Studying biomolecules using photoactivation and mass spectrometry techniques

A thesis submitted to the University of Manchester for the degree of Doctor
of Philosophy in the Faculty of Science and Engineering

2021

Rachelle Black

Department of Chemistry

Table of Contents

Final Word Count: 58175

Table of Figures

List of Tables

Abbreviations

ADH – Alcohol dehydrogenase

Ado – 5’-deoxyadenosyl ligand

AdoCbl – 5’-deoxyadenosylcobalamin

AES – Atomic Emission Spectroscopy

aIMS – Activated ion mobility spectrometry

ATD – Arrival time distribution

AUT – Analytical Ultracentrifuge

BmCarH – CarH from Bacillus megaterium

CCS – Collision cross section

CCSD – Collision cross section distribution

CD – Circular dichromism

CE – Collision energy

CID – Collision induced dissociation

CSD – Charge state distribution

DMB – 5,6-dimethylbenzimidazole

DTIMS – Drift tube ion mobility spectrometry

ECD – Electron capture dissociation

ESI – Electrospray ionisation

ETD – Electron transfer dissociation

FEL – Free-electron laser

FLWT – Full-length wild type

FMN – Flavin mononucleotide

FPOP – Fast photochemical oxidation of proteins

FRET – Förster resonance energy transfer

G192Q – G192Q variant of TtCarH

HCD – Higher-energy collisional dissociation

IDP – Intrinsically disordered protein

IDR – Intrinsically disordered region

IMS – Ion mobility spectrometry

IMMS – Ion mobility mass spectrometry

IR – Infrared

IRMPD – Infrared multiphoton photodissociation

IVR – Intramolecular vibrational energy redistribution

LC – Liquid chromatography

LED – Light emitting diode

MALDI – Matrix-assisted laser desorption ionisation

MD – molecular dynamics

MeCbl – methylcobalamin

MS – Mass spectrometry

MVA – Multivariate analysis

MxCarH – CarH from Myxococcus xanthus

nESI – Nano-electrospray ionisation

NC – Naphthyl based 18C6 reagent

OPO – optical parametric oscillator

PCA – Principal component analysis

PCB – Printed circuit board

PD – Photodissociation

REMPI – Resonant Enhanced MultiPhoton Ionization

PCB – Printed circuit board

PISA – Photoisomerisation action spectroscopy

PLS-DA - Partial least squares-discriminant analysis

PTM – Post-translational modification

PTR – Proton-transfer reactions

RPSB – Retinal protonated Schiff base

RMSD – Root mean squared deviation

SASE – Self amplified spontaneous emission

SAXS – Small-angle X-ray scattering

SEC – size exclusion chromatography

SIC – Selected ion count

SID – Surface induced dissociation

TIC – Total ion count

ToF – Time-of-flight

TOFTOF – Tandem time of flight

TtCarH – Thermus thermophiles CarH

TtCarHCt – C-terminal domain of TtCarH

TWIMS – Travelling wave ion mobility spectrometry

$^{TW}CCS_{N2}$ – T-wave collision cross section measured in N2 drift gas

UV – Ultraviolet

UVPD – Ultraviolet photodissociation

UVR8 – UV RESISTANCE LOCUS8

WAXS – Wide-angle X-ray scattering

WT – wild-type

XFEL – X-ray free-electron laser

XUV – Extreme-ultraviolet

Abstract

The ideal experiment for the understanding of structural biology will yield information on both structure and function in a single experiment. Standard approaches for the investigation of biomolecules usually includes NMR and X-Ray Crystallography, these come with limitations including complex spectra from NMR and in many cases proteins will not crystallise where they contain intrinsically disordered regions. Other methods include UV/Vis and IR investigations which will give information on the secondary structure but are limited when it comes to the understanding of the quaternary structure. Ion mobility mass spectrometry (IMMS) has increasingly been utilised to investigate protein structure and overcome some of these issues. IMMS provides both stoichiometric and structural data that can be used to probe the stability of global fold and the non-covalent interactions within a given protein complex. The data produced is also far less complex than results from NMR making it an easy process to apply and can be applied to a range of biomolecules with minimal changes to the settings.

This thesis describes the use of IMMS to interrogate the functional fold of proteins *via* a modified Synapt G2-S that incorporates light in the form of high powered pulsed UV lasers as well as LED. The UV lasers have been implemented to allow photodissociation of proteins (Ultraviolet Photodissociation, UVPD) either pre- and post-ion mobility. These arrangements allow us to link the fragments to the protein fold and hence gain a better understanding of the perturbation of protein structure. On the same IMMS instrument, a rig was constructed to permit LED illumination of analytes within the electrospray (ESI) tip, and to then use the mass spectrometer to report on the results of such in solution photoactivation. This is applied to the investigation of photoactive proteins and a methodology has been developed to perform this work under minimal light activation.

UVPD-IMMS investigation of well-studied monomeric proteins produces distinctive fragments that can be related to the non-covalent interactions which determine the global fold. UVPD-IMMS experimental approach was extended to large multimeric proteins, where further information of the native fold could be determined. Alongside UVPD-IMMS, collision induced dissociation (CID) IMMS measurements were also performed on well-studied monomeric proteins, distinctive fragments could relate to the

precursor charge state. The use of multiple fragmentation strategies produces complex rich datasets and these were compared using PLS-DA to produce scores plots. From these comparisons were made between conformationally different proteins fragmented *via* UVPD and between different charge states fragmented *via* CID. We observed that the conformations and charge states yielded significantly different fragmentation patterns that could be separated *via* PLS-DA and give the option to get further information than the standard fragment assignments.

The use of LEDs to activate proteins is described in a protocol aimed at those with general knowledge of MS but unaware of the added benefit of applying this to photoreceptor proteins. We describe the procedure for the construction of the apparatus, show exemplar data and acquisition settings that have been developed on a range of different mass spectrometers. Exemplar data for photoreceptor proteins and a commercially available small molecule is also shown for the set-up. The photoreceptor proteins chosen cover a range of the UV-visible spectrum to show the flexibility of the LED set-up.

Finally, we have utilised the improved capabilities to work with photoreceptors and light-reactive proteins to investigate how the tetramer formation of WT TtCarH when in the presence of Adenosylcobalamin. This was performed under minimal photoactivation and required special care when handling. The step-wise potential mechanism was determined through a combination of MS, stop-flow and molecular dynamics simulations. The sensitivity of IMMS allowed us to observe partially-ligated versions of the dimer and tetramer, which suggests that pre-assembly of tetramer can occur without a full complement of bound Adenosylcobalamin. To further investigate the role of Adenosylcobalamin in the photoactivation mechanism we used methylcobalamin with WT TtCarH and adenosylcobalamin with the TtCarH variant G192Q.

The contents of this thesis go to show the benefit of incorporating multiple photoactivation techniques with native MS and IMMS, this has been shown on standard, well-studied proteins and more complex photoreceptor proteins. The photo-IMMS process has also been shown to work well in conjunction with computational data processing methods such as PLS-DA to gain information on structural dependence on fragmentation, and molecular dynamics to aid the determination of mechanisms involved within biological systems. The understanding of the protein fold and its effects on UVPD has been extended, an LED rig has been set up to aid in the determination of functional biological behaviour of photoreceptors and *via* IMMS measurements performed in the dark, distinct new information could be obtained on the tetramer formation of TtCarH when bound to cobalamin B_{12}.

Declaration

Preliminary data I obtained for the work described in Chapter 2 were included in the co-author's (AT) thesis, however the data presented here are a continuation of that work and, as all other chapters, have not been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses

Acknowledgements

Firstly, I would like to thank my supervisor for providing me the opportunity to be a part of such an exciting, knowledgeable group and allowed me to work in numerous different research areas, with a variety of collaborators and on a range of projects.

I would also like thank the Barran Research Group for being absolutely fantastic and encouraging through thick and thin! The photolab has been my little subgroup since the beginning, thank you to the photolab team Alina, Bruno, Lennart, Joe, and Ellen, through alignment woes and instrument migrations. To Emily and Emma, the Elsa and Olaf to my Anna. To Caitlin, giving me the light-hearted relief during those hard labs days and group coffee breaks. To my PhD cohort Dale and Jack, it doesn't really feel like it's been 4 years on this journey but here we are, thank you for being there every step (even though one of you ran to Singapore for 2 years). And finally to the newbies of the group, Emilia and Niklas, enjoy this PhD ride and make the most of every opportunity given!

To all my friends from University and home, you've all been there and see the other side of this PhD! Lisa, thanks for listening to me moan for 26 years! My safe chat girls, Flo and Amy, for the gossips and film/TV recommendations. Arpan and Luke, the trashy memes and gifs, "*Hey, look at us? Who woulda thought?".* To my longest friend, Sally-Ann, thank you for sticking around all this time! To my Dukeries girls, Giorgina, Mia, Kayleigh, Sam, Bethany, Laura and Aishling!!

Thank you to the Bodybarre girls soft-side crew Emily, Olivia, Katie, Anna, Meenie and Rachel. To the fabulous instructors, Penny, Kaz and Flo, you are all fabulous and amazing, thank you for giving me a sanctuary!

Finally, thanks my family, my mum and dad for always supporting me through these past 8 years. To my sister Stacey, for letting me drag her to all the films I want to watch and for keeping me sane throughout the past 4 years. To Henry and George, the best of dogs both driving me mad and making my day a little brighter. To my Great Grandma, Edna, who couldn't see me reach this final year but has left me with stories to tell for the rest of my life.

“There's nothing more exciting than science. You get all the fun of sitting still, being quiet, writing down numbers, paying attention. Science has it all.”

Seymour Skinner, The Simpsons

1 Introduction – Photoactivation and dissociation in biomolecular mass spectrometry

1.1 Declaration

This chapter has been reproduced from a published review book chapter that I wrote:

R. Black, L. A. I. Ramakers, and P. E. Barran, “Photoactivation and dissociation in biomolecular mass spectrometry” in *Advanced Fragmentation Methods in Biomolecular Mass Spectrometry: Probing Primary and Higher Order Structure with Electrons, Photons and Surfaces,* First Edition., F. Lermyte, Royal Society of Chemistry, 2021, pp. 235-280.

As the first author of the book chapter, I drafted and edited the manuscript and reproduced all Figures. All sections and layouts for this chapter were determined by myself, LAIR took the lead for the IR sections and myself for the UV and the visible sections.

Sections 1.9 and 1.10 have been added to link the book chapter to the thesis project.

Photoactivation and dissociation in biomolecular mass spectrometry

R. Black, P. E. Barran and L. A. I. Ramakers

Manchester Institute of Biotechnology, Department of Chemistry, University of Manchester, Manchester, M1 7DN

*corresponding email address: perdita.barran@manchester.ac.uk

1.2 Abstract/Web Summary

In this chapter, we highlight recent developments using laser irradiation to probe ions in mass spectrometers. We will primarily cover the structural characterisation of biomolecules. This chapter will introduce the principles behind laser irradiation and describe some of the common instrument configurations, with an emphasis on benchtop laser systems coupled to instruments in laboratories. We have ordered the narrative in this chapter according to what has been achieved with three different irradiation wavelengths namely UV, visible and infrared. Instrumentation developed to permit laser irradiation along with mass spectrometry is described in each section along with some examples of recent applications primarily focused on the analysis of peptides and proteins.

1.3 Introduction and recent history of laser coupling to mass spectrometry

The well-established and preeminent method used to fragment ions in a mass spectrometry (MS) experiment is collision-induced dissociation (CID).[1] This is a standard operating procedure in most commercial mass spectrometers with undoubted benefits in reproducibility especially in ‘omic applications.[2–4] In a classic CID experiment, a given molecular ion is accelerated through a region of the mass spectrometer filled with an inert gas. Fragmentation proceeds via collisions which results in the transfer of kinetic energy to break the weakest bonds in a given analyte. This preference for weaker bonds can be disadvantageous, for example in cleaving post-translational modifications (PTMs), and when applied to noncovalent protein complexes, the energy deposited prior to fragmentation substantially perturbs the tertiary structure. CID is common in bottom-up proteomics investigations where peptide ions cannot inform on the tertiary or quaternary structure of the protein but do provide valuable information to map the primary structure.

As well as CID several other activation techniques are available to the mass spectrometrist, from electron-based methods such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) to photon-based methods such as ultraviolet photodissociation (UVPD) and infrared multiphoton photodissociation (IRMPD). As for CID, the electron- and photon-mediated techniques have principally been employed for primary structure determination, but since the capture of an electron or a photon occurs on a faster timescale than vibrational motions, the process is nonergodic – although the departure of a cleaved portion may take much longer and may need to be facilitated by additional collisional activation.[5,6] These methods therefore present the opportunity to provide information on the noncovalent interactions in the biomolecule of interest as well as to retain information on more labile bonds.[7–9]

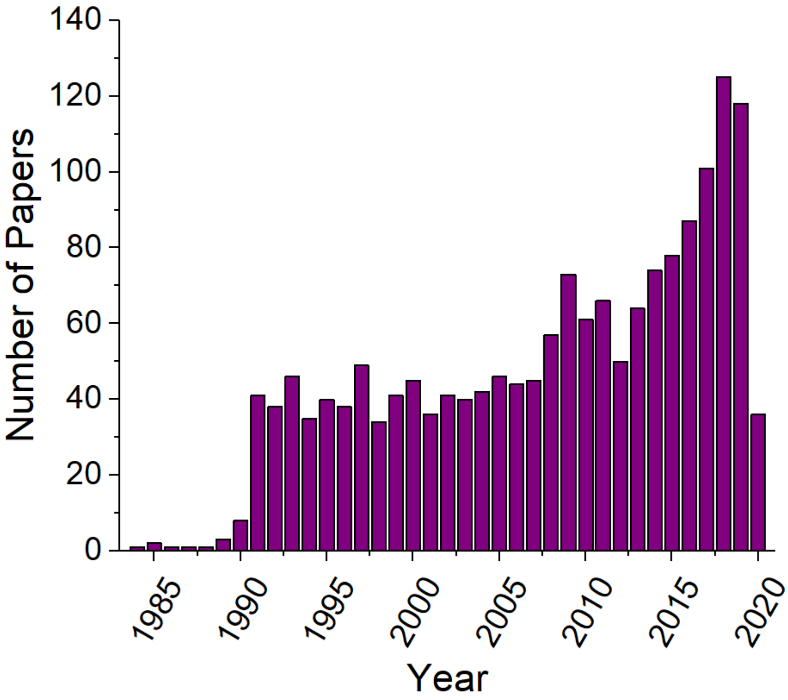


Figure 1.1: Results from a search on Web of Knowledge for the words photodissociation and mass spectrometry appearing in the topic field. The number of published papers per year is plotted. Data correct as of June 2020.

The first coupling of a laser to a mass spectrometer for fragmentation was reported over 40 years ago.[10] The advent of soft ionisation methods, and the subsequent developments in mass spectrometry based ‘omics research led to a significant increase into the research and characterisation of intact biomolecules. For many years the introduction of laser irradiation for spectroscopy or dissociation to MS was principally performed by academic scientists using homebuilt instrumentation. This involved laser sources at purpose-built beamlines as well as benchtop lasers, and in either case the mass spectrometers were either completely homemade or modified commercial instruments. In early studies, the majority of analytes were of low molecular weight (<2000 amu). A considerable subset of the work has been performed using free-electron laser (FEL) facilities, where multiple irradiation techniques are housed under one roof, for example FELIX in Utrecht and now Nijmegen; the Fitz-Haber Institute (FHI) in Berlin, and SOLEIL in Paris. With wide wavelength ranges, these lasers have many uses for publically accessible science, only some of which incorporate MS. It is this limitation in laser access that delayed the uptake in the utilisation of photodissociation methods coupled with mass spectrometry (see Figure 1.1). With the access to benchtop laser systems and these purpose-built institutes improving, the number of publications implementing these techniques has rapidly increased.

Much of the early work utilised action-IR spectroscopy and focused on determining the similarity between molecules in solution and their ionic counterparts in the gas phase. In these experiments, the ion of interest is irradiated with light from the infrared region of the electromagnetic spectrum, and photo-induced fragmentation occurs only when the frequency of the infrared radiation is resonant with a vibrational mode of the ion. Monitoring the fragmentation yield as a function of the IR wavelength allows the action-IR spectrum of the ion to be recorded, and this action spectrum is generally considered to be analogous to the gas phase IR absorption spectrum. From solution-based studies, the IR spectrum is known to be highly sensitive to the specifics of the underlying geometrical structure of the molecule, leading to action-IR spectroscopy being widely used to determine the structure of ionic species and the population of different conformers being analysed by mass spectrometry. Additionally, the ability to calculate the gas-phase IR spectra of molecules, utilising *ab initio* methods,[11,12] and to obtain complementary results from homebuilt two laser Resonant Enhanced MultiPhoton Ionization (REMPI) set-ups, primarily focussed on neutral species,[13–17] means action-IR spectroscopy is a popular method to directly compare the experimental ionic geometries with candidate structures. For proteins and peptides, regions associated with the C=O stretching mode (amide I) and the N-H bending modes (amide II) are known to be excellent probes of secondary structure. Action-IR spectroscopy on these modes has allowed changes in secondary structure due to the transition to the gas phase to be monitored, as well as the impact of the charge state on the native structure of these biomolecules to be probed. Such a structural study was carried out in early work by Von Helden and Oomens, where the action IR spectra of cytochrome c was compared across different charge states. To date, this protein is still widely studied by action-IR spectroscopy, leading to novel insights in recent work[18,19] and still represents one of the largest molecules to have been examined with IR-FEL experiments. Other complex non-biological molecules have been examined with such free-electron set ups, at FELIX Rijs *et al.* utilised this approach to perform detailed spectroscopic studies on isolated interlocked molecular assemblies.[20–22]

As well as these experiments in purpose-built laser facilities, there have been many pioneering studies using benchtop lasers, including early work performed on sector mass spectrometers to test the stability of inorganic complexes. A significant focus of this was the stability of d-block complexes including copper, silver and gold containing complexes, often leading to very rare stable complexes that would not usually be observed in the condensed phase.[23–25] It was also used to monitor the gas-phase coordination of d-block ions when oxygen and nitrogen-containing ligands were present.[26,27] As the affordability of benchtop laser systems has improved, more effort has gone in to coupling them to higher resolution mass spectrometers. Bruker first incorporated IRMPD into their FTICR instruments in the early 2000s and more recently, Thermo (Waltham, MA, USA) has started to market instruments which incorporate a UV laser for UVPD experiments following work performed on biological systems by Brodbelt, Barran, Boyarkin,[14,28–30] and others. This chapter will now focus on instrumentation and applications of mass spectrometers that incorporate benchtop laser systems.

1.4 Background to the principles of laser irradiation and the production of light from far UV to IR.

Laser sources utilise a lasing medium housed within an optical cavity containing at least one highly reflective mirror and one partially reflective mirror, sometimes referred to as the output coupler. Whilst the exact design of this optical cavity depends on the laser system, the basic operation of most benchtop laser systems follows the model first proposed by Einstein in 1916. Namely, the lasing medium is pumped (*e.g.* using a flash lamp or high-voltage electrodes) to generate a population inversion (a state in which the population of the first excited state is much greater than the ground state).[31] Such a population inversion is transient and will start to relax back to the ground state via spontaneous emission. The photons emitted by the spontaneous emission events travel through the medium triggering stimulated emission. These emitted photons then travel around the optical cavity before being incident onto the lasing medium again. During this round trip, the lasing medium is pumped again to regenerate the population inversion allowing the light to be amplified by stimulated emission on each pass of the optical cavity. At the end of each pass around the optical cavity a small percentage of the photons are emitted through the output coupler. This emitted portion of the photons forms the emitted laser beam. The first operational laser source was presented by Maiman[32] and Schawlow[33] in 1960. Since this first laser, many more have been developed, utilising a range of different lasing media (*e.g.* TiS crystals, Nd:YAG crystals, gas mixtures, semiconductors *etc.*) and different optical cavity designs, allowing them to be utilised in a range of applications.

In addition to these benchtop laser systems, the development of free-electron lasers (FEL), first reported in the 1970's,[34] provides an intense and tuneable narrowband source of laser light. FELs require a synchrotron to generate light and as such generate light utilising self-amplified spontaneous emission (SASE).[35] Briefly, relativistic electron pulses are injected into an undulator, consisting of a series of alternating magnets, from a synchrotron. The oscillation of these electron pulses results in the incoherent emission of electromagnetic radiation. The subsequent interactions between the electrons in the pulse and the emitted radiation leads to the electrons bunching, at the radiation wavelength, leading to coherent emission of the radiation in question and the production of a laser pulse.[35] Tuning the undulator parameters allows the output wavelength to be altered. The requirement of relativistic electron pulses and long undulators means that such laser sources require a dedicated institute.[35]

Since the late 1970s, lasers have been coupled to mass spectrometers leading to the development of a wide range of different techniques. These techniques include photo-fragmentation methods, ion spectroscopies as well as novel ionisation methodologies, including MALDI[36] and laser vaporisation.[37] Developments in photofragmentation and ion spectroscopies have led to mass spectrometers being coupled with a range of different laser sources including but not limited to excimer, solid state, Nd:YAG, dye, and CO_2 lasers.[10] These different laser sources fall into two classes based on whether they are fixed- (such as an excimer laser) or tuneable-wavelength sources (such as a CO_2 laser) allowing a wide range of action spectroscopies and photofragmentation processes to be studied. For UVPD, excimer (gas mixture tuned to produce 193 nm) or 213 nm solid state lasers systems are most commonly incorporated into mass spectrometers due to their ability to produce intense nanosecond deep-UV laser pulses at high repetition rates (100 Hz – 1 kHz). These two wavelengths are commonly used in UVPD studies due to their ability to fragment a wide range of biological samples. Additionally, optical parametric oscillator (OPO) laser systems, capable of producing intense nanosecond laser pulses over a wide range of wavelengths, have also been incorporated into mass spectrometers to perform both UV and IR action spectroscopies. These laser systems normally utilise a nanosecond pump laser (often a Nd:YAG laser) to pump a series of nonlinear crystals, housed in the OPO optical cavity. Tuning the resonance of the OPO cavity allows the output laser wavelength to be selected,[38] allowing photofragmentation of ions to be performed over a wide range of wavelengths, spanning most of the UV-visible region of the electromagnetic spectrum. The addition of a difference frequency generator then allows the output of such a laser system to be tuned into the mid-IR region of the spectrum.[39] Due to this flexibility, such OPO laser systems have become the most common benchtop laser system used to carry out UV and IR action spectroscopies. Though it is important to note that UV and IR irradiation result in different molecular response to illumination, UV/visible resulting in excitation of electrons with higher energy UV resulting in bond breakage whereas IR irradiation results in the excitation of bonds, with more IR absorption required to obtain any bond breakages. This will be covered in more detail throughout this chapter.[40]

1.5 Ultraviolet region

The ultraviolet electromagnetic spectrum comprises the wavelength range of 10 - 400 nm. From 156-400 nm there are several commercially available benchtop laser systems often with fixed wavelengths and the first reports of UV laser coupling to MS was almost 30 years ago.[41–43] The use of UVPD for sequencing peptides or proteins, analogous to ECD or CID, has been somewhat curtailed by the lack of UV absorption by the common bonds of the peptide backbone above 200 nm. In some cases, nature provides the chromophore for improved absorption; for example, the presence of aromatic amino acid groups, such as tryptophan, tyrosine, and phenylalanine in a peptide chain allows for the absorption of 266 nm UV radiation.[44,45] 266 nm also has the added benefit of being the 4th harmonic on a Nd:YAG laser, resulting in a relatively straightforward MS coupling and gives some flexibility to other harmonic wavelengths of that are obtainable.

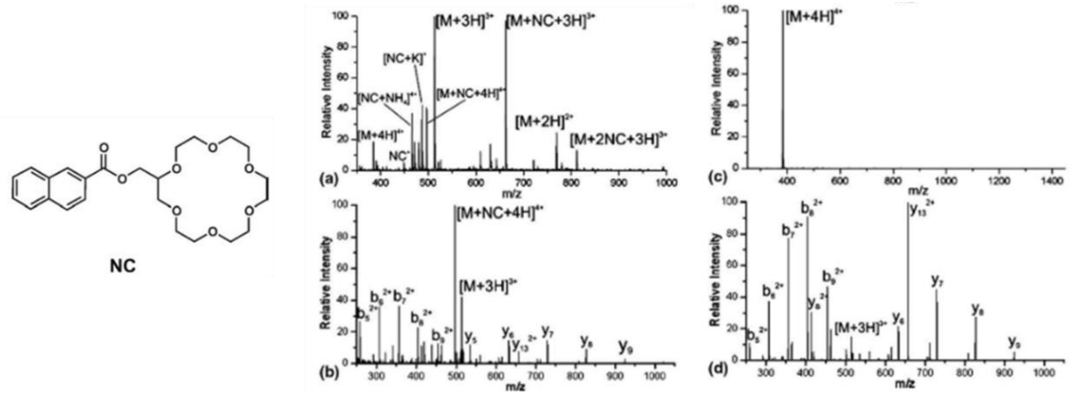


Figure 1.2: Coordination of crown ethers to act as chromophores allows for activation under wavelengths not traditionally absorbed. (a)ESI mass spectrum for naphthyl-based 18-crown-66 reagent (NC) and a peptide, KP13K. (b) 266 nm photodissciation spectrum for $[M+NC+4H]^{4+}$*. (c) 266 nm photoactivation of the peptide, KP13K. (d) CID spectrum for* $[M+3H]^{3+}$*. Adapted from ref. 45 with permission from American Chemical Society, Copyright 2011.* [46]

One method for getting around this limitation is through the addition of a chromophore to the biomolecule that will permit the absorption >200 nm that may be easily attainable. Brodbelt *et al.* investigated the use of crown ethers as noncovalent chromophores with a strong absorption around 355 nm that might coordinate to a peptide and promote UVPD. By coordinating 18-crown-6-dipyrrolylquinoxaline to a non-chromogenic peptide sequence, 355 nm and 266 nm can be absorbed and produce significantly more peptide sequence fragments than with either CID and IRMPD. [47] Julian *et al.* have implemented a similar coordination procedure utilising a naphthyl-based 18-crown-6 reagent (NC), (see Figure 1.2). This yielded similar results to those observed by Brodbelt, with a single laser pulse from a 266 nm Nd:YAG laser being sufficient to induce fragmentation in peptides where no fragments were originally observed. Although there were still some limitations in the lack of complete sequence coverage, this implementation of chromophores provides an opportunity for fragmentation in a collisionless environment and with further development could lead to improved sequence coverage.[46]

Utilising UV lasers that can access wavelengths <200 nm has become more commonplace in recent years and this alleviates the requirement of a universal chromophore for any peptide or protein. Primarily this has been performed with the use of excimer lasers, for example ArF (193 nm) on several different platforms including a linear ion trap[48] and an orbitrap.[49] Reilly demonstrated the use of 157 nm excimer (F_2*) comparing the production of fragments from peptides to those produced by CID using a matrix-assisted laser desorption ionisation (MALDI) source coupled to a tandem time-of-flight (TOFTOF)[50,51] as well as with electrospray ionisation (ESI) coupled to a tandem linear ion trap.[52] 157 nm has also been applied to oligosaccharides produced *via* MALDI on a time-of-flight mass spectrometer[53]. The benefit of this deep-UV wavelength is that it is able to cleave bonds with a single photon. However, 193 nm excimer lasers are somewhat more user-friendly with respect to alignment optics and the ability to transmit the laser beam through air.[54]

1.5.1 Understanding the mechanism of photoactivation

UV photons span energies from 3 to 1000 eV which can lead to single-photon bond cleavages and a richer fragmentation yield without the time for energy redistribution as found in CID or the multiphoton approaches utilised for IRMPD (see Section 1.7).[46]

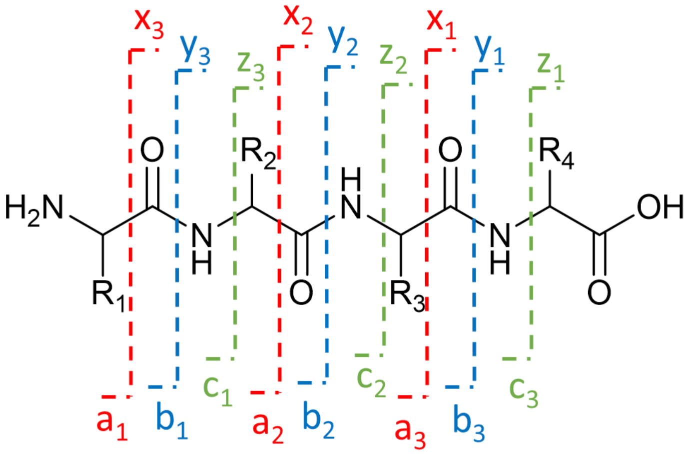


Figure 1.3: Nomenclature for peptide fragmentation via mass spectrometry activation. Reproduced from ref. 54 with permission from the Royal Society of Chemistry. [55]

The mobile proton mechanism, pioneered by Gaskell and Wysocki, underpins much of our understanding of CID fragmentation. This proceeds by two potential routes, the first the charge-directed one where a labile proton on the backbone of the protein directs cleavage and gives rise to the formation of *b*- and *y*- sequence ions (see Figure 1.3). The second route is the charge-remote route whereby the proton is sequestered by a basic site and will yield fragments that do not all correspond to *b*- and *y*- sequence ions. [56–58]

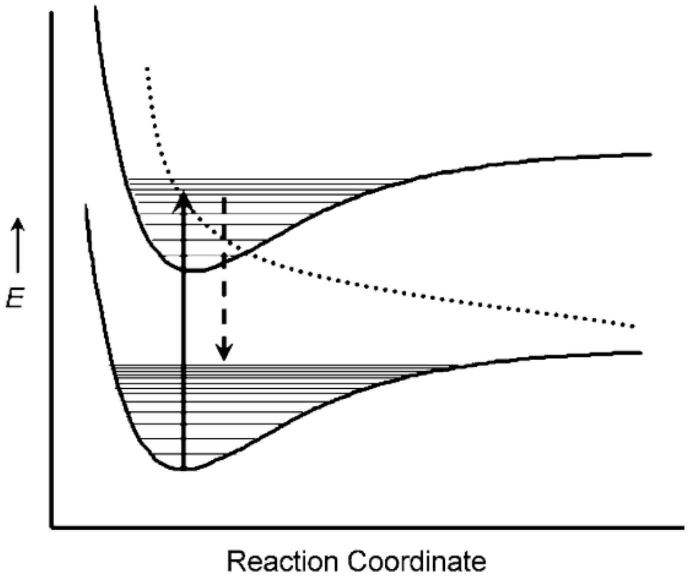


Figure 1.4: A simplified diagram representing the process for direct dissociation, reaction coordinate refers to the bonding state, where the lowest point is the ideal bond length and at the energy minimum. Absorption (solid arrow) of photon energy promotes an electronic transition into an excited electron state; this can then couple to an antibonding state (dotted line) where the electron density in the orbital is concentrated outside the bonding region weakening the chemical bond and leading to direct dissociation, shown here by the decreasing energy in the increasing bond length and therefore dissociation. Alternatively, the excited state can redistribute the energy into vibrational modes and result in fragmentation resembling CID. Reproduced from ref. 43 with permission from John Wiley and Sons, Copyright 2009. [44]

For UVPD, two potential paths have been distinguished, the first of these is direct dissociation,[28,59] where the photon energy is higher than the bond strength of the peptide.[54] Direct dissociation occurs at the femtosecond timescale, the photon induces electronic motion to a higher energy level and then relaxation into a dissociative orbital to produce sequence ions without any prior energy redistribution. This process is shown in Figure 1.4and for peptides, this will cleave the C_α-carbonyl carbon bond which will yield distinctive *a*- and *x*-type fragments (see Figure 1.3) [44] Direct dissociation of this bond, means that there is very little energy remaining to redistribute and break noncovalent bonds that will be holding the fragments in place. In order for these sequence ions to be detected, there must therefore be a second process that works alongside direct dissociation to target the noncovalent interactions and allow the fragments to dissociate away.[60]

The second dissociation method involves the conversion of the energy absorbed through to the vibrational modes in an approach called intramolecular vibrational energy redistribution (IVR).[61] This redistribution allows for the molecule to become "hot" and randomise the energy to induce dissociation obtaining fragments that are similar to both CID and IRMPD *via* the mobile proton mechanism.[56–58] This method produces *b-/y*-type ions as in CID and IRMPD and also targets the noncovalent interactions that form the secondary and tertiary structure of the ion of interest.

The generation of ions similar to those produce by CID is the primary proof for this second UVPD mechanism; however, there are some other pieces of information to support this.[60] CID can disrupt noncovalent interactions, which from a multimeric complex will give rise to dimers and monomers where the charges from the parent ion partition asymmetrically into the ensuing fragments. By contrast, UVPD has been shown to produce fragments where the charge is symmetrically distributed.[62] Another unique feature of high energy UVPD is the production of unique fragment ions, *b+2*, *a+2* and *y-2* that occur when the protein probed contains proline residues. The *y-2* ion is yielded when the homolytic cleavage of the C-N peptide bond close to a proline residue occurs and a simultaneous rearrangement and elimination of CO and two H radicals occurs. Whereas the formation of *b+2* and *a+2* requires a C-C or C_α-C bond breaking, respectively, and a proton transfer procedure that in the case of *a+2* will again result in the release of CO.[63]

The resulting spectra yielded *via* UVPD implies that neither direct dissociation or internal conversion are the sole mechanism for UVPD and instead a combination of the mechanism must occur. When photons are absorbed by the protein, cleavage of bonds *via* direct dissociation results in fragments that are only able to depart the parent ion if the internal conversion mechanism can cause disruption of the noncovalent interactions. This duality is a unique feature of UVPD and allows for the observation of fragment ions, related to the unactivated gas phase structure and not readily observed in CID alone.[60] Interestingly, the extent of direct dissociation has been reported to relate directly to the wavelength of the UV laser utilised, where a shorter wavelength (*i.e.* higher-energy photons) is more likely to result in direct dissociation. For example, at 266 nm direct dissociation is less feasible than at 213 nm but the extent of direct dissociation drastically increases when going from 213 nm to 193 nm and even more so at 157 nm UVPD.[51,54] Reilly has previously proposed that 157 nm UVPD occurs primarily *via* photolytic radical cleavage (Norrish Type I) for the peptide backbone at C_α and carbonyl carbon bonds to produce *a* and *x* ions. After this initial cleavage, radical migration and other secondary processes occur to produce the varied assortment of fragment ions, primarily those which are observed through direct dissociation.[51]

1.5.2 Instrument developments for UVPD

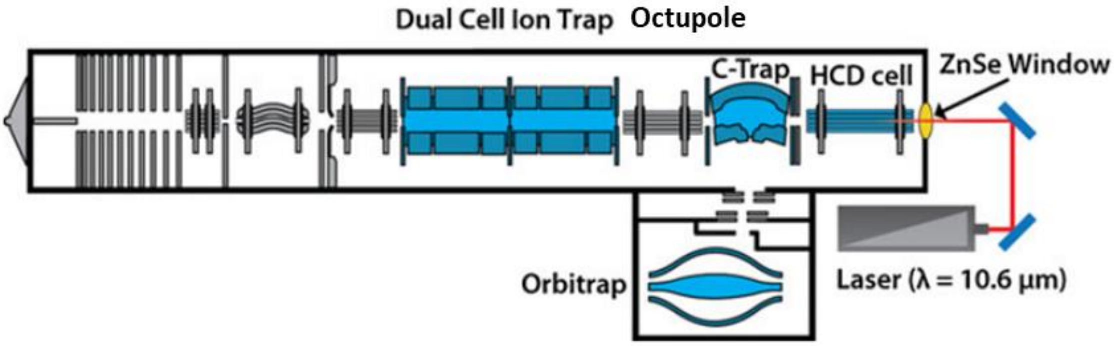


Figure 1.5: A modified Orbitrap Elite to incorporate IRMPD and 193 nm UVPD into the HCD cell. Reproduced from ref. 63 with permission from American Chemical Society, Copyright 2011. [64]

As previously mentioned, UVPD has been routinely implemented into linear ion trap mass spectrometers and Orbitrap instruments. The Orbitrap Fusion Lumos is a commercial instrument that is equipped with a 213 nm laser for UVPD. The Brodbelt group previously modified an Orbitrap Elite via the higher-energy collisional dissociation (HCD) cell of a dual pressure ion to incorporate IRMPD, and subsequently 193 nm UVPD (see Figure 1.5).[64,65] In a standard Orbitrap arrangement, the HCD cell is used for beam-type collision-induced dissociation (CID) whereby a higher RF voltage retains the ions in the C-trap before introducing the ions into the mass analyser.[66] In modified instruments, the HCD cell has been altered to allow the irradiation of the ions with the laser beam. Photodissociation product ions are subsequently directed into the C-trap and then into the mass analyser. The laser beam enters the HCD cell trapping region through a ZnSe window and enlarged aperture in the HCD endplate electrode. A nitrogen line was fitted into the C-trap and the top and bottom of the C-trap were fitted with ceramic plates to decrease gas conductance out of the C-trap. This modification allowed for reduced the gas flow needed to achieve desired trapping efficiencies in the C-trap whilst minimising the N_2 partial pressure in the quadrupole linear ion trap. Further modifications were made to facilitate photodissociation, of the three approaches taken, a slight increase gas flow to the C-trap and no flow in the HCD cell led to acceptable precursor-to-product conversion efficiency and the best product ion sensitivity.[64]

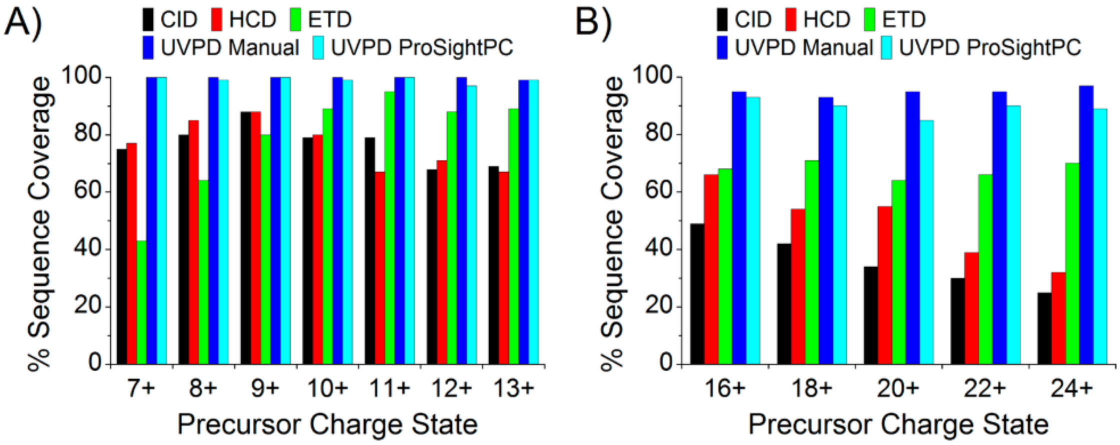


Figure 1.6: Comparison of CID, HCD, ETD, and UVPD monitoring sequence coverage as a function of precursor ion charge state for (A) ubiquitin and (B) myoglobin. Adapted from ref. 64 with permission from American Chemical Society, Copyright 2013. [65]

A comparison to collision-induced and electron-mediated dissociation shows that a single 5 ns laser pulse at 193 nm was able to yield a mass spectrum with higher fragmentation content for the proteins ubiquitin and myoglobin, (see Figure 1.6).[65] Further developments include the ability to perform UVPD and electron capture dissociation (ECD) on the same instrument by Shaw *et al*.[67] The octupole located after the quadrupole and before the C-trap is replaced with an electromagnetostatic ECD cell (e-MSion, Corvallis, OR, USA). This dual fragmentation strategy has been applied to the large multimeric protein, alcohol dehydrogenase (ADH) to reveal structural changes which occur with increasing ion source activation energy by monitoring the intensity of fragments produced, their position in the sequence and the charge they yield, (see Figure 1.7).[67]

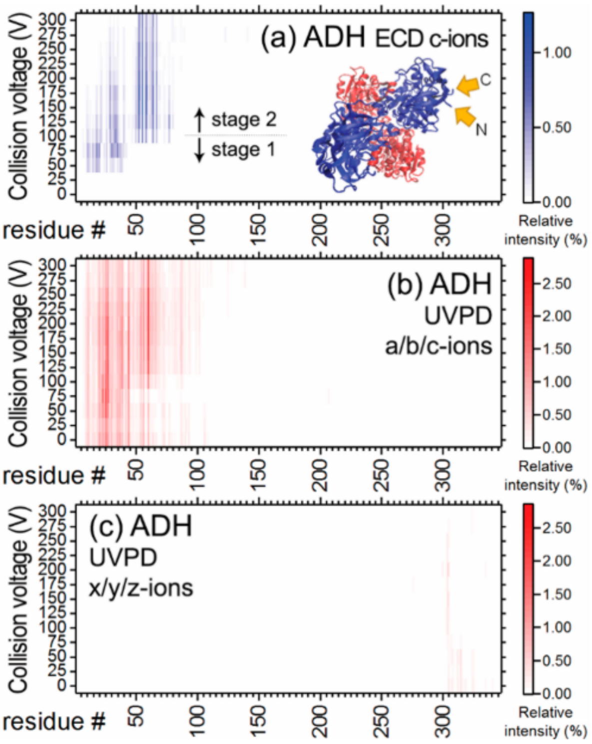


Figure 1.7: Monitoring fragmentation with increasing source activation. (a) ECD c-ion intensity map. Inset shows the PDB structure of ADH 4mer (Protein Data Bank ID: 4w6z), (b) UVPD N-terminal ion intensity map, and (c) UVPD C-terminal ion intensity map. This clearly shows two stages of structural change present due to the quick change in the fragmentation pattern. Reproduced from ref. 66 with permission from American Chemical Society, Copyright 2019. [67]

Another mass spectrometry platform modified to incorporate photoactivation is the Waters (Milford, MA, USA) Synapt G2-S,[68,69] which adds the separation capabilities of ion mobility (IMMS) to the analytical workflow. Ions are pulsed into a drift tube filled with inert gas, such as nitrogen, in the presence of static and/or travelling wave electric fields. The system measures the time it takes an ion to traverse the drift tube along with its *m/z*. As a given ion travels through the gas it will experience collisions that impede its progress, and the larger an ion is, the more collisions it will experience. The arrival time of an ion after passing through a drift tube in a given gas can be converted to a rotationally averaged collision cross section which in turn can help to describe its three dimensional structure.[70,71] Synapt instruments incorporate travelling wave ion mobility

(TWIMS) wherein ions are “carried” through the drift tube and separated by a traveling DC wave of a given magnitude and velocity.[72,73]

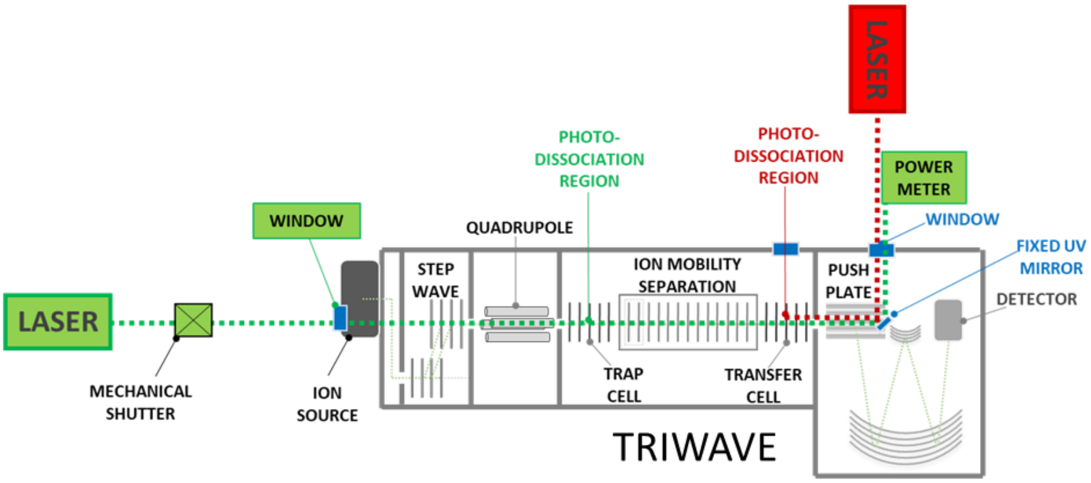


Figure 1.8: A modified Synapt G2-S incorporating UVPD and IMMS. This has been modified in-house with flexibility in operation modes. The laser beam can be introduced either via the TOF facilitating irradiation post drift cell, or via the ion source, which is possible with Step-wave ion optics. The latter mode reduces the loss of laser power and this configuration permits several workflows which combine photoactivation with ion mobility separation and mass detection.

