen transfer (i.e., $z+1$ and, eventually, $c+1$ ions), or even ions with monoisotopic peak misassigned during deconvolution by Xtract (with a mass shift of 1.003 Da). Notably, automated assignment within the 20 ppm mass tolerance returned 27 and 59 unique cleavage sites for light and heavy chain, respectively, whereas the remaining product ions were assigned after manual validation-supported enlarging the mass tolerance to 4.2 Da.

The comparison of the results with MS-Deconv/MS-Align+ shows that this combination produces results similar to Xtract/ProSight PC for what concerns the heavy chain, whereas some discrepancies are noticed in regard to

the light chain, with lower sequence coverage obtained with MS-Deconv/MS-Align+. Importantly, MS-Deconv/MS-Align+ results are fully automated, and the manual validation is limited to those cleavage sites identified by a single ion and to the removal of eventual cleavage sites at the N-terminus of Pro. Furthermore, disulfide-free loops of both heavy and light chains show a highly similar sequence coverage with both software packages, although no manual ion assignment was needed in the case of MS-Deconv/MS-Align+. Note, other data sets acquired with Orbitrap Elite (such as some EThcD mass spectra) show slight advantage in favor of MS-Deconv/MS-Align+ even after the second search step in ProSight PC (data not shown), but most importantly they show good accordance between the cleavage sites identified by the two algorithms.

A possible explanation of the lower number of identified unique cleavage sites in fully-automatic search with Xtract/ProSight PC relative to MS-Deconv/MS-Align+ might derive from the differences in the product ion "envelope selection" procedure adopted by Xtract and MS-Deconv: the former selects iteratively the highest scoring envelopes whereas the latter uses a dedicated algorithm to determine a "heaviest path" in an envelope graph (comprising all the calculated theoretical envelopes) which finally results in optimized selection of product ion envelopes for highly convoluted MS/MS spectra with overlapping product ion clusters, like those produced by ETD/EThcD MS/MS of IgGs. Moreover, differently from MS-Deconv, Xtract combines close monoisotopic masses derived from distinct envelopes into a

unique monoisotopic mass, which might be limiting for highly convoluted MS/MS data of very large proteins.

As manual validation is a very time-consuming and operator-dependent procedure, after the tests MS-Deconv/MS-Align+ combination was chosen for the analysis of the data sets presented in this paper for a reason of the overall higher analysis speed and good reliability of identified ions, especially highly charged ones. We cannot exclude, nevertheless, that manual validation of data for finding additional product ions could lead to the discovery of additional cleavage sites, particularly those represented by product ions affected by hydrogen transfer, as efficiently demonstrated by the analysis with ProSight PC under large mass tolerance settings.

E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-R-S-L-R-L-S-C-A-A-S
G-F-T-F-D-D-Y-A-M-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-S-A
I-T-W-N-S-G-H-I-D-Y-A-D-S-V-E-G-R-F-T-I-S-R-D-N-A
K-N-S-L-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S
Y-L-S-T-A-S-S-L-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K
G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V
K-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P
A-V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q
T-Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D
K-T-H-T-C-P-P-C-P-A-P-E-L-L-G-G-P-S-V-F-L-F-P-P-K
P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-V-V-V-D-V-S-H-E-D-P
E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-A-K-T-K-P-R-E-E-Q-Y
N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-D-W-L-N-G-K-E-Y-K-C
K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-S-K-A-K-G-Q-P-R-E-P
Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-Q-V-S-L-T-C-L-V-K-G
F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-P-E-N-N-Y-K-T-T-P-P
V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-V-D-K-S-R-W-Q-Q-G-N
V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-T-Q-K-S-L-S-L-S-P-G
K

D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A
S-Q-G-I-R-N-Y-L-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-A
A-S-T-L-Q-S-G-V-P-S-R-F-S-G-S-G-S-G-T-D-F-T-L-T-I
S-S-L-Q-P-E-D-V-A-T-Y-Y-C-Q-R-Y-N-R-A-P-Y-T-F-G-Q
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L
K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V
D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L
S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G
L-S-S-P-V-T-K-S-F-N-R-G-E-C

Light: 60 unique cleavage sites (27 auto)
Heavy: 112 unique cleavage sites (59 auto)

E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-R-S-L-R-L-S-C-A-A-S
G-F-T-F-D-D-Y-A-M-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-S-A
I-T-W-N-S-G-H-I-D-Y-A-D-S-V-E-G-R-F-T-I-S-R-D-N-A
K-N-S-L-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S
Y-L-S-T-A-S-S-L-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K
G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V
K-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P
A-V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q
T-Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D
K-T-H-T-C-P-P-C-P-A-P-E-L-L-G-G-P-S-V-F-L-F-P-P-K
P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-V-V-V-D-V-S-H-E-D-P
E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-A-K-T-K-P-R-E-E-Q-Y
N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-D-W-L-N-G-K-E-Y-K-C
K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-S-K-A-K-G-Q-P-R-E-P
Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-Q-V-S-L-T-C-L-V-K-G
F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-P-E-N-N-Y-K-T-T-P-P
V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-V-D-K-S-R-W-Q-Q-G-N
V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-T-Q-K-S-L-S-L-S-P-G
K

D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A
S-Q-G-I-R-N-Y-L-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-A
A-S-T-L-Q-S-G-V-P-S-R-F-S-G-S-G-S-G-T-D-F-T-L-T-I
S-S-L-Q-P-E-D-V-A-T-Y-Y-C-Q-R-Y-N-R-A-P-Y-T-F-G-Q
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L
K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V
D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L
S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G
L-S-S-P-V-T-K-S-F-N-R-G-E-C

Light: 41 unique cleavage sites
Heavy: 101 unique cleavage sites

Figure S2. Mass spectra of intact IgGs. **A**, broadband spectrum of Adalimumab. Indicated in red are the maximum and minimum charge states included in the isolation window chosen for ETD MS/MS, 2400-3000 *m/z*. **B**, zoomed-in view of 56+ charge state of Adalimumab, showing baseline-resolved glycoforms. **C**-**D**, zoomed-in views of selected charge states of Trastuzumab and Panitumumab with indicated glycoforms. Mass spectra acquired with resolution 15'000 at *m/z* 400.
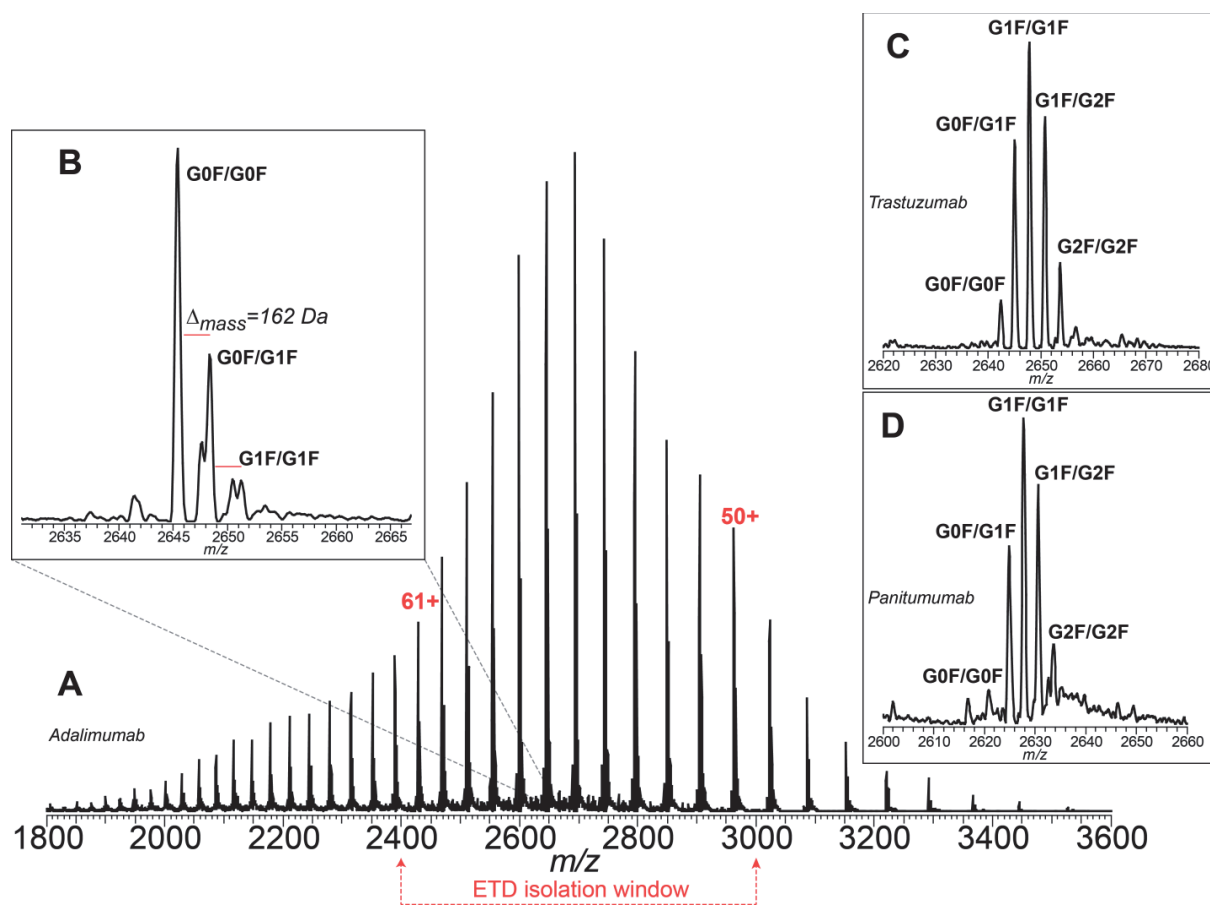
Figure S3. Evaluation of ETD MS/MS of 24+ precursor of horse myoglobin, 2 ms duration, at different HCD gas pressures. Each panel shows on top the results achieved with Δp=0.35E-10 torr ("high" pressure, or standard Δp condition), and on bottom with Δp=0.1E-10 torr ("low" pressure). **A**, broadband view. With high pressure the base peak in the spectrum is a product ion, whereas with low pressure remaining precursor and charge reduced species are dominating. Panels **B**-**D**: expanded views on selected ions. It is apparent how, at low m/z and for light, singly-charged ions, the signal-to-noise ratio (SNR) is similar between the two pressure conditions, and slightly higher when the pressure is high. Panel D shows that, on the contrary, heavier ions are preferentially transmitted when the pressure in the C-trap is lowered. All the displayed mass spectra are single scans, with each scan composed of 10 microscans. The MS/MS AGC was set to 1E6, similarly to the case of the IgG. Nevertheless, the IgG is 10 times heavier than myoglobin, and the largest identified IgG product ion exceeds in mass that of intact myoglobin. ETD of IgG produces many heavy, highly charged product ions which compete with low product ions for being transferred from LTQ to Orbitrap. In the case of myoglobin, using HCD trapping brings back the SNR of light ions to the levels normally achieved under high pressure condition (e.g., z_3^+ has SNR of ~9.5 with HCD trapping enabled).
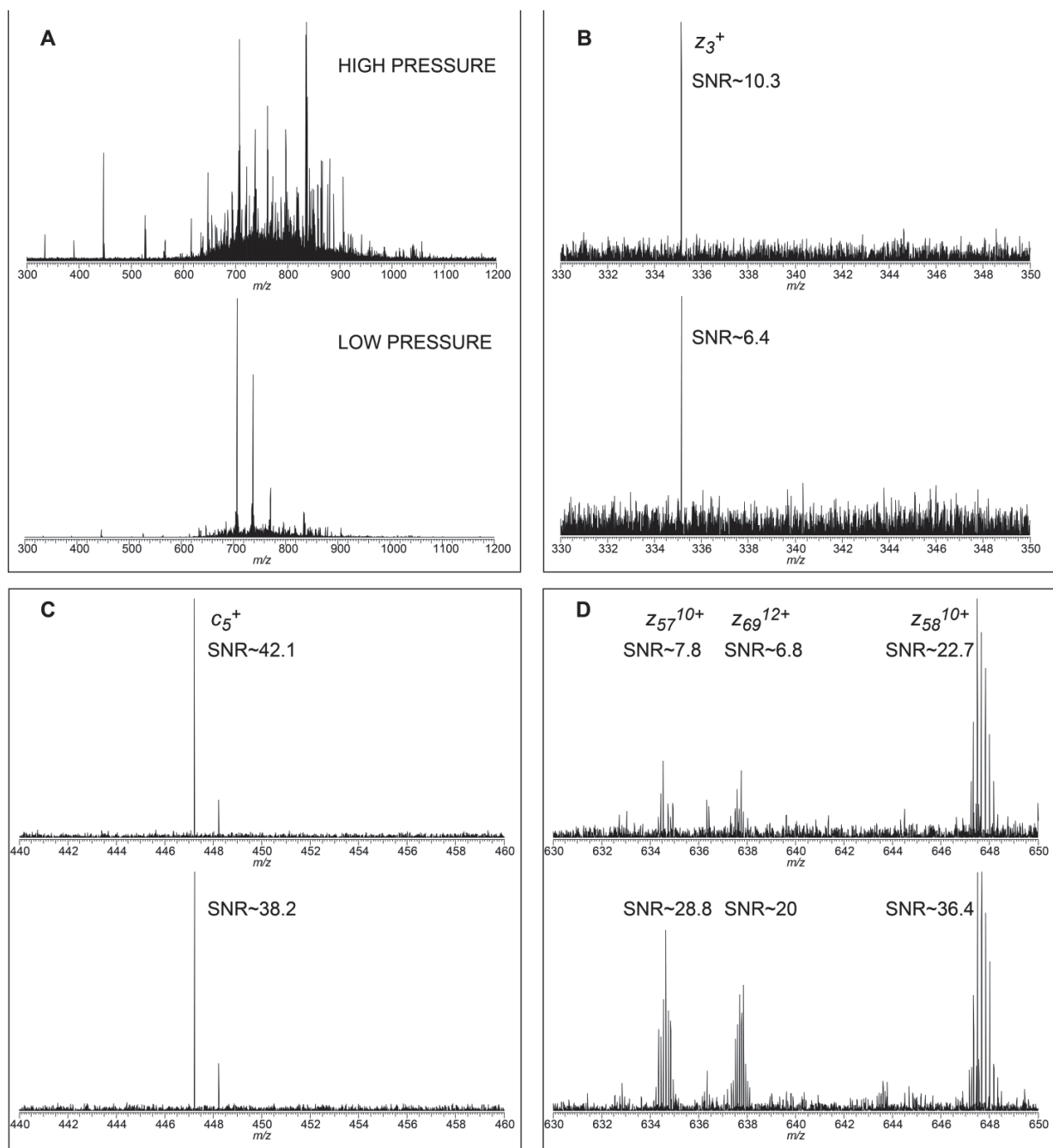
Figure S4. Fragmentation map of top-down ETD MS/MS of Trastuzumab

E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-G-S-L-R-L-S-C-A-A-S
G-F-N-I-K-D-T-Y-I-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-A-R
I-Y-P-T-N-G-Y-T-R-Y-A-D-S-V-K-G-R-F-T-I-S-A-D-T-S
K-N-T-A-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-S-R-W-G
G-D-G-F-Y-A-M-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K-G
P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V-K
D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P-A
V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q-T
Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D-K
T-H-T-C-P-P-C-P-A-P-E-L-L-G-G-P-S-V-F-L-F-P-P-K-P
K-D-T-L-M-I-S-R-T-P-E-V-T-C-V-V-V-D-V-S-H-E-D-P-E
V-K-F-N-W-Y-V-D-G-V-E-V-H-N-A-K-T-K-P-R-E-E-Q-Y-N
S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-D-W-L-N-G-K-E-Y-K-C-K
V-S-N-K-A-L-P-A-P-I-E-K-T-I-S-K-A-K-G-Q-P-R-E-P-Q
V-Y-T-L-P-P-S-R-E-E-M-T-K-N-Q-V-S-L-T-C-L-V-K-G-F
Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-P-E-N-N-Y-K-T-T-P-P-V
L-D-S-D-G-S-F-F-L-Y-S-K-L-T-V-D-K-S-R-W-Q-Q-G-N-V
F-S-C-S-V-M-H-E-A-L-H-N-H-Y-T-Q-K-S-L-S-L-S-P-G

D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A
S-Q-D-V-N-T-A-V-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-S
A-S-F-L-Y-S-G-V-P-S-R-F-S-G-S-R-S-G-T-D-F-T-L-T-I
S-S-L-Q-P-E-D-F-A-T-Y-Y-C-Q-Q-H-Y-T-T-P-P-T-F-G-Q
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L
K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V
D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L
S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G
L-S-S-P-V-T-K-S-F-N-R-G-E-C

Figure S5. Combination of ETD fragmentation maps of Adalimumab achieved in the present work and in the previous study based on the LTQ Orbitrap Velos Pro (i.e., standard Orbitrap). To the product ions assigned in the current study, signed in red, those derived from the old study were added (only where not already present), and signed in black. Note, that only c- and z-type ions were considered. Both light and heavy chain show primarily black product ions (i.e., from the Velos Pro data) located at the termini.

