UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Analysis of protein complexes *in vivo* by native mass spectrometry

Margarida Gonçalves Isidro

Mestrado em Bioquímica
Bioquímica Médica

Dissertação orientada por:

Doutor Carlos Cordeiro
Doutora Marta Sousa Silva

2020

Acknowledgements

First, I would like to thank my supervisors. Prof. Carlos Cordeiro, for the opportunity of joining his lab and for the knowledge shared on a wide variety of topics, for all the guidance throughout this year, for showing me the good and the bad of working in a scientific research environment with sensible instruments and for the ensure of our security in pandemic times. Prof. Marta Sousa Silva for teaching me so much and always being available to answer my questions in such a great work environment, for the support, help and revision of all my progress reports, for so very thorough corrections to this thesis and for being a driving force for its improvement. And thanks for sharing the stories of our four-legged friends.

I would also like to thank my colleagues, Francisco Traquete, Mónica Soeiro and Mariana Louro. To Francisco for discussing several theories and helping me during the writing of the thesis. To Mónica for make me company during lunch hours and to Mariana, for being my successor, for accompanying me in this last phase and for having supported and helped me in the most difficult parts.

A special thanks to my friends, Bárbara Nobre, Mariana Marques and Pedro Ribeiro, who have listened my problems and support me in this experience. Thank you for being such great friends, without you my academic experience wouldn't have been the same!

I also want to express my deepest grateful to my best friend Sara Maciel, for being with me since the 10th grade, for knowing all my secrets and always help me, for continuing to have patience for every talks.

To my partner in crime, who was always there when I needed, for all the love, patience and understanding when I couldn´t do anything because I was dedicated to my work.

My greatest thanks are for the most important people in my life, my parents, for making this opportunity possible and for always believing in me and who always worked to give me everything I needed. For having the greatest patience in the world and for calming me down. And for my sister, for being the best and giving me the best gift in the world, my nephews.

Work supported by the Portuguese Mass Spectrometry Network (LISBOA-01-0145-FEDER-022125), and by the Project EU_FT-ICR_MS, funded by the Europe and Union's Horizon 2020 research and innovation programme under grant agreement nr. 731077.

Resumo

Conhecer a estrutura de uma proteína é uma das características mais importantes para perceber qual a atividade biológica da mesma. Sendo a estrutura tridimensional de uma proteína uma das características mais importantes para compreender a sua função, várias técnicas foram desenvolvidas ao longo do tempo para obter informação sobre a mesma. Entre estas destacam-se a cristalografia de raio-X, a ressonância nuclear magnética e a cryo-EM. Contudo, a necessidade de utilizar uma grande quantidade de amostra e existência de limitações para complexos quaternários de elevada massa molecular eram algumas das desvantagens destas técnicas. No entanto, nos últimos anos, a espetrometria de massa tem vindo a tornar-se uma abordagem cada vez mais usada em biologia estrutural e molecular, devido ao grande avanço tecnológico dos instrumentos utilizados. O desenvolvimento de técnicas de ionização suaves e o surgimento do nano *electrospray* permitiu que pequenas quantidades de amostra fossem analisadas através desta técnica, trazendo muitas vantagens em comparação com as técnicas disponíveis para o estudo da estrutura de uma proteína. Com o aparecimento da espetrometria de massa nativa passou a ser possível a análise de complexos proteicos intactos e, com esta abordagem, informações como a massa molecular, a estequiometria e a topologia dos complexos proteicos passou a ser imediatamente conhecida.

Uma das características da espetrometria de massa é a necessidade de purificação da amostra para que esta pudesse ser analisada e produzisse um espetro claro. No entanto, os tampões utilizados durante a purificação de complexos proteicos contêm sais e agentes solubilizantes incompatíveis com espetrometria de massa. Para além disso, durante o processo de purificação, o complexo passa por vários passos que podem levar à perda das ligações não covalentes e à alteração da conformação do mesmo. A purificação é também uma desvantagem no estudo de complexos proteicos, uma vez que não têm em conta o efeito de *crowding* macromolecular que controla as interações entre proteínas na célula. Assim, a purificação limita a informação obtida sobre estas interações. Contudo, recentemente foi desenvolvido um método que permite a análise de proteínas diretamente do extrato celular, sem necessidade de purificação. Para tal, é necessário que a proteína seja sobre expressa e a sua abundância relativa seja superior às proteínas endógenas para que o sinal que produz seja dominante no espetro de massa. Adicionalmente, para estudos estruturais é necessário usar a ionização por nanoESI, ainda mais suave e eficiente que o ESI mas cuja implementação é muito variável entre diferentes espetrómetros de massa e requer considerável otimização para o estudo de proteínas no seu estado nativo.

Deste modo, o estudo de complexos proteicos no estado nativo a partir do extrato celular através de espetrometria de massa exige a necessidade de ajustar parâmetros para ser possível ionizar a proteína, mas não destruir as ligações não covalentes que mantêm o complexo proteico. Deste modo, o objetivo principal do trabalho é a otimização de um método que permita melhorar a eficiência de ionização e a transmissão de complexos proteicos diretamente de um extrato celular crude e que possa ser usado em diferentes espetrómetros de massa. Para tal foi necessário o desenvolvimento de capilares adequados, que maximizassem a ionização de complexos proteicos, a otimização da expressão proteica e da preparação da amostra e por fim, o ajuste dos parâmetros para a transmissão de iões com *m/z* elevado num Q-TOF e num FT-ICR. Estes dois espetrómetros de massa partilham algumas características comuns ao nível da sua ótica de iões, mas têm características muito distintas entre si e que podem ser exploradas no âmbito de estudos estruturais. Como modelo foi usado o complexo proteico glioxalase I de *Leishmania infantum*, uma vez que é um homodímero, cujas interações são mantidas por ligações não covalentes, e a sua estrutura determinada por cristalografia é conhecida.

Para otimizar a transmissão do sinal e melhorar o espetro dos complexos proteicos, um dos aspetos mais importantes é o processo de ionização. A forma do capilar utilizado influência a qualidade do espetro, uma vez que controla o fluxo. Quanto menor for o diâmetro da ponta do capilar, melhor a eficiência de ionização. Assim, foram produzidos diversos capilares, tendo sido obtido a variação da corrente iónica e os respetivos espetros. Como amostras foram utilizadas duas proteínas, a mioglobina, uma proteína monomérica com cerca de 17 kDa e o álcool desidrogenase, uma proteína tetramétrica com cerca de 147 kDa, ambas no estado desnaturado e nativo. Durante o estudo com o álcool desidrogenase foi possível acompanhar a formação da estrutura tetramérica, uma vez que ao longo do tempo se viu a diminuição da intensidade da espécie monomérica e o aumento da espécie tetramérica. O facto de não se detetar a espécie dimérica sugere que a cinética de formação desta espécie é demasiado rápida para ser detetada por espetrometria de massa com o método utilizado.

Para otimizar a qualidade do espetro, as condições de expressão da amostra foram também otimizadas. Foram testados dois meios de cultura, LB e TB. O meio TB, sendo um meio nutricionalmente rico foi capaz de induzir a expressão da proteína, mesmo sem a adição de indutor e a quantidade produzida foi suficiente para a proteína dominar o espetro. A temperatura de expressão também foi otimizada e observou-se que a expressão foi maior quando o crescimento foi induzido a 28ºC, temperatura a qual ocorre a expressão no parasita, em comparação com o crescimento a 37ºC.

Assim, para a otimização dos parâmetros que permitem o estudo dos complexos proteicos a partir do extrato celular nos espectrómetros de massa, a expressão da proteína foi induzida a 28ºC em meio TB. Apesar de o Q-TOF ser um espetrómetro de massa modificado para a transmissão de iões com elevado *m/z*, parâmetros como a temperatura da fonte e a voltagem da energia de colisão foram otimizados para melhorar a transmissão de complexos proteicos a partir do extrato celular. A temperatura demonstrou não influenciar a conformação da proteína, não havendo diferença na distribuição dos estados de carga obtidos com a temperatura da fonte a 37º e a 100ºC. Uma voltagem de 30 V demonstrou ser a que melhora a qualidade do espetro, uma vez que diminui a intensidade de outros picos não correspondentes à proteína, aumentando a intensidade do sinal pretendido. Por sua vez, uma energia de colisão demasiado baixa como 10 V, levou a um aumento da intensidade de outros picos e uma voltagem demasiado alta levou à alteração da conformação da proteína, uma vez que ocorre o aparecimento de estados de carga superiores. No entanto, 100 V não foi suficiente para destruir as ligações não covalentes, uma vez que a espécie monomérica não apareceu no espetro. Com estes resultados foi possível concluir que a elevada temperatura da fonte não foi capaz de destruir as ligações não covalentes, produzindo o melhor espetro quando combinada com uma energia de colisão de 30 V.

Para a análise do complexo proteico através do FT-ICR foi utilizada uma fonte de nano-electrospray em desenvolvimento. Os espetros obtidos apresentavam uma distribuição dos estados de carga menores e mais ruído, quando comparados com os obtidos por Q-TOF, devido à presença de aductos típicos de espetros nativos. No entanto, foi possível observar a presença do dímero e não foram detetados picos correspondentes ao monómero, indicando que não ocorreu dissociação durante o processo de ionização, embora tenha sido usada uma energia de colisão de 30 V, que melhorava a qualidade do espetro.

Demonstrou-se ser possível estudar a estrutura de uma proteína a partir do extrato celular por espetrometria de massa nativa, sem necessidade de purificação, preservando as condições do complexo no interior da célula, tanto no Q-TOF previamente modificado para transmissão de iões de elevado *m/z*, como no FT-ICR através do ajuste de parâmetros que permitem a transmissão e análise de iões de elevado *m/z*.

Palavras chave: Complexos proteicos; extrato celular, espetrometria de massa nativa; Q-TOF; FT-ICR

Abstract

Knowing the structure of a protein is one of the most important characteristics to understand its biological activity. However, when a protein is analysed in a purified form there are a lot of information that is not obtained. The crowded intracellular environment is essential in the control of protein interactions and all the information about these is lost when a protein pass through a purification system. In this way, characterization of intact protein complexes must be done with the minimum change in the environment of the protein. This goal can be achieved using the analysis of protein direct from crude cell lysate, without purification steps, by native mass spectrometry. For this it is necessary that the protein is overexpressed and its relative abundance is superior to the endogenous proteins so that the produced signal dominates the spectrum.

To study protein complexes and protein interactions directly from crude cell lysate in Q-TOF and FT-ICR mass spectrometers, instrumental workflow optimisation is needed. So, the main objective is the optimisation of a method that improves the efficiency of ionisation and transmission of protein complexes from a cell extract in different mass spectrometers. For that, it was necessary the development of adequate emitter tips, which maximize the ionisation of protein complexes, the optimisation of protein expression and sample preparation and tunning parameters from high *m/z* ions transmission. As a model was used glyoxalase I from *Leishmania infantum*, since it is a homodimer with non-covalent bonds and the structure determined by crystallography is known.

The ionisation process is one of the most important aspects in optimising transmission of protein complexes. The shape of the emitter tip used influence the quality of the spectrum, smaller the tip diameter, better the efficiency. Several emitter tips were produced, having obtained the variation of ion current over time and the respective spectra. Myoglobin and alcohol dehydrogenase were used as samples to test the efficiency of ionisation. To optimise the expression conditions, two medium cultures (LB and TB) and two temperatures (28ºC and 37ºC) were tested. The conditions that maximized expression were the growth in TB medium at 28ºC. With this condition it was possible to obtain enough protein to dominate the spectrum, maximizing the ionisation and transmission of protein complex. Although the Q-TOF is a mass spectrometer modified for high *m/z* transmission, parameters such as source temperature and collision energy showed changing the quality of spectrum, having been observed that source temperature of 100ºC and 30 V of collision energy improved the transmission of protein complexes, maintaining their structure. For the characterization of complex in FT-ICR mass spectrometer, a source under optimisation was used. Although a lower distribution of charge states was showed when compared with Q-TOF, it was possible to observe the presence of the dimer without dissociation, even though collision energy of 30 V was used.

With this work, it was possible to characterize the structure of a protein complex from the cell extract by native mass spectrometry using two different mass spectrometers, without the need of purification, preserving the conditions of the complex inside the cell.

Keywords: Protein complexes; cell extract; native mass spectrometry; Q-TOF; FT-ICR

Table of Contents

List of figures

List of Tables

List of Symbols and Abbreviations

ADH – alcohol dehydrogenase (EC 1.1.1.1)

CAD – collisionally activated dissociation

CCS - collisional cross-sections

CID – collision-induced dissociation

CRM – charge residue model

CSD – charge-state distribution

DTT – dithiothreitol

DC – direct current

ECD – electron capture dissociation

*Ec*GLO1 - glyoxalase I from *Escherichia coli*

EDTA - ethylenediamine tetraacetic acid

EM – electron microscopy

ESI – electrospray ionisation

ESI-MS – electrospray ionisation mass spectrometry

FT-ICR - Fourier transform ion-cyclotron resonance

FT-ICR-MS – Fourier transform ion-cyclotron resonance mass spectrometry

GLO1 – glyoxalase I (EC 4.4.1.5)

GLO2 – glyoxalase II (EC 3.1.2.6)

GSH – glutathione

HDX - hydrogen/deuterium exchange

HisGLO1 - glyoxalase I with His-tag

His*Li*GLO1 - glyoxalase I with His-tag from *Leishmania infantum*

*Hs*GLO1 - glyoxalase I from *Homo Sapiens*

ICR-MS - ion-cyclotron mass spectrometry

IEM – ion evaporation model

IPTG - isopropyl β-D-thiogalactopyranoside

I.D – inner diameter

LB – Luria-Bertani

*Li*GLO1 – glyoxalase I from *Leishmania infantum*

*Lm*GLO1 – glyoxalase I from *Leishmania major*

MALDI – matrix-assisted laser desorption/ionisation

MDa – megadalton

MHz – megahertz

MPD – 2 – methyl-2,4-pentanediol

MS – mass spectrometry

MW - molecular weight

m/z – mass-to-charge ratio

NanoESI – nano electrospray ionisation

NaTFA – sodium trifluoroacetate

NMR - nuclear magnetic resonance

OA – orthogonal acceleration

OD – optical density

O.D – outer diameter

PBS – phosphate buffered saline

PTM – post-translational modifications

Q-TOF – quadrupole time of flight

RF – radio frequency

SDL-GSH – S-D-lactoyglutathione

SDS- PAGE - sodium dodecyl sulphate -polyacrylamide gel electrophoresis

S/N – signal-to-noise ratio

TB – Terrific Broth

TOF- time of flight

TP – turbo-molecular pumps

UHV – ultra high vacum

1. Introduction

1.1. Protein Structure

In 2001, Charles Tanford and Jacqueline Reynolds, both distinguished scientists, wrote a book entitled "Nature's Robots". This book tells us the story of proteins and the first pages are used to explain the origin of the book name - **"***robots are automatons—you don't need to tell them what to do, they already know"***,** as well as proteins. For every task in a living organism, for every little step in every task, there is a protein designed to carry it out and every protein knows what to do. Proteins are amazingly versatile molecules as they transmit signals in the body, they identify and kill foreign invaders and are responsible for the processes and reactions that keep us alive [1]. Therefore, it is essential to understand the mechanism and function of these natural robots. Knowing the sequence and structure that allow the protein to have biological activity, in addition to knowing the function and interactions that proteins establish with each other, is essential to understand the complexity of biological processes.

The three-dimensional structure of a protein is an important part of understanding how the protein functions. A protein can be classified based on its primary, secondary, tertiary or quaternary structures, as exemplify in Figure 1.1. A description of all covalent bonds, mainly peptide and disulfide bonds, linking amino acid residues in a polypeptide chains is denominated primary structure, being the sequence the most important element. The secondary structural element is defined as particularly stable arrangements of amino acids residues giving raise to recurring structural patterns. Tertiary structure describes all aspects of the three-dimensional folding of a polypeptide. The last one, the quaternary structure, can be defined as its interactions and associations with other macromolecules and ligands that defined its biological function, so most proteins require a specific three-dimensional structure, hence, when a protein has two or more polypeptide subunits, their arrangement in space is denominated as quaternary structure [2].

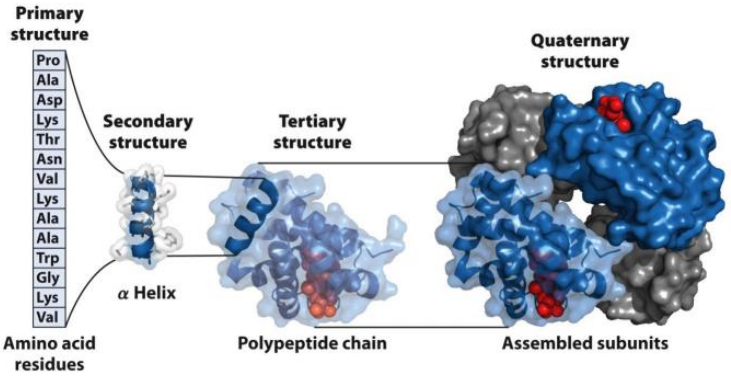


Figure 1.1 Levels of protein structure. Many proteins have multiple polypeptide subunits and these association can serve a variety of functions. Proteins with more than one subunit are also known as multimers. These can be formed by nonidentical subunits and the overall structure of the protein can be asymmetric. However, the majority of multimers have identical subunits or repeating groups of nonidentical subunits in symmetric arrangements. These protein-protein interactions occur at almost every level of cell function (Adapted from Nelson and Cox, 2008).

Proteins possess a relationship between their structure and function, with folded protein structures generating recognition motifs for the binding of other proteins and ligands. The protein folds into a particular three-dimensional shape, determined by its amino acid sequence and stabilized primarily by non-covalent interactions. Although the final shape of the folded protein is dictated by its amino acid sequence, the folding may be aided by molecular chaperones. The precise three-dimensional structure,

or native conformation, of the protein is crucial to its function. The association between proteins and other macromolecules can form supramolecular complexes such as chromosomes, ribosomes and membranes [3].

Although protein sequences carry all necessary information for achieving their native conformation, accurate folding and self-assembly also require the right cellular environment taking account pH, ionic strength, metal ion concentrations and other parameters [4]. One physicochemical phenomenon responsible by the attraction of macromolecules to each other at high concentrations is known as macromolecular crowding. When two macromolecules contact each other, the pressure of smaller molecules will maintain their proximity to each other; there are molecules that will impact them from the outside, but there is no force that would push them from the inside or pull them in the outward direction [5–7].

Within a cell, macromolecular crowding is generated by large biomolecules, such as proteins, and nucleic acids. It has been proposed that macromolecular crowding serves as a driving force for the assembly of large macromolecular complexes and nuclear compartments such as nuclear bodies. The effect of macromolecular crowding on reaction rates and equilibrium constants may arise from change of reactants' mobility, depletion interaction (excluded volume effect) that results in attraction between reactants and non-specific chemical interactions with crowders. Non-specific repulsive interactions that occur as a result of minimizing surface exposure, such as during protein folding and association, lead to a stabilizing effect and these can enhance the effects of volume exclusion. However, non-specific attractive interactions of a protein with the surroundings can have quite the opposite effect and result in destabilization, which can completely negate the stabilizing effects of volume exclusion [8,9].

In this way, within the crowded intracellular environment, individual proteins are constantly coming into physical contact with other proteins and biological macromolecules. There is a huge diversity in the frequency, specificity, and duration of these interactions. Their identification provides a global picture of cellular functions and biological processes.

The accurate identification of the set of interacting proteins in an organism is useful for deciphering the molecular mechanisms underlying biological functions, as well as for assigning functions to unknown proteins based on their interacting partners. Proteins can have structural roles, like the virus coat proteins, or regulatory roles in which the binding of small molecules may affect the interaction between subunits, causing changes in the protein's activity. For example, in the yeast proteome, approximately 70% of the more than 4000 proteins identified by mass spectrometry have protein interaction partners [10].

A large fraction of proteins forms long-lived homomeric – a protein complex formed from self-assembly of a single type of a subunit - or heteromeric assemblies – a protein complex formed from multiple distinct subunits - and have well-defined molecular functions [2]. On the other hand, cells are full of many short-lived and promiscuous interactions due mostly to intracellular crowding. However, many transient interactions are of crucial functional importance, particularly in cell signalling [11]. Knowledge of protein structure can help to interpret results from interaction discovery experiments.

Structural biology elucidates structures of biomolecules such as proteins, nucleic acids, and their complexes to understand the mechanism of their function with atomic detail. Structural biology methods provide information on protein structure in macromolecular assemblies.

Three main experimental methods have been used to determine protein structure complexes: x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (EM).

X-ray crystallography uses diffraction patterns generated upon irradiation of a crystallized protein with X-rays to construct 3D structures. Proteins are dynamic molecules, exhibiting structural heterogeneity and discrete conformational substates, often of functional importance. In protein structure determination by X-ray crystallography, the observed diffraction pattern results from the scattering of X-rays by an ensemble of heterogeneous molecules, ordered and oriented by packing in a crystal lattice. There is no size limit on the proteins studied using this technique, and the majority of large complexes known to atomic resolution (i.e. where molecular details of the interaction are known) have been solved by this method. However, it is often difficult to obtain sufficient material for crystallization, and large complexes require high quality crystals for diffraction at high resolution. For example, ribosome structure took 25 years to be obtained [12,13].

Nuclear Magnetic Resonance (NMR) spectroscopy is the second most common method for protein structure determination and measures transitions between different nuclear spin states within a magnetic field, which provide information about distances between atoms within a macromolecule. A distinct advantage of NMR is the determination of protein structure in solution and even inside de cells, if the protein is the most abundant one. Currently the technique is limited to proteins of up to 40 kDa, meaning that atomic structures for large complexes cannot be determined, although it has an increasing role in studying interaction interfaces between structures determined independently. Most NMR structures deal with single polypeptide chains. Therefore, more monomeric proteins had their structures solved by NMR than protein complexes because smaller proteins are much easier to characterize structurally than larger proteins. It is also conceivable that less consideration is given to correctly assigning quaternary structure during NMR structure determination, so some of the monomeric NMR structures might actually exist as homomers in solution [14].

Cryo-electron microscopy provides 3D structural information by imaging non-crystalline specimens of frozen biological molecules and assemblies kept at cryogenic temperatures. Techniques for averaging particles observed after staining have existed for many years. More recently, cryo techniques, where samples are fast-frozen before study, have provided structures with 2-4 Å resolution range on a regular basis and are then used to build atomic models [15,16]. In addition, new methods of tomography are increasingly providing shape information by combining multiple images of a single particle. For some EM images it is possible to determine the interaction interfaces between subunits for which individual structures [3,17].

The development of accurate and reliable methods for identifying proteins interactions has very important impacts in several protein research. These methods have similarly originated from condensed matter physics and have been gradually adapted to the needs of structural biology.

1.2. Mass spectrometry as a tool to study protein structure

In recent years, mass spectrometry (MS) has emerged as a powerful method to study proteins and their structure. Its speed, accuracy and sensitivity are unmatched by conventional analytical techniques. There has been an explosion in instrumentation, particularly in the field of protein and peptide analysis, and its applications to the study of large biomolecules.

1.2.1. Soft ionisation techniques

The principle of mass spectrometry is the measurement of the mass-to-charge ratio (*m/z*) of gas-phase ions. Only a couple of decades ago, the method was limited to relatively low molecular mass volatile molecules. However, the development of soft ionisation techniques like matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) allowed the extension of mass spectrometry to several fields, such as biochemistry, biotechnology, pharmacology, microbiology, proteomic and genomic.

The ionisation techniques mentioned above have made it possible to ionise large thermally labile biomolecules and transfer them to the gas-phase without dissociation. Non-covalent complexes of biomolecules can nowadays be analysed in the megadalton (MDa) range [18,19].

As mentioned, mass spectrometry measures the *m/z* of ionised molecules in the gas-phase. Hence, the analytes need to be transferred and ionised to the gas-phase prior to analysis. The ionisation technique known as **matrix assisted laser desorption/ ionisation (MALDI)** was developed nearly simultaneously by two research groups. First, by Tanaka and co-workers in Japan [20] and then by Karas and Hillenkamp in Germany [21]. This revolutionary approach has provided a unique opportunity to apply mass spectrometry to the analysis of proteins and other biomolecules with masses in excess of 200 kDa. MALDI is initiated by mixing the sample solution with a large molar excess of the host matrix material and depositing the mixture on a specially designed MALDI sample target. After evaporation of the solvent, the sample-matrix crystals are irradiated with a laser beam of high irradiance power and short pulse widths to simultaneously desorb and ionise the sample and matrix molecules into the gas-phase. MALDI produces mainly singly charged ions, and this feature means it is excellently suited for analysis of complex biological mixtures such as protein digests [22,23].

Electrospray ionisation (ESI) is the most popular ionisation technique for the study of biomolecules and is applicable to a wide range of liquid-phase samples. This technique was developed by John Fenn [24]. As the name implies, electrospray ionisation is a process that produces a fine spray of highly charged droplets under the influence of an intense electric field. Evaporation of the solvent converts those charged droplets into gas-phase ions. The sample solution in a suitable solvent mixture flows continuously through a stainless-steel capillary tube whose tip is held at a high potential with respect to the walls of the surrounding atmospheric-pressure region, called a *counter-electrode*. The potential difference between the tip of the capillary and the *counter-electrode* produces an electrostatic field that is sufficiently strong to disperse the emerging solution into a fine mist of charged droplets. During the process of droplet evaporation, some of the dissolved ions are released into the atmosphere. The resulting ions are transferred from the atmospheric-pressure region to the high vacuum of the mass analyser [23].

When a protein becomes ionised by ESI, it produces a series of consecutive ion signals at different *m/z* values, forming a Gaussian distribution, also called the charge state envelope or charge-state distribution (CSD).

Once electrospray ionisation produces multiply charged ions, being observed many different charge states, it is a very suitable method for combining chromatographic methods because the electrospray process transfers ions from the solution phase to the gas-phase at atmospheric pressure [22].

This technique must be recognized as the first method successfully used to study intact protein complexes by mass spectrometry. The first measurements of proteins in their folded state were carried out by ESI-MS in the early 1990s [25–28].

In principle, there are two theories explaining the formation of protein ions from solution. The first one, the ion evaporation model (IEM), is mainly thought to account for generating ions from small compounds [29]. The second, the charged residue model (CRM) is generally accepted to account for protein complex ion formation. According to this model, the solvent will start to evaporate from the droplet. At the point where charged droplet reaches the point of Coulombic repulsion, the droplet will undergo fission. This process repeats until most solvent molecules are evaporated, creating multiple charged ions containing a single protein complex [30].