The Synapt G2-S was modified to incorporate a laser beam whilst maintaining the IMMS capabilities (see Figure 1.8), allowing additional operation modes beyond those of photoactivation mass spectrometers without IM. This has been demonstrated with pulsed lasers of several wavelengths and frequencies. An early inception of this instrument positioned the laser such that it entered through a CaF_2 window above the orthogonal time-of-flight (TOF) analyser and reflected along the ion axis via a mirror in a machined slot at 45 ° to the ion pusher plate. External optics (not shown) allow the laser beam to be aligned along the central axis of the transfer cell, the TWIMS cell and the trap cell. Using a customised control script, WREnS (Waters research enabled software), the laser pulse can be gated with an optical shutter and triggered to interact with ions at the different ion storage cells. A typical trapping protocol consists of four steps (see Figure 1.9): (1) Standard beam checking (to monitor the ion signal), (2) Gating and filling (ions are selectively trapped after the IM drift time), (3) Trapping and activation (trapping the ions, where the laser pulse interacts with the trapped ion cloud), (4) Extraction (empty the trap and transfer ions to the TOF mass analyser).[69]

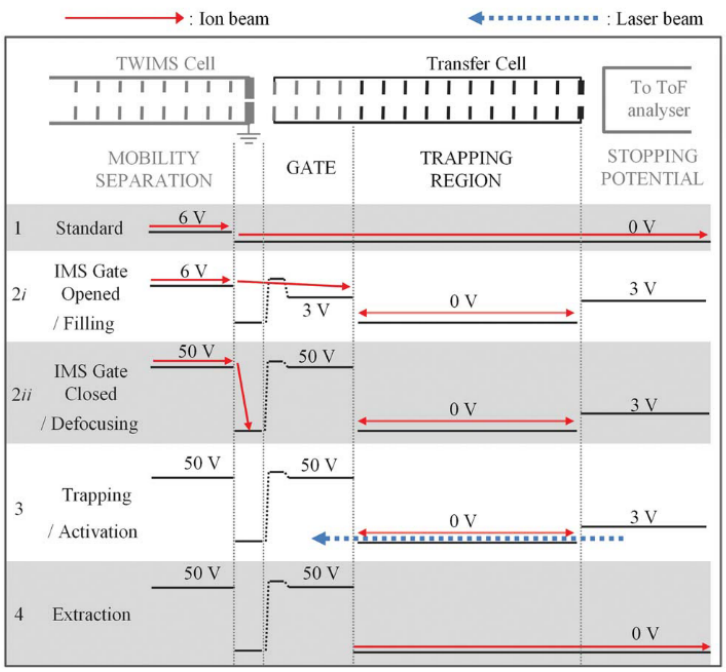


Figure 1.9: The trapping script for the operation of the UVPD-enabled Synapt G2-S. This allows ions to accumulate in the relevant region where they are trapped for a selected time period whilst being irradiated by the laser. Product ions are then released to progress through to the TOF mass analyser. Reproduced from ref. 68 with permission from the Royal Society of Chemistry. [69]

Irradiating within the transfer cell allows isolation of an ion *via m/z* (quadrupole) and/or following a specific mobility separation. UVPD on this instrument was first demonstrated with small molecules, peptides and the 20 amino acid mini-protein Trp cage. UVPD with a 266 nm laser showed that different conformations for the 5+ charge state of the peptide melittin could be selected after mobility separation and each yielded distinct fragmentation patterns.[68]

A later iteration of the Synapt platform employed a modified ion source block, allowing the laser beam to enter the instrument at the ion source end, (see Figure 1.8). In this scenario, the laser enters through a CaF_2 window directly into the upper level of the step-wave ion optics. This permits the laser to travel along the ion beam axis through the quadrupole and through the TriWave assembly made up of the trap cell, ion mobility cell and the transfer cell. This arrangement retains more laser power and permits ion irradiation in the source region, the TWIMS as well as the trap and transfer cells (before and after the drift cell).

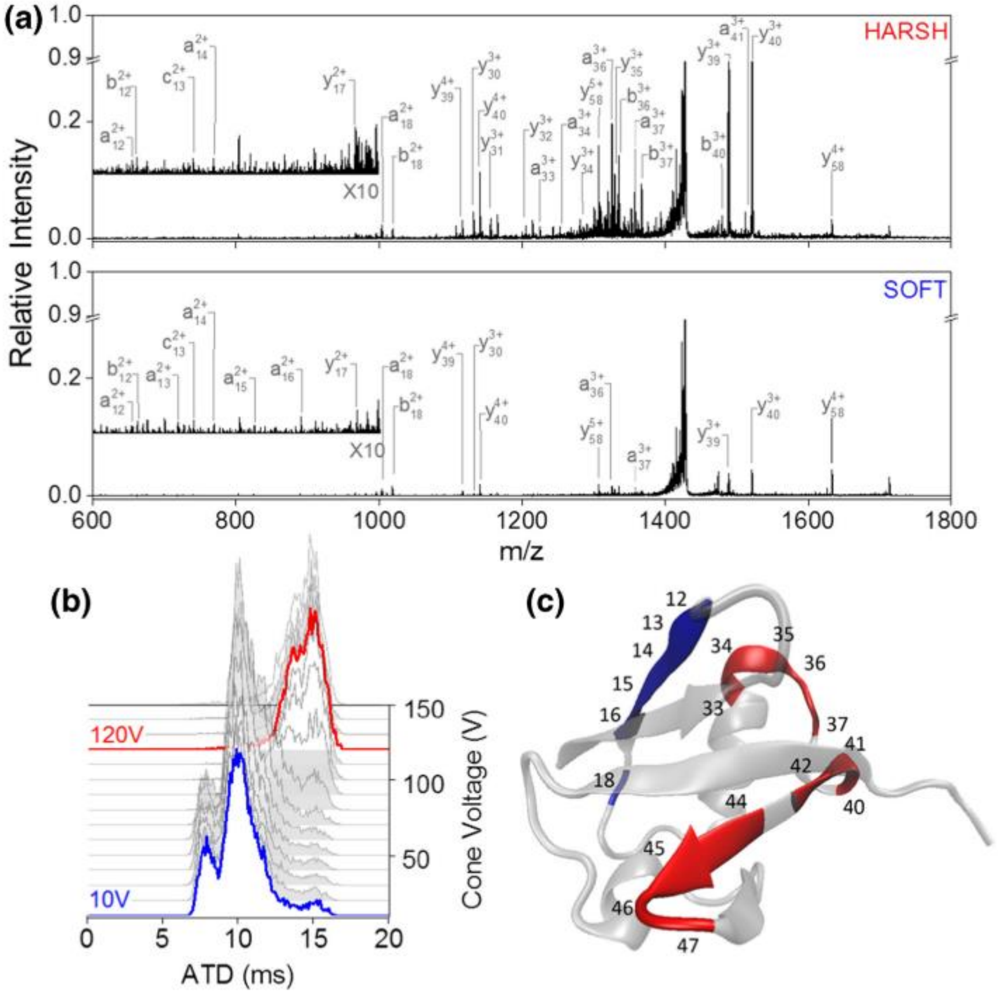


Figure 1.10: 213 nm UVPD of two different conformational arrangements of $[M+6H]^{6+}$ ubiquitin. (a) UVPD fragmentation spectrum obtained at “soft” ionisation conditions (cone voltage 10 V, blue) and “harsh” ionisation conditions (cone voltage 120 V, red). (b) Arrival time distributions for a range of cone voltages ranging from 10 V to 120 V. (c) UVPD cleavage sites overlaid onto the crystal structure (PDB ID 4Z9S) of ubiquitin. Sites that showed a higher yield under the harsh conditions are represented in red and the sites for which the obtained yield was higher under the soft conditions are represented in blue. Reproduced from ref. 73 with permission from American Chemical Society, Copyright 2018.[74]

Much of the work on this instrument has involved native MS approaches. This approach is shown nicely by activating the ions produced from the proteins ubiquitin, cytochrome c, and myoglobin in source (see Figure 1.10). Ion mobility reveals substantial conformational rearrangement; performing UVPD on *m/z*-selected activated ions can then uncover unfolding pathways *via* changes in their fragmentation yield and preferential position of fragmentation.[74]

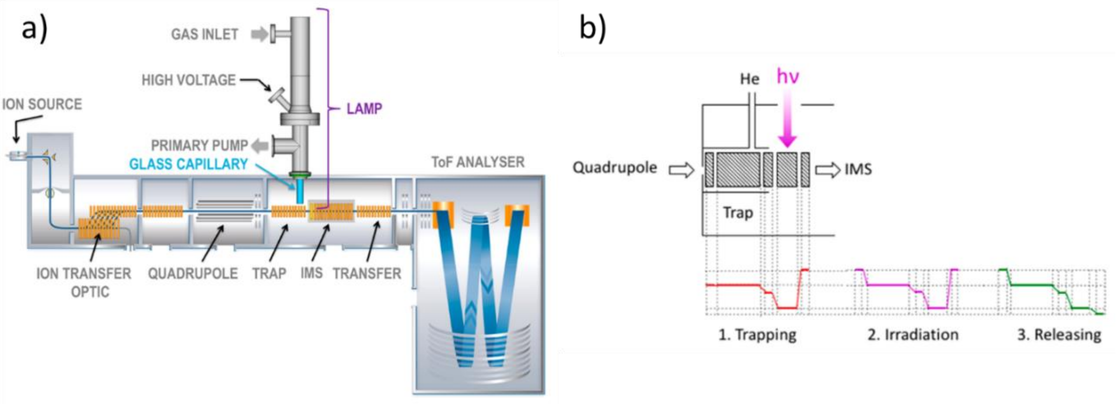


Figure 1.11: Implementation of XUV into a Synapt G2-Si, a) schematic of the Synapt G2-Si showing the ion path and the arrangement of the XUV lamp in the trap region b) Trap T-wave region showing the XUV irradiation region, different potentials are applied for trapping (red), irradiation (purple) and releasing (green) of the ions. Reproduced from ref. 74 with permission from American Chemical Society, Copyright 2018. [75]

Other recent developments in UVPD-MS include the implementation of extreme-ultraviolet (XUV) lamps and the inclusion of light-emitting diodes (LEDs). The first reporting of the inclusion of XUV lamps was in 2018 by Green *et al.* The lamp is implemented into a Synapt G2-Si into the trap cell prior to mobility and irradiates an area equating to ~10 mm with higher photon-energy XUV 40-80 nm UVPD (See Figure 1.11). This setup showed significant fragmentation yield in comparison to standard CID on small singly charged ions and the small doubly charge biomolecule, Glu-fibrinopeptide B. This photon source is an inexpensive alternative to lasers and requires less safety concerns.[75]

LEDs are another alternative to lasers and are also available at a fraction of the price and can be found in different wavelengths covering many regions of the electromagnetic spectrum. Specifically, UV LEDs tend to come in the wavelength range of 250 nm to 385 nm with a spectral purity of ~5-10 nm. Unlike fixed wavelength laser sources with highly collimated beams, for LEDs the light is not collimated and needs to be corrected with a focusing lens. This limits the use of LEDs to fixed wavelength irradiation as a set up for a scanning wavelength approach with multiple LEDs would require differing focal lengths which if utilising the same lens is not optimum for irradiation. A method round this is the use of mounted LEDs that already contain the ball lens and would not require a universal lens to collimate the beam.

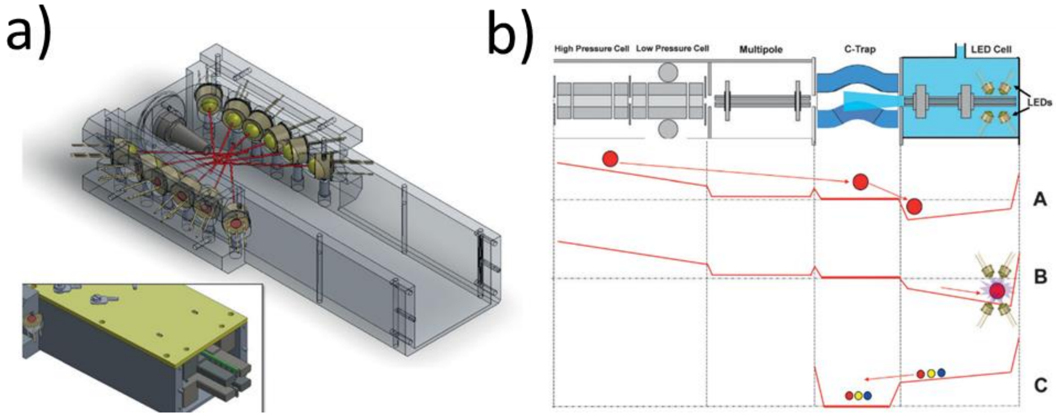


Figure 1.12: LEDs implemented into the higher-energy collisional dissociation (HCD) cell of an Orbitrap instrument. a) LED-HCD cell with housing and holding rods removed, fixed alignment of LEDs represented by red lines. Inset shows the cell arrangement with the cover and trapping rods installed. b) Applied potentials during experimental steps, ion being transferred to the front of the cell (A), ions transferred to the back of the cell and irradiated (B) and ions being transferred to the C-trap (C). Reproduced from ref. 75 with permission from John Wiley and Sons, Copyright 2016.[76]

UV LEDs were first integrated into a mass spectrometer by the Brodbelt group in 2016 (see Figure 1.12), here they couple “ball focusing” lens LEDs with wavelengths ranging 255-275 nm. The HCD cell of the orbitrap was replaced and redesigned to incorporate up to 12 LEDs, these were orientated to the back of the HCD cell. The ion cloud can then be pushed to the back of the instrument using printed circuit boards (PCBs), this allows for some tuning to effectively compress the ion cloud and increase the LED beam coverage. The arrangement has produced good fragmentation coverage on the highly conjugated flavin mononucleotide (FMN), which has a maximum absorption at

268 nm making it the ideal candidate for this LED arrangement, yielding comparable data to utilising a laser system.[68,76]

1.5.3 Further developments in the use of UV light sources with mass spectrometry

Native MS experiments performed on macromolecular (and commonly protein) complexes aim to minimise the disruption of noncovalent interactions to provide useful structural information. The nonergodic nature of UVPD means it is well set up to provide a reproducible readout for the structure of these ions wherein fragmentation patterns will relate to the gas phase conformation.[74] An extension of this would be to apply UVPD to probe multicomponent mixtures, for example proteins bound to a panel of ligands or to DNA. A drawback to doing this with direct infusion native approaches is a highly congested fragment mass spectrum. Proton-transfer reactions (PTR) coupled with UVPD have been developed to alleviate some of this congestion *via* charge state reduction whereby overlapping ions are moved to lower charge states at higher *m/z* and congestion is alleviated.[77,78,79]

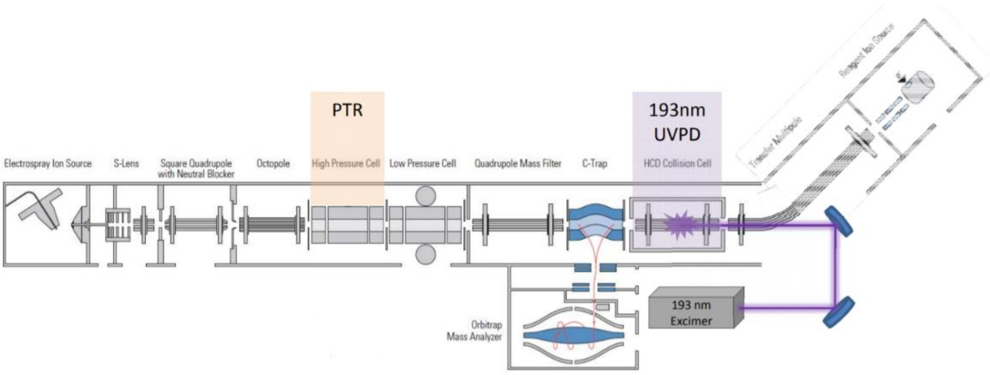


Figure 1.13: Modified Orbitrap Elite mass spectrometer to enable 193 nm ultraviolet photodissociation and PTR within the high-pressure ion trap. Reproduced from ref. 78 with permission from American Chemical Society, Copyright 2016.[79]

An instrument that incorporates both PTR and UVPD has been constructed by Brodbelt *et al* on an Orbitrap mass spectrometer (see Figure 1.13). This setup has been used to show that PTR can be implemented on a precursor ion, allowing for direct UVPD of species that would have had the same *m/z* (without PTR). The origin of fragments is elucidated by performing UVPD with and without PTR.[80] Such an approach is also applicable for product ions, and precursor ions located in heavily congested regions can be re-isolated using an *m/z* window and exposed to a PTR reagent. This allows identification of additional fragments from larger intact proteins that were previously poorly characterised through standard UVPD techniques.[81]

Although most UV-MS work utilises the absorption by the peptide backbone to induce fragmentation, sometimes biomolecules have a unique response to sunlight in order to fulfil their role, and these are often termed optically active biomolecules. This is the case for the photoreceptor protein UV RESISTANCE LOCUS8 (UVR8), which when kept in the dark adopts a homodimeric structure, however UV-B light (280-315 nm) is absorbed by clusters of tryptophan residues at the dimer interface and induces dissociation to form monomers.[82–84]

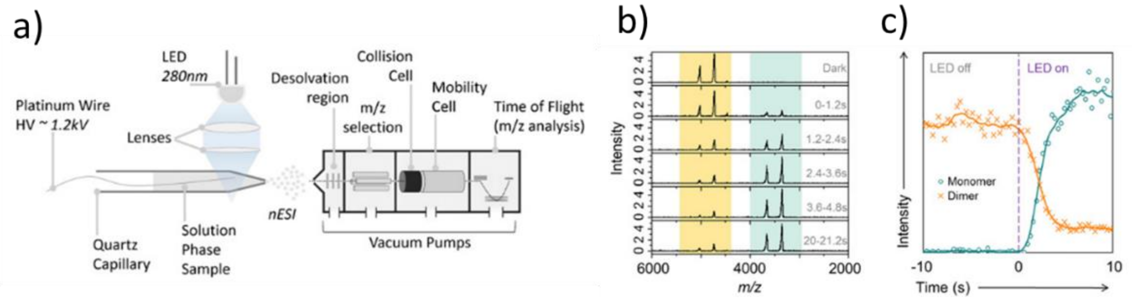


Figure 1.14: UV LED implementation for the investigation of protein UVR8 (a) the instrumental set up for the illumination of the solution phase prior to ionisation in the nano-electrospray tip. (b) the mass spectra of UVR8$^{12\text{-}381}$ as a function of illumination time using high-power LED. Dimer signal, pale orange; monomer signals, pale green. (c) Ion chromatogram extracted for the UVR8$^{12\text{–}381}$ dimer (orange) and monomer (green) as a function of data acquisition time. Adapted from ref. 84, https://doi.org/10.1073/pnas.1813254116, under the terms of the CC BY 4.0 licence, https://creativecommons.org/licenses/by/4.0/. [85]

To investigate conformational changes, the ion source of a Synapt G2-Si ion mobility mass spectrometer has been modified such that the light from an LED (280 nm) can be focused on the nano-electrospray (nESI) capillary tip (see Figure 1.14a). When the solution containing the UVR8 protein is illuminated any ensuing change in the structure or stoichiometry or both can be observed *via* IMMS. The MS signal readily confirmed the cleavage of the dimeric interface upon illumination accompanied by substantial conformational change attributed to extension in flexible intrinsically disordered regions of the protein sequence (see Figure 1.14b & c).[85]

Finally, mass spectrometry has also been coupled with a UV laser to determine the chirality of DNA quadruplexes in the gas phase, leading to the development of mass-resolved electronic circular dichroism (CD) ion spectroscopy.[86] UV CD spectroscopy is routinely used in solution-phase studies to probe secondary structures within proteins and peptides, as well as the chirality of DNA duplexes and quadruplexes.[87–89] This versatile and established spectroscopic method is widely applied in experimental studies,[90–93] but like many solution-phase techniques is unable to unambiguously assign features to single conformations in the type of heterogeneous samples often encountered in biology. By integrating a circularly polarised laser source into a trapping mass spectrometer, Daly *et al*.[86] have shown that is it possible to record mass-selected CD spectra of DNA complexes. As with UVPD, this method relies on the detection of photoproducts formed upon the irradiation of the ions with UV light. In this action spectroscopy experiment the UV laser pulse is passed through a series of optical components so that the light becomes circularly polarised, specifically a Rochon prism and an achromatic broadband quarter wave plate. The angle between the Rochon prism and the fast axis of the quarter wave plate (Fig X 15.(a))[94] determines whether the polarisation vector of the light rotates in the clockwise or anticlockwise direction. The difference between the photo-fragmentation yield upon irradiation with these two different circularly polarised laser pulses can then be used to obtain an action-CD spectrum for the gaseous ions under investigation. A schematic of the optical setup utilised by Daly *et al.*,[86] is shown in Figure 1.15.

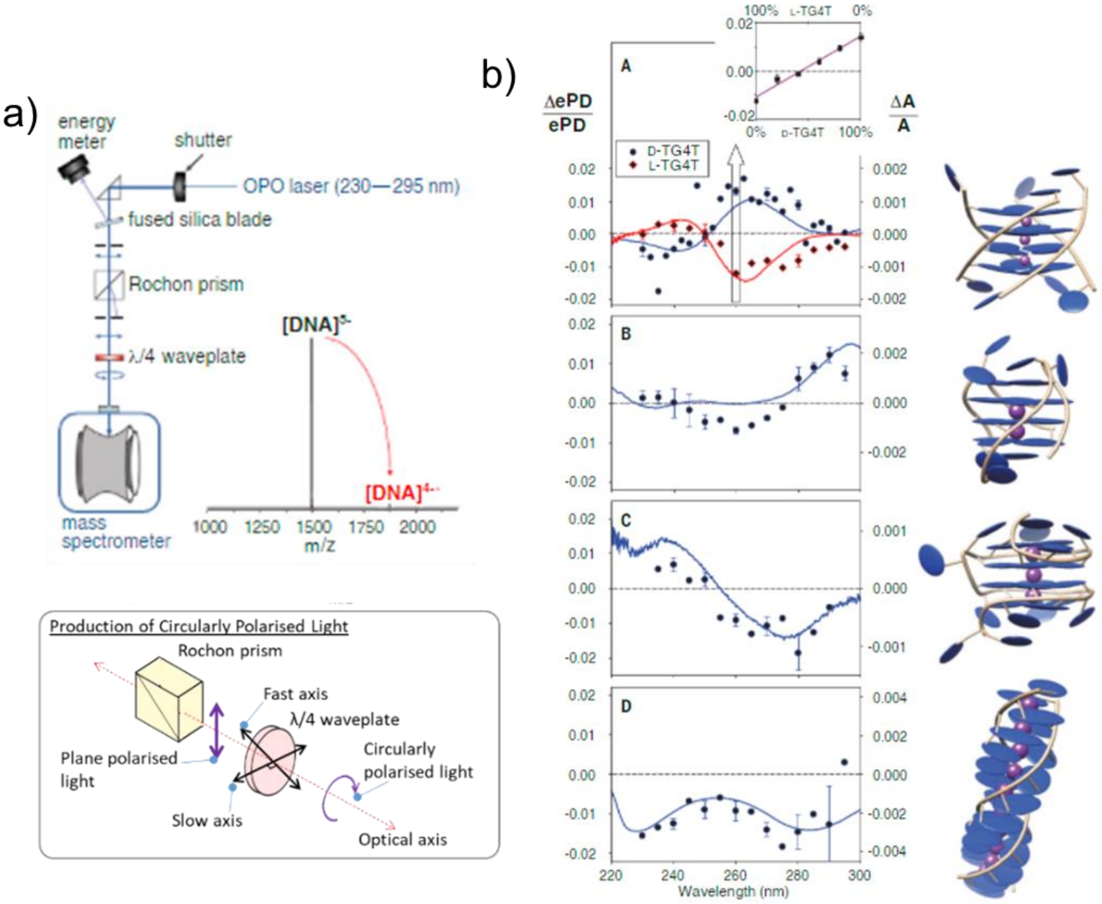


Figure 1.15: a) A schematic of the optical path employed in a novel CD-MS instrument, with a diagram showing how the circularly polarised light is produced in the optical path and b) the action-CD spectra (ΔePD/ePD) obtained for a range of different DNA duplexes and quadruplexes with comparison to their solution phase CD spectra (ΔA/A). Adapted from ref. 85 with permission from The American Association for the Advancement of Science, Copyright 2020. [86]

DNA duplexes and quadruplexes were chosen as the CD spectroscopy of these molecules is well understood and the dominant photo-product of negatively changed DNA ions is known to be photo-electron detachment, across most of the UV-visible wavelengths utilised in this spectroscopic method. The gas phase CD spectra are directly compared to their solution phase counterparts and overall were found to closely resemble each other (Figure 1.15b), demonstrating that this new form of UV action spectroscopy is highly sensitive to the chiral structure of the DNA ions and indicating that this property of the DNA complexes is conserved upon the transition of these molecules into the gas phase. This is further supported by the inversion of the spectrum when the natural D-sugars in the DNA strands are replaced by their non-natural counterparts (blue and red lines in Figure 1.15b panel A, respectively). This study shows that the photo-fragmentation of these molecular structures can be directly related to the chirality of the molecule, allowing for the development of novel analysis techniques for challenging biological samples by interfacing this spectroscopy with other mass spectrometry techniques.

1.6 Mass spectrometry experiments with light from the visible region

Beyond the lower frequency end of the ultraviolet region of the spectrum sits an interesting band of electromagnetic radiation, commonly referred to as the visible region. This region of the spectrum, situated between 400 nm and 800 nm, corresponds to the peak of the solar emission spectrum, leading to the natural selection of organisms able to detect or utilise this radiation. Within science this part of the electromagnetic spectrum holds a special significance, as experiments utilising these wavelengths can be observed by the human eye. One such experiment was performed in the 17th century by Isaac Newton,[31] who demonstrated that light was made up of individual wavelengths (colours). In the following decades this work was expanded upon until in 1860 Kirchoff and Bunsen developed the first flame spectrophotometer.[31] Continued experimental and theoretical studies led to the development of a number of new analytical techniques, that we now refer to collectively as spectroscopy. From these early days, with the development of better optics and detectors, a multitude of different applications utilising visible radiation have emerged as powerful analytical techniques based on the emission, absorption, or scattering of visible light. As with the UV part of the electromagnetic spectrum, a range of experiments that integrate visible light with mass spectrometers, to provide deeper insights into the structure of gas phase ions have been explored.

1.6.1 Understanding and mechanism

In order to understand how these techniques can be combined with mass spectrometers it is helpful to consider the underlying mechanisms. As with all other forms of spectroscopy, the interaction between light and matter requires two properties. Firstly, the molecular or atomic system must contain a transition (such as an electronic transition within a molecule) with an energy gap matching a photon in the visible part of the spectrum. The excitation of this transition must exhibit a non-zero transition moment allowing it to couple to the oscillating electric field of the photon. If these two conditions are met, the irradiation of the material with the appropriate wavelength will lead to the absorption, emission or scattering of light. Normally, when a molecule or one of its moieties meets these conditions these groups are referred to as chromophores. The most common interactions between molecules and electromagnetic radiation in the UV and visible regions of the spectrum can be summarised using Jablonski diagrams.[95] The most common processes are summarised in the Jablonski diagrams shown in Figure 1.16.

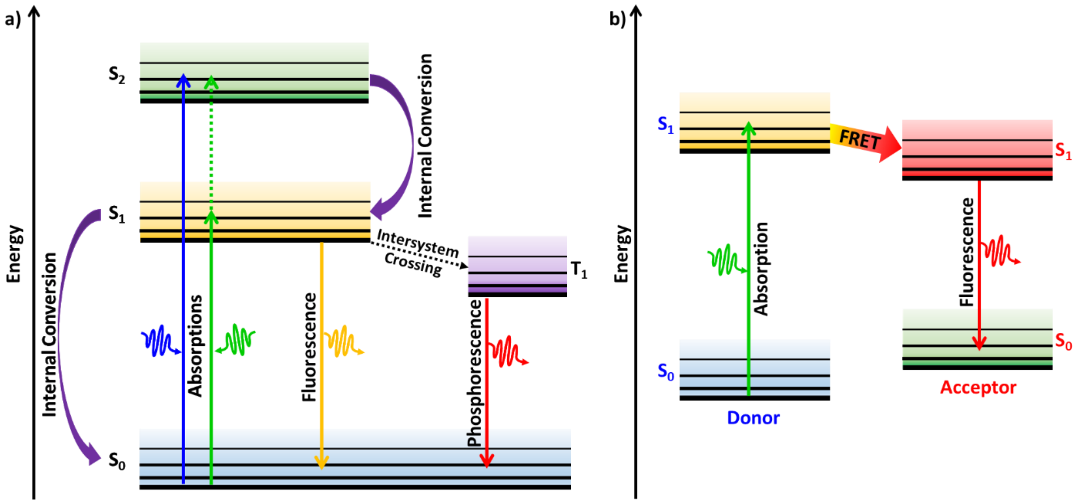


Figure 1.16: Jablonski diagrams summarising the interactions between light with a) a single molecule and b) the molecular pair often utilised in FRET. S_0 is the singlet ground state, S_1 is the singlet first excited state, S_2 singlet second excited state and T_1 is triplet excited state.

These diagrams are used to summarise the individual electronic state transitions that molecules can undergo upon interaction with UV and visible radiation. The simplest of these to understand is absorption, shown in Figure 1.16a. Here the incoming photons are resonant with the energy gap between two electronic states within the molecule. The light is absorbed and the molecule becomes excited; due to the fact that each electronic energy level has a series of vibrational modes associated with it a band of wavelengths around the energy gap between the states can be absorbed. This process yields a UV/vis absorption spectrum of the chromophore within the irradiated molecular system. Once the molecule is in the excited state, it can either transition back into the ground state by a radiative pathway (*e.g.* fluorescence), a non-radiative pathway (*e.g.* a structural rearrangement) or in the case of some biomolecular structures this excitation energy is used to perform chemical reactions (*e.g.* chlorophyll). These types of biomolecules are widely observed in nature and often found in plants to aid in their development and growth.

In order for a biomolecule to be monitored *via* the visible section of the electromagnetic spectrum, it must contain a chromophore that absorbs light within this range. Such chromophores are often cofactors, which in turn can convey the action of the light to a variety of larger biomolecular hosts. One important example of such a cofactor is retinal, which is found with the photoreceptor proteins opsins.[96] Another family of chromophores are bilins, found in mammals, vertebrates, invertebrates, and plants, including green plants and red algae. In plants, bilins in phytochrome proteins convey a response to light which can alter the growth of the plant, focusing the energy on either growing the stem to reach more light or growing the leaves to absorb the light that it is experiencing.[97,98]

In addition to utilising naturally occurring chromophores, biomolecular systems can be studied by the incorporation of fluorescent dyes. This approach has been used to study DNA melting, reformation and its interaction with a range of small molecules, by replacing the naturally occurring adenosine nucleotide with one of its fluorescent derivatives.[99,100–103] If a pair of fluorescent dyes are chosen such that they have approximately parallel transition dipole moments, the fluorescent emission of one of the dyes (the donor dye) overlaps with the absorption spectra of the other dye (the acceptor dye) then such a pair can undergo a Förster resonance energy transfer (FRET) process.[104] Here, the excitation of the donor dye leads to a resonant energy transfer from the donor to the acceptor dye, leading to the presence of fluorescence of the acceptor dye (summarised in Figure 1.16b). FRET is known to be highly sensitive to the distance between the two dyes, allowing this distinct form of fluorescent tags to be used as a molecular ruler.[105–107] The versatility of these spectroscopic methods has led to a concerted effort to integrate these techniques with mass spectrometry. Recently, these efforts have led to the development of action-FRET spectroscopies based on measuring the FRET transmission efficiency in the gas phase. The development of these action-FRET techniques, based on detecting either the fluorescence or the photo-fragmentation of the dye pair, has allowed intermolecular distances within different conformers in the gas phase to be explored. The development of these techniques is outlined in the following section.

1.6.2 Instrument developments

One of the first examples of the integration of FRET into a mass spectrometer was demonstrated by Parks *et al.* in 2003.[108,109] In these studies, a mass spectrometer designed to carry out highly sensitive laser-induced-fluorescence spectroscopy was used to carry out FRET efficiency measurements during the thermal dissociation of short DNA duplexes. The changes in FRET fluorescence with increasing temperature illustrated that the duplex gas-phase melting pathway follows an unzipping mechanism, containing a meta-stable frayed duplex intermediate state.[108,109] This approach, of directly measuring the FRET fluorescence, has also been used by Jockush *et al.*[110–112] to study the conformation and stability of peptides[110] and small protein ions[111] in the gas-phase, by utilising both steady-state and lifetime FRET efficiency measurements. These measurements showed that as the charge state increases both peptides[110] and proteins[111] become more elongated and unfolded, respectively, due to the increases in the coulombic repulsion. Additionally, the lifetime measurements indicated the presence of multiple conformations for the peptides and protein studied, even at low charge states.[110,111]

The success of the FRET studies utilising laser-induced-fluorescence spectroscopy instruments led to the development of an action-FRET spectroscopy analogue to this technique. This action-FRET analogue was first demonstrated by Dugourd *et al*. in 2014.[113] The layout of the adapted mass spectrometer used in the development of this type of action spectroscopy is shown in Figure 1.17.

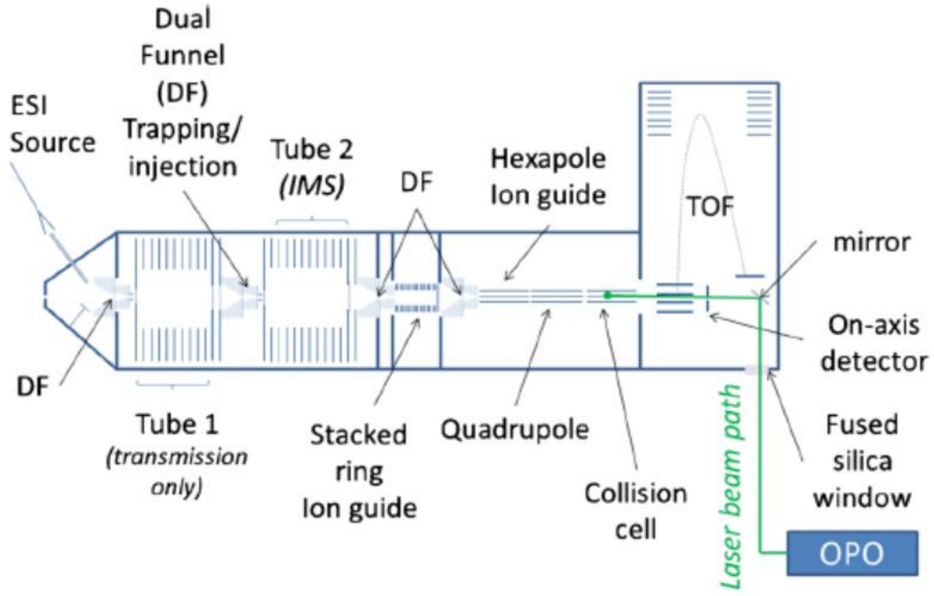


Figure 1.17: Diagram of the ion mobility enabled Bruker maXis Impact QToF instrument modified to incorporate an OPO laser for action FRET spectroscopy. Reproduced from ref. 113 with permission from American Chemical Society, Copyright 2018. [114]

The overall layout is similar to instruments described above that have been adapted for UVPD. Briefly, the ToF head plate of the ion mobility-enabled Bruker maXis Impact QToF instrument, was modified to include a fused silica window and a small mirror in the ToF head. This allows a laser beam to enter the mass spectrometer and be aligned with the ion beam. A tuneable visible OPO laser system is deployed and the action-FRET response of the ions can be investigated.

In proof-of-principle experiments, dye pairs were attached to a series of synthetic peptides of different lengths[113] as well as a peptide fragment of amyloid β,[115] thought to be important in the aggregation of this protein seen in Alzheimer's disease. These studies demonstrated that, as with solution-based FRET measurements, the dye pairs must be carefully chosen, such that there is a suitable amount of overlap between the fluorescence emission spectrum of donor dye and the absorption spectrum of the acceptor dye. Additionally, the resonant energy transfer was improved if the transition dipole moments of the two dyes are approximately parallel. In action-FRET spectroscopy the energy transfer between the two dyes can only be detected by fragmentation of the acceptor dye on the irradiation of the donor dye. This detection methodology imposes an additional set of requirements on the dye pair to be utilised in action-FRET. Namely, the resonant energy transfer between the dye pair must occur before the donor dye undergoes relaxation *via* either fluorescent emission of light or energy redistribution throughout the molecule, *via* an intramolecular energy redistribution pathway (such as IVR) leading to fragmentation of the molecule. Upon transfer of the energy from the donor dye to the acceptor dye, the acceptor dye must undergo fragmentation on a faster timescale than deactivation through one of the radiative pathways, such as fluorescence or phosphorescence.

Fragments of the acceptor dye produced via the action-FRET process must be distinct from fragments produced through other fragmentation methodologies, such as CID. This final requirement is particularly important as it signifies that fragmentation of the acceptor dye occurred as a result of the electronic excitation of this tag. In the first of these two studies a pair of dyes, specifically QSY7 *N*-succinimidyl ester (QSY) and 5-carboxyrhodamine 575 *N*-succinimidyl ester (rh575), were extensively probed by both collision- and laser-induced fragmentation methods to determine their suitability for use in action-FRET spectroscopy. These experiments showed a substantial overlap between the action spectra of these two dyes, and that upon laser-induced dissociation, QSY had a pair of distinct fragments. Grafting these two dyes onto the ends of a short peptide trimer, these two distinct fragments of QSY were found to present in the action spectrum obtained upon the excitation of the rh575 dye, proving the FRET mechanism occurs in the gas phase. Using QSY as the acceptor and rh575 as the donor dye, a series of peptides of varying lengths were studied. Comparing the FRET efficiency of these different model peptide systems to computationally predicted conformations and results from separate IMMS experiments demonstrated that action-FRET also shows a strong dependence on the distance between the acceptor and donor dyes.

Action-FRET was also applied to examine the formation of amyloid β dimers, to gain insight into their role in Alzheimer's disease.[116] A combination of action-FRET, ion mobility measurements, along with molecular dynamics simulations, were brought to bear on a fragment of the amyloid β peptide. Samples were prepared where the peptides either contained the acceptor or the donor dye. Action-FRET and ion mobility measurements were then performed on dimers possessing one of each dye. After initial unbiased sampling, replica-exchange molecular dynamic simulations, constrained by experimental data, provided geometries of plausible dimers structures, shown in Figure 1.18.

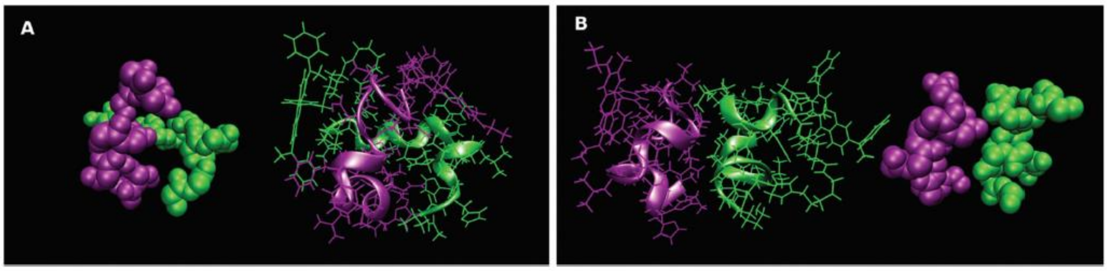


Figure 1.18: Optimised amyloid $\beta_{12\text{-}28}$ *dimer geometries obtained from a) unguided and b) experimentally-guided replica-exchange molecular dynamics calculations. Reproduced from ref. 115 with permission from the PCCP Owner Societies.* [116]

The experimental data guides the molecular dynamics simulation to converge on a more weakly bound, higher-energy dimer structure (compare Figure 1.18b *versus* Figure 1.18a). The combined action-FRET and ion mobility measurements were consistent with this weakly bound dimer being the only populated state observed in the experiment, and further computational work, indicated a large energy barrier between these two dimer geometries in both the gas and the solution phase. This suggests that the weakly bound dimer represents a kinetically trapped state in the formation of amyloid $\beta_{12\text{-}28}$ dimer. This study clearly demonstrated novel and important insights which could be obtained by applying action-FRET to the study of protein aggregation.

Other notable developments have been studies showing self-quenching action-FRET measurements, where one dye is used as both the acceptor and donor.[117] In addition, "hole-burning" IMS-resolved action-FRET spectroscopy, can be used to directly probe structural differences in subpopulations with very similar conformations, as an intensity depletion in the arrival time distribution observed between laser-on and laser-off ion mobility measurements.[114] This approach has been used to probe and monitor the collision-induced unfolding pathway of ubiquitin.[114]

Evan Bieske and valiant co-workers built an IMMS instrument to perform photo-isomerisation action spectroscopy. This incorporates tuneable laser interaction in two positions to irradiate drifting ions. In the first position, the transverse laser pulse is timed to intercept the ions downstream from the ion gate on alternate ion injections whereas the second arrangement allows for much lower light intensities, the ions are irradiated continuously and a mechanical shutter is used to record the laser-on laser-off measurements (see Figure 1.19).[118]

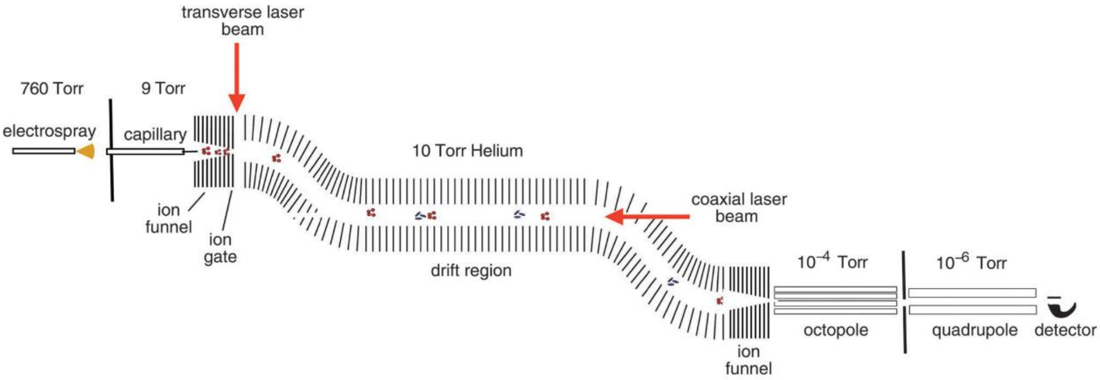


Figure 1.19: Diagram of the ion mobility instrument modified by the Bieske group to incorporate an OPO laser for photoisomerisation action spectroscopy. Reproduced from ref. 117 with permission from the PCCP Owner Societies. [118]

This layout has been utilised extensively on dyes,[119] retinal Schiff bases and its derivatives[120–122] and azobenzenes.[123,124] In this instrument an action spectrum can be produced through two different approaches, the first occurs by monitoring the production of a photo-fragment intensity as a function of the laser wavelength, this will then go on to show the absorption spectrum of the ion in the gas phase. This set-up also incorporates IMMS which lead to further development of a new technique for determining the gas-phase absorption spectrum called photoisomerisation action spectroscopy (PISA).