E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-R-S-L-R-L-S-C-A-A-S

G-F-T-F-D-D-Y-A-M-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-S-A

I-T-W-N-S-G-H-I-D-Y-A-D-S-V-E-G-R-F-T-I-S-R-D-N-A

K-N-S-L-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S

Y-L-S-T-A-S-S-L-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K

G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V

K-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P

A-V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q

T-Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D

K-T-H-T-C-P-P-C-P-A-P-E-L-L-G-G-P-S-V-F-L-F-P-P-K

P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-V-V-V-D-V-S-H-E-D-P

E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-A-K-T-K-P-R-E-E-Q-Y

N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-D-W-L-N-G-K-E-Y-K-C

K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-S-K-A-K-G-Q-P-R-E-P

Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-Q-V-S-L-T-C-L-V-K-G

F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-P-E-N-N-Y-K-T-T-P-P

V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-V-D-K-S-R-W-Q-Q-G-N

V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-T-Q-K-S-L-S-L-S-P-G

K

D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A

S-Q-G-I-R-N-Y-L-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-A

A-S-T-L-Q-S-G-V-P-S-R-F-S-G-S-G-S-G-T-D-F-T-L-T-I

S-S-L-Q-P-E-D-V-A-T-Y-Y-C-Q-R-Y-N-R-A-P-Y-T-F-G-Q

G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L

K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V

D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L

S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G

L-S-S-P-V-T-K-S-F-N-R-G-E-C

Table S1. Charge states of product ions identified at the C-terminus (z-ions) of Adalimumab and Trastuzumab for 10 and 25 ms ETD MS/MS. The reported values include: average charge state of all C-terminal product ions; average charge state of product ions belonging only to the disulfide-free loop connecting C_H2 and C_H3; highest charge state of an assigned product ion.

	Adalimumab 10 ms ETD	Trastuzumab 10 ms ETD	Trastuzumab 25 ms ETD
Average, C-term	6.2	6.2	5.3
Average, loop	6.9	6.9	5.7
Maximum	12	12	9

Paper VII

Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap FTMS

Luca Fornelli[1+], Daniel Ayoub[1+], Konstantin Aizikov[2], Alain Beck[3], and Yury O. Tsybin[1*]

[1] Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[2] Thermo Fisher Scientific GmbH, 28199 Bremen, Germany

[3] Centre d'Immunologie Pierre Fabre, 74160 St Julien-en-Genevois, France

[+] These authors contributed equally to this work

*Correspondence should be addressed to Prof. Yury O. Tsybin, EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland. E-mail: yury.tsybin@epfl.ch

Running title: Middle-down ETD Orbitrap™ FTMS of monoclonal antibodies

To be submitted to: *Analytical Chemistry*

Current manuscript date: 13 November 2013

(Abstract)

The rapid growth of approved biotherapeutics, e.g., monoclonal antibodies or immunoglobulins G (IgGs), demands improved techniques for their quality control. Traditionally, proteolysis-based bottom-up mass spectrometry (MS) has been employed. However, the long, multi-step sample preparation protocols required for bottom-up MS are known to potentially introduce artifacts in the original sample. For this reason, a top-down MS approach would be preferable. The current performance of top-down MS of intact monoclonal IgGs, though, enables reaching only up to ~30% sequence coverage, with incomplete sequencing of the complementarity determining regions which are fundamental for IgG's antigen binding. Here, we describe a middle-down MS protocol based on the use of immunoglobulin G-degrading enzyme of *Streptococcus pyogenes* (IdeS), which is capable of digesting IgGs in only 30 minutes. After chemical reduction, the obtained ~25 kDa proteolytic fragments were analyzed by reversed phase liquid chromatography (LC) coupled on-line with an electron transfer dissociation (ETD)-enabled hybrid Orbitrap Fourier transform mass spectrometer (Orbitrap Elite™ FTMS). Upon optimization of ETD and product ion transfer parameters, results show that up to ~50% sequence coverage for selected IgG fragments is reached in a single LC run, and up to ~70% when data obtained by distinct LC-MS runs are averaged. Importantly, we demonstrate the potential of this middle-down approach in the identification of oxidized methionine residues. The described approach shows a particular potential for the analysis of IgG mixtures.

Introduction

With the tremendous progress in the development of protein engineering technologies, the use of recombinant protein therapeutics has expanded significantly in recent years with monoclonal antibodies (mAbs) becoming the fastest growing class of human therapeutics.[1] Currently, there are more than 40 approved mAbs and mAbs derivative products and 30 others in advanced clinical investigations. MAbs are indicated for the treatment of a variety of diseases including cancer.[2] The success of mAbs as therapeutics is mainly attributed to their specificity to targets and their favorable pharmacokinetics. All approved therapeutic mAbs for clinical use belong to the immunoglobulin G (IgG) class. IgGs are tetrameric glycoproteins with molecular weights near 150 kDa. They consist of four polypeptide chains: two heavy chains (Hc) of ~50 kDa each and two light chains (Lc) of ~25 kDa each, linked together by disulfide bonds to form the characteristic Y-shaped complex.[3]

MAbs are required to be well characterized structurally to ensure their safety, efficiency, batch-to-batch consistency, and stability. With some approved therapeutic mAbs coming off patents, biosimilar antibodies are starting to be filed for approval. Biosimilar antibodies are "generic" versions of "innovator" (or "originator") antibodies produced through different manufacturing processes and from different clones.[4] For a biosimilar to be approved by the regulation agencies, it has to prove similar physicochemical characteristics, functional properties, and clinical efficiency to those of the innovator product.[5] One of the most important criteria required is for the

amino-acid sequence to be strictly identical.[6] Bottom–up liquid chromatography – tandem mass spectrometry (LC-MS/MS) sequencing of mAbs is widely used in pharmaceutical laboratories. Although bottom-up methods provide the most structural information, they suffer a number of drawbacks such as artifact introduction and lengthy sample preparation.[7,8] Top-down MS sequencing constitutes an interesting alternative and might become a method of choice as it is a fast and convenient way to obtain useful protein sequence information.[9-11] It has the advantage of implying limited sample preparation and therefore minimal artifact introduction due to sample processing. The advent of high resolution MS instruments coupled to electron transfer dissociation (ETD)[12] or electron capture dissociation (ECD)[13,14] enabled easier access to top-down analysis. In recent papers our team and others showed that top-down MS allows the sequencing of IgG terminal regions as well as variable domains and the characterization of major glycoforms.[15-17] Up to 30% sequence coverage was achieved using ETD or ECD fragmentation on Orbitrap[17] and ion cyclotron resonance (ICR)[16] Fourier transform (FT) instruments, respectively. However, the current state of the art does not allow overcoming this 30% limit primarily due to the gas phase retention of highly structured areas in the IgG, mainly in correspondence of the immunoglobulin-like domains and the disulfide bond protected areas.

By analogy to top-down and bottom-up approaches, the terms middle-up and middle-down were introduced.[18] Middle-up refers to the mass measurement of large fragments or subunits of a protein after limited

proteolysis while middle-down also includes the MS/MS-based sequencing of these subunits.[6,18] The terms top-down and middle-down are sometimes misused in the literature when referring to mass measurements of intact IgG or protease generated IgG subunits respectively, without performing MS/MS. Subunits of mAbs can be obtained through the chemical reduction of disulfide bonds thus yielding free heavy chains (50 kDa) and light chains (25 kDa), or by limited proteolysis in non-denaturing conditions with cleavage of the hinge region of the heavy chain yielding Fab (~50 kDa) or (Fab)'$_2$ (~100 kDa) and Fc (~50 kDa) fragments, **Figure 1.** Chemical reduction in denaturing conditions of these proteolysis-generated subunits results in three ~25 kDa fragments: the light chain and two half heavy chains (the Fc/2 and the Fd). These 25 kDa fragments better match the performance characteristics of the state-of-the-art LC-MS/MS methods and techniques compared to intact mAbs targeted with top-down approaches.[3,6] Several proteases such as papain,[19] Lys-C,[20] and pepsin[21] have been described for cleaving in the IgG hinge region under non-denaturing and controlled conditions. These proteases suffer however from limited specificity leading to nonspecific cleavages, complicating therefore the data analysis and interpretation.[22] Recently, IdeS (Immunoglobulin G-degrading enzyme of *Streptococcus pyogenes*) has been reported to specifically cleave between the two consecutive glycine residues under the hinge region.[23] It has the advantage of being rapid (30 min), low material consuming, and active in mAbs formulation buffers.[22] The reduced complexity of IdeS-generated ~25 kDa subunits allows an accurate profiling of N-glycans site by site and the

improved compared to intact IgG identification of various IgG microvariants (proteoforms) such as C-terminal lysine cleavage, cyclization of N-terminal glutamine and others.[6,24]

In this paper, we investigated the utility of ETD coupled with high resolution high-field Orbitrap FTMS[25] for the sequencing and characterization of approved therapeutic monoclonal antibodies in a middle-down fashion, Figure 1. The limited sample processing when using IdeS to generate medium-sized subunits allows minimizing artifact introduction. On the other hand, their relatively small size and the reduction of the disulfide bonds give access to the fragmentation of the S-S protected areas that are not efficiently sequenced by top-down MS. To benchmark the efficiency of the presented method in monitoring post-translational modifications, we chose the oxidation of labile amino-acid side chains as a case study. Oxidation is one of the major challenges for improving the stability profile in the development of monoclonal antibodies.[26] Oxidation of methionine residues from the sulfhydryl to the sulfoxide form is one of the common modifications known to occur in mAbs during the manufacturing, formulation, and/or storage process.[27] It could decrease bioactivity and stability of IgGs which result in reduced serum half-life and limited shelf life. Occurring mainly in Fc, it leads to decreased binding to the neonatal Fc receptor (FcRn) and loss of protection for catabolism.[28]

Hydrophobic interaction chromatography and ion exchange are the most commonly used LC techniques to detect the presence of oxidized mAbs proteoforms.[26,27] Bottom-up peptide mapping is then used to identify oxidation

sites. We employed mild oxidation conditions to generate oxidized IgG proteoforms that we analyzed by middle-down MS. Overall, the results reported here indicate that IdeS-based middle-down ETD MS can constitute a complementary or even an alternative method to bottom-up oxidation assessment in IgG quality control.

Materials and methods

Chemicals. Water, acetonitrile (ACN), trifluoroacetic acid (TFA) and isopropanol (IPA) were purchased in LC-MS purity grade. Water and ACN were obtained from Fluka Analytical (Buchs, Switzerland), formic acid (FA) from Merck (Zug, Switzerland), IPA from Thermo Fisher Scientific (Switzerland), and guanidinium chloride (GdnCl) from Carl Roth (Germany). *Tris*(2-carboxyethyl)phosphine, *tert*-butyl hydroperoxide, hydrogen peroxide and iron chloride were purchased from Sigma Aldrich (Buchs, Switzerland).

Chemical mild oxidation of IgGs. Therapeutic monoclonal antibodies of the IgG1 class, Adalimumab (Humira, Abbot Laboratories), Bevacizumab (Avastin, Genentech/Roche), and Trastuzumab (Herceptin, Genentech) were obtained as the European Medicines Agency approved versions and formulations, available commercially to the general public. To benchmark PTM identification by middle-down ETD MS, Adalimumab was oxidized prior to IdeS proteolysis. Two previously described protocols were employed.[26] The first one required the incubation of the IgG with 1 mM of hydrogen peroxide and 60 μM of $FeCl_3$ at

37° C for 19 hours. In the second, the IgG was treated with 1.4% of *tert*-butyl hydroperoxide for 19 hours at room temperature.

IdeS digestion. IdeS (FabRICATOR, Genovis, Lund, Sweden) digestion of IgGs was performed in formulation buffers. One unit of IdeS was added to each μg of IgG and left to react for 30 minutes at 37° C. Then, IgGs were denatured and reduced by incubation with 6 M GdnCl and 30 mM TCEP at room temperature for 30 minutes. Finally, the reaction was quenched by acidifying the solution to 1 % TFA. For analysis samples were diluted with 0.1% FA in water to a final concentration of 1μg/μl.

Liquid chromatography – mass spectrometry. The chromatographic separation of IgG proteolytic fragments was performed using an Ultimate 3000 LC system (Thermo Scientific, Amsterdam, The Netherlands) under UPLC conditions. A combination of reversed phase C4 trap-column (Acquity UPLC PrST C4 VanGuard pre-column, 2.1x5 mm, particle size 1.7 μm, pore size 300 Å, Waters, Baden-Dättwil, Switzerland) and C4 column (Acquity UPLC PrST C4, 1x150 mm, particle size 1.7 μm, pore size 300 Å, Waters) was employed to ensure on-line IgG fragment desalting and separation. For each injection, 1 μg of digestion product was loaded on the column, heated at 65° C. After initial loading at 5% solution B (organic phase), a gradient of solution B from 15 to 45% in 15 minutes was used at a flow rate of 100 μl/min. Solution A consisted of 0.1% of FA in water, whereas solution B was composed of 39.9% IPA, 60%

ACN, and 0.1% FA. The LC column outlet was on-line coupled with the electrospray (ESI) source of the mass spectrometer. MS experiments were performed on an ETD-enabled hybrid linear ion trap high-field Orbitrap FT mass spectrometer (LTQ Orbitrap Elite FTMS, Thermo Scientific, Bremen, Germany). Separate experiments were dedicated to record broadband mass spectra and ETD tandem mass spectra. Instrumental parameters were set as follows: S-lens RF level was set to 70%, the temperature of heated transfer capillary was 350° C, microESI source (IonMax source, Thermo Scientific) was used with a 3.7 kV potential, and sheath gas was set to 20 and auxiliary gas to 10 arbitrary units. All the mass spectra were acquired using ion detection in the Orbitrap FTMS, in the m/z range 200-2000. For broadband and tandem mass spectrometry, we both reduced the gas (N_2) "delta pressure" in the Orbitrap detector region to 0.1E-10 torr, and applied "HCD trapping", which is a temporary ion storage in the HCD cell before ion transmission to the Orbitrap mass analyzer through the C-trap.[29] Broadband mass spectra were recorded with either 15'000 or 120'000 resolution at 400 m/z, with a target value for the automatic gain control (AGC) of 1 million charges in either MS or MS/MS modes. For ETD experiments, precursor ions were isolated in the high pressure chamber of the LTQ and subsequently subjected to ETD MS/MS. The AGC target value for fluoranthene radical anions was set to 7-8E5 charges, with anion maximum injection time of 50 ms. ETD duration (i.e., ion-ion interaction time) was progressively increased from 3 ms to 9 ms in consecutive experiments. Product ion detection in the Orbitrap mass analyzer was

performed with 120'000 resolution at 400 *m/z* (eFT enabled). All Orbitrap FTMS scans were recorded averaging 10 microscans to improve the signal-to-noise ratio (SNR). Isolation windows for ETD of IgG fragments included one charge state per precursor ion (isolation width: 15 Th) in the case of Bevacizumab and Trastuzumab, or multiple charge states for Adalimumab (isolation width: 100 Th and wider).

Data processing and tandem MS analysis. Data were analyzed both as single LC-MS/MS runs, and after additional data processing aimed at improving SNR of tandem mass spectra. In the latter case, time-domain (transient) signals recorded in separate LC-MS/MS experiments were processed as previously described for top-down LC-MS/MS of mAbs.[17] Briefly, Orbitrap FTMS transient signals were first recorded in MIDAS .dat format;[30] these were then grouped according to the fragment type and duration of ETD MS/MS, averaged, and finally subjected to time-to-frequency conversion with the FT procedure. The resulting standard Thermo .RAW files could be then opened and processed with commercial XCalibur software (Thermo Scientific), and were thus ready for the data analysis. For each ETD duration, a summed mass spectrum for each IgG fragment was obtained; in addition, a total tandem mass spectrum was built by averaging all the transients (i.e., transients derived from different ETD duration experiments) available for a single IgG fragment.