1.2.2. Mass analysis and ion detection

When the ionised molecules have been transferred to the gas-phase, a mass analyser can be employed to determine their *m/z*. These can be divided into two classes. The first class, electric-field mass analysers, consist of the quadrupole mass filter, the quadrupole ion trap (Paul trap) and the time of flight (TOF) mass analyser. The second class, magnetic-field mass analysers, comprise magnetic sectors and ion cyclotron resonance mass analysers.

The quadrupole instrument is one of the most widely used types of mass spectrometer. This instrument consists of four precisely matched parallel metal rods. The mass separation is accomplished by the stable vibratory motion of ions in a high frequency oscillating electric field that is created by applying direct current (DC) and radio frequency (RF) potentials to these electrodes. Under a set of defined DC and RF potentials, ions of a specific *m/z* value pass through the geometry of quadrupole rods. A mass spectrum is obtained by changing both the DC and RF potentials while keeping their ratio constant [31].

The quadrupole ion trap it is a three-dimensional analogue of a quadrupole mass filter and it is also called three-dimensional ion trap to distinguish it from the two-dimensional ion trap. Mass separation is achieved by storing the ions in the trapping space and by manipulating their motion in time rather than in space. This task is accomplished with an oscillating electric field that is created within the boundaries of a three-electrode structure. The mass spectrum is acquired by changing the applied RF field to eject ions sequentially from the trapping field [32].

The TOF is a pulsed analyser and separates ions based on the difference in the time various ions take to reach the detector in a field-free drift tube. The flight tube operates under vacuum, free of electrical fields, which is important, as now the velocities of the ions do not change during their flight time. Because the most TOF analysers work by accelerating a set of ions with a constant kinetic energy in an orthogonal direction compared to the ion beam axis, the velocity of the ions is also independent of their previous velocity. At the start of a pulse, all ions gain the same kinetic energy, but due to their different *m/z* ration, they reach the detector at difference times [33].

The magnetic sector instruments are the oldest type of mass spectrometers. A magnetic sector separate ions of different *m/z* values via momentum dispersion and directional focusing. When ions of different *m/z* but identical kinetic energy enter the magnetic field, they travel in different circular paths. Ions of higher *m/z* value move in a large trajectory and ions of lower *m/z* value move in a lower trajectory and thus are separated from each other. Currently, there are two types of magnetic-sector instruments, single and double focusing. Simple-focusing are constructed with only a magnet field, whereas in double-focusing sector instruments, an electrostatic analyser is incorporated along with a magnetic analyser [23].

Fourier transform ion-cyclotron mass spectrometry (FT-ICR-MS) is also an ion-trap. It is based on the measurement of the frequency of ion rotation in the magnetic field – cyclotron frequency. Ion *m/z* ratio is inversely proportional to the cyclotron frequency and it is possible to determine the *m/z* ratio if we know this frequency and the magnetic field strength. This technique demonstrates the highest mass resolution and mass measurement accuracy, in excess of 20000000 resolving power and a mass accuracy in the tens of ppb. The main part of the FT-ICR spectrometer is a measuring cell which is a Penning trap. In this cell ions get trapped by electric field along the magnetic field vector and by magnetic field in the perpendicular direction. By applying RF electric field containing resonant frequency in the plane of ion cyclotron motion we are exciting their synchronous cyclotron motion. Rotating ions induce image charges in detection electrodes resulting in the current in electrical circuit of a signal amplifier [34]. More than 10^6 ions can be trapped in the ion trap and their cyclotron frequencies can be measured simultaneously [10].

The choice of mass analyser will reflect the needs of the specific application, and will take mass accuracy, resolution, sensitivity and cost into account. Two or more analysers can be combined to form a tandem mass spectrometer, such as a triple quadrupole or a hybrid instrument (consisting of two different analysers), e.g. quadrupole-TOF (Q-TOF) or magnetic sector-TOF.

1.2.3. Tandem Mass Spectrometry

In many cases the mass spectrum alone does not yield sufficient information on the investigated biomolecule. This applies, in particular, for larger proteins, as the peaks broaden and binding to small ligands cannot necessarily be identified accurately. To obtain information about ligand stoichiometry or complex stability, tandem mass spectrometry is applied (MS^2 or MS/MS). The general procedure involves the isolation and subsequent activation of a precursor ion species and finally the detection of the product ions. The activation of the ion can be induced either in small steps (low energy activation) or in a single fast event (high energy activation). Multitude tandem MS techniques are available. Often, two or more techniques are combined to obtain complementary data, and thus increases the information content.

Tandem mass spectrometry has attained an enviable status as an analytical tool to identify and quantify compounds in complex mixtures. MS/MS refers to the coupling of two stages of mass analysis, either in time or space. This concept involves the action of two mass spectrometry systems. The first one performs the mass selection of a desired target ion from a stream of ions produced in the ion source. This mass-selected ion undergoes fragmentation or a chemical reaction in the intermediate region. The second MS system performs the mass analysis of the product ions that are formed in the intermediate step. By convection, the mass-selected ion is called the precursor ion and its fragments are called product ions. Because of the incontrovertible link between the precursor ion and all of its product ions, the molecular specificity of MS/MS approaches is unambiguity. This unique attribute of tandem mass

spectrometry is a highly useful feature that plays a role in the unequivocal identification of a target compound in real-words samples [23].

Tandem mass spectrometry is not restricted to two stages of mass analysis, it is also possible to perform multistage MS experiments, abbreviated as MS^n. These experiments can determine the genealogical relation between a precursor and its ionic products. Multistage MS experiments are performed mostly with ion-trapping instruments [35].

- **Collision-induced dissociation (CID)**

The most common ion activation and dissociation of method is collision-induced dissociation (CID), also known as collisionally activated dissociation (CAD). This technique, introduced in 1968 is a two-step process: collision activation and unimolecular dissociation. In the collision activation step, the fast-moving mass selected precursor ions are excited to higher-energy states by collisions with atoms of an inert gas such as helium, argon or xenon. During this excitation process, a part of the initial translational energy of the fast-moving precursor ion is converted into the excitation energy of the ion to cause it to fragment. The maximum amount of kinetic energy that can be converted into internal energy during collision depends on the masses of the collision partners [35,36].

- **Electron Capture Dissociation (ECD)**

Another technique of fragmentation, specifically used in FT-ICR mass spectrometers is electron capture dissociation (ECD), that is based on the dissociative recombination of multiply protonated polypeptide molecules with low energy electrons. This method of fragmentation provides information on protein sequence and post-translational modifications (PTM). The mechanism of ECD can be described as a polypeptide that capture an electron in a high orbit, which is followed by charge neutralization, leading to an exited radical species that undergoes bond cleavage. The presence of a radical makes this fragmentation highly bond-specific. ECD is useful in top-down sequence characterization, *de novo* sequencing and PTM mapping. However, the loss of signal owing to partial charge reduction is a disadvantage of this method. Whereas collisional excitation leads to polypeptide chain unfolding and losses of labile groups, ECD preserves the secondary structure and the labile groups [37].

1.3. Nanoelectrospray Ionisation (nanoESI)

A miniaturized version of standard ESI, designed to operate at submicroliter flow rates was developed [38,39]. This technique is called nanoelectrospray ionisation (nanoESI). NanoESI is a form of electrospray, with the same fundamental ionisation process of droplet formation followed by multiple uneven Rayleigh divisions, and finally desorption of pre-formed ions from the droplet, but there are practical differences between this mode and ESI. First, nanoESI uses a gold-coated fine-drawn capillary with a spraying tip aperture of only 1 to 3 µm rather than the 100 µm capillary used in a standard ESI source. About 3 µL of solvent is loaded directly into the capillaries. Another difference is related to the stability of the spray that is achieved at much lower capillary voltages (500-1000 V). No external pumping is required to maintain a continuous solvent flow because the solution emerges from the tip itself at the rate of 10 to 50 $nLmin^{-1}$. This process of ionisation is stable at a flow rate that seems to be lower than that possible or practical in a pumped system, probably because of the very small diameter of the spraying capillary. The emerging droplets are small, typically less than 200 nm in diameter, about 100-200 smaller than the volume of 1-2 µm droplets generated by conventional electrospray sources. At concentrations of 1 pmol/µL such droplets contain on average only one analyte molecule [23,40].

The name reflects the low nanoliter per minute flow and the droplet size in the nanometer rather than the micrometer range. To improve ionisation, desolvation and transfer efficiency, the nanoESI capillary is usually mounted within 1 to 2 mm of the nozzle orifice of an ESI source.

The theoretical model for the electrostatic dispersion of liquid in electrospray, described in Electrospray and Taylor-Cone theory, predicts a proportionality between the two thirds power of the flow rate and the sized of droplets produced from the tip of a stable Taylor cone - the cone observed in electrospraying process, as represented in Figure 1.2. Small droplets have several desirable analytical properties like high surface to volume ratio which makes a large proportion of analyte molecules available for desorption [39].

Figure 1.2 The formation of electrospray. Left: at a low voltage, the drop is almost spherical. Centre: with a higher voltage, but below the onset potential (1-2 kV), the drop elongates under the pressure of the charges accumulating at the tip. Right: at onset voltage, the pressure is higher than the surface tension, drop shape changes to a Taylor cone and small droplets highly charged are released. (Adapted from Dass 2007).

With the introduction of nanoflow electrospray the smaller size of the initial droplets facilitates the use of pure aqueous solutions and gentle desolvation conditions such that non-covalently bound complexes are more preserved than when conventional ESI is used.

Within the advantages, it can be highlighted the complete absence of any cross contamination, few opportunities for sample loss, the possibility of long signal averaging to obtain accurate mass measurements in protein mixtures. The stability of the nanoESI source helps in measuring protein masses accurately when protein mixtures and/or low protein concentrations are involved. In this way, the nanoESI source has been shown to possess a number of unique analytical features. These include freedom to choose the solvent in a wide range of composition and pH and an unprecedented tolerance to salt contamination [39].

For ESI-MS analysis, peptide mixtures are usually first separated by reversed phase HPLC and are then detected on-line. This necessity of peptide separation results from the complexity of peptide mixture spectra when convoluted with the charge state distributions of the peptides. Therefore, for the rapid analysis of unseparated mixtures, ESI-MS was not considered a method of choice. However, with the nanoESI source, more than 30 minutes of measurement time is available on 1 μL of sample, opening the possibility of determining the identity of most of the peaks in complex peptide mass map by fragmentation, in simple samples.

The overall efficiency in ESI -MS can be defined as the number of analyte ions recorded at the detector divided by the number of analyte molecules sprayed. It is thus a combination of desolvation, ionisation and transfer efficiencies into the vacuum system. The nanoESI source improves all these factors in comparison to conventional sources [41].

However, there are some disadvantages, since this technique required a very high level of training for handling the glass capillaries such performing the sample loading and adjust the capillary tip in the right place to have a constant signal. On-line coupling with nanoscale separation methods or automation of sample loading is not yet reliably possible – nanoESI still needs a skilled operator. Moreover, commercial spray capillaries vary wildly in quality, even if the same pulling conditions are chosen. NanoESI requires dedicated ion-sources that have either to be manufactured in-house or adapted from commercial products that generally show lower performance. For contaminated samples nanoESI may become a time-consuming procedure, because high-quality results are often obtained after a pronounced initial period in which only noisy mass spectra are acquired [42].

1.4. Native Mass Spectrometry

Mass spectrometry can also be used for determining the mass of intact assemblies, their precise stoichiometry, direct interactions between subunits, the relative position (core vs. periphery) of subunits within an assembly and the strength of inter-subunit interactions, using conditions that preserve these non-covalent interactions. This technique is called **native mass spectrometry (native-MS)** and it is used to analyse samples that vary greatly in mass, degree of flexibility and symmetry. Second, multiple oligomeric states can be analysed simultaneously, providing specific information about each individual species (Figure 1.3).

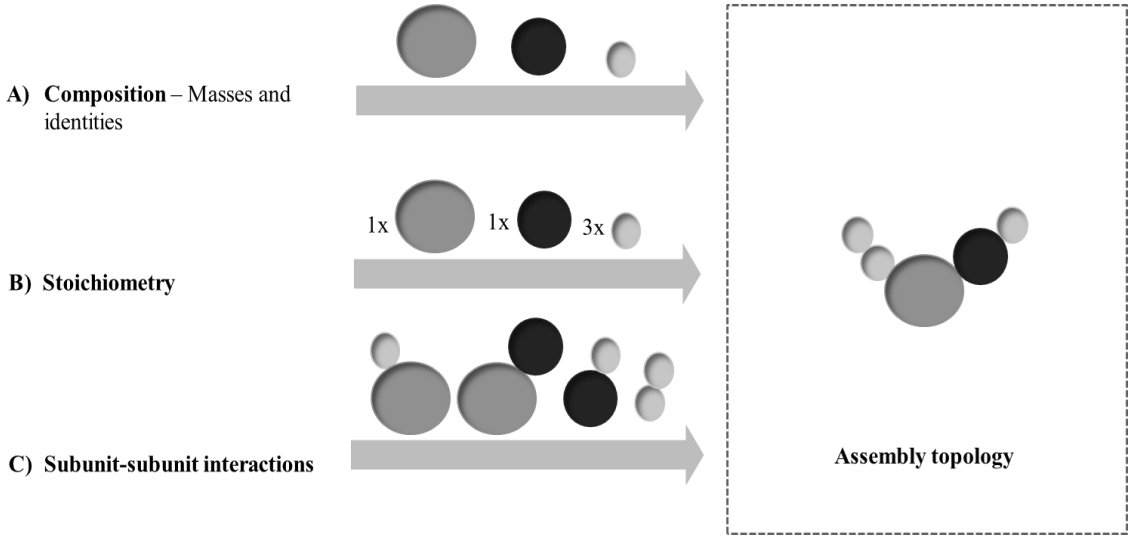


Figure 1.3 Levels of information obtained using MS. Beyond the obtained information on the mass of each subunit, it is also possible to determine the stoichiometry of complexes. Using conditions that preserve non-covalent interactions, native-MS allows the identification of all the interactions between subunits and reconstitute the assembly topology of protein in study.

Native-MS is a robust and reliable tool for the analysis of non-covalent complexes. This label-free analytical technique also referred to as non-covalent, non-denaturing or supramolecular MS, utilizes nano-electrospray to transfer and ionise non-covalent assemblies from solution to the gas-phase. Accurate measurement of the molecular mass of an intact complex provides valuable insights into biomolecule interaction, including stoichiometry, reversibility, site-specificity, affinity and polar contributions to the binding [43].

This method allows the dynamics of quaternary structure to be studied in real time. Besides, native-MS is highly sensitive. In several cases successful analyses required only a few microlitres of sample at relatively low (μM) concentration. Finally, native-MS does not require samples to be chemically labelled or crosslinked, although that may be necessary for some studies. In some cases, the biochemical steps involved in sample preparation may need to be optimised so that the native state of the complex is preserved during buffer exchange. Unlike other types of ESI-MS analysis, neither acidic conditions nor organic solvents are used [43,44].

Native-MS can also provide information on the dynamic behaviour of complexes and assess the presence of distinct forms that could detect allostery. Another exciting application is the possibility to follow the kinetics of a reaction by incubating mixtures of differently labelled forms of non-covalent complexes [45,46].

This technique is a particular approach based on electrospray ionisation whereby the biological analytes are sprayed from a non-denaturing solvent. However, it is important to refer that the term "native" in its purest definition would mean the state of a protein in its natural environment, in the cell and in the vicinity of its natural molecular partners. In this case, native-MS being a gas-phase method certainly cannot live up to that definition, so this expression describes the biological status of the analytes in solution, before to the ionisation. The observation that proteins electrosprayed from aqueous solution exhibited narrower charge state distributions compared with proteins sprayed from organic solvents was used as a proof that native protein structure can be preserved in the gas-phase. Only solvent exposed basic residues could pick a charge. In this way, a less structure or unfolded protein in its most expanded form will pick up more charges during ionisation process than a folded protein [47].

A most remarkable experience using the tobacco mosaic virus was also performed [48]. The virus was electrosprayed from solution and then collected and visualized by transmission electron microscopy. After flight through the mass spectrometer, the virus was structurally viable and capable to infect tobacco plants, proved that the virus capsid structure survived upon the transition in the gas-phase.

Another important finding supporting the ability of MS to mirror the state of a reaction in solution comes from a set of simulation studies. These analyses indicate that the absence of water "freezes" the protein in a conformation not far from the native state, yielding stable gas-phase conformations similar to the structures in solution. However, much care still needs to be taken to avoid over-interpretation of native-MS data since protein conformation can change upon transition from solution into the gas-phase [46].

Nevertheless, native-MS has some drawbacks. MS is performed under vacuum conditions and the detection of macromolecular complexes takes place in the gas-phase. Consequently, the relative abundance of detected complexes may deviate from that in solution because of the distinct ionisation, transmission and detection probabilities of different complexes. Moreover, in the gas-phase hydrophobic and electrostatic interactions become weaker and stronger, respectively, than in solution. This may render the detection of certain assemblies impossible without prior crosslinking. Nevertheless, ionisation efficiency between proteins and peptides is very similar and computational and experimental evidence indicates that the transition from solution to the gas-phase does not drastically alter biomolecules [49].

Moreover, resolution on individual charge states obtained with native-MS are in general much lower than the inherent mass resolution of the used mass analysers. Yet, native-MS has been proven to be a powerful method enabling the analysis of MDa protein complexes, the investigation of subunit arrangement and exchange within non-covalent assemblies as well as the identification of PTMs and ligand binding with great mass accuracy. Native analysis has been applied to a range of analytes with diverse masses, including small proteins, ribosomes and virus capsids. In these analyses, new understanding of their natural form and their non-covalent interactions with proteins, nucleic acids, ligands, cofactors, or metal ions can be gained [50].

Recently, there has been a strong interest in combining native-MS with more traditional top-down proteomics methods [51,52] and x-ray crystallography [53]. These applications and advances are possible due to the native structure of the protein that remains intact when electrosprayed with non-harsh buffer conditions.

The main difference between traditional denaturing and native electrospray ionisation mass spectrometry is the composition and pH of the solution. Denaturing solutions containing organic solvents and non-neutral pH conditions produce unfolded molecules by disrupting non-covalent interactions, causing molecules to unfold and expose sites of protonation. On the other hand, native

solutions with minimal concentrations of organic solvents and a neutral pH (~ 7) preserve non-covalent interactions, including those between subunits of protein complexes, as these molecules transition to the gas-phase [54].

Charge state and intensity during electrospray ionisation are determined by the number and chemical properties of ionizable sites, solvent surface tension, intramolecular interactions, ion surface area, and Columbic repulsion. ESI-MS of denatured species typically produces a normal distribution of highly charged ions in a wider charge-state distribution, usually with a consistent charge density centre around one charge every 9-10 residues. Native-MS produces ions with lower charge states and a narrower charge-state distribution. Largely, this is due to the globular nature of folded proteins, limiting solvent accessible residues to only those on the surface of the protein/complex. In turn, globular, native proteins with less exposed surface residues will have smaller collisional cross-sections (CCS) when compared to their denatured counterparts [55].

Time-of-flight is the mass analyser appropriate and commonly used for measuring the high *m/z* protein assembly ions introduced via native-ESI. The popularity of TOF tubes for high-mass analysis is largely due to their wide mass range, only limited by decreasing detector sensitivity and the fact that there is only a very shallow decline in mass resolution with increasing *m/z*.

The study of a protein complex by MS usually involves two sets of experiments: proteomics-MS and native-MS, which are often performed separately due to sample complexity and technical limitations. Proteomics-MS, either at the peptide level (bottom-up) or at the intact protein level (top-down), allows the identification of individual protein components in the protein complex as well as the mapping of any PTM´s, while native-MS studies the upper levels of protein organization and intact protein complexes, and provides structural information, namely stoichiometry, stability and topology which are complementary to information obtained from conventional biophysical techniques for structural biology [56].

1.4.1. Native-MS of recombinant proteins from crude cell lysates

Originally protein complexes were purified using classical biochemical methods, that was a time-consuming process for low-abundance protein complexes, and they required purification steps to be tailored to individual complexes. The development of epitope tagging and affinity purification techniques, performed under native conditions, has produced vast networks of protein interactions [57].

Until very recently, a prerequisite for acquiring native-MS data was the prior purification of the protein complex. However, most common buffers used during purification of protein complexes contain salts and solubilizing agents such as imidazole, DTT and EDTA that are incompatible with MS, requiring a buffer exchange before analysis [57]. Furthermore, during the multiple-step process of purification, non-covalent association of the protein with ligands and cofactors are likely to be lost.

In 2017, Michal Sharon and co-workers presented a novel method of direct analysis of protein complexes from crude cell lysates using native mass spectrometry, showing that highly resolved native-MS data can be generated, wherein the intact recombinant protein becomes the dominant component in the mass spectra [58]. The high expression levels that are necessary can be obtained in systems like *E. coli*. In this way, native-MS measurement allows the analysis of intact protein assemblies under non-denaturing conditions from the direct crude extract providing structural characterization of the

overexpressed protein and the method reduces the time lag between production and characterization, while preserving the holo-protein forms thus being a better approximation to the *in vivo* conditions.

This simple and rapid method enables to assess sequence variations, molecular weight, folding, stability, topological arrangements and associations of ligands avoiding purification and buffer exchanges procedures that can cause modifications in the protein of interest [58].

The high dynamic range of cellular protein levels is beyond the intrinsic limitations of mass spectrometry sensitivity, resulting in the masking of proteins found at lower abundance by those found at higher abundance. The high protein expression level overlaps the signal of proteins of lower abundance and lead to the detection of multiple proteoforms of the proteins of interest. Direct analysis of crude cell lysate allows monitoring the protein expression and detect modifications in these patterns [58–60].

However, the reduction in the relative abundance of the produced protein at longer growth periods can indicate the generation of a non-soluble fraction, like aggregates or inclusion bodies, which affects the protein ability to ionise, not being detected in native-MS. The relative abundance of background peaks also provides information on the culture state. To control the presence of background peaks that exhibit molecular masses highly similar to that of the overexpressed protein is necessary to acquire a reference spectrum from a sample prepared without induction [60].

To preserve the natural state of protein complexes and to ensure compatibility with native-MS requirements, analysis is generally carried out using millimolar concentrations of aqueous solutions at physiological pH's. Solutions close to neutral and high volatility like ammonium acetate and ammonium carbonate must be used [60].

1.4.2. Improving high *m/z* transmission

To improve high *m/z* transmission, detection and spectra quality in terms of transmission and protein complex desolvation, there are aspects that need to be optimised, like source ion guide sleeves [61], gas-bleed valves in to the source hexapole ion optics, low frequency quadrupoles and lower frequency time-of-flight pusher optics [62].

Control of the pressure gradient in the different pumping stages in a quadrupole time-of-flight mass spectrometer has been found to be essential for the detection of macromolecular particles [63–65]. Conventional tandem mass spectrometers have limitations for studying such complexes, since the first mass analyser is usually a quadrupole with limited *m/z* range. In addition, the different pressure regions are not optimised for the transmission of large macromolecular complexes. The energy input required for the phase transfer should not exceed the dissociation threshold of the complexes. The energy input can be controlled by careful adjustment of the acceleration voltages in the different stages of the instrument. Increased pressure provides improved collisional cooling and focusing of large ions in the quadrupole guides, and therefore, better transmission through the quadrupole and TOF. The analyte undergoes collisions with background gas atoms, by which it is activated and subsequently loses residual, attached solvent molecules. The amount of activation per collision depends on the kinetic energy of the ions, which is determined by the accelerating electric field as well as the average collision frequency with background gas atoms according to the pressures in different stages of the instrument. Raising the background pressure in the initial vacuum stages appears to be a prerequisite for the detection of intact non-covalent complexes.

The low frequency quadrupole extends the *m/z* range of ions that can be selected for tandem mass spectrometry experiments as well as increases the transmission of high-*m/z* species into the time-of-flight analyser. Although the mass range of a TOF is in theory unlimited, quadrupoles can transmit ions only to a certain upper limit, which is dependent on the RF amplitude. The resolution that could be obtained with these instruments was significantly reduced when compared to that of a quadrupole operating at higher RF but with the benefit of a better transmission [62].

Electrospray ions are generated at atmospheric pressure and introduced into the vacuum chamber of the mass spectrometer in a flowing gas stream, usually through a nozzle and a skimmer, with sequential stages of pressure reduction [23]. The energy acquired from the gas jet may have both axial and radial components and may result in various types of transmission losses depending on the particular analyser. For example, in quadrupole mass filters radial oscillations of ions lead to reduced transmission through inter quadrupole apertures, while excessive axial energy will result in reduced resolution. In trapping mass analysers, such as FT-ICR, energy variation with mass may lead to mass discrimination effects during trapping. In TOF instruments with orthogonal injection, ions having excessive axial energy may partially or completely miss the detector [61]. To overcome this problem, Douglas and French, proposed the use of quadruple ion guides at elevated pressures in order to thermalize ions in collisions with gas molecules, usually nitrogen. Collisional cooling results in significant reduction of both axial and radial kinetic energy components. Loss of radial component causes the ions to move to the minimum of the effective potential at the center of quadrupole. Ions are therefore more efficiently transmitted through the exit or inter quadrupole apertures and into the downstream mass analysing quadrupole [66].

If an ion is not cooled well enough in a collisional quadrupole, a poor transmission through apertures connecting the quadrupoles results, while insufficient axial cooling makes ions miss the TOF detector. One of the simplest solutions of this problem is to increase the pressure in ion guide to fit the stopping path within the existing quadrupole length [61].

Studies using a quadrupole time-of-flight shows that ions with MW > 100 kDa cannot be recorded efficiently unless pressure is increased to ~15 mTorr (~ 0.020 mBar). Also, the pressure required for thermalization depends not only on mass but also on the declustering voltages, i.e, the orifice and skimmer potentials, indicating that large ions may acquire significant energy both from the gas jet and from electric field applied in the interface area. Heavy ions that are also compact appear to need additional collisions in order to be fully colisionally cooled and focused in a RF ion guide under standard conditions [61].

The use of FT-ICR is less frequent in the analysis of native-MS but a few groups have demonstrated its utility for large proteins assemblies. However, ion optics and vaccum system components may be modified within the FT-ICR instrument which my afford improved levels of desolvatation such as source ion transfer capillary temperature, source skimmer voltage, collision cell voltage and ICR-cell trapping times. Therefore, it is necessary tunning the mass spectrometers to improve the transmission of the assembly state intact and perform analysis of direct crude cell lysates.