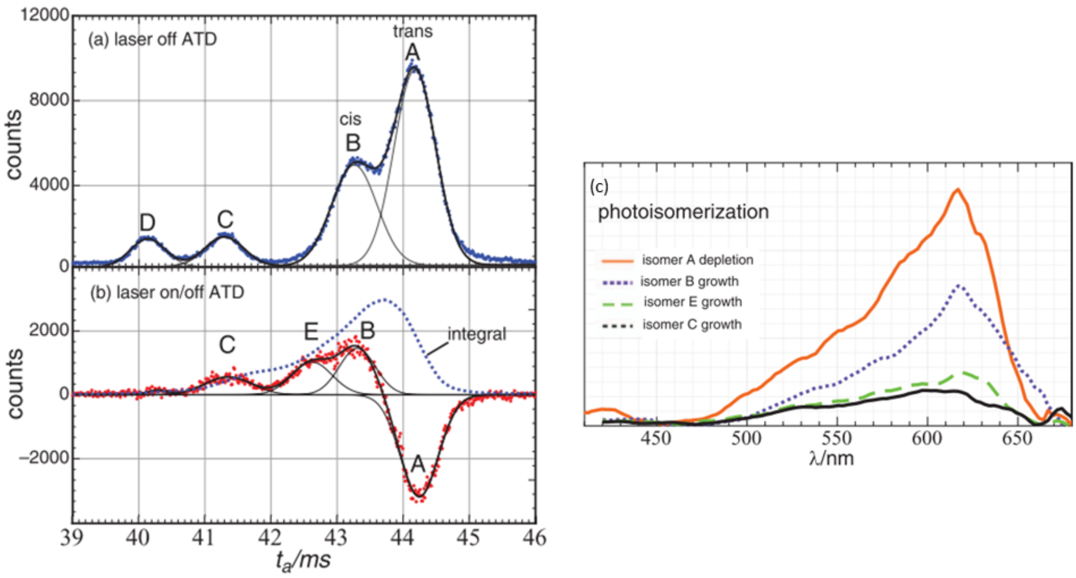


Figure 1.20: Example of retinal protonated Schiff base PISA. (a) ATD of electrosprayed RPSB. (b) Difference between the laser-off and laser on ATDs. (c) PISA plot of the RPSB over the range of the wavelength covered. Adapted from ref. 119 with permission from AIP Publishing, Copyright 2014. [120]

PISA monitors the change in the arrival time distribution (ATD) of a selected ion with laser on/off, an example for this is retinal protonated Schiff base (RPSB), which is a common target that includes an n-butylamine group which will tether the retinal moiety to the surrounding native protein. A PISA spectrum can be produced when irradiation causes a significant change in the ions mobility, Figure 1.20 shows the results of the PISA experiment of RPSB, here Figure 1.20a shows the ATD of RPSB prior to any irradiation of the ion; however, Figure 1.20b shows change in the ATD when under 590 nm pulsed irradiation; here there is a marked decrease in the intensity of the isomer A. These data are then collated as shown in Figure 1.20c for the entire wavelength range investigated to give the unique PISA spectrum which gives a greater representation of the wavelength dependence of the photoactivity of RPSB.[120] This work has been extended to focus on one wavelength that induces fragmentation in the RPSB using the same apparatus; the focus is then placed on determining the mechanism for this photofragmentation.[121]

1.7 Mass spectrometry experiments with light from the infrared region

The infrared[125] (IR) part of the electromagnetic spectrum spans the wavelengths between visible light and microwaves (~800 nm to 1 mm). Due to this large span, many different spectral features are located in the infrared[125] range leading to this part of the spectrum being utilised in applications including astronomy,[126] process analytical techniques,[127] and the exploration of biological processes.[128] In addition to these applications, a subset, between ~2.5-15 μm, of the infrared region contains frequencies that are resonant with the fundamental vibrational transitions found in all molecules. This region, normally referred to as the mid-IR region, is widely probed in a range of IR and Raman spectroscopies and the location of a range of different vibrational modes is summarised in Figure 1.21.

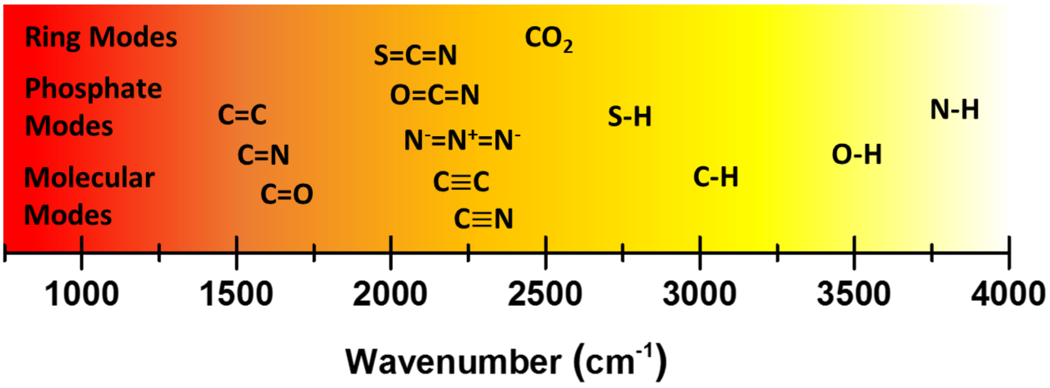


Figure 1.21: Illustration of the mid-IR range from the low frequency (750 cm^{-1}) to the high frequency (4000 cm^{-1}) end with well-known vibrational modes positioned at their approximate wavenumber. The colours are included to indicate the high- and low-frequency ends of the mid-IR spectral range. Adapted from ref. 128 with permission from author, Copyright 2017. [129]

The low end of the mid-IR range (800 – 1500 cm^{-1}) is dominated by a plethora of low-energy modes involving warping of the entire molecular structure, a multitude of different ring modes, and C-H bends.[125] The large number of vibration modes in this region means that this part of the spectrum is highly congested and these modes are often referred to as the intramolecular vibrational bath modes.[125] Vibrational modes associated with double bonds and bonds to hydrogen are found in the 1500 – 1800 cm^{-1} and the 2550 – 4000 cm^{-1} spectral windows, respectively.[125] A wide range of molecular systems contain such bonds, meaning that these modes are found in many solvents, as well as chemical and biological systems.[125] The ubiquitous nature of these modes have led to them being widely used as naturally occurring IR probes to explore the structure,[130] local environments,[131,] vibrational coupling and dynamics of a wide range of different systems.[125,132] For example, the amide I mode, [132,133,133] present in all peptides and proteins, has been used to probe their secondary structure, and the DNA base modes have been used to probe the structural dynamics of DNA and its complexes.[134,135] The remaining part of the mid-IR region (1800 – 2550 cm^{-1}) is populated by the stretching modes of the azides, nitriles, alkynes, cyanates and thiocyanates. These modes have been widely utilised as non-natural IR probes to enable the further exploration of complicated chemical and biological systems/environments.[136,137,138,139]

The ability of these spectroscopic methods to probe the structure, as well as intra- and intermolecular interactions in molecular systems, along with the development of tuneable IR lasers lead to the development of IRMPD spectroscopy. The development of this spectroscopy has allowed the IR spectra of ions to be measured, which are hard to measure *via* traditional absorption spectroscopy due to their very low densities, which are typically less than 10^8 cm^{-1}. In IRMPD the low density of the ions is overcome in a similar way to other ‘action’ spectroscopies, namely the absorption of the light by the ions is not directly measured but the effect of the absorption is measured. As outlined in the introduction (see Section 1.1), the IRMPD spectrum of the gas-phase ion is obtained by monitoring the overall photo-fragmentation yield as the IR wavelength is scanned over the mid-IR region of the electromagnetic spectrum. In the mid-1970’s, with the development of tuneable mid-IR CO_2 lasers (9.2 to 10.8 µm), a range of different experimental studies were carried out on the photofragmentation of molecular species in the gas phase. In 1978, Beauchamp *et al.*[10] demonstrated one of the first implementations of IRMPD on ionic species utilising a trapped ion cyclotron resonance mass spectrometer modified to include a CO_2 laser source.

1.7.1 Understanding and mechanism

This type of photofragmentation process was recognised in the late 1970s to constitute a new family of photophysical effects leading to a great deal of interest into exploring the theoretical underpinnings of this process.[140,141] Through this theoretical work the interaction between a molecular species and IR radiation leading to photofragmentation, led to the identification of at least two different fragmentation pathways. One of these pathways is often referred to as the 'ladder-climbing' model. In this model the targeted vibrational mode of the molecular species coherently absorbs IR photons leading to it stepping through its vibrational energy levels (*e.g.* $\nu_0 \rightarrow \nu_1 \rightarrow \nu_2 \rightarrow \nu_3 \rightarrow$ *etc.*) until its energy has increased to the stage that dissociation of the underlying bond occurs. The main issue with this model is associated with the anharmonicty of vibrational potential energy surfaces. This anharmonicity leads to a decrease in the spacing between successive energy levels, leading to a gradual loss of resonance between the laser pulse and the vibrational mode, preventing further absorption of the IR radiation. This phenomenon is often referred to as the anharmonic bottleneck. Whilst this bottleneck means that pathway is not accessible when the incident radiation from the IR laser is a simple narrowband pulse, the development of ultrafast laser sources capable of producing femtosecond, broadband IR pulses and pulse shaping methodologies[142] has allowed the vibrational 'ladder-climbing' pathway to be accessed.[143] To date such experiments have not been used to generate photofragments but are attempting to utilise highly excited vibrational modes to direct chemical reactivity.[144]

A schematic summarising of the other possible pathway underlying the IR photofragmentation of molecular systems is shown in Figure 1.22.

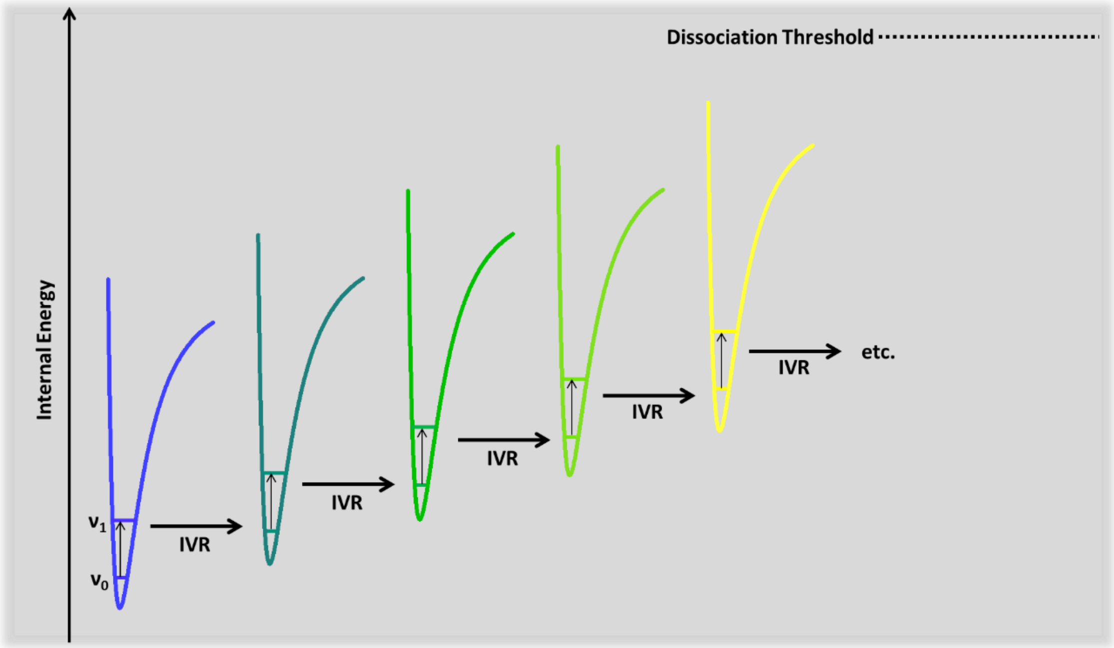


Figure 1.22: Diagram showing the dissociation mechanism underpinning IRMPD via multiple absorption events and IVR. Here the photon is absorbed and quickly dissipates into the vibrational levels by IVR leading to the increased internal energy of the ion, potential re-absorption of a photon and eventual fragmentation. Adapted from ref. 144 with permission from the Royal Society of Chemistry. [145]

In this pathway the absorption of IR light is followed by intra-molecular vibrational energy redistribution (IVR) into the low energy vibrational bath modes, increasing the internal energy of the molecular system. This process is repeated sequentially, gradually heating the molecule and leading to dissociation channels being accessed. The redistribution of the energy, leads to the relaxation of the targeted mode back to the ground state (ν_0), allowing this pathway to avoid the anaharmonic bottleneck. There are several different physical consequences imposed on the nature of the photo-fragmentation due to this mechanism. The first is that the produced photo-fragments are not specific to the IR wavelength being used to excite the molecules, meaning that choosing a wavelength resonant with a specific vibrational mode in the molecule does not lead to only fragments based around the substituents of the moiety vibrating. This consequence has been widely observed experimentally, lending support to this mechanism as underlying IRMPD. Due to this the measurement of the IR action spectrum utilises the total fragmentation yield of the molecule as a function of wavelength, with the fragmentation seen being noted to be broadly similar to CID fragmentation. Additionally, this mechanism depends on the efficient and rapid transfer of the vibration energy of the mode to the molecular bath states. IVR has been noted to be much more efficient in larger molecules, as larger molecules have many more molecular bath states capable of accepting energy from higher lying vibrational modes within the molecule. The implication from this work is that the IRMPD mechanism will be more efficient on larger molecules, although the ‘action’ will be diluted into a greater number of possible fragments.[145] However, the incorporation of mid-IR lasers into mass spectrometers has a rich history and has been utilised to gain new insights into the structures of ionic species in the gas phase.

1.7.2 Instrument developments

The original IRMPD capable instrument was pioneered by Beauchamp *et al.* in the late 1970s,[10] who incorporated a mid-IR laser into a mass spectrometer containing a trapped ion cyclotron resonance cell. In this adapted spectrometer the ions were irradiated within the source of this ICR trap leading to the IRMPD process occurring on a static ion cloud, enhancing the photo-fragmentation yield.[10] This increase in yield, leading to a stronger IRMPD signal, is attributed to the multiphoton absorption mechanism and the fact that such a process is easier to achieve in a static cloud of ions which can be subjected to longer irradiation times. This hybrid instrument allowed the exploration of the potential energy surfaces, photochemistry and photo-physics of a range of small charged molecules and proton-bound dimers of diethyl ether upon IR irradiation as well as allowing multiphoton absorption events to be probed experimentally.[10] This demonstration of the versatility of IRMPD to probe molecules led to the development of a wide range of different instruments, adapted to incorporate a mid-IR laser. Whilst the mid-IR laser sources, ion optics and detectors utilised in these new hybrid instruments varies, most of these instruments incorporate an ion trap to boost the IRMPD fragmentation yield. The most common geometries of ion traps, and how IR lasers are incorporated into these traps are shown in Figure 1.23.

In ion traps with a cylindrical geometry, the ions are stored in a short region of the ion path containing ion optics which can create a local potential minimum for storage. This can be achieved in a range of different ion optical arrangements including but not limited to a series of stacked rings, a segmented hexapole, or a penning trap.[146] As shown in Figure 1.23a, such a trap leads to the formation of an elongated ellipsoidal ion cloud. This means that the most efficient way to irradiate the ions in such a trap geometry, and enhance the overlap between the laser beam and ion cloud, is to incorporate the laser beam along the ion path. An example of this is the incorporation of a CO_2 laser into an orbitrap instrument, as demonstrated by Brodbelt *et al*.[79] (Figure 1.13), where the laser propagates in the opposite direction to the ion beam. The opposite approach is the incorporation of a mid-IR laser into an instrument via the source block, leading to the laser and ions beams propagating in the same direction (Figure 1.8).

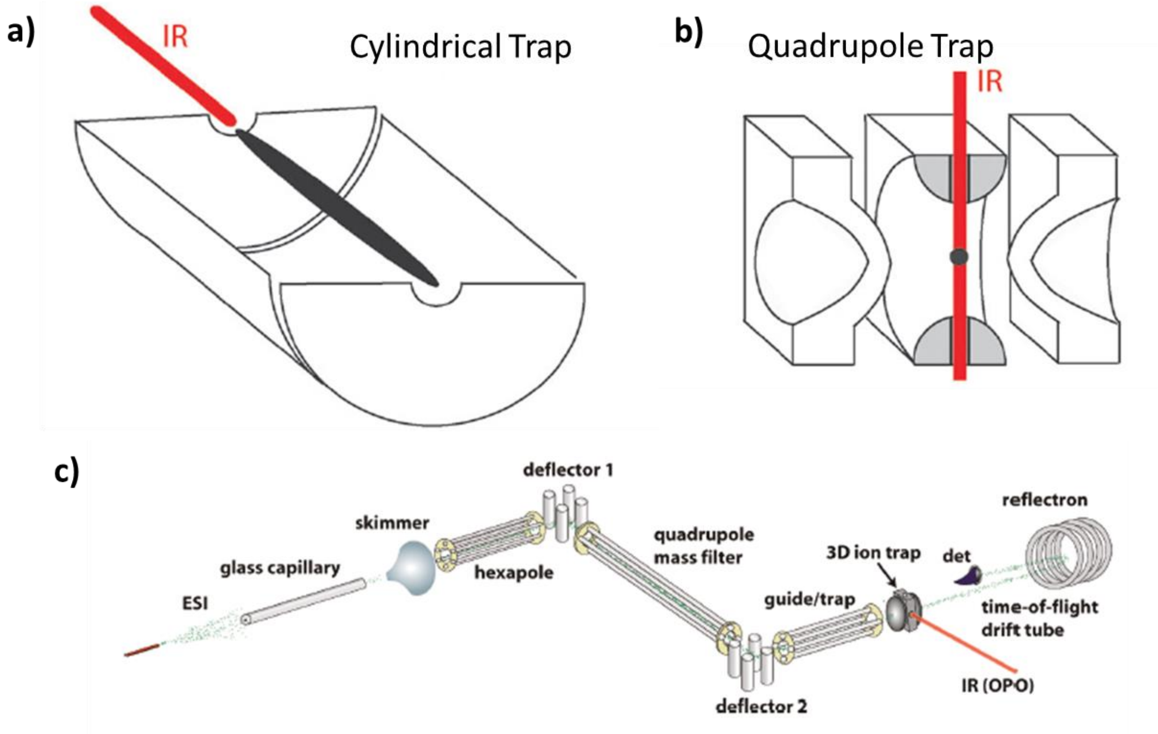


Figure 1.23: Incorporation of an IR laser source into a) cylindrical, b) quadrupole ion trap geometries commonly found in IRMPD-MS instruments and c) the ion optics and geometry of an IRMPD-MS instrument. Ion clouds and IR laser beams are shown in black and red, respectively. Adapted from ref. 146 with permission from the Royal Society of Chemistry and ref. 144 with permission from American Chemical Society, Copyright 2011. [147]

Whilst the incorporation of the laser beam into the mass spectrometer with this trap geometry is more compatible with a lot of the commercial instrument currently available, the major disadvantage is down to issues associated with the focusing of the IR beam within the elongated ion cloud and the need for precise alignment of the laser beam. This precise alignment is due to the fact that the inner diameter of ion optics is narrow, required for the efficient transmission of ions, and even a small misalignment can lead to the beam being partially blocked leading to a loss of IRMPD yields. Additionally, the elongated shape of the ion cloud means that the focusing of the beam throughout the cloud will vary. Since IRMPD is a multiphoton process, and such processes are highly dependent on the photon density, this means that the IRMPD efficiency will vary throughout the ion cloud leading to less overall fragmentation. Due to these two issues, some of the instruments which have been adapted to incorporate an IR laser source utilise a quadrupole trap geometry, shown in Figure 1.23b. The quadrupole potential generated in a trap utilising this geometry leads to a much more spatially confined ion cloud (as shown in Figure 1.23b). This trapping geometry has been incorporated into MS instruments, as demonstrated by Polfer *et al.*[147] (Figure 1.23c). In these traps the IR laser beam is introduced into the trap by drilling holes in the central annular electrode, and so the laser can be directly overlapped with the highly localised trapped ion cloud. Due to this the alignment of the beam into the trapping region of the MS instrument is easier, as the beam does not need to travel along the ion path. Additionally, it is easier to focus the IR beam equally throughout the entire ion cloud. The combination of these two aspects means that the IRMPD fragmentation yield is enhanced over the cylindrical trap geometry allowing less intense vibrational modes of the molecular ions to be probed. However, there are now fewer commercially available MS instruments which incorporate a quadrupole ion trap, and this approach requires more significant modification to incorporate it into an instrument.

Utilising these trapping geometries and benchtop OPO and dye laser systems, IRMPD has been used to obtain insights into a range of different systems. Such systems include but are not limited to probing and understanding the structures of short peptides as well as their dimers in the gas phase[148] and yielding experimental evidence for the formation of hydration rings around hexahydrated sulfate dianions.[149] One recent study utilised IRMPD spectroscopy to analyse heterogeneous samples of monosaccharides.[150] The relative populations of anomers and conformers normally present in these samples are difficult to analyse by MS methods as these structures are isobaric. In this study the IRMPD spectra of three different monosaccharides were experimentally obtained and reconstructed as a linear combination of the simulated IR spectra of the lowest energy geometries of their different anomers (summarised in Figure 1.24).

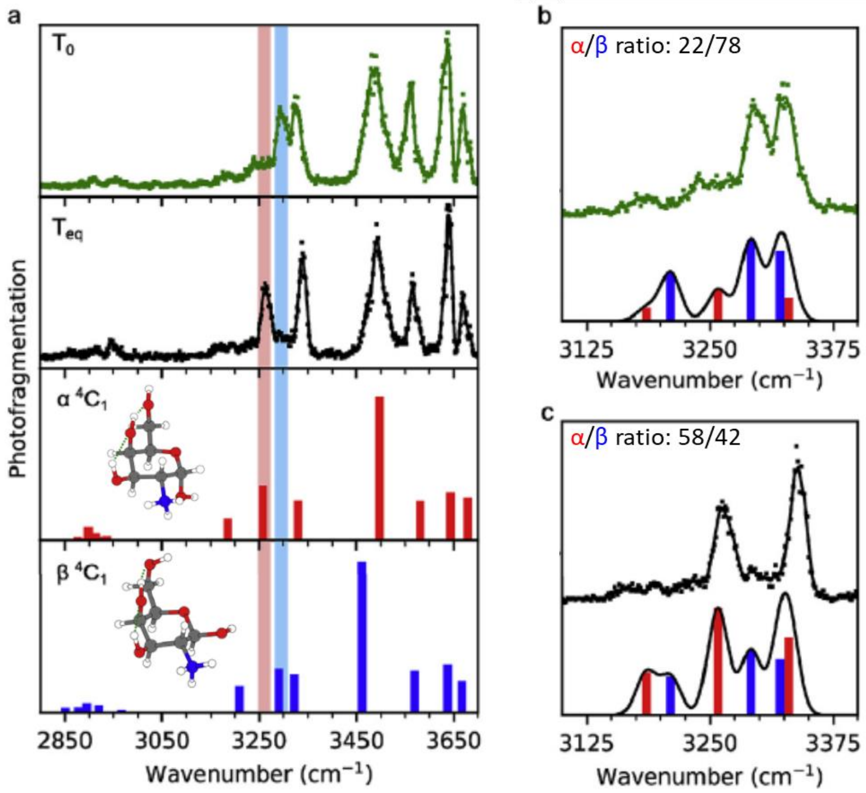


Figure 1.24: a) Comparison between the experimentally obtained IRMPD spectra of galactosamine (GaIN), before (green, T_0) and after (black, T_{eq}) the mutarotation equilibrium is established upon dissolution, and the simulated IR spectra of the lowest energy conformations of the α- (red) and β-glactosamine (blue) anomers. Comparisons between the experimental and mixture of the simulated spectra, constructed by minimising the RMSD, for the GaIN solution b) at T_0 yielding a 22/78 α/β ratio and c) at T_{eq} yielding a 58/42 α/β ratio. Adapted from ref. 149 with permission from Elsevier, Copyright 2020. [150]

The resulting populations of the different anomers were then validated by NMR spectroscopy on the same samples. This application of IRMPD is one of the first examples of this action spectroscopy being employed as a powerful way to determine the presence and populations within heterogeneous samples containing isobaric species, without the need for additional separation techniques to be implemented. The development of these hybrid IRMPD instruments and these studies showing the additional insights that can be obtained by IR action spectroscopy have led to further developments in this field.

1.7.3 Further developments in the field

Due to the success of IRMPD to probe these systems, there has been a continual effort to further develop this methodology. One of these developments has focused on probing traditionally weaker IRMPD vibrational modes. Whereas other developments, have aimed at gaining further insights into the molecular structures and interactions in gas-phase species, as can be routinely performed on their solution phase counterparts *via* IR spectroscopic methods. Such work has focused on the incorporation of additional non-resonant laser sources, controlling ion temperature, development of higher resolution and quantitative IRMPD methodologies, probing dissociation pathways accessed by the multiphoton absorption and increasing the coverage of the mid-IR spectral range.[145]

Studying vibrational modes has led to a range of insights into structural aspects of different molecules as well as allowing clustering interactions to be probed. Incorporating the insights obtained by probing such vibrational modes, has allowed differences in molecular structure between ion mobility separated conformers to be explored. Unfortunately, it has been noted that IR photo-fragmentation yields vary with the identity of the vibrational mode being probed. One example of these traditionally weaker modes in IRMPD is the N-H stretch,[145,151] noted to be intense in solution-based IR spectroscopies. It is thought that the multi-photon absorption underlying the fragmentation mechanism is responsible for the low intensities observed in IRMPD spectra for these modes. Specifically, it is thought that the irradiation of these traditionally weaker IRMPD modes (such as the N-H asymmetric stretching mode in amino acids) with an IR laser leads to the production of a population of ions excited to varying levels of internal energies. This means that a much smaller portion of these ions have been excited to such a level as to allow them to access the dissociation pathways, when compared to stronger IRMPD modes.[145] This leads to only a very small amount of ions fragmenting upon exposure to the laser pulse and so the signal observed from this mode is very weak. One way to overcome this issue is to control the temperature of the ions upon irradiation with the IR laser. If the temperature of the ions is raised, less energy is required for these ions to reach the dissociation thresholds required to undergo IRMPD.[148,152] This methodology can be particularly useful for experiments involving strongly bound complexes.[148,152] Another approach which has been explored to overcome the issue of a majority of ions only reaching an intermediate level of internal energy is to alter the optical setup to allow the ions to be irradiated with both a resonant and a non-resonant laser source, commonly a CO_2 laser which is non-resonant with the modes of a wide range of molecules. Whilst this laser source is not resonant with any of the vibrational modes of the ions when they are not excited, it is noted that when their internal energy is increased by the absorbance of photons from the resonant laser source the density of states of the ion increases.[145] This change in the density of states means that ions excited to an intermediate level can absorb the non-resonant photons, supplied by the two laser setup, allowing these ions to reach the energy required to overcome the dissociation threshold. This approach was first implemented to probe a range of gaseous ions,[153] including carbonium ions.[146] One recent study,[151] incorporating both an OPO laser system and a CO_2 laser, allowed the N-H and O-H stretches of earth alkaline metal-tryptophan complexes to be studied (Figure 1.25).

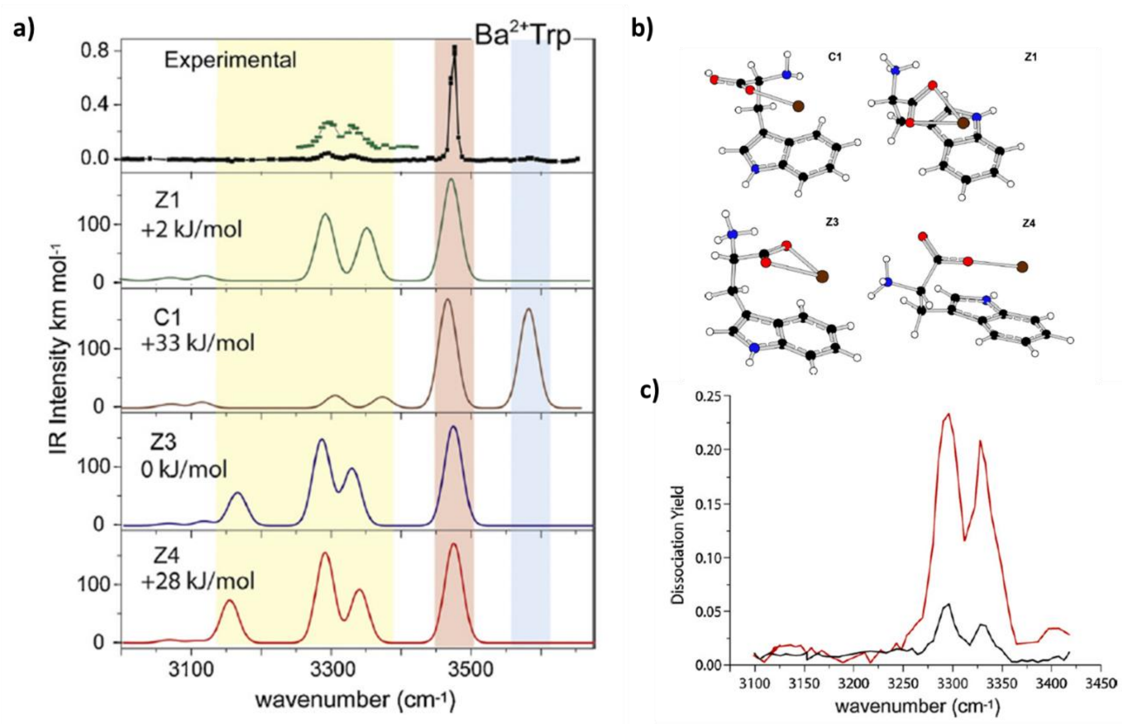


Figure 1.25: a) Comparison between the experimental IRMPD spectrum of the Barium-Tryptophan complex and the simulated IR spectra for low energy charge-solvated (C1) and zwitterionic (Z1, Z3 and Z4) conformers, b) structures of low energy charge-solvated (C1) and zwitterionic (Z1, Z3 and Z4) conformers of the Barium-Tryptophan complex and c) the photo-fragmentation yields of the N-H stretches with (red) and without (black) the non-resonant CO_2 laser, illustrating the 5-fold increase in yield. Adapted from ref. 150 with permission from Elsevier, Copyright 2020. [151]

In this study the inclusion of the non-resonant laser increased the fragmentation yield for the N-H stretches of these complexes by 5-fold to ~25% (Figure 1.25c).[145] This increase in yield allows the conformation of these complexes to be unambiguously assigned with an instrument utilising a benchtop laser system, by comparison to the computationally predicted IR spectra for the different possible conformers (Figure 1.25a & b). The success of this two-colour laser system approach is also thought to be applicable to systems currently installed at FEL facilities to allow weaker vibrational modes to be probed over a larger wavelength range.

In addition to enhancing the signal strength of a range of different vibrational modes that are traditionally weak in IRMPD, by carefully controlling the temperature of the ion upon irradiation or employing a non-resonant laser source to boost the fragmentation yield, ion "tagging" has also been explored.[145,154] Normally a technique employed in cold ion spectroscopy, this methodology relies on forming a complex between the ions of interest and a small tagging molecule. The underlying mechanism behind IRMPD leads to an increase in internal energy until the dissociation threshold is met, meaning that the weakest bonds in the molecular ion fragment first upon irradiation. In the case of the "tagged" ions, the weaker noncovalent interactions underpinning the complex have a much lower dissociation threshold, increasing the fragmentation yield upon irradiation with IR lasers.[145] This significantly enhances the IRMPD signal for all of the vibrational modes within the tagged molecular ions. At room temperature ions are normally tagged using metal cations, allowing molecular ions to form stable "tagged" complexes. This approach has been used to probe the amide I and II bands of a number of different cytochrome c charge states simultaneously, due to the ability to identify pairs of photo-fragments and their precursor ions (Figure 1.26).[155]

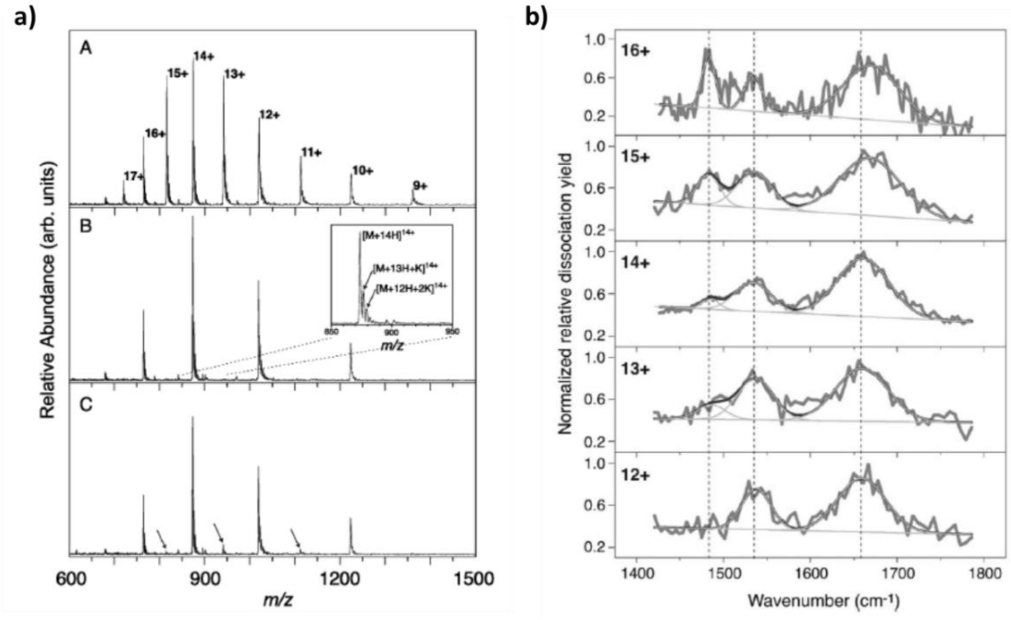


Figure 1.26: a) The full mass spectrum of cytochrome c (panel A), as recorded on a FTICR mass spectrometer, the mass spectrum of cytochrome c (panel B) obtained upon applying mass selective isolation, set to only retain even charge states, and the IRMPD fragment spectrum of the even charge states (panel C) showing the dissociation of K^+ at 6 μm. b) The fitted action-IR spectrum recorded for the different charge states of cyctochrome c highlighting bands at 1483, 1535 and 1660 cm^{-1}. Adapted from ref. 154 with permission from the PCCP Owner Societies. [155]

These vibrational modes are well-characterised in solution-phase IR spectroscopy and are known to be excellent reporter modes on the secondary structure of proteins and peptides. Interestingly, these bands in the lower charge state ions show the presence of α-helix motifs within the ions which is lost as the amount of charge held on the ion increases (Figure 1.26b). This study shows the potential versatility in understanding how much of the native structure of these types of biomolecular systems is conserved upon the transition to the gas phase.

Finally, over recent years IRMPD on a range of different systems has been probed using FEL sources such as CLIO, in France, and FELIX, in the Netherlands. These FEL sources are becoming increasingly popular for IR action spectroscopy due to the wider range of IR frequencies that can be accessed by these sources. These experiments have been used to gain deeper understanding of the structures of sodium bound hetero- and homo-dimers,[156] gaining insights into the structure of the dissociation product of adamantine, the adamantyl cation, as well as experimentally measuring its IR spectrum for the first time,[157] and to probe the structures of metal-bound uracil dimers as well as their hydrated counterparts.[158] One notable IRMPD study carried out utilising a FEL source was carried out recently into understanding the structures of protonated mismatched DNA base-pairs involving guanidine.[159] Understanding these structures is important in gaining a deeper understanding of their biologically as they are thought to occur upon natural mutagenesis and it is known that such structures are often missed by the natural cellular checking mechanism.[159] In this study the obtained IRMPD spectrum of the protonated guanidine homodimer and well as the protonated heterodimers it forms with adenine and thymine were measured and compared to the computationally predicted lowest-energy conformers to obtain structures for these mismatched pairs for the first time.[159] Such experiments allow the stability of these protonated mismatched pairs to be explained with an aim of understanding how such structures go unnoticed by the cellular protein responsible for identifying DNA mutations. This study demonstrates how the development of these FEL sources is allowing IRMPD to be expanded to a range of additional molecular systems.

1.8 Outlook

Since its emergence the in the late 1970s, photodissociation and action spectroscopies have become ubiquitous as MS fragmentation techniques. Employing many different wavelengths and adopting a wide range of spectroscopic strategies, these methodologies have unique abilities to interrogate ions in the gas phase. The development in these techniques has further been boosted by developments in benchtop lasers systems, FEL sources, and computational methodologies, allowing increasingly more complicated action spectroscopies and samples to be explored and understood. Overall, this has yielded important insights into the processes underlying important chemical and biological systems. In the UV region of the electromagnetic spectrum, these techniques range from allowing proteins to be sequenced *via* a top-down UVPD fragmentation approach, to the development of CD-action spectroscopy allowing MS techniques to access the chirality of large macromolecules, such as DNA duplexes. Additionally, the incorporation of LEDs to activate samples in the electrospray source opens up the possibility of studying photo-responsive systems and their kinetics for the first time by MS. In the visible region, the development of action-FRET has expanded the capabilities of IMMS by supplying information on the site-to-site distances and how these vary within different conformers of an ion. Additionally, the development of photo-isomerisation action spectroscopy allows the function of photo-switching molecules to be probed in the gas-phase, allowing computational models, of these important molecules to be rigorously tested experimentally. Finally, the development and understanding of the mechanism underlying IRMPD has allowed the measurement of action-IR spectra over large parts of the mid-IR region. This allows the structure of clusters, hydration states and secondary structures of proteins in different charge states to probed. Furthermore, the developments in IRMPD allow populations in heterogeneous samples containing isobaric species to be determined and the structures of protonated DNA mismatched base pairs to be explored. Taken together, this exciting rapidly developing set of advanced fragmentation techniques, offers a range of otherwise inaccessible insights, allowing the versatility of MS to be expanded into non-traditional applications and helping to further enhance this rich analytical technique.

1.9 Other methods for the investigation of biological structure

1.9.1 Biological Mass Spectrometry

Mass Spectrometry is not only useful for the identification of small molecules but can also be used for the structural analysis of large biomolecular molecules such as proteins. This sort of investigation requires soft ionisation to transfer the biomolecule into the gas-phase without disrupting the non-covalent interactions and altering the structure. It is these soft-ionisation developments that have caused the recent growth in “native-MS”. This approach describes the protein being kept in conditions similar to what it is found in its natural environment. However, as MS is a gas-phase technique, there is already a stark difference to the natural environment, therefore in this case native refers to the environment prior to ionisation.[160]

This native-MS approach is important to get valuable information on the native fold of biomolecules that may directly relate to their function, and in turn, could relate misfolding and aggregation to disease development. It is therefore important to understand the interactions between native proteins before going onto drug design and the using the developed technology in parallel present methods for disease diagnostics.[160]

Soft ionisation techniques allow the analyte enough energy to get into the gas phase but to not give it too much energy to alter the native-like conformation it will adopt, throughout this thesis the ionisation technique used is ESI. [161,162] ESI was first developed by John Fenn in 1989, where he explained it as a method of producing gaseous ions in-vacuum from large complex molecules in solution phase. Specifically, ESI has the ability to produce gaseous ions of large, fragile biomolecules which require additives to stabilise their structure.[163,164]

Native-MS tends to utilise nESI, this gives the added benefits of a lower sample consumption, lower flow rates and smaller droplets. [163] It has also been shown that the process of nESI is a gentler approach than standard ESI, mainly down to the small starting droplet allowing for few collisions and less energy in the collisions in order to desolvate. [165]

1.9.2 Ion Mobility Spectrometry

The theory behind ion mobility spectrometry (IMS) has been known for decades[166] and has been utilised industrially in airport security where they routinely screen samples against explosives and drugs.[167] The main procedure for separating and characterising ions *via* IMS is based on their gas-phase motion down an electric gradient.[71] The first coupling of IMS and MS was performed by Kebarle in 1965, who successfully coupled a drift tube to the mass spectrometer thus creating IMMS.[168,169] However, in recent years these combinations are becoming more common place due to the increasing number of commercially available instruments as well as a group’s ability to design and create their own instrument.[170]

The simplest IMS instrumentation is drift tube. This comprises of a metal tube that will be filled with an inert gas such as helium or nitrogen. Ions produced at the ionisation source will pass through this tube *via* a voltage gradient that is applied across the tube. An ion will travel down this voltage gradient until they experience collisions with the inert gas molecules that are present, it is these molecules that will work against the pull of the voltage gradient and therefore work as a frictional-like force. The more collisions the molecule experiences, the longer it will take to pass through the drift tube to the detector leading to an arrival time distribution (ATD) of the ion.[167] An ion with a large ATD will have a lower mobility through the drift gas.[171]

The shape of the ion will dictate the ATD due to the number of collisions it will experience as it passes through the tube due to the rotationally averaged area that is observed for the structure.

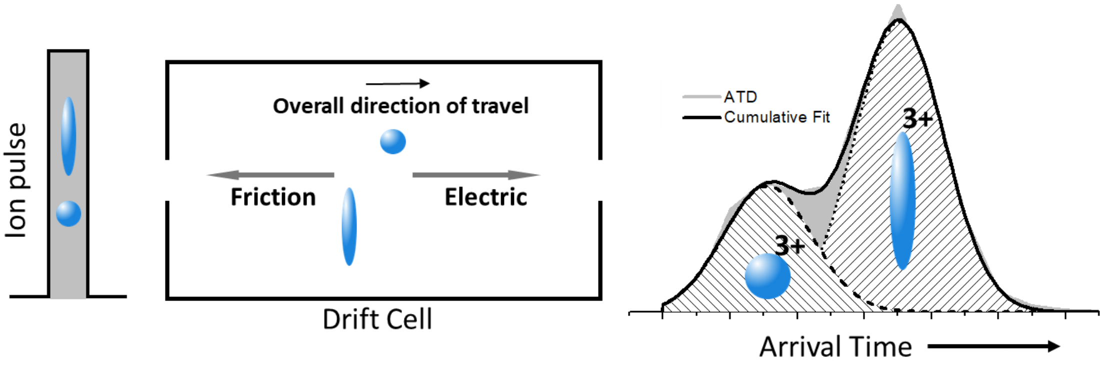


Figure 1.27: Concept of ion mobility spectrometry. The shape of the ion will determine an ions mobility or drift time through an electric field and can be utilised to separate out different conformations of identical ions. The more extended conformation will have a much slower drift time due to the increased collisions with the drift gas, whereas the more compact conformation will have a much quicker drift time. The drift time is effected by collisions with the drift gas which will work as a friction force to slow the ion down, the voltage gradient will allow the ion to move through the drift cell therefore the higher the charge of the ion the quicker the drift time.