Data analysis was performed using Xtract and ProSightPC 3.0 (Thermo Scientific).[31] First, Xtract was used for tandem mass spectra deconvolution,

peak centroiding, and peak picking. Then, cleavage sites were assigned with ProSightPC using 15 ppm tolerance. Both methionine and tryptophan were considered as possible oxidation sites. Complementary c- and z-type product ions were searched separately from b- and y-type product ions.

Results and discussion

LC-MS of IdeS-derived IgG fragments. Monoclonal antibodies digested with IdeS were subjected to chemical denaturation and disulfide bond reduction prior to LC-MS analysis. Notably, no alkylation of reduced thiols of cysteine residues was necessary, as the MS analysis was performed immediately after sample preparation and IgG fragments were maintained under acidic conditions (see Material and Methods), which helped preventing the re-formation of disulfide bridges. **Figure 2** shows a typical total ion current (TIC) chromatogram of the three proteolytic fragments, namely Fc/2, Fd, and Lc, of Adalimumab. These are baseline separated and, as displayed in the figure insets, present over the selected m/z window (i.e., m/z 200-2000) extended charge state envelopes centered around charge states 24+-26+, with maximum charge state exceeding 30+. A similar LC separation has been achieved for fragments derived from Bevacizumab and Trastuzumab, although, as expected, retention times of Lc and Fd fragments, containing IgG variable regions, are different among the three antibodies (Supporting Information, Figure S1). Importantly, the applied chromatographic conditions, with the UPLC C4 column heated at 65° C, ensured high elution reproducibility (data not shown).

The three fragments have masses in the range of 23-25 kDa. In the case of Adalimumab, for instance, theoretical average masses (neutral) of non-modified fragments are the following: Fc/2, 23887.08 Da; Lc, 23412.17 Da; and Fd, 25458.63 Da. We observed and determined by MS[1] two classical modifications, well described in literature, located on the Fc/2 fragment (of all the three antibodies), which is N-glycosylated and presents often C-terminal lysine clipping. Major glycosylations could be detected: in the case of Adalimumab, the most intense peak corresponds to Fc/2 with G0F glycosylation, followed by G1F with an intensity of about one third of the base peak.[32] The mass accuracy achieved with the adopted MS settings (see Material and Methods) at 15'000 resolution (at m/z 400) was ~1 Da. With regard to mass accuracy, on the LC time scale it was also possible to fully resolve the isotopic distribution of the fragments using a resolution setting of 120'000 (at m/z 400), as shown in Figure 2 for Fc/2 fragment with G0F glycosylation of Adalimumab. Nevertheless, as a general consideration about the current state of available instrumentation, it is fundamental to recall that the advantage of isotopically resolved distributions of polypeptides relies mainly on the determination of the monoisotopic mass. According to calculations based on model proteins, the dynamic range needed for simultaneously detecting the most intense peak in the isotopic distribution and the monoisotopic one, is of ~E32 for an intact IgG and ~6E4 for the ~29 kDa carbonic anhydrase. Therefore, considering the molecular weight of the IgG fragments, close to that of carbonic anhydrase, the utility of isotopically resolving the here studied

fragments is currently limited given that no current FTMS mass analyzer exceeds four orders of magnitude in spectral dynamic range. In addition, further complications are due to the restricted number of mass spectra recorded during a single LC experiment, the possible co-elution of protein adducts (e.g., Na^+ adducts), and presence of proteoforms with small differences in mass. As an example, Chen *et al.* showed that the commonly employed method for monoisotopic mass estimation based on averagine (i.e., average amino acid with the formula $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$)[33] can lead to an error of ~0.3 Da for isotopically resolved 15 kDa protein RNase A.[34]

ETD MS/MS of IgG fragments. ETD of IdeS-produced IgG subunits was performed in two different fashions. Proteolytic subunits derived from Bevacizumab and Trastuzumab were subjected to ETD in a "proteomic" fashion, which corresponds to the isolation of precursor ions in a single charge state. These experiments proved the efficiency of ETD and facilitated tuning the instrumental parameters for further studies described below.

Conversely, ETD MS/MS of Adalimumab was aimed at maximizing the sequence coverage, for direct comparison with top-down MS results, and was thus performed exploiting the high chromatographic reproducibility of subunit elution. The instrument was operating in MS/MS mode only during selected time windows corresponding to the elution times of the three fragments. In this case, a larger isolation window was used, for including ~5 highly charged precursors and increasing ETD efficiency. Isolation windows were selected on

the base of the charge state envelopes of each fragment (i.e., using the information previously obtained by MS[1]), and centered as indicated by the arrows in the bottom left inset of Figure 2. Note, in top-down and middle-down experiments, the sequence coverage is calculated as the ratio of assigned cleavage sites to the total number of possible backbone cleavage sites, whereas in bottom-up experiments, the identified peptides are used without regard to the number of cleavage sites assigned after evaluation of MS/MS data.

As expected, the highest sequence coverage was obtained with the acquisition parameters previously optimized for Adalimumab analysis. The corresponding fragmentation map is represented in **Figure 3**. As reported in the Figure, a final coverage of almost 70% was obtained for Fc/2 and Lc fragments, whereas a slightly lower coverage, of nearly 60%, was obtained for the heavier Fd fragment, presumably because of different charge localization and lower charge-over-mass unit ratio relative to Fc/2. The final sequence coverage was calculated accounting all the cleavage sites assigned through the analysis of different experimental datasets, derived from experiments where the ETD duration spanned from 3 to 9 ms. For each dataset, 4 to 10 LC runs were acquired. In each LC run, up to 90 transients (microscans) per fragment were recorded (the number varies among the fragments, being dependent on the elution time frame). In total, ~3000 transients (microscans) were averaged. Importantly, all the complementarity determining regions (CDRs) were sequenced, and the position of the glycosylation site on the Fc/2 was confirmed, as well as the Lys-clipping on the same fragment, Figure 3.

For Bevacizumab and Trastuzumab, the averaging of a similar number of transients resulted in slightly lower sequence coverage (Table S1, Supporting Information). Nevertheless, sequencing of CDRs and confirmation of the glycosylation site as well as Lys-clipping were obtained (Figure S2, Supporting Information).

Finally, it is noteworthy that the sequence coverage achieved with a single LC run (related ETD mass spectra are reported in Figure S3, Supporting Information) is considerably high for ~25 kDa polypeptides with sharp elution peaks (of 20-40 s): almost 50% sequence coverage is reached for Fc/2 and Lc and ~30% for Fd of Adalimumab (Table S2, Supporting Information). Importantly, the advantage in terms of sequence coverage obtainable operating the mass spectrometer under reduced HCD cell pressure and with HCD trapping enabled is dramatic for a single ETD LC-MS/MS experiment, as summarized in Table S2 and visually reported in Figure S4, Supporting Information.

Middle-down LC-MS/MS analysis of oxidized IgG. The chromatographic separation of fragments derived from IdeS digestion of oxidized Adalimumab resulted in baseline-resolved elution peaks. **Figure 4** illustrates a comparison between fragments of IgG oxidized with *tert*-buthyl hydroperoxide (red line) and those of the "native" IgG (black line). It is apparent how only the Fc/2 shows a different chromatographic behavior in the two cases, Figure 4 top panel. Effectively, whereas Lc and Fd where not oxidized (as determined by MS[1]

analysis), Fc/2 presented one or two oxidations, with the doubly oxidized species being about four fold more abundant than the mono-oxidized species, Figure 4 middle panel. Note that in the control experiment, no oxidation was detected which proves that no artifact oxidation was induced by the IdeS sample preparation (data not shown). By a single 3 ms ETD LC-MS/MS run it was possible to determine the sites of oxidation, Figure 4 bottom panel. These are located at two methionine residues, namely Met_{16} and Met_{192}, as schematically indicated on the right inset in Figure 4 bottom panel and further detailed in Figure S5, top panel, Supporting Information. A more extended sequence coverage was then obtained by averaging 300 transients (microscans) deriving from ETD experiments with duration of 3 ms and 300 other transients (microscans) from 5 ms ETD experiments, Figure S5, bottom panel, Supporting Information. These results can be compared to the higher-energy collision dissociation (HCD)[35] MS/MS results reported in the literature for similar ~25 kDa antibody subunits.[36] In this study, the two oxidized forms of the Fc/2 were also detected by MS[1]. However, HCD MS/MS yielded **8 %** sequence coverage of the Fc/2 subunits with only one distant *b* fragment (b_{48}) assigning the oxidation at Met_{30} and nine distant *y* fragments assigning oxidation at Met_{206}. Our results show that ETD fragmentation increased significantly sequence coverage to more than 30% in one run with 3 ms ETD to 47.1% when averaging ~10 runs of 3 and 5 ms ETD. Note that these percentages are influenced by the contemporary analysis of the two forms of oxidized Fc/2. Cleavage sites adjacent to the oxidized methionines and other potential oxidation sites

(tryptophanes) are obtained giving more confident assignment to the oxidation sites.

Middle-down analysis of IgG mixtures. As displayed in **Figure 5**, we applied the above described LC-MS method for the analysis of an equimolar mixture of monoclonal IgGs digested with IdeS. Adalimumab, Bevacizumab, and Trastuzumab were first pooled together and subjected to IdeS digestion, reduction and subsequent LC-MS analysis (with 15'000 resolution at *m/z* 400). This validation experiment was aimed at verifying the actual possibilities offered by reversed phase UPLC for the future analysis of either IgG mixtures of biological origin (e.g., polyclonal antibodies) or so-called therapeutic "IgG cocktails". While Fc/2 fragments share a very high sequence homology (over 90%) and are therefore not separated, most of the other fragments can be isolated. Two fragments, Lc of Bevacizumab and Fd of Trastuzumab are co-eluting, but this does not impair the recording of high SNR broadband mass spectra for the two (Figure 5, left inset). Single charge state precursor ion isolation and sequencing with ETD MS/MS should thus be possible as indicated by the data presented here.

Conclusions.

To improve structural analysis of monoclonal antibodies, IgGs, we introduce a middle-down mass spectrometry approach consisting in quick (about 1 hour) IgG sample preparation protocol followed by 25 kDa polypeptide analysis with

reversed phase LC-based protein separation on-line coupled to high-resolution Orbitrap FTMS with ETD capability. The reported results show that the application of the described middle-down approach doubles the IgG sequence coverage previously obtained with top-down MS. The particular advantage of the approach described here consists in the limited sample preparation reducing potential artifacts introduction. The described method proved to be efficient in detecting and assigning IgG oxidation sites, which makes it a quick and easily implemented technique for mAbs' oxidation assessment. Finally, we demonstrate that the 25 kDa molecular weight range of the IgG fragments under study should enable an efficient structural analysis of simple, up to three IgGs considered here, antibody mixtures. We envision application of the developed approach for structural analysis of more complex mAbs mixtures, as required by modern drug discovery strategies.

References

(1) Walsh, G. *Nat Biotechnol* **2010**, *28*, 917-924.

(2) Debaene, F.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. *Anal Chem* **2013**, *85*, 9785-9792.

(3) Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal Chem* **2013**, *85*, 715-736.

(4) Beck, A.; Diemer, H.; Ayoub, D.; Debaene, F.; Wagner-Rousset, E.; Carapito, C.; Van Dorsselaer, A.; Sanglier-Cianférani, S. *TrAC Trends in Analytical Chemistry* **2013**, *48*, 81-95.

(5) Weise, M.; Bielsky, M. C.; De Smet, K.; Ehmann, F.; Ekman, N.; Giezen, T. J.; Gravanis, I.; Heim, H. K.; Heinonen, E.; Ho, K.; Moreau, A.; Narayanan, G.; Kruse, N. A.; Reichmann, G.; Thorpe, R.; van Aerts, L.; Vleminckx, C.; Wadhwa, M.; Schneider, C. K. *Blood* **2012**, *120*, 5111-5117.

(6) Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner, E.; Suckau, D.; Beck, A. *MAbs* **2013**, *5*, 699-710.

(7) Ren, D.; Pipes, G. D.; Liu, D. J.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. *Anal Biochem* **2009**, *392*, 12-21.

(8) Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. *Anal Chem* **2006**, *78*, 6645-6650.

(9) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O. *Analytical and bioanalytical chemistry* **2013**, *405*, 8505-8514.

(10) Kellie, J. F.; Tran, J. C.; Lee, J. E.; Ahlf, D. R.; Thomas, H. M.; Ntai, I.; Catherman, A. D.; Durbin, K. R.; Zamdborg, L.; Vellaichamy, A.; Thomas, P. M.; Kelleher, N. L. *Mol Biosyst* **2010**, *6*, 1532-1539.

(11) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M. X.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254-U141.

(12) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *P Natl Acad Sci USA* **2004**, *101*, 9528-9533.

(13) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J Am Chem Soc* **1998**, *120*, 3265-3266.

(14) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal Chem* **2000**, *72*, 563-573.

(15) Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. *Anal Chem* **2011**, *83*, 8919-8927.

(16) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal Chem* **2013**, *85*, 4239-4246.

(17) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. *Molecular & cellular proteomics* **2012**, *11*, 1758-1767.

(18) Zhang, Z.; Pan, H.; Chen, X. *Mass Spectrom Rev* **2009**, *28*, 147-176.

(19) Yan, B.; Valliere-Douglass, J.; Brady, L.; Steen, S.; Han, M.; Pace, D.; Elliott, S.; Yates, Z.; Han, Y.; Balland, A.; Wang, W.; Pettit, D. *J Chromatogr A* **2007**, *1164*, 153-161.

(20) Gadgil, H. S.; Bondarenko, P. V.; Pipes, G. D.; Dillon, T. M.; Banks, D.; Abel, J.; Kleemann, G. R.; Treuheit, M. J. *Anal Biochem* **2006**, *355*, 165-174.

(21) Gadgil, H. S.; Bondarenko, P. V.; Pipes, G.; Rehder, D.; McAuley, A.; Perico, N.; Dillon, T.; Ricci, M.; Treuheit, M. *J Pharm Sci* **2007**, *96*, 2607-2621.

(22) Chevreux, G.; Tilly, N.; Bihoreau, N. *Anal Biochem* **2011**, *415*, 212-214.

(23) Ryan, M. H.; Petrone, D.; Nemeth, J. F.; Barnathan, E.; Bjorck, L.; Jordan, R. E. *Mol Immunol* **2008**, *45*, 1837-1846.

(24) Wang, B.; Gucinski, A. C.; Keire, D. A.; Buhse, L. F.; Boyne, M. T. *Analyst* **2013**, *138*, 3058-3065.

(25) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Muller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J. J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. *Molecular & cellular proteomics* **2012**, *11*, O111 013698.

(26) Boyd, D.; Kaschak, T.; Yan, B. *J Chromatogr B Analyt Technol Biomed Life Sci* **2011**, *879*, 955-960.

(27) Teshima, G.; Li, M. X.; Danishmand, R.; Obi, C.; To, R.; Huang, C.; Kung, J.; Lahidji, V.; Freeberg, J.; Thorner, L.; Tomic, M. *J Chromatogr A* **2011**, *1218*, 2091-2097.

(28) Kuo, T. T.; Aveson, V. G. *MAbs* **2011**, *3*, 422-430.

(29) Rosati, S.; Rose, R. J.; Thompson, N. J.; van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. *Angewandte Chemie* **2012**, *51*, 12992-12996.

(30) Senko, M. W.; Canterbury, J. D.; Guan, S.; Marshall, A. G. *Rapid Commun Mass Spectrom* **1996**, *10*, 1839-1844.

(31) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. *Nucleic Acids Res* **2007**, *35*, W701-W706.

(32) Nallet, S.; Fornelli, L.; Schmitt, S.; Parra, J.; Baldi, L.; Tsybin, Y. O.; Wurm, F. M. *New Biotechnology* **2012**, *29*, 471-476.

(33) Senko, M. W.; Beu, S. C.; Mclafferty, F. W. *J Am Soc Mass Spectr* **1995**, *6*, 229-233.

(34) Chen, Y. F.; Chang, C. A.; Lin, Y. H.; Tsay, Y. G. *Anal Biochem* **2013**, *440*, 108-113.

(35) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nature methods* **2007**, *4*, 709-712.