-

1.5. Glyoxalase I from *Leishmania infantum* as a protein complex model

To study protein interactions directly from crude cell lysate and optimise the parameters that allows the maintenance of the native structure intact, the model protein chosen was the enzyme glyoxalase I (S-D-lactoylglutathione methylglyoxal-lyase, EC 4.4.1.5) from *Leishmania infantum* (*Li*GLO1). The glyoxalase I is the first enzyme of the glyoxalase pathway, and in trypanosomatids it uses a trypanothione-dependent hemithioacetal as substrate, as represented in Figure 1.4. The glyoxalase pathway, comprising the glyoxalase I (GLO1) and glyoxalase II (S-2-hydroxyacylglutathione hydrolase, EC 3.1.2.6, GLO2), is considered the main catabolic system of 2-oxoaldehydes, such as methylglyoxal, a compound produced in every organism. Methylglyoxal, a mutagenic, toxic and inhibitor of glycolytic enzymes, is converted to D-lactate using reduced glutathione (GSH) as a specific co-factor. Methylglyoxal reacts non-enzymatically with glutathione, forming an hemithioacetal that is isomerized by glyoxalase I to the thiolester, S-D-lactoylglutathione (SDL-GSH). This thiolester is then hydrolysed to D-lactate by glyoxalase II, regenerating GSH [67,68]. Functionally, trypanothione replaces glutathione in trypanosomatids, being the glutathione-dependent enzymes replaced by functionally analogue enzymes using trypanothione, that being a thiol has two reducing disulfide groups available for reaction, which is equivalent to two glutathione molecules. Hence, is kinetically more favourable as a reductant than GSH [69].

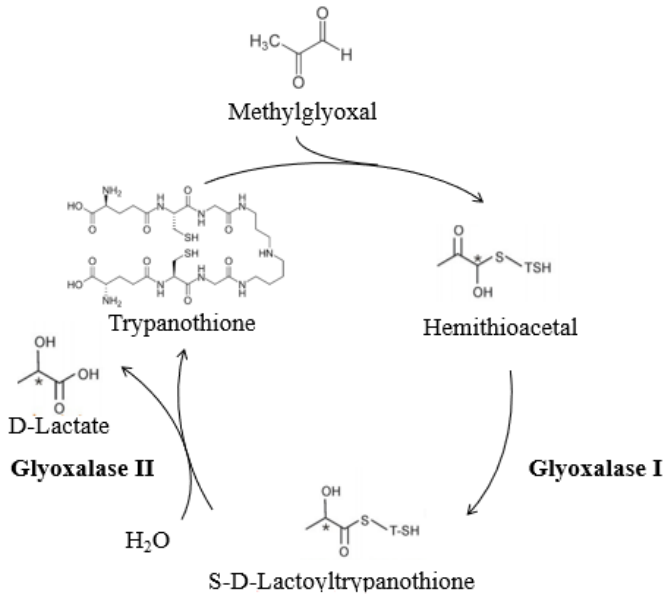


Figure 1.4 Glyoxalase pathway in *Leishmania infantum.* Methylglyoxal which is formed non-enzimatically, reacts also non-enzimatically with reduced trypanothione, forming an hemithioacetal. This hemithioacetal is isomerised by glyoxalase I, forming the thiolester S-D-Lactoyltrypanothione, subsequently hydrolysed to D-lactate by glyoxalase II and regenerating trypanothione.

In trypanosomatids, GLO1 is specific for trypanothione and also catalyses the reaction of the glutathione-methylglyoxal hemithioacetal, while GLO2 is specific for lactoyltrypanothione [70].

Glyoxalase I is present in most organisms, either prokaryotes or eukaryotes. Typically, eukaryotic GLO1 contains Zn^{2+} at the active site and prokaryotic GLO1 contains Ni^{2+} [71–73]. Structural information of GLO1 in different organisms, including *Escherichia coli* (*Ec*GLO1; PDB entries 1f9z, 1fa5, 1fa6, 1fa7and 1fa8)[74], *Homo sapiens* (*Hs*GLO1; PDB entries 1fro, 1qin, 1qip and 1bh5)[73] and

Leishmania major (*Lm*GLO1; PDB entry 2C21)[75], is available through X-ray structure. In most eukaryotes and prokaryotes the enzyme is dimeric but in yeast the enzyme is monomeric [76].

In *Leishmania infantum* GLO1 is formed by 141 residues with 16361 Da (Figure 1.5 A). The asymmetric unit of the orthorhombic crystal form of *Li*GLO1 contains six polypeptide chains, which form three dimers. Each protein dimer has two metal ions and two MPD (2- methyl-2,4-pentanediol) molecules. A total of 285 water molecules were found in the structure [77].

*Li*GLO1 is a homodimeric protein consisting of two subunits, since there are no cysteines in its sequence, the dimer is maintained by non-covalent bonds. Each monomer composed of two βαββββ domains and the residues from each subunit contributing to the active site pocket. An eight-stranded β-sheet with pseudo twofold symmetry emerges due to the dimer formed by the interaction of the first domain of one monomer with the second domain of the other. The structure of *Li*GLO1 (Figure 1.5 B) was calculated by molecular replacement using the related structure of *Lm*GLO1 (PDB entry 2C21). *Li*GLO1 is very similar to GLO1 of *Leishmania major* (*Lm*GLO1), differing only in three residues located at the protein surface: Gln39, Ala50 and Met126 of *Li*GLO1, corresponding to Glu39, Gly50 and Thr126 in *Lm*GLO1 [77].

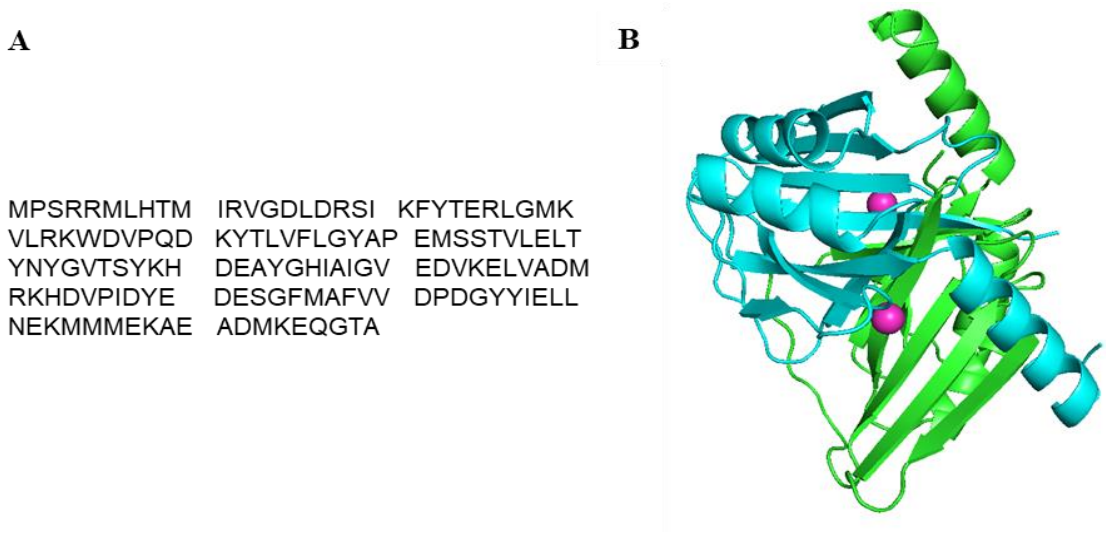


Figure 1.5 Glyoxalase I from *Leishmania infantum*. A) *Li*GLO1 sequence. B) Structure of dimeric *Li*GLOI. One chain is represented in blue and the other in green. Each subunit is composed of two βαββββ domains. The residues from each subunit contribute to the active site pocket. Metal ions are represented in pink. Structure represented by PyMOL [78].

Native-MS measurement of crude cell lysate, will allow the analysis of intact protein assemblies under non-denaturing conditions, then provides structural characterization of the overexpressed protein, avoiding purification steps. Properties such as solubility, molecular weight, folding, assembly state as topological arrangements can be revealed. This approach can also be expanded to address other biological questions such as the strength of protein interactions. In this way, to improve transmission and allow the study of protein complexes there is a set of parameters ranging from sample preparation to detection, which can be optimised.

2. Aims

The main objective of this work was to optimise the instrumental workflow that allows the characterization of protein complexes by native mass spectrometry, including the analysis of protein complexes *in vivo* trough crude cell lysates.

To detect protein complexes and study protein interactions, instrumental parameters need to be carefully optimised to minimize in-source dissociation and non-specific interaction during the nanoESI process. Also, sample preparation is key to the successful detection of weakly bound complexes with adequate signal-to-noise (S/N) values in mass spectra. Challenges for native-MS of intact proteins and protein complexes include developing a protocol to increase the transmission of high *m/z* ions through mass spectrometers. For such, a better understanding of how emitter tips influence protein structure and ionisation and how the solution conditions and protein expression affect the spectrum are needed.

Hence, to optimise the instrumental workflow for the study of protein complexes, the first task consisted in producing adequate emitter tips to increase the transmission of large ions and obtain a clear spectrum. Once produced, these were tested using different proteins and protein complexes, mainly on a well-established platform like a modified Q-TOF for native-MS using a Micromass nanoESI source. Later, the emitters were used on a custom built nanoESI source, under development, for the FT-ICR.

After optimising the emitter tips, it was necessary to develop a method that allowed the study of native protein complexes in crude cell lysate, increasing the transmission of homogenous complexes in Q-TOF and FT-ICR mass spectrometers. As a model, a homodimeric protein was analysed - glyoxalase I from *Leishmania infantum* and different expressing and solution conditions were tested to understand the effect of each one in native spectra and maximize the detection of protein complex in crude cell lysate of *E. coli* overexpressing it.

3. Materials and Methods

3.1. Instrumentation

3.1.1. Q-TOF Ultima II

The Q-TOF is a hybrid quadrupole time of flight mass spectrometer, as represented in Figure 3.1. This instrument utilises a high performance, research grade quadrupole mass analyser and orthogonal acceleration (OA) TOF mass spectrometer.

The quadrupole is operated as an ion transport element for MS experiments and as a mass selection device for MS/MS experiments. A collision cell is situated after the quadrupole to induce fragmentation from the MS/MS experiments. Beyond the quadrupole collision cell assembly is an orthogonal acceleration cell in which the ions are either pushed downward into the TOF analyser or allowed to drift onward into the post acceleration photomultiplier detector. This instrument was modified for native-MS work. Briefly, the quadrupole ion transmission characteristics were enhanced for a high mass range. Pressure control and readback are available for the source region and the collision cell. Detector sensitive was enhanced and maximum collision energy range available was doubled.

The ions that are pushed out of the OA describe either a V or a W path travelling through the reflectron lens at the end of the instrument. The general tuning parameters for V and W mode operation are very different. In W-mode, the pusher rates approximately double due to the increase in flight time of the ions in the TOF chamber. In the V-mode the ion beam has to travel the length of the flight tube twice, whereas in W-mode this is four times. All the samples were analysed in V-mode (Figure 3.1 A).

This instrument is connected to a nanoESI source that use a borosilicate metal coated glass capillary. The nanoflow source consists of a sample sprayer mounted on a three-axis manipulator. The sprayer assembly can be rotated through 90 degrees for ease of access, when changing the glass capillary as shows the Figure 3.1 B.

The ions are then detected with a microchannel plate detector and ion counting system and this output is then recorded as a mass spectrum on the control console using MassLynx software that allows control the instrument. This configuration allows both interactive and fully control of the system.

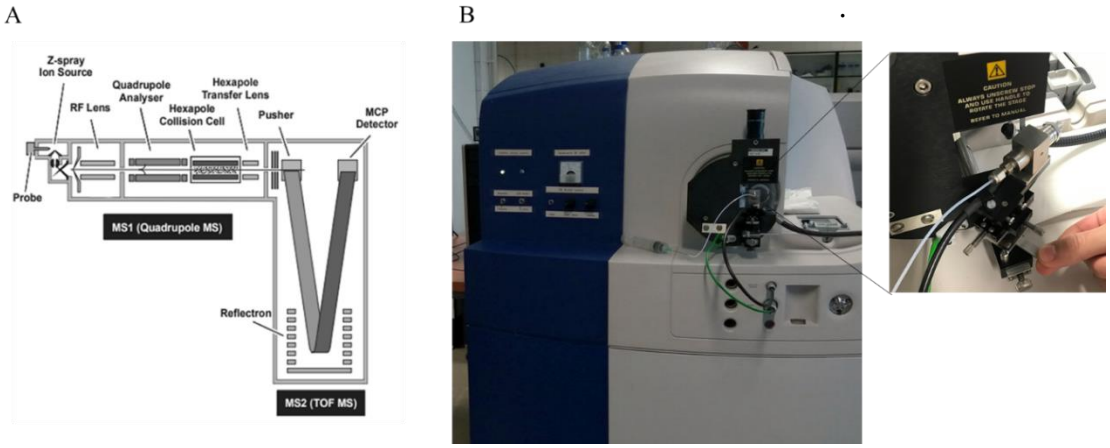


Figure 3.1 Q-TOF mass spectrometer. A) Ion optics of Q-TOF mass spectrometer. The first part consists of a quadrupole and a collision cell, followed by a time of flight (TOF) analyser. B) Modified Micromass Q-TOF Ultima II with a nanoESI source installed. Notice the manual controls available that are absent on standard instruments.

3.1.2. FT-ICR 7 Tesla SolariX XR

The solariX XR series FT-ICR mass spectrometer (Brüker Daltonics, Bremen, Germany) is composed of two major components: the FT-ICR cart and the superconducting magnet of 7 Tesla. The ion path can be divided into 4 sections: the nanoESI source, the Qh-Interface, the ion transfer optics and the detector. Six differential pumping stages maintained by four turbo-molecular pumps (TP1 – TP4) and two mechanical roughing pumps allow the introduction of ions at atmospheric pressure while providing the ultrahigh vacuum of $\leq 10^{-10}$ mbar necessary to operate the analyser, as represented in Figure 3.2 A.

Ions entering the funnel stage are focused into an ion beam using DC and RF voltages. The ions then enter the second funnel in vacuum stage 2 at ~ 0.1 mbar. Here they are decelerated before they experience further cooling in vacuum stage 3 within the source multipole. Before they enter the Qh-Interface, ions can be accumulated in this region for a defined time or simply passed through. Ions are subsequently focused into the mass-selective quadrupole. This allows selection of a desired mass window around a given *m/z* to be passed through into the collision cell where the ions can be stored for a selected time. The collision cell is at an elevated pressure of $\sim 1 \times 10^{-3}$ mbar of argon. If desired, the conditions can be changed to induce fragmentation of the mass-selected precursor ions. After thermalization and, if desired, fragmentation, ions are pulsed out of the collision cell and focused by several DC lens elements. The first of these is equipped with a closure device ("Shutter"), which only opens when ions are transferred towards the analyser. After passing through the beam valve that separates the Qh-Interface from the ultrahigh vacuum (UHV) region ($\leq 10^{-8}$ mbar), ions are directed by the multipole ion guide through the inhomogeneous fringe field of the superconducting magnet before they reach the analyser. Instrument control and data acquisition was carried out by Compass ftmsControl (Brüker Daltonics, Bremen,Germany).

The used nanoESI source was initially developed by Petr Novák, head of Group of Protein Structure Characterization by Advanced Mass Spectrometry in Prague, Czech Republic. This source is still under development and aims to provide a reliable nanoESI source, circumventing the shortcomings of the Bruker nanoESI sources. In this case, a conductive wire is placed in direct contact with the sample solution, and an external power source is used to supply high voltage. The tip of the needle is placed at about 1-2 mm in front of the orifice of the gas diverter, either at 180 degreeses or slightly tilted. The tips used were homemade, as in Figure 3.2 B.

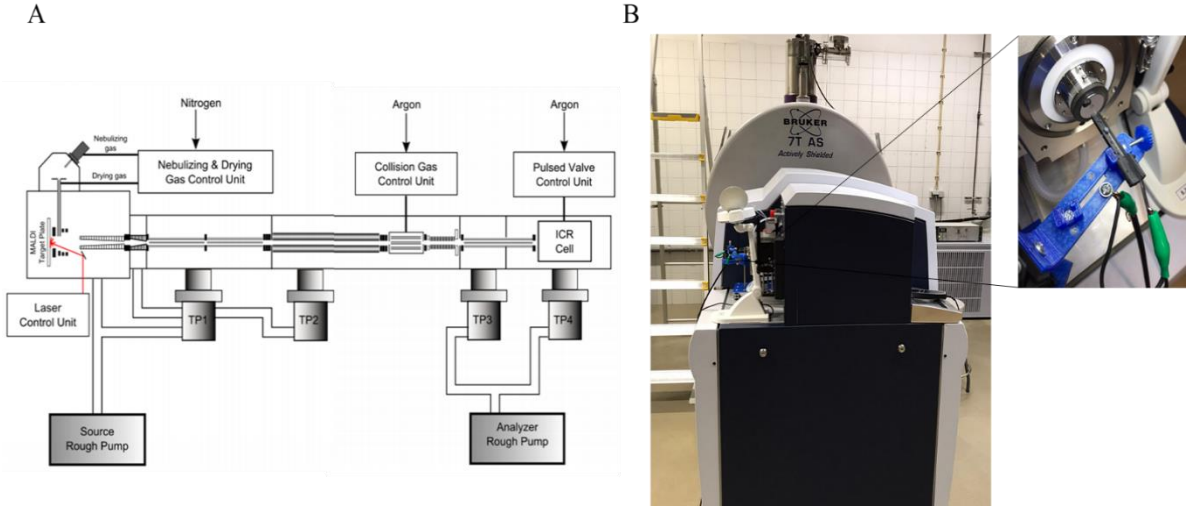


Figure 3.2 FT-ICR spectrometer. A) Scheme of the constitution of a FT-ICR. The source, the Qh-Interface, the ion transfer optics and the detector. The six differential pumping stages are maintained by four turbo-molecular pumps (TP1 – TP4). B) 7 Tesla SolariX XR mass spectrometer (Bruker Daltonics, Bremen,Germany) with nanoESI source, still under optimisation.

3.2. Emitter tip optimisation

3.2.1. Emitter tip production

The emitter tips were prepared in-house using a micropipette puller from Sutter Instrument. It was necessary optimise the tip to produce an orifice diameter with increased efficiency of ionisation. Two types of borosilicate glass with capillaries were used, one with 1.0 mm O.D (outer diameter) x 0.78 mm I.D (inner diameter) and other with 1.2 mm O.D x 0.69 mm I.D. A Sutter Instrument micropipette puller, model P-1000 with the filament FB255B (2.5 mm square box filament, 2.5 mm wide), was used (Figure 3.3).

Figure 3.3 Sutter Instrument P-1000. The pipette puller Model P-1000 is designed for the specific use of creating micropipettes. The filament is in the center, protected by rectangular box. The puller holds the capillary.

To produce the tips, some parameters that can be controlled in the Sutter Instrument were optimised, namely heat, pull, velocity, time or delay, and pressure.

Heat represent the amount of current supplied to the filament. The value does not represent the temperature, but indicates how much current is delivered to the filament. The type and size of filament, such as the size of glass determine the range of values that can be used.

Pull represent the hard pull introduced to the glass after the glass has softened. This pull value determines the amount of current supplied to the pull solenoid to create the hard pull.

Velocity is the rate of separation of the puller bars when the glass first starts to melt. The P-1000 is a 'velocity sensing' puller. This patented feature allows the puller to indirectly sense the viscosity of the glass, giving the P-1000 the ability to pull pipettes from all glasses except quartz. The velocity is detected by a transducer inside the puller, a patented feature, and the velocity has a direct correlation to the viscosity of the glass. High velocity makes longer taper and smaller tip and low velocity makes shorter taper and larger tip. The Model P-1000 supplies a blast of air to cool the filament area after the heating segment of a pull cycle.

Time is the duration of air used to cool the glass and the filament as the glass is being pulled. When using the Time mode, the glass softens, the velocity trip point is reached, and then the glass is cooled at pulled simultaneously. 1 unit is equivalent to 0.5 milliseconds of cooling air.

Delay is the alternative mode of cooling which provides a longer duration of cooling (300 ms) than the Time mode. In the Delay mode, the glass softens, the velocity trip point is reached, and then 300 ms of cooling is initiated. The Delay value determines how long the glass is cooled before the hard pull is engaged and continues cooling the glass as the glass is pulled. So, by cooling the glass while delaying the hard pull, you can determine how viscous the glass is when it is being pulled, higher delay for more viscous glass.1 unit is equivalent to 1 millisecond.

Pressure is the pressure of air used to cool the filament and the glass. The default Pressure setting is 500 units, which represents 2 psi of cooling air. Both the Time and Delay modes (duration of cooling) work in conjunction with the Pressure to cool the glass. A decrease in the Pressure will reduce the cooling to the glass and allow for longer and more gradual tapers.

The program can have more than one loop, when the glass does not separate in one stage the puller will return back and read line one over again or read the next line if it is programmed until the glass separates.

A variety of programs have been tested and optimised. In the end, just three of the seven programs developed (Table 3.1) formed a tip with the desired shape. As was already mentioned, these parameters have arbitrary units in the instrument. However, using the conversion provided by the manufacturer it was possible to express them in the conventional units (Supplementary table ***7.1*** in annexes).

Table 3.1 Programs and respective parameters for the emitter formation with different tips shapes. The values presented here are the ones used directly in the program, in arbitrary units. Heat represent the amount of current supplied to the filament. Pull represent the hard pull introduced to the glass after the glass has softened. Velocity is the rate of separation of the puller bars when the glass first starts to melt. Time is the duration of air used to cool the glass and the filament as the glass is being pulled. Delay is the alternative mode of cooling which provides a longer duration of cooling. Pressure is the pressure of air used to cool the filament and the glass.

Program	Glass (O.D x I.D)	Heat	Pull	Velocity	Time	Delay	Pressure
1	1.2 mm x 0.69 mm	450	-	155	80	-	500
		400	-	15	50	-	
		450	200	20	80	-	
2		450	-	8	250	-	
		400	200	6	200	-	
3		450	-	8	150	-	
		400	200	3	250	-	
4		470	130	80	100	-	
5		470	150	80	-	100	
6		450	200	60	-	160	400
7	1.0 mm x 0.78 mm	400	225	60	-	160	

The programs that resulted in a tip with a proper shape were tested and the signal efficiency was analysed (Supplementary figure 7.1 in annexes).

The prepared emitter tips for nanoESI capillary using the Q-TOF mass spectrometer were gold-coated using an automated sputter coater JEOL JFC 1200. For the analysis by nanoESI in the FT-ICR mass spectrometer, a conductive wire was placed in direct contact with the sample, using an external source to supply high voltage, being unnecessary the gold coating for the emitter tip. Both methods produce similar results but since the gold coating tens to erode, the wire maybe more reliable in the long run.

3.2.2. Emitter tip efficiency test

For testing the emitter tips two different samples were analysed, myoglobin from equine heart (SIGMA M1882, PDB entry 1AZI), a monomer around 17 kDa, and alcohol dehydrogenase (ADH) from *Saccharomyces cerevisiae* (SIGMA A7011, EC1.1.1.1, PDB entry 5ENV), a tetramer with 147 kDa. Both proteins were prepared in native buffer (150 mM of ammonium acetate) and in denaturing buffer (1:1 water and acetonitrile with 1% of formic acid). For native analysis, no purification was used and the samples were directly dissolved in the final buffer. For denatured analysis, the sample was prepared in water, being acetonitrile and formic acid added later. The sample of denatured ADH was filtered using a Amicon-Ultra-0.5 10K (Millipore) before the addition of acetonitrile and formic acid, resting over-night at room temperature before analysis. The final protein concentration for nanoESI-MS measurements was 7 µM for ADH and 15 µM for myoglobin, measured using nanodrop.

3.2.2.1. Q-TOF analysis

The gold-coated tips were tested in a custom modified Micromass Q-TOF Ultima II using ADH and myoglobin. Three µL of sample were introduced in the mass spectrometer through nanoESI source, using the different produced emitter tips. In total, around 300 scans were accumulated. Typical spray conditions were needle voltage 1.5 kV, cone voltage 132 V. The source temperature was 37ºC for native and 100ºC for denatured analysis. The source pressure was 5.5 10^{-3} mbar. To monetarize total ion current, for each sample was used a single nano-spray needle and leaving all experimental parameters constant. The backing Pirani, analyser penning and TOF penning were 2.0 mbar, 1.9 10^{-5} and 3.1 10^{-7} mbar, respectively. The mass spectrometer was externally calibrated with cesium iodide clusters (25 mg/mL).

Experimentally measured molecular weights (MW) were determined using UniDec software, a suite of computational tools built around a core Bayesian deconvolution algorithm [79,80]. Theoretical MWs were calculated based on the appropriate UniProt sequences [81].

3.2.2.2. FT-ICR analysis

Denatured ADH was also analysed in a FT-ICR-MS. Three µL of sample were introduced in the mass spectrometer through the nanoESI source. The nanoESI capillary voltage was set to 1.0-1.2 kV. The temperature of dry gas was 100°C and the flow rate was 1.6 L/min. The RF amplitude of the ion-funnels was 150 Vpp, and the applied voltages were 150 V and 6 V for funnels 1 and 2, respectively. The voltage of skimmer 1 was 15 V. A 22 % of sweep excitation power was used. The values of RF frequencies in all ion transmission regions were: octupole (5 MHz), quadrupole (2 MHz), and transfer hexapole (6 MHz). Ions were accumulated for 0.1 ms in the hexapole collision cell before being transmitted to the cell. A time-of-flight of 1 ms was used. Vacuum pressures for different regions were 3 mbar for the source region, 6×10^{-6} mbar for the quadrupole region, and 2×10^{-10} mbar for the UHV-chamber pressure. Various types of activation techniques were performed to optimise the signal. CID was performed in the hexapole collision cell with a collision energy of 10 V. The mass spectrometer was externally calibrated with NaTFA. Data was collected in magnitude mode with 100 transients from *m/z* 900 – 2500 averaged for each spectrum and recorded at 1M data points.