A larger more extended structure such as a denatured protein, shown by the rod-like structure in Figure 1.27, will move and spin through the tube resulting in a larger rotationally average area. This will result in an increase in the number of collisions that the ion will feel as it travels through the drift tube. Therefore, the extended structure will have a larger rotationally averaged area that the inert gas molecules can collide with. Whereas a smaller more compact structure, such as the sphere or a globular protein, will have a much smaller rotationally averaged area and therefore there is a smaller chance of the inert gas molecules colliding with the ion.[163]

Another factor which will affect the ATD of an ion is the charge of the ion itself. The ion travels *via* the voltage gradient that the whole system is under, therefore the larger the charge on the ion, the greater the effect of the voltage gradient. Highly charged ions will faster down the voltage gradient than the less charged ions due to the coulombic attraction it feels from the charge at the detector.

The Mason-Schamp equation is used to convert the ATD into a rotationally averaged collision cross section (CCS) value, whereby it will take the charge of the ion, temperature and pressure of the system into consideration.

$$\Omega = \frac{1}{K_0}\frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_bT)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{1}{N}$$

Where Ω is the CCS of the ion, k_0 is the reduced mobility which takes into account the standard temperature and pressure conditions, T is the temperature in kelvin, k_b is the Boltzmann constant, m_I is the mass of the ion, m_B is the mass of the buffer gas, z is the charge state, e is the elementary charge and N is the number density of the buffer gas. [167]

The buffer gas used can also affect the ATD and separation. Nitrogen itself is much larger than helium, this limits the level of detail that the buffer gas can observe. On the other hand, nitrogen is much more polarisable than the helium inert gas, resulting in nitrogen having a larger retarding interaction than helium due to the attractive forces holding the ion back. Both of these properties cause differing ATDs between nitrogen and helium.[172]

CCS information is very important for understanding how the three-dimensional (3D) structure of a protein relates to the function. Structural information can be determined from X-ray crystallography and nuclear magnetic resonance (NMR) structural investigations however these are limited to homogenous samples and are performed in the condensed phase. IMS on the other hand is compatible with heterogeneous samples, and the ions are analysed in the gaseous phase. Theoretical CCS values can also be determined computationally *via* data attained from X-ray crystallography or NMR and compared to gas-phase experimental CCS.[70,173–175]

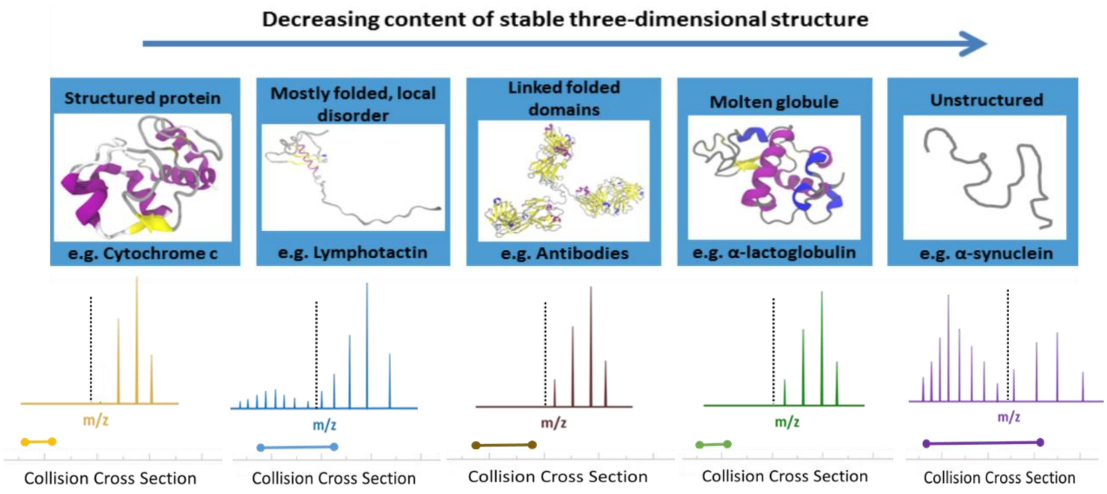


Figure 1.28: Schematic showing the effects of 3D structure on the charge state distribution (CSD) and collision cross section (CCSD). Adapted from ref. 175 with permission from Elsevier, Copyright 2018.[176]

As shown in Figure 1.28, the more extended and disordered a protein is, the larger the range of the CSD and a higher number of charges observed such as for the protein α-synuclein. The more extended a structure, the more charges it can accommodate upon ionisation whilst satisfying the repulsive forces. The opposite is true for more compact structures such as the globular protein Cytochrome-c. The more compact structure will be unable to satisfy the repulsive forces of a large number of charges, therefore the globular proteins will observe a smaller CSD.[176,177]

Ion mobility measurements add another scope for the analysis of protein structure, the MS data works orthogonally to IMS data to give more information on the conformation of the proteins. As shown in Figure 1.28, the more extended proteins result in larger CCS values and tend to occupy a wider range of CCS. More specifically IMS is able to distinguish between the different conformation where there may be limited difference in the charge state distribution.[74,176]

1.9.3 Hydrogen-Deuterium Exchange

An extension to MS, is the use of hydrogen-deuterium exchange (HDX). Analogues to the implementation of the exchange in NMR, here a protein can be exposed to a deuterated buffer for a given time and isotope exchange is fastest for regions that are solvent-exposed. Readily accessible hydrogens will exchange with the deuterated buffer and therefore result in a shift in mass, the exchange rates for these hydrogens are often widely different depending on their location on the protein, backbone hydrogen exchange should be complete within a minutes. For example, an unfolded protein is expected to have a quick exchange rate for all labile hydrogens, but a tightly folded protein will have far fewer number of hydrogens that are readily accessible to the solvent and will exchange rapidly.[178,179] This gives a plot that can yield information on the extent of unfolding of a protein by the number of exposed positions. Combining this methodology with digestive enzymes can give an even more refined image of the location of these heavily exchanged locations and therefore give a clearer image of the regions whereby the system is more exposed and unfolded. [180]

This technique allows for further information to be yielded and can be combined with MS alone to give information on the extent of exchange or combined along with IMS to give a more detailed image of the global structure. Compact conformers can protect some hydrogens that become exposed in more extended conformers, this is similar to what is observed in solution phase, whereby compact conformers are compared to with denatured conformers.[179] In many cases, due to the limitations of the drift gas there may be some finer information missed by IMS where HDX can give information on the accessibility such as the previously mentioned fragmentation capabilities.

1.9.4 Molecular Dynamics and Modelling

Typically, to further understand the results obtained *via* the techniques applied throughout this thesis, molecular dynamics measurements can be utilised throughout to gain further information on the numerical values obtained and what they potentially mean in terms of structure and conformation. As IMMS yields a global CCS value there is little information means other than an increase or a decrease implying an unfolding or folding of the structure respectively. Therefore, molecular dynamics is often applied to gain further understanding on what these CCS values actually describe. For small molecules this process involves utilising X-ray crystallography structures as an approximation for the conformation, this conformation can then be optimised and utilising software such as MOBCAL the theoretical CCS can be determined in the specified drift gas, such as helium.[181] For larger structures such as proteins and nucleic acids, this initial crystal data may be difficult to obtain, therefore entirely molecular modelling is applied and theoretical CCS can be yielded on these conformations.[182]

A prime example of this the application of modelling when applied to the photoreceptor UVR8. In the full length UVR8 protein, it was observed that there were higher charge states present as well as the larger CCS measurements for the monomer, these were not observed when just looking at the core UVR8. Through molecular modelling, a series of structures could be modelled that could yield a similar CCS to what was achieved experimentally, and what was derived from this was the presence of the disordered tail that would allow the formation of the higher charge states and larger CCS measurements. [85]

1.9.5 Stopped-Flow Fluorescence

An alternative methodology applied in this thesis is stopped-flow fluorescence measurements, this does not incorporate MS is the use of stop-flow fluorescence measurements. In this thesis, this was specifically applied to the binding of cofactors and was able to yield kinetic information that cannot be obtained through MS methodology. Stopped-flow fluorescence involves mixing two reagents together and then stopping this in an observation cell, in this thesis this is the protein and its cofactor.[183] This cell is then irradiated with light, e.g. 280 nm, and the signal is taken as the reaction proceeds, this is typically absorbance and fluorescence but other methods such as circular dichromism have been used in stopped-flow.[184,185] Other techniques have been coupled with stopped-flow including NMR,[186,187] fluorescence lifetime measurements [188,189] and small-angle X-ray scattering.[190]

After response, the data can be fitted to a set of reaction models and used to obtain rate constants for the interaction. The conversion of a single type of molecule from one form to another, biomolecular interactions or a multistep reaction can all be modelled in this method.[183] The timescale of these experiments covers a timescale that is quicker than MS measurements and therefore offers another technique that can yield complementary information.

The application of this technique can be seen in Chapter 4 on the mixing of the protein and its cofactor to form a tetrameric species, this is combined with IMMS and molecular modelling techniques to yield a greater understanding of the system. Therefore, is a prime example of the synergy of these techniques and the value added when using them in combination.

1.10 Developments in photoactivation and dissociation methods applied to mass spectrometry and biomolecules

As described in the previous sections there are a variety of different approaches to monitor the photoactivation of biomolecules, with many being applied to MS. The focus of this thesis is the application of UVPD and LED illumination for the photoactivation of proteins and to monitor structural and conformation changes that are exhibited under these conditions. This has been approached through the implementation of a 213 nm laser and LED source illumination onto a Waters Synapt G2-S with further applications to different mass spectrometers.

In this thesis the aim is to expand the current implementation of UVPD and LED irradiation and apply them to the mass spectrometers including the ion mobility enabled Synapt G2-S. Current UVPD methodology apply standard fragmentation assignment of the McLuckey series and using these to determine the sequence coverage. The first part of this thesis aims to improve the ability to monitor the effect of precursor changes, such as conformation and charge state (and potentially others untested here), and it’s effect the change in the fragmentation pattern. Thus from the fragmentation pattern can be group *via* multiply component analysis (MVA) and similarities and difference can be determined, such as global conformational changes, charge states and cofactor binding.

Previous work utilised well-known commercial native-like proteins and the source activation was either kept at soft or harsh conditions, these different source conditions resulted in folded and unfolded conformations, respectively. 213 nm UVPD was then applied to these different conformations and resulted in different fragmentation yields, this work can be found in Appendix B. Chapter 2 describes an extension to this work, whereby photoactivation alongside traditional collisional activation techniques were performed on a series of well understood commercial proteins. This comparison work was also used as a test to show the benefits of utilising multiple component analysis (MVA) in order to determine structural information from the sequence fragments and unassignable fragments. The work also expands on previous work with the implementation of collisional activation post photoactivation in an attempt to demonstrate the conformational variance.

Previous incorporation of LEDs included a very basic rig using mounted LEDs that were more expensive to buy and therefore the LEDs available were sparser in their coverage of the UV/visible spectrum. For good coverage of the spectrum and initial investigations, it is important that the LEDs are easily mounted and do not require a significant investment to test their applicability. This has been implemented in this thesis and a protocol has been described for the applicability of the LEDs into multiple mass spectrometers. With quick installation and minimal optimisation between optically active proteins, mass spectrometry provides an excellent analytical technique that can show added value and understanding to their behaviour.

Chapter 3 is a methods chapter that explains the implementation of LEDs into the source of a mass spectrometer in order to investigate the behaviour of photoreceptor proteins. Mass spectrometry is not currently a go to method for the understanding of photoreceptors proteins, this methods chapter aims to put MS forward as another technique to add to the arsenal of techniques in the investigations of photoreceptor proteins. Previous work involving the implementation of LEDs has been performed in the group involving mounted LEDs,[85,191] this chapter strives to expand on this to include the use of unmounted LEDs and the creation of the LED holder than can be used with these for installation into the mass spectrometer source. This also includes the implementation and arrangement of multiple LEDs for photoreceptor proteins with many absorption wavelengths.

The added value given from the implementation of IMMS is demonstrated perfectly in the following work in Chapter 4 on the photoactive protein, CarH. By incorporating IMMS techniques with standard molecular dynamics and stop-flow methodology, further understanding was yielded in the uptake of cofactors required for photoactive behaviour. This investigation includes implementing relatively standard approaches for the use of IMMS which are not widely applied outside the realm of those working in IMMS, therefore it is an excellent demonstration of the benefit of utilising this technique.

Chapter 4 involves the investigation of chromophore binding onto a monomeric protein triggering a structural change and a photoreactive response under illumination. The purpose of this investigation was to understand the formation of the multimeric species on binding with the chromophore. The entire process was performed under red illumination to minimise the photoactivation of the resulting photoactive tetramer species. It is a highly collaborative project that involves computational molecular dynamic (MD) simulations and stop-flow measurements to piece together the mechanism of the oligomerisation. It provides an example of the use and benefit of the mass spectrometry in the investigations of photoreceptor proteins.

1.11 Acknowledgements

R.B is grateful to the BBSRC (BB/M011208/1) and Waters Corp for their support, and we note the award from BBSRC that allowed us to build our Photoactivation Ion Mobility Mass Spectrometer (BB/L002655/1). P.E.B and L.A.I.R are grateful to the MS SPIDOC project funded by the European Union’s Horizon 2020 FET-OPEN research and innovation programme (Grant agreement No. 801406) for support of ongoing photo activation IMMS activities. We also acknowledge Dr. Bruno Bellina for all his inputs to optical activation mass spectrometry over the years.

1.12 References

1 S. A. McLuckey, S. Verma, R. G. Cooks, M. J. Farncombe, R. S. Mason and K. R. Jennings, *Int. J. Mass Spectrom. Ion Phys.*, 1983, **48**, 423–426.

2 J. R. Yates, S. F. Morgan, C. L. Gatlin, P. R. Griffin and J. K. Eng, *Anal. Chem.*, 1998, **70**, 3557–3565.

3 J. K. Eng, A. L. McCormack and J. R. I. Yates, *J. Am. Soc. Mass Spectrom.*, 1994, **5**, 976–989.

4 D. F. Hunt, J. R. Yates, J. Shabanowitz, S. Winston and C. R. Hauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 6233–6237.

5 N. C. Polfer, K. F. Haselmann, P. R. R. Langridge-smithy and P. E. Barran, *Mol. Phys.*, 2005, **103**, 1481–1489.

6 R. Black, A. Barkhanskiy, L. A. I. Ramakers, A. Theisen, J. M. Brown, B. Bellina,

D. K. Trivedi and P. E. Barran, *Int. J. Mass Spectrom.*, 2021, **464**, 116588.

7 S. R. Harvey, M. Porrini, R. C. Tyler, C. E. MacPhee, B. F. Volkman and P. E. Barran, *Phys. Chem. Chem. Phys.*, 2015, **17**, 10538–10550.

8 S. R. Harvey, M. Porrini, A. Konijnenberg, D. J. Clarke, R. C. Tyler, P. R. R. Langridge-Smith, C. E. MacPhee, B. F. Volkman and P. E. Barran, *J. Phys. Chem. B*, 2014, **118**, 12348–12359.

9 H. Li, Y. Sheng, W. McGee, M. Cammarata, D. Holden and J. A. Loo, *Anal. Chem.*, 2017, **89**, 2731–2738.

10 R. L. Woodin, D. S. Bomse and J. L. Beauchamp, *J. Am. Chem. Soc.*, 1978, **100**, 3248–3250.

11 S. K. Kim, T. Ha and J. P. Schermann, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3334–3335.

12 A. M. Rijs, G. Ohanessian, J. Oomens, G. Meijer, G. Von Helden and I. Compagnon, *Angew. Chemie - Int. Ed.*, 2010, **49**, 2332–2335.

13 T. R. Rizzo, Y. D. Park, L. A. Peteanu and D. H. Levy, *J. Chem. Phys.*, 1986, **84**, 2534–2541.

14 J. A. Stearns, C. Seaiby, O. V. Boyarkin and T. R. Rizzo, *Phys. Chem. Chem. Phys.*, 2009, **11**, 125–132.

15 J. K. Carr, A. V. Zabuga, S. Roy, T. R. Rizzo and J. L. Skinner, *J. Chem. Phys.*

16 K. K. Irikura, J. W. Hudgens and R. D. Johnson, *J. Chem. Phys.*, 1995, **103**, 1303–1308.

17 E. G. Robertson and J. P. Simons, *Phys. Chem. Chem. Phys.*, 2001, **3**, 1–18.

18 A. I. González Flórez, E. Mucha, D. S. Ahn, S. Gewinner, W. Schöllkopf, K. Pagel and G. Von Helden, *Angew. Chemie - Int. Ed.*, 2016, **55**, 3295–3299.

19 S. Warnke, W. Hoffmann, J. Seo, E. De Genst, G. von Helden and K. Pagel, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 638–646.

20 A. M. Rijs, I. Compagnon, A. Silva, J. S. Hannam, D. A. Leigh, E. R. Kay and P. Dugourd, *Phys. Chem. Chem. Phys.*, 2010, **12**, 12556.

21 A. M. Rijs, I. Compagnon, J. Oomens, J. S. Hannam, D. A. Leigh and W. J. Buma, *J. Chem. sicence*, 2009, **131**, 2428–2429.

22 A. M. Rijs, N. Sandig, M. N. Blom, J. Oornens, J. S. Hannam, D. A. Leigh, F. Zerbetto and W. J. Buma, *Angew. Chemie - Int. Ed.*, 2010, **49**, 3896–3900.

23 N. R. Walker, R. R. Wright, P. E. Barran, J. N. Murrell and A. J. Stace, *J. Am. Chem. Soc.*, 2001, **123**, 4223–4227.

24 L. Puskar, P. E. Barran, R. R. Wright, D. A. Kirkwood and A. J. Stace, *J. Chem. Phys.*, 2000, **112**, 7751–7754.

25 N. R. Walker, R. R. Wright, P. E. Barran, H. Cox and A. J. Stace, *J. Chem. Phys.*, 2001, **114**, 5562–5567.

26 N. Walker, M. P. Dobson, R. R. Wright, P. E. Barran, J. N. Murrell and A. J. Stace, *J. Am. Chem. Soc.*, 2000, **122**, 11138–11145.

27 L. Puskar, P. E. Barran, B. J. Duncombe, D. Chapman and A. J. Stace, *J. Phys. Chem. A*, 2005, **109**, 273–282.

28 A. V. Zabuga, M. Z. Kamrath, O. V. Boyarkin and T. R. Rizzo, *J. Chem. Phys.*, 2014, **141**, 1–12.

29 A. Y. Pereverzev, V. N. Kopysov and O. V. Boyarkin, *J. Phys. Chem. Lett.*, 2018, **9**, 5262–5266.

30 A. Y. Pereverzev, Z. Koczor-Benda, E. Saparbaev, V. N. Kopysov, E. Rosta and O. V. Boyarkin, *J. Phys. Chem. Lett.*, 2020, **11**, 206–209.

31 J. M. Hollas, *Modern Spectroscopy*, John Wiley & Sons Ltd, Fourth., 2004.

32 T. H. Maiman, 1960, **187**, 493–494.

33 A. L. Schawlow and C. H. Townes, *Phys. Rev.*, 1958, **112**, 1940–1949.

34 C. Pellegrini, *Eur. Phys. J. H*, 2012, **37**, 659–708.

35 C. Feng and H. X. Deng, *Nucl. Sci. Tech.*, 2018, **29**, 1–15.

36 M. Karas and F. Hillenkamp, *Anal. Chem.*, 1988, **60**, 2299–2301.

37 H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, *Nature*, 1985, **318**, 162.

38 J. Sperling and K. Hens, 2018, 22–24.

39 M. Gerhards, *Opt. Commun.*, 2004, **241**, 493–497.

40 J. S. Brodbelt, *Chem. Soc. Rev.*, 2014, **43**, 2757–83.

41 W. D. Bowers, S. S. Delbert, R. L. Hunter and R. T. Mclver, *J. Am. Chem. Soc.*, 1984, **106**, 7288–7289.

42 E. R. Williams, J. J. P. Furlong and F. W. McLafferty, *J. Am. Soc. Mass Spectrom.*, 1990, **1**, 288–294.

43 D. F. Hunt, J. Shabanowitz and J. R. Yates, *J. Chem. Soc. Chem. Commun.*, 1987, 548–550.

44 T. Ly and R. R. Julian, *Angew. Chemie - Int. Ed.*, 2009, **48**, 7130–7137.

45 J. P. Reilly, *Mass Spectrom. Rev.*, 2009, **28**, 425–447.

46 G. K. Yeh, Q. Sun, C. Meneses and R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 385–393.

47 J. J. Wilson, G. J. Kirkovits, J. L. Sessler and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 257–260.

48 J. A. Madsen, D. R. Boutz and J. S. Brodbelt, *J. Proteome Res.*, 2010, **9**, 4205–4214.

49 J. R. Cannon, C. Kluwe, A. Ellington and J. S. Brodbelt, *Proteomics*, 2014, **14**, 1165–1173.

50 L. Zhang and J. P. Reilly, *Anal. Chem.*, 2009, **81**, 7829–7838.

51 M. S. Thompson, W. Cui and J. P. Reilly, *Angew. Chemie - Int. Ed.*, 2004, **43**, 4791–4794.

52 T.-Y. Kim, M. S. Thompson and J. P. Reilly, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 1657–1665.

53 A. Devakumar, M. S. Thompson and J. P. Reilly, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 2313–2320.

54 M. S. Thompson, W. Cui and J. P. Reilly, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1439–1452.

55 P. Roepstorff and J. Fohlman, *Biol. Mass Spectrom.*, 1984, **11**, 601–601.

56 A. R. Dongré, J. L. Jones, Á. Somogyi and V. H. Wysocki, *J. Am. Chem. Soc.*,

1996, **118**, 8365–8374.

57 V. H. Wysocki, G. Tsaprailis, L. L. Smith and L. A. Breci, *J. Mass Spectrom.*, 2000, **35**, 1399–1406.

58 A. G. Harrison and T. Yalcin, *Int. J. Mass Spectrom. Ion Process.*, 1997, **165/166**, 339–347.

59 P. Y. Cheng, D. Zhong and A. H. Zewail, *Chem. Phys. Lett.*, 1995, **237**, 399–405.

60 R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1823–1826.

61 P. R. Stannard and W. M. Gelbart, *J. Phys. Chem.*, 1981, **85**, 3592–3599.

62 L. J. Morrison and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2016, **138**, 10849–10859.

63 M. Girod, Z. Sanader, M. Vojkovic, R. Antoine, L. MacAleese, J. Lemoine, V. Bonacic-Koutecky and P. Dugourd, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 432–443.

64 L. A. Vasicek, A. R. Ledvina, J. Shaw, J. Griep-Raming, M. S. Westphall, J. J. Coon and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2011, **22**, 1105–1108.

65 J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2013, **135**, 12646–12651.

66 J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann, *Nat. Methods*, 2007, **4**, 709–712.

67 M. Zhou, W. Liu and J. B. Shaw, *Anal. Chem.*, 2019, **92**, 1788–1795.

68 A. Theisen, B. Yan, J. M. Brown, M. Morris, B. Bellina and P. E. Barran, *Anal.*

Chem., 2016, **88**, 9964–9971.

69 B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon and P. E. Barran, *Analyst*, 2014, **139**, 6348–51.

70 E. Jurneczko, P. E. Barran, J. Kalapothakis, I. D. G. Campuzano, M. Morris and P. E. Barran, *Analyst*, 2011, **136**, 20–28.

71 A. A. Shvartsburg and R. D. Smith, *Anal. Chem.*, 2008, **80**, 9689–9699.

72 K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose and R. H. Bateman, *Rapid Commun. Mass Spectrom.*, 2004, **18**, 2401–2414.

73 T. Sugai, in *Fundamentals of Mass Spectrometry*, 2013, vol. 1, pp. 89–107.

74 A. Theisen, R. Black, D. Corinti, J. M. Brown, B. Bellina and P. E. Barran, *J. Am. Soc. Mass Spectrom.*, 2018, **30**, 24–33.

75 A. Giuliani, J. P. Williams and M. R. Green, *Anal. Chem.*, 2018, **90**, 7176–7180.

76 D. D. Holden, A. Makarov, J. C. Schwartz, J. D. Sanders, E. Zhuk and J. S. Brodbelt, *Angew. Chemie - Int. Ed.*, 2016, **55**, 12417–12421.

77 G. E. Reid, H. Shang, J. M. Hogan, G. U. Lee and S. A. McLuckey, *J. Am. Chem. Soc.*, 2002, **124**, 7353–7362.

78 S. A. McLuckey and J. L. Stephenson, *Mass Spectrom. Rev.*, 1998, **17**, 369–407.

79 D. D. Holden, W. M. McGee and J. S. Brodbelt, *Anal. Chem.*, 2016, **88**, 1008–1016.

80 D. D. Holden and J. S. Brodbelt, *Anal. Chem.*, 2016, **88**, 12354–12362.

81 J. D. Sanders, C. Mullen, E. Watts, D. D. Holden, J. E. P. Syka, J. C. Schwartz and

J. S. Brodbelt, *Anal. Chem.*, 2020, **92**, 1041–1049.

82 Di Wu, Q. Hu, Z. Yan, W. Chen, C. Yan, X. Huang, J. Zhang, P. Yang, H. Deng, J. Wang, X. Deng and Y. Shi, *Nature*, 2012, **484**, 214–219.

83 J. M. Christie, A. S. Arvai, K. J. Baxter, M. Heilmann, A. J. Pratt, A. O. Hara, S. M. Kelly, M. Hothorn, B. O. Smith, K. Hitomi, G. I. Jenkins and E. D. Getzoff, *Science*, 2012, **335**, 1492–1497.

84 L. Rizzini, J. J. Favory, C. Cloix, D. Faggionato, A. O'Hara, E. Kaiserli, R. Baumeister, E. Schäfer, F. Nagy, G. I. Jenkins and R. Ulm, *Science*, 2011, **332**, 103–106.

85 I. S. Camacho, A. Theisen, L. O. Johannissen, L. A. Díaz-ramos, J. M. Christie, G. I. Jenkins, B. Bellina, P. E. Barran and A. R. Jones, *Proc. Natl. Acad. Sci.*, 2019, **116**, 1116–1125.

86 S. Daly, F. Rosu and V. Gabelica, *Science*, 2020, **368**, 1465–1468.

87 M. J. McLeish, K. J. Nielsen, L. V. Najbar, D. J. Craik, J. D. Wade, F. Lin and M. B. Doughty, *Biochemistry*, 1994, **33**, 11174–11183.

88 A. J. Miles and B. A. Wallace, *Chem. Soc. Rev.*, 2006, **35**, 39–51.

89 J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713–1725.

90 A. S. Ladokhin, M. Fernández-Vidal and S. H. White, *J. Membr. Biol.*, 2010, **236**, 247–253.

91 T. Yuan, T. A. Mietzner, R. C. Montelaro and H. J. Vogel, *Biochemistry*, 1995, **34**, 10690–10696.

92 N. Poklar, J. Lah, M. Salobir, P. Maček and G. Vesnaver, *Biochemistry*, 1997, **36**, 14345–14352.

93 G. Zhang, X. Hu and J. Pan, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 2011, **78**, 687–694.

94 A. Shah and P. Ghalsasi, *Eur. J. Phys.*, , DOI:10.1088/1361-6404/ab2fef.

95 A. Jablonski, *Nature*, 1933, **131**, 839–840.

96 G. Wald, *Science*, 1968, **162**, 230–239.

97 M. Legris, C. Klose, E. S. Burgie, C. C. Rojas, M. Neme, A. Hiltbrunner, P. A. Wigge, E. Schäfer, R. D. Vierstra and J. J. Casal, *Science*, 2016, **354**, 897–900.

98 N. C. Rockwell and J. C. Lagarias, *ChemPhysChem*, 2010, **11**, 1172–1180.

99 A. C. Jones and R. K. Neely, *Q. Rev. Biophys.*, 2015, **48**, 244–279.

100 T. Lenz, E. Y. M. Bonnist, G. Pljevaljčić, R. K. Neely, D. T. F. Dryden, A. J. Scheidig, A. C. Jones and E. Weinhold, *J. Am. Chem. Soc.*, 2007, **129**, 6240–6248.

101 L. Ma, X. Wu, G. G. Wilson, A. C. Jones and D. T. F. Dryden, *Biochem. Biophys. Res. Commun.*, 2014, **449**, 120–125.

102 E. Y. M. Bonnist, K. Liebert, D. T. F. Dryden, A. Jeltsch and A. C. Jones, *Biophys. Chem.*, 2012, **160**, 28–34.

103 M. Bood, A. F. Füchtbauer, M. S. Wranne, J. J. Ro, S. Sarangamath, A. H. El-Sagheer, D. L. M. Rupert, R. S. Fisher, S. W. Magennis, A. C. Jones, F. Höök, T. Brown, B. H. Kim, A. Dahlén, L. M. Wilhelmsson and M. Grøtli, *Chem. Sci.*, 2018, **9**, 3494–3502.

104 B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumkell and W. A. Eaton, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 2754–2759.

105 M. L. Wallace, L. Ying, S. Balasubramanian and D. Klenerman, *J. Phys. Chem. B*, 2000, **104**, 11551–11555.

106 E. Boura, B. Rózycki, D. Z. Herrick, H. S. Chung, J. Vecer, W. A. Eaton, D. S. Cafiso, G. Hummer and J. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 9437–9442.

107 B. Saccà, R. Meyer, U. Feldkamp, H. Schroeder and C. M. Niemeyer, *Angew. Chemie - Int. Ed.*, 2008, **47**, 2135–2137.

108 A. S. Danell and J. H. Parks, *Int. J. Mass Spectrom.*, 2003, **229**, 35–45.

109 A. S. Danell and J. H. Parks, *J. Am. Soc. Mass Spectrom.*, 2003, **14**, 1330–1339.

110 F. O. Talbot, A. Rullo, H. Yao and R. A. Jockusch, *J. Am. Chem. Soc.*, 2010, **132**, 16156–16164.

111 M. F. Czar, F. Zosel, I. König, D. Nettels, B. Wunderlich, B. Schuler, A. Zarrine-Afsar and R. A. Jockusch, *Anal. Chem.*, 2015, **87**, 7559–7565.

112 M. F. Czar and R. A. Jockusch, *Curr. Opin. Struct. Biol.*, 2015, **34**, 123–134.

113 S. Daly, F. Poussigue, A. L. Simon, L. Macaleese, F. Bertorelle, F. Chirot, R. Antoine and P. Dugourd, *Anal. Chem.*, 2014, **86**, 8798–8804.

114 S. Daly, L. MacAleese, P. Dugourd and F. Chirot, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 133–139.

115 A. Kulesza, S. Daly, L. Macaleese, R. Antoine and P. Dugourd, *J. Chem. Phys.*, 2015, **143**, 025101.

116 A. Kulesza, S. Daly, C. M. Choi, A. L. Simon, F. Chirot, L. Macaleese, R. Antoine and P. Dugourd, *Phys. Chem. Chem. Phys.*, 2016, **18**, 9061–9069.

117 S. Daly, C. M. Choi, F. Chirot, L. Macaleese, R. Antoine and P. Dugourd, *Anal. Chem.*, 2017, **89**, 4604–4610.

118 B. D. Adamson, N. J. A. Coughlan, R. E. Continetti and E. J. Bieske, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9540–9548.

119 B. D. Adamson, N. J. A. Coughlan, G. Da Silva and E. J. Bieske, *J. Phys. Chem. A*, 2013, **117**, 13319–13325.

120 N. J. A. Coughlan, K. J. Catani, B. D. Adamson, U. Wille and E. J. Bieske, *J. Chem. Phys.*, 2014, **140**, 164307–164310.

121 N. J. A. Coughlan, B. D. Adamson, K. J. Catani, U. Wille and E. J. Bieske, *J. Phys. Chem. Lett.*, 2014, **5**, 3195–3199.

122 N. J. A. Coughlan, M. S. Scholz, C. S. Hansen, A. J. Trevitt, B. D. Adamson and E. J. Bieske, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1483–1490.

123 M. S. Scholz, J. N. Bull, N. J. A. Coughlan, E. Carrascosa, B. D. Adamson and E. J. Bieske, *J. Phys. Chem. A*, 2017, **121**, 6413–6419.

124 J. N. Bull, M. S. Scholz, E. Carrascosa and E. J. Bieske, *Phys. Chem. Chem. Phys.*, 2018, **20**, 509–513.

125 P. Hamm and M. Zanni, *Concepts and methods of 2D infrared spectroscopy*, 2012.

126 Y. Zhou, D. Apai, L. R. Bedin, B. W. P. Lew, G. Schneider, A. J. Burgasser, E. Manjavacas, T. Karalidi, S. Metchev, P. A. Miles-Páez, N. B. Cowan, P. J.

Lowrance and J. Radigan, *Astron. J.*, 2020, **159**, 140.

127 K. Pontius, H. Junicke, K. V. Gernaey and M. Bevilacqua, *Appl. Microbiol. Biotechnol.*, 2020, 5315–5335.

128 J. P. Lomont, K. L. Rich, M. Maj, J. J. Ho, J. S. Ostrander and M. T. Zanni, *J. Phys. Chem. B*, 2018, **122**, 144–153.

129 L. A. I. Ramakers, Strahclyde University, 2017.

130 M. Koziński, S. Garrett-Roe and P. Hamm, *J. Phys. Chem. B*, 2008, **112**, 7645–7650.

131 M. M. Waegele, R. M. Culik and F. Gai, *J. Phys. Chem. Lett.*, 2011, **2**, 2598–2609.

132 M. Cho, *Chem. Rev.*, 2008, **108**, 1331–1418.

133 J. N. Bandaria, S. Dutta, S. E. Hill, A. Kohen and C. M. Cheatum, *J. Am. Chem. Soc.*, 2008, **130**, 22–23.

134 L. A. I. Ramakers, G. Hithell, J. J. May, G. M. Greetham, P. M. Donaldson, M. Towrie, A. W. Parker, G. A. Burley and N. T. Hunt, *J. Phys. Chem. B*, 2017, **121**, 1295–1303.

135 G. Hithell, L. A. I. Ramakers, G. A. Burley and N. T. Hunt, *Applications of 2D-IR Spectroscopy to Probe the Structural Dynamics of DNA*, Elsevier Inc., 2018.

136 M. M. Waegele, M. J. Tucker and F. Gai, 2010, **478**, 249–253.

137 R. Shrestha, A. E. Cardenas, R. Elber and L. J. Webb, *J. Phys. Chem. B*, 2015, **119**, 2869–2876.

138 Z. Getahun, C. Y. Huang, T. Wang, B. De León, W. F. DeGrado and F. Gai, *J. Am.*

Chem. Soc.*, 2003, **125**, 405–411.

139 A. J. Stafford, D. L. Ensign and L. J. Webb, *J. Phys. Chem. B*, 2010, **114**, 15331–15344.

140 J. P. Simons, in *Gas Kinetics and Energy Transfer*, ed. R. J. Ashmore, P. G., Donovan, Volume 2., 1977, pp. 58–95.

141 S. Mukamel and J. Jortner, *J. Chem. Phys.*, 1976, **65**, 5204–5225.

142 G. M. H. Knippels, A. F. G. van der Meer, R. F. X. A. M. Mols, P. W. van Amersfoort, R. B. Vrijen, D. J. Maas and L. D. Noordam, *Opt. Commun.*, 1995, **118**, 546–550.

143 D. J. Maas, D. I. Duncan, R. B. Vrijen, W. J. Van Der Zande and L. D. Noordam, *Chem. Phys. Lett.*, 1998, **290**, 75–80.

144 T. Witte, T. Hornung, L. Windhorn, D. Proch, R. De Vivie-Riedle, M. Motzkus and K. L. Kompa, *J. Chem. Phys.*, 2003, **118**, 2021–2024.

145 N. C. Polfer, *Chem. Soc. Rev.*, 2011, **40**, 2211–2221.

146 L. I. Yeh, J. M. Price and Y. T. Lee, *J. Am. Chem. Soc.*, 1989, **111**, 5597–5604.

147 W. K. Mino, K. Gulyuz, D. Wang, C. N. Stedwell and N. C. Polfer, *J. Phys. Chem. Lett.*, 2011, **2**, 299–304.

148 H. Bin Oh, C. Lin, H. Y. Hwang, H. Zhai, K. Breuker, V. Zabrouskov, B. K. Carpenter and F. W. McLafferty, *J. Am. Chem. Soc.*, 2005, **127**, 4076–4083.

149 M. F. Bush, R. J. Saykally and E. R. Williams, *J. Am. Chem. Soc.*, 2007, **129**, 2220–2221.

150 L. Barnes, A. R. Allouche, S. Chambert, B. Schindler and I. Compagnon, *Int. J. Mass Spectrom.*, 2020, **447**, 116235.

151 W. K. Mino, J. Szczepanski, W. L. Pearson, D. H. Powell, R. C. Dunbar, J. R. Eyler and N. C. Polfer, *Int. J. Mass Spectrom.*, 2010, **297**, 131–138.

152 M. F. Bush, J. T. O'Brien, J. S. Prell, R. J. Saykally and E. R. Williams, *J. Am. Chem. Soc.*, 2007, **129**, 1612–1622.

153 D. M. Peiris, M. A. Cheeseman, R. Ramanathan and J. R. Eyler, *J. Phys. Chem.*, 1993, **97**, 7839–7843.

154 M. Okumura, L. I. Yeh, J. D. Myers and Y. T. Lee, *J. Chem. Phys.*, 1986, **85**, 2328–2329.

155 J. Oomens, N. Polfer, D. T. Moore, L. van der Meer, A. G. Marshall, J. R. Eyler, G. Meijer and G. Von Helden, *PCCP*, 2005, **7**, 1345–1348.

156 K. Le Barbu-Debus, D. Scuderi, V. Lepère and A. Zehnacker, *J. Mol. Struct.*, 2020, **1205**, 127583.

157 N. Polfer, B. G. Sartakov and J. Oomens, *Chem. Phys. Lett.*, 2004, **400**, 201–205.

158 B. Power, V. Haldys, J. Y. Salpin and T. D. Fridgen, *Int. J. Mass Spectrom.*, 2018, **429**, 56–65.

159 R. Cheng, T. D. Fridgen, E. Loire and J. Martens, *Phys. Chem. Chem. Phys.*, 2020, **22**, 2999–3007.

160 A. C. Leney and A. J. R. Heck, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 5–13.

161 A. El-Aneed, A. Cohen and J. Banoub, *Appl. Spectrosc. Rev.*, 2009, **44**, 210–230.

162 E. de Hoffman and V. Stroobant, in *Mass Spectrometry: Principles and Applications*, John Wiley & Sons, Chichester, Third Edit., 2007, pp. 1–14.

163 R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E. Macphee and P. E. Barran, *Anal. Chem.*, 2014, **86**, 10979–10991.

164 J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.

165 J. L. P. Benesch, B. T. Ruotolo, D. A. Simmons and C. V Robinson, *Chem. Rev.*, 2007, **107**, 3544–3567.

166 H. E. Revercomb and E. A. Mason, *Anal. Chem.*, 1975, **47**, 970–983.

167 C. J. Gray, B. Thomas, R. Upton, L. G. Migas, C. E. Eyers, P. E. Barran and S. L. Flitsch, *Biochim. Biophys. Acta - Gen. Subj.*, 2016, **1860**, 1688–1709.

168 P. Kebarle and A. M. Hogg, *J. Chem. Phys.*, 1965, **42**, 668–674.

169 A. M. Hogg and P. Kebarle, *J. Chem. Phys.*, 1965, **43**, 449–456.

170 F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281–294.

171 H. H. Hili, F. William, R. H. S. R. H. S. Louis, D. Mcminn, H. H. Hill, W. F. Siems and R. H. S. R. H. S. Louis, *Anal. Chem.*, 1990, **62**, 1201–1209.

172 C. Bleiholder, N. R. Johnson, S. Contreras, T. Wyttenbach and M. T. Bowers, *Anal. Chem.*, 2015, **87**, 7196–7203.

173 S. H. Chen and D. H. Russell, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1433–1443.

174 B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung and C. V

Robinson, *Nat. Protoc.*, 2008, **3**, 1139–1152.

175 Y. Sun, S. Vahidi, M. A. Sowole and L. Konermann, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 31–40.

176 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, **42**, 177–185.

177 A. C. Susa, Z. Xia, H. Y. H. Tang, J. A. Tainer and E. R. Williams, *J. Am. Soc. Mass Spectrom.*, 2016, **21**, 971–978.

178 V. Katta, B. T. Chait and S. Carr, *Rapid Commun. Mass Spectrom.*, 1991, **5**, 214–217.

179 S. J. Valentine and D. E. Clemmer, *J. Am. Chem. Soc.*, 1997, **119**, 3558–3566.

180 L. Konermann, J. Pan and Y. H. Liu, *Chem. Soc. Rev.*, 2011, **40**, 1224–1234.

181 C. Lapthorn, F. S. Pullen, B. Z. Chowdhry, P. Wright, G. L. Perkins and Y. Heredia, *Analyst*, 2015, **140**, 6814–6823.

182 V. D'Atri, M. Porrini, F. Rosu and V. Gabelica, *J. Mass Spectrom.*, 2015, **50**, 711–726.

183 X. Zheng, C. Bi, Z. Li, M. Podariu and D. S. Hage, *J. Pharm. Biomed. Anal.*, 2015, **113**, 163–180.

184 K. M. Youngman, D. B. Spencer, D. N. Brems and M. R. DeFelippis, *J. Biol. Chem.*, 1995, **270**, 19816–19822.

185 H. Roder, K. Maki, H. Cheng and M. C. Ramachandra Shastry, *Methods*, 2004, **34**, 15–27.

186 J. Balbach, V. Forge, N. A. J. Van Nuland, S. L. Winder, P. J. Hore and C. M.

Dobson, *Nat. Struct. Biol.*, 1995, **2**, 865–870.

187 C. Frieden, *Biochemistry*, 2003, **42**, 12439–12446.

188 B. E. Jones, J. M. Beechem and C. R. Matthews, *Biochemistry*, 1995, **34**, 1867–1877.

189 J. M. Beechem, E. A. James and L. Brand, *Time-Resolved Laser Spectrosc. Biochem. II*, 1990, **1204**, 686.

190 D. J. Segel, A. Bachmann, J. Hofrichter, K. O. Hodgson, S. Doniach and T. Kiefhaber, *J. Mol. Biol.*, 1999, **288**, 489–499.

191 R. J. Kutta, S. J. O. Hardman, L. O. Johannissen, B. Bellina, H. L. Messiha, J. M. Ortiz-Guerrero, M. Elías-Arnanz, S. Padmanabhan, P. Barran, N. S. Scrutton and A. R. Jones, *Nat. Commun.*, 2015, **6**, 7907.

2 Characterization of native protein structure with ion mobility mass spectrometry, multiplexed fragmentation strategies and multivariant analysis

2.1 Declaration

This chapter consists of one published article:

This article has been reproduced in an unchanged format except for minor adjustments to incorporate it in to this thesis.

As the first author on this publication I performed all UVPD mass spectrometry experiments for concanavalin a and haemoglobin, I also performed the initial UVPD experiments for cytochrome c alongside AT. I therefore prepared Figures 2.1 – 2.5 alongside SI Figures 2, and 6-10. I also drafted the manuscript. AB collected the CID mass spectrometry experiments and DT performed MVA experiments. Experiment/Project planning was performed by myself, PB and LAIR, large multimeric proteins and proteins chosen for MVA analysis were chosen by myself.