(36) Zhang, J.; Liu, H.; Katta, V. *J Mass Spectrom* **2010**, *45*, 112-120.

Figure caption

Figure 1. Schematics of the proposed middle-down workflow. Sample preparation lasts one hour, and it is followed by LC-MS/MS and ProSight PC analysis. The presented scheme can be applied to single IgGs or mixtures.

Figure 2. Total ion current chromatogram of IdeS-digested Adalimumab. The three subunits (from left to right: Fc/2, Lc, and Fd) are baseline resolved. Broadband Orbitrap FTMS mass spectra recorded with 15'000 resolution at *m/z* 400 show the charge state distributions of the three subunits, insets. Arrows indicate the center of the employed precursor ion isolation window for ETD MS/MS. The bottom left inset displays a baseline resolved single charge state of the Fc/2 subunit with G0F glycosylation, recorded by LC-MS with 120'000 resolution at *m/z* 400.

Figure 3. Fragmentation maps of IdeS digested Adalimumab after spectral SNR improvement by averaging of transients from different LC runs. Product ions assigned from different ETD datasets are represented according to the color legend. Both *c*- and *z*-type as well as *y*-type product ions were assigned. CDRs are highlighted in yellow, whereas the N-glycosylation site on the Fc/2 fragment is indicated in green. Right bottom inset shows summary of sequence coverage for the corresponding fragments of Adalimumab.

Figure 4. Middle-down MS-based analysis of IgG oxidation. Top panel: comparison between TIC chromatograms of control (black line) and oxidized (red line) Adalimumab. The oxidized Fc is eluting earlier than the non-modified counterpart. Middle panel: expanded view of broadband mass spectra showing the isotopically resolved Fc/2 subunit (mass spectra recorded with 120'000 resolution at *m/z* 400). This panel, made by overlapping mass spectra of "control" and "oxidized" IgGs, illustrates the presence of both singly and doubly oxidized Fc/2 subunits. Bottom panel: 3 ms ETD Orbitrap FTMS mass spectrum (single LC run) of oxidized Fc (isolation window 120 Th). The left inset shows a product ion confirming the position of oxidized Met_{16}. The right inset is a scheme of identified product ions, with the positions of oxidized Met indicated by arrows.

Figure 5. TIC chromatogram of an equimolar mixture of IdeS cleaved monoclonal antibodies Adalimumab, Bevacizumab, and Trastuzumab. The inset shows the Orbitrap FTMS mass spectrum of co-eluting Lc of Bevacizumab and Fd of Trastuzumab.
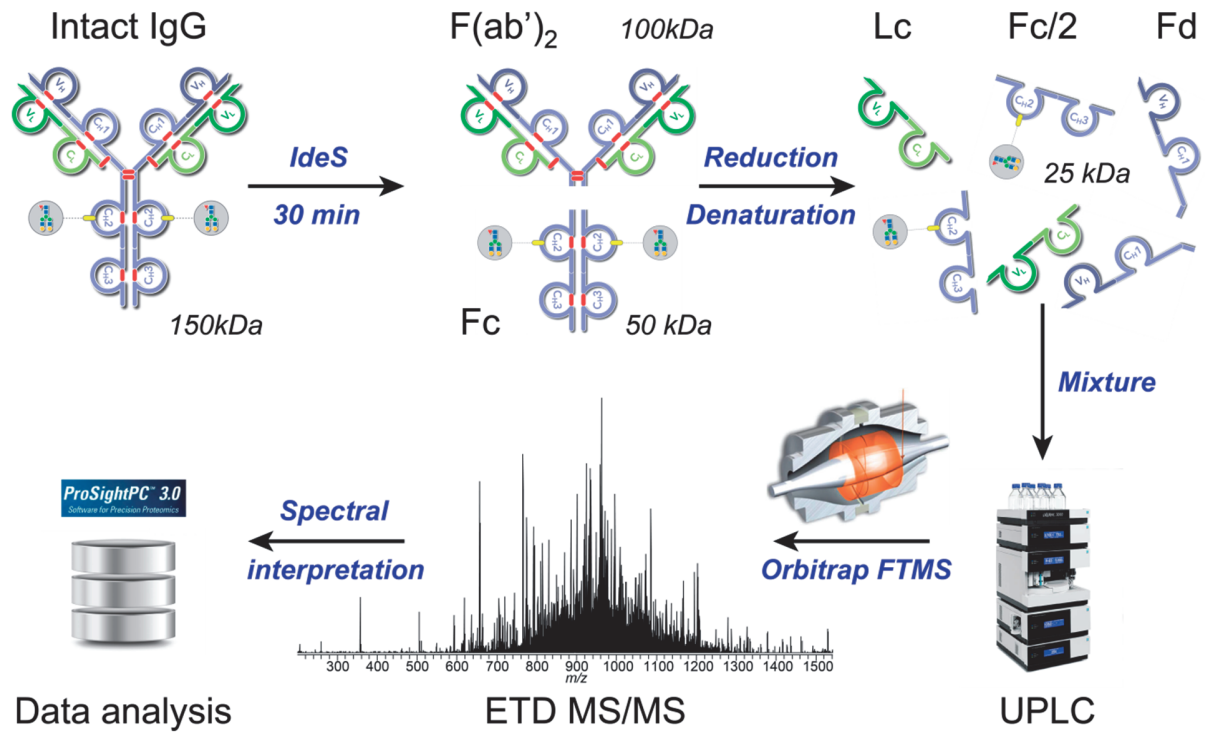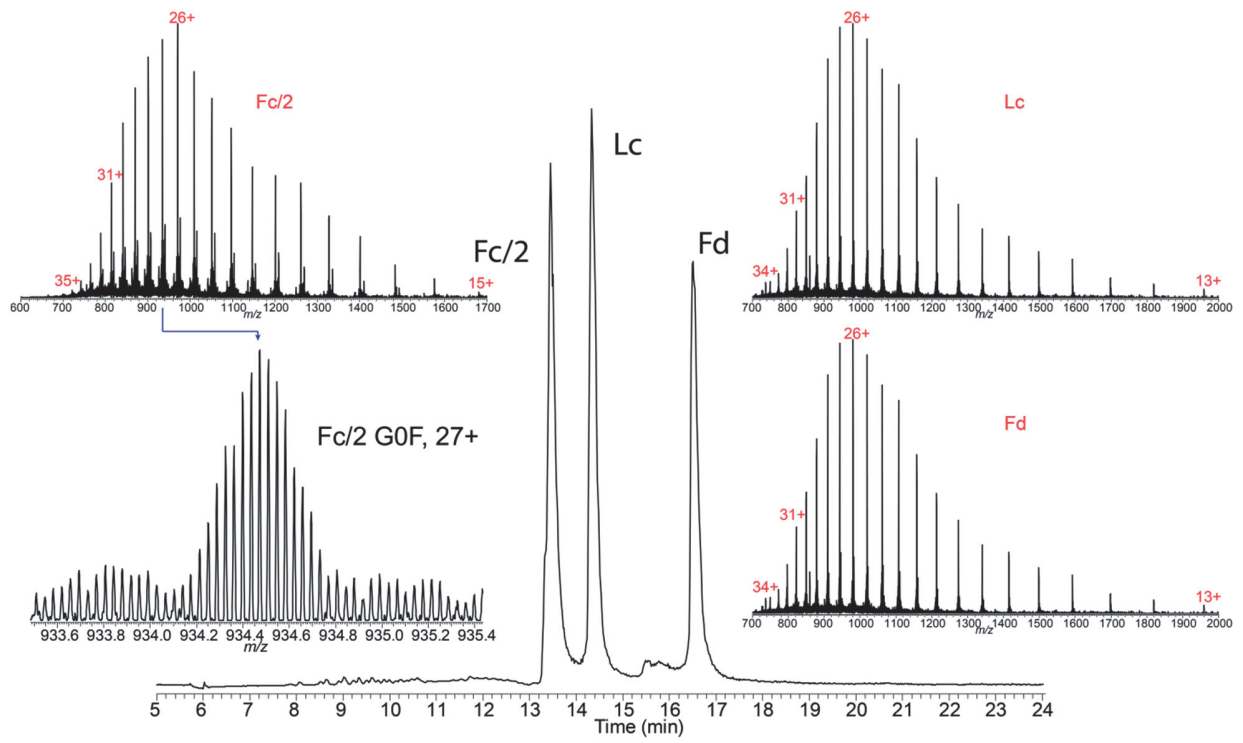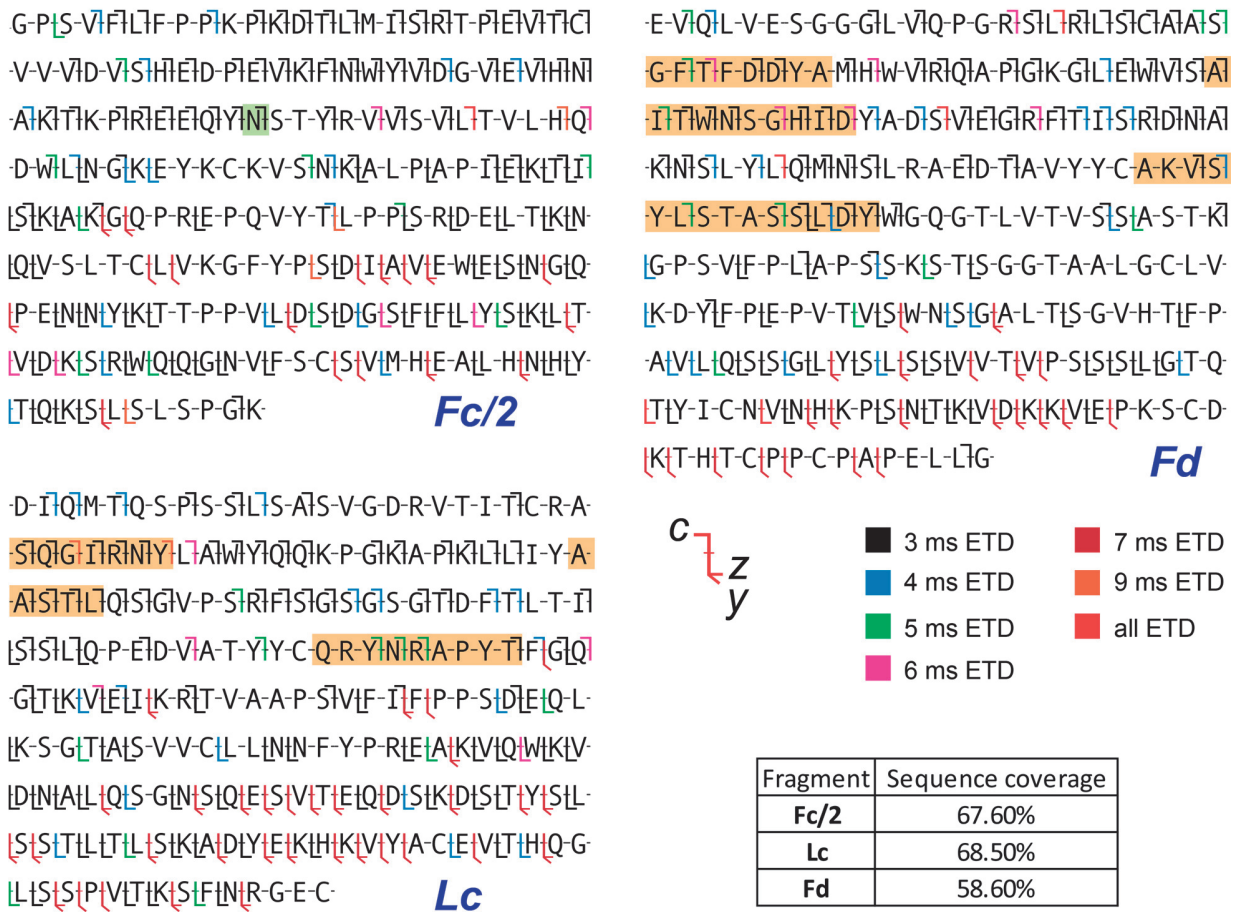
Figure 1.

Figure 2.

Figure 3.

G-P-S-V-F-L-F-P-P-K-P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-
V-V-V-D-V-S-H-E-D-P-E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-
A-K-T-K-P-R-E-E-Q-Y-N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-
D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-
S-K-A-K-G-Q-P-R-E-P-Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-
Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-
P-E-N-N-Y-K-T-T-P-P-V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-
V-D-K-S-R-W-Q-Q-G-N-V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-
T-Q-K-S-L-S-L-S-P-G-K-   ***Fc/2***

D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A-
S-Q-G-I-R-N-Y-L-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-A-
A-S-T-L-Q-S-G-V-P-S-R-F-S-G-S-G-S-G-T-D-F-T-L-T-I-
S-S-L-Q-P-E-D-V-A-T-Y-Y-C-Q-R-Y-N-R-A-P-Y-T-F-G-Q-
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L-
K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V-
D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L-
S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G-
L-S-S-P-V-T-K-S-F-N-R-G-E-C-   ***Lc***

E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-R-S-L-R-L-S-C-A-A-S-
G-F-T-F-D-D-Y-A-M-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-S-A-
I-T-W-N-S-G-H-I-D-Y-A-D-S-V-E-G-R-F-T-I-S-R-D-N-A-
K-N-S-L-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S-
Y-L-S-T-A-S-S-L-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K-
G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V-
K-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P-
A-V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q-
T-Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D-
K-T-H-T-C-P-P-C-P-A-P-E-L-L-G-   ***Fd***

c ⌐ z
     y

■ 3 ms ETD	■ 7 ms ETD
■ 4 ms ETD	■ 9 ms ETD
■ 5 ms ETD	■ all ETD
■ 6 ms ETD	

Fragment	Sequence coverage
Fc/2	67.60%
Lc	68.50%
Fd	58.60%

Figure 4.