3.2.Protein expression and analysis

3.3.1. Protein expression and crude cell preparation

The protein used as a model was glyoxalase I from *Leishmania infantum*. This protein was expressed in *Escherichia coli* BL21Plus cells, already carrying the expressing plasmids [77]. Two different plasmids were used to produce *Li*GLO1: the vector pET28a containing *Li*GLO1, leading to the expression of the protein with a His6-tag in the N-terminal (His*Li*GLO1); and the pET23a containing *Li*GLO1, leading to the expression of the protein without any tag. For each analysis, the crude cell lysate without purification was used. The cell lysate from untransformed *E. coli* BL21Plus was used as negative control, while for positive control the purified recombinant protein of *Li*GLO1 was analysed, in native and denatured states.

For overexpression of *Li*GLO1 with His-tag, *E. coli* transformants were grown in two different culture medium: Terrific Broth (TB), a nutritionally rich medium for the growth of bacteria (24 g/L yeast extract, 20 g/L tryptone, 4 mL/L glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4), and Luria-Bertani (LB) medium (5 g/L yeast extract, 10 g/L tryptone and 10 g/L NaCl) supplemented with 1% of glucose, both containing 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol for pET28a, and 100 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol for pET23a. For overexpression of *Li*GLO1 without tag *E. coli* transformants were grown in the culture medium that gave the best results for native His*Li*GLO1 analysis from cell crude extracts.

Cell cultures of transformants with *Li*GLO1 with and without His-tag were grown at two different temperatures, with growth at 28ºC and 37ºC being tested. Following an over-night preculture at 37ºC, cells were diluted in culture medium to an OD_{600nm} of 0.1 and grown for about 1-1.5h at 37ºC. Expression of *Li*GLO1 was induced when the culture reached an OD_{600nm} of 0.6 with 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 3h at 37ºC or over-night at 28ºC. Cells were harvested by centrifugation at 7000 g for 5 minutes at 4°C. Cell pellets were frozen at -80°C.

To get the crude cell lysate, the cell pellet was washed two times in water and centrifuged at 7000 g for 5 minutes at 4ºC to remove the culture medium. Then, the pellet was resuspended in native buffer (ammonium acetate 500 mM), or in PBS and lysed by sonication four times at 70% of amplitude and 0.7s pulses with breaks between cycles at 4°C. After centrifugation at 15000 g for 40 minutes, the soluble fraction was saved. Total protein was quantified using a nanodrop. To analyse the levels of the expressed proteins, 25 µg of cell extracts were run on 12% polyacrylamide-SDS gels, together with 4.5 µg of recombinant protein and the molecular weight marker (NZYColour Protein Marker II, MB090). After the addition of loading buffer 4 times concentrated (62.5 mM Tris-HCl pH 6.8, 2.5% (w/v) SDS, 0.002% (w/v) Bromophenol Blue, 0.7135 M (5% (w/v)) β-mercaptoethanol and 10% (v/v) glycerol) to all samples, they were boiled at 100ºC for 10 minutes. After running, gels were stained with Coomassie blue.

3.3.2. Crude cell lysate and purified recombinant protein analysis in Q-TOF

For the native analysis samples were analysed in Q-TOF mass spectrometer. All spectra were externally calibrated, using cesium iodide (25 mg/mL). Typically, an aliquot of 3 µL of sample was loaded into a gold-coated nanoESI capillary prepared in-house, as previously described. Conditions within the mass spectrometer were adjusted to preserve non-covalent interactions, with the source operating in positive mode. In total around 300 scans were accumulated.

The capillary voltages were adjusted to maximize ionisation efficiency for each protein complex and varied between 1.4 kV and 1.7 kV, and cone voltage 132 V. The source pressure was maintained between 5.5 10^{-3} mbar and 6.5 10^{-3} mbar.

For the native analysis of purified recombinant protein, the source temperature was 37ºC, for denatured analysis of purified recombinant protein the source temperature was adjusted for 100ºC. For crude cell lysate both temperatures were also tested. Collision energy varied between 10 kV – 100 kV. The backing Pirani, analyser penning and TOF penning was 2.0 mbar, 1.9 10^{-5} mbar and 3.1 10^{-7} mbar, respectively.

Data processing and experimentally measured MWs were determined using UniDec [79,80]. Theoretical MWs were calculated based on the appropriate UniProt sequences [81].

3.3.3. Crude cell lysate and purified recombinant protein analysis in FT-ICR

For the native analysis of samples using FT-ICR, a 7 Tesla SolariX XR mass spectrometer was used. All spectra were externally calibrated using NaTFA. An aliquot of 3 µL of sample was loaded into the nanoESI capillary prepared in-house, as previously described.

The nanoESI capillary voltage was set to 1.2-1.5 kV. The temperature of dry gas was 37°C and the flow rate was 1.0 L/min. The RF amplitude of the ion-funnels was 150 Vpp, and the applied voltages were 150 V and 6 V for funnels 1 and 2, respectively. The voltage of skimmer 1 was 15 V. A 20% of sweep excitation power was used. The values of RF frequencies were optimised to transmit a better signal for high *m/z* ions. The final values for the analysis of crude cell lysate were the same value in all ion transmission regions: octupole (2 MHz), quadrupole (2 MHz), and transfer hexapole (2 MHz). Ions were accumulated for 0.1 ms in the hexapole collision cell before being transmitted to the cell. The time-of-flight of 2.5 ms was used. Vacuum pressures for different regions were 3 mbar for the source region, 3×10^{-6} mbar for the quadrupole region, and 2×10^{-10} mbar for the UHV-chamber pressure. CID was performed in the hexapole collision cell. Data was collected in magnitude mode with 100 scans from *m/z* 2800 – 5000 averaged for each spectrum and recorded at 512k data points.

For purified recombinant protein the temperature of dry gas has was 37ºC and RF frequencies were 5 MHz in octupole, 1.4 MHz in quadrupole and 2 MHz in transfer hexapole. The time-of-flight of 1.0 ms was used. Data was collected in magnitude mode with 50 scans. The remaining parameters were maintained.

4. Results and discussion

4.1. Emitter tips optimisation

4.1.1. Emitter tips production

Analysis of proteins is usually performed by means of nanoflow electrospray ionisation, using glass or quartz capillaries which have been pulled to a fine tip (~1 μm inner diameter), and coated with conductive material, such as gold when necessary. These capillaries are available ready-to-use from commercial sources, but can also be prepared in-house [82].

In nanoESI, the orifice diameter of the emitter is typically reduced to between 1 and 50 μm, which results in several important benefits for MS analysis. The dependence of ESI characteristics on the diameter of the emitter orifice has been examined by several investigators and some conclusions can be retained. First, as the emitter diameter decreased from 5 to 1 μm, ion populations shifted to higher charge states. This shift in charge-state distribution was attributed to the higher surface-to-volume ratio with decreasing emitted droplet diameter. Second, higher-than-expected ion currents are detected for the smaller tip diameters. For instance, a 1 μm diameter tip was observed to produce higher ion currents than a 2 μm diameter tip for the same solution. Higher ion current was attributed to the significantly higher electric field at the smaller tip, which produced greater droplet charging even though the volumetric flow rate was smaller [39,83].

Therefore, the shape and diameter of tips are essential to increase the electrospray ionisation and obtain a stable signal. Emitter tips previously produced in the laboratory were mainly intended for the study of purified and denatured protein, so were not suited for very high mass protein complexes. Hence, the first task was optimising the tips shape of the emitters to increase ion current and the signal of high *m/z* ions, maintaining non-covalent interactions with adequate S/N values in mass spectra.

The most important characteristic of an emitter is the shape of the tip, which include glass size, taper length and tips size. The most common causes of difficulty in producing the right shape of emitter tip can be attributed to the use of poor parameter settings. Just one poorly adjusted parameter setting in the program - too high or too low - can lead to a lot of variability in tip size and taper length [82,84].

The shape of the tip will determine the flow rate of the sample. By varying sample flow rate, initial droplets of different size can be generated. Protein molecules in small droplets form gas-phase ions earlier than the ones in large droplets. Previous studies demonstrate significant alterations to the charge state distributions with increasing flow rate. For example, in the presence of ammonium acetate, at low flow rates, lower charge states of proteins showed high intensities, while at high flow rates, ions related to higher charge states of proteins dominated the spectra. At low flow rates the protein molecules follow charged residue model of ionisation mechanism, and at high flow rates—due to structural changes in protein molecules in large ESI droplets—the charged residue and chain ejection models can possibly co-exist [85].

To produce a high ionisation efficiency tip, it is important to have a long tip with a fine orifice, to form a small droplet. A short tip with a wide orifice produces a higher droplet, which decreases the signal efficiency. Hence, it is necessary to adjust the parameters to produce a long tip. A lot of programs were created but only the emitters produced with an ideal shape (from programs 1, 5 and 6) were coated and tested (Table 3.1 and Supplementary figure 7.1). Programs 2, 3 and 4 produced tips too short. Two

different glass sizes (1.2 mm O.D x 0.69 mm I.D vs 1.0 mm x 0.78 mm) were used but the glass with 1.2 mm O.D x 0.69 mm I.D had increased resistance. Therefore, this glass was chosen for program optimisation, excluding program 7.

In Figure 4.1 two different capillaries are presented, being possible to observe the difference in the shape of the tip.

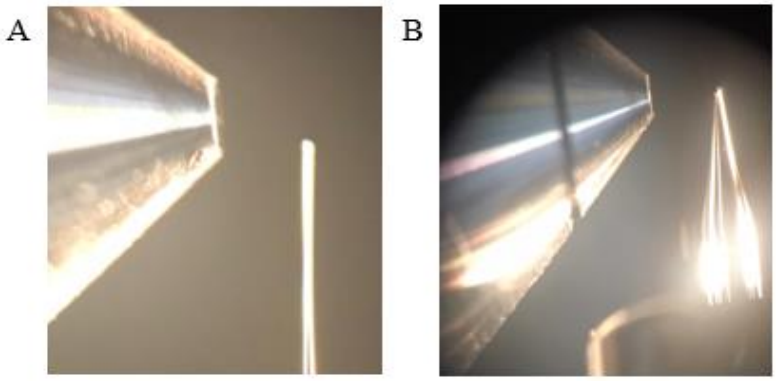


Figure 4.1 Emitter tips. A) Example of an emitter tip with adequate shape for the formation of a small enough droplet to allow efficient ionisation of the sample. B) Emitter tip with a large diameter, which makes a droplet too big, decreasing the ionisation efficiency.

Therefore, to analyse the capillaries with the best shape for protein analysis, two different proteins, myoglobin and alcohol dehydrogenase, were used. To choose the best emitter tip was taken in account, beyond the shape of the tips, the signal stability that was rated in total ion current and the noise present in the spectra. Another important aspect is the fact that the tips should be open. Since, the objective is to produce a small orifice, sometimes the orifice can be too small and the tip produced is closed and there is no flow.

4.1.2. Emitter tips efficiency test with myoglobin in Q-TOF

To test the efficiency of the emitter tips model proteins were analysed. The first one was myoglobin from equine heart, since is widely used as calibration standard or test compounds for various mass spectrometric methodologies [86].

Myoglobin (PDB entry 1AZI) is a globular all-helical protein with a molecular mass of 16951 Da (for the apo form) and binds heme (616 Da) in its native state, which corresponds to a total mass of 17567 Da, (Figure 4.2 and Supplementary figure 7.2). In denatured conditions, the prosthetic group is known to dissociate due to titration of the heme binding histidine residues [87]. As the severity of the denaturing conditions is increased, the interaction between the globin and the cofactor can be weakened, with possible separation of the cofactor from the protein. Multiply charged ions observed in the positive-ion spectra are produced by proton attachment to basic and deprotonated acidic sites in the protein and reflect, to some extent, the degree of protonation in solution. Under mild electrospray source conditions, myoglobin ions are highly charged showing a charge state distribution from +8 to +30 [88].

Figure 4.2 Myoglobin structure. The myoglobin polypeptide chain is represented in green, the heme group in blue and the water molecules in red. The structure was built with the software PyMOL (PDB entry 1AZI) [78].

The acid denaturation of myoglobin has been studied extensively under a variety of condition and denaturation occurs in the pH range 4.5-3.5 and depends strongly on the ionic strength of the solution [89,90]. It has been shown that some of the histidine's are buried in the native state of myoglobin and do not react with hydrogen ions [91]. The effective pK's of the remaining histidines and those of the glutamic and aspartic acid side chains are lowered considerably, compared to their intrinsic pK values, because of the electrostatic interactions in the native conformation.

As a result, the net positive charge on the native protein is lower than would be expected from the intrinsic pK values of the side chains. At lower pH values, the unfolding of the protein results in complete protonation of the buried histidines, disruption of the heme-globin interaction, and changes in the effective pK values of the side chains toward the intrinsic pK's. Thus, the denatured state may be characterized by the absence of heme and a relatively higher positive charge on the globin. Conversely, the native state of myoglobin may be characterized by the presence of heme in the hydrophobic pocket and by a lower net positive charge on the protein molecule [92,93].

- **Native myoglobin**

All tips used in the analysis of myoglobin were open and all the spectra had a low noise. The acquired spectra indicated a well-resolved charge state series between 1500 *m/z* and 3000 *m/z* in all samples tested, being the distribution of charge state centered at +7. The measured mass and the low number of charge states, indicated that myoglobin is in native form. The Figure 4.3 shows the variation of the total ion current over time and the respective spectra of each emitter tip tested.

Using the emitter tip produced by program 1, despite showing a stable ion current, an error of 10 Da in the mass calculated was observed, with a 7 Da deviation from the predicted theoretical mass (17567 Da), (Figure 4.3 A).

The program 5 produced an emitter tip that had a stable ion current with an error in mass calculated of just 1 Da and a deviation from the theoretical mass of just 3 Da (calculated mass 17570 Da), (Figure 4.3 B).

In its turn, the emitter tip produced by program 6 showed large fluctuations in total ion current, being the signal marked by a decrease at 2 minutes and the signal intensity was only recovered around the 5 minutes of analysis (Figure 4.3 C). The signal recovery was due to the application of some pressure on the emitter tip to increase the flow. Even so, the signal was lost before the 10 minutes of analysis, demonstrating to less stability when compared to the other emitter tips. Nevertheless, the mass calculated and the error were the same obtained with program 5.

In conclusion, of the three emitter tips tested with native myoglobin, the emitter tip from program 5 showed the best results concerning ion current stability and mass deviation.

Differences between theoretical and zero-charge molecular weights are attributed to unresolved, non-specific adducts that are common features of native mass spectra.

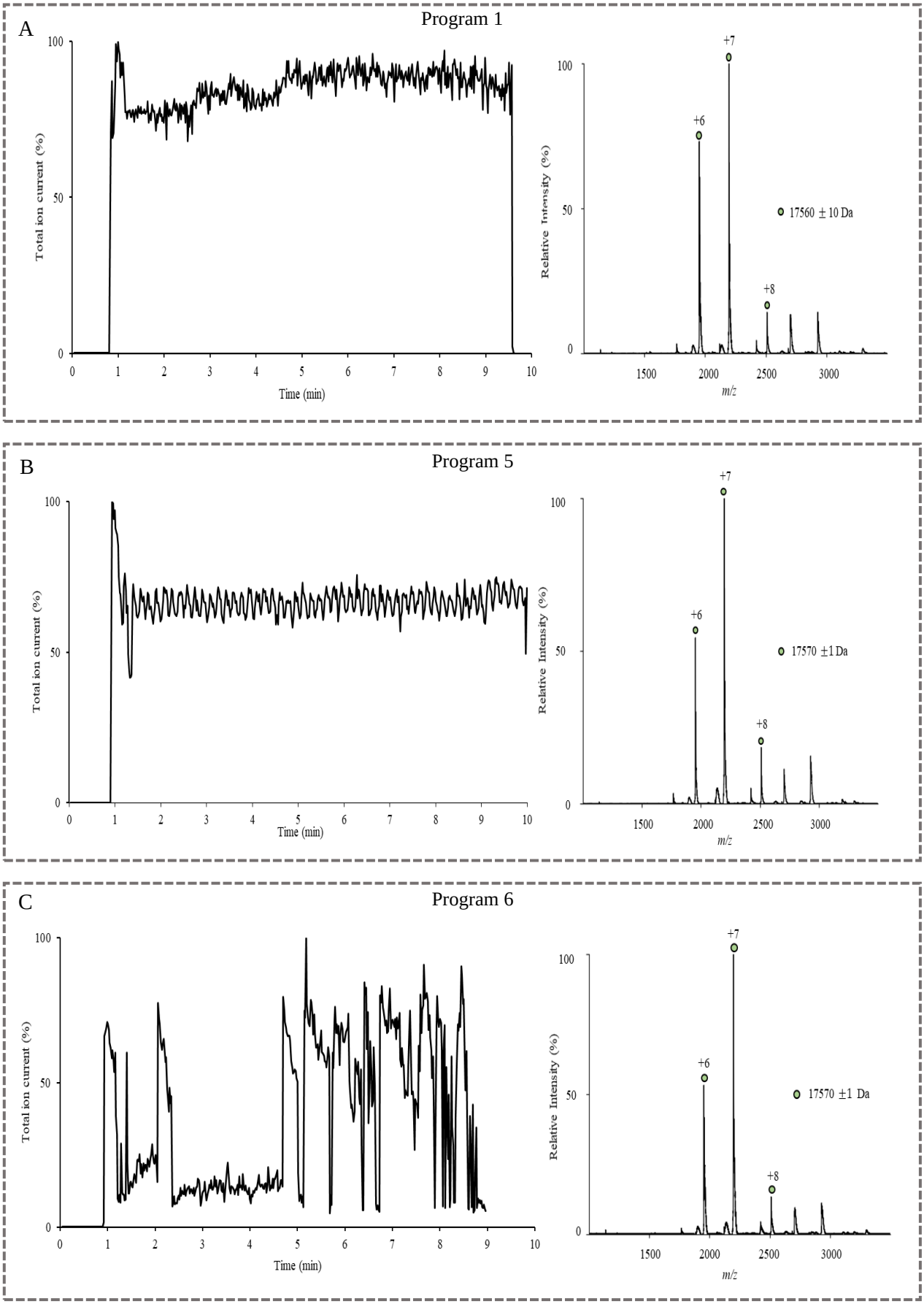


Figure 4.3 Myoglobin in native state using the emitter tips produced by different programs. In the left side of panels, A, B and C are represented the variation of total ion current (in %) over time (min) of emitter tips produced by programs 1, 5 and 6, respectively. The emitter tip produced by program 5 has the most stable signal. In the right side of each panel are the respective mass spectrum. Since the analysis was performed in native conditions, the determined mass for myoglobin is 17.5 kDa, accurate with 1 Da.

- **Denatured myoglobin**

Emitter tips produced using the same programs were also used to analyse the denatured form of myoglobin.

The first emitter tip tested produced by program 1 was closed but all the others produced by the different programs were opened. Figure 4.4 shows the data obtained from the different emitter tips with denatured myoglobin. The measured mass, the high number of charge states and heme detection indicated that myoglobin is in denatured form.

The emitter tip produced by program 1 had a fluctuation in total ion current and the signal was lost for a few minutes. The spectrum had some noise comparing with the other spectra obtained with this sample. However, it was possible to observe the presence of heme group, that appears a *m/z* 616, since the sample tested is denatured myoglobin. There was an error of 3 Da in the mass calculated (16960 Da) and this one had a deviation from the predicted mass of 9 Da. The acquired spectra indicated a well-resolved charge state series (+10 to +15) between 600 *m/z* and 1800 *m/z* (Figure 4.4 A).

The program 5 formed an emitter tip that produce a stable ion current. The spectrum has less noise than the other two spectra obtained with the same sample. The mass calculated had an error of 4 Da, and the deviation between the calculated (16960 Da) and the predicted mass (16951 Da) is equal to the previous, 9 Da. It was also possible to observe the heme group at *m/z* 616 and the distribution of charge state varied also between +10 and +15 (Figure 4.4 B).

The emitter tips produced by program 6 had a stable peaked at around 6 minutes in total ion current, because some pressure was placed on the emitter tip to increase the flow. During application of pressure, the total ion current was higher but decayed again when the pressure was no longer applied. The justification found to this phenomenon is due the presence of a sample squirt when the pressure is applied, instead of producing a nano-electrospray. The formation of this uneven jet is not expectable when the emitter tip is appropriate. Hence, ionisation efficiency is decreased, once small droplets have several desirable analytical properties like high surface to volume ratio which makes a large proportion of analyte molecules available for desorption. Since the small droplet is not formed, the proportion of charged molecules produced decreases. Besides, the mass calculated had an error of 5 Da and due to the lower ionisation efficiency, it is not possible to observe the heme group. The mass calculated was 16950 Da, which represents a deviation of 1 Da to the predicted mass. The spectrum has more noise than the one produced by emitter tip of program 5 but less than the one produced by program 1. Furthermore, the obtained ions had a lower distribution of charge state comparing with the others, varying only between +10 and +12 (Figure 4.4 C).

To conclude, of the three emitter tips tested with denatured myoglobin, the one produced using program 5 gave the best results, as also observed in the analysis with the myoglobin sample in the native state.

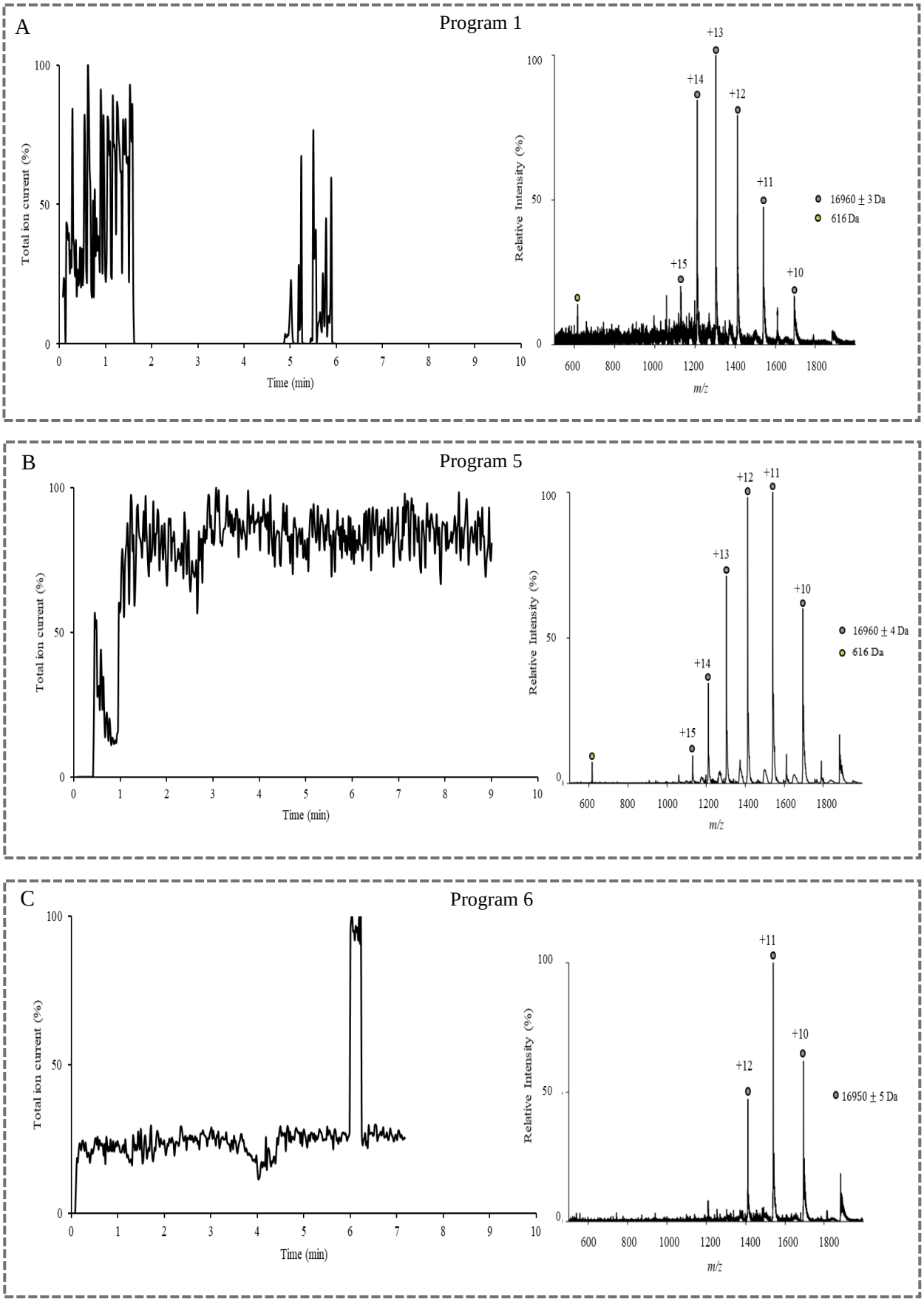


Figure 4.4 Myoglobin in denatured state using the emitter tips produced by different programs. In the left side of panels, A, B and C are represented the variation of total ion current (in %) over time (min) of emitter tips produced by programs 1, 5 and 6, respectively. The emitter tip produced by program 5 had the most stable signal. In the right side of each panel are the respective mass spectrum. Since the analysis was performed in denatured conditions, the mass obtained was 16,9 kDa and in spectra of panel A and B it was possible to observe the heme group.

4.1.3. Emitter tips efficiency test with alcohol dehydrogenase in Q-TOF

Since the goal of this work was optimising conditions from protein complexes transmission and myoglobin is a protein with only one polypeptide chain easily ionised, the efficiency of the tips was also tested using a protein complex, the yeast tetrameric alcohol dehydrogenase (EC1.1.1.1, PDB entry 5ENV). This protein is a zinc metalloenzyme responsible for the reduction of acetaldehyde to ethanol during fermentation of glucose [94–96]. The ADHs in higher eukaryotes (plants and animals) are usually dimeric, whereas those in prokaryotes and lower eukaryotes (yeast) are tetrameric [97,98]. Yeast and mammalian ADHs are distinctly homologous, yet only 25% of all residues are conserved.

Several isoforms are described for yeast ADH [99]. Sequence analysis suggests that the protein was a mixture of at least three different isoforms in yeast. Several microheterogeneities in the protein are indicated and one apparent amino acid exchange suggest that different types of subunits can occur [100]. As mentioned, yeast ADH is a tetrameric protein and, since there are different isoforms described for this protein, the mass of the monomer can vary between 36732 Da and 41124 Da and the tetramer between 146928 Da and 164496 Da. Previous studies indicate the presence of a dimer species for yeast ADH by MALDI-MS, gel electrophoresis and experiments with cross linking agents [101,102]. However, there are also other reports on studies of the quaternary structure of yeast ADH, using electrophoresis, ultracentrifugation and ESI-MS, which do not mention the presence of the dimer [103–105].