This work follows on from work presented in Appendix 2 and presents a continuation with the addition of collisional activation post UVPD and MVA data processing approaches.

Supporting Information

Additional Figures are provided in the Supporting Information (SI) in section 2.9 of this thesis.

International Journal of Mass Spectrometry, DOI: 10.1016/j.ijms.2021.116588

Characterization of native protein structure with ion mobility mass spectrometry, multiplexed fragmentation strategies and multivariant analysis

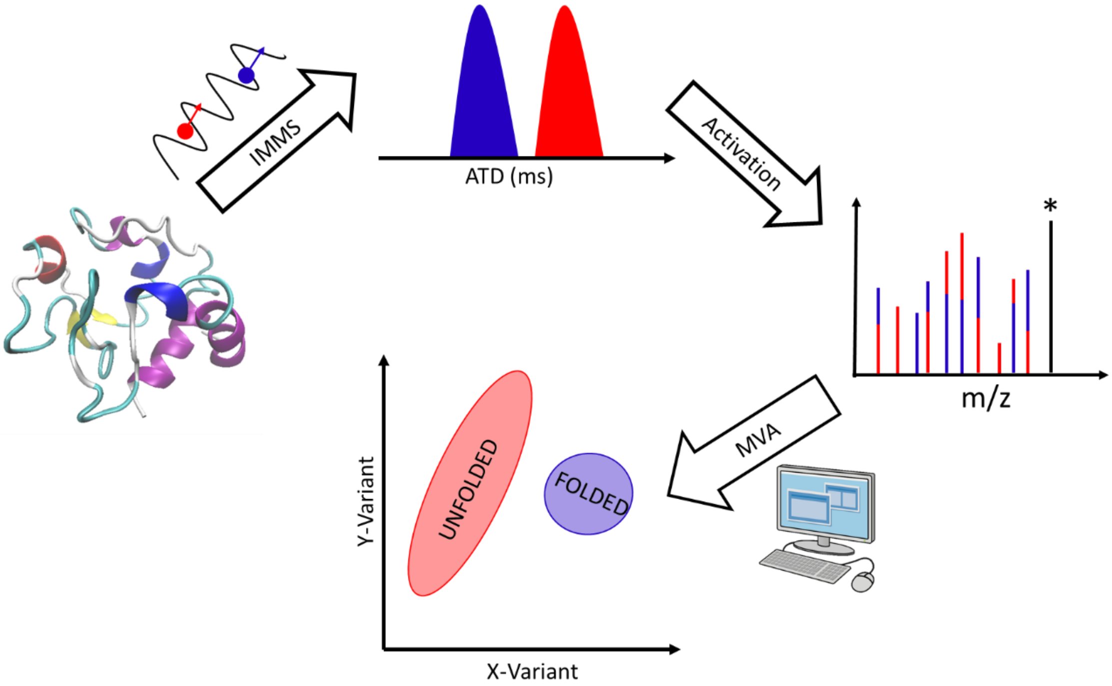
Rachelle Black[1], Alexey Barkhanskiy[1], Lennart A. I. Ramakers[1], Alina Theisen[1], Jeffery M. Brown[2], Bruno Bellina[1], Drupad Trivedi[1], Perdita E. Barran[1]

[1]Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[2] Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, United Kingdom

Corresponding author: perdita.barran@manchester.ac.uk

2.2 Abstract



Activated ion mobility measurements provide insights to the stability of tertiary and quaternary structures of proteins and adding fragmentation can delineate which part(s) of the primary sequence is disrupted from a folded structure. In this work, we use 213 nm photodissociation coupled with ion mobility mass spectrometry and collisional activation to probe the conformational landscape of model proteins. UVPD experiments are performed on proteins following in source activation and on collisionally activated photoproducts post ion mobility separation. For all observed conformations of cytochrome c, there is a significant increase in the fragmentation yield with collisional activation post mobility. Similar strategies are deployed with the multimeric proteins, concanavalin a, and haemoglobin. For these complexes' CID results in 'classical' asymmetric charge distribution in subunit products, which when preceded by UV irradiation, yields fragments from within subunits. Combining these strategies provides complex multidimensional datasets, rich in information, which here we explore with multivariate analysis (MVA). This approach readily determines differences in UVPD and CID fragmentation patterns as a function of conformation and reveals diagnostic information about the precursor native structure without the need to assign backbone cleavage sites.

2.3 Introduction

Ion mobility mass spectrometry (IMMS) is widely used to investigate protein structure and dynamics. Native mass spectrometry (MS), performed under gentle ionizing conditions and using salty aqueous solutions, provides information in the form of charge state distributions, and ion mobility contributes size, charge, and shape separation. Together these data map the conformational landscape a gas-phase protein occupies under given experimental conditions.[1–4] Drift times can be converted into rotationally averaged collision cross-sections, a buffer-gas dependent global size measurement which in turn can be compared to values obtained computationally from crystal structures or NMR experiments.[5] A large body of work has advanced our understanding of how solution structures and gas-phase structure might relate, and it is generally accepted that at least some aspects of the native fold and topology can be preserved upon transfer into the gas-phase.[5–13]

Several fragmentation strategies have been developed that are complementary to IMMS, informing on structure, sequence and potential interactions.[10,14–16] Collision-induced dissociation, the most widely available and commonly used technique, involves slow heating of the molecule via collisions with a neutral buffer gas until a dissociation threshold is reached, cleaving the most labile bonds first.[17] Due to the statistical redistribution of energy before dissociation, it is difficult to obtain topological information as proteins will unfold before fragmentation occurs.

Over recent years, ultraviolet photodissociation (UVPD) has become more widely available for the analysis of biomolecules including lipids and sugars to proteins and non-covalent complexes.[18–22] While different wavelengths enable different experimental strategies, most ‘top-down’ experiments are carried out at 213 or 193 nm. For proteins sprayed from denaturing solutions this can provide good sequence coverage throughout the entire protein.[23,24] UVPD is sensitive to the rigidity of protein secondary structural elements and to proline isomerization, inviting exploitation of UV photons for structural characterization.[25–27] At lower UV wavelength <220 nm, deposition of a single photon can be sufficient to cause electronic excitation and dissociation of protein molecules as the amide backbone becomes an efficient chromophore. However, the fragmentation observed cannot be accounted for by direct dissociation only and it has therefore been proposed that a mix of direct dissociation and dissociation via internal conversion occurs.[28]

Fragments from peptides and proteins via CID or UVPD often arise due to backbone cleavages and therefore conventionally assigned using the Roepstorff and Fohlman nomenclature.[29] In many workflows fragments are usually analysed either *via de novo* sequencing or comparing to a database of known proteins and their fragmentation patterns but restricted to backbone cleavages that contain either the N and C termini.

These approaches are well utilised throughout academia and industry, although it has been predicted that they ignore up to 90% of the fragment ion intensity, which can arise from several sources including side-chain losses, the presence of PTMs including non-canonical, internal fragments and adduction.[30] When applied to top-down sequencing and especially that for proteins in a native MS workflow, these exceptions become more prevalent, often necessitating manual assignment.[31] Further to this, no automated assignment workflows take into account the conformation of the precursor ion. This is despite widespread evidence that the fragment spectra, and in particular the intensity of the fragment ions can be diagnostic of secondary and tertiary structure.[10,32–34]

It is well known that the charge state distribution of a protein can vary according to the solvated conformation[35] and also IMMS measurements have shown that a single precursor ion can adopt numerous conformations even within a single charge state.[3,36] Fragmentation approaches tend to focus only on understanding the primary sequence of the protein and may go on to relate this to a 3D structure determined from X-ray crystallography,[37] however, few approaches determine how the fragmentation yield alters depending on the original conformation or charge state. As previously reported, 213 nm UVPD can probe protein conformation, yielding different fragment spectra depending on the initial conformations according to IMMS measurements.[38] The charge state of an ion typically results in a conformational difference when monitored *via* IMMS,[39] when these different charge states are fragmented *via* CID there is an observable difference.[40]

Although CID is the most common approach for inducing fragmentation, compaction and asymmetric fragmentation of subunits,[41,42] it can limit the insights to multimeric proteins and alternative fragmentation methods may have more merit for structural studies.[20,43–46]

Non-covalently bound fragments in ECD show as a reduced molecular ion and can, therefore, be readily distinguished from the precursor by m/z alone [12]; in UVPD however, there may be m/z non-covalently bound to the precursor ion and therefore not easily identifiable in the mass spectrum. While no studies known to us have yet investigated whether these ‘sticky’ fragments exist directly, Halim *et al.* used 213 nm UVPD on ubiquitin, either followed by or coincident with IR irradiation. For the +13 precursor ion, they reported an improvement in sequence coverage, although the yield of a-type fragments did not increase, indicating that UVPD did not result in ‘masked cleavage’ due to fragments remaining bound by non-covalent interactions.[28,47] For Ubiquitin, the 13+ ion has been shown to exist in a fully extended conformation, and likely does not possess the tight network of non-covalent bonds present in compact, native-like structures and therefore these types of fragments should not be expected.[48,49]

Principal component analysis (PCA) is commonly used as a data mining tool and forms the basis for multivariate data analysis. The main aim of such a method is to extract statistically relevant features from rich multidimensional datasets.[50] It is not limited by analytical technique and is widely used in metabolomics, however for proteomics, perhaps due to the success of algorithms that only consider backbone cleavages, it is rarely used. This is even though a significant portion of the mass spectra from top-down and bottom-up proteomics remains ‘dark’. PatternLab 4.0 has started to solve this problem; the software clusters similar proteomic profiles together and with a scoring system.[51] Picotti and Reiter also recently introduced ‘chemoproteomics’ wherein they utilise machine learning to discern features indicative of drug binding and score them to identify protein targets for small molecules.[52]

PatternLab 4.0 has been extended with the DiagnoProt module, which generates a PCA plot for each MS analysis. This also utilises machine learning to assess the spectral-peak intensities of tandem mass spectrometry data that originates from HCD on precursor ion.[53] DiagnoProt can provide a quick view as to how similar two biological conditions are to one another via clustering of the fragmentation yield.[54,55] These MS PCA workflows are independent of precursor conformations although PCA has been utilised for monitoring protein folding dynamics albeit with molecular dynamics and NMR.[50]

In this work, we show the utility of applying multiple activation strategies to probe the structure of proteins. Since such datasets are highly complex, we explore the use of multivariate analysis (MVA) for such analyses with model investigations that reveal that the fragment spectra robustly identify the difference between the ions from adjacent charge states and also between conformations adopted by a single charge state.

2.4 Experimental section

2.4.1 Materials

Methanol was obtained from Sigma Aldrich (UK) with a purity of >99.9%. Ammonium acetate was purchased from Fisher Scientific (Loughborough, UK). Ultrapure water was produced by a Milli-Q Advantage A10 ultrapure water filtration system (Merck Millipore, Darmstadt, Germany). Bovine ubiquitin and bovine cytochrome c were purchased from Sigma Aldrich (UK) as lyophilized powders with purities of ≥98% and 95% respectively. The multimeric proteins, jack bean concanavalin A and human haemoglobin were purchased from Sigma Aldrich (UK) and were dissolved in 200 mM ammonium acetate to a final concentration of 10 µM for UVPD and CID experiments. Ubiquitin and cytochrome c were dissolved in 200 mM ammonium acetate and diluted to a final protein concentration of 10 µM for UVPD and CID experiments. Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US) were used to desalt the Haemoglobin.

2.4.2 UVPD-IM-CID experiments

The experimental setup for UVPD-IMMS has been described in detail previously,[38] a schematic of the instrument can be found in SI Figure 1. Briefly, a laser beam is introduced into a Synapt G2-S through a CaF_2 window in the source block and directed onto the same pathway as the ion beam. Mass-selected ions are accumulated in the trap cell region of the instrument before ion mobility separation using a series of DC potential gates;[56] a mechanical shutter is then opened allowing the laser beam to interact with the trapped ions for a specified amount of time and all photo products are subsequently released into the IMS cell for ion mobility separation. Optionally, a collision voltage can be applied in the transfer cell post-IM to activate the mobility-separated photoproducts.

2.4.3 UVPD workflow

All samples were ionized using nano-electrospray (nESI) in positive ion mode. Monomeric proteins were sprayed from borosilicate glass capillaries (World Precision

Instruments, Stevenage, UK) pulled in-house on a Flaming/Brown P-1000 micropipette puller (Sutter Instrument Company, Novato, CA, US) into which a platinum wire was inserted to apply a voltage to the solution. Multimeric proteins were sprayed from quartz glass capillaries on a Sutter Micropipette pulled (P-2000; Sutter Instrument Co.).

All source voltages were tuned for gentle ionization with a typical capillary voltage of 1.1-1.3 kV. Source temperature was altered for the species investigated, for monomeric proteins a source temperature of 40 °C for the native conditions, additionally, the trap bias setting was reduced to 30 V – 35 V to minimise protein activation. For multimeric the source temperature remained at 40 ˚C and trap gas was increased to improve the extent of multimeric species observed.

After ionization, the charge state of interest was m/z selected in the quadrupole and accumulated in the trap cell for 2 seconds to achieve an ion count of ~ 2e3. Monomeric protein ions were then photoactivated for 2 seconds by a 213 nm laser (CryLas, UK) operating at 1 kHz with an average pulse energy of ~1 µJ. Photoproducts were then released into the IMS cell for mobility separation. Post-IM, ions were optionally activated by increasing the transfer collision energy (CE) voltage before proceeding to the ToF. Multimeric proteins were not trapped before irradiation, ubiquitin data shows identical ATD for fragments produced via trapping and no trapping protocols (SI Figure 2). The multimeric protein was irradiated by a 213 nm laser (CryLas, Germany) which operates at 1 kHz and 25 Hz and delivers 213 nm photons with an average pulse energy of 27 µJ. Multimeric proteins were investigated under a cone voltage of 10 V to yield the most folded tetrameric structure. The fragmentation difference was monitored over the course a series of different activation methods.

UVPD-IM-CID was repeated at different cone voltages ranging from 10 V to 150 V for each monomeric protein. Experiments were performed in duplicate to confirm any observed trends.

Fragmentation yield was calculated according to the relationship:

$$Normalised\ fragmentation\ yield\ = \frac{\sum fragment\ ions}{\sum fragment\ ions + \ precursor}$$

Data was analysed using MassLynx v4.1 (Waters Corporation, US), OriginPro 9.1 (OriginLab Corporation, US) and Microsoft Excel 2010 (Microsoft, US).

UVPD-MVA experiments were repeated at different cone voltages ranging from 10 V to 120 V to yield the folded and extended conformation. To obtain reliable results from the MVA analysis, the mass spectrum was calibrated using 2 µg/µL NaI and CsI made up in 1:1 Water:Isopropanol, these were run on the instrument before the UVPD analysis. The calibration graph is shown in SI Figure 3. These were then processed via an MVA workflow.

2.4.4 CID workflow

All samples were ionized using nESI in positive ion mode. These were sprayed from quartz glass capillaries on a Sutter Micropipette pulled (P-2000; Sutter Instrument Co.) into which a platinum wire was inserted to apply a voltage to the solution.

All samples were infused directly into a commercial non-modified Waters Synapt G2-S mass spectrometer *via* nESI, all settings were maintained to be as non-activating and soft as possible. The mass spectrometer was calibrated with a NaI solution at 2 µg/µL using an internal fluidics system over the range of m/z 50-2000 and threshold value of 1 ppm. All source voltages were tuned to maintain native-like conformations, the capillary voltage was maintained at 1.1-1.3 kV with a source temperature of 60 °C and the sampling cone at 40 V. Trap gas was maintained at 2 mL/min for both protein samples. The charge state of interest was selected, for ubiquitin, this was the 5+ and 6+ ions, these were collisionally activated at 60 V and 50 V respectively. For concanavalin a, dimer and tetramer ions were selected and collisionally activated at 60 V and 80 V respectively.

2.4.5 Multivariate analysis workflow

Raw data were processed using R[57] and figures were produced using the package ggplot2.[58] Data points were calculated by combining scans and creating a single spectrum for each injection. To compare masses across different replicates, masses were binned to two decimal accuracies and combined into a single matrix. The final matrix was *m* x *n* structure where each row was a binned *m/z* and each column was corresponding ion count for each replicate. Data were analysed in R using MetaboAnalystR package. Features that had a constant or single value across samples were removed and remaining missing values were imputed in the remainder of data. If there were more than 50% missing values these features were removed and for the remaining data, missing values were replaced by one fifth of the minimum positive value for each feature. The features were normalised by total ion count for each sample and data was then mean centred prior to MVA.

PCA was carried out with this new normalised data and scores plots were generated (SI Figure 12 and SI Figure 14). Additionally, univariate analysis consisting of a t-test and fold change calculations was carried out. All the significant ($p<0.05$) *m/z* values with fold change greater than 1.5 or -1.5, along with their corresponding assignments are listed in SI Table 4 and SI Table 5. PCA is a multivariate analysis (MVA) technique used mainly for dimensionality reduction and also for clustering. The components PC1 and PC2 are new axes on which data is plotted such that PC1 represents the biggest variation within the data and PC2 shows the second biggest variation within the data. For data that are influenced by many unknown parameters it is often impossible to label what PC1 and PC2 are in terms of experimental parameters.[59]

2.5 Results and discussion

2.5.1 Collisional activation post UVPD to probe protein restructuring.

Cytochrome c

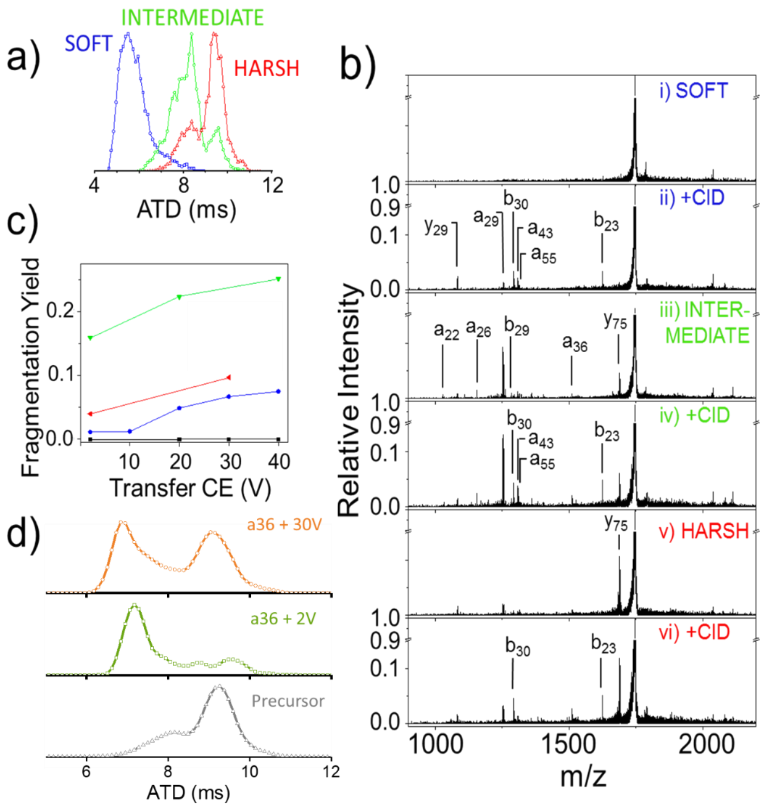


Figure 2.1: UVPD-IM-CID experiments performed on three conformational arrangements of [M+7H]$^{7+}$ of Cytochrome C. a) Arrival time distribution at soft (cone 10 V), intermediate (cone 85 V) and harsh (cone 120 V) source conditions. b) i) UVPD spectrum at soft conditions without and ii) with additional transfer CE of 30 V. iii) UVPD spectrum at intermediate conditions iv) with additional transfer CE of 40 V. v) UVPD spectrum at harsh conditions vi) with additional transfer CE of 30 V. c) Fragmentation yield of a-type fragments as a function of transfer collision energy in soft (blue), intermediate (green), harsh (red) conditions and the CID-only control (black). d) Arrival time distribution of fragment a36 from harsh conditions compared between transfer CE 2 V (green) and transfer CE 30 V (orange), the precursor is shown in grey.

Cytochrome c presents predominantly as ions [M+6H]$^{6+}$ and [M+7H]$^{7+}$ when sprayed from 200 mM ammonium acetate with most intensity in the latter. [M+7H]$^{7+}$ was mass selected for analysis and arrival time distributions were recorded for a range of cone voltages (Figure 2.1a). Under soft source conditions, [M+7H]$^{7+}$ exists as a single, compact conformer. Increasing the cone voltage to 85 V (‘intermediate’) resulted in a shift away from the compact conformer to more extended forms. Further increase of the cone voltage to 120 V (‘harsh’) moves the intensity to the most extended form. Each cone voltage tells us about the conformation and structure of the protein.

At each discrete cone voltage, $[M+7H]^{7+}$ was subjected to UVPD producing a, b, c and y-type fragments, a full list of all fragments observed can be found in SI Table 1. The main observation from these differing conditions is the changed intensities of certain fragments as a function of activation (Figure 2.1b). When collisional activation was applied post ion mobility to the compact conformer photoproducts an increase in the intensity of all fragments is observed. This is concomitant with the appearance of additional a- and b-type fragments from cleavage within the primary sequence from Phe36 to Met65. This region in the crystal structure comprises a loop and two short alpha-helices, [60] which are located deep in the protein core. It follows that collisional activation causes a partial unravelling of these secondary structural elements, breaking stabilising non-covalent interactions and allowing the release of cleaved fragments.

When the intermediate conformer was subjected to UVPD (prior to any activation), we observe an increase in the summed fragment ion intensity as well as in the number of fragments (Figure 2.1biii). UVPD-IM-CID further increases the fragment yield, and releases ions containing residues Lys22 to Lys55. These results indicate that the helix between Asp50 and Gly56 is disrupted before the helix containing Met65. Increasing the cone voltage to harsh conditions decreases the overall fragment yield (Figure 2.1bv). When collisional activation was applied post-IM, a b29 ion is observed, suggesting again that noncovalent interactions remain even under harsh source conditions and extra collisional energy is required to release this UVPD fragment. Interestingly, under soft, intermediate and harsh conditions fragments that contain the C- and N-terminal stretches are observed (Asp2 to Cys14 and Lys87 to Thr103 respectively). No shorter fragments that contain these regions are ever seen even after additional post-IM activation, supporting prior observations that these are the most stable structural elements[61].

The overall fragmentation yield increases with increasing transfer CE (SI Figure 4) for all source conditions, as does the yield of only a-type fragments (Figure 2.1c). Since the ATD of the intermediate and the harsh source conditions both contain the same conformer, we used ion mobility to determine whether the increase in fragmentation yield (harsh) is a result of the remaining intermediate family or originates from the most extended one. As collisional activation occurs post ion mobility, we can use the ion mobility of fragments to determine where they dissociated and whether this was due to the initial UVPD activation or due to the collisional activation allowing the release of the fragment. The a36 ion has a distinct ATD with little overlap with the ATD of the intact protein (Figure 2.1d lower and middle panel). Upon additional collisional activation post mobility, there is an increase in the fragmentation yield of this ion and its ATD now contains a distribution that resembles that of the intact protein (Figure 2.1d lower and top panel). This supports the assertion that increasing the transfer CE allows the a36 ion, caused by UVPD, to depart the precursor, but of course it retains the same ATD, since it traversed the drift cell prior to this departure, the difference in ion intensity suggests that most is retained until the CE activation. This is observed for other a-ions (SI Figure 5).

Overall, collisional activation post IM increases the fragmentation yield presumably by disrupting non-covalent interactions, that keep cleaved regions together. Ubiquitin shows similar behaviour (SI Figure 6).

2.5.2 UVPD analysis of large multimeric proteins

Concanavalin A

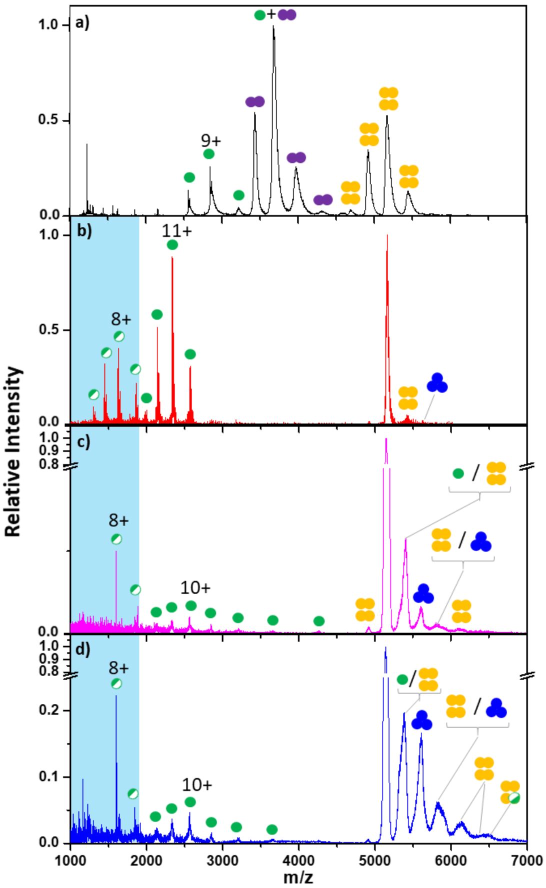


Figure 2.2: a) The full MS spectrum of the Concanavalin A (black), labelled to show the subunits along with the final tetrameric structure. b) CID spectra (red), at a collisional energy of 1.6 keV, of the [4M+20H]$^{20+}$ tetramer in the trap region of the mass spectrometer. UVPD spectra at a repetition rate of c) 250 Hz (pink) and d) 1 kHz (blue), of the [4M+20H]$^{20+}$ tetramer in the trap region of the mass spectrometer. Monomer sequence ions as well as monomer, dimer, trimer and tetramer species indicated by half-filled green, green, purple, blue and yellow circles, respectively.

Concanavalin A was subjected to similar experiments as described above for cytochrome c (Figure 2.2, and SI Figures 7 and 8). Collisional activation of the tetramer (Figure 2.2b) produced significant monomer, as previously described. Alongside these monomeric species, CID of [4M+20H]$^{20+}$ yields low mass fragments (1000-1800 m/z) potentially from ejected monomer units (blue, Figure 2.2b). Product ions with higher m/z than the $[4M+20H]^{20+}$ precursor are also observed, and are assigned as charge stripped species as previously described.[62] Collisional activation was also performed on the dimeric form yielding mainly monomer species over a wide charge state distribution (CSD), some charge stripping and low intensity backbone cleaved fragments (SI Figure 7).

The $[4M+20H]^{20+}$ ion was isolated and fragmented *via* UVPD at 250 Hz (Figure 2.2c) and 1 kHz (Figure 2.2d) to monitor the effect of the repetition rate on the fragmentation spectrum. Similar to CID, the UVPD fragmentation spectra shows monomer, trimer, charge stripped tetramer and a range of backbone cleavage ions. There are some evident differences between CID and UVPD spectra, the distributions of ejected monomers in UVPD whilst centred on $[M+10H]^{10+}$ contain lower charge states than those generated by CID (Figure 2.2c). The abundance of these low charge species is greater when UVPD is carried out at 1 kHz (Figure 2.2d) and is also accompanied by more backbone cleavage ions. The higher photon flux is acting to promptly dissociate the monomer, akin to SID, and also to generate more fragments than reported with CID.[20,62,63] Whilst the low signal to noise and the large number of possible fragments which may include side chain losses, internal fragments and even non-covalent interactions from two subunits, means these ions are hard to assign. Here we note dominant fragments from the 7+ → 10+ charge states of the $\{a_{120}\text{-}NH_3\}$, a_{120}, b_{120}, b_{121}, $\{a_{122}\text{-}NH_3\}$ and $\{b_{122}\text{-}NH_3\}$ ions (SI Figure 8). UVPD also produces charged stripped tetrameric and trimeric species, and the photo-fragmentation of the $[4M+20H]^{20+}$ concanavalin A ion yields a 5+ monomer subunit, not seen with CID. From this UVPD data, we can conclude that single monomers can be dissociated from the tetramer even with a lower photon flux, and the CSD infers that they are less unfolded than with CID, and may retain some of the native structure, although we do not see dimer.

Similar findings are obtained with haemoglobin (SI Figure 9). Here CID leads to the production of α and β globin monomer units, a portion of which are thought to be formed by the secondary ejection of the heme group after monomer ejection as well as a range of holo- and apo-trimeric species (SI Figure 9b). UVPD also leads to formation of monomer and trimer subunits along with a greater amount of primary sequence ions (SI Figure 9c and d) confirming that large multicomponent protein complexes are amenable to UVPD within this instrument.

As with many multimeric protein complexes, the dominant products of collisional activation of tetrameric concanavalin and haemoglobin are highly charged monomeric species, and low charged trimer, where both undergo substantial structural rearrangement, as shown by IM data and the charge of the fragments prior to dissociation.[20,64,65] By contrast, UVPD facilitates the observation of fragments from the intact tetramer (Figure 2.2) which can provide more information about the native fold and how it is disrupted. This approach is perhaps most informative on combining UVPD with collisional activation as shown here for cytochrome c (Figure 2.1), but here even for a simple monomeric protein, the wealth of multidimensional data produced, prevents a full assessment of this related to structure. In order to guide interpretation of these multidimensional datasets, we decided to explore the use of MVA, with an initial focus on small monomeric proteins.

2.5.3 MVA strategies applied to fragmentation experiments in native mass spectrometry.

2.5.3.1 Ubiquitin following CID comparing native charge states

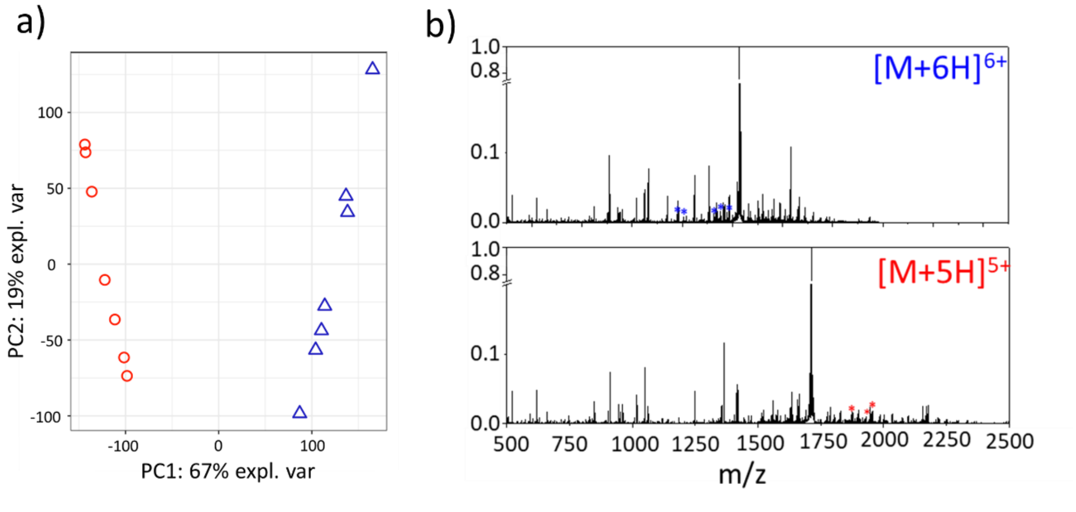


Figure 2.3: nESI-MS/MS-MVA of ubiquitin ions [M+6H]$^{6+}$ and [M+5H]$^{5+}$. a) Score plot for the tandem MS of the 5+ (red) and 6+ (blue) charge states following collisional activation. b) CID spectra of [M+6H]$^{6+}$ (top) and [M+5H]$^{5+}$(bottom) with discriminant ions determined with MVA labelled on the spectra.

To develop a method to mine the complex data from native top-down experiments, an MVA workflow was trained on Ubiquitin. Experiments were performed in 7 replicates.

MVA analysis was performed on these precursor ions to monitor for significant changes in the fragmentation pattern between the two charge states. CID was also performed at concentration 5 µM and 20 µM (SI Figure 10) and showed no significant difference in the fragmentation pattern that was produced.

The activated charge states yield different fragmentation patterns when exposed to collisional activation, this is evident on manual comparison of the MS (SI Figure 11 and SI Table 3) and *via* MVA when a sPLS-DA plot is produced (Figure 2.3a), yielding distinct separation between the two charge states. The sPLS-DA plot shows good separation and clustering for each charge state, with a significant variance within each charge state, attributable to the slight differences between the repeats perhaps due to conformational differences. To account for any spray changes and fluctuations within the total ion count (TIC) the MVA data was also normalised to the sum of all the intensities (SI Figure 12), again the separation between the two charge states remains significant. This separation implies that even for adjacent charge states of a protein ion produced under soft ionisation conditions, the precursor charge state restructures differently which in turn influences the fragmentation behaviour.

An advantage to an MVA method is the significance of any spectral feature can be determined, upon performing a t-test the fragments with greater significance ($p<0.05$) and a fold change greater than 1.5 or -1.5 (SI Table 4) have been chosen and identified (Figure 2.3b). We find eight significant fragments that are highly dependable on the precursor protein charge, (five for $[M+6H]^{6+}$, and three for $[M+5H]^{5+}$ Figure 2.3b and SI Table 4). These infer that specific fragments can be linked to precursor ion charge states *a priori*. Unlike the fully annotated MS spectrum, we highlight fragments here where there are statistically significant different intensities between the data from $[M+6H]^{6+}$ in comparison to $[M+5H]^{5+}$. Although some of these differences can be observed manually, some fragment ions hide within the “dark” regions of the MS that may not have been taken into consideration. Whilst the example of ubiquitin is trivial, this approach could find utility in the top-down measurement of more ‘exotic’ proteins, single point mutants, or *apo*- and *holo*-protein complexes.

2.5.3.2 Cytochrome c following in source activation and UVPD

To further explore the rationale for an MVA strategy we now apply this to a 213 nm UVPD fragmentation workflow for cytochrome c, with a focus on variation between conformations (SI Figure 13) found for a single charge state.

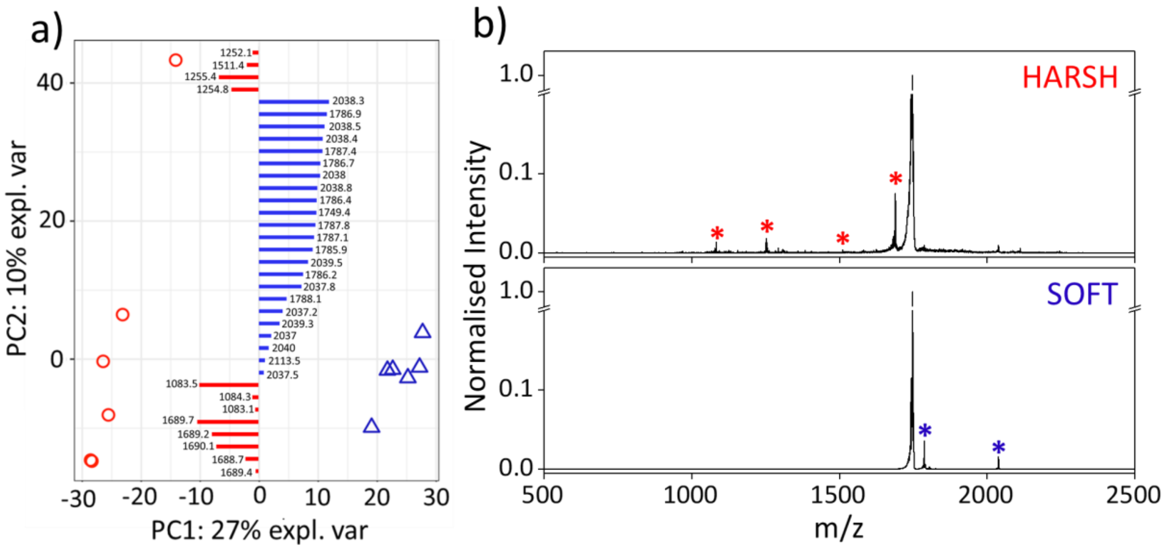


Figure 2.4: UVPD-IM-MVA of two conformations of [M+7H]$^{7+}$ Cytochrome c. a) Score plot for the soft (blue) and harsh (red) conditions exposed to 213 nm UVPD with MVA determined features/fragmentation peaks overlayed showing condition where they are more prevalent and significant difference in the fragmentation yield between soft and harsh, b) UVPD spectrum at soft (cone 10 V) and harsh (cone 120 V) conditions without any additional transfer CE, labelled are the most prominent changes according to the MVA results.

Following this soft and harsh in-source activation the UVPD fragmentation yields were processed *via* the MVA protocol which involved monitoring the fragmentation patterns and then grouping the replicates. To overcome discrepancies caused by spray quality and nESI tips, the MVA was performed on a repeated MS to minimise these fluctuations between runs. The MS applied to the MVA was a summation of the entire 20 minutes' acquisition under different in-source activation, after MVA analysis was performed all identified features were confirmed to be "real" before continuing. The two in-source activation conditions remain remarkably separate from one another in the resulting score plot following this approach (Figure 2.4a). Data obtained under soft conditions clusters closely between experimental replicates, indicating little fluctuation in the precursor ion structure or little fluctuation due to the background noise. Conversely, the data obtained following harsher in-source activation shows significant variance between the repeats, this may be due to the inherent structural flexibility under harsher conditions or due to the TIC variation owing to spray stability or slightly different capillary fluctuations. To account for the TIC variations, the MVA data was normalised to account for this (SI Figure 14), whilst it appears that the harsh source conditions result in a higher population of noise, this does not affect the separation of the two conditions. Therefore, the use of MVA to identify the conformation of the protein remains meaningful however the difference between the distribution between each data point is reduced and therefore we may be limited in the assignment of features to individual conformers following activation.

This separation in the sPLS-DA plot occurs because of the differing UVPD fragments from conformers that are produced by different activation conditions. This paves the way to use MVA to highlight features of characteristic *m/z* values (Figure 2.4a) that can be used to monitor the folded state of a protein. The fragments chosen here represent those that have greater significance ($p<0.05$) and a fold change greater than 1.5 or -1.5 (SI Table 5) the features first identified do include isotopes and different charge states of the same ion (Figure 2.4b) although this redundancy can be managed manually and in the future computationally. MVA applied to these top-down fragmentation data sets from native protein mass spectra finds features that are diagnostic of protein conformation. The identified features can also be compared to assignment ions (SI Table 1) to confirm their location in the protein fold.

This MVA procedure enables a data directed approach to assess the conformational landscape utilising the UVPD fragmentation information. The fragments that show the greatest change between the two conditions can act as the identifier features to indicate when and where structural perturbation of a protein has occurred. The fragments identified *via* MVA that cannot be assigned are just as important and provide new reproducible features that signify the conformational state, here it is less about the need to assign the entire sequence and more about utilising all the information available in the “dark regions” of the MS to gauge about the protein conformation from the fragmentation yield. MVA shows its clear benefits of monitoring the fragmentation yield and relating this to the conformation *via* its position on the sPLS-DA plot.

2.5.3.3 Combining MVA and standard assignment methods

Multivariate approaches are widely utilised in a range of analytical and physical techniques,[66–69] to allow data to be extracted from congested spectra containing features which are hard to assign. Combining MVA approaches with such techniques allows tools to be developed that are capable of extracting useful information from complicated spectra, including important structural aspects of DNA binding interactions,[66] development of diagnostic tests for endemic diseases,[67] particle size during crystallisation[68] and allowing for rapid quality control to be performed.[69] The overall aim of the MVA applied here was to develop a novel tool to study the conformational state of proteins.

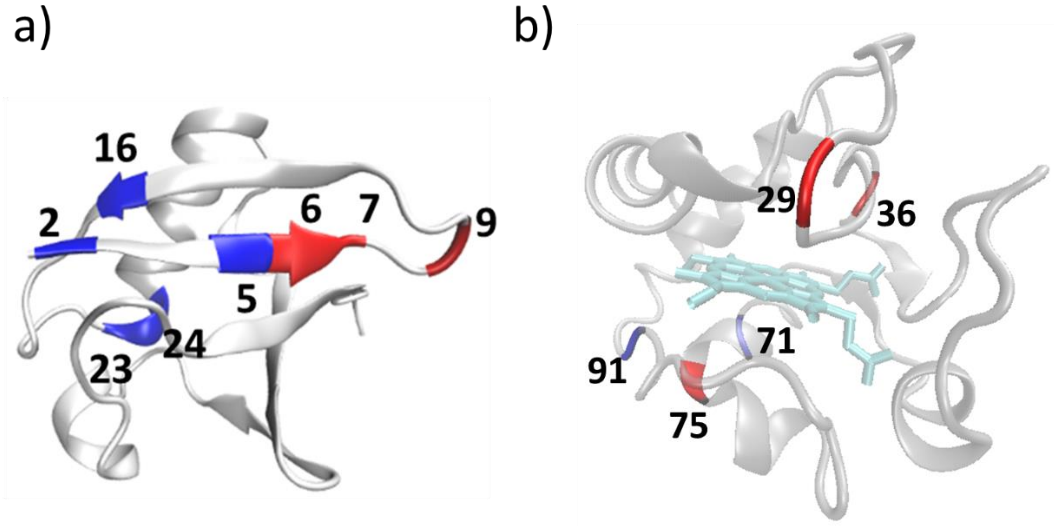


Figure 2.5: The top m/z identified through MVA are mapped onto the crystal structure of the protein, a) Ubiquitin (PDB structure: 4Z9S [70]), the regions highlighted show where significant fragments are present between the two charge states $[M+5H]^{5+}$ (red) and the $[M+6H]^{6+}$ (blue). b) Cytochrome C (PDB structure: 2B4Z [60]), the regions highlighted show where there is significant fragments between the two different conformations from soft (blue) and harsh (red) in-source activation.

Following MVA analysis, the identified features that have the highest confidence ($p<0.05$) for ubiquitin (SI Table 4) are then assigned *via* Roepstorff and Fohlman nomenclature and mapped onto the crystal structure of the protein (Figure 2.5a). For each fragment, the corresponding part of structure is highlighted and colour coded to match the initial charge state of the precursor ion. Interestingly, fragments from the higher charge state precursor originate deeper in the tertiary fold, with a significant portion coming from the β-sheets. Whilst CID is known to be a slow heating fragmentation technique that would result in unfolding before fragmentation[71] this data indicates that the higher charge state is already more disrupted. Whereas the regions that are preferentially cleaved in $[M+5H]^{5+}$ consist of the turn (Thr9) and the end of the first β-sheet (Val6 and Lys7), and not from any part of the protein with helical or β-sheet secondary structure.

MVA determined features from UVPD, were also mapped to the structure of cytochrome c (Figure 2.5b). Again, the significant features are colour coded. As expected, regions shaded in blue which are due to fragments that are found from soft conditions, are found mainly in the loop and turn regions of the protein (Pro71 and Arg91), where there are little non-covalent interactions that could prevent the dissociation. On the other hand, many of the identified features from data taken under harsh conditions arise from regions of the sequence close to, and in one case within, α-helices (Gly29, Phe36 and Ile75). As for ubiquitin, harsher activation causes unfolded conformations with fewer non-covalent interactions therefore permitting dissociation of such fragments. The identified regions are comparable to those seen previously, where we use IMMS alone to examine the effect of conformation on the 213 nm UVPD fragmentation patterns[38] illustrating the effectiveness of the MVA approach to identify structurally significant fragments *a priori*.