Figure 5

1. Lc Adalimumab
2. Lc Trastuzumab
3. Lc Bevacizumab/
   Fd Trastuzumab
4. Fd Bevacizumab
5. Fd Adalimumab

Fc/2

° Lc, Bevacizumab
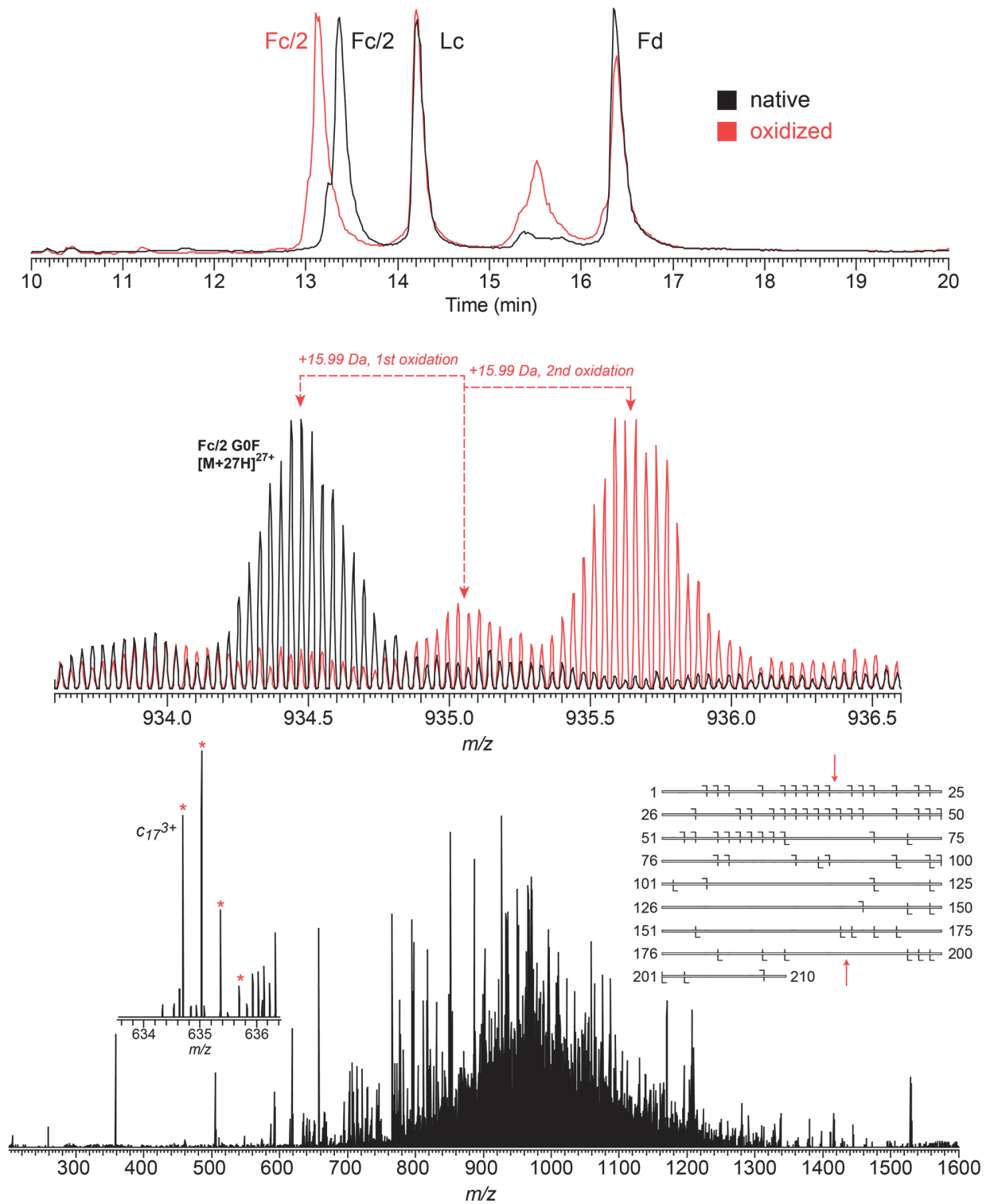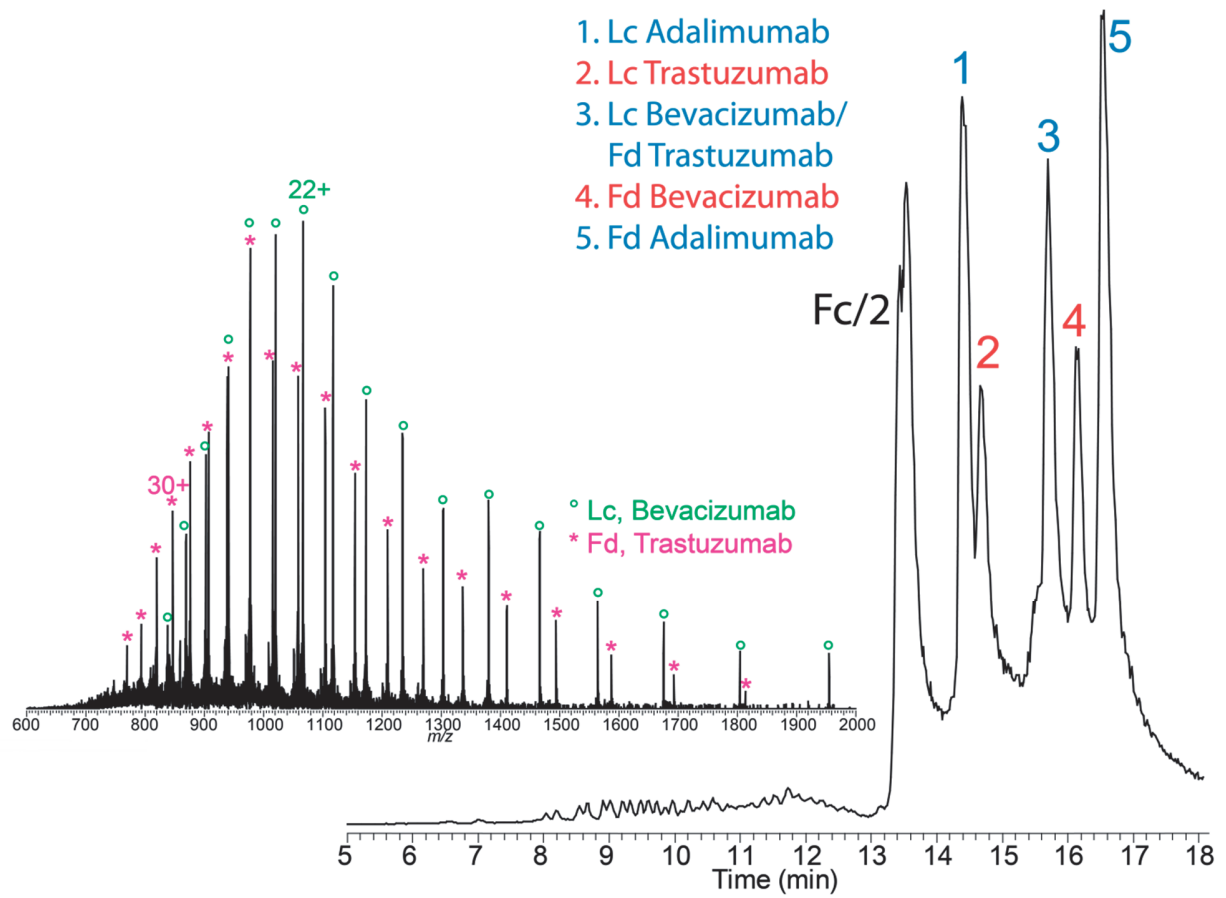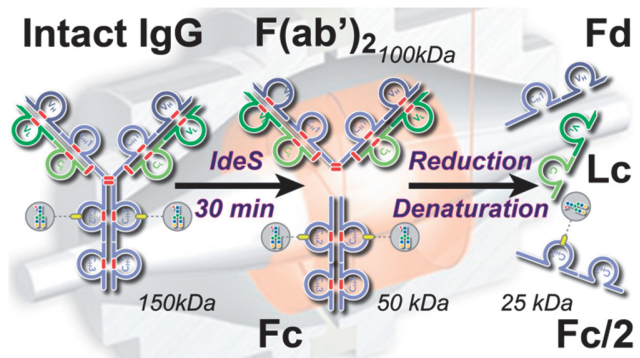* Fd, Trastuzumab

22+

30+

m/z

Time (min)

For TOC only

Supporting information.

Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap FTMS

Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, Alain Beck, and Yury O. Tsybin

Table S1. Sequence coverage for IdeS-derived fragments of Trastuzumab and Bevacizumab after spectral SNR improvement by averaging of transients from different LC runs.

Fragment	Trastuzumab	Bevacizumab
Fc/2	59.80%	52.86%
Lc	58.21%	58.21%
Fd	44.53%	45.64%

Table S2. Comparison between sequence coverage of IdeS-obtained fragments of Adalimumab reached in a single LC run for high and low HCD cell pressure difference Δp.

Fragment	High Δp	Low Δp
Fc/2	30.47%	49.04%
Lc	28.63%	49.29%
Fd	15.89%	29.28%

Figure S1. TIC chromatograms of IdeS-derived fragments of monoclonal IgGs Bevacizumab (top panel) and Trastuzumab (bottom panel). Fragments are generally baseline separated, with the exception of the Lc and Fd of Bevacizumab, which have close retention times even under UPLC conditions with the applied gradient. Nevertheless, the identification of the single fragments by MS1 and the isolation of selected precursor for tandem MS is not affected.
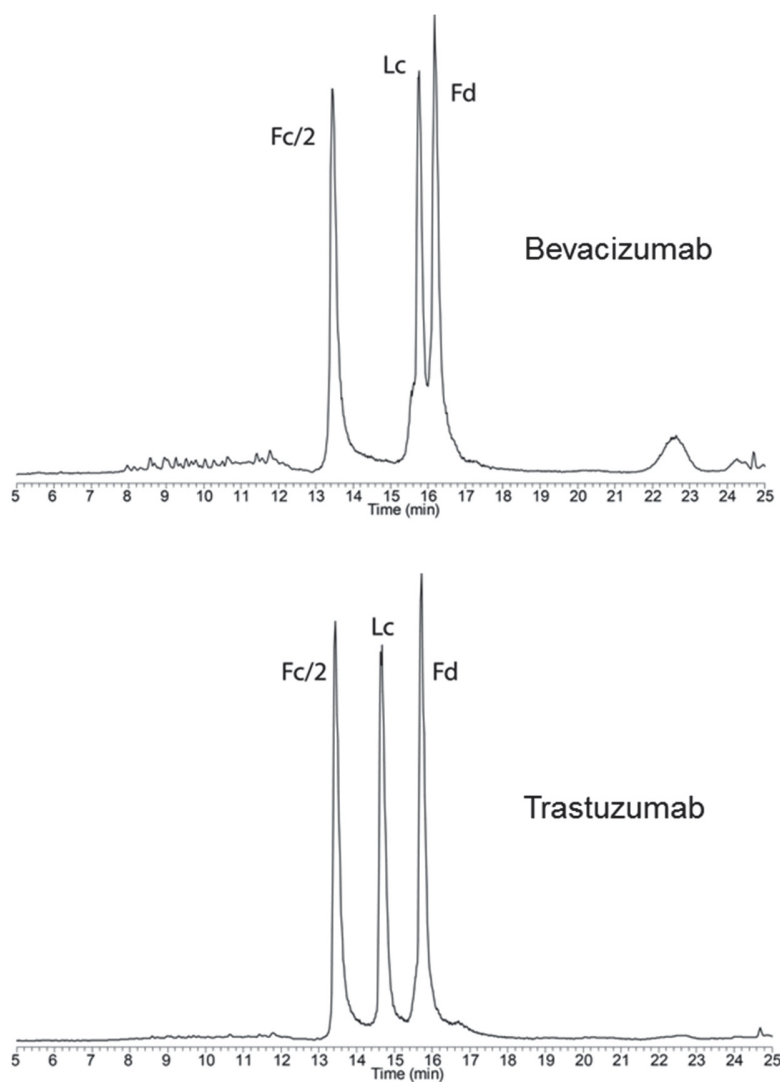
Figure S2. Fragmentation map of Trastuzumab. Color legend is used for distinguishing cleavage sites identified from a specific ETD duration. CDRs are highlighted in yellow. The results for Bevacizumab are comparable.

-G-P-S-V-F-L-F-P-P-K-P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-
-V-V-V-D-V-S-H-E-D-P-E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-
-A-K-T-K-P-R-E-E-Q-Y-N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-
-D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-
-S-K-A-K-G-Q-P-R-E-P-Q-V-Y-T-L-P-P-S-R-E-E-M-T-K-N-
-Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-
-P-E-N-N-Y-K-T-T-P-P-V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-
-V-D-K-S-R-W-Q-Q-G-N-V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-
-T-Q-K-S-L-S-L-S-P-G- *Fc/2*

-D-I-Q-M-T-Q-S-P-S-S-L-S-A-S-V-G-D-R-V-T-I-T-C-R-A-
-S-Q-D-V-N-T-A-V-A-W-Y-Q-Q-K-P-G-K-A-P-K-L-L-I-Y-S-
-A-S-F-L-Y-S-G-V-P-S-R-F-S-G-S-R-S-G-T-D-F-T-L-T-I-
-S-S-L-Q-P-E-D-F-A-T-Y-Y-C-Q-Q-H-Y-T-T-P-P-T-F-G-Q-
-G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I-F-P-P-S-D-E-Q-L-
-K-S-G-T-A-S-V-V-C-L-L-N-N-F-Y-P-R-E-A-K-V-Q-W-K-V-
-D-N-A-L-Q-S-G-N-S-Q-E-S-V-T-E-Q-D-S-K-D-S-T-Y-S-L-
-S-S-T-L-T-L-S-K-A-D-Y-E-K-H-K-V-Y-A-C-E-V-T-H-Q-G-
-L-S-S-P-V-T-K-S-F-N-R-G-E-C- *Lc*

-E-V-Q-L-V-E-S-G-G-G-L-V-Q-P-G-G-S-L-R-L-S-C-A-A-S-
-G-F-N-I-K-D-T-Y-I-H-W-V-R-Q-A-P-G-K-G-L-E-W-V-A-R-
-I-Y-P-T-N-G-Y-T-R-Y-A-D-S-V-K-G-R-F-T-I-S-A-D-T-S-
-K-N-T-A-Y-L-Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-S-R-W-G-
-G-D-G-F-Y-A-M-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K-G-
-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V-K-
-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P-A-
-V-L-Q-S-S-G-L-Y-S-L-S-S-V-V-T-V-P-S-S-S-L-G-T-Q-T-
-Y-I-C-N-V-N-H-K-P-S-N-T-K-V-D-K-K-V-E-P-K-S-C-D-K-
-T-H-T-C-P-P-C-P-A-P-E-L-L-G *Fd*

■ 3 ms ETD	■ 7 ms ETD
■ 4 ms ETD	■ 9 ms ETD
■ 5 ms ETD	■ all ETD
■ 6 ms ETD	

Figure S3. ETD Orbitrap FTMS mass spectra obtained in a single LC run for Fc/2 (top), Lc (middle), and Fd (bottom) of Adalimumab. Mass spectra averaged in the time-frames of the fragment elution were seven, seven, and five, respectively. As an example of the good SNR obtained, three product ions of Fc/2 are displayed in the top right inset. Most of product ions are highly charged and comprised in the m/z 800-1100 window.
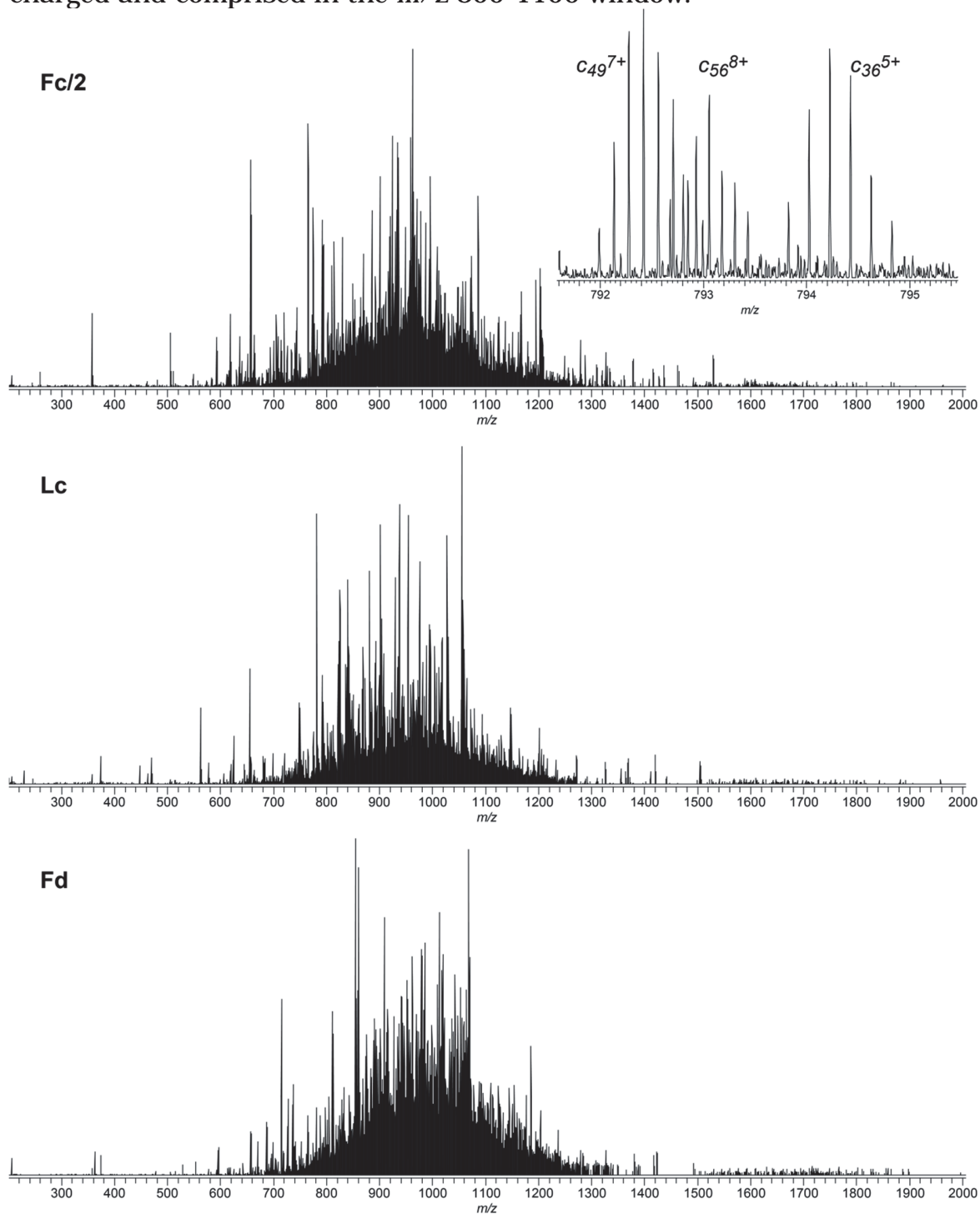
Figure S4. Fragmentation maps obtained from single 3 ms ETD LC-MS/MS run of IdeS-derived fragments of Adalimumab. The left column shows results of a run performed under "standard" HCD cell pressure conditions (Δp=0.35E-10 torr), whereas the right column presents the corresponding fragmentation map for an experiment performed using a lower HCD cell pressure (Δp=0.1E-10 torr). Product ions considered include c-, z-, and y-types.