A structure for yeast ADH was determined by X-ray crystallography at 2.4 Å resolution [106]. The asymmetric unit contains four different subunits, arranged as similar dimers named AB and CD. The A and C subunits in each dimer are structurally similar, with a closed conformation, bound to the coenzyme (NAD^+). In contrast, the B and D subunits have an open conformation with no bound coenzyme, as represented in Figure 4.5, [106].

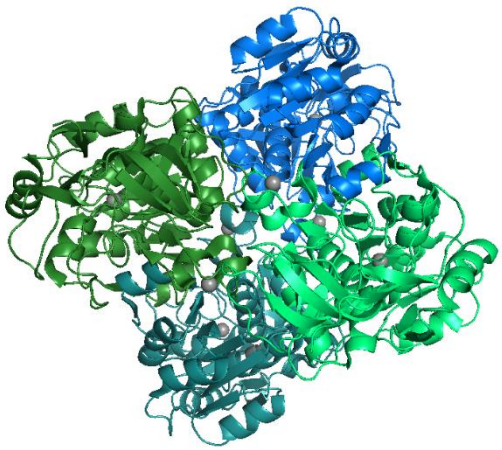


Figure 4.5 Structure of alcohol dehydrogenase. Alcohol dehydrogenase in yeast is a tetrameric complex arranged as similar dimers, named AB and CD. The A and C subunits are represented in blue in each dimer. The subunits B and D are represented in green. Metal ions are shown in grey (Zn^{2+}). The structure was built using the software PyMOL (PDB entry 5ENV) [78].

- **Native alcohol dehydrogenase**

ADH in native buffer was analysed using emitter tips produced by programs 1,5 and 6. Figure 4.6 shows the data obtained from different emitter tips with ADH in native buffer - 150 mM of ammonium acetate. It can be observed that the ADH sample appeared between 3000 *m/z* and 7000 *m/z*. The emitter tip produced by program 1, despite showing an unstable ion current during the first 5 minutes, ended up stabilizing. In the spectrum, it was possible to observe the presence of two species, a monomer with 36950 Da and a tetramer with 147810 Da, both with an error of 4 Da in the calculated mass. This spectrum had more noise than the other two spectra obtained with this sample. Narrow charge state series was clearly resolved for monomer (+11 to +13) and for tetramer (+24 to +27), (Figure 4.6 A).

Program 5 produced an emitter tip that had a stable ion current starting from 3 minutes. The mass through the spectra had an error of 4 Da for both the monomer with 36950 Da and the tetramer with 147530 Da. The distribution of charge state varied between +11 and +13 for monomer and between +24 and +28 for the tetramer, being possible to observe one more charge state than in the spectrum obtained with the previous emitter tip, (Figure 4.6 B).

The emitter tip produced by program 6 showed instability peaks, that are followed by minutes of stability in total ion current. In the spectrum, it was possible to observe the presence of a monomer with 36450 Da and a tetramer with 147650 Da. The mass calculated had an error of 20 Da for both species. The distribution of charge state varied between +10 and +13, exhibiting one more charge state than the other samples for monomer and between +23 and +27 for tetramer. The presence of the state of charge +10 in monomer that was not possible to observe in the remaining samples may be related with the fact that signal intensity was higher for the monomeric species than for the tetrameric species (Figure 4.6 C).

Although there are several isoforms with different masses described for the yeast ADH protein, the results obtained for the mass of the monomer, revealed that the isoform present in higher quantities had a mass between 36000 and 37000 Da. These results allowed to exclude some isoforms with higher masses. Therefore, the isoform with 36732 Da and 36849 Da are the most probable to be present in large amounts in this sample. To verify this hypothesis, MS/MS should be performed, to discover the sequence and know which isoform corresponds. Since it is not known which isoform is present in higher quantities, the deviations between the predicted and the obtained mass was not be calculated.

Nevertheless, some deviations in the mass calculated could be found, since several post translation modifications are reported for this protein, such as serine 213 and 279 phosphorylation of and threonine 223 or serine 2 acetylation [100,107].

In conclusion, as observed in the previous results with myoglobin, of the three emitter tips tested with native alcohol dehydrogenase, the one that presents the best results was the emitter tip 5.

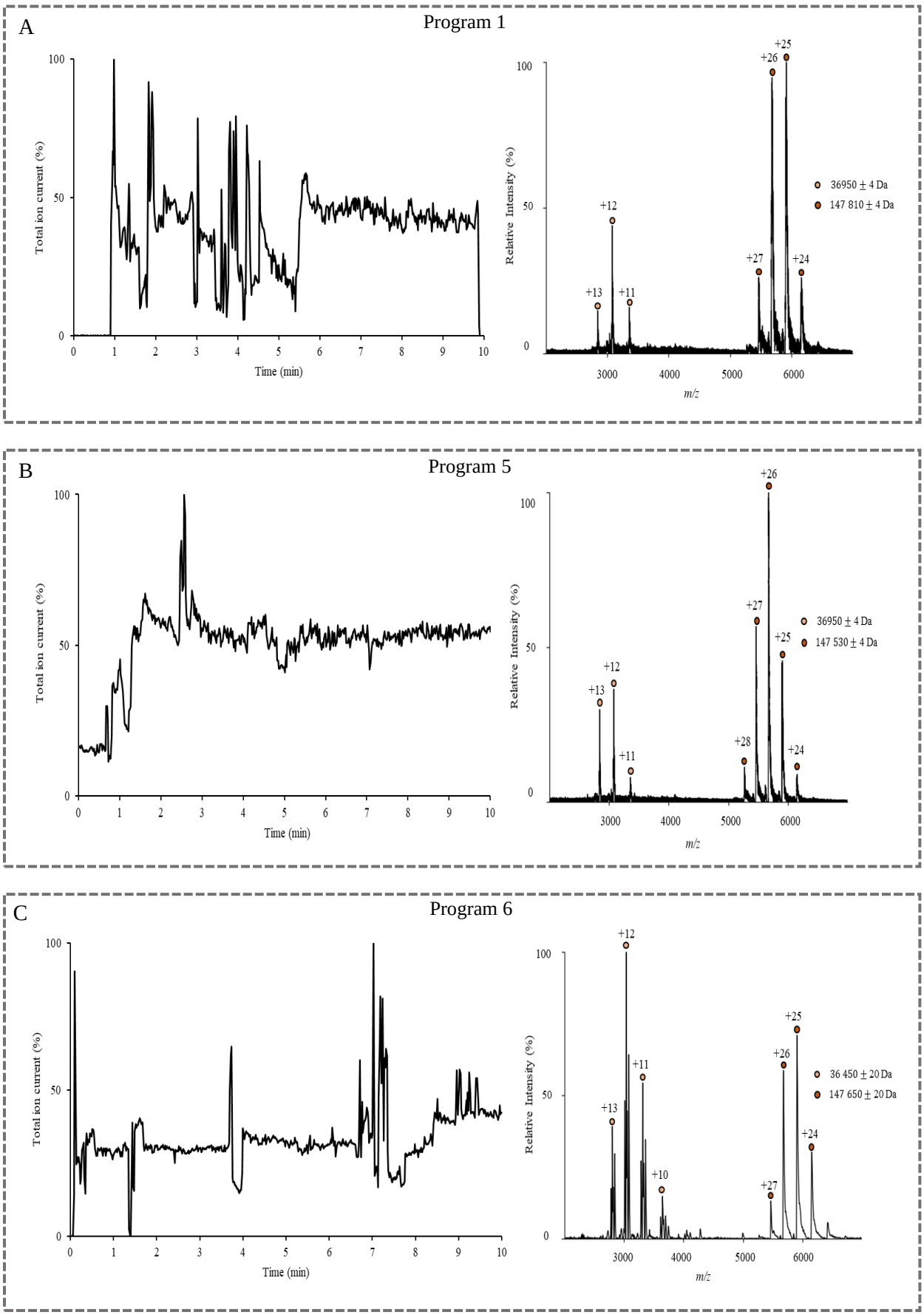


Figure 4.6 Alcohol dehydrogenase in native state using the emitter tips produced by different programs. In the left side of panels, A, B and C are represented the variation of total ion current (in %) over time (min) of emitter tips produced by programs 1, 5 and 6, respectively. In the right side of each panel are the respective mass spectrum. Since the analysis was performed in native conditions, the mass obtained was 147 kDa. However, it was still possible to observe the presence of monomeric species with 36 kDa.

- **Denatured alcohol dehydrogenase**

Again, denatured alcohol dehydrogenase was also analysed using emitter tips produced by the same programs 1, 5 and 6 (Figure 4.7). These spectra were obtained following an incubation of the ADH sample in denaturing buffer over-night and filtered. This step was crucial to observe a clear distribution of charge state.

The sample appears between 1500 *m/z* and 3500 *m/z*. The emitter tip produced by program 1 had an unstable total ion current, with various fluctuations over time. In the spectrum, it was possible to observe the presence of the monomer with 36740 Da with an error of 8 Da in the mass calculated. The distribution of charge state varied between +10 and +25, (Figure 4.7 A).

Program 5 formed an emitter tip that had a stable ion current starting from 3 minutes, with an error in mass calculated of 5 Da. The mass calculated from monomer was 36760 Da. The distribution of charge state varied between +10 and +26, (Figure 4.7 B).

The emitter tip produced by program 6 had instability peaks, as previously observed, and the signal was lost before 6 minutes. The mass calculated (36760Da) has an error of 10 Da. The distribution of charge state varied between +12 and +28. This spectrum had a bigger noise than the previous two, (Figure 4.7 C).

Comparing the native (Figure 4.6 right side) and the denatured ADH spectra (Figure 4.7), it is possible to observe the difference in the folding of the monomeric species. Although both correspond the same polypeptide chain, with the native buffer, the distribution of charge state was lower than the denatured buffer. When the protein is correctly folded, there are less amino acids exposed, acquiring less charge. In accordance with the charged residue model (CRM) [30], native mass spectra of folded proteins or protein complexes show a series of few charge states with Gaussian distribution and the number of acquired charges usually correlates with the surface area of the globular protein.

From these results, it was possible to conclude that the emitter tip produced by the program 5 had a more stable total ion current in all samples analysed, producing a lower noise spectrum with a higher ionisation efficiency, proved to be adequate for the study of protein complexes either in the native or denatured state.

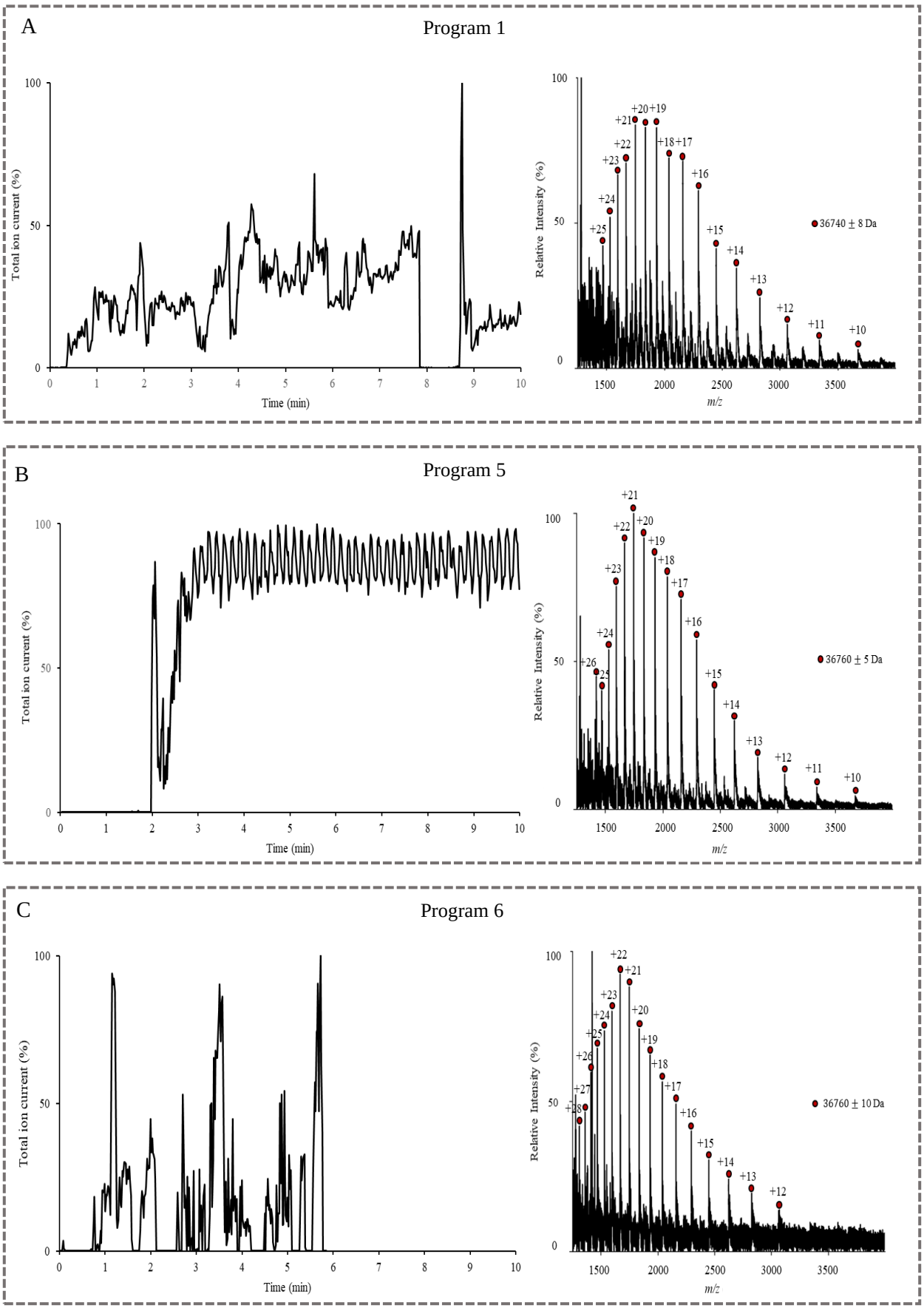


Figure 4.7 Alcohol dehydrogenase in denatured state using the emitter tips produced by different programs. In the left side of panels, A, B and C are represented the variation of total ion current (in %) over time (min) of emitter tips produced by programs 1, 5 and 6, respectively. The signal obtained by emitter tip produced by program 5 was the more stable. In the right side of each panel are the respective mass spectrum. Since the analysis was performed in denatured conditions, the mass obtained for the monomer was 36 kDa.

4.1.4. Emitter tip test with alcohol dehydrogenase in FT-ICR

Since the work also aims to optimise the conditions for the study of protein complexes, both in Q-TOF and FT-ICR mass spectrometers, after concluding that the emitter tip formed by program 5 produced a more stable total ion current in nanoESI source of Q-TOF, the same emitter tip was tested in the nanoESI source of the FT-ICR mass spectrometer. Since, at this point the parameters for the transmission of complexes in native state had not yet been optimised, only the spectrum of denatured protein was obtained.

The most unique advantage of FT-ICR as a mass analyser is that ion mass-to-charge ratio is experimentally manifested as a frequency. Because frequency can be measured more accurately than any other experimental parameter, ICR-MS offers inherently higher resolution (and thus, higher mass accuracy) than any other type of mass measurement. The benefit of measuring a compound's mass with adequately high accuracy can directly determine its elemental composition. Accurate mass thus acts as a powerful "filter" useful for confirming the identity of a compound or even identification of an unknown. Although the resolution is inversely proportional to the square root of the mass in FT-ICR, it is greater than that of Q-TOF [34,108]. Hence, the analysis of the denatured ADH using this powerful instrument could confirm unequivocally the isoform present in the sample.

After optimising the parameters for the transmission of the protein signal and the accumulation of 100 transients, the spectrum for denatured ADH protein was obtained, (Figure 4.8).

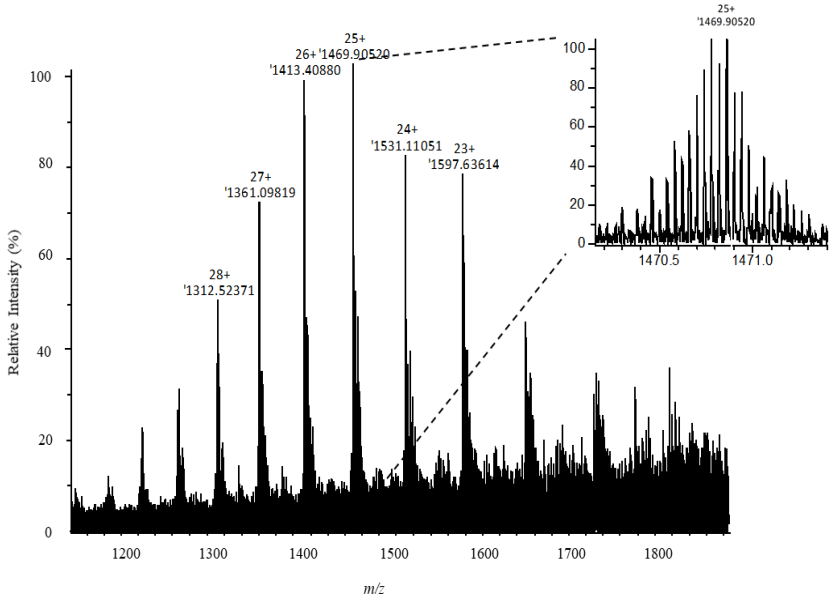


Figure 4.8 Alcohol dehydrogenase in denatured state analysed by FT-ICR-MS. The sample was infused using the emitter tip produced by program 5. The monomeric mass obtained was 36725 Da.

Observing the experimental mass obtained, 36725.43780 Da, and comparing with the results obtained by Q-TOF analysis it was possible to confirm which isoform was present in the sample. The previous results allowed to exclude isoforms of higher mass, but two possibilities could not be excluded, the isoforms with 36732 Da and 36849 Da. However, due to the resolving power of the FT-ICR, it was possible to conclude that the isoform present had 36725 Da (Supplementary figure 7.3), excluding the isoform with 36849 Da.

4.2. Analysis of complex formation of alcohol dehydrogenase

As mentioned before, ADH in native state was prepared directly from the vial and dissolved in the analysis buffer. The results obtained in the analysis of emitter tips with native ADH demonstrated the presence of the monomeric species even 2 to 3 hours after preparation, what was not expected.

In Figure 4.6 it was possible to observe that the intensity of monomeric species decreased with time, with an increase of the signal of tetramer, since the sample in emitter tip produced by program 6 was the first analysed, followed by the emitter tip produced by program 5 and finally, the emitter tip produced by program 1.

Since it was not possible to observe the presence of a dimer, these results are in agreement with previously published results where is mentioned the transition from monomer to tetramer without the presence of a dimer [10,101].

Various MS-based strategies may be used to gather information on the multimeric protein assembly mechanism. A straight-forward approach is to track the biogenesis process in real time, and monitor the decrease in the intensity of the monomers, and the corresponding increase in the intensity of the intermediate and assembled states [109]. The generated results enable extraction of information on the timescale and kinetics of the assembly process, and identification of the presence or absence of rate-limiting steps. In general, MS does not constitute a quantitative approach, owing to differences in ionisation efficiency, charging and transmission of protein ions across the instrument. However, the relative abundance of the protein can be monitored.

To study the formation of tetrameric ADH, the protein was analysed, using the gold-coated emitter tip produced by program 5, 1h and 3h after sample preparation (Figure 4.9).

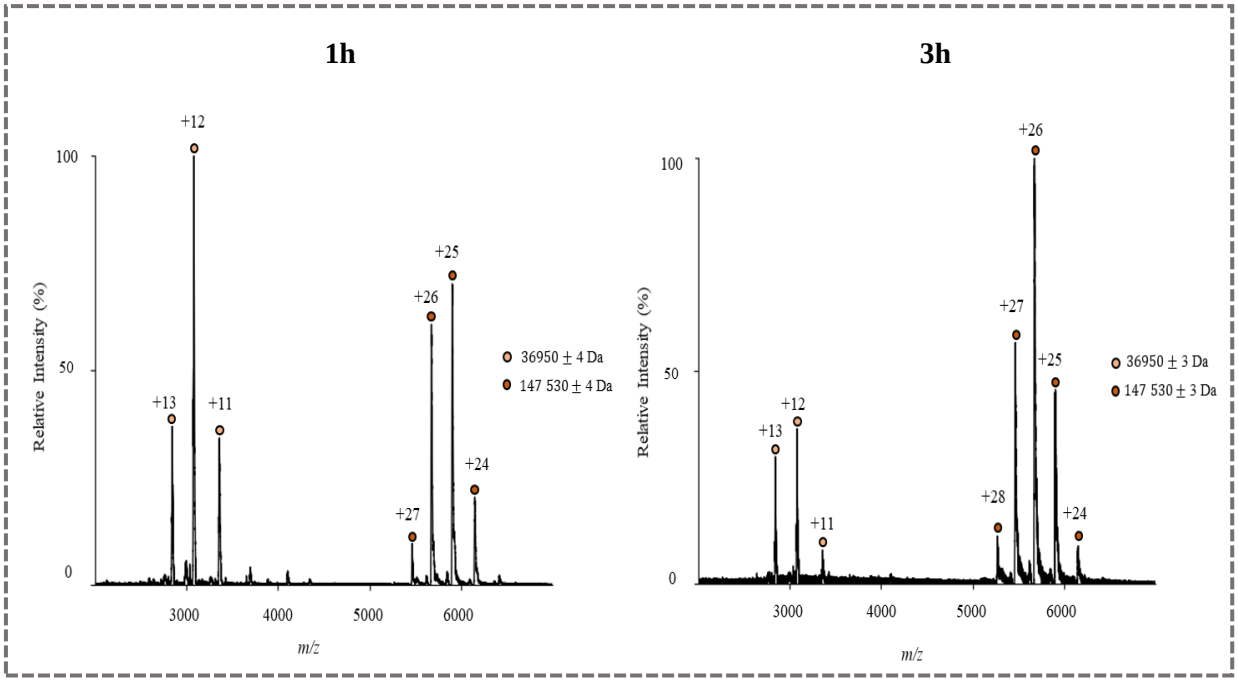


Figure 4.9 Complex formation of alcohol dehydrogenase over time. 1h after the preparation of native ADH, the monomeric species present was higher than the tetrameric species. 3h after the preparation, the amount of monomeric species decreased and the tetrameric species became present in higher quantities. Dimeric species was not detected. Zero-charge molecular weights determined from the experimental charge-state distributions are shown next to the spectra.

The results obtained confirmed that, in 1h after sample preparation, the most intense signal corresponded to the monomeric species. However, at the end of 3h, although the monomeric species was still detected, it was much less intense. Again, these results confirmed that the presence of dimer was not detected.

Despite of the kinetically forming of the quaternary structure of yeast ADH isn't known, structural studies report that this structure is composed by two similar dimers, being subunits A and C very similar to one another, as well as subunits B and D [106]. The packing of the symmetry related dimers in the unit cell could produce two different tetramers: a "back-to-back" dimer of dimers, AB:AB and CD:CD, in which the active sites are on the "front" sides and fully available to bind substrates; or "front-to-front" dimers, AB:CD, in which the active sites would be opposed and less accessible to solvent [106]. Therefore, these observations suggest that there is formation of the dimeric species before the formation of the tetrameric species. However, the formation of this species was not detected, suggesting that the transition from dimer to tetramer occurs very quickly.

It is becoming increasingly clear that dissecting the protein's structural architecture is not enough and a complete description of the biomolecular activity must also include the dimension of time. Transitions in structure are continuous, and occur at different levels and time-scales. For example, entire subunits rearrange and large-scale motions such as folding/unfolding and association/dissociation takes place. These protein dynamics, including structural kinetics and conformational transitions, is very broad. The simplicity, the ability to detect multiple populations of protein assemblies at the same time and the unlimited size of proteins made mass spectrometry very useful not only in structural and protein dynamics, but also in kinetics studies.

There are multiple options for monitoring structural kinetics by MS. The most straightforward option is to follow a reaction continuously in real time. Such an analysis will reveal the directionally and progression of the process and capture its kinetic properties. However, has the disadvantage of occurring over an extended period of time inside the ESI capillary, where it is often difficult to maintain a stable flow over the entire duration of the experiment. Moreover, prolonged spaying can cause electrochemical changes in the solution, which can affect the chemistry of the reaction.

Repeated manual sampling of the reaction at different time points may be also performed. In this case, attention should be given to the issue of capillary-capillary irreproducibility, which can be addressed by normalization approaches. When the availability of the reactants is limited, even acquisition of a single spectrum during steady-state can provide valuable information. Such a spectrum can reveal the co-existing species and their relative abundance, providing clues to the reaction mechanism [46].

How proteins folds continue to be the subject of many studies, and a major experimental challenge. The intermediates are transient, with estimated lifetimes in the millisecond or even microsecond range. Amide hydrogen/deuterium exchange (HDX) is a powerful strategy for capturing such short-lived folding intermediates. This method measures the rate of exchange between hydrogen and deuterium atoms. Hydrogens in backbone and side-chain are in constant equilibrium with the hydrogen atoms in the surrounding solution. Hydrogens with a low degree of solvent accessibility, or those involved in stable hydrogen bonding, will undergo exchange much more slowly than those that do not satisfy either of these conditions [110].

Another strategy for monitoring conformational transformations is by covalent labelling of solution-exposed surfaces. This approach was employed to monitor the time-resolved structural changes during refolding of acid-denatured cytochrome c [111].

The results obtained shows that Q-TOF mass spectrometer can be used to study the formation of protein complexes in a fast and reproducible manner, with minimal sampling handling and allowing the time course analysis of protein complex formation.

For future work, the formation kinetics of the tetrameric ADH complex will be studied following the reaction continuously in real time, to reveal the directionally and progression of the process and to capture its kinetic properties.

4.3. Protein expression optimisation

The protein used as model for the study of protein complexes *in vivo* was glyoxalase I from the parasite *Leishmania infantum*. As mentioned, this protein is homodimeric and since that has no cysteines in its sequence (Figure 1.5 A), the complex is formed by non-covalent bonds, together with the fact that the structure determined by crystallography is known, make this protein a good model for optimising complex transmission conditions in mass spectrometers.

The *Li*GLO1 enzyme is formed by 141 amino acids and the monoisotopic mass obtained through the sequence is 16350 Da, with an average mass of 16362 Da (UniProt ID: Q2PYM9). Glyoxalase I is a functional dimer with 32724 Da, predicted by sequence.