The proteins and structural changes investigated here are well studied and understood allowing them to act as model systems for the development of MVA as an analytical tool. Our results indicate that this approach could be applied to other less common variations such as PTMs or cofactor binding, to gain a deeper understanding of their structural impact on protein systems. It is also important to note, that following standard Roepstorff and Fohlman (and without IMS knowledge), the difference between both conditions (Figure 2.4) would have been overlooked. It then becomes important to consider all the data that has been obtained, even those in the “dark” unassignable region of the MS can give valuable information on the protein structure.

2.6 Conclusions

We have shown that multiple fragmentation strategies coupled with ion mobility mass spectrometry can provide insights to protein structure and stability for model monomeric and tetrameric proteins. Collisional activation of photoproducts post ion mobility separation releases non-covalently linked fragments even from compact (native) conformers. Multiple protein conformers co-exist in intermediate and harsh source conditions, and these strategies combined with the ability to map fragments to the ion mobility data facilitate a conformer directed data analysis approach. This can yield insights to the stability of the fold; for Cytochrome c even the most extended conformer released a-type (UVPD) fragments following UVPD-CID.

Similarly, when UVPD is applied to larger multimeric proteins, we have shown here that *via* UVPD workflows we're able to get more information on the native-fold of larger multimeric proteins *via* the observed fragments from the intact tetramer.

The CID-MVA procedure has permitted the separation of the two initial precursor charge states, both of which originate from soft ionisation conditions and believed to be "native-like", however, both undergo a different fragmentation pathway. We propose here that MVA has allowed more information to be taken from this data and has shown us that the precursor charge state does have an effect on the fragmentation of the ion.

MVA was also successfully applied to UVPD data where it has been previously shown that the precursor ion conformation dictates which fragments are observed. Here, we showed how to utilise MVA strategies on this data to yield sPLS-DA plots that show the direct difference between the two conformations utilising only the MS fragmentation data. It is hoped that in future work this will allow one to determine the point at which the conformation of a protein has been disrupted without the requirement of IMS.

We have shown here the benefits of utilising multiple activation methods to gain structural and conformational information on a set of model proteins. The purpose of this paper is to demonstrate that MVA can be used to determine, if a given conformation is identical or different to another, without the extra effort of fragment assignment based on limited predicted fragments. This allows for the "dark" regions of the MS to be accounted for even when assignment may not previously be possible. Given the sensitivity of modern mass spectrometers, such an approach could readily be applied as a screening tool, for example, to map the structural features that alter due to PTMs, cofactor binding, or in the directed evolution of enzymes.

2.7 Acknowledgements

This work was supported by BBSRC grants BB/L002655/1, BB/L016486/1 and BB/M01108/1 as well as studentship awards to AT, RB with additional financial support from Waters Corp who have also part supported AB and PB.

PB and LAIR are grateful to the MS SPIDOC project funded by the European Union's Horizon 2020 FET-OPEN research and innovation programme (Grant agreement No. 801406) for support of ongoing photo activation IMMS activities.

AB acknowledges the EPSRC, BBSRC and AstraZeneca plc. for funding under the Prosperity Partnership EP/S005226/1.

2.8 References

1 D. E. Clemmer, R. R. Hudgins and M. F. Jarrold, *J. Am. Chem. Soc.*, 1995, **117**, 10141–10142.

2 T. Wyttenbach, G. Von Helden and M. T. Bowers, *J. Am. Chem. Soc.*, 1996, **118**, 8355–8364.

3 R. Beveridge, L. G. Migas, K. A. P. Payne, N. S. Scrutton, D. Leys and P. E. Barran, *Nat. Commun.*, 2016, **7**, 1–9.

4 R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E. Macphee and P. E. Barran, *Anal. Chem.*, 2014, **86**, 10979–10991.

5 E. Jurneczko and P. E. Barran, *Analyst*, 2011, **136**, 20–28.

6 Z. Hall, A. Politis and C. V Robinson, *Structure*, 2012, **20**, 1596–609.

7 J. A. Loo, J. X. He and W. L. Cody, *J. Am. Chem. Soc.*, 1998, **120**, 4542–4543.

8 Z. Deng, N. Thontasen, N. Malinowski, G. Rinke, L. Harnau, S. Rauschenbach and K. Kern, *Nano Lett.*, 2012, **12**, 2452–2458.

9 D. Hewitt, E. Marklund, D. J. Scott, C. V Robinson and A. J. Borysik, *J. Phys. Chem. B*, 2014, **118**, 8489–95.

10 N. C. Polfer, K. F. Haselmann, P. R. R. Langridge-smithy and P. E. Barran, *Mol.*

Phys., 2005, **103**, 1481–1489.

11 S. R. Harvey, M. Porrini, A. Konijnenberg, D. J. Clarke, R. C. Tyler, P. R. R. Langridge-Smith, C. E. MacPhee, B. F. Volkman and P. E. Barran, *J. Phys. Chem. B*, 2014, **118**, 12348–12359.

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

13 T. Wyttenbach and M. T. Bowers, *Annu. Rev. Phys. Chem.*, 2007, **58**, 511–33.

14 J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz and D. F. Hunt, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9528–33.

15 R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter and F. W. McLafferty, *Anal. Chem.*, 2000, **72**, 563–73.

16 A. S. Galhena, S. Dagan, C. M. Jones, R. L. Beardsley and V. H. Wysocki, *Anal. Chem.*, 2008, **80**, 1425–36.

17 N. L. Kelleher, *Anal. Chem.*, 2004, **76**, 196 A-203 A.

18 M. R. Robinson, J. M. Taliaferro, K. N. Dalby and J. S. Brodbelt, *J. Proteome Res.*, 2016, **15**, 2739–2748.

19 A. Racaud, R. Antoine, L. Joly, N. Mesplet, P. Dugourd and J. Lemoine, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1645–1651.

20 L. J. Morrison and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2016, **138**, 10849–10859.

21 T. Ly and R. R. Julian, *J. Am. Chem. Soc.*, 2010, **132**, 8602–8609.

22 M. A. Halim, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd and M. Girod, *J. Am. Soc. Mass Spectrom.*, 2017, 1–14.

23 J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, *J. Am. Chem. Soc.*, 2013, **135**, 12646–12651.

24 J. R. Cannon, M. B. Cammarata, S. A. Robotham, V. C. Cotham, J. B. Shaw, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher and J. S. Brodbelt, *Anal. Chem.*, 2014, **86**, 2185–2192.

25 T.-Y. Kim, S. J. Valentine, D. E. Clemmer and J. P. Reilly, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 1455–65.

26 S. Warnke, C. Baldauf, M. T. Bowers, K. Pagel and G. Von Helden, *J. Am. Chem. Soc.*, 2014, **136**, 10308–10314.

27 M. B. Cammarata and J. S. Brodbelt, *Chem. Sci.*, 2015, **6**, 1324–1333.

28 R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1823–1826.

29 P. Roepstorff and J. Fohlman, *Biol. Mass Spectrom.*, 1984, **11**, 601–601.

30 J. A. Taylor and R. S. Johnson, *Anal. Chem.*, 2001, **73**, 2594–2604.

31 A. S. Phillips, A. F. Gomes, J. M. D. Kalapothakis, J. E. Gillam, J. Gasparavicius, F. C. Gozzo, T. Kunath, C. MacPhee and P. E. Barran, *Analyst*, 2015, **140**, 3070–3081.

32 J. M. D. Kalapothakis, Y. Berezovskaya, C. G. Zampronio, P. A. Faull, P. E. Barran and H. J. Cooper, *Chem. Commun.*, 2014, **50**, 198–200.

33 B. T. Ruotolo, S. J. Hyung, P. M. Robinson, K. Giles, R. H. Bateman and C. V. Robinson, *Angew. Chemie - Int. Ed.*, 2007, **46**, 8001–8004.

34 J. T. Seffernick, S. R. Harvey, V. H. Wysocki and S. Lindert, *ACS Cent. Sci.*, 2019, **5**, 1330–1341.

35 L. Konermann, E. Ahadi, A. D. Rodriguez and S. Vahidi, *Anal. Chem.*, 2012, **85**, 2–9.

36 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, **42**, 177–185.

37 M. Schennach and K. Breuker, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1059–1067.

38 A. Theisen, R. Black, D. Corinti, J. M. Brown, B. Bellina and P. E. Barran, *J. Am. Soc. Mass Spectrom.*, 2018, **30**, 24–33.

39 M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson and B. T. Ruotolo, *Anal. Chem.*, 2010, **82**, 1–9.

40 G. E. Reid, J. Wu, P. A. Chrisman, J. M. Wells and S. A. McLuckey, *Anal. Chem.*, 2001, **73**, 3274–3281.

41 J. C. Jurchen and E. R. Williams, *J. Am. Chem. Soc.*, 2003, **125**, 2817–2826.

42 J. C. Jurchen, D. E. Garcia and E. R. Williams, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 1408–1415.

43 S. Tamara, A. Dyachenko, K. L. Fort, A. A. Makarov, R. A. Scheltema and A. J. R. Heck, *J. Am. Chem. Soc.*, 2016, **138**, 10860–10868.

44 S. N. Sipe and J. S. Brodbelt, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9265–9276.

45 M. Sharon, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 487–500.

46 I. Sinelnikov, E. N. Kitova and J. S. Klassen, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 617–631.

47 M. A. Halim, M. Girod, L. MacAleese, J. Lemoine, R. Antoine and P. Dugourd, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1435–1442.

48 J. Li, J. A. Taraszka, A. E. Counterman and D. E. Clemmer, *Int. J. Mass Spectrom.*, 1999, **185**–**187**, 37–47.

49 H. Oh, K. Breuker, S. K. Sze, Y. Ge, B. K. Carpenter and F. W. McLafferty, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15863–8.

50 G. G. Maisuradze, A. Liwo and H. A. Scheraga, *J. Mol. Biol.*, 2009, **385**, 312–329.

51 P. C. Carvalho, D. B. Lima, F. V. Leprevost, M. D. M. Santos, J. S. G. Fischer, P. F. Aquino, J. J. Moresco, J. R. Yates and V. C. Barbosa, *Nat. Protoc.*, 2016, **11**, 102–117.

52 I. Piazza, N. Beaton, R. Bruderer, T. Knobloch, C. Barbisan, L. Chandat, A. Sudau, I. Siepe, O. Rinner, N. de Souza, P. Picotti and L. Reiter, *Nat. Commun.*, 2020, **11**, 1–13.

53 D. B. Lima, A. R. F. Silva, M. Dupré, M. D. M. Santos, M. A. Clasen, L. U. Kurt, P. F. Aquino, V. C. Barbosa, P. C. Carvalho, J. Chamot-Rooke and J. Hancock, *Bioinformatics*, 2019, **35**, 3489–3490.

54 A. R. F. Silva, D. B. Lima, A. Leyva, R. Duran, C. Batthyany, P. F. Aquino, J. C. Leal, J. E. Rodriguez, G. B. Domont, M. D. M. Santos, J. Chamot-Rooke, V. C. Barbosa and P. C. Carvalho, *Bioinformatics*, 2017, **33**, 1883–1885.

55 J. M. Silva, H. H. Wippel, M. D. M. Santos, D. C. A. Verissimo, R. M. Santos, F. C. S. Nogueira, G. A. R. Passos, S. L. Sprengel, L. A. B. Borba, P. C. Carvalho

and J. de S. d. G. Fischer, *Sci. Rep.*, 2020, **10**, 1–11.

56 B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon and P. E. Barran, *Analyst*, 2014, **139**, 6348–51.

57 R. C. Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, http://www.r-project.org/, (accessed 13 August 2020).

58 H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer US, New York, 2009.

59 J. Lever, M. Krzywinski and N. Altman, *Nat. Methods*, 2017, **14**, 641–642.

60 N. Mirkin, J. Jaconcic, V. Stojanoff and A. Moreno, *Proteins*, 2008, **70**, 83–92.

61 M. M. G. Krishna, H. Maity, J. N. Rumbley and S. W. Englander, *Protein Sci.*, 2007, **16**, 1946–56.

62 M. Zhou, S. Dagan and V. H. Wysocki, *Analyst*, 2013, **138**, 1353–1362.

63 K. Pagel, S. J. Hyung, B. T. Ruotolo and C. V. Robinson, *Anal. Chem.*, 2010, **82**, 5363–5372.

64 R. S. Quintyn, M. Zhou, S. Dagan, J. Finke and V. H. Wysocki, *Int. J. Ion Mobil. Spectrom.*, 2013, **16**, 133–143.

65 Z. Hall, H. Hernández, J. A. Marsh, S. A. Teichmann and C. V. Robinson, *Structure*, 2013, **21**, 1325–1337.

66 R. Fritzsch, P. M. Donaldson, G. M. Greetham, M. Towrie, A. W. Parker, M. J. Baker and N. T. Hunt, *Anal. Chem.*, 2018, **90**, 2732–2740.

67 P. Heraud, P. Chatchawal, M. Wongwattanakul, P. Tippayawat, C. Doerig, P. Jearanaikoon, D. Perez-Guaita and B. R. Wood, *Malar. J.*, 2019, **18**, 1–11.

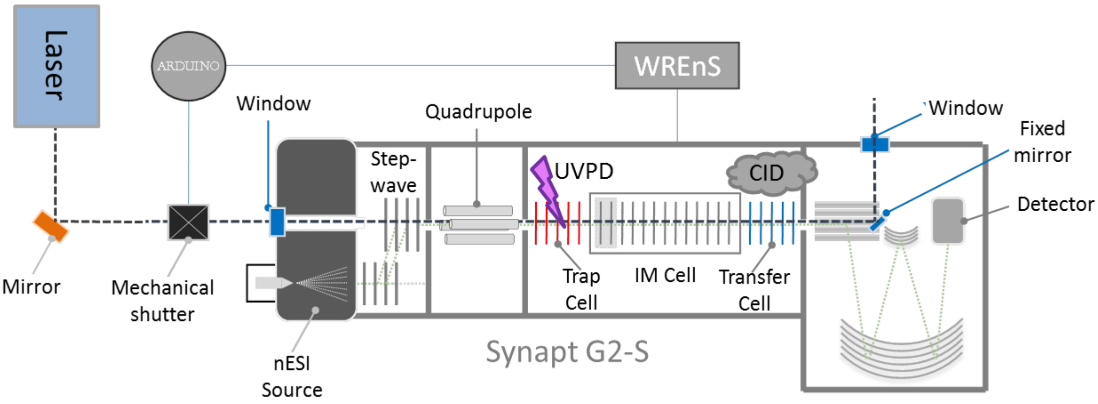
68 C. Ferreira, J. Cardona, O. Agimelen, C. Tachtatzis, I. Andonovic, J. Sefcik and Y. C. Chen, *Powder Technol.*, 2020, **376**, 1–11.

69 A. Alberti, T. P. Machado dos Santos, A. A. Ferreira Zielinski, C. M. Eleutério dos Santos, C. M. Braga, I. M. Demiate and A. Nogueira, *LWT - Food Sci. Technol.*, 2016, **65**, 436–443.

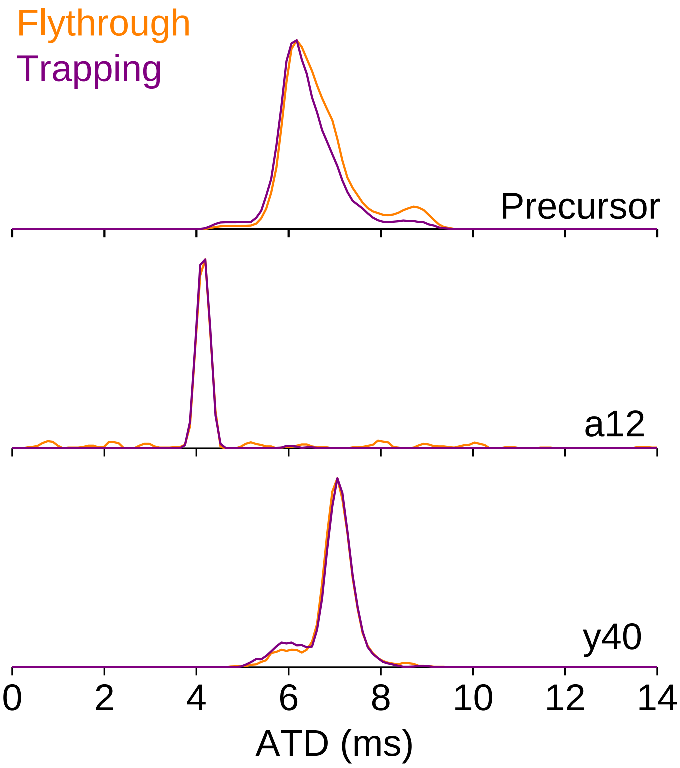
70 O. Levin-Kravets, N. Shohat and G. Prag, *Biochemistry*, 2015, **54**, 4704–4710.

71 L. G. Migas, A. P. France, B. Bellina and P. E. Barran, *Int. J. Mass Spectrom.*, 2018, **427**, 20–28.

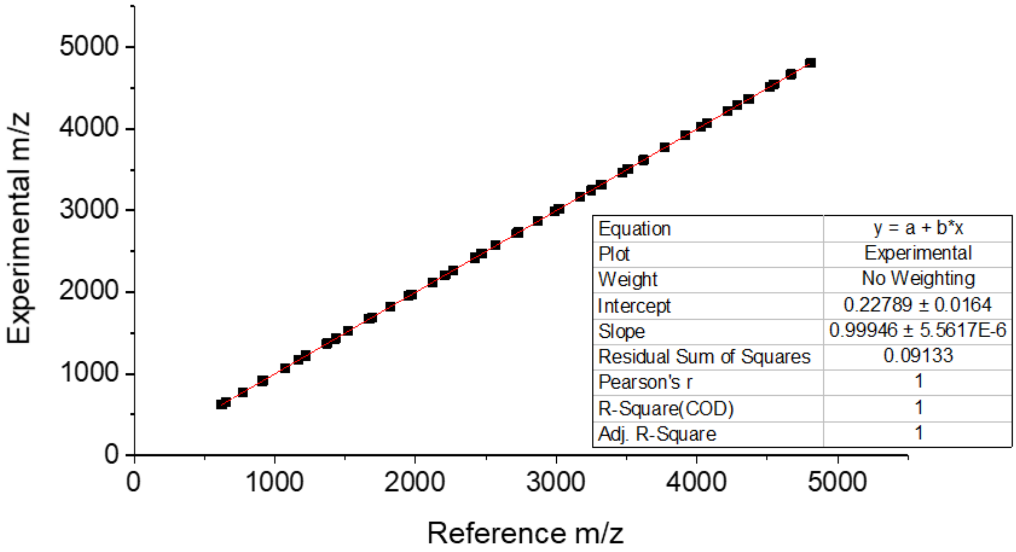
2.9 Supporting information



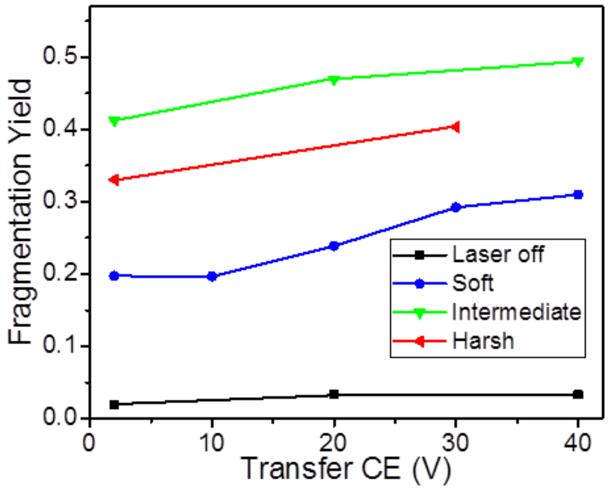
SI Figure 1: Schematic of the Photo-Synapt G2-S instrument. For further details on this customised instrument see refs [56,71]



SI Figure 2: Selected ATDs for Ubiquitin 6+ 213 nm UVPD in both flythrough (orange top) and trapping mode (pink top), shown are the precursor ion, a12 fragment and y40 fragment ion.



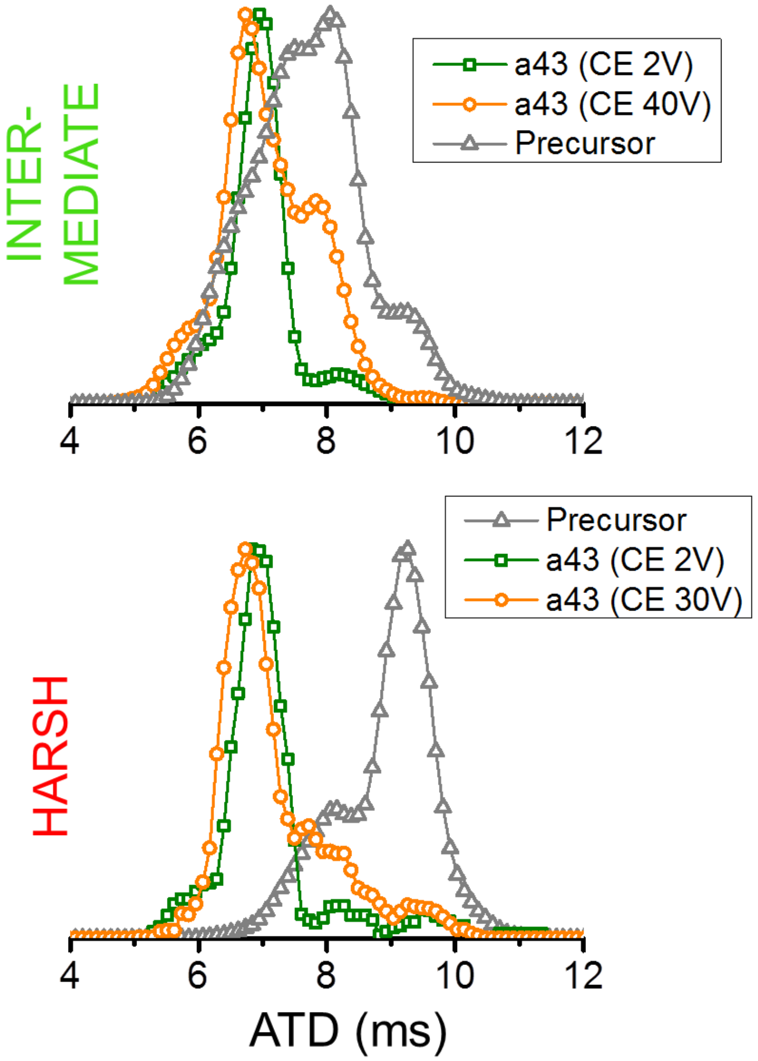
SI Figure 3: NaI and CsI calibration graph that was applied to the mass spectrometry measurements utilised for the UVPD-MVA analysis.



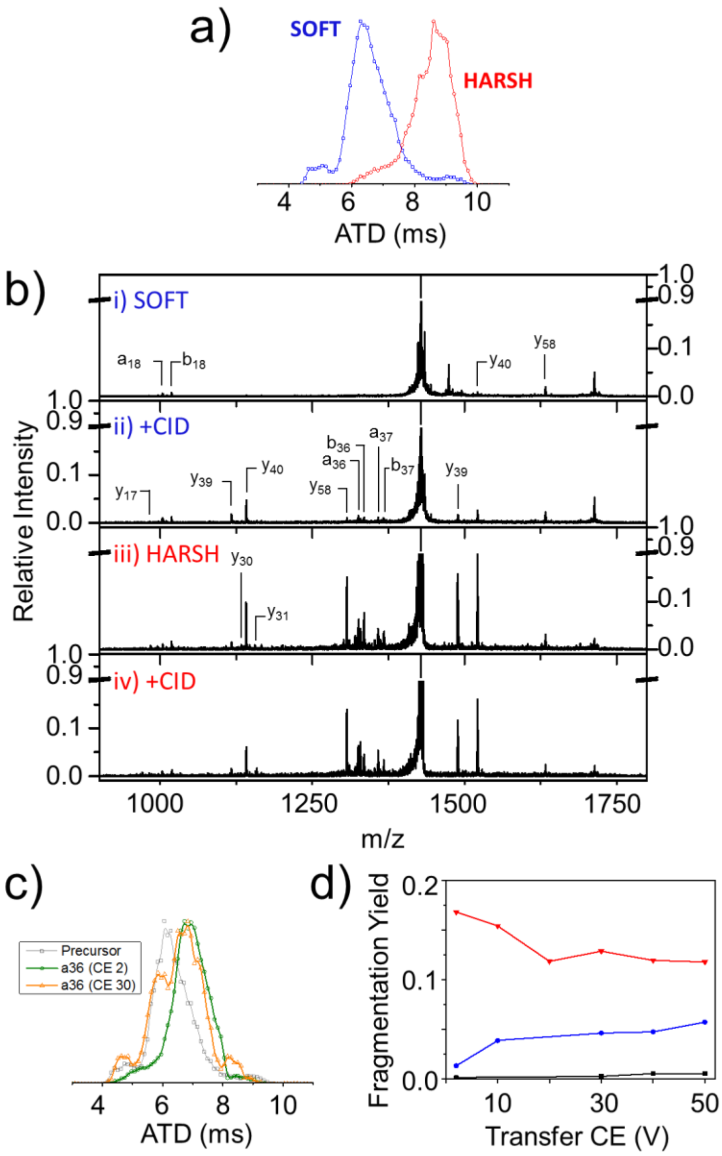
SI Figure 4: Overall fragmentation yield of $[M+7H]^{7+}$ of Cytochrome C including all fragment ions.

m/z	z	Assignment	Theoretical m/z	Soft		Intermediate		Harsh	
				CE 2V	CE 30V	CE 2V	CE 40V	CE 2V	CE 30V
1029.067	3+	a22	1029.7681		🟩	🟩	🟩	🟩	🟩
1067.141	3+	a24	1067.7824				🟩	🟩	🟩
1077.581	3+	y29-H2O	1077.9356	🟩	🟩	🟩	🟩	🟩	🟩
1083.151	3+	y29	1083.9391	🟩	🟩	🟩	🟩	🟩	🟩
1110.472	3+	a25	1110.4807		🟩	🟩	🟩		
1125.498	3+	c25	1125.4879		🟩	🟩	🟩		
1134.912						🟩	🟩		
1156.075	3+	a26	1156.167	🟩	🟩	🟩	🟩	🟩	🟩
1198.956	3+	a27	1198.8654		🟩	🟩	🟩		
1232.784	3+	a28	1232.5479			🟩	🟩		
1241.651	3+	b28	1241.8796			🟩	🟩		
1245.899	3+	a29-NH3	1245.8796	🟩	🟩	🟩	🟩	🟩	🟩
1251.485	3+	a29	1251.5551	🟩	🟩	🟩	🟩	🟩	🟩
1253.793	3+	b29-H2O	1254.8832			🟩	🟩		🟩
1261.144	3+	y33	1261.3849	🟩		🟩	🟩	🟩	🟩
1284.311	3+	a30	1283.906		🟩	🟩	🟩	🟩	🟩
1293.279	3+	y34	1293.7358		🟩		🟩		🟩
1293.279	3+	b30	1293.2376		🟩		🟩	🟩	🟩
1308.527	4+	a43	1308.1169		🟩	🟩	🟩	🟩	🟩
1311.372	5+	a55	1311.2201		🟩	🟩	🟩	🟩	🟩
1315.313	4+	b43	1315.1156		🟩	🟩	🟩	🟩	🟩
1321.926	3+	a31	1321.9203		🟩	🟩	🟩		
1359.799	3+	a32	1359.615		🟩	🟩	🟩		
1383.618	4+	a46	1383.4025		🟩		🟩	🟩	🟩
1405.555	3+	a33	1405.3013	🟩	🟩	🟩	🟩	🟩	🟩
1510.967	3+	a36	1511.0259		🟩	🟩	🟩	🟩	🟩
1625.680	2+	y29	1625.405		🟩		🟩		🟩
1690.116	5+	y75	1689.6859			🟩	🟩	🟩	🟩
1742.730	4+	y61	1742.6568			🟩	🟩	🟩	🟩
1786.980	6+	b91	1787.047	🟩		🟩		🟩	
1790.933	5+	a75	1790.2554		🟩		🟩		🟩
2038.614	6+	MH	2038.1792	🟩	🟩	🟩	🟩	🟩	🟩
2082.319	5+	a89	2081.6272		🟩		🟩		🟩
2111.921	4+	b71	2111.4819	🟩	🟩	🟩	🟩	🟩	🟩

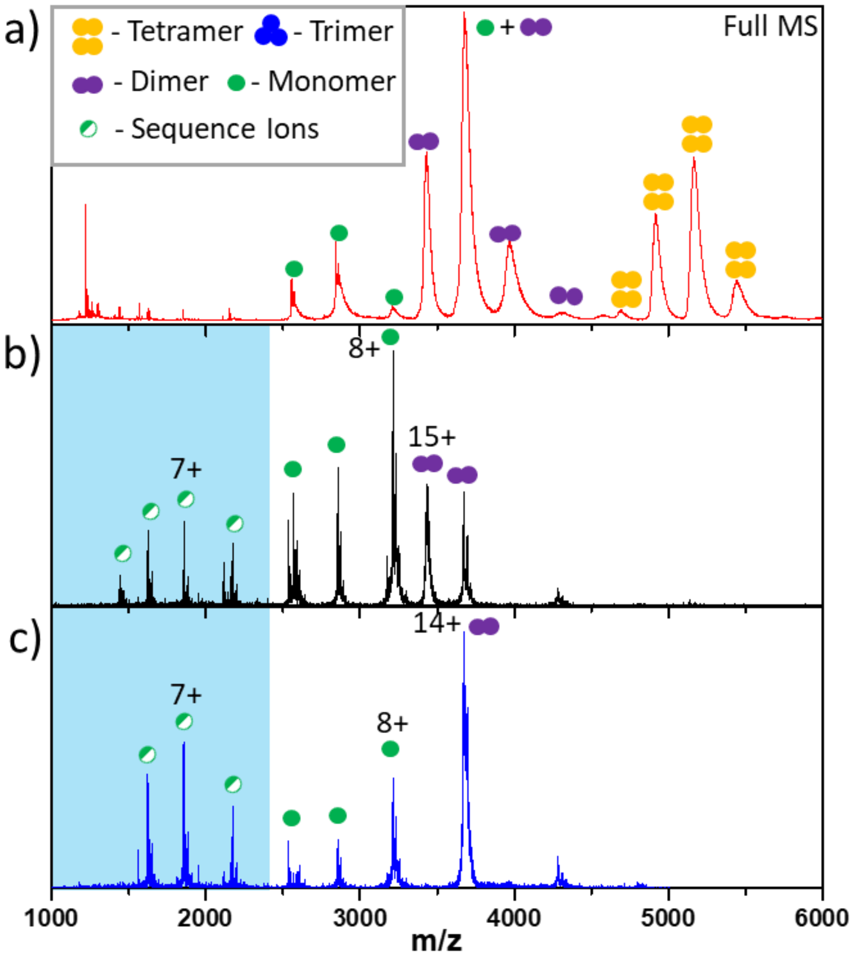
SI Table 1: List of all UVPD fragments identified from $[M+7H]^{7+}$ of Cytochrome C, green represents the fragment being present and identified in the fragmentation spectrum.



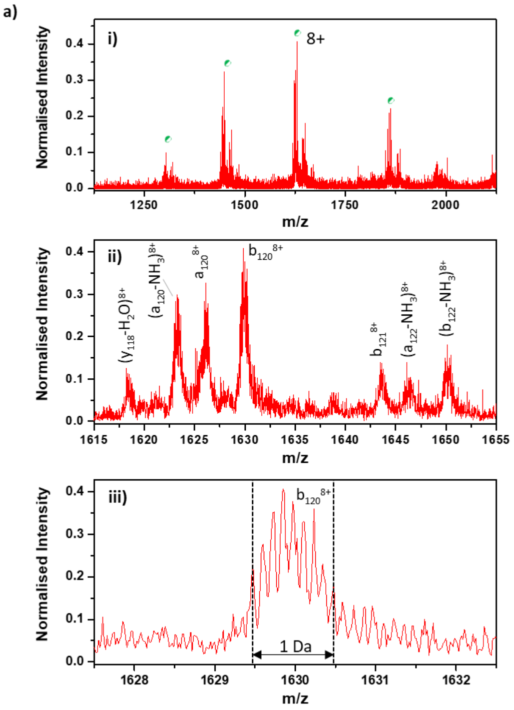
SI Figure 5: Arrival time distribution of fragment a43 under intermediate (top) and harsh (bottom) source conditions as defined in the text.



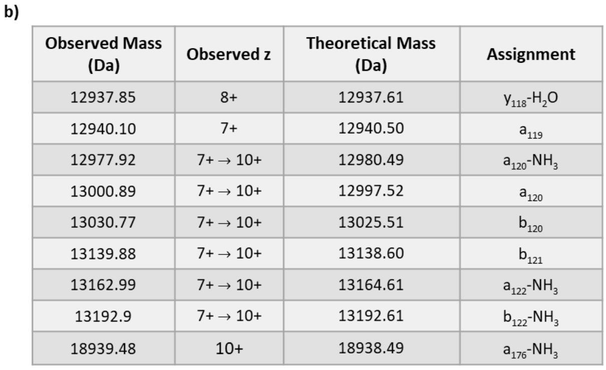
SI Figure 6: UVPD-IM-CID of two different conformations of [M+6H]$^{6+}$ of Ubiquitin. a) Arrival time distribution under gentle source conditions ('soft', blue) and activating source conditions ('harsh', red) corresponding to a cone voltage of 10 V and 120 V respectively. b) i) UVPD spectrum of the compact structural ensemble without additional CID activation post ion mobility and ii) with a transfer collision energy of 30 V. iii) UVPD spectrum of the extended conformational arrangement without additional activation and iv) with a transfer collision energy of 30 V. c) Arrival time distribution of the a36 ion at a transfer CE of 2 V (green) and 30 V (orange) compared to the precursor at CE 30V (grey). d) Fragmentation yield of a-type fragments as a function of transfer CE at soft conditions (blue), harsh conditions (red) and CID-only control (black).



SI Figure 7: a) The full MS spectrum of the Concanavalin A (red), labelled to show the subunits along with the final tetrameric structure. b) CID spectra (black), at a collisional energy of 900 eV, of the $[2M+15H]^{15+}$ dimer in the trap region of the mass spectrometer and c) CID spectra (blue), at a collisional energy of 840 eV, of the $[2M+14H]^{14+}$ dimer in the trap region of the mass spectrometer.

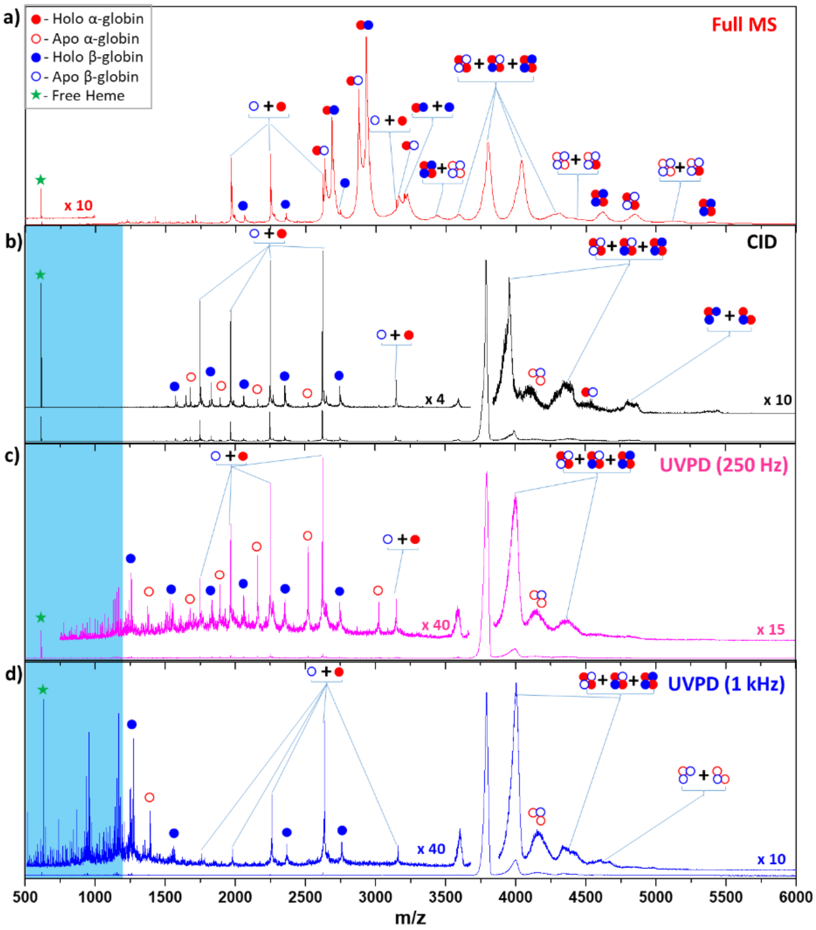


b)



Observed Mass (Da)	Observed z	Theoretical Mass (Da)	Assignment
12937.85	8+	12937.61	y_{118}-H_2O
12940.10	7+	12940.50	a_{119}
12977.92	7+ → 10+	12980.49	a_{120}-NH_3
13000.89	7+ → 10+	12997.52	a_{120}
13030.77	7+ → 10+	13025.51	b_{120}
13139.88	7+ → 10+	13138.60	b_{121}
13162.99	7+ → 10+	13164.61	a_{122}-NH_3
13192.9	7+ → 10+	13192.61	b_{122}-NH_3
18939.48	10+	18938.49	a_{176}-NH_3

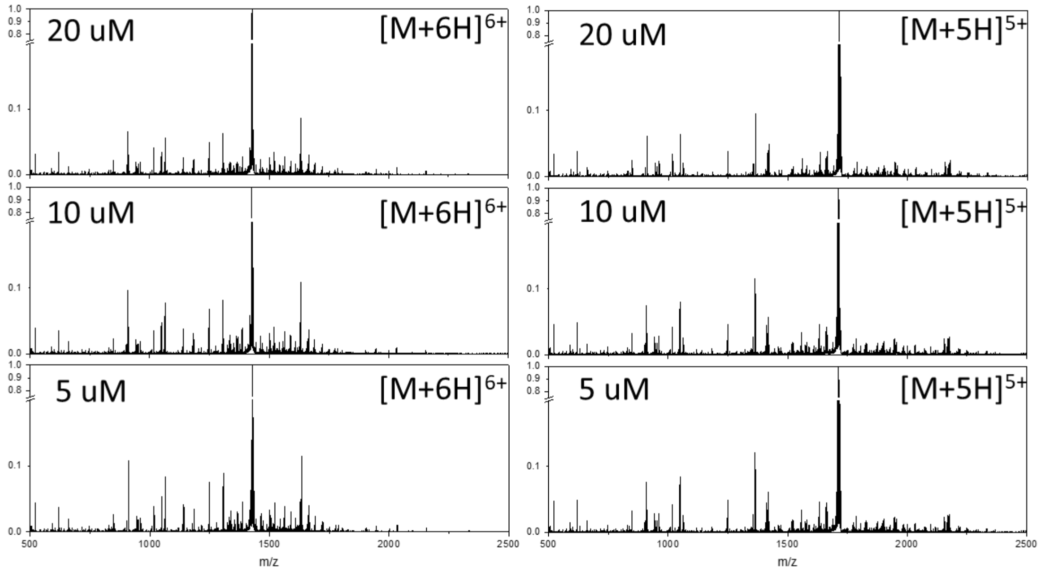
SI Figure 8: a) CID spectrum of the $[4M+20H]^{20+}$ tetramer species Concanavalin A (red), i) showing the charge state distribution of the sequence ions obtained at a CID energy of 1.6 keV, ii) the assigned sequence fragments in 8+ peak of the sequence ions and ii) showing the isotopic distribution of the dominant b_{120} fragment in the 8+ ion series. b) a table summarising the assigned sequence ions as observed in the CID spectrum of the Concavanalin A 20+ tetramer at 1.6 keV.



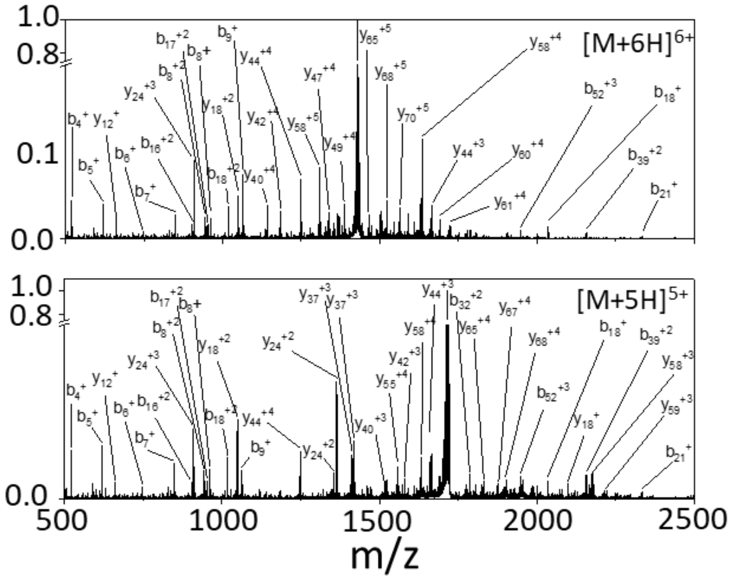
SI Figure 9: UVPD analysis of [4M+17H]$^{17+}$ of Haemoglobin. a) The full MS spectrum of Human Haemoglobin (red), labelled to show the subunits along with the final tetrameric structure. b) CID (black) activation at a collision voltage of 60 V, showing the formation of the highly charged monomer subunits and a small amount of the trimer. The UVPD fragmentation yield using the 213 nm laser, average pulse energy of 27 µJ and a repetition rate of c) 250 kHz (pink) and b) 1 kHz (blue), showing the formation of monomer and trimer subunits along with a significant amount of sequence ions. All spectra were also observed to contain free Heme and were obtained for desalted (using Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US) 10 µM Human Haemoglobin in 200 mM ammonium acetate.

m/z	z	Assignment	CID at 1.6 keV	UVPD at 250 Hz	UVPD at 1 kHz
1299.049	10+	$a_{120}\text{-}NH_3$	🟩		
1300.752	10+	a_{120}	🟩		
1303.551	10+	b_{120}	🟩		
1314.860	10+	b_{121}	🟩		
1317.461	10+	$a_{122}\text{-}NH_3$	🟩		
1320.261	10+	$b_{122}\text{-}NH_3$	🟩		
1443.277	9+	$a_{120}\text{-}NH_3$	🟩		
1445.169	9+	a_{120}	🟩		
1448.279	9+	b_{120}	🟩		
1460.844	9+	b_{121}	🟩		
1463.735	9+	$a_{122}\text{-}NH_3$	🟩		
1466.845	9+	$b_{122}\text{-}NH_3$	🟩		
1623.562	8+	$a_{120}\text{-}NH_3$	🟩	🟩	🟩
1625.690	8+	a_{120}	🟩	🟩	🟩
1629.189	8+	b_{120}	🟩	🟩	🟩
1643.325	8+	b_{121}	🟩	🟩	🟩
1646.577	8+	$a_{122}\text{-}NH_3$	🟩	🟩	🟩
1650.076	8+	$b_{122}\text{-}NH_3$	🟩	🟩	🟩
1855.356	7+	$a_{120}\text{-}NH_3$	🟩	🟩	🟩
1857.788	7+	a_{120}	🟩	🟩	🟩
1861.788	7+	b_{120}	🟩	🟩	🟩
1877.943	7+	b_{121}	🟩	🟩	🟩
1881.659	7+	$a_{122}\text{-}NH_3$	🟩	🟩	🟩
1885.658	7+	$b_{122}\text{-}NH_3$	🟩	🟩	🟩
1985.072	13+	Monomer	🟩		
2150.412	12+	Monomer	🟩	🟩	🟩
2345.813	11+	Monomer	🟩	🟩	🟩
2580.294	10+	Monomer	🟩	🟩	🟩
2866.882	9+	Monomer		🟩	🟩
3225.118	8+	Monomer		🟩	🟩
3685.706	7+	Monomer		🟩	🟩
4299.824	6+	Monomer		🟩	
5159.588	5+	Monomer		🟩	🟩
5528.059	14+	Trimer	🟩	🟩	🟩
5953.217	13+	Trimer		🟩	🟩
4913.941	21+	Tetramer		🟩	
5431.093	19+	Tetramer	🟩	🟩	🟩
5732.765	18+	Tetramer		🟩	🟩
6069.927	17+	Tetramer		🟩	🟩
6449.235	16+	Tetramer		🟩	🟩
6446.412	14+	$\text{Tetramer-}\{y_{118}\text{-}H_2O\}^{6+}$			🟩

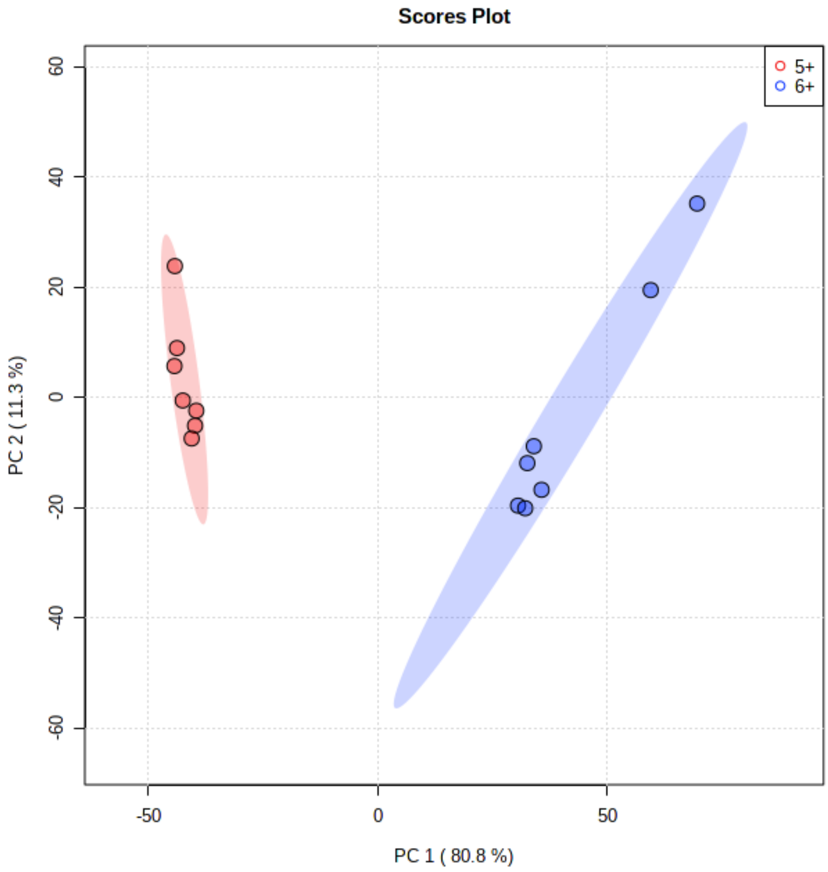
SI Table 2: List of all CID and UVPD fragments identified from $[4M+20H]^{20+}$ of Concanavalin A, green represents the fragment being present and identified in the fragmentation spectrum.