Δp=0.35E-10 torr

Fc/2

-G-P-S-V̅F̅L̅F-P-P̅K-P̅K̅D̅T̅L̅M-I̅S-R̅T-P-E-V-T-C-
V-V-V̅D-V-S-H-E̅D-P-E̅V̅K̅F̅N̅W̅Y̅V-D-G̅V̅E-V-H̅N-
A̅K̅T̅K-P̅R̅E̅E̅Q̅Y̅N̅S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-
D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-
S-K-A-K̅G-Q-P-R-E-P-Q-V-Y-T-L-P-P-S-R̅D-E-L-T-K̅N-
Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I̅A-V-E-W-E̅S̅N-G̅Q-
P̅-E̅N-N-Y-K-T-T-P-P-V̅L-D-S̅D-G-S̅F̅F̅L-Y-S-K̅L̅T-
V-D-K̅S-R̅W̅Q̅Q̅G̅N-V-F-S-C̅S̅V-M-H-E-A-L̅H̅N̅H̅Y-
T̅Q̅K-S-L-S-L-S-P-G-K

Lc

-D-I-Q-M-T-Q-S-P-S-S̅L-S-A-S-V-G-D-R-V-T-I-T̅C-R-A-
S-Q̅G-I̅R̅N̅Y̅L-A̅W̅Y̅Q̅Q̅K-P-G̅K̅A-P̅K̅L̅L̅I-Y-A-
A-S-T̅L̅Q̅S̅G̅V-P-S-R̅F-S-G-S-G-S-G-T-D-F-T-L-T-I-
S̅S-L-Q-P-E-D-V-A-T-Y-Y-C-Q-R-Y-N-R-A-P-Y-T-F-G-Q-
G-T-K-V-E-I-K-R-T-V-A-A-P-S-V-F-I̅F-P-P-S-D-E-Q-L-
K-S-G-T̅A-S-V-V-C-L-L-N̅N̅F-Y-P-R-E-A̅K-V-Q-W̅K-V-
D̅N̅A-L-Q-S-G̅N̅S-Q̅E̅S-V̅T-E-Q-D-S-K-D-S̅T-Y̅S-L-
S̅S̅T̅L-T-L-S̅K̅A̅D̅Y̅E̅K-H̅K̅V̅Y̅A-C-E̅V-T-H̅Q-G-
L̅S̅S-P-V-T̅K̅S-F̅N-R-G-E-C

Fd

-E-V-Q-L-V-E-S-G-G-G-L-V̅Q-P-G-R-S̅L-R-L̅S̅C-A-A̅S-
G-F-T-F-D-D̅Y-A-M-H-W-V-R̅Q̅A-P-G̅K-G̅L-E̅W̅V̅S-A-
I-T-W̅N-S-G-H-I-D-Y-A-D-S-V-E̅G-R-F-T-I-S-R̅D̅N̅A-
K-N̅S-L-Y-L-Q-M̅N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S-
Y-L-S-T-A-S-S-L-D-Y-W-G-Q-G-T-L-V-T-V-S-S-A-S-T-K-
G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V-
K-D-Y-F-P-E-P-V-T-V-S-W-N-S-G-A-L-T-S-G-V-H-T-F-P-
A-V-L-Q-S̅S-G-L-Y̅S-L-S-S-V̅V-T-V̅P-S-S̅S̅L-G-T-Q-
T-Y-I-C-N-V̅N̅H̅K-P̅S̅N̅T̅K̅V̅D̅K-K-V̅E-P-K-S-C̅D-
K-T-H-T-C̅P-P-C-P-A-P-E-L-L-G

Δp=0.1E-10 torr

Fc/2

-G-P-S-V̅F̅L̅F-P-P-K-P̅K̅D-T̅L̅M-I̅S̅R̅T-P̅E-V̅T̅C̅
V-V-V̅D-V̅S-H̅E̅D-P̅E̅V̅K̅F̅N̅W̅Y̅V̅D-G-V̅E-V-H̅N̅
A-K̅T̅K-P-R̅E̅E-Q̅Y-N̅S-T-Y-R-V-V-S-V̅L-T-V̅L-H-Q-
D-W̅L̅N-G̅K-E-Y-K-C-K-V̅S̅N̅K̅A̅L-P-A-P-I-E̅K̅T-I̅
S̅K̅A̅K̅G̅Q-P-R̅E-P̅Q̅V-Y̅T̅L-P-P-S-R-D-E-L-T̅K̅N̅
Q̅V-S-L̅T̅C-L-V-K-G-F-Y-P-S̅D̅I̅A̅V̅E̅W̅E̅S-N-G̅Q-
P-E̅N̅N-Y̅K̅T-T-P-P-V-L̅D-S̅D-G̅S-F̅F̅L-Y-S-K̅L̅T-
V-D-K-S-R̅W̅Q̅Q̅G̅N-V-F-S-C̅S̅V-M-H̅E-A-L̅H̅N̅H̅Y-
T̅Q̅K̅S-L-S-L-S-P̅G-K

Lc

-D-I-Q-M-T-Q-S-P-S-S-L̅S-A-S-V-G-D-R-V-T-I-T̅C-R-A-
S̅Q-G-I̅R̅N-Y̅L-A̅W̅Y̅Q-Q̅K-P-G̅K̅A-P̅K̅L̅L̅I̅Y-A-
A̅S̅T̅L̅Q̅S̅G̅V-P-S-R̅F̅S-G̅S-G-S-G̅T̅D̅F̅T̅L-T-I̅
S̅S̅L̅Q-P-E-D-V̅A-T-Y-Y-C-Q-R-Y-N-R̅A-P-Y̅T-F-G̅Q-
G̅T̅K-V̅E̅I̅K̅R̅T-V-A-A-P-S-V̅F-I̅F-P-P-S-D̅E-Q-L̅
K-S-G̅T-A̅S-V̅V-C-L-L̅N̅N-F-Y-P-R̅E-A̅K-V̅Q̅W̅K-V-
D̅N̅A̅L̅Q-S-G-N̅S̅Q̅E̅S-V̅T̅E̅Q̅D̅S-K̅D̅S̅T̅Y̅S̅L-
S̅S-T̅L-T̅L̅S̅K̅A̅D̅Y̅E̅K̅H̅K̅V-Y̅A-C-E̅V̅T-H-Q-G-
L̅L̅S̅S̅P-V-T̅K̅S-F̅N-R-G-E-C

Fd

-E-V-Q-L-V-E-S-G-G-G-L-V̅Q-P-G-R-S̅L-R̅L̅S-C̅A̅A-S-
G-F̅T-F-D̅D-Y-A-M-H̅W-V̅R̅Q̅A-P̅G̅K-G-L̅E-W̅V̅S̅A-
I-T̅W̅N̅S-G-H̅I̅D-Y̅A̅D̅S-V-E̅G-R-F̅T̅I-S-R̅D̅N-A-
K̅N̅S-L-Y-L̅Q-M-N-S-L-R-A-E-D-T-A-V-Y-Y-C-A-K-V-S-
Y-L-S-T-A-S-S̅L-D-Y̅W-G-Q-G-T-L-V̅T-V-S-S-A̅S-T-K-
G-P-S-V-F-P-L-A-P-S-S-K-S-T-S-G-G-T-A-A-L-G-C-L-V-
K-D-Y-F-P̅E-P-V-T-V-S̅W-N-S̅G-A-L-T-S-G-V-H-T-F-P-
A̅V-L-Q̅S̅S̅G̅L̅Y̅S-L-S̅S̅V-V-T-V̅P-S̅S̅S̅L-G-T-Q-
T-Y̅I-C-N-V̅N̅H̅K-P̅S̅N̅T̅K̅V̅D̅K-K̅V̅E-P̅K-S-C-D-
K-T-H-T-C-P-P-C-P-A-P-E-L-L-G

239

Figure S5. Fragmentation map of Fc/2 of oxidized Adalimumab. Top, fragmentation map derived from a single 3 ms ETD LC-MS/MS run. Bottom, fragmentation map obtained from the averaging of transients from 10 LC-MS/MS runs, equally divided into 3 ms ETD and 5 ms ETD LC-MS/MS runs.

single LC run

-G-P-S-V-F-L-F-P-P-K-P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-
-V-V-V-D-V-S-H-E-D-P-E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-
-A-K-T-K-P-R-E-E-Q-Y-N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-
-D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-
-S-K-A-K-G-Q-P-R-E-P-Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-
-Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-
-P-E-N-N-Y-K-T-T-P-P-V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-
-V-D-K-S-R-W-Q-Q-G-N-V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-
-T-Q-K-S-L-S-L-S-P-G-K-

multiple LC runs summed

-G-P-S-V-F-L-F-P-P-K-P-K-D-T-L-M-I-S-R-T-P-E-V-T-C-
-V-V-V-D-V-S-H-E-D-P-E-V-K-F-N-W-Y-V-D-G-V-E-V-H-N-
-A-K-T-K-P-R-E-E-Q-Y-N-S-T-Y-R-V-V-S-V-L-T-V-L-H-Q-
-D-W-L-N-G-K-E-Y-K-C-K-V-S-N-K-A-L-P-A-P-I-E-K-T-I-
-S-K-A-K-G-Q-P-R-E-P-Q-V-Y-T-L-P-P-S-R-D-E-L-T-K-N-
-Q-V-S-L-T-C-L-V-K-G-F-Y-P-S-D-I-A-V-E-W-E-S-N-G-Q-
-P-E-N-N-Y-K-T-T-P-P-V-L-D-S-D-G-S-F-F-L-Y-S-K-L-T-
-V-D-K-S-R-W-Q-Q-G-N-V-F-S-C-S-V-M-H-E-A-L-H-N-H-Y-
-T-Q-K-S-L-S-L-S-P-G-K-

Chapter 7: Conclusions

7.1 Summary of results

The experimental studies presented in this Thesis advance the mass spectrometry (MS)-based in-depth structural investigation of polypeptides with sizes ranging from the peptide level to intact proteins, reaching an apex in the analysis of intact ~150 kDa IgGs. Indispensable analytical characteristics for all the here reported studies are high mass resolution and mass accuracy, as well as efficient biomolecular ion activation and dissociation. Particular attention is paid to development and optimization of biomolecular sample preparation and separation techniques, as well as methods of data analysis.

The importance of the analytical characteristics for biomolecular MS is first demonstrated by studies involving peptides, specifically the neuropeptide substance P and IgG-derived tryptic glycopeptides. In the former case, deamidation and transamidation of substance P by the enzyme tissue transglutaminase (tTGase) were found to be competitive processes as revealed by electron capture dissociation (ECD) FT-ICR MS (*paper I*). Interestingly, the employed LC-MS/MS setup allowed not only the *differentiation* between non-deamidated peptides and the singly or doubly deamidated counterparts, but also to define the *precise order* according to which the two glutamine residues present in substance P are sequentially converted into glutamic acid by the catalytic activity of the enzyme. The difference in reactivity of the two glutamines as tTGase substrates was further confirmed by the analysis of transamidation: substance P dimers with different deamidation degrees, fragmented by ECD *without rupture* of the newly formed amide bond, were found to be originated only from one of the two glutamines, the same which showed higher propensity to deamidation.

Glycopeptides derived from tryptic digestion of a transiently expressed IgG were on the contrary identified on the base of MS^1 thanks to high mass accuracy offered by Fourier transform MS and the important observation at the base of glycopeptide databases which consider only biologically relevant forms of oligosaccharides for a selected organism (*paper II*). This elegant methodology was considered, nevertheless, as only one part of the necessary glycan characterization for establishing the batch-to-batch consistence of IgG transient expression. Additional data, which can be considered as a prelude of the remaining study described in this Thesis, included *intact*

mass measurements of IgGs by high resolution qTOF MS, and were applied for a qualitative definition of major glycoforms (i.e., combinations of two N-linked glycans to the IgG sequence) in the different IgG batches.

From there, efforts have been focused on the developments of methodologies for targeted top-down structural analysis of large intact proteins, particularly monoclonal IgGs, by electron transfer dissociation (ETD) MS. A detailed assessment was performed for the fragmentation of ~79 kDa serotransferrin, which was analyzed in both oxidized and reduced forms, both resulting in ~10% sequence coverage with *complementary fragmentation* patterns in ETD qTOF MS experiments (*paper III*). These data elucidated the relevance of disulfide bridges in maintaining the protein in a conformation not easily fragmented in the gas phase, with most of the fragments derived from disulfide-free regions. At the same time, what had been previously reported for peptides, which is that radical-driven ion dissociation methods can *cleave disulfide bonds*, was also confirmed for ETD of transferrin, with a few product ions derived from disulfide-protected regions. Moreover, results also show the potential of high resolution TOF mass analyzers for ETD-based top-down MS.

Papers IV, *V*, and *VI* were entirely dedicated to the top-down analysis of IgGs by ETD MS, employing both TOF and Orbitrap mass analyzers. Although a precise comparison of the performances of which different MS platforms were capable is impossible due to the heterogeneity of samples analyzed, both hybrid qTOF and LTQ-Orbitrap mass spectrometers allowed reaching protein sequence coverage not only higher than that previously reported for single antibody chains fragmented *via* CID, but also more extended relative to that of the largest protein previously analyzed in intact form by ECD, the 142 kDa cardiac myosin binding protein C. Specifically, ~21% sequence coverage was reached for a murine IgG1 on the qTOF instrument (*paper IV*), whereas ~32% was obtained in a more extensive study for a human IgG1 by Orbitrap FTMS (*paper V*), which also allowed the identification of pyroglutamic acid at the N-terminus of the IgG heavy chain. Dedicated strategies were also developed and applied for improving the signal-to-noise ratio of tandem mass spectra. Besides further confirming the ETD capabilities of cleaving disulfide bridges and, as proven in *paper V*, generating a relatively high number of *y*-ions also during the fragmentation of large proteins, these two research projects highlighted that the *multi-domain nature* of IgGs or, more specifically, of their heavy and light chains reverberate

on the observed fragmentation pattern, which alternates high sequencing of disulfide-free areas and poor fragmentation of disulfide protected regions. Effectively, *immunoglobulin domains*, whose structure resembles that of beta barrels, have a compact conformation in solution which is potentially retained even in the gas phase due to the presence of an *inter-molecular disulfide bridge*. This hypothesis found new support in the results described in *paper VI*. The comparison of fragmentation patterns of IgGs belonging to the same subclass (IgG1) obtained on the same platform and under the same experimental conditions strongly suggested that a recurring fragmentation pattern and comparable total sequence coverage between the IgGs were consequence of a similar gas phase conformation: in fact, areas highly fragmented in both IgGs are the same and belong to the so-called *complementarity determining regions,* which differ from an antibody to the other, whereas among the non-sequenced regions we found wide areas with complete sequence homology.

In addition to this, results of *paper VI* were also analyzed considering the product ion abundances. PIA analysis not only aided explaining the differences observable in ETD mass spectra recorded applying different ion-ion interaction times, but also reinforced the idea that, independently from the ETD parameters in use, ETD is effective only on loops interconnecting consecutive immunoglobulin domains. Finally, fragmentation of a monoclonal IgG2 served the purpose of elucidating the role of *inter-molecular disulfide bridges* in IgG fragmentation. From the analytical chemistry point of view, though, the most relevant achievements reported in the paper are the fine tuning of ETD and, particularly, *ion transmission parameters*, and the introduction for the top-down analysis of large proteins of an hybrid, two-stage fragmentation technique combining ETD and HCD, *EThcD*.