One of the most important aspects for obtaining a good spectrum is sample preparation. Since the complex was studied from the cell extract, optimising the sample preparation include optimising the protein complex expression. It is difficult to provide exact numbers for the overexpression ratio or the protein concentration required for direct-MS analysis because the method is highly dependent on the abundance, molecular weight and ionisation efficiency of both generated and background proteins. As a rule, following induction, if the produced protein appears as a dominant band in a Coomassie-stained SDS-PAGE gel, or is far greater or smaller, in terms of size, than other proteins, it is likely to be detected by direct-MS [58]. Hence, the expression of *Li*GLO1 was monitored, by SDS-PAGE.

To study this protein *in vivo*, the first step was to produce the protein in a cell system capable of overexpressing it, such as *E. coli*. To obtain a higher expression of this protein, different culture media and expression temperatures were tested.

4.3.1. Culture medium optimisation

Two culture media were tested to select the best conditions for *Li*GLO1 protein expression and direct MS analysis from the crude extract. The two media tested were LB (Luria-Bertani), a culture medium suitable for the growth and maintenance of *E. coli* strains, and TB (terrific broth), a highly enriched medium, containing a higher concentration of yeast extract also used for bacteria growth. The optimisation of the culture medium started with *E. coli* cells transformed with plasmid pET28a – *Li*GLO1 (expressing a His*Li*GLO1 protein), since we observed these cells grow faster, than those containing pET23a – *Li*GLO1 plasmid. After induction, both soluble and insoluble (pellet) protein fractions were analysed, to discard the possibility of forming inclusion bodies. Protein extracts from non-induced cells (without IPTG) were also analysed (Figure 4.10).

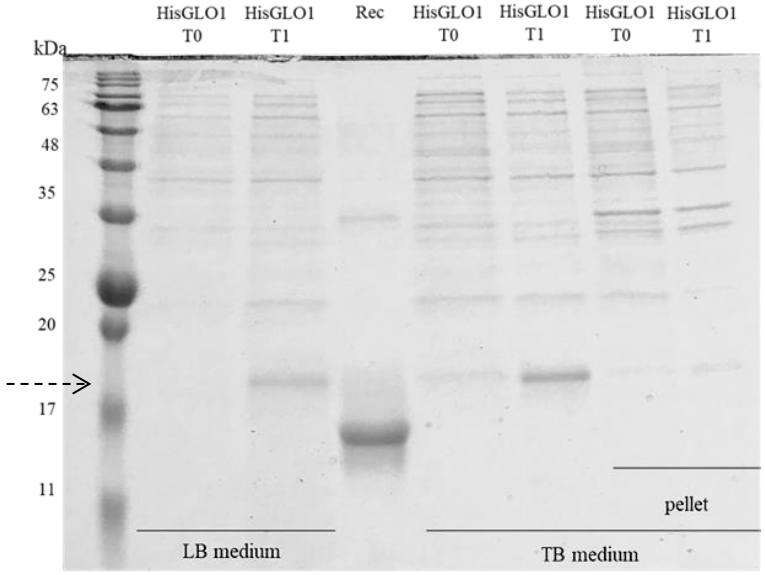


Figure 4.10 Expression of His*Li*GLO1 in LB and TB medium. HisGLO1 T0 – crude cell lysate of *E. coli* overexpressing *Li*GLO1 with His-tag without the addition of IPTG. HisGLO1 T1 – crude cell lysate of *Li*GLO1 with His-tag with 0.1 mM of IPTG. Lysates from cells grown in both LB and TB medium were analysed. The *Li*GLO1 recombinant protein (Rec), without His-tag, was used as control. Since *Li*GLO1 with His-tag was higher mass, migrated above the recombinant protein. Discontinues arrow indicates the correspondent band of His*Li*GLO1.

*Li*GLO1 with His-tag migrated above the recombinant protein. The molecular mass of His*Li*GLO1 protein is 18161 Da (calculated from the sequence and including the 1818 Da from the His-tag). Therefore, the band between the markers of 17 and 20 kDa corresponded to the His-tagged protein, marker by discontinues arrow and the band between 11 and 17 kDa in the 4th lane (Rec) corresponded to the purified protein *Li*GLO1 (Figure 4.10). In the recombinant protein sample, can also be observed the presence of a band at around 35 kDa, that corresponded to the dimeric species of this protein, demonstrating that even despite the denaturing conditions and after being boiled for more than 10 minutes, it was still possible to observe the presence of this dimeric species.

The first two lanes contained the supernatant of the cells grown in LB medium (Figure 4.10). In the first one, the protein expression was not induced, that is, no IPTG was added, and no band corresponding to His*Li*GLO1 protein was observed, as expected. In turn, in the next lane, the addition of IPTG during the growth, led to the induction of His*Li*GLO1 expression, confirmed by the presence of a band between 17 and 20 kDa.

The two lanes following the recombinant protein contained the supernatant of the cells grown in TB medium. Contrary to what was observed when growth occurs in LB medium, it was possible to detect the expression of His*Li*GLO1 even without the addition of IPTG. In the supernatant of the cells in which inducer was placed the protein expression was higher, as verified by the presence of a more intense band between 17 and 20 kDa (Figure 4.10).

As mentioned before, to rule out the hypothesis of the presence of His*Li*GLO1 in inclusion bodies, the insoluble fractions were also analysed. Although it was possible to observe a band corresponding to His*Li*GLO1 in these samples, the bands are less intense then in the supernatant (Figure 4.10, last lane).

The structural studies by mass spectrometry started with the analysis of the recombinant protein *Li*GLO1 using the Q-TOF mass spectrometer. The recombinant glyoxalase I was analysed in denaturing and native conditions, as represented in Figure 4.11.

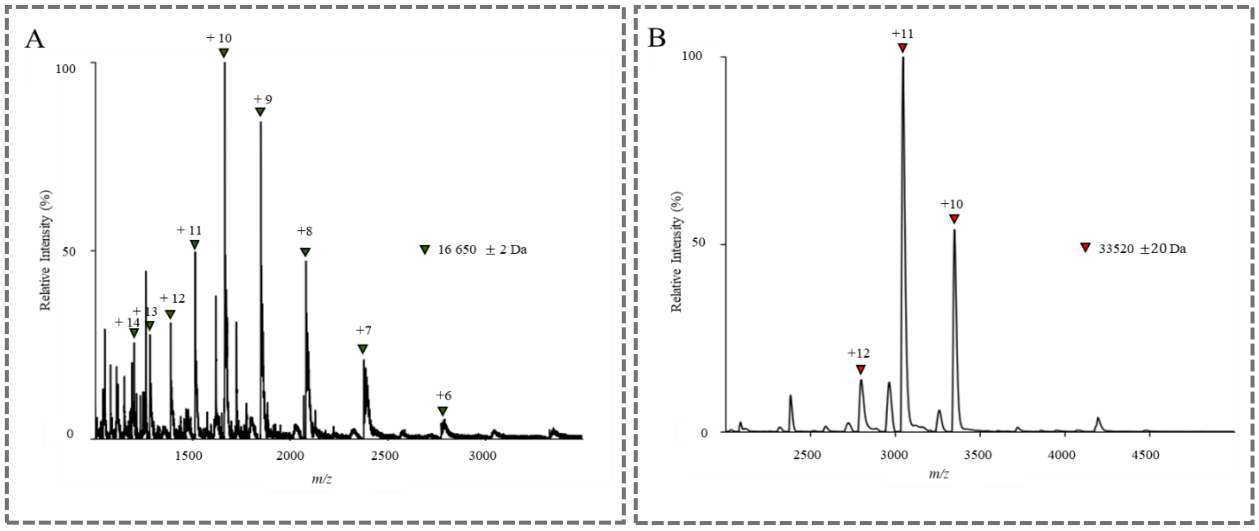


Figure 4.11 Spectra of Glyoxalase I from *Leishmania infantum* in Q-TOF mass spectrometer. In panel A is represented the spectrum of denatured *Li*GLO1. In panel B the spectrum of native *Li*GLO1. Zero-charge molecular weights determined from the experimental charge-state distributions are shown next to the spectra. The mass obtained in the native spectrum demonstrate that the protein is dimeric.

Denatured *Li*GLO1 was detected with a molecular mass of 16650 Da, with a distribution of charged states between +6 and +14. The difference between the calculated mass and the predicted mass by sequence was 289 Da (Figure 4.11 A). Since the recombinant protein was previously purified, mass shift can be caused by changes during purification steps, such as the presence of adducts resulting from the buffers used.

In the native state, purified *Li*GLO1 appears as a dimer with 33520 Da (Figure 4.11 B), being this information in agreement with structural information previously published about this protein, since studies using preliminary Static Light Scattering (SLS), based on measuring the average intensity of light scattered by a protein solution of defined concentration confirmed that this protein is dimeric in solution [77].

Since the monomer has a mass of 16650 Da, the predicted mass for the dimer would be 33300 Da, and we detected the recombinant dimer with a difference of 220 Da from the observed value.

Crystallography studies showed that each dimer had two metal binding sites and the metal binding sites were located in the interface of two monomers building up the dimer [76]. This study also reveals that zinc was the main metal bound to *Li*GLO1, being nickel and cobalt present in very small amounts. When zinc was present, also were detected low quantities of copper and manganese. In the *Li*GLO1

structure, Zn^{2+} is coordinated by residues His8 and Glu59, from one monomer, and His77 and Glu120, from another monomer, and a water molecule [76].

Biochemical and structural studies have shown that glyoxalases are metalloenzymes that require either nickel (II) or zinc (II) for catalytic activity. It has been suggested that the metal specificity may be predicted based on the amino-acid sequence and the length, given that zinc (II)-dependent glyoxalases contain short additional regions that are absent in the nickel (II)-dependent enzymes. On the other hand, structural analyses indicate that only the enzymes in which the metal cofactor displays an octahedral coordination geometry are catalytically active. Such an octahedral geometry arises from the metal coordination to four conserved residues present in His/Glu/Gln/Glu or His/Glu/His/Glu metal-binding motifs, completed by two solvating water molecules [112].

Since the spectrum was obtain using native conditions, the mass differences between the observed and the predicted can be explained by the presence of metal ions. In the structure obtained by crystallography a total of 285 water molecules were found [77], some of which may be present inside the dimer, contributing to the increase in the mass of the dimer observed in the MS analysis (Figure 4.11 B).

However, note that during native-MS, the experimentally determined mass of a protein or protein complex is usually higher than its theoretical mass. This mass shift is caused by incomplete desolvation and remaining water, buffer and salt adducts. In contrast to a folded protein, a denatured protein in solution shows a larger distribution of charge states with higher charges. The observed charge states in native-MS spectra are therefore indicative for the folding state of the proteins/protein complexes, i.e. folded or unfolded states. Typically, 0.5 M - 1 M ammonium acetate at pH ~7 is used. For most proteins, this solution results in minimal accumulation of adducts and high signal-to-noise ratios during measurements. However, lower concentrations can also be used [60].

As a control, a reference spectrum from a non-transformed bacteria must be acquired. Specifically, for bacterial expression systems, this control can be a sample of a cell lysate before induction of expression or the same *E. coli* cells without any plasmid. For this purpose, untrasformed *E. coli* cells were analysed and the obtained spectrum is represented in Figure 4.12.

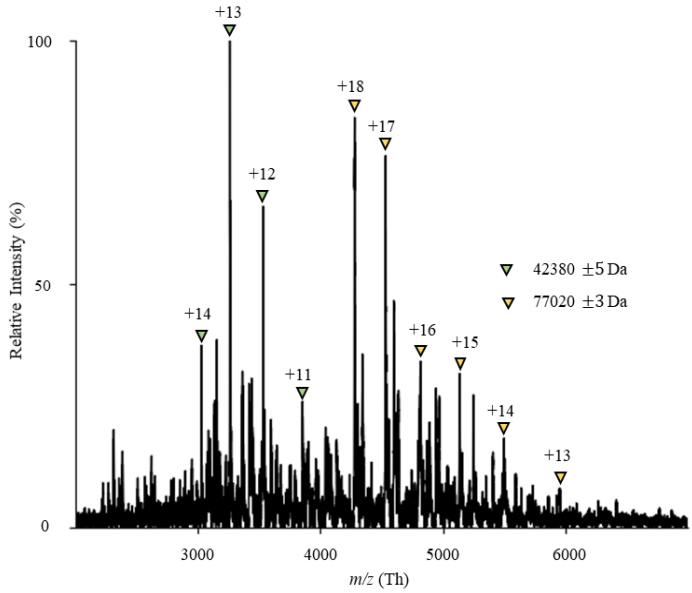


Figure 4.12 Spectrum of crude cell lysate of bacterial cell. In the spectrum of a non-transformed bacteria it was possible to observe two main proteins, with 42 kDa and 77kDa, that dominated the signal. These are probably the most abundant proteins in these bacteria. Zero-charge molecular weights determined from the experimental charge-state distributions are shown next to the spectra.

Altought the obtained spectrum was complex and noisy, two major proteins with 42380 Da and 77020 Da were observed. The identification and comparison of these proteins in the SDS-PAGE gel (Figure 4.10) was not possible, since the gel was run in denaturing conditions and the protein's quarternary structure may consist of more than one polypeptide chain. No attempt was made to identify these proteins. While the specific sequence of these proteins remains unknown, similar signals have been reported previously from samples that were prepared in a similar manner to that performed [58].

However, these results show that through this technique it is possible to study proteins in bacteria systems even if they are not being overexpressed. The expression of these endogenous proteins was sufficient to be detected, even there isn't a very marked band on the gel. The same was not observed when studying the yeast extract, in which the spectrum obtained does not allow the identification of any endogenous protein most likely because of the higher complexity of the yeast proteome.

The cell lysates of *E. coli* overexpressing His*Li*GLO1 grown in different culture media were compared. The expression of His*Li*GLO1 was induced for 3h since, over long expression periods, a reduction in signal intensity might occur due to the formation of protein aggregates, which affect the protein's solubility and its ability to ionise, or due to extensive masking of the signal by endogenous proteins that accumulate in the growth medium because of continuous cell death.

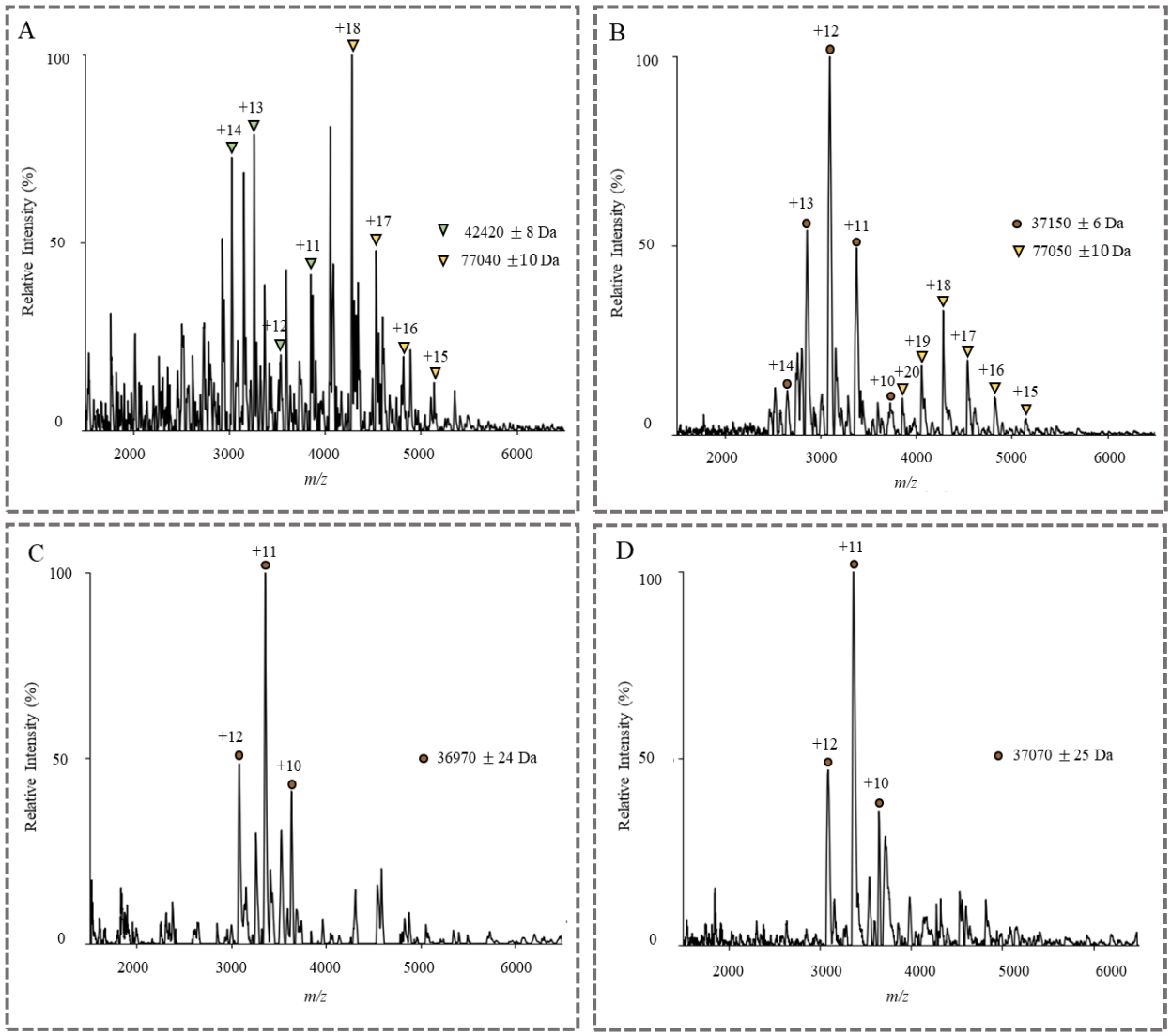


Figure 4.13 Crude cell lysate spectra of *E. coli* overexpressing His*Li*GLO1 grown in different culture media. Panel A contains the spectrum of crude cell lysate in LB medium without IPTG. In this sample, the protein was not detected. Panel B shows the spectrum of crude cell lysate overexpressing His*Li*GLO1, in LB medium (with IPTG). Beside His*Li*GLO1, it was still observed the presence of the protein with 77 kDa. Panel C contains the spectrum of crude cell lysate in TB medium without IPTG. The basal expression of His*Li*GLO1 present in this rich medium is enough to dominate the spectrum. In panel D is shown the spectrum of the crude cell lysate overexpressing His*Li*GLO1 in TB medium (with IPTG).

When cells were grown in LB medium without the addition of IPTG, the protein His*Li*GLO1was not detected, as expected, since the expression was not induced (Figure 4.13 A). These results are in accordance with what was observed in gel (Figure 4.10). It was possible to detect two proteins, one with 42420 Da and another with 77040 Da, as observed in the sample used as a control (Figure 4.12), in which the bacteria had not been transformed.

In cells grown in LB medium with the addition of IPTG, it was already possible to detect the presence of His*Li*GLO1 with 37150 Da, indicating that this becames one of the most abudant proteins in cell (Figure 4.13 B). Thus, the monomer had a mass of 18575 Da, showing a deviation of 414 Da for the predicted mass (by sequence, plus the His-tag). Since the cell extract was diretly analysed without applying any purifications steps, it was expected that deviations in masses of protein would be found. Besides metal ions, there may be other molecules attached to the protein when it is in dimeric form in the cell. However, it was still possible to observe the presence of the protein with about 77 kDa (Figure 4.13 B), also confirmed by SDS-PAGE (Figure 4.10).

When cells grown in TB medium, the His*Li*GLO1 protein was expressed and produced a signal that dominated the spectrum, being other proteins no longer detected (Figure 4.13 C, D). The results corroborated the observed in SDS-PAGE gel (Figure 4.10), the protein was expressed even without IPTG (Figure 4.13 C). The distribution of charge states was the same with or without the induction of protein expression. In fact, being a highly enriched medium, TB medium induced the protein expression in *E. coli* carrying the pET28a-His*Li*GLO1, in sufficient quantities to produce a good spectrum.

Comparing the results of the gel (Figure 4.10) and the spectra (Figure 4.13), it was possible to observe that the low amounts of the protein that accumulated after 3h of induction were enough to obtain good measurements. This was achieved because the crude growth medium displays a relatively low-complexity background, which did not mask the signals of *Li*GLO1.

Therefore, the sample that produced the best spectrum was His*Li*GLO1 grown in TB medium, since the overexpressed protein dominate the signal, no other proteins were detected and the complex presented a clear distribution of charge states.

4.3.2. Induction temperature optimisation

Besides the culture medium, the effect of the induction temperature on protein expression was also tested. The His*Li*GLO1 expression was induced at 28ºC and 37ºC, simulating the temperatures of the growth cycle of the *Leishmania infantum* parasite. Inside the female sandfly, parasites are maintained at lower temperatures, around 26-28ºC, whereas when infecting an human host, the growth temperature is close to 37ºC. The supernatants of the cells grown at these temperatures were extracted and analysed by SDS-PAGE (Figure 4.14). *E. coli* cells either overexpressing His*Li*GLO1 or untagged *Li*GLO1 proteins were grown in TB medium. His*Li*GLO1 appeared between the markers 17 and 20 kDa (discontinues arrow in Figure 4.14) and *Li*GLO1 between 11 and 17kDa (continues arrow in Figure 4.14), as expected given the molecular mass for both proteins.

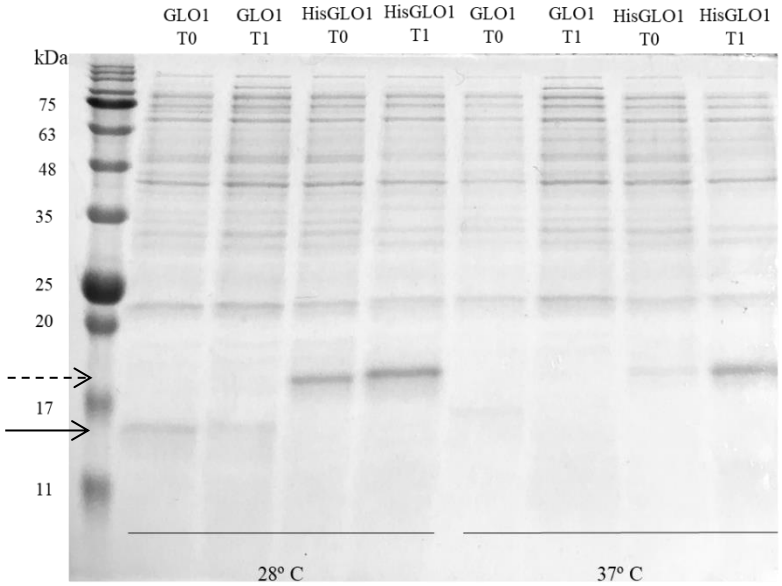


Figure 4.14 Expression of untagged *Li*GLO1 and His*Li*GLO1 in *E. coli* cells grown in TB medium at different induction temperatures (28ºC and 37ºC). GLO1 T0 - crude cell lysate of *Li*GLO1 without IPTG. GLO1 T1 - crude cell lysate of *Li*GLO1 with IPTG. HisGLO1 T0 – crude cell lysate of His*Li*GLO1 without IPTG. HisGLO1 T1 – crude cell lysate of His*Li*GLO1with IPTG. Continues arrow indicates the band of *Li*GLO1 and discontinues arrow indicates the correspondent band of His*Li*GLO1.

As previously observed, transformed cells grown in this culture medium resulted in a higher protein expression, even without the addition of the inducer (Figure 4.14 T0 lanes). In all samples, the expression of tagged and untagged protein, was higher at 28ºC than at 37ºC, with better results obtained with the His-tagged protein (Figure 4.14).

Since the protein overexpression at 28ºC proved to be better for the present research purpose, only the spectra of samples grown at this temperature were obtained.

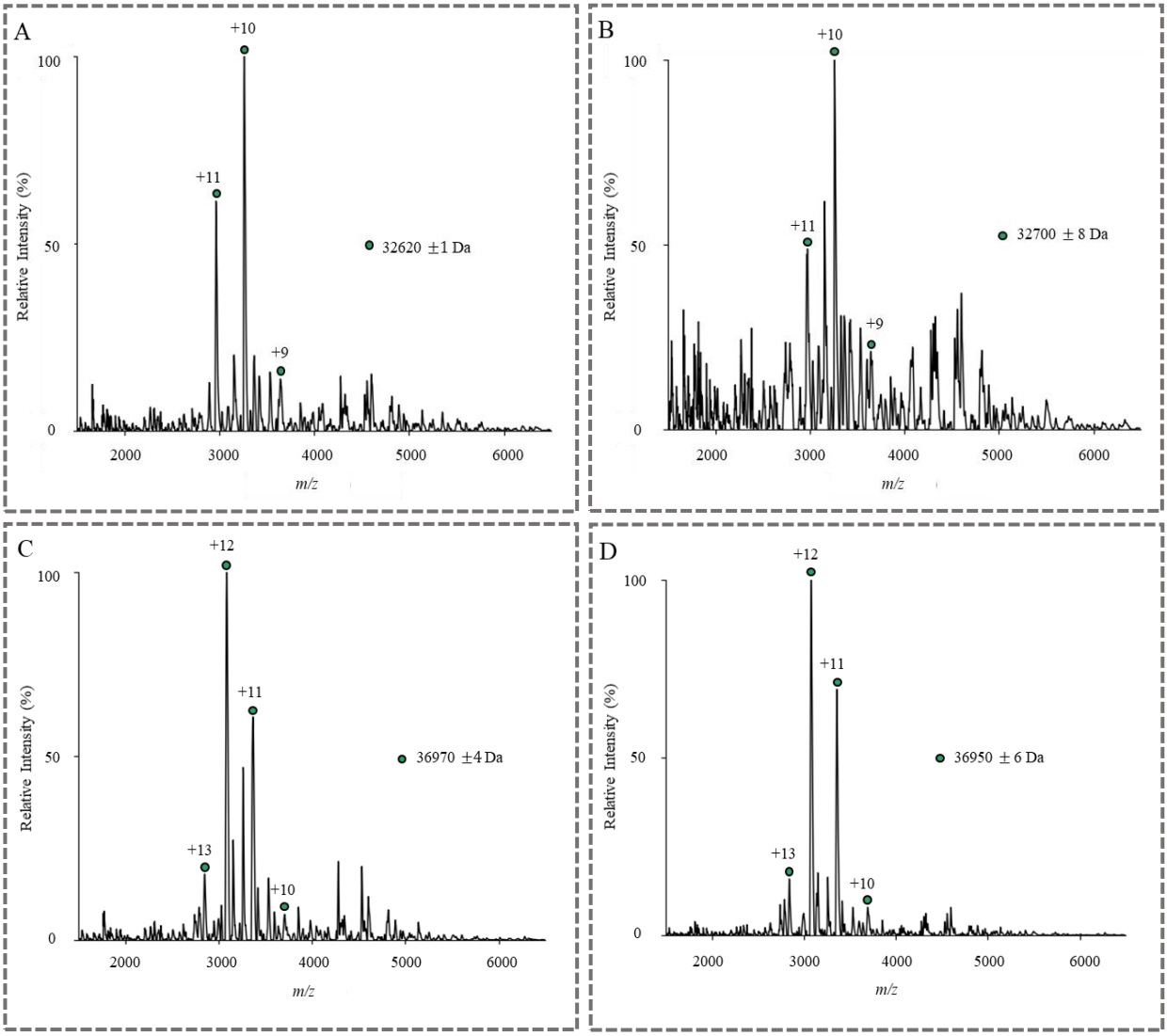


Figure 4.15 Spectra of *E. coli* cell lysates overexpressing either *Li*GLO1 or His*Li*GLO1 at 28ºC. Panel A shows the spectrum of crude cell lysate *Li*GLO1 without IPTG. Panel B shows the spectrum of crude cell lysate overexpressing *Li*GLO1 with 0.1 mM of IPTG. Panels C and D contain the spectrum of crude cell lysate of His*Li*GLO1 without and with IPTG, respectively.