SI Figure 10: Ubiquitin CID performed at concentrations 5 µM, 10 µM and 20 µM for the $[M+6H]^{6+}$ and $[M+5H]^{5+}$.



SI Figure 11: Annotated nESI-MS/MS fragmentation spectra of ubiquitin ions $[M+6H]^{6+}$ (top) and $[M+5H]^{5+}$ (bottom) at 50 V and 60 V, respectively. Fragment ions with a neutral loss are denoted by the terminal part only. The y-axis represents the normalised intensity and in both cases have identical breaks between 0.2 and 0.8 to clearly show the fragmentation.



SI Figure 12: Scores plot for ubiquitin CID MVA data after normalising all the fragments by the TIC i.e. sum intensities of that sample.

m/z	z	Assignment	Theoretical m/z	5+	6+
520.2576	1+	b4	520.2588	🟩	🟩
619.3238	1+	b5	619.3272	🟩	🟩
661.4112	2+	y12	661.4068	🟩	🟩
674.8855	2+	b12	674.8865	🟩	
747.4189	1+	b6	747.4222	🟩	🟩
848.4716	1+	b7	848.4699		🟩
903.0129	2+	b16	903.0157	🟩	🟩
909.5187	3+	y24	909.5069	🟩	🟩
943.5513	1+	b8-H2O	943.5434	🟩	
952.5579	2+	b17	952.5499	🟩	🟩
961.5532	1+	b8	961.5539	🟩	🟩
1003.08	2+	a18	1003.0737	🟩	🟩
1008.064	2+	b18-H2O	1008.0659	🟩	🟩
1017.081	2+	b18	1017.0712	🟩	🟩
1040.593	2+	y18-NH3	1040.0947	🟩	
1044.582	1+	b9-H2O	1044.591	🟩	🟩
1049.102	2+	y18	1049.1	🟩	🟩
1062.608	1+	b9	1062.6016	🟩	🟩
1064.089	4+	y37	1064.0871	🟩	
1101.613	1+	b10-H2O	1101.6125	🟩	
1119.625	1 +	b10	1119.6231	🟩	🟩
1135.444	5+	y50	1135.4256		🟩
1141.38	4+	y40	1141.37	🟩	🟩
1166.637	2+	b21	1166.6271	🟩	🟩
1169.637	4+	y41	1169.6412		🟩
1178.088	5+	y52	1178.048		🟩
1183.639	4+	y42	1183.897	🟩	🟩
1203.867	5+	y53	1203.8564		🟩
1206.671	2+	y21	1206.6715		🟩
1216.169	4+	y43	1216.1572		🟩
1248.198	4+	y44	1248.181	🟩	🟩
1248.198	4+	y44	1248.181	🟩	🟩
1257.207	2+	y22	1257.1954	🟩	🟩
1303.792	1+	y12-H2O	1303.7957	🟩	
1306.521	5+	y58	1306.505		🟩
1321.794	1+	y12	1321.8063	🟩	
1324.589	6+	y71	1324.557		🟩
1328.724	5+	y59-H2O	1328.7115		🟩
1332.339	5+	y59	1332.3136		🟩
1337.252	4+	y47	1337.223		🟩

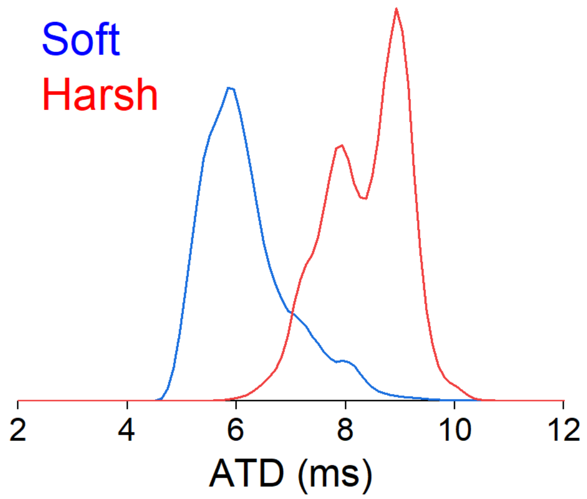
m/z	z	Assignment	Theoretical m/z	5+	6+
1348.755	1+	b12	1348.7657	🟩	
1352.14	5+	y60	1352.1273		🟩
1354.759	2+	y24-H2O	1354.7514	🟩	🟩
1363.756	2+	y24	1363.7566	🟩	🟩
1369.26	4+	y48	1369.2471		🟩
1384.406	6+	y74	1384.4269		🟩
1387.023	4+	y49	1387.006	🟩	🟩
1412.455	3+	y37-H2O	1412.4435	🟩	🟩
1418.467	3+	y37	1418.447	🟩	
1461.842	1+	b13	1461.8498	🟩	
1463.595	5+	y65	1463.5885		🟩
1467.807	4+	y52-H2O	1467.8053	🟩	🟩
1472.31	4+	y52	1472.308		🟩
1485.827	5+	y66-NH3	1485.8021		🟩
1489.228	5+	y66	1489.2075		🟩
1504.582	4+	y53	1504.5686		🟩
1515.491	3+	y40-H2O	1515.4877	🟩	🟩
1520.85	5+	y68	1520.821		🟩
1553.612	4+	y55-H2O	1553.5989	🟩	
1558.142	4+	y55	1558.102	🟩	🟩
1563.696	5+	y70	1563.648		🟩
1578.214	3+	y42	1578.193	🟩	🟩
1632.898	4+	y58	1632.88	🟩	🟩
1632.898	4+	y58	1632.88	🟩	🟩
1663.918	3+	y44	1663.9056	🟩	
1689.931	4+	y60	1689.907		🟩
1722.192	4+	y61	1722.168		🟩
1786.485	2+	b32	1786.4682	🟩	🟩
1824.738	4+	y65	1829.2338	🟩	
1829.291	4+	y65	1829.234	🟩	🟩
1843.004	3+	y49-H2O	1843.0026		🟩
1849.007	3+	y49	1849.0061		🟩
1870.999	4+	y67-H2O	1871.0102	🟩	
1875.534	4+	y67	1875.513	🟩	
1900.823	4+	y68	1900.775	🟩	
1929.079	4+	y69	1929.046	🟩	
1945.74	3+	b52	1945.713	🟩	🟩
1949.825	4+	y70-H2O	1949.805		
1981.842	4+	y71-H2O	1981.8288	🟩	
2033.131	1+	b18	2033.1351	🟩	🟩
2047.883	4+	y73	2047.8657	🟩	
2071.143	3+	y55-H2O	2071.1295	🟩	
2097.194	1+	y18	2097.1927	🟩	🟩
2118.179	2+	y37-H2O	2118.1616	🟩	

m/z	z	Assignment	Theoretical m/z	5+	6+
2127.171	**2+**	**y37**	**2127.1669**	🟩	
2154.664	**2+**	**b39**	**2154.656**	🟩	
2170.867	**3+**	**y58-H2O**	**2170.833**	🟩	🟩
2176.851	**3+**	**y58**	**2176.8369**	🟩	
2194.2	**1+**	**y19-H2O**	**2194.2091**	🟩	
2213.877	**3+**	**y59-H2O**	**2213.848**	🟩	
2332.263	**1+**	**b21**	**2332.2469**	🟩	
2486.346	**2+**	**y44-H2O**	**2486.3494**	🟩	
2675.441	**3+**	**y72-NH3**	**2675.4535**	🟩	🟩

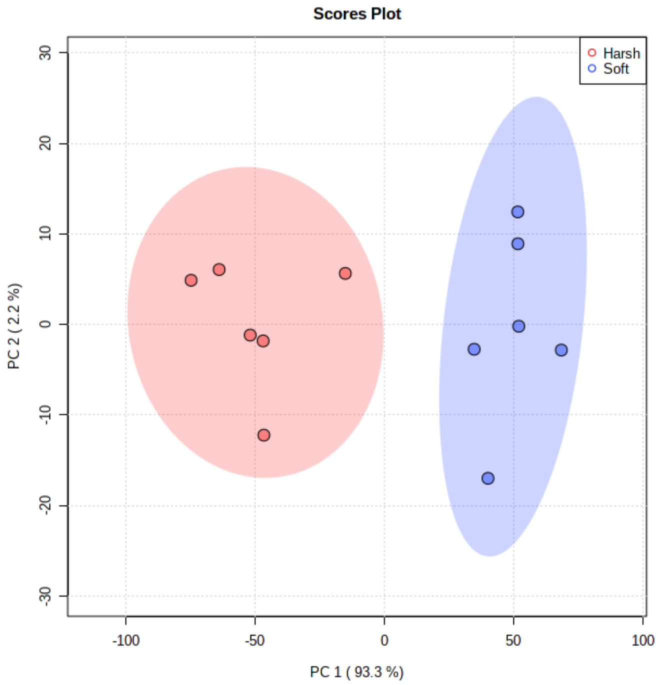
SI Table 3: List of all CID fragments identified from $[M+6H]^{6+}$ and $[M+5H]^{5+}$ of Ubiquitin, *green represents the fragment being present and identified in the fragmentation spectrum.*

m/z	FC	log2(FC)	raw.pval
1705.8	61.935	5.9527	8.57E-13
1705.6	56.511	5.8205	6.25E-13
1705.4	54.269	5.7621	6.59E-13
1705.2	53.829	5.7503	7.41E-13
1707.6	50.424	5.656	5.73E-13
1707.8	41.407	5.3718	3.49E-13
1705	40.935	5.3553	1.31E-12
1929.6	26.41	4.723	1.62E-12
1702	25.47	4.6707	5.24E-13
1702.2	23.158	4.5334	1.16E-12
1702.4	19.64	4.2957	1.59E-12
1703.4	14.731	3.8808	1.08E-12
1871.9	13.754	3.7818	1.97E-12
1950.7	12.487	3.6423	1.80E-12
1698	7.7978	2.9631	3.35E-13
1352.8	0.066932	-3.9012	1.84E-12
1353.6	0.047656	-4.3912	1.19E-12
1352.4	0.045687	-4.4521	9.57E-13
1385.3	0.025705	-5.2818	1.20E-12
1179.3	0.024829	-5.3318	1.10E-12
1352.6	0.020121	-5.6351	1.72E-12
1325.6	0.019173	-5.7048	1.56E-12
1178.5	0.018682	-5.7422	8.38E-13
1325.1	0.018636	-5.7457	9.36E-13
1378.6	0.017304	-5.8527	1.80E-12
1204.5	0.012356	-6.3386	8.74E-13
1425.5	0.004784	-7.7075	1.40E-12

SI Table 4: Univariate analysis of ubiquitin consisting of a t-test and fold change calculation. Significant ($p<0.05$) m/z values with a fold change greater than 1.5 or -1.5. All positive fold changes = high in 5+, low in 6+ and all negative fold changes = low in 5+, high in 6+.



SI Figure 13: ATD for the precursor ion for bovine cytochrome c under both soft (blue) and harsh (red) conditions. The y-axis represents the intensity of the two species, these have not been normalised here and therefore this is difference due to different acquisition lengths and overall signal.



SI Figure 14: Scores plot for cytochrome c UVPD MVA data after normalising all the fragments by the TIC i.e. sum intensities of that sample.

m/z	FC	log2(FC)	raw.pval
1078.6	23.518	4.5557	5.30E-06
2246.1	16.608	4.0538	5.75E-06
1689.7	14.107	3.8184	1.97E-08
1083.5	12.139	3.6016	2.02E-09
1255.4	10.68	3.4169	3.82E-07
1689.2	8.9418	3.1606	3.84E-08
1287.8	8.7678	3.1322	6.39E-06
1690.1	7.9869	2.9976	1.06E-06
1688.7	7.9354	2.9883	2.43E-06
1683.8	5.7725	2.5292	2.34E-07
1685.7	5.0111	2.3251	1.89E-06
1685.9	4.8657	2.2827	5.36E-06
1680.3	4.2749	2.0959	4.88E-06
2038.5	0.28264	-1.823	5.41E-07
2038.3	0.28122	-1.8302	3.94E-07
1786.9	0.21195	-2.2382	3.15E-07
1786.7	0.18455	-2.4379	5.85E-06

SI Table 5: Univariate analysis of cytochrome c consisting of a t-test and fold change calculation. Significant (p<0.05) m/z values with a fold change greater than 1.5 or -1.5. All positive fold changes = high in harsh, low in soft and all negative fold changes = low in harsh, high in soft.

3 “Light Footprinting” – The investigation of photoreceptor proteins under photo responsive environment

3.1 Declaration

This chapter consists of one drafted article awaiting submission to Nature Protocols:

This article has been reproduced in an unchanged format except for minor adjustments to incorporate it in to this thesis.

As first author on this publication I performed all MS and IMS experiments on all three photoreceptor proteins, some work for UVR8 had been performed previously by AT. I compiled and prepared all the figures. IC expressed and purified WT Tt-CarH and UVR8. DJH expressed and purified R. Palustris phytochrome.

All protocol layout was determined by myself with editing performed by PEB and LAIR. Initial LED installation was performed by AT and was further developed by myself, from this I have developed this entire protocol and flexible LED set up along with installation to multiple MS instruments.

Supporting information

Additional Figures are provided in the Supporting Information (SI) in section 3.12 of this thesis

"Light Footprinting" – The use of light and mass spectrometry to investigate photoreceptors

Rachelle Black[1], Lennart A. I. Ramakers[1], Inês, S. Camacho,[3] Bruno Bellina[1], Jeffery M. Brown[2], Derren J. Heyes[1], Alex R. Jones [3], Perdita E. Barran[1]

[1]Manchester Institute of Biotechnology and Department of Chemistry, The Unversity of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[2] Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, United Kingdom

[3] Biometrology, Department of Chemical and Biological Sciences, National Physical Laboratory, Teddington, Middlesex, TW11 0LW

Corresponding author: perdita.barran@manchester.ac.uk

3.2 Abstract

Understanding the mechanism of photoreceptor proteins, ubiquitous in nature, is a growing field both to understand fundamentals of the processes and to exploit this behaviour for biotechnology applications. Here we present a mass spectrometry-based method that determines photoresponsive mechanisms in terms of the change in stoichiometry and conformation on light activation. We describe the addition of broad-spectrum LEDs to irradiate the sample before electrospray ionisation, and subsequent mass spectrometry experiments and data acquisition. This protocol is aimed at researchers with an interest in light-sensitive proteins and a general awareness of mass spectrometry, with access to mass spectrometry instrumentation. We utilise standard, native ion mobility mass spectrometry techniques on three commercially available platforms, and we show the ease of installation of LEDs in each source, permitting light activation experiments on photoreceptors across the visible region. This protocol is demonstrated on a Synapt G2-S, Thermo Q-exactive UHMR, and an Agilent 6560 instrument with a series of photoreceptor proteins activated by light from across the near UV and visible regions of the spectrum.

3.3 Introduction

This method stems from the growing uptake of native mass spectrometry (MS) techniques to biological and biophysical researchers. The ability to investigate large multimeric photoreceptor proteins with widely available mass spectrometry apparatus and utilise LEDs rather than more expensive laser systems permits a facile implementation of this method with minimal alignment. To demonstrate the method, we focus on naturally occurring photoreceptor proteins from plant and bacterial sources that absorb wavelengths across the UV, visible, and near-IR spectrum (200-800nm).

3.3.1 Applications of the method

Previous work has described the use of native MS to study photoreceptor proteins widely located throughout the biological environment. Investigations have examined proteins from plants involved with phototropism[1] and opsin proteins which possess the retinal chromophore involved in vision.[2] The prevalence of photoreceptors in the natural world, as well as the promiscuous biological behaviour that allows them to perform multiple processes, means that they provide a rich and underexplored subject for investigation. In addition, natural photoreceptors have evolved under terrestrial light from the sun and are therefore, activated by wavelengths from across the visible and near-UV. This renders them potentially useful as tools for optogenetics, phototherapeutics, and other areas of biotechnology. [3,4]

Optogenetics involves the use of light to control biological processes, by genetically modifying host DNA to express light-sensitive proteins. This allows for targeted control of events in biological systems precisely defined in time and space.[4] Further understanding of the mechanism, light-triggered structural changes, and conformational diversity of the target photoreceptor will allow refinement and further development of optogenetic tools. To this end, biophysical approaches are required to understand the photoactivation mechanisms, including kinetics, the role of co-factors, and the protein environment as well as quaternary changes.

The behaviour of photoreceptor proteins is dependent on their chromophores, which respond to the absorption of specific wavelengths and undergo photochemical change, such as electron transfer, photoisomerisation, bond cleavage, or adduct formation. This is then followed by local reorganisation of H-bonds, salt-bridges, etc. that leads to larger scale conformational change. In this way, the wavelength-dependent activation of the chromophore is propagated through the photoreceptor protein allowing it to initiate a myriad of biological processes.[1,5] Native MS, which can preserve large biomolecular assemblies into the gas phase, is exquisitely able to determine the conformational change that will occur as a response to light by photoreceptors, Figure 3.1.

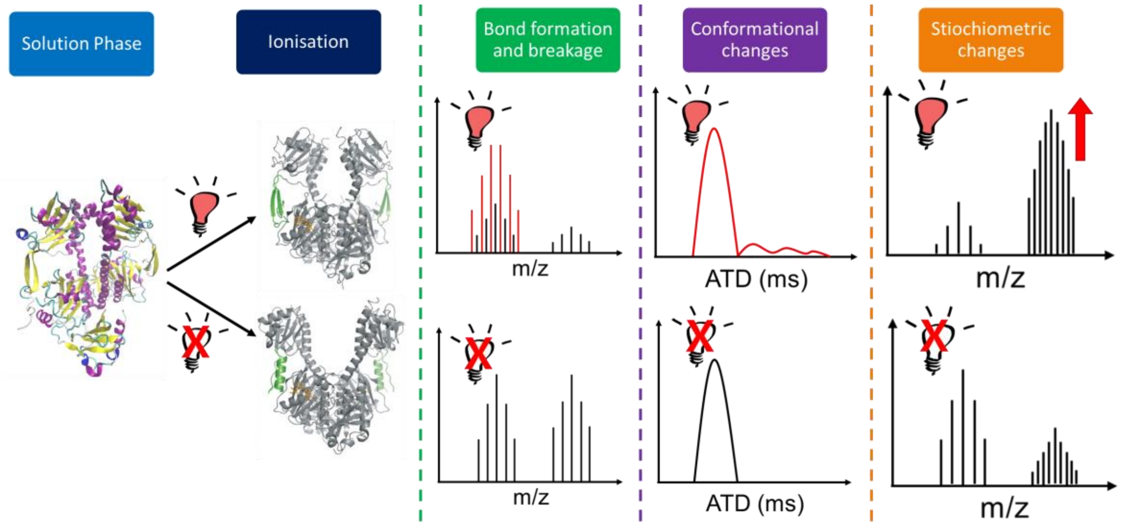


Figure 3.1: Experimental schematic of light footprinting to investigate photoreceptor proteins utilising in electrospray source LED irradiation and ion mobility-mass spectrometry. Using this methodology, the action of light on the protein can be observed in the mass spectrum and the ion mobility data. Upon irradiation in solution phase, the mass spectrum can alter due to bond formation and/or cleavage, photo-activated conformational change, and stoichiometric changes. These can all be monitored as exemplified here.

3.3.2 Development of the protocol

We initially implemented the method with investigations into the photoreceptors TtCarH and full-length UVR8 alongside a truncated form of UVR8.[6,7] CarH is an AdoCbl-dependent protein, which uniquely utilises the coenzyme B_{12} cofactor, with an upper axial 5’-ado group, for light-sensing. Initial experiments on the TtCarH photoreceptor complemented UV-visible transient absorption data and aided the determination of the photochemical mechanism. [6] It was determined that in the irreversible illumination of TtCarH tetramer the 5’-deoxyadenosyl group becomes displaced by H132,[8] which results in the TtCarH tetramer dissociating into monomers, each with a cob(III)alamin covalently bound. These photoconverted monomers are thus of a different mass to both the apomonomer (i.e., with no B_{12} cofactor bound) and the 5'-deoxyadenosylcoablamin bound monomer units that make up the tetramer. They are therefore readily identifiable *via* mass spectrometry measurements.[6] Further work focussed on the assembly of the photoactive tetrameric species upon binding of AdoCbl in the dark.[9] We discovered that, when AdoCbl is scarce, low abundance tetrameric species form with sub-stoichiometric AdoCbl. The conformation of these species is similar to the TtCarH tetramers with all four AdoCbl-binding sites occupied, suggesting each monomer does not need a chromophore bound.

The UVR8 photoreceptor employs a cluster of tryptophan residues as a photoactive chromophore.[10] Our investigations, therefore, utilised a 280 nm mounted LED that was directed towards the source, section 3.4.4.[7] The method was implemented into a range of available mass spectrometers and provided both stoichiometric and structural insights. The work investigated the full-length wild type (FLWT) UVR8 and the core domain UVR8 (residue 12-381) used for previous structural studies, where the disordered tails are truncated.[7] Both are photoactive (*i.e.*, they undergo a dimer to monomer transition when illuminated), but the role of the disordered tails present in the full-length protein was dissected in a manner that was lacking from prior structural work. The tails were revealed to adopt both compact and extended conformations and, surprisingly, destabilise the core fold in the monomers, presumably to aid with binding to downstream partners. Results from ion mobility mass spectrometry were augmented with computational modelling, which provided candidate conformations to assist our understanding of the mechanism of photoactivation as well as subsequent downstream processes.

Further validation of the technique was provided by studies on a third class of photoreceptors, the phytochromes. We incorporated two LEDs that allow for irradiation at two wavelengths and reversible photoconversion between two states, section 3.4.4, again with facile changes to the electrospray ion source. The phytochromes are a class of photoreceptor proteins that are found in plants, bacteria, and fungi and comprise two identical chains.[11] Each of these chains contains a PAS, GAF, and a PHY domain, which are conserved domains that are present in the N-terminal photosensory region of all of the different species.[5] The chromophore, a bilin derivative, is covalently bound within the GAF domain of this N-terminal region and gives rise to the typical light sensitivity of phytochromes in the red and far-red regions of the visible spectrum. They can be split into Type I or Type II, referring to being activated by far-red light or activated by red light, respectively.[12] The bilin chromophore[5,13] undergoes photoisomerisation, which is then amplified through the protein to a shift from β-sheet to α-helix in the tongue that allows for the opening of the dimer[14–16] This behaviour is well researched, although details of the structures that constitute the transient "open" and "closed" forms are not always tractable with current methods, and as such these represent excellent demonstrators for light-foot printing MS experiments.

In summary, the growing interest in photoreceptor proteins and the more widespread availability of native MS presents an opportunity for this light-footprinting method to be implemented in many laboratories. We here describe a protocol to do this on commercially available mass spectrometry instrumentation.

3.3.3 Comparison with other methods

Common methods for investigating dynamical and structural change in proteins include X-ray crystallography, Cryo-electron microscopy, and NMR. In every case, experimental data are coupled with computational methods to resolve structural details.[17–19] For photoreceptors, such approaches, necessitate distinct illumination conditions for the pre-activated and activated states, which can be cumbersome or impossible to implement. Optical methods are utilised to provide the absorption maximum for photoreceptors,[6,20], and to obtain time-resolved information and aid in the determination of the mechanism, but do not provide structural information on the global fold. [20–24] In addition, for X-ray crystallography, large-scale conformational changes, such as those observed in many photoreceptor proteins, can destroy the crystal altogether.[25] Crystallisation of proteins remains a challenge, especially for proteins that contain intrinsically disordered regions or conformational diversity, to avoid this issue there have been significant developments in the study of biological macromolecule dynamics *via* X-ray scattering of liquid solutions. X-ray scattering approaches (small-angle X-ray scattering, [SAXS] and wide-angle X-ray scattering, [WAXS]) can eliminate this requirement for the crystallisation of the photoreceptor protein and monitor any overall structural change.[25] For such X-ray scanning approaches access to synchrotrons and X-ray free-electron lasers (XFELs) is required, which are not usually 'down the hall' as with MS instrumentation. The unique selling point of native MS is the ability to research these intrinsically disordered or conformationally diverse proteins.[26–30]

Since the development of soft ionisation techniques, such as electrospray ionisation (ESI) and nano-ESI (nESI), the use of MS has been valuable in the structural investigation of proteins.[31] The transition from solution to gas-phase and the ability to maintain structure, has been discussed routinely and different methodologies have been utilised to show the perseverance of the solution structure.[32–37] For the MS experiment to be native, the ionisation conditions must be 'soft' and the resulting mass spectrum is likely to consist of a narrow charge state distribution (CSD) consisting of low charge states, indicating that non-covalent interactions have been preserved, for complexes this is also indicated by the mass recorded. The combination of MS and IMMS with optical data in the infrared (IR) and with ultraviolet photodissociation (UVPD) and electron capture dissociation (ECD) are consistent with the preservation of secondary structure into the gas phase.[38]

There have been notable studies that monitor photo-induced conformational change within the vacuum environment of the mass spectrometer,[39–43] and there is evidence that some photoreceptor proteins require an aqueous environment to enable the light-activated structural response.[7] As a consequence, in this protocol activation, is performed *via* illumination of the photoreceptor protein in an aqueous environment in the nESI tip and the structural changes then occur before transfer to the gas phase. This is akin to hydrogen-deuterium exchange (HDX) and fast photochemical oxidation of proteins (FPOP) measurements,[44] wherein chemical changes in solution provide a mass footprint of structural changes that can be recorded in the mass spectrometer and hence we term our analogous approach 'light foot-printing'.

3.3.4 Expertise needed to implement the protocol

It is important that the user has a good understanding of native MS and how to perform MS whilst maintaining the non-covalent interactions. Methodology for native MS is well reported now due to its increased applicability; a guide for the typical native MS settings has been previously published and covers a range of ion mobility instrumentation and can be adapted accordingly.[45] Other groups have also shown the use of the technique to characterise overproduced proteins with minimal preparation required.[46]

3.4 Overview of the procedure

3.4.1 Timeline

The main purpose of this protocol is to demonstrate the effectiveness of MS as an analytical technique for the investigation of photoreceptor proteins; therefore, a summary of the methodology to produce the photoreceptors utilised here is explained. The MS-specific methodology is the main focus here, full details of the steps can be found in Section 3.8. The first step of which is the preparation of the photoreceptor proteins into an 'MS-friendly' buffer,[46,47] which takes between 1 hour to 1 day depending on the method that is taken (steps 2-3). The MS and IMS experimental work takes approximately a day for each protein with technical challenges described in the troubleshooting section (steps 4-12). The data processing will take approximately a day for each protein (steps 13-18).

3.4.2 Photoreceptor preparation

The preparation of photoreceptor proteins varies depending on the photoreceptor, with the general process involving the use of plasmids encoded with the photoreceptor protein gene. The method for the synthesis is often found in literature, as is the case of the photoreceptors utilised here, UVR8,[7,48] TtCarH[9,49] and *R. Palustris* phytochrome.[50]

3.4.3 Mass spectrometry preparation

The photoreceptor proteins are desalted into an MS-friendly salty solution, commonly ammonium acetate or bicarbonate, these are safe to use with the mass spectrometer and will not contaminate or "clog" up the ion optics of the instrument which will then require significant maintenance or cleaning. Such salts are also volatile and depart the protein with the aqueous solution leaving the protein as a desolvated ion.[51] The experiments described here use working protein concentrations that are 2-10 µM, depending on the sensitivity of the instrument and the mass heterogeneity of the analyte. In house produced quartz nESI tips are used for photoreceptors where the absorption is in the UV region of the electromagnetic spectrum, where this is not required borosilicate nESI tips are used. Both are produced using a Sutter P-2000 Laser-Based Micropipette. The settings are then tuned according to the photoreceptor by visual inspection of the tip but mainly by the spray quality, stability and ability to spray the biomolecule at low capillary voltage. Tune settings that make the most difference in the tip shape are the pull, velocity, and heat, which should be changed in steps, with their effect on the tip shape monitored and tested. Tune settings may differ between photoreceptor proteins. This is due to the inherent difference between the proteins, as they often have different masses, extent of disorder and conformational fold, which all impact the potential spray settings. nESI tips are filled in the dark or appropriate illumination *via* back filling. Tips are visually checked for the presence of bubbles in the solution, which may block the spray. These can easily be removed by flicking the tip to force the bubbles to rise away from the nESI tapered tip or with use of a centrifuge.

3.4.4 Mass spectrometry acquisition

Initial MS data acquisition is best performed with the photoreceptor solution under conditions where it is not activated, this will either be in the dark under a black-out curtain or under ambient light of a non-activating wavelength, SI Figure 1. Under these conditions, native MS measurements can be performed, which include checking that the spray is stable over the time period that the experiment will run (hours), that the signal intensity is sufficient, and that the ensuing charge state distribution and stoichiometry are invariant over at least 30 minutes. In addition, tandem MS measurements can be made, such as collision-induced activation to investigate the stability and topology of the protein.[52,53] Where possible, ion mobility can also be implemented, ions of interest can be m/z selected and measurements such as selected arrival time distribution (ATD) to gain information on the global conformation and activated ion mobility (aIMS), which will enable the unfolding of the ion to be monitored.

After native MS measurements have been completed in the dark, the LED can be switched on to monitor any change in signal. Running an acquisition file whilst the switch from dark to LED is being performed will allow for the change to be monitored and recorded to yield an ion chromatogram, Figure 3.2a. This file will also contain the mass spectrum both before and after the illumination procedure, Figure 3.2b; therefore direct comparison can be made between the two. To yield mobility data, the full mass spectrum should be acquired again under the new illumination, which will allow for ion mobility measurements to be compared to the dark spectrum, Figure 3.2c.

3.4.5 Data processing

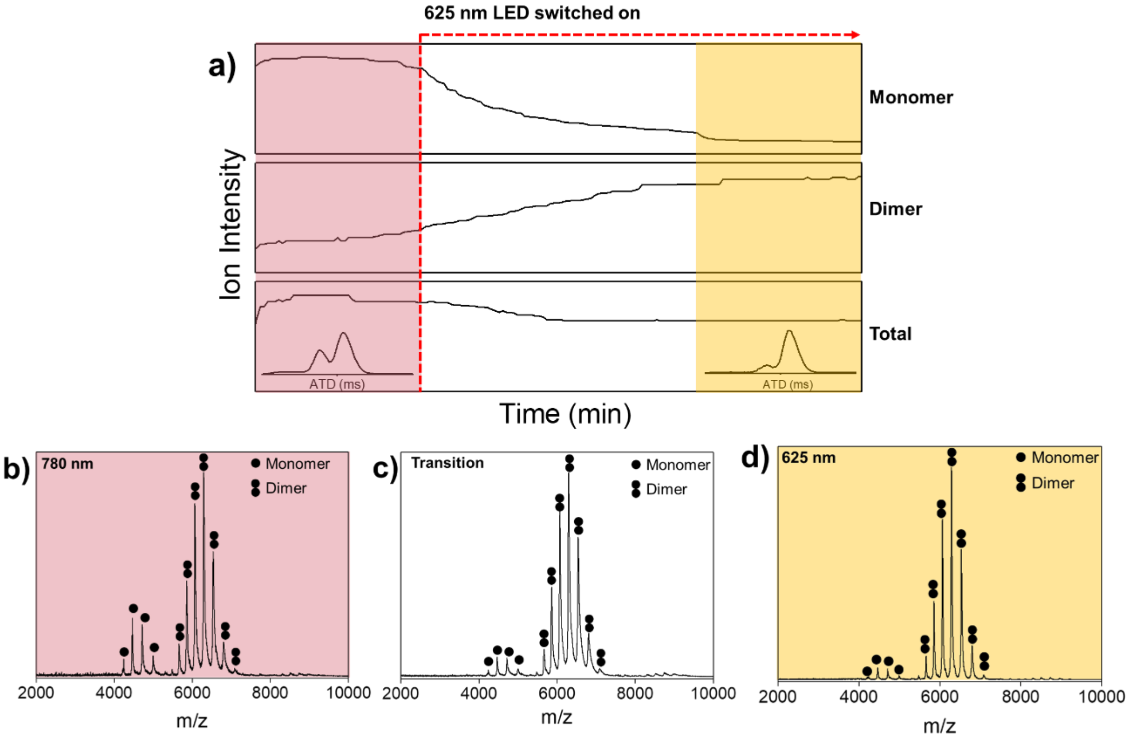


Figure 3.2: Example data obtained for the R. Palustris phytochrome initially under 780 nm before switching to 625 nm illumination, a) The ion count for total mass spectrum, selected for the monomer species (m/z range of ~4100 – ~5200) and the dimer species (m/z range ~5400 – ~7300), time versus the ion intensity, arrow represents the point where is switched to 625 nm. Inset is the ATD for 780 nm and 625 nm. b, c and d) The full mass spectrum under 780 nm and 625 nm illumination respectively, green circles represent monomer species and double green circles represent the dimer species.

Data processing can be performed using software provided by the instrument manufacturer as well as with open-source software that can be manufacturer invariant and possess greater functionality.[54–56] The total ion count (TIC) can be used to determine if there is any change in the ionisation yield as a function of illumination, for example due to the relative ionisation efficiency of monomers versus dimers, or if the sample aggregates. If the TIC has a large variance it may be harder to determine whether the observed shift in the species is real or an effect of the TIC variation. This may make it important to normalise the plots to the TIC at each time point to remove this issue. Selected ion counts (SIC) are also useful, as depicted in, Figure 3.2a to demonstrate the shift in stoichiometry as a function of illumination alongside the mass spectrum, Figure 3.2b and c. If ion mobility is available, the total arrival time distribution (ATD) can also report on changes in stoichiometry since monomeric forms will most often have different drift times to multimers. The more common use of IM data is to monitor the ATD of m/z selected species to determine conformational differences, Figure 3.2a inset. The experimental ATDs can be converted to CCS distributions as described previously, to allow comparison to conformations that have been measured with other methods.[57]

3.5 Materials

Biological Reagents:

- Coenzyme B_{12} ≥97.0% (Merck, cat. no. C0884)
- Core UVR8$^{12\text{-}381}$ and FLWT UVR8
- *R. Palustris* Phytochrome
- WT Tt-CarH

Reagents:

- Ammonium acetate solution (Fisher Scientific, cat. no. A/3440/53)
- Ultrapure water (Milli-Q Advantage A10; Merck Millipore, Darmstadt, cat no.: Z00Q0V0WW)

Instruments and equipment:

- Mass spectrometers. Synapt G2-S HDMS instruments (Waters MS Technologies) 6560 IM-QToF (Agilent Technologies) and Q-exactive UHMR UHMR mass spectrometry (Thermo Fisher Scientific)
- Varian Cary 50 Bio UV-Visible Spectrophotometer (Agilent Technologies)
- P-2000 Laser-Based Micropipette Puller (Sutter Instruments)
- Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, cat. no. 7326221)

- Slide-A-Lyzer Dialysis Cassettes, 3.5K MWCO, 0.5 mL (Thermo Fisher Scientific, cat. no. 66333)
- Hamilton GASTIGHT syringe, 1700 series (Merck, cat. no. 20972)
- Platinum wire, 99.95%, diameter 0.125 mm (± 10%) (Goodfellow, cat. no. PT005117)
- Borosilicate glass capillaries with filament, outer diameter 1.2 mm, inner diameter 0.69 mm, length 10 cm (Science Products, cat. no. Z-BF-120-69-10)

Software

- ORIGAMI[54]
- CIUSuite[55]
- UniDec[56]
- Origin 2017
- Microsoft Office

LED custom set-up:

- Laser safety fabric (ThorLabs, cat. no. LPCM1)
- Threaded mount for TO-18, TO-39, TO-46, or T-1 ¾ LEDs (ThorLabs, cat. no. S05LEDM)
- Lens tube, 2-inch thread depth (ThorLabs, SM05L20)
- Lens tube, 3-inch thread depth (ThorLabs, SM05L30)
- N-BK7 Bi-Convex lens, diameter 1/2-inch, f=25.0 mm, uncoated (ThorLabs, cat. no. LB1014)
- N-BK7 Bi-Convex lens, diameter 1/2-inch, f=30.0 mm, uncoated (ThorLabs, cat. no. LB1258)
- Slip ring for SM2 lens tubes, M4 tap (ThorLabs, cat. no. SM2RC/M)
- ThorLabs mounted LEDs
    - 780 nm IR mounted LED (ThorLabs, 200 mW, 800 mA, cat. no. M780L3)
    - 625 nm Fibre-coupled LED (ThorLabs, 13.2 mW, 1000 mA, cat. no. M625F2)
    - 530 nm single colour cold visible mounted LED (ThorLabs, 370 mW, 1000 mA, M530L4)

    - 455 nm single colour cold visible mounted LED (ThorLabs, 1150 mW, 1000 mA, cat. no. M455L4)
    - 405 nm UV mounted LED (ThorLabs, 1000 mW, 1000 mA, cat. no. M405L4)
- ThorLabs unmounted LEDs
    - 625 nm LED with a glass lens, 12 mW, TO-18 (ThorLabs, 12 mW, cat. no, LED625L)
    - 505 nm LED with a glass lens (ThorLabs, 4 mW, cat. no. LED505L)
- Arduino Uno (Arduino, cat. no A000066)
- T-Cube LED driver, 1200 mA Max Drive Current (ThorLabs, cat. no. LEDD1B)

3.6 Hardware requirements for illumination pre-ionisation

The majority of the work was performed on a Synapt HDS G2-S which possesses ion mobility capabilities, although the method is amenable to other mass spectrometers and has also been demonstrated on a Q-exactive UHMR UHMR and a 6560 IM-QToF mass spectrometer. [58–61]

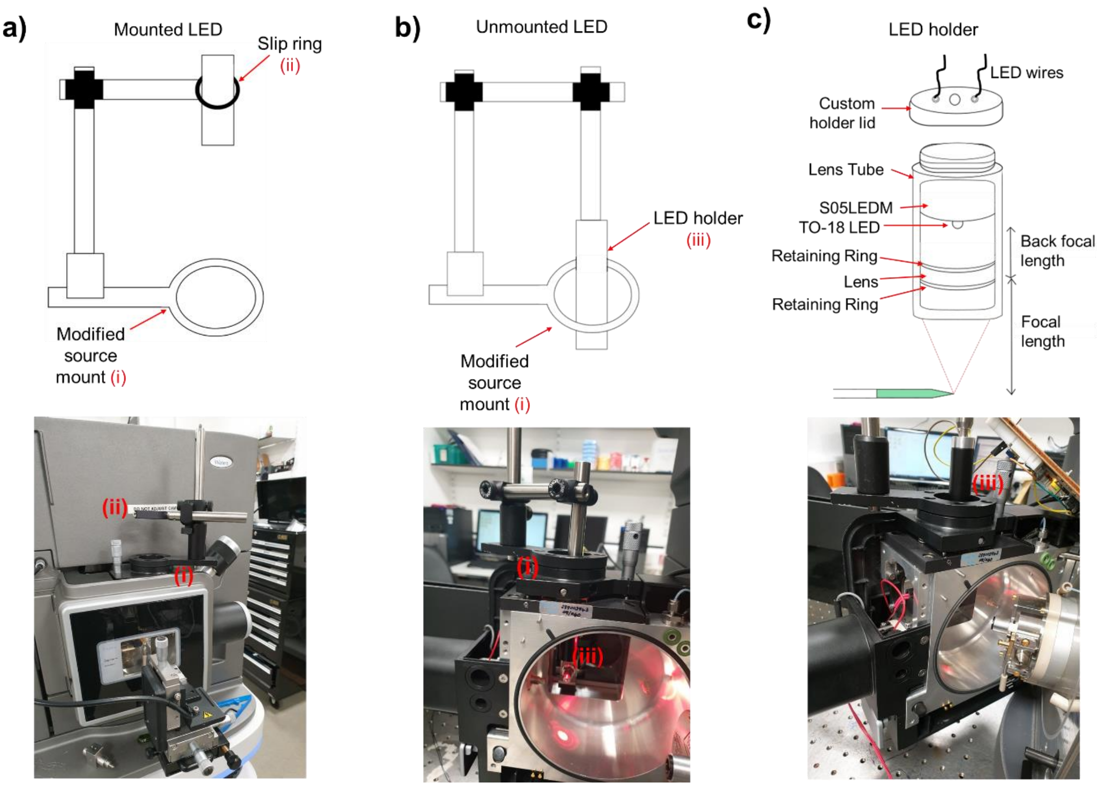


Figure 3.3: Schematics of the LED holder utilised for the inclusion into MS. a) The schematic for the mounted LEDs purchased from ThorLabs including the slip ring and modified source mount. A labelled photo of the mounted LED arrangement is shown underneath the schematic. b) The schematic for the unmounted LEDs purchased from ThorLabs including the lens tube and custom holder lid. A labelled photo of the unmounted LED arrangement is shown underneath the schematic. c) Holder for unmounted TO-18 LEDs including a lens into focus the illumination onto the end of the nESI tip (green). A photo of the unmounted LED lens tube in position on the Waters source.