Middle-down MS could represent one of the ways for overcoming the limit set to ~30% of sequence coverage in the case of the analysis of intact IgGs, and possibly obtaining sequence confirmation for all the CDR regions (*paper VII*). The developed middle-down method was based on a rapid, robust protocol centered on the *bacterial enzyme IdeS*, whose proteolytic products were analyzed by ETD MS/MS. This approach was demonstrated capable of yielding almost 50% sequence coverage in a single LC-MS/MS run on a recent generation LTQ-Orbitrap instrument (Orbitrap Elite FTMS). Importantly, this method, being based on the analysis of large IgG fragments of ~25 kDa each, took advantage of all the advances developed and successfully

employed for TD MS (time-domain (transient) signal averaging for increasing SNR, optimization of ion transmission by reduction of the gas pressure in the HCD cell and HCD trapping). The size of the analyzed fragments resulted is also compatible with the *identification and localization of PTMs*, specifically of methionine oxidation.

7.2 Concluding remarks and future perspectives

Without the presumptuous idea of following Hegel's statement *"contradictio est regula veri, non contradictio falsi"*, on which his Dialectic is grounded, my personal opinion is that, nevertheless, relevant achievements contained in this work (and, in general, in a research that intends to propose or improve analytical methods) arise not only from *positive experimental results* but also, if not primarily, by the identification of *limitations* in the currently available techniques for top-down mass spectrometry.

Among the former, the most important is surely represented by the record value of ~32% sequence coverage reached for an intact IgG1 by ETD Orbitrap FTMS (*paper V*). Disclosing the possibility of exploiting the theoretical advantages of TD MS over the other alternative approaches (e.g., MD MS and BU MS) in the analysis of therapeutic antibodies by finely tuning instrumental parameters will probably be beneficial in the close future for the biotechnology industry, where this technology will contribute in elevating quality control standards.

Furthermore, beyond the high sequence coverage achieved by MD and TD MS, it is important to recall that deamidation of Asn and Gln as well as iso-Asp formation are PTMs commonly found in antibodies. Their relevance is due to the fact that their presence in the CDRs can modify the three-dimensional structure of the antigen binding site, thus impairing the biological activity of the antibody. For instance, about one fourth of therapeutic IgG Trastuzumab administered to patients is known to present an isomerized Asp on a CDR that completely prevents the antigen binding.[111, 112] On the other hand, in the pharmaceutical and biotechnological industry the interest around antibody-drug conjugates (ADCs) is rapidly increasing.[113] Several types of chemical reactions have been used so far to covalently bind small organic molecules to immunoglobulins, and potentially the bonds connecting amino acid side chains with drugs can be cleaved by vibrational-based ion activation. Altogether, therefore, results achieved with radical-

driven tandem mass spectrometry even at the peptide level might disclose new opportunity for the future of fine antibody and ADC analysis.

On the other hand, the challenging analysis of a high molecular weight, multichain protein imposed the optimization of specific settings for all the required experimental steps, namely: precursor ion charge state selection for efficient electron transfer dissociation, product ion transfer to the high resolution mass analyzer, and averaging of multiple mass spectra or time-domain transient signals for SNR improvement. Therefore, the results presented in this Thesis, although based on targeted studies, can be interpreted in a wider horizon, as they open the possibility to concretely undertake the study of any large, multichain proteins in a top-down fashion. Pushing the upper limit of the molecular weight of proteins to analyze them in their intact forms is of fundamental importance in the perspective of challenging initiatives as the cellular-based Human Proteome Project, which aims at the definition of cellular types on the base of their respective proteomes conceived as peculiar pools of specific proteoforms, and therefore is indissolubly linked to TD proteomics.[114] It is noteworthy, however, that the use of proteolytic enzyme has been part of research studies presented in this Thesis, as the current contribution of proteolysis-based MS analysis of proteins is still fundamental and will most likely continue to complement for a very long time TD MS in the future.

With regard to the limitations emerged during top-down MS of IgGs, the most apparent one is surely the divisions of both light and heavy chains in areas that are almost fully sequenced and others which are essentially not affected by fragmentation. Beyond this phenomenological description, a deeper reconsideration of obtained data was achieved by combining results of product ion abundance analysis and structural information derived from biophysical techniques. As a conclusion we formulated the plausible hypothesis that the retention of a compact, partially folded conformation in the gas phase is hindering the protein fragmentation efficiency. Waiting for the verification of this hypothesis by other MS-based techniques such as ion mobility MS, some facts already manifest, such as the characteristic of disulfide bonds of reducing drastically the sequence coverage, despite the capability of ETD of cleaving them and the possibility of applying strategies for improving spectral SNR. Understanding the reasons for the limited sequence coverage will surely facilitate the design of new ion fragmentation techniques. One attempt we already finalized consisted in exploiting the hybrid design of one of our mass

spectrometers to combine two different activation techniques, namely ETD and HCD, to gain new product ions and facilitate the separation of undissociated complexes. In addition to that, beneficial results might derive from pre-activation of ions via IR photons or collisions with neutral molecules, which are established methods for unfolding peptides and proteins before ECD and ETD fragmentation, that have not been applied yet to IgGs. Similarly, a new radical-driven ion activation method, 193 nm ultraviolet photodissociation, has demonstrated protein sequencing capabilities even superior to ETD, but so far it has not been applied to disulfide bond-rich proteins like IgGs. A new possible solution might derive from a combination of photodissociation and hybrid design for MS^2: Fung $et\ al.$[115] demonstrated that 157 nm photons can easily cleave disulfide bridges without relevant charge reduction of precursor cations, and hence an application of a 157 nm laser in combination with ECD or ETD could be envisioned.

Finally, in parallel with instrumentation and method development, an improvement for the tools dedicated to the analysis of top-down MS data of large proteins is required. In tandem MS, the number of available fragmentation channels increases rapidly with the length of the polypeptidic chain. In the case of large protein, some of these channels lead to the formation of non-canonical fragments, such as internal fragments or, in ECD and ETD, product ions with atypical number of hydrogens (e.g., the so called "$z^{.}+n$" ions). Therefore, given that current versions of software for top-down MS analysis do exclusively consider canonical product ion pairs (e.g., b- and y-type or c- and z-type), the actual risk in the analysis of high molecular weight protein is not only that good part of the information obtainable from tandem mass spectra remains unused, but, paradoxically, that increasing the number of product ions which are detected but not identified might result in lowering the identification scores.

Effectively, as described in particular in $papers\ III$ and V, specific data analysis strategies must be applied to the analysis of ETD TD MS mass spectra, especially for large proteins containing multiple disulfide bonds. The data collected for the here described research studies surely contributed to the understanding of the possibilities currently offered by available algorithms for TD data treatment.

Besides the analytical aspects to be improved, however, it is reasonable to consider the methods for targeted studies as mature and sophisticated enough for being applied to large-scale investigations, i.e., proteomic studies, potentially with a focus on drug discovery. We envision

the use of a combination of middle-down and top-down MS, performed in different variants (e.g., from accurate mass measurement to sequence confirmation by MS^2), to determine the presence in the serum of immunogenized animals of IgGs active against a specific antigen. Genetic information will provide a large pool of candidate IgGs, and MS-based analysis will be carried out to verify their actual presence.

This project resembles a typical proteomic survey, with a database, in this case limited to IgGs, which is searched for finally achieving protein identification. Compared to what already performed, this study will be likely based primarily on the analysis of ~100 kDa $F(ab')_2$ fragments, as they maintain the variable regions of light and heavy chains which are responsible for IgG diversification and, conversely, lack of the Fc region which is highly conserved and is a source of additional sample heterogeneity (i.e., it is N-glycosylated). My personal hope is that this will be just the first one of a series of large-scale biological applications involving large, >50 kDa, proteins to be performed by top-down mass spectrometry.

References

[1]   Schmutz, J., Wheeler, J., Grimwood, J., Dickson, M. et al., Quality assessment of the human genome sequence. *Nature* 2004, *429*, 365-368.

[2]   Voet, D., Voet, J. G., *Biochemistry*. 4th ed. 2011, Hoboken, NJ: John Wiley & Sons. xxv, 1428, I-1453 p.

[3]   Ambrogelly, A., Palioura, S., Soll, D., Natural expansion of the genetic code. *Nature Chemical Biology* 2007, *3*, 29-35.

[4]   Lewin, B., *Genes IX*. 9th ed. 2008, Sudbury, Mass.: Jones and Bartlett Publishers. xvii, 892 p.

[5]   Smith, L. M., Kelleher, N. L., Proteomics, C. T. D., Proteoform: a single term describing protein complexity. *Nature Methods* 2013, *10*, 186-187.

[6]   Edman, P., Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica* 1950, *4*, 283-293.

[7]   Hoffmann, E. d., Stroobant, V., *Mass spectrometry : principles and applications*. 3rd ed. 2007, Chichester, England ; Hoboken, NJ: J. Wiley. xii, 489 p.

[8]   Gonzalez-Porta, M., Frankish, A., Rung, J., Harrow, J., Brazma, A., Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol* 2013, *14*, R70.

[9]   Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.

[10]   Zhang, Y. Y., Fonslow, B. R., Shan, B., Baek, M. C., Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* 2013, *113*, 2343-2394.

[11]   Wu, C., Tran, J. C., Zamdborg, L., Durbin, K. R. et al., A protease for 'middle-down' proteomics. *Nature Methods* 2012, *9*, 822-824.

[12]   Xu, P., Peng, J. M., Characterization of polyubiquitin chain structure by middle-down mass spectrometry. *Analytical Chemistry* 2008, *80*, 3438-3444.

[13]   Siuti, N., Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nature Methods* 2007, *4*, 817-821.

[14]   Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E. et al., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, *480*, 254-U141.

[15]   Cui, W. D., Rohrs, H. W., Gross, M. L., Top-down mass spectrometry: Recent developments, applications and perspectives. *Analyst* 2011, *136*, 3854-3864.

[16]   McLafferty, F. W., Turecek, F., *Interpretation of mass spectra*. 4th ed. 1993, Mill Valley, Calif.: University Science Books. xviii, 371 p.

[17]   McNaught, A. D., Wilkinson, A., International Union of Pure and Applied Chemistry., *Compendium of chemical terminology : IUPAC recommendations*. 2nd ed. 1997, Oxford Oxfordshire ; Malden, MA: Blackwell Science. vii, 450 p.

[18]   Zubarev, R. A., Demirev, P. A., Hakansson, P., Sundqvist, B. U. R., Approaches and Limits for Accurate Mass Characterization of Large Biomolecules. *Analytical Chemistry* 1995, *67*, 3793-3798.

[19]   Zubarev, R. A., Hakansson, P., Sundqvist, B., Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Analytical Chemistry* 1996, *68*, 4060-4063.

[20]   Downard, K., Royal Society of Chemistry (Great Britain), *Mass spectrometry : a foundation course*. 2004, Cambridge: Royal Society of Chemistry. xvi, 210 p.

[21]   Nomenclature and Symbolism for Amino-Acids and Peptides - Recommendations 1983. *Journal of Biological Chemistry* 1985, *260*, 14-42.

[22]   Fruton, J. S., Early theories of protein structure. *Ann N Y Acad Sci* 1979, *325*, xiv, 1-18.

[23]   Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology* 1963, *7*, 95-99.

[24]   Kleywegt, G. J., Jones, T. A., Phi/psi-chology: Ramachandran revisited. *Structure* 1996, *4*, 1395-1400.

[25]   Kumar, S., Nussinov, R., Close-range electrostatic interactions in proteins. *Chembiochem* 2002, *3*, 604-617.

[26]   Arora, N., Jayaram, B., Strength of hydrogen bonds in alpha helices. *Journal of Computational Chemistry* 1997, *18*, 1245-1252.

[27]   Dill, K. A., MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* 2012, *338*, 1042-1046.

[28]   Brändén, C.-I., Tooze, J., *Introduction to protein structure.* 2nd ed. 1999, New York: Garland Pub. xiv, 410 p.

[29]   Wilkinson, K. D., The discovery of ubiquitin-dependent proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102*, 15280-15282.

[30]   Hay, R. T., SUMO: A history of modification. *Molecular Cell* 2005, *18*, 1-12.

[31]   Burda, P., Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochimica Et Biophysica Acta-General Subjects* 1999, *1426*, 239-257.

[32]   Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray Ionization-Principles and Practice. *Mass Spectrometry Reviews* 1990, *9*, 37-70.

[33]   Tanaka, K., Waki, H., Ido, Y., Akita, S. et al., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 1988, *2*, 151-153.

[34]   Karas, M., Bachmann, D., Hillenkamp, F., Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Analytical Chemistry* 1985, *57*, 2935-2939.

[35]   Karas, M., Hillenkamp, F., Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical Chemistry* 1988, *60*, 2299-2301.

[36]   Taflin, D. C., Ward, T. L., Davis, E. J., Electrified Droplet Fission and the Rayleigh Limit. *Langmuir* 1989, *5*, 376-384.

[37]   Kebarle, P., Verkerk, U. H., Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev* 2009, *28*, 898-917.

[38]   Glish, G. L., Mcluckey, S. A., Ridley, T. Y., Cooks, R. G., A New Hybrid Sector Quadrupole Mass-Spectrometer for Mass-Spectrometry Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* 1982, *41*, 157-177.

[39]   Glish, G. L., Burinsky, D. J., Hybrid mass spectrometers for tandem mass Spectrometry. *Journal of the American Society for Mass Spectrometry* 2008, *19*, 161-172.

[40]   Scigelova, M., Hornshaw, M., Giannakopulos, A., Makarov, A., Fourier Transform Mass Spectrometry. *Molecular & Cellular Proteomics* 2011, *10*.

[41]   March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry* 1997, *32*, 351-369.

[42]   Ouyang, Z., Wu, G. X., Song, Y. S., Li, H. Y. et al., Rectilinear ion trap: Concepts, calculations, and analytical performance of a new mass analyzer. *Analytical Chemistry* 2004, *76*, 4595-4605.

[43]   Schwartz, J. C., Senko, M. W., Syka, J. E. P., A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 2002, *13*, 659-669.

[44]   Marshall, A. G., Hendrickson, C. L., Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 1998, *17*, 1-35.

[45]   Michalski, A., Damoc, E., Lange, O., Denisov, E. et al., Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Molecular & Cellular Proteomics* 2012, *11*.

[46]   Second, T. P., Blethrow, J. D., Schwartz, J. C., Merrihew, G. E. et al., Dual-Pressure Linear Ion Trap Mass Spectrometer Improving the Analysis of Complex Protein Mixtures. *Analytical Chemistry* 2009, *81*, 7757-7765.

[47]   Makarov, A., Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry* 2000, *72*, 1156-1162.

[48]   Zubarev, R. A., Makarov, A., Orbitrap Mass Spectrometry. *Analytical Chemistry* 2013, *85*, 5288-5296.

[49]   Ahlf, D. R., Compton, P. D., Tran, J. C., Early, B. P. et al., Evaluation of the Compact High-Field Orbitrap for Top-Down Proteomics of Human Cells. *Journal of Proteome Research* 2012, *11*, 4308-4314.

[50]   Michalski, A., Damoc, E., Lange, O., Denisov, E. et al., Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 2012, *11*, O111.013698.

[51]   Lange, O., Damoc, E., Wieghaus, A., Makarov, A., Enhanced Fourier Transform for Orbitrap Mass Spectrometry. *Proc. 59th Conf. Amer. Soc. Mass Spectrom.,* Denver, June 5-9, 2011.

[52]   Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L. et al., A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics* 2009, *8*, 2759-2769.

[53]   Guilhaus, M., Principles and Instrumentation in Time-of-Flight Mass-Spectrometry - Physical and Instrumental Concepts. *Journal of Mass Spectrometry* 1995, *30*, 1519-1532.

[54]   Vorm, O., Roepstorff, P., Mann, M., Improved Resolution and Very High-Sensitivity in Maldi Tof of Matrix Surfaces Made by Fast Evaporation. *Analytical Chemistry* 1994, *66*, 3281-3287.

[55]   Kaplan, D. A., Hartmer, R., Speir, J. P., Stoermer, C. et al., Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Communications in Mass Spectrometry* 2008, *22*, 271-278.

[56]   Chernushevich, I. V., Loboda, A. V., Thomson, B. A., An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 2001, *36*, 849-865.

[57]   Johnson, R. S., Martin, S. A., Biemann, K., Collision-Induced Fragmentation of (M+H)+Ions of Peptides - Side-Chain Specific Sequence Ions. *International Journal of Mass Spectrometry and Ion Processes* 1988, *86*, 137-154.

[58]   Wells, J. M., McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Biological Mass Spectrometry* 2005, *402*, 148-185.