The analysis of the crude cell lysates of *E. coli* cells transformed with plasmid pET23a-*Li*GLO1 confirmed that it was not necessary to add IPTG for protein expression in TB medium. Thus, the existing basal expression is sufficient to obtain a clear spectrum. The existing of basal protein expression, even without induction, had already been observed in results published by other authors [58].

Moreover, the observation of peaks corresponding to the native, homodimeric form, in the spectra A and B of Figure 4.15 also indicated that the protein was correctly folded. In fact, the spectra obtained from crude cell lysates was comparable to one recorded for purified *Li*GLO1, (Figure 4.11 B).

The spectra of crude cell lysates of *E. coli* transformed with plasmid pET28a-His*Li*GLO1 (Figure 4.15 C,D) are comparable to the samples of the Figure 4.13 that were grown at 37ºC. At 28ºC there was the appearance of one more charge state (+13) and it was observed that the error in the calculation of the mass was smaller. These results are due to the greater amount of protein produced at this temperature, as can be seen in the gel (Figure 4.14). Overall, all spectra acquired revealed highly resolved peaks,

narrow charge state distributions and the anticipated stoichiometry, regardless of the high-mass spectrometer used.

In conclusion, the conditions of expression that maximize the transmission of *Li*GLO1 were the growth in TB medium at 28ºC overnight, since with this conditions, the amount of protein expression is higher.

4.4. Analysis of protein complex in crude cell lysates in Q-TOF

After optimising the emitter tip and the sample preparation conditions of expression in *E. coli* cells that formed a clear spectrum, some parameters were optimised in the mass spectrometers to allow the native study of the complexes from crude cell lysate, namely source temperatures and voltage of collision energy, CID.

As mentioned above, the Q-TOF is already modified to transmit higher *m/z*. Major modifications of the Q-TOF mass spectrometer include: changes in the pressure gradient in the different pumping stages of the instrument and reducing the radio-frequency of the quadrupole mass analyser. When compared with conventional Q-TOF instruments, the pressure of all pumping stages is increased and, as a consequence, transmission of high *m/z* ions is improved. This is mainly achieved by collisional cooling and, at the same time, collisional focusing of high *m/z* ions in the hexapole ion guide of the initial vacuum stage. Quadrupoles transmit ions to an upper limit which depends on the radio frequency and amplitude as well as the diameter (or inner radius) between the quadrupole rods [113].

In addition to the above discussed changes in the pressure gradient and the quadrupole, the aperture of the collision-gas cell entrance and exit is decreased to allow higher collisional pressure; at the same time, the orifice is optimised for transmission of precursor and product ions obtained in native-MS experiments [84].

Temperature and gas flow rate are parameters that affect the desolvation efficiency. Improper settings may result in loss of signal [114]. However, since these are native experiments, it is necessary to use parameters that maintain the non-covalent interactions and reduce the noise caused by all the other proteins in the extract.

CID is one of the most common methods used to dissociate protein complexes in the gas phase, primarily because of its ease of implementation. In this process, as already mentioned, ions from a protein complex are accelerated in an electric field into a gas-filled region of a mass spectrometer, where they undergo multiple low-energy collisions with inert gas molecules, converting their translational energy to internal energy. The gradual increase in internal energy eventually causes disruption of non-covalent interactions and ejection of a highly charged monomeric protein. Mechanistic details governing the release of various subunits within a protein complex remain under investigation. The current consensus, however, is that the key factors in determining which subunit is ejected are the local concentration of surface charges on the complex and the propensity for the ejected subunit to unfold [115].

In this way, the stability of crude cell lysate of cells expressing *Li*GLO1 were analysed and the spectra showed in Figure 4.16.

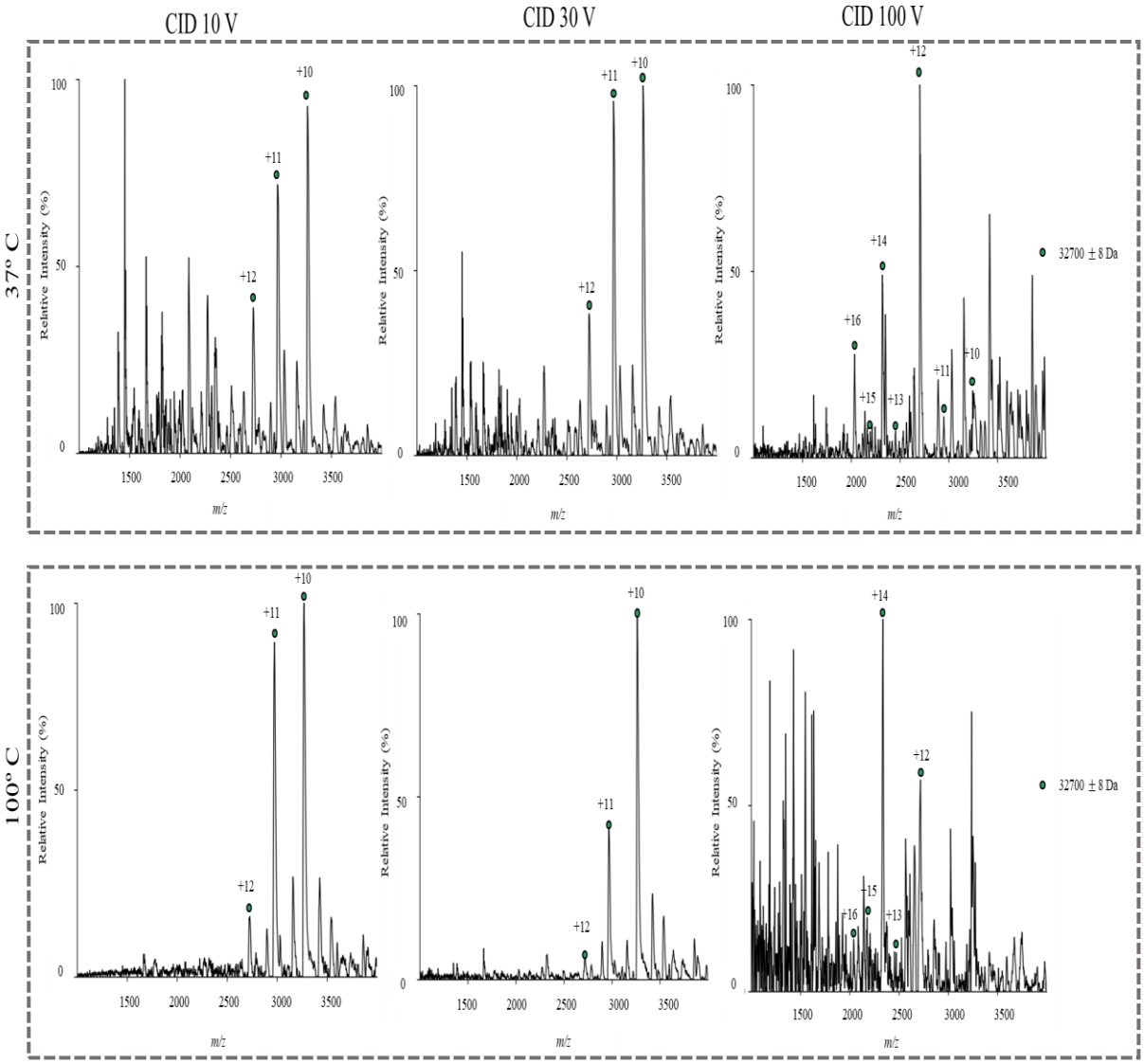


Figure 4.16 The effect of source temperature and CID in spectra of protein complex in crude cell lysates. Increasing the temperature does not lead to changes in charge state distribution. However, a high collision energy (100 V) led to an increase in distribution of charge states, indicating changes in protein conformation. The parameters that increase the signal were a source temperature of 100ºC and CID 30 V, since these values produced the best spectrum and maintained the native conformation of protein complex.

Increase of source temperature (37ºC to 100ºC) resulted in reduced levels of adducting and improved ion transmission is also observed. For lower collision energy (10 V and 30 V) there was no change in charge state distribution, demonstrating that an increase in source temperature improved the transmission of protein overexpressed and did not lead to perceptible conformation changes.

A too high collision energy, as 100 V led to an increased distribution of charge states, indicating that protein conformation changed. However, this voltage was not high enough to destroy the dimeric complex, since it continued to dominate the spectra and the monomeric species was not detected.

4.5. Analysis of protein complex in crude cell lysates in FT-ICR

Although less frequently used for this propose, and restricted to a handful of groups worldwide, FT-ICR mass spectrometers are also appropriate to investigate intact proteins and protein assemblies. However, unlike Q-TOF, FT-ICR mass spectrometer was not optimised for high *m/z* transmission. Therefore, for the study of protein complexes through crude cell lysates in an FT-ICR mass spectrometer was necessary optimise several steps, from the nanoESI source to the detection of high *m/z* ions [116,117].

Therefore, after optimising the sample expression conditions, a crude cell lysate of cells growth at 28ºC in TB medium was obtained, and this sample was injected in a FT-ICR using the emitter tip produced by program 5 in nanoESI source. Since the values used for the study of denatured protein are too aggressive and fail in maintain non-covalent interactions, new parameters were adjusted to produce a native spectrum.

When the FT-ICR mass spectrometer is used for the analysis of intact proteins, the performance can be affected by space-charge affects, being these a consequence of the mutual repulsion between particles of like charge and are significant due to the high number of charges per analyte that is introduced in the mass analyser. However, these multiply charged analytes induce larger image currents allowing even single molecule detection.

In FT-ICR trap, the magnetic field induces next to the cyclotron motion also high-magnitude magnetron motions especially for high mass ions, limiting their trapping efficiency. Resolved isotopes generate beat patterns within the transients recorded in FT-MS. The increased number of isotopes for high-mass ions results in narrower beats with longer in-between time intervals without significant signal, consequently leading to a signal decrease. That way, acquiring high resolution in FT-ICR requires sufficiently long measurement of the image current transients, which may go up to minutes to obtain isotope resolution for a large protein. This required stable oscillations and stable ion current of the analyte ions but field imperfections, space-charge effects and collisions with background gas ions contribute to a rapid decay [118].

FT-ICR is capable of higher resolution than any other mass spectrometer, TOF instruments included. Chemical background is reduced in FT-MS since background ions need to oscillate in stable transients to be detected by the FT algorithm. Mass resolution is therefore not impaired by low signal intensity. Signal-to-noise ratios will however increase when individual transients are summed up.

Nevertheless, the inherent mass resolution of each mass analyser, i.e. the maximum instrument mass resolution, is found to deviate substantially from that obtained experimentally on native proteins. Apart from the maximum instrument mass resolution, there are additional, mass analyser independent factors that influence the native resolution, in particular incomplete desolvation of the analyte and/or unwanted association buffer or salt molecules as well as the width of the analyte's natural isotope distribution.

The apparent mass resolution observed for protein in native-MS is predominantly influenced by the three factors - inherent mass resolution of the mass analyser; width of the natural isotope distribution of the analyte ion and efficiency of desolvation in the ESI process, i.e. adduct removal from the analyte. The apparent resolution and the offset between calculated and experimental protein ion masses in native-MS are more affected by the efficiency of adduct elimination and analyte desolvation than by the inherent mass resolution of the current analysers [113,118].

Most adducts are stripped off in the source region and the collision cell of an instrument. Most importantly, elimination of small molecules adducts can be enhanced by collisional activation of protein ions.

Native-MS experiments are often performed with acceleration voltages that only just prevent subunit dissociation, but under such condition's conformational changes within the analysed biomolecular complex may already occur. In these cases, careful weighing of the benefit of improved desolvation against the danger of structural deterioration is required.

Some investigators studied several factors influencing the association of small molecules during the ESI process of proteins and the results suggest that the binding process is purely random and therefore not influenced by inherent protein characteristics such as size and structure. For the analysis of larger proteins by native-MS it is almost inevitable that some undesired adducts remain attached. Mostly, desolvation efficiency decreases with increasing protein complex size, causing the apparent resolution to decay irrespective of the inherent mass analyser [119].

The first aspect than need to be optimise is the level of ion desolvation and transmission. Previous studies refer the importance of three essential factors: source temperature, source skimmer and collision cell voltage. The authors conclude that reduction in charge state adducting and therefore the best desolvation can be achieved by increasing all these three factors. However, skimmer 1 voltage proved to be less efficient at removing adducts. Desolvation efficiency is greatly affected by the construction of both source region and the collision cell since most adducts may be stripped off all these stages. Despite the drawbacks of adduct elimination by CID, it is still important to exploit the full potential of this approach. Depending on the protein system and purpose of a native-MS study, the available collision energy and hence the degree of analyte desolvation may be insufficient to answer the scientific questions [113].

Therefore, the native purified recombinant *Li*GLO1 in FT-ICR was analysed, with the best results obtained when 30 V of collision energy was applied (Figure 4.17).

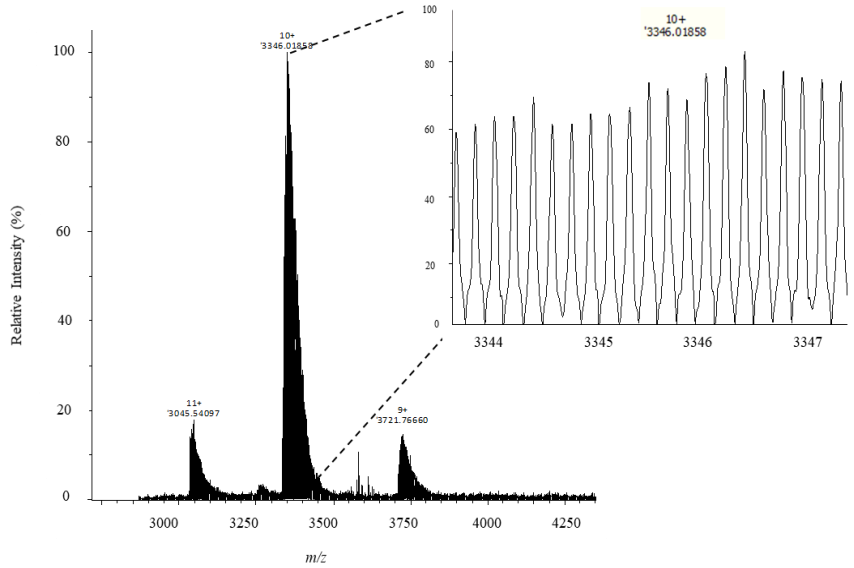


Figure 4.17 Spectrum of native recombinant *Li*GLO1 in FT-ICR mass spectrometer. The purified dimeric protein *Li*GLO1 was analysed using the emitter tip produced by program 5. The best spectrum was obtained by using 30 V of collision energy.

In this obtained spectrum it was possible observe the distribution of three charge states (+9 to +11) and a mass of 33460.1858 Da. These results are in agreement with those obtained in the Q-TOF mass spectrometer (Figure 4.11) and it was possible to observe that the protein complex was maintained.

However, during the analysis of this sample, it was observed that the position of emitter tip in nanoESI source is a very important factor that influences the ionisation efficiency, since the quality of the spectrum improves a lot when small adjustments were made. The source used was not fixed, unlike the nanoESI source used in Q-TOF, so was not possible to ensure that the emitter position was the same in each analysis, an issue that need to be solved and is related to both source and FT-ICR cart design, which generates some irreproducibility of results and requires sub millimetric adjustments. In the future it is necessary to solve this problem, fixing the nanoESI source, to allow only a fine adjustment in the position of emitter tip.

After finding the parameters that allow the transmission of high *m/z* and preserve the purified protein complex, the spectrum of *Li*GLO1 in crude cell lysate growth at 28ºC in TB medium was obtained.

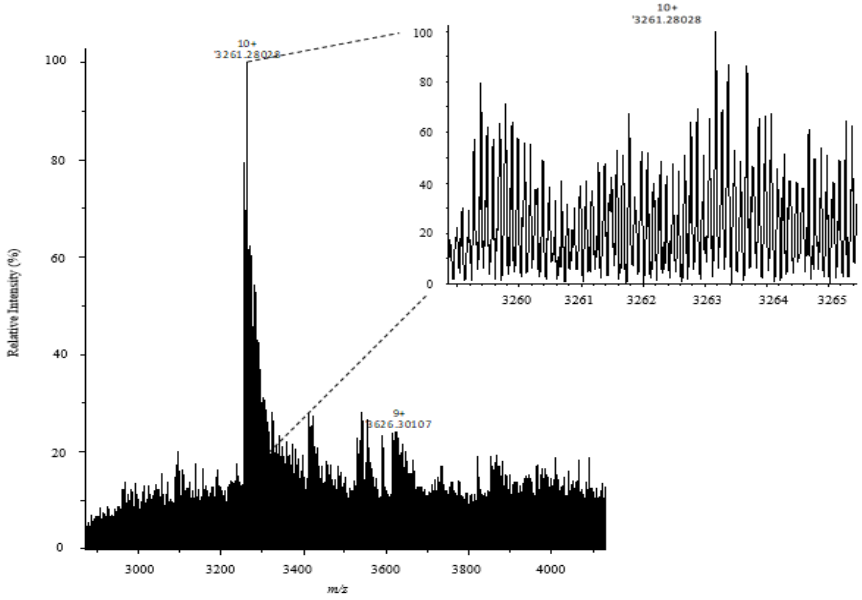


Figure 4.18 Spectrum of *E. coli* cell lysate expressing *Li*GLO1 in FT-ICR mass spectrometer. The crude cell lysate of cells with plasmid pET23a-LiGLO1was analysed in FT-ICR mass spectrometer using the nanoESI source and emitter tips produce in-house by program 5. A collision energy of 30 V was used.

In the analysis of *E. coli* cells transformed with plasmid pET23a-*Li*GLO1 in FT-ICR mass spectrometer was possible to observed a spectrum with some noise and only was detected two charges states (+9 and +10). However, a protein with 32612.2808 Da was obtained (Figure 4.18). A collision energy of 30 V proved to increase the quality of spectrum.

These results show that was possible to observe a protein complex directly from crude cell lysate, at isotopic resolution, in a FT-ICR mass spectrometer without dissociation of non-covalent bounds. To obtained this spectrum was necessary accumulate more 50 transients, comparing with the spectrum of purified *Li*GLO1 (Figure 4.17).

Beyond the spectrum of *Li*GLO1 in crude cell lysate growth at 28ºC in TB medium, was also obtained the spectrum of His*Li*GLO1 under the same conditions.

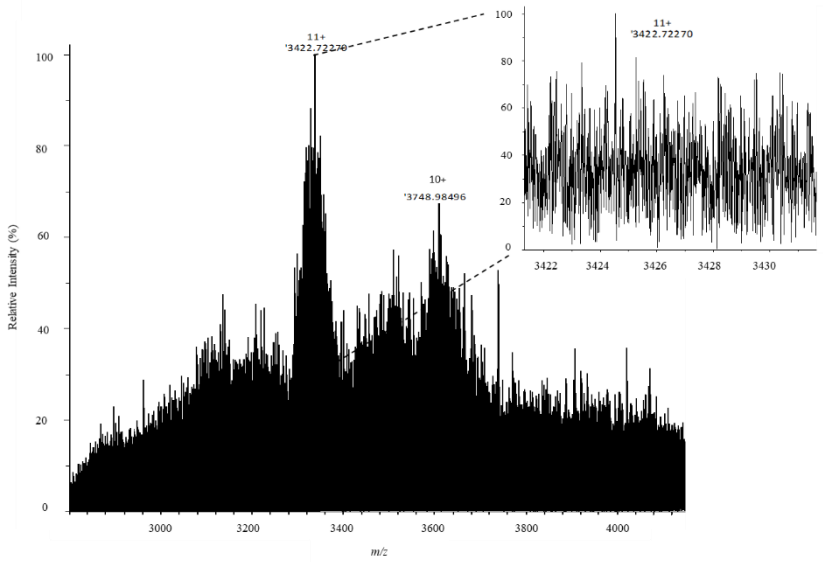


Figure 4.19 Spectrum of *E. coli* cell lysate expressing His*Li*GLO1 in FT-ICR mass spectrometer. The crude cell lysate of cells with plasmid pET28a-HisLiGLO1was also analysed in FT-ICR mass spectrometer using the nanoESI source and the emitter tips produce in-house by program 5. A collision energy of 30 V increases the signal transmission.

In the analysis of *E. coli* cells transformed with plasmid pET28a-His*Li*GLO1 in FT-ICR mass spectrometer (Figure 4.19) was possible to observed the presence of just two charge state (+11 and +10), similar to what was obtained with *E. coli* cells transformed with plasmid pET23a-*Li*GLO1. It was obtained a mass of 37649.9497 Da, which was higher than the mass obtained in Q-TOF mass spectrometer (Figure 4.15 D). As mentioned, this mass difference can be justified by the presence of adducts, which also justifies the high noise observed.

As mentioned, reducing the radio-frequency of the quadrupole mass analyser improves the transmission of higher *m/z* values. Lower RF frequencies in ion transmission regions proved to be better to transmit the signal of protein complexes in crude cell lysate.

All these results together show that was possible characterize the protein complex expressed in *E. coli* directly from the crude cell lysate in both mass spectrometers used. This approach allows the study and characterization of different protein complexes in a similar way to that found in the cell. In the future, this approach can be applied not only in characterization of homogenous complexes but also in heterogenous and can be used to study their stability. The next step being the use of MS/MS methods and their optimisation for complex characterizations, most notably ECD.

5. Conclusion

A simple method to characterize protein complexes from crude cell lysate without purification by native mass spectrometry was optimised. Using Glyoxalase I from *Leishmania infantum* as example, it was demonstrated that rich information can be obtained in a direct analysis of crude cell lysate overexpressing the protein complex with native MS experiments using Q-TOF and Fourier transform ion cyclotron mass spectrometers.

For MS of protein complexes, in contrary to other MS technologies, the protein assembly of interest is not enzymatically digested before the analysis but the assembly is ionised and is transferred into the gas phase of the mass spectrometer as a whole, directly from crude cell lysate. This is achieved by a gentle ionisation process at atmospheric pressure, typically from buffered solutions. Additionally, careful optimisation of instrument conditions, namely pressure and accelerating voltages, is required to safeguard the traversal of the intact complex through the instrument without dissociation. The obtained mass spectra from crude cell lysates reveal information on subunit interactions and topology as well as heterogeneity of the assemblies. The mass range spans from individual subunits or ligands up to large assemblies giving clues on the stability of the complexes as well as preferred interactions. Native-MS, consequently, takes an important role in the structural analysis of proteins and protein complexes by MS.

The use of adequate nanoESI emitter tips with a stable ion current, leads to clearly resolved charge-state distributions of proteins and protein complexes. The formation of electrospray from the tip of the emitter proved to be an important factor in the study of protein complexes, since it influences the distribution of charge states. When squirt occurs, instead of spray, a small droplet is not formed and the proportion of charged molecules produced decreases, lowering the efficiency of ions formed.

The fundamental requirement for the direct analysis of cell extract is that the overexpressed target became the dominant protein, such that outperforms the levels of endogenous background proteins. Beside the concentration, the method is also highly dependent on the abundance, molecular weight (MW) and ionisation efficiency of both generated and background proteins. In this way, was necessary to optimise the protein expression to maximize the quality of spectra obtained in native analysis of protein complexes. Growth at 28ºC in TB medium proved to increase protein expression of *Li*GLO1, improving the quality of the spectrum obtained, since the protein complexes dominate the signal.

After optimising the emitter tips and the sample preparation that increased the transmission and the signal of protein complexes, the parameters in each mass spectrometers were optimised. In Q-TOF a collision energy of 30 V and a source temperature of 100ºC produced the best spectrum, without having dissociation of protein complex. In FT-ICR, the ionisation proved to be very sensitive to the emitter tip position. Since the source is not fixed, emitter tip position varied in each analysis, decreasing reproducibility. It was possible to observe the protein complex from crude cell lysate, with isotopic resolution, without purification.

This method is equally applicable to other protein complexes overexpressed in *E. coli*. Future studies will include the application of this method to larger and heterogenous complexes and optimisation of MS/MS methods for complex characterizations.

6. Bibliography

1. Wood EJ. Nature's robots: A history of proteins. *Biochem Mol Biol Educ*. 2004;32(5):282-282. doi:10.1002/bmb.2004.494032059999

2. Nelson DL, Cox MM. *Lehninger Principles Of Biochemistry*. 5th ed.; 2008.

3. Marsh JA, Teichmann SA. Structure , Dynamics , Assembly , and Evolution of Protein Complexes. 2014;(November):1-25. doi:10.1146/annurev-biochem-060614-034142

4. Mittal S, Chowhan RK, Singh LR. Macromolecular crowding: Macromolecules friend or foe. *Biochim Biophys Acta - Gen Subj*. 2015;1850(9):1822-1831. doi:10.1016/j.bbagen.2015.05.002

5. Minton AP. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*. 1983;55(2):119-140. doi:10.1007/BF00673707

6. Minton AP. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J Biol Chem*. 2001;276(14):10577-10580. doi:10.1074/jbc.R100005200

7. Ellis RJ. Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol*. 2001;11(1):114-119. doi:10.1016/S0959-440X(00)00172-X

8. Tabaka M, Kalwarczyk T, Szymanski J, Hou S, Holyst R. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front Phys*. 2014;2(September):1-14. doi:10.3389/fphy.2014.00054

9. Kim YC, Best RB, Mittal J. Macromolecular crowding effects on protein-protein binding affinity and specificity. *J Chem Phys*. 2010;133(20). doi:10.1063/1.3516589

10. Zhang H, Cui W, Wen J, Blankenship RE, Gross ML. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal Chem*. 2011;83(14):5598-5606. doi:10.1021/ac200695d

11. Acuner Ozbabacan SE, Engin HB, Gursoy A, Keskin O. Transient proteinprotein interactions. *Protein Eng Des Sel*. 2011;24(9):635-648. doi:10.1093/protein/gzr025

12. McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr Sect FStructural Biol Commun*. 2014;70:1445-1467. doi:10.1107/S2053230X14019670

13. Schluenzen F, Tocilj A, Zarivach R, et al. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*. 2000;102(5):615-623. doi:10.1016/S0092-8674(00)00084-2

14. Markwick PRL, Malliavin T, Nilges M. Structural biology by NMR: Structure, dynamics, and interactions. *PLoS Comput Biol*. 2008;4(9). doi:10.1371/journal.pcbi.1000168

15. Fan X, Wang J, Zhang X, et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat Commun*. 2019;10(1):1-11. doi:10.1038/s41467-019-10368-w

16. Wu M, Lander GC, Herzik MA. Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J Struct Biol X*. 2020;4(February):100020. doi:10.1016/j.yjsbx.2020.100020

17. Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. *Cell*. 2015;161(3):438-449. doi:10.1016/j.cell.2015.03.050

18. Snijder J, Rose RJ, Veesler D, Johnson JE, Heck AJR. Studying 18 MDa virus assemblies with

native mass spectrometry. *Angew Chemie - Int Ed*. 2013;52(14):4020-4023. doi:10.1002/anie.201210197

19. Forsberg E, Fang M, Siuzdak G. Staying Alive: Measuring Intact Viable Microbes with Electrospray Ionization Mass Spectrometry. *J Am Soc Mass Spectrom*. 2017;28(1):14-20. doi:10.1007/s13361-016-1440-y

20. Tanaka K, Waki H, Ido Y, et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*. 1988;2(8):151-153. doi:10.1002/rcm.1290020802

21. Karas M, Hillenkamp F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal Chem*. 1988;60(20):2299-2301. doi:10.1021/ac00171a028

22. Jonsson AP. Mass spectrometry for protein and peptide characterisation. *Cell Mol Life Sci*. 2001;58(7):868-884. doi:10.1007/PL00000907

23. Dass C. *Fundamentals of Contemporary Mass Spectrometry*.; 2006. doi:10.1002/9780470118498

24. Fenn JB, Mann M, Meng CKAI, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-)*. 1989;246(6):64-71. doi:10.1126/science.2675315

25. Ganem B. Observation of Noncovalent Enzymesubstrate and Enzyme-Product Complexes by Ion-Spray Mass Spectrometry. 1991;(8):7818-7819. doi:10.1021/ja00020a085

26. Miranker A, Robinson C V., Radford SE, Aplin RT, Dobson CM. Detection of transient protein folding populations by mass spectrometry. *Science (80-)*. 1993;262(5135):896-900. doi:10.1126/science.8235611

27. Mirza UA, Cohen SL, Chait BT. Heat-Induced Conformational Changes in Proteins Studied by Electrospray Ionization Mass Spectrometry. *Anal Chem*. 1993;65(1):1-6. doi:10.1021/ac00049a003

28. Heck AJR, Van Den Heuvel RHH. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev*. 2004;23(5):368-389. doi:10.1002/mas.10081

29. Iribarne J V., Thomson BA. On the evaporation of small ions from charged droplets. *J Chem Phys*. 1976;64(6):2287-2294. doi:10.1063/1.432536

30. Fernandez De La Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta*. 2000;406(1):93-104. doi:10.1016/S0003-2670(99)00601-7

31. Dawson PH. Quadrupole mass analyzers: Performance, design and some recent applications. *Mass Spectrom Rev*. 1986;5(1):1-37. doi:10.1002/mas.1280050102

32. Graham Cooks R, Glish GL, McLuckey SA, Kaiser RE. Ion trap mass spectrometry. *Chem Eng News*. 1991;69(12):26-41. doi:10.1021/cen-v069n012.p026

33. Guilhaus M. Principles and Instrumentation in Time-of-flight Mass Spectrometry. *J Mass Spectrom*. 1995;30(November):1519-1532. doi:DOI: 10.1002/jms.1190301102

34. Marshall AG, Hendrickson CL, Jackson GS. *Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*.; 2006. doi:10.1002/9780470027318.a6006m

35. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom*. 2004;39(10):1091-1112. doi:10.1002/jms.703

36. Jennings KR. Colllision-Induced Decompositions of Aromatic Ions. *Int J Mass Spectrom*.

1968;1:227-235. doi:10.1016/0020-7381(68)85002-8

37. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*. 2004;15(1):12-16. doi:10.1016/j.copbio.2003.12.002

38. Emmett MR, Caprioli RM. Ultra-High-Sensitivity Analysis of Peptides and Proteins. *J Am Soc Mass Spectrom*. 1994;5(7):605-613. doi:10.1016/1044-0305(94)85001-1

39. Wilm MS, Mann M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int J Mass Spectrom Ion Process*. 1994;136(2-3):167-180. doi:10.1016/0168-1176(94)04024-9

40. El-Faramawy A, Siu KWM, Thomson BA. Efficiency of nano-electrospray ionization. *J Am Soc Mass Spectrom*. 2005;16(10):1702-1707. doi:10.1016/j.jasms.2005.06.011

41. Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. *Anal Chem*. 1996;68(1):1-8. doi:10.1021/ac9509519

42. Karas M, Bahr U, Dülcks T. Nano-electrospray ionization mass spectrometry: Addressing analytical problems beyond routine. *Fresenius J Anal Chem*. 2000;366(6-7):669-676. doi:10.1007/s002160051561

43. Heck AJR. Native mass spectrometry: A bridge between interactomics and structural biology. *Nat Methods*. 2008;5(11):927-933. doi:10.1038/nmeth.1265

44. Susa AC, Xia Z, Williams ER. Native Mass Spectrometry from Common Buffers with Salts That Mimic the Extracellular Environment. *Angew Chemie - Int Ed*. 2017;56(27):7912-7915. doi:10.1002/anie.201702330

45. Puglisi R, Boeri Erba E, Pastore A. A Guide to Native Mass Spectrometry to determine complex interactomes of molecular machines. *FEBS J*. 2020:febs.15281. doi:10.1111/febs.15281

46. Ben-Nissan G, Sharon M. Capturing protein structural kinetics by mass spectrometry. *Chem Soc Rev*. 2011;40(7):3627-3637. doi:10.1039/c1cs15052a

47. Leney AC, Heck AJR. Native Mass Spectrometry: What is in the Name? *J Am Soc Mass Spectrom*. 2017;28(1):5-13. doi:10.1007/s13361-016-1545-3

48. Siuzdak G, Bothner B, Yeager M, et al. Mass spectrometry and viral analysis. *Chem Biol*. 1996;3(1):45-48. doi:10.1016/S1074-5521(96)90083-6

49. van Duijn E. Current limitations in native mass spectrometry based structural biology. *J Am Soc Mass Spectrom*. 2010;21(6):971-978. doi:10.1016/j.jasms.2009.12.010

50. Erba EB, Petosa C. The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci*. 2015;24(8):1176-1192. doi:10.1002/pro.2661

51. Skinner OS, Haverland NA, Fornelli L, et al. Top-down characterization of endogenous protein complexes with native proteomics. *Nat Chem Biol*. 2018;14(1):36-41. doi:10.1038/nchembio.2515

52. Muneeruddin K, Nazzaro M, Kaltashov IA. Characterization of Intact Protein Conjugates and Biopharmaceuticals Using Ion-Exchange Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry and Top-Down Tandem Mass Spectrometry. *Anal Chem*. 2015;87(19):10138-10145. doi:10.1021/acs.analchem.5b02982

53. Uetrecht C, Lorenzen K, Kitel M, et al. Native mass spectrometry provides sufficient ion flux for XFEL single-particle imaging. *J Synchrotron Radiat*. 2019;26(3):653-659. doi:10.1107/S1600577519002686

54. Loo JA. Electrospray ionization mass spectrometry: A technology for studying noncovalent macromolecular complexes. *Int J Mass Spectrom*. 2000;200(1-3):175-186. doi:10.1016/S1387-3806(00)00298-0

55. Kafader JO, Melani RD, Schachner LF, et al. Native vs Denatured: An in Depth Investigation of Charge State and Isotope Distributions. *J Am Soc Mass Spectrom*. 2020. doi:10.1021/jasms.9b00040

56. Li H, Nguyen HH, Loo RRO, Campuzano IDG, Loo JA. An integrated native mass spectrometry and topdown proteomics method that connects sequence to structure and function of macromolecularcomplexes. *Nat Chem*. 2018;10(2):139-148. doi:10.1038/NCHEM.2908

57. Robinson C V, Herna H. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. 2007;2(3):715-726. doi:10.1038/nprot.2007.73

58. Gan J, Ben-nissan G, Arkind G, et al. Native Mass Spectrometry of Recombinant Proteins from Crude Cell Lysates. 2017. doi:10.1021/acs.analchem.7b00398

59. Ben-Nissan G, Vimer S, Warszawski S, et al. Rapid characterization of secreted recombinant proteins by native mass spectrometry. *Commun Biol*. 2018;1(1):1-12. doi:10.1038/s42003-018-0231-3

60. Vimer S. Direct characterization of overproduced proteins by native mass spectrometry. *Nat Protoc*. 2019;10. doi:10.1038/s41596-019-0233-8

61. Chernushevich I V, Thomson BA. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal Chem*. 2004;76(6):1754-1760. doi:10.1021/ac035406j

62. Sobott F, Hernández H, McCammon MG, Tito MA, Robinson C V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem*. 2002;74(6):1402-1407. doi:10.1021/ac0110552

63. Rostom AA, Robinson C V. Detection of the intact GroEL chaperonin assembly by mass spectrometry. *J Am Chem Soc*. 1999;121(19):4718-4719. doi:10.1021/ja990238r

64. Sanglier S, Leize E, Dorsselaer A Van, Zal F. Comparative ESI-MS study of ~2.2 MDa native hemocyanins from deep-sea and shore crabs: From protein oligomeric state to biotope. *J Am Soc Mass Spectrom*. 2003;14(5):419-429. doi:10.1016/S1044-0305(03)00131-4

65. Tahallah N, Pinkse M, Maier CS, Heck AJR. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun Mass Spectrom*. 2001;15(8):596-601. doi:10.1002/rcm.275

66. Douglas DJ, French JB. Collisional focusing effects in radio frequency quadrupoles. *J Am Soc Mass Spectrom*. 1992;3(4):398-408. doi:10.1016/1044-0305(92)87067-9

67. Thornalley PJ. The glyoxalase system: New developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem J*. 1990;269(1):1-11. doi:10.1042/bj2690001

68. Thornalley PJ. The glyoxalase system in health and disease. *Mol Aspects Med*. 1993;14(4):287-371. doi:10.1016/0098-2997(93)90002-U

69. Sousa Silva M, Ferreira AEN, Gomes R, Tomás AM, Ponces Freire A, Cordeiro C. The glyoxalase pathway in protozoan parasites. *Int J Med Microbiol*. 2012;302(4-5):225-229. doi:10.1016/j.ijmm.2012.07.005

70. Silva MS, Ferreira AEN, Tomás AM, Cordeiro C, Freire AP. Quantitative assessment of the glyoxalase pathway in Leishmania infantum as a therapeutic target by modelling and computer simulation. *FEBS J*. 2005;272(10):2388-2398. doi:10.1111/j.1742-4658.2005.04632.x

71. Sukdeo N, Clugston SL, Daub E, Honek JF. Distinct classes of glyoxalase I: Metal specificity of the Yersinia pestis, Pseudomonas aeruginosa and Neisseria meningitidis enzymes. *Biochem J*. 2004;384(1):111-117. doi:10.1042/BJ20041006

72. Aronsson AC, Marmstål E, Mannervik B. Glyoxalase I, a zinc metalloenzyme of mammals and yeast. *Biochem Biophys Res Commun*. 1978;81(4):1235-1240. doi:10.1016/0006-291X(78)91268-8

73. Ridderström M, Cameron AD, Jones TA, Mannervik B. Involvement of an active-site Zn2+ ligand in the catalytic mechanism of human glyoxalase I. *J Biol Chem*. 1998;273(34):21623-21628. doi:10.1074/jbc.273.34.21623

74. He MM, Clugston SL, Honek JF, Matthews BW. Determination of the structure of Escherichia coli glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry*. 2000;39(30):8719-8727. doi:10.1021/bi000856g

75. Ariza A, Vickers TJ, Greig N, Fairlamb AH, Bond CS. Crystallization and preliminary X-ray analysis of Leishmania major glyoxalase I. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2005;61(8):769-772. doi:10.1107/S174430910502169X

76. Barata L, Sousa Silva M, Schuldt L, et al. Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of glyoxalase i from Leishmania infantum. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010;66(5):571-574. doi:10.1107/S1744309110010754

77. Barata L. Methylglyoxal Metabolism in Leishmania infantum Lídia Isabel Sebastião Barata. 2010.

78. Schrödinger L. *The {PyMOL} Molecular Graphics System, Version 2.*; 2015.

79. Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson C V. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal Chem*. 2015;87(8):4370-4376. doi:10.1021/acs.analchem.5b00140

80. Reid DJ, Diesing JM, Miller MA, et al. MetaUniDec : High-Throughput Deconvolution of Native. 2018:118-127. doi:10.1007/s13361-018-1951-9

81. Bateman A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515. doi:10.1093/nar/gky1049

82. Nguyen GTH, Tran TN, Podgorski MN, Bell SG, Supuran CT, Donald WA. Nanoscale Ion Emitters in Native Mass Spectrometry for Measuring Ligand-Protein Binding Affinities. *ACS Cent Sci*. 2019;5(2):308-318. doi:10.1021/acscentsci.8b00787

83. Li Y, Cole RB. Shifts in Peptide and Protein Charge State Distributions with Varying Spray Tip Orifice Diameter in Nanoelectrospray Fourier Transform Ion Cycltron Resonance Mass Spectrometry. *Anal Chem*. 2003;75(21):5739-5746. doi:10.1021/ac0301402

84. Susa AC, Xia Z, Williams ER. Small Emitter Tips for Native Mass Spectrometry of Proteins and Protein Complexes from Nonvolatile Buffers That Mimic the Intracellular Environment. *Anal Chem*. 2017;89(5):3116-3122. doi:10.1021/acs.analchem.6b04897

85. Prabhu GRD, Ponnusamy VK, Witek HA, Urban PL. Sample Flow Rate Scan in Electrospray Ionization Mass Spectrometry Reveals Alterations in Protein Charge State Distribution. *Anal Chem*. 2020;92(19):13042-13049. doi:10.1021/acs.analchem.0c01945

86. Zaia J, Annan RS, Biemann K. The correct molecular weight of myoglobin, a common calibrant for mass spectrometry. *Rapid Commun Mass Spectrom*. 1992;6(1):32-36. doi:10.1002/rcm.1290060108

87. Konermann L, Rosell FI, Mauk AG, Douglas DJ. Acid-induced denaturation of myoglobin studied by time-resolved electrospray ionization mass spectrometry. *Biochemistry*. 1997;36(21):6448-6454. doi:10.1021/bi970353j

88. Li YT, Hsieh YL, Henion JD, Ganem B. Studies on heme binding in myoglobin, hemoglobin, and cytochrome c by ion spray mass spectrometry. *J Am Soc Mass Spectrom*. 1993;4(8):631-637. doi:10.1016/1044-0305(93)85027-U

89. Flanagan M a, E BG, Friend SH, Feldmann RJ, Scouloudi H, Gurd FRN. Contributions of Individual Amino Acid Residues o the Structural Stability. *Differences*. 1983:6027-6037. doi:10.1021/bi00294a051

90. Friend SH, Gurd FRN. Electrostatic Stabilization in Myoglobin. pH Dependence of Summed Electrostatic Contributions†. *Biochemistry*. 1979;18(21):4612-4619. doi:10.1021/bi00588a023

91. Antonini Eraldo, Brunori M. *Hemoglobin and Myoglobin in Their Reactions with Ligands*. Amsterdam : North-Holland; 1971. doi:10.1126/science.178.4058.296

92. Ray SS, Singh SK, Balaram P. Sulfonate (ANS) Binding to Proteins. *Data Process*. 2001;0305(01). doi:10.1016/S1044-0305(01)00206-9

93. Chait VK and BT. Observation of the Hemdlobin Complex in Native Myoglobin by Electrospray-Ionization Mass Spectrometry. 1991;25000(13):8534-8535.

94. Eklund H, Nordström B, Zeppezauer E, et al. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J Mol Biol*. 1976;102(1):27-59. doi:10.1016/0022-2836(76)90072-3

95. Ganzhorn AJ, Green DW, Hershey AD, Gould RM, Plapp B V. Kinetic characterization of yeast alcohol dehydrogenases. Amino acid residue 294 and substrate specificity. *J Biol Chem*. 1987;262(8):3754-3761.

96. Jörnvall H, Hedlund J, Bergman T, Kallberg Y, Cederlund E, Persson B. Origin and evolution of medium chain alcohol dehydrogenases. *Chem Biol Interact*. 2013;202(1-3):91-96. doi:10.1016/j.cbi.2012.11.008

97. Eklund H, Samama J-P, Brändén C-I, ÅkesonT Å, Wallén L, Jones A. Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. *J Mol Biol*. 1981;146(4):561-587. doi:10.1016/0022-2836(81)90047-4

98. Eklund H, Ramaswamy S. Medium- and short-chain dehydrogenase/reductase gene and protein families: Three-dimensional structures of MDR alcohol dehydrogenases. *Cell Mol Life Sci*. 2008;65(24):3907-3917. doi:10.1007/s00018-008-8589-x

99. Lutstorf U, Megnet R. Multiple forms of alcohol dehydrogenase in Saccharomyces cerevisiae. I. Physiological control of ADH-2 and properties of ADH-2 and ADH-4. *Arch Biochem Biophys*. 1968;126(3):933-944. doi:10.1016/0003-9861(68)90487-6

100. Jornvall H. The Primary Structure of Yeast Alcohol Dehydrogenase. *Eur J Biochem*. 1977;72(3):425-442. doi:10.1111/j.1432-1033.1977.tb11267.x

101. Moniatte M, Lesieur C, Vécsey-Semjén B, et al. Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry in the subunit stoichiometry study of high-mass non-covalent complexes. *Int J Mass Spectrom Ion Process*. 1997;169-170(97):179-199. doi:10.1016/s0168-1176(97)00213-9

102. Farmer TB, Caprioli RM. Assessing the multimeric states of proteins: Studies using laser desorption mass spectrometry. *Biol Mass Spectrom*. 1991;20(12):796-800. doi:10.1002/bms.1200201209

103. Bühner M, Sund H. Yeast Alcohol Dehydrogenase: –SH Groups, Disulfide Groups, Quaternary

Structure, and Reactivation by Reductive Cleavage of Disulfide Groups. *Eur J Biochem*. 1969;11(1):73-79. doi:10.1111/j.1432-1033.1969.tb00741.x

104. Loo JA. Observation of large subunit protein complexes by electrospray ionization mass spectrometry. *J Mass Spectrom*. 1995;30(1):180-183. doi:10.1002/jms.1190300127

105. Kägi JHR, Vallee BL. The role of zinc in alcohol dehydrogenase. *J Biol Chem*. 1960;235(11):3188-3192.

106. Raj SB, Ramaswamy S, Plapp B V. Yeast Alcohol Dehydrogenase Structure and Catalysis. *Biochemistry*. 2014;53(36):5791-5803. doi:10.1021/bi5006442

107. Chi A, Huttenhower C, Geer LY, et al. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A*. 2007;104(7):2193-2198. doi:10.1073/pnas.0607084104

108. Marshall AG, Hendrickson CL. Fourier transform ion cyclotron resonance detection: Principles and experimental configurations. *Int J Mass Spectrom*. 2002;215(1-3):59-75. doi:10.1016/S1387-3806(01)00588-7

109. Sharon M, Witt S, Glasmacher E, Baumeister W, Robinson C V. Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J Biol Chem*. 2007;282(25):18448-18457. doi:10.1074/jbc.M701534200

110. Engen JR. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal Chem*. 2009;81(19):7870-7875. doi:10.1021/ac901154s

111. Stocks BB, Konermann L. Time-dependent changes in side-chain solvent accessibility during cytochrome c folding probed by pulsed oxidative labeling and mass spectrometry. *J Mol Biol*. 2010;398(2):362-373. doi:10.1016/j.jmb.2010.03.015

112. Turra GL, Agostini RB, Fauguel CM, et al. Structure of the novel monomeric glyoxalase I from Zea mays. *Acta Crystallogr Sect D Biol Crystallogr*. 2015;71(2015):2009-2020. doi:10.1107/S1399004715015205

113. Lössl P, Snijder J, Heck AJR. Boundaries of mass resolution in native mass spectrometry. *J Am Soc Mass Spectrom*. 2014;25(6):906-917. doi:10.1007/s13361-014-0874-3

114. Paíga P, Silva LMS, Delerue-Matos C. Optimization of the Ion Source-Mass Spectrometry Parameters in Non-Steroidal Anti-Inflammatory and Analgesic Pharmaceuticals Analysis by a Design of Experiments Approach. *J Am Soc Mass Spectrom*. 2016;27(10):1703-1714. doi:10.1007/s13361-016-1459-0

115. Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, et al. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J Am Soc Mass Spectrom*. 2016;27(3):507-519. doi:10.1007/s13361-015-1308-6

116. Campuzano IDG, Nshanian M, Spahr C, et al. High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *J Am Soc Mass Spectrom*. 2020;31(5):1155-1162. doi:10.1021/jasms.0c00030

117. Tolmachev A V., Errol W. Robinson SW, Paša-Tolić L, Smith RD. FT-ICR MS optimization for the analysis of intact proteins. *Mas spectrom*. 2009. doi:10.1016/j.ijms.2008.10.010.FT-ICR

118. Barth M, Schmidt C. Native mass spectrometry—A valuable tool in structural biology. *J Mass Spectrom*. 2020;55(10):0-1. doi:10.1002/jms.4578

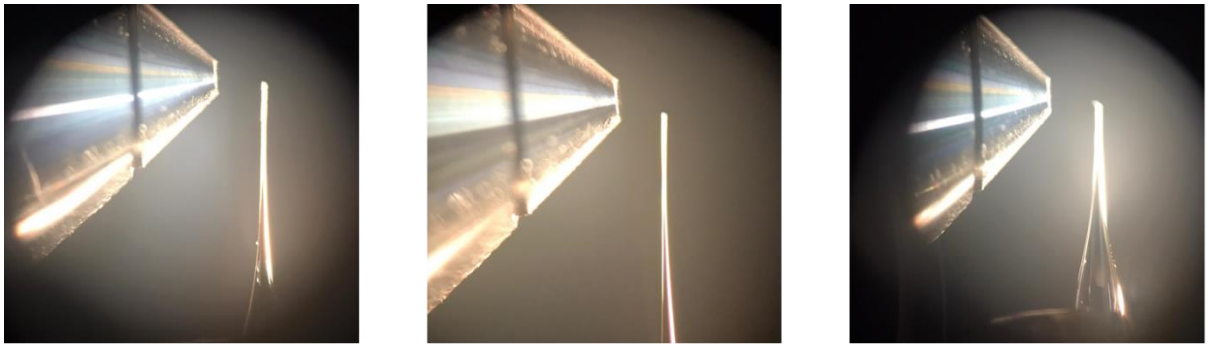
119. Sun N, Soya N, Kitova EN, Klassen JS. Nonspecific Interactions Between Proteins and Charged Biomolecules in Electrospray Ionization Mass Spectrometry. *J Am Soc Mass*

Spectrom. 2010;21(3):472-481. doi:10.1016/j.jasms.2009.12.002

7. Annexes

Supplementary table 7.1 Programs and respective parameters for the emitter formation with different tips shapes, with unit corrections. The values presented in the instrument are in arbitrary units. In this table are the values in known units. The heat is the current applied, in Amperes, pull and velocity are given in units. Time and delay are in ms and pressure in psi.

Program	Glass	Heat (A)	Pull	Velocity	Time (ms)	Delay (ms)	Pressure (psi)
1	1.2 mm x 0.69 mm	450	-	155	40	-	2
		400	-	15	25	-	
		450	200	20	40	-	
2		450	-	8	125	-	
		400	200	6	100	-	
3		450	-	8	75	-	
		400	200	3	125	-	
4		470	130	80	50	-	
5		470	150	80	-	100	
6		450	200	60	-	160	1.6
7	1.0 mm x 0.78 mm	400	225	60	-	160	



Emitter tip 1 Emitter tip 5 Emitter tip 6

Supplementary figure 7.1 Shape of each emitter tip. Dependence of ESI characteristics on the diameter of the emitter orifice. To produce a high ionisation efficiency, it is important to have a long tip with a fine orifice, in order to produce a small droplet.

MGLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDL
KKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGAD
AQGAMTKALELFRNDIAAKYKELGFQG

Supplementary figure 7.2 Myoglobin sequence. The sequence of myoglobin from equine heart is formed by 154 amino acids. The initiator methionine is removed, which result in a mass with 16951 Da.

MSIPETQKAIIFYESNGKLEHKDIPVPKPKPNELLINVKYSGVCHTDLHAWHGDWPLPTKLPL
VGGHEGAGVVVGMGENVKGWKIGDYAGIKWLNGSCMACEYCELGNESNCPHADLSGYTH
DGSFQEYATADAVQAAHIPQGTDLAEVAPILCAGITVYKALKSANLRAGHWAAISGAAGGL
GSLAVQYAKAMGYRVLGIDGGPGKEELFTSLGGEVFIDFTKEKDIVSAVVKATNGGAHGIIN
VSVSEAAIEASTRYCRANGTVVLVGLPAGAKCSSDVFNHVVKSISIVGSYVGNRADTREALD
FFARGLVKSPIKVVGLSSLPEIYEKMEKGQIAGRYVVDTSK

Supplementary figure 7.3 Alcohol Dehydrogenase sequence. The sequence of monomeric isoform 2 ADH from yeast is formed by 348 amino acids. The initiator methionine is removed, which result in a mass with 36732 Da.