[59]   Olsen, J. V., Macek, B., Lange, O., Makarov, A. et al., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* 2007, *4*, 709-712.

[60]   Little, D. P., Speir, J. P., Senko, M. W., Oconnor, P. B., Mclafferty, F. W., Infrared Multiphoton Dissociation of Large Multiply-Charged Ions for Biomolecule Sequencing. *Analytical Chemistry* 1994, *66*, 2809-2815.

[61]   Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 1998, *120*, 3265-3266.

[62]   Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, *101*, 9528-9533.

[63]   Satake, H., Hasegawa, H., Hirabayashi, A., Hashimoto, Y., Baba, T., Fast multiple electron capture dissociation in a linear radio frequency quadrupole ion trap. *Analytical Chemistry* 2007, *79*, 8755-8761.

[64]   Ding, L., Brancia, F. L., Electron capture dissociation in a digital ion trap mass spectrometer. *Analytical Chemistry* 2006, *78*, 1995-2000.

[65]   Zubarev, R. A., Kruger, N. A., Fridriksson, E. K., Lewis, M. A. et al., Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *Journal of the American Chemical Society* 1999, *121*, 2857-2862.

[66]   Creese, A. J., Cooper, H. J., The effect of phosphorylation on the electron capture dissociation of peptide ions. *J Am Soc Mass Spectrom* 2008, *19*, 1263-1274.

[67]   Mirgorodskaya, E., Roepstorff, P., Zubarev, R. A., Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem* 1999, *71*, 4431-4436.

[68]   Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L. et al., Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 2000, *72*, 563-573.

[69]   Zeller, M., Mueller, M., Damoc, E., Denisov, E. et al., Systematic evaluation of ultrahigh resolution MS instrument parameter to optimize topdown analysis. *Proceedings of Sanibel Mass Spectrometry Conference,* Clearwater, January 24-27, 2013.

[70]   Amster, I. J., Fourier transform mass spectrometry. *Journal of Mass Spectrometry* 1996, *31*, 1325-1337.

[71]   Senko, M. W., Beu, S. C., Mclafferty, F. W., Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply-Charged Ions. *Journal of the American Society for Mass Spectrometry* 1995, *6*, 52-56.

[72]   Horn, D. M., Zubarev, R. A., McLafferty, F. W., Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of the American Society for Mass Spectrometry* 2000, *11*, 320-332.

[73]   Liu, X. W., Sirotkin, Y., Shen, Y. F., Anderson, G. et al., Protein Identification Using Top-Down. *Molecular & Cellular Proteomics* 2012, *11*.

[74]   Senko, M. W., Beu, S. C., Mclafferty, F. W., Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *Journal of the American Society for Mass Spectrometry* 1995, *6*, 229-233.

[75]   Salih, E., Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. *Mass Spectrometry Reviews* 2005, *24*, 828-846.

[76]   Ong, S. E., Mittler, G., Mann, M., Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature Methods* 2004, *1*, 119-126.

[77] Choudhary, C., Mann, M., Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular Cell Biology* 2010, *11*, 427-439.

[78] Zhang, K. L., Williams, K. E., Huang, L., Yau, P. et al., Histone acetylation and deacetylation - Identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. *Molecular & Cellular Proteomics* 2002, *1*, 500-508.

[79] Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L. et al., Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* 2009, *325*, 834-840.

[80] Stensland, M., Holm, A., Kiehne, A., Fleckenstein, B., Targeted analysis of protein citrullination using chemical modification and tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* 2009, *23*, 2754-2762.

[81] Hurtado, P. P., O'Connor, P. B., Differentiation of isomeric amino acid residues in proteins and peptides using mass spectrometry. *Mass Spectrometry Reviews* 2012, *31*, 609-625.

[82] Stephenson, R. C., Clarke, S., Succinimide Formation from Aspartyl and Asparaginyl Peptides as a Model for the Spontaneous Degradation of Proteins. *Journal of Biological Chemistry* 1989, *264*, 6164-6170.

[83] Arentz-Hansen, H., Korner, R., Molberg, O., Quarsten, H. et al., The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *Journal of Experimental Medicine* 2000, *191*, 603-612.

[84] Molberg, O., Mcadam, S. N., Korner, R., Quarsten, H. et al., Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease (vol 4, pg 713, 1998). *Nature Medicine* 1998, *4*, 974-974.

[85] Li, X. J., Cournoyer, J. J., Lin, C., O'Cormora, P. B., Use of O-18 labels to monitor deamidation during protein and peptide sample processing. *Journal of the American Society for Mass Spectrometry* 2008, *19*, 855-864.

[86] Li, X. J., Lin, C., O'Connor, P. B., Glutamine Deamidation: Differentiation of Glutamic Acid and gamma-Glutamic Acid in Peptides by Electron Capture Dissociation. *Analytical Chemistry* 2010, *82*, 3606-3615.

[87] Cournoyer, J. J., Pittman, J. L., Ivleva, V. B., Fallows, E. et al., Deamidation: Differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. *Protein Science* 2005, *14*, 452-463.

[88] Yang, H., Fung, Y. M. E., Zubarev, A. R., Zubarev, R. A., The Use of ECD for Proteomics-wide Identification and Quantification of iso-Asp Residues. *Molecular & Cellular Proteomics* 2009, S13-S13.

[89] O'Connor, P. B., Cournoyer, J. J., Pitteri, S. J., Chrisman, P. A., McLuckey, S. A., Differentiation of aspartic and isoaspartic acids using electron transfer dissociation. *Journal of the American Society for Mass Spectrometry* 2006, *17*, 15-19.

[90] Li, X. J., Yu, X., Costello, C. E., Lin, C., O'Connor, P. B., Top-Down Study of beta(2)-Microglobulin Deamidation. *Analytical Chemistry* 2012, *84*, 6150-6157.

[91] Hanisch, F. G., O-glycoproteomics: site-specific O-glycoprotein analysis by CID/ETD electrospray ionization tandem mass spectrometry and top-down glycoprotein sequencing by in-source decay MALDI mass spectrometry. *Methods Mol Biol* 2012, *842*, 179-189.

[92] Zhang, H., Li, X. J., Martin, D. B., Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology* 2003, *21*, 660-666.

[93] Carr, S. A., Huddleston, M. J., Bean, M. F., Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci* 1993, *2*, 183-196.

[94] Zhu, Z., Su, X., Clark, D. F., Go, E. P., Desaire, H., Characterizing O-linked glycopeptides by electron transfer dissociation: fragmentation rules and applications in data analysis. *Anal Chem* 2013, *85*, 8403-8411.

[95] Lingg, N., Zhang, P. Q., Song, Z. W., Bardor, M., The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis. *Biotechnology Journal* 2012, *7*, 1462-1472.

[96] Grass, J., Pabst, M., Chang, M., Wozny, M., Altmann, F., Analysis of recombinant human follicle-stimulating hormone (FSH) by mass spectrometric approaches. *Analytical and Bioanalytical Chemistry* 2011, *400*, 2427-2438.

[97] Jefferis, R., Isotype and glycoform selection for antibody therapeutics. *Archives of Biochemistry and Biophysics* 2012, *526*, 159-166.

[98] Go, E. P., Rebecchi, K. R., Dalpathado, D. S., Bandu, M. L. et al., GlycoPep DB: A tool for glycopeptide analysis using a "smart search". *Analytical Chemistry* 2007, *79*, 1708-1713.

[99] Taylor, M. E., Drickamer, K., *Introduction to glycobiology*. 2nd ed. 2006, Oxford ; New York: Oxford University Press. xix, 255 p.

[100] Janeway, C., *Immunobiology : the immune system in health and disease*. 6th ed. 2005, New York: Garland Science. xxiii, 823 p.

[101] Jefferis, R., Antibody therapeutics: isotype and glycoform selection. *Expert Opinion on Biological Therapy* 2007, *7*, 1401-1413.

[102] Nemazee, D., Receptor editing in lymphocyte development and central tolerance. *Nature Reviews Immunology* 2006, *6*, 728-740.

[103] Bork, P., Holm, L., Sander, C., The Immunoglobulin Fold - Structural Classification, Sequence Patterns and Common Core. *Journal of Molecular Biology* 1994, *242*, 309-320.

[104] Jiang, X. R., Song, A., Bergelson, S., Arroll, T. et al., Advances in the assessment and control of the effector functions of therapeutic antibodies. *Nature Reviews Drug Discovery* 2011, *10*, 101-110.

[105] Liu, H. C., Gaza-Bulseco, G., Faldu, D., Chumsae, C., Sun, J., Heterogeneity of monoclonal antibodies. *Journal of Pharmaceutical Sciences* 2008, *97*, 2426-2447.

[106] Chelius, D., Jing, K., Lueras, A., Rehder, D. S. et al., Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Analytical Chemistry* 2006, *78*, 2370-2376.

[107] Chelius, D., Xiao, G., Nichols, A. C., Vizel, A. et al., Automated tryptic digestion procedure for HPLC/MS/MS peptide mapping of immunoglobulin gamma antibodies in pharmaceutics. *Journal of Pharmaceutical and Biomedical Analysis* 2008, *47*, 285-294.

[108] Ayoub, D., Jabs, W., Resemann, A., Evers, W. et al., Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *MAbs* 2013, *5*, 699-710.

[109] Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., Zhang, Z. Q., Mass Measurement and Top-Down HPLC/MS Analysis of Intact Monoclonal Antibodies on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* 2009, *20*, 1415-1424.

[110] von Pawel-Rammingen, U., Johansson, B. P., Bjorck, L., IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *Embo Journal* 2002, *21*, 1607-1615.

[111] Harris, R. J., Kabakoff, B., Macchi, F. D., Shen, F. J. et al., Identification of multiple sources of charge heterogeneity in a recombinant antibody. *Journal of Chromatography B* 2001, *752*, 233-245.

[112] Diepold, K., Bomans, K., Wiedmann, M., Zimmermann, B. et al., Simultaneous Assessment of Asp Isomerization and Asn Deamidation in Recombinant Antibodies by LC-MS following Incubation at Elevated Temperatures. *Plos One* 2012, *7*.

[113] Hamblett, K. J., Senter, P. D., Chace, D. F., Sun, M. M. C. et al., Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clinical Cancer Research* 2004, *10*, 7063-7070.

[114] Kelleher, N. L., A Cell-Based Approach to the Human Proteome Project. *Journal of the American Society for Mass Spectrometry* 2012, *23*, 1617-1624.

[115] Fung, Y. M. E., Kjeldsen, F., Silivra, O. A., Chan, T. W. D., Zubarev, R. A., Facile disulfide bond cleavage in gaseous peptide and protein cations by ultraviolet photodissociation at 157 nm. *Angewandte Chemie-International Edition* 2005, *44*, 6399-6403.

Curriculum Vitae

Luca Fornelli

Address	Chemin Auguste-Pidou 5, 1007 Lausanne, Switzerland
Phone number	+41 76 75 60 513
Date of birth	30th October 1983
Nationality	Italian
Email address	luca.fornelli@gmail.com
Skype	Luca Fornelli *(Switzerland)*

Education and training

Philosophy Doctorate, Chemistry (May 2009 – Dec 2013)

- Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
- Doctoral Program in Chemistry and Chemical Engineering (EDCH)
- Thesis title: "Top-down and middle-down mass spectrometry structural characterization of monoclonal antibodies"
- Supervisor: Prof. Yury O. Tsybin

Visiting Scholar (May 2012)

- Northwestern University, Evanston, IL, USA
- Host: Prof. Neil L. Kelleher

Master of science, Industrial Biotechnology (Oct 2005 – Dec 2007)

- University of Padova, Italy
- Molecular biology, biochemistry, chemistry, bioinformatics, cell biology
- Major in "Genomics and Proteomics"
- Thesis title: "Synthesis and functionalization of gold nanoparticles for biotechnological applications"
- Supervisor: Prof. Flavio Maran
- Final mark: 110/110 cum laude

Bachelor of Science, Biotechnology (Sep 2002 – Sep 2005)

- University of Padova, Italy
- Mathematics, physics, biochemistry, molecular biology, physical chemistry, organic chemistry
- Thesis title: "Study of membranes by AFM-SECM"
- Supervisor: Prof. Flavio Maran
- Final mark: 110/110 cum laude

Publications and awards

Publications

- **Fornelli L***, Schmid AW*, Grasso L, Vogel H, Tsybin YO. Deamidation and transamidation of substance P by tissue transglutaminase revealed by electron capture dissociation Fourier transform mass spectrometry. *Chemistry – Eur. J.* (2011) 17(2), 486-497
- Tsybin YO, **Fornelli L**, Stoermer C, Luebeck M, Parra J, Nallet S, Wurm FM, Hartmer R. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal Chem* (2011) 83(23), 8919-8927
- Tsybin YO, **Fornelli L**, Kozhinov AN, Vorobyev A, Miladinović SM. High-resolution and tandem mass spectrometry--the indispensable tools of the XXI century. *Chimia* (2011) 65(9), 641-645
- Nallet S*, **Fornelli L***, Schmitt S, Parra J, Baldi L, Tsybin YO, Wurm FM. Glycan variability on a recombinant IgG antibody transiently produced in HEK-293E cells. *New Biotechnology* (2012) 29(4), 471-47
- Miladinović SM, **Fornelli L**, Lu Y, Piech KM, Girault HH, Tsybin YO. In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Anal Chem* (2012) 84(11), 4647-4651
- Gorshkov MV, **Fornelli L**, Tsybin YO. Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom* (2012) 26(15), 1711-1717
- **Fornelli L**, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics* (2012) 11(12), 1758-1767
- Zhurov KO, **Fornelli L**, Wodrich MD, Laskay ÜA, Tsybin YO. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews* (2013) 42(12), 5014-5030
- Laskay ÜA, Srzentić K, **Fornelli L**, Upir O, Kozhinov AN, Monod M, Tsybin YO. Practical considerations for improving the productivity of mass spectrometry-based proteomics. *Chimia,* (2013) 67(4), 244-249
- **Fornelli L**, Parra J, Hartmer R, Stoermer C, Luebeck M, Tsybin YO. Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry, *Anal Bioanal Chem*, (2013) 405(26), 8505-8514

Awards

- Excellence in teaching award 2011, Doctoral School of Chemistry at EPFL (EDCH)
- SCS/SCNAT Travel award 2013
- Best oral presentation award, Analytical chemistry division of SCS, Lausanne, 2013
- SNF "Early PostDoc Mobility" Fellowship

Oral presentations

- Swiss Group for Mass Spectrometry (SGMS) meeting, 2010, Beatenberg, Switzerland
- Swiss Proteomics Society (SPS) PhD student Symposium, 2012, Basel, Switzerland
- American Society for Mass Spectrometry (ASMS) Conference, 2013, Minneapolis, USA
- Fall meeting of the Swiss Chemical Society, 2013, Lausanne, Switzerland

Poster presentations

- European Fourier Transform Mass Spectrometry (EFTMS) Workshop, 2010, Lausanne, Switzerland
- American Society for Mass Spectrometry (ASMS) Conference, 2010, Salt Lake City, USA
- Fall meeting of the Swiss Chemical Society, 2010, Zurich, Switzerland
- Swiss Proteomics Society (SPS) PhD student Symposium, 2010, Basel, Switzerland
- American Society for Mass Spectrometry (ASMS) Conference, 2011, Denver, USA
- HUPO World Congress, 2011, Geneva, Switzerland
- Fall meeting of the Swiss Chemical Society, 2011, Lausanne, Switzerland
- American Society for Mass Spectrometry (ASMS) Conference, 2012, Vancouver, Canada

Technical skills

- Work on FT-based mass spectrometers, FT-ICR and Orbitrap; familiar with time-of-flight and ion trap mass spectrometry; top-down analysis of large proteins
- Protein purification and/or separation techniques: liquid chromatography (HPLC, FPLC); SDS-PAGE; dialysis
- Sample preparation for bottom-up proteomics: protein enzymatic digestion (trypsin, Lys-C etc), reduction and alkylation
- Immunoglobulin G purification and digestion with different enzymes
- Working knowledge of analytical techniques acquired during didactic workshops: spectroscopy (IR, UV, ^{1}H-NMR and ^{13}C-NMR); genomics (DNA extraction and purification, PCR and DNA chip preparation); solid and solution-phase synthesis of peptides

Personal skills

- Fluent in English, basic knowledge of French and German
- Supervisor of Master and Bachelor students
- Ability to work in an international research team and in a multidisciplinary environment