To perform the light-footprinting of the photoreceptors, a rig to secure the relevant LEDs is installed within the source region of the mass spectrometer external to the desolvation region. It is possible to purchase both mounted LEDs which retail at c. £250/$342 and unmounted LEDs that retail at a fraction of this price c. £10/$14. Here we have shown both types. Either type of LED requires a modified source mount to allow the LEDs to illuminate the source region. For mounted LEDs, a slip ring is required and the beam, which passes through a lens can then be aligned vertically to a focal point within the nESI tip (which can also be aligned with respect to the light beam). Figure 3.3a, whereas for unmounted LEDs a lens tube that acts as a LED holder is required with a home manufactured aluminium lid for the LED as well as machined holes in the LED body both of which allow coupling to the optical posts whilst maintaining access to the LED through the connected wires, Figure 3.3b and c. Both set ups utilise optical mounting posts from ThorLabs.

In all cases, the LEDs were focused onto the end of the nESI tip. This required precise measurements of both the back focal length in the installation of the lens into the lens tube and the front focal length of its positioning within the source, Figure 3.4a and b. These were determined using the information of the lens used and put into a position depending on the front focal length, this was then checked in the source.

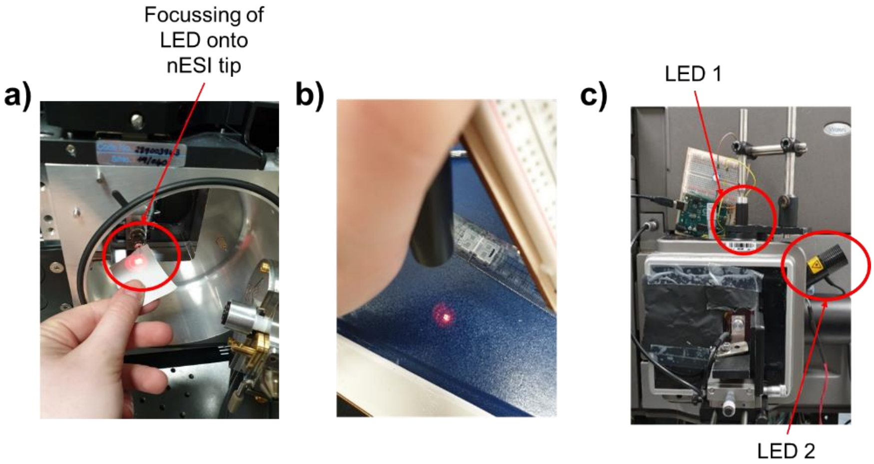


Figure 3.4: Light footprinting using more than one LED, a) focussing of LED1, aligned so it is focussed onto the tip. b) The focal spot of LED1 when using the lens within the tube. c) LED1 aligned above the tip holder and LED2 directed at the tip from the side

Light footprinting can be performed with two or more LEDs of differing wavelengths, Figure 3.4c. Such multiple LED arrangements permit monitoring of a photo-responsive species under a change in illumination wavelength.

3.7 Exemplar results

To test the LED system, it is important to utilise biomolecules that are widely available and have easily observable changes.

3.7.1 Small molecule chromophores

Small molecule chromophores can be illuminated with the LED setup described above. One example of this is 5’-deoxyadenosylcobalamin (coenzyme B_{12}, AdoCbl). This is a biologically active form of vitamin B_{12}, absorbs light across the visible and near UV spectrum at wavelengths shorter than ~ 600 nm and with peaks in the green and UVA. Our previous work coupling photo activation with MS allowed a mechanism to be derived that included the dissociation of the carbon-cobalt bond upon illumination, resulting in the formation of a deoxyadenosyl radical. This causes a mass difference making it the perfect test subject for this LED arrangement, Figure 3.5a. AdoCbl was dissolved in water:methanol (50:50) and diluted to 5 µM before injection into the mass spectrometer.

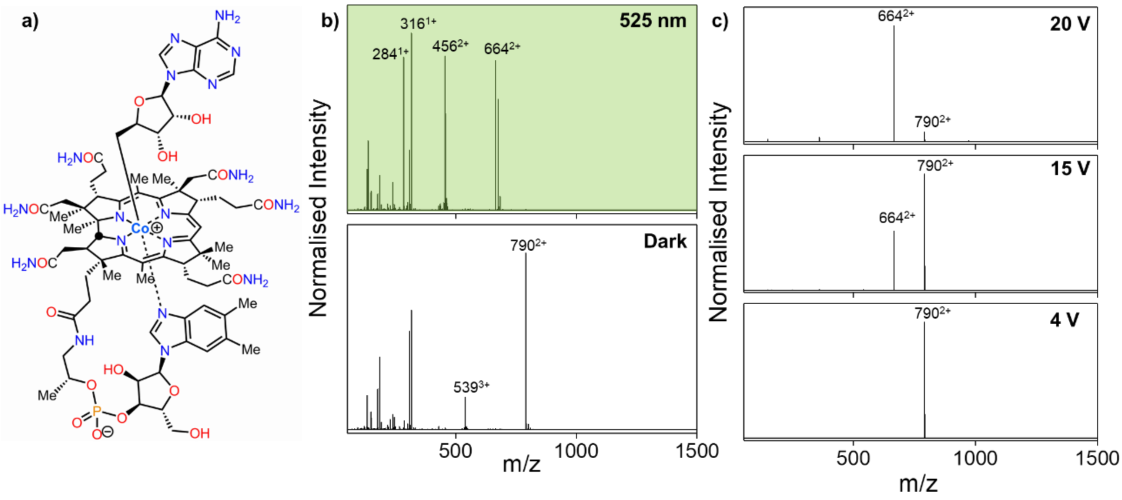


Figure 3.5: AdoCbl chromophore utilised for the LED set up on the Synapt G2-S, a) chemical structure, b) full mass spectrum under dark and 530 nm LED illumination, and c) trap collisional activation of the $[M+2H]^{2+}$ *ion under dark conditions.*

Initially, data was acquired with the sample (and ionisation source) held in the dark followed by illumination (here at 530 nm), Figure 3.5b. In this case, under the dark conditions, AdoCbl is identifiable at *m/z* 790; however, once this is illuminated at 530 nm, the loss of the 5’-deoxyadenosyl group, results in the product ion with *m/z* 664.

To observe whether this reaction also can occur via collisional activation, the precursor ion was subjected to increasing collision energy in the trap cell of the Synapt G2-S,

Figure 3.5c. This also resulted in the formation of the product at 664 m/z as well as some smaller ions likely to be caused by the further fragmentation of the AdoCbl molecule after the loss of the 5’-deoxyadenosyl group.

3.7.2 Photoreceptor proteins

3.7.2.1 Mass Spectrometry

Expanding this out to chromophores within photoreceptor proteins, the flexibility of the LEDs has been shown by implementing the setup on a range of photoreceptor proteins that absorb at a variety of wavelengths.

Three photoreceptor proteins were chosen to test the LED system, UVR8, TtCarH, and *R. Palustris* phytochrome. These were run under conditions that wouldn’t initiate a photoresponse to obtain initial MS data; illumination conditions were then switched to the activating wavelength, Figure 3.6ai, bi, and ci. These photoreceptor proteins were chosen to cover a range of wavelengths across the near UV and visible spectrum. In each case, the action observed is a change in stoichiometry. Core UVR8$^{12\text{-}381}$ was investigated and also yielded a shift in dimer to monomer species under UV-B illumination, SI Figure 2. This was identical to the full-length UVR8 however the full-length UVR8 also exhibited a low abundance of high charge states for both the monomer and the dimer, inferring that the full-length conformation was more diverse and that the action of light was dependent on the conformation of the precursor.[7] The shift in the oligomeric states can be seen *via* the MS CSD; it can also be monitored *via* the SIC where the increase/decrease in the relevant m/z range can be monitored, Figure 3.6aii, bii and cii. In all cases, the SIC is produced for the different oligomeric species as well as displaying the total TIC, which will monitor the spray efficiency throughout the acquisition.

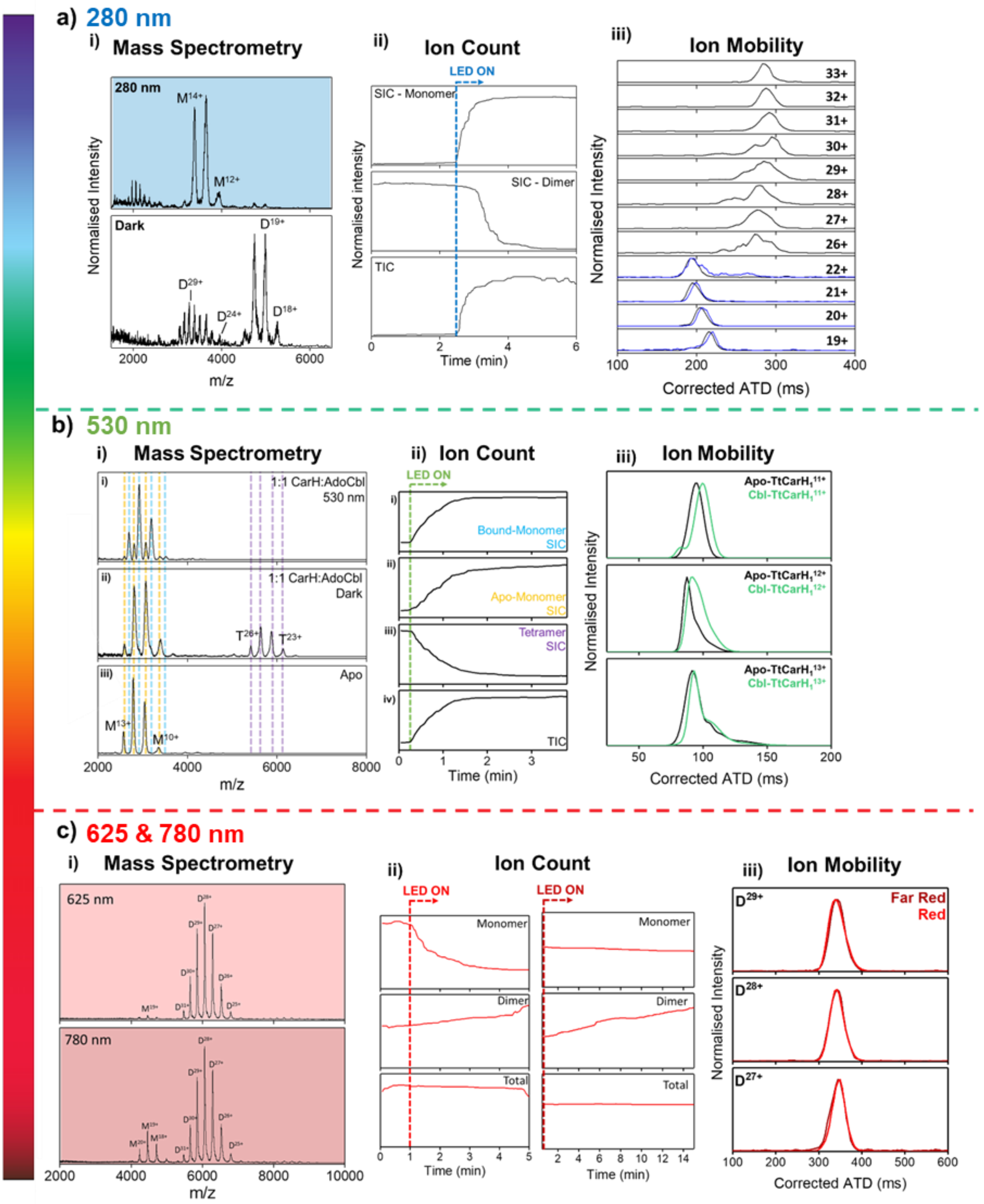


Figure 3.6: Example IMMS results following illumination of three different proteins under three different wavelengths a) FLWT UVR8 at 280 nm, b) WT Tt-CarH and 530 nm, and c) R. Palustris phytochrome 625 and 780 nm. Data shown includes (i) full MS data pre-and post-LED illumination, (ii) ion count monitoring and (iii) drift time corrected mobility data. The coloured stripe to the left represents the visible spectrum and where these photoreceptors absorb.

To demonstrate the versatility of the setup, it has also been configured with a Synapt G2-S and a Q-exactive UHMR UHMR, SI Figure 3a-d. *R. Palustris* phytochromes have been utilised for the Q-exactive UHMR UHMR application due to their reversibility and therefore flexibility if exposed to light. In both mass spectrometers, the nESI source provides an ideal setup for the installation of one LED with a second LED requiring more modifications or temporary holders, demonstrated here with the use of a clamp stand. The same change in the population of monomer and dimer was also observed on the Q-exactive UHMR UHMR with the monomer species increasing in the 780 nm illumination than the 625 nm, SI Figure 3e. Some low intensity species due to a tetrameric form of the protein are observed in the Q-exactive UHMR UHMR data, this is likely caused by in spray oligomerisation. At a lower concentration of the protein (5 μM) SI Figure 3f there is a reduction in the intensity of this tetramer signal, implying that this species was produced *via* non-specific aggregation.

Activation techniques such as CID can be performed alongside native MS approaches. This has been performed on the apo- and holo-WT TtCarH monomer and the *R. Palustris* phytochrome, SI Figure 4a and b respectively. Holo-TtCarH shows the clear loss of the AdoCbl when CID is applied, the backbone fragmentation then yields identical fragmentation to the apo-TtCarH. *R. Palustris* phytochrome CID was performed at a range of increasing voltages and to multiple charge states, SI Figure 5, with minimal dimer to monomer dissociation observable. Interestingly, this difficulty to produce monomer with CID was observed previously with the UVR8 photoreceptor.[7] It was explained previously, that the interfacial region of each monomer is highly charged and the dimeric structure is maintained by cross-dimer salt bridges in the dark. It was previously thought that a water molecule was ejected from the interfacial region following photoexcitation of the tryptophan chromophore, which would then result in the disruption of the network of H-bonds and cross-dimer salt bridges.[62] Therefore, for dimers to produce subsequent monomers without immediately re-dimerising, previous experiments have indicated that additional water molecules are required to enter the interfacial regions and shield the complementary charges from one another. These water molecules are absent in the gas phase and therefore, prevents the formation of the monomers.[7]

3.7.2.2 Ion Mobility-Mass Spectrometry

IMS data can be acquired following the same light foot-printing in source and simultaneously to the MS data; *i.e.*, ATDs can be acquired first in the dark or with a base wavelength and subsequently upon illumination. The ATD can then be converted to a CCS distribution for each charge state as well as to global CCS distributions scaled with respect to MS signal intensity.

IMMS was performed for the three example photoreceptor proteins discussed above, Figure 3.6aiii, biii and ciii. For TtCarH, there is a difference between the ATD plots although it is marginal and at the cusp of experimental error. When comparing the respective monomers both before AdoCbl has been introduced and after illumination, there is a small mass change post-illumination that accounts for the binding of the adenosylcobalamin, resulting in tetramer formation in the dark and subsequent breaking under illumination. This process causes a local structural arrangement but this was not readily observable in the IMS data on the entire protein. For *R. palustris*, there is no difference in the ATD pre- and post-illumination. The ATDs are converted into the CCS distribution following calibration, SI Figure 6.[57] For the UVR8, the dimer CSD was much broader, therefore indicating multiple conformations, which was confirmed by the ATDs that are observed across the charge states, 19+ through to 33+. Interestingly, for the UVR8$^{12\text{-}381}$ the CSD is much narrower and ATDs account for only one conformation, which is a significant difference in the ATDs and thus indicates that the C- and/or N- terminal chains that had not been present in the core UVR8$^{12\text{-}381}$ must themselves be disordered and present thee different conformations. It was previously deduced through molecular dynamic measurements that the C-terminal chain, which is substantially longer than the N-terminal chain, is the reason behind this conformational flexibility[7]

The unfolding behaviour of a protein is dependent on the non-covalent interactions of the native form and activated-IMS can monitor this. In such experiments protein ions are mass selected and subjected to stepwise collisional activation whilst monitoring their ATDs. Such a procedure was performed on the three *R. palustris* phytochrome dimer ions, 28+, 27+ and 26+ under both red and far-red illumination conditions, SI Figure 7. The results showed a gradual unfolding from a drift time of 10.5 ms to 14 ms. For all three charge states, when the lab frame energies are increased to ~1540 V, ~1500

V, and ~1950 V respectively, there is then a rapid shift to the unfolded conformer with a drift time of 16 ms. This trend is highly similar with both red and far-red illumination indicating that there is little difference in the non-covalent interactions between the two dimers, and that the opening of the dimer results in the formation of the monomer.

3.8 Procedure

Photoreceptor preparation

1. Photoreceptors are produced using standard methodology through cellular expression. A brief explanation of the methods used for the exemplar photoreceptors is given below with more details in the relevant references.:
    a. The gene encoding full-length UVR8 from *Arabidopsis thaliana* (residues 1-440) was cloned into modified pET28a (*Novagen*). The protein was expressed and purified as previously described according to Camacho *et al.*, 2019.[7]
    b. The gene encoding the CarH from *Thermus thermophilus* was synthesised and subcloned into pET15b (*Novagen*). The protein was expressed and purified as previously described according to Camacho *et al.*, 2021.[9]
    c. The codon-optimized gene encoding the photosensory region (PAS-GAF-PHY) of phytochrome from Rhodopseudomonas palustris was synthesised and subcloned into pET21a (Life Technologies). The protein was expressed and purified as previously described according to Choudry *et al.*, 2018.[50]

Photoreceptor mass spectrometry preparation (Timing 1 hour – 1 day)

2. If protein is provided in a buffer that is not compatible with native MS, then it should be buffer exchanged before analysis. Depending on the photoreceptor, the desalting/buffer exchange procedure will differ, it is beneficial to try both procedures outlined below to determine which results in a well resolved mass spectrum. These two approaches are as follows:
    a. Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US). Standard protocol was performed following the instructions provided. (Timing 1 hour)

    b. Slide-A-Lyzer dialysis cassettes (Thermo Fisher Scientific, MA, US). Standard protocol was performed following the instructions provided. (Timing 1 day)

3. UV/Vis absorbance measurements should always be performed to confirm that the protein is in the dark state, prior to starting the MS experiment.

Mass spectrometry experiments (Timing 1 day)

4. All proteins were then diluted to a final concentration of 10 µM for mass spectrometry investigations.

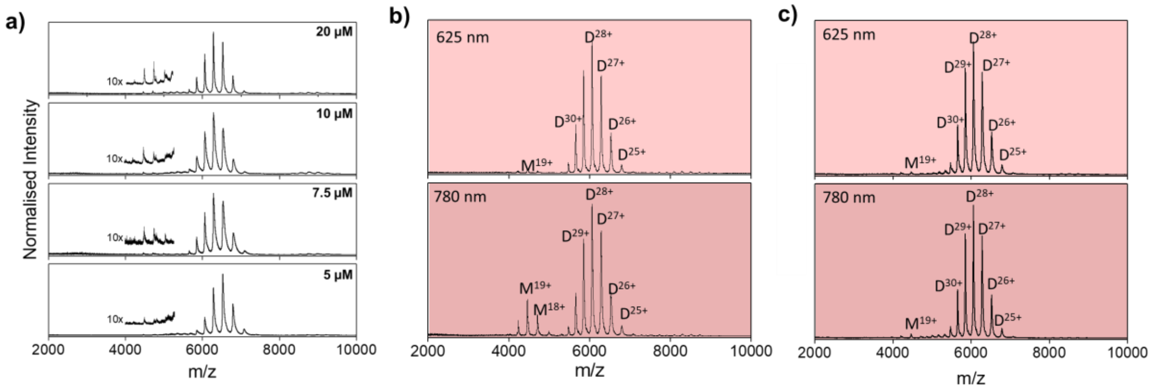


Figure 3.7: Monitoring the full MS of the R. Palustris phytochrome under different protein concentrations (a) and different ammonium acetate concentrations, 40 mM ammonium acetate (b) and 200 mM ammonium acetate (c).

    a. The concentration of the protein becomes very important when using nESI. If the concentration is too high, then the small orifice used can become blocked and prevent the spray, if the concentration is too low, then it will become difficult to see the mass spectrum. It is therefore important to test a range of concentrations, Figure 3.7a. The appropriate concentration will also vary between instruments due to differences in sensitivity. Of the concentration range shown here, the 10 µM concentration was taken forward for future measurements. Although 7.5 µM and 5 µM do appear to show the clear mass spectrum, the spray was not stable in both of these cases, whereas 10 µM provided sufficient concentration to obtain a clear mass spectrum and maintain a stable spray.

    b. The buffer conditions depend on protein and testing is required for optimisation (shown for *R. Palustris* phytochrome in both 40 mM and 200 mM ammonium acetate, Figure 3.7b, and c). These investigations showed that 200 mM ammonium acetate reduces the response to the change in illumination conditions. Previous work has shown that the concentration of ammonium acetate alters the ionic strength and therefore affects the affinity of receptor-ligand binding.[63] It is therefore likely that the ionic strength has a similar effect on the monomer-dimer interactions observed here. These effects mean that it is very important to choose the concentration of the ammonium acetate carefully.

5. All sample preparation is performed in the dark or under the illumination of a non-activating wavelength. Samples are loaded into the nESI tips under these conditions and a single Pt wire is installed into the tip to apply the voltage.
6. Initially, data are acquired in the dark to get the baseline MS before any photo-activation (in some cases, such as the phytochromes this is performed under a constant illumination of a specific wavelength to yield one of the quaternary arrangements). The MS settings can then be tuned to get the best signal intensity and maintain native-like conditions.
    a. Example instrument settings can be found in Table 3.1. These are for a modified Synapt G2-S in both MS only and mobility settings. Further example instrument settings for the Agilent 6560 and Thermo Q-exactive UHMR UHMR can be found in SI Table 1 and 2, respectively.

Table 3.1: Example instrument parameters on the Synapt G2-S. The instrument used here was a modified Synapt G2-S therefore settings may vary due to modifications that are present.

	MS	**IMS**
Source voltages (V) and gas flows (mL/min)		
Capillary (kV)	<1.50	<1.50
Source Temperature (°C)	60-80	60-80
Sampling Cone	10-60	10-60
Source Offset	20	20
Trap Collision Energy	4	4
Transfer Collision Energy	2	2
Trap Gas Flow (mL/min)	2-4	2-4
Helium Cell Gas Flow		180-200

	MS	**IMS**
IMS Gas Flow (mL/min)		75-90
DC voltages		
Trap DC Entrance	0	3
Trap DC Bias	2	35-45
Trap DC	-2	0
Trap DC Exit	0	0
IMS DC Entrance	-20	20
Helium Cell DC	1	50
Helium Exit	-20	-20
IMS Bias	2	3
IMS DC Exit	20	0
Transfer DC Entrance	5	4
Transfer DC Exit	15	15
RF voltages (V)		
Trap	350	300
IMS	300	300
IMS Mobility	250	250
Transfer	350	350
Wave velocities (ms^{-1}) and heights (V)		
Trap Wave Velocity (m/s)	300	450
Trap Wave Height (V)	0	4
IMS Wave Velocity (m/s)	300	170
IMS Wave Height (V)	0	15
Transfer Wave Velocity (m/s)	249	300
Transfer Wave Height (V)	0	10
Pressures (mbar)		
Backing	1.33E+00	1.58E+00
Source	6.69E-03	7.66E-03
Sample Plate	1.00E-06	1.00E-06
Trap	1.37E-02	2.74E-02
Helium Cell	1.00E-04	1.36E+03
IMS	1.49E-04	3.29E+00
Transfer	9.05E-03	2.25E-02
TOF	9.57E-07	1.19E-06

b. Key parameters to minimise are the capillary voltage, source temperature, and sampling cone. A lower trap bias will also allow for a more native-like photoreceptor; however, it will also have a significant effect on the transmission. For larger photoreceptors, a higher trap gas flow will allow for better transmission of the photoreceptor. A higher helium gas flow and lower IMS gas flow will also soften the mobility experiments and allow a more native-like transmission.

7. Along with the native MS, the m/z of the ions of interest should be mass-selected and exposed to further activation to compare the fragmentation pattern to the illuminated form.
8. Where available, IMS measurements should also be performed on the full mass spectrum along with selected ions, the mobility of the ion will depend on the charge state and conformation of the ion.
    a. One thing to consider is the potential of the biomolecule to undergo charge stripping. This involves the ion losing a proton whilst travelling through the instrument post-IMS separation, therefore arriving at the detector as a lower charge state. If this happens it can result in a different ATD readout that matches that of the ion with a higher charge state, SI Figure 8. This can be avoided by performing m/z selection of the ion before ion mobility measurements.
9. Activated IMS (aIMS) measurements can also be performed, which includes recording the ATD whilst increasing the activation applied to the ion before ion mobility measurements. The software Origami,[54] allows for this to be performed in a single data file at set increment measurements.
10. Once the MS has been acquired in the dark/base wavelength, the different LED illumination can be monitored. It is good practice to acquire the data whilst the switch in illumination occurs, this allows for any light-induced changes to the MS to be monitored and recorded throughout.
    a. It is important here to make sure that the LED is illuminating the solution within the tip as the majority of the photoreceptors require the solution phase to perform their structural change.
11. When under the new illumination, the analysis can be repeated under the different wavelength illumination. This includes full MS acquisition, collisional activation on the selected m/z, IMS measurements, and activated ion mobility.

Data processing (Timing 1 day)

12. There is an abundance of open-source software available to aid in the analysis of MS and IMMS data. UniDec aids in the determination of oligomeric species and identifying the mass of the charge state distribution[56] and CIUSuite aids in the analysis of activated ion mobility data.[55] Origami is another open-source software

that has incorporated both UniDec's MS capabilities and the ability to analyse activated ion mobility data collected either manually or *via* the Origami instrument addition software.[54]

13. The MS data for the dark and illuminated samples can be compared directly and provide a clear visual difference in the dominant species.
14. If an acquisition file has been performed whilst the wavelength was switched, then the ion count can represent the TIC, the SICs on dominant species, and ions of interest, such as to monitor the switch between monomer and dimer.
15. Yielding collision cross-section (CCS) measurements from the ATDs collected will depend on the ion mobility instrument utilised.
    a. A travelling-wave ion mobility spectrometer (TWIMS) is more commonplace, with the Waters Synapt operating using this approach. TWIMS calibration is required using known protein standards where the mass of the standard protein and the charge states it presents bracket those of the photoreceptor protein. The procedure for calibration has been explained in-depth previously. [57]
    b. If a drift time ion mobility spectrometer (DTIMS) has been utilised, then calibration is not required. Instead, the mobility measurement will be performed across a range of voltages, which can then be used to produce a line where the gradient will represent the mobility of the ion. This can then be inserted into the Mason-Schamp equation to determine the CCS of the ion.[64,65] Commercial instruments will vary in their methodology but can often perform this using built-in software. Instruments that have been modified to contain linear ion mobility cells, such as the RF-confining drift tube in a Synapt G1 or G2 will often require manual processing of this data.[66,67]
16. Where calibration isn't available or not wanted, the ATD can be corrected to account for the charge of the ion. To do this the drift time is multiplied by the charge state of the ion. This allows for the direct comparison between the different charge states and corrects for the influence of the charge state in the ion mobility separation,
17. Traditionally, aIMS is presented as heat maps that represent the shift in the conformation when the activation in the source has increased. Origami can minimise

the amount of extra processing, protocols on how to use this open-source software and what you can gain from utilising it have been published previously.[54]

3.9 Troubleshooting

Step	Problem	Possible reason	Solution
6	No ions were detected across the m/z range.	Blocked capillary No capillary voltage applying.	1 – Knock the tip of the capillary. 2 – Reposition the capillary closer to the cone and increase the capillary then move. 3 – If capillary continues to become blocked then sample dilution is required.
6	Unstable spray.	Capillary position. Capillary voltage. Protein concentration. Protein aggregation. Damaged nESI tip.	Reposition the capillary. Increase/decrease the capillary voltage. Increase/decrease the sample concentration and run through the instrument at a higher concentration. 1 - Check the tip under the microscope for any aggregation at the end of the tip that is blocking the spray. 2 – Dilute the protein spray solution.
6	Charge state distribution is unexpected and background signals are the most significant ions.	Settings for the MS are not native.	1 - Check the instrument parameters are retaining the native-like structure, if using ion mobility then check the IMMS ATD results. 2 - Reduce the voltages to the lowest they can transmit using the above values as an example.

Step	Problem	Possible reason	Solution
10	Charge state distribution is more similar to that of the post-activation species.	Photoreceptor has been activated during sample preparation, likely the photoreceptor was exposed to the activating wavelength.	1 - If the photoresponse is unknown, then check the absorption on the UV/Vis as a comparison before and after activation. 2 - All photoreceptors investigated will have a safe illumination working wavelength. This allows for the room light to be switch to this wavelength without activating the photoreceptor before analysis. 3 - If the photoreceptor is exposed before analysis then depending on the photoreceptor and its behaviour, the whole process will likely have to restart. 4 - Keep several aliquots of the same photoreceptor available, so in the event of early photoactivation, there are spare aliquots available. 5 - If the photoreceptor is reversible, then they are easily illuminated back to their desired state
12	No change is observed in the charge state distribution or ion mobility under different illumination.	LED not focused onto the tip correctly. The conformational change is too quick or small to monitor *via* this approach.	Reposition the LED onto the nESI tip. Check the expected behaviour of the photoreceptor, it is required to have a significant structural change to observe a difference before and after illumination.

[a] This is simply brushing the end of the tip using the cone when there is no capillary voltage applied.

3.9.1 Advantages and limitations

As with any method, there are advantages and limitations with the use of mass spectrometry that apply to this protocol. A list of limitations for the mass spectrometry of macromolecular assemblies has been previously published by the Robinson group.[68] A list of further limitations relating specifically to the photoreceptor proteins and the advantages of this protocol is as follows:

Table 3.2: The advantages and limitations of the MS protocol for the investigation and monitoring of photoreceptor proteins.

Advantages		Limitations	
Quick identification of features and changes.	Monitoring of the entire MS spectrum allows for the process to be examined in its entirety, this includes clear identification of oligomeric species that may be formed.	**Instrument sensitivity.**	With a significant number of these photoreceptors being large proteins, instrument settings must be tuned for the transmission of these to get the best response available. It is very important to obtain clear and assignable MS data to be able to understand and describe the process that is occurring.
Straightforward instrument modification.	The LED installation is cheap and easy to add to any commercial mass spectrometer.	**Scanning the visible spectrum.**	In its current arrangement, this setup is not able to perform scanning experiments often performed by laser systems. This setup is available for specific wavelength irradiation of the source however future developments may be made for the ability to scan across the wavelengths.
Full flexible coverage of the visible spectrum.	LEDs are available at wavelengths covering the entire visible spectrum and can be	**Change must be a shift in mass or a significant conformational**	MS-only experiments require a full mass shift, such as a monomer-dimer shift

	easily swapped for a different wavelength when required.	**rearrangement.**	or vice versa to be observed. However, IMS can observe conformational change, this can also be determined from a change in charge state distribution, whereby a larger more extended conformation will exhibit a higher charge state distribution. IMS experiments are limited to monitoring global conformational change therefore the shift is required to be significant to observe this.
Applicable to photoreceptor proteins unable to crystallise.	MS technology is easily applicable across different photoreceptors with minimal modification to instrument acquisition settings.	**Monitors the gas-phase interactions and behaviour.**	Would need to be performed alongside other technique to demonstrate gas-phase relation to the solution phase.

3.10 Anticipated results

Following this protocol, structural mass spectrometry of photoreceptor proteins can be yielded. The protocol requires very little sample and requires little parameter changes between the different photoreceptors. It therefore, provides major benefits as a fairly rapid and relatively straightforward technique that yields important structural and conformation details about the photoreceptor protein of interest. The LED flexibility allows for the installation onto multiple mass spectrometer interfaces, and the type of mass spectrometer will mean the resolution, sensitivity, or ion mobility capabilities can be tested and utilised. This can yield information such as structural multimeric changes in the photoreceptor protein and even smaller conformational changes in the photoreceptor through the inclusion of ion mobility. It should also be noted that for a difference to be observed, the structural changes must be significant, such as oligomeric formation or dissociation or a significant conformational change.

3.11 References

1 S. G. Kong and K. Okajima, *J. Plant Res.*, 2016, **129**, 111–114.

2 A. Terakita, *Genome Biol.*, 2005, **6**, 1–9.

3 H. M. Beyer, S. Naumann, W. Weber and G. Radziwill, *Biotechnol. J.*, 2015, **10**, 273–283.

4 K. Deisseroth, *Nat. Methods*, 2011, **8**, 26–29.

5 C. Bongards and W. Gärtner, *Acc. Chem. Res.*, 2010, **43**, 485–495.

6 R. J. Kutta, S. J. O. Hardman, L. O. Johannissen, B. Bellina, H. L. Messiha, J. M. Ortiz-Guerrero, M. Elías-Arnanz, S. Padmanabhan, P. Barran, N. S. Scrutton and A. R. Jones, *Nat. Commun.*, 2015, **6**, 7907.

7 I. S. Camacho, A. Theisen, L. O. Johannissen, L. A. Díaz-ramos, J. M. Christie, G. I. Jenkins, B. Bellina, P. E. Barran and A. R. Jones, *Proc. Natl. Acad. Sci.*, 2019, **116**, 1116–1125.

8 M. Jost, J. Fernández-Zapata, M. C. Polanco, J. M. Ortiz-Guerrero, P. Y. T. Chen, G. Kang, S. Padmanabhan, M. Elías-Arnanz and C. L. Drennan, *Nature*, 2015, **526**, 536–541.

9 I. S. Camacho, R. Black, D. J. Heyes, L. O. Johannissen, L. A. I. I. Ramakers, B. Bellina, P. E. Barran, S. Hay and A. R. Jones, *Chem. Sci.*, 2021, **12**, 8333–8341.

10 L. Rizzini, J. J. Favory, C. Cloix, D. Faggionato, A. O’Hara, E. Kaiserli, R. Baumeister, E. Schäfer, F. Nagy, G. I. Jenkins and R. Ulm, *Science*, 2011, **332**, 103–106.

11 H. Takala, A. Björling, M. Linna, S. Westenhoff and J. A. Ihalainen, *J. Biol. Chem.*, 2015, **290**, 16383–16392.

12 S. Von Horsten, S. Straß, N. Hellwig, V. Gruth, R. Klasen, A. Mielcarek, U. Linne, N. Morgner and L. O. Essen, *Sci. Rep.*, 2016, **6**, 1–8.

13 H. Hanzawa, T. Shinomura, K. Inomata, T. Kakiuchi, H. Kinoshita, K. Wada and M. Furuya, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4725–4729.

14 H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen and S. Westenhoff, *Nature*, 2014, **509**, 245–248.

15 N. C. Rockwell and J. C. Lagarias, *ChemPhysChem*, 2010, **11**, 1172–1180.

16 M. Carrillo, S. Pandey, J. Sanchez, M. Noda, I. Poudyal, L. Aldama, T. N. Malla, E. Claesson, W. Y. Wahlgren, D. Feliz, V. Šrajer, M. Maj, L. Castillon, S. Iwata, E. Nango, R. Tanaka, T. Tanaka, L. Fangjia, K. Tono, S. Owada, S. Westenhoff, E. A. Stojković and M. Schmidt, *Structure*, 2021, **29**, 743-754.e4.

17 P. Koehl, *Rev. Comput. Chem.*, 2006, **22**, 1–55.

18 G. A. Petsko, D. Ringe and G. A. P. D. Ringe, *Annu. Rev. Phys. Chem.*, 2004, **23**, 165–192.

19 H. A. Scheraga, M. Khalili and A. Liwo, *Annu. Rev. Phys. Chem.*, 2007, **58**, 57–83.

20 S. J. O. Hardman, D. J. Heyes, I. V. Sazanovich and N. S. Scrutton, *Biochemistry*, 2020, **59**, 2909–2915.

21 D. J. Heyes, S. J. O. Hardman, M. N. Pedersen, J. Woodhouse, E. De La Mora, M. Wulff, M. Weik, M. Cammarata, N. S. Scrutton and G. Schirò, *Commun. Biol.*, 2018, **2**, 1–8.

22 J. Tenboer, S. Basu, N. Zatsepin, K. Pande, D. Milathianaki, M. Frank, M. Hunter, S. Boutet, G. J. Williams, J. E. Koglin, D. Oberthuer, M. Heymann, C. Kupitz, C. E. Conrad, J. Coe, S. Roy-Chowdhury, U. Weierstall, D. James, D. Wang, T. Grant, A. Barty, O. Yefanov, J. Scales, C. Gati, C. Seuring, V. Srajer, R. Henning, P. Schwander, R. Fromme, A. Ourmazd, K. Moffat, J. J. Van Thor, J. C. H. Spence, P. Fromme, H. N. Chapman and M. Schmidt, *Science*, 2014, **346**, 1242–1246.

23 C. Kupitz, S. Basu, I. Grotjohann, R. Fromme, N. A. Zatsepin, K. N. Rendek, M. S. Hunter, R. L. Shoeman, T. A. White, D. Wang, D. James, J. H. Yang, D. E. Cobb, B. Reeder, R. G. Sierra, H. Liu, A. Barty, A. L. Aquila, D. Deponte, R. A. Kirian, S. Bari, J. J. Bergkamp, K. R. Beyerlein, M. J. Bogan, C. Caleman, T. C. Chao, C. E. Conrad, K. M. Davis, H. Fleckenstein, L. Galli, S. P. Hau-Riege, S. Kassemeyer, H. Laksmono, M. Liang, L. Lomb, S. Marchesini, A. V. Martin, M. Messerschmidt, D. Milathianaki, K. Nass, A. Ros, S. Roy-Chowdhury, K. Schmidt, M. Seibert, J. Steinbrener, F. Stellato, L. Yan, C. Yoon, T. A. Moore, A. L. Moore, Y. Pushkar, G. J. Williams, S. Boutet, R. B. Doak, U. Weierstall, M. Frank, H. N. Chapman, J. C. H. Spence and P. Fromme, *Nature*, 2014, **513**, 261–265.

24 N. Coquelle, M. Sliwa, J. Woodhouse, G. Schirò, V. Adam, A. Aquila, T. R. M. Barends, S. Boutet, M. Byrdin, S. Carbajo, E. D. La Mora, R. B. Doak, M. Feliks, F. Fieschi, L. Foucar, V. Guillon, M. Hilpert, M. S. Hunter, S. Jakobs, J. E. Koglin, G. Kovacsova, T. J. Lane, B. Lévy, M. Liang, K. Nass, J. Ridard, J. S. Robinson, C. M. Roome, C. Ruckebusch, M. Seaberg, M. Thepaut, M. Cammarata, I. Demachy, M. Field, R. L. Shoeman, D. Bourgeois, J. P. Colletier, I. Schlichting and M. Weik, *Nat. Chem.*, 2018, **10**, 31–37.

25 M. Levantino, B. A. Yorke, D. C. F. Monteiro, M. Cammarata and A. R. Pearson, *Curr. Opin. Struct. Biol.*, 2015, **35**, 41–48.

26 R. Beveridge, A. S. Phillips, L. Denbigh, H. M. Saleem, C. E. MacPhee and P. E. Barran, *Proteomics*, 2015, **15**, 2872–2883.

27 R. Beveridge, Q. Chappuis, C. Macphee and P. Barran, *Analyst*, 2013, **138**, 32–42.

28 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, **42**, 177–185.

29 C. Bleiholder and M. T. Bowers, *Annu. Rev. Anal. Chem.*, 2017, **10**, 365–386.

30 C. Bleiholder, N. F. Dupuis, T. Wyttenbach and M. T. Bowers, *Nat. Chem.*, 2011, **3**, 172–177.

31 J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.

32 J. L. P. Benesch, B. T. Ruotolo, D. A. Simmons and C. V Robinson, *Chem. Rev.*, 2007, **107**, 3544–3567.

33 A. J. R. Heck, *Nat. Methods*, 2008, **5**, 927–933.

34 K. Muneeruddin, J. J. Thomas, P. A. Salinas and I. A. Kaltashov, *Anal. Chem.*, 2014, **86**, 10692–10699.

35 S. Mehmood, T. M. Allison and C. V. Robinson, *Annu. Rev. Phys. Chem.*, 2015, **66**, 453–474.

36 S. H. Chen and D. H. Russell, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 1433–1443.

37 S. Vahidi, B. B. Stocks and L. Konermann, *Anal. Chem.*, 2013, **85**, 10471–10478.

38 J. Seo, W. Hoffmann, S. Warnke, M. T. Bowers, K. Pagel and G. von Helden, *Angew. Chemie - Int. Ed.*, 2016, **55**, 14173–14176.

39 B. D. Adamson, N. J. A. Coughlan, R. E. Continetti and E. J. Bieske, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9540–9548.

40 N. J. A. Coughlan, M. S. Scholz, C. S. Hansen, A. J. Trevitt, B. D. Adamson and E. J. Bieske, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1483–1490.

41 N. J. A. Coughlan, K. J. Catani, B. D. Adamson, U. Wille and E. J. Bieske, *J. Chem. Phys.*, 2014, **140**, 164307–164310.

42 M. S. Scholz, J. N. Bull, N. J. A. Coughlan, E. Carrascosa, B. D. Adamson and E. J. Bieske, *J. Phys. Chem. A*, 2017, **121**, 6413–6419.

43 J. N. Bull, M. S. Scholz, E. Carrascosa and E. J. Bieske, *Phys. Chem. Chem. Phys.*, 2018, **20**, 509–513.

44 B. C. Gau, J. S. Sharp, D. L. Rempel and M. L. Gross, *Anal. Chem.*, 2009, **81**, 6563–6571.

45 A. P. France, L. G. Migas, E. Sinclair, B. Bellina and P. E. Barran, *Anal. Chem.*, 2020, **92**, 4340–4348.

46 S. Vimer, G. Ben-Nissan and M. Sharon, *Nat. Protoc.*, 2020, **15**, 236–265.

47 L. Konermann, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1827–1835.

48 J. M. Christie, A. S. Arvai, K. J. Baxter, M. Heilmann, A. J. Pratt, A. O. Hara, S. M. Kelly, M. Hothorn, B. O. Smith, K. Hitomi, G. I. Jenkins and E. D. Getzoff, *Science*, 2012, **335**, 1492–1497.

49 J. M. Ortiz-Guerrero, M. C. Polanco, F. J. Murillo, S. Padmanabhan and M. Elías-Arnanz, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 7565–7570.

50 U. Choudry, D. J. Heyes, S. J. O. Hardman, M. Sakuma, I. V. Sazanovich, J. Woodhouse, E. De La Mora, M. N. Pedersen, M. Wulff, M. Weik, G. Schirò and N. S. Scrutton, *ChemBioChem*, 2018, **19**, 1036–1043.

51 L. Konermann, E. Ahadi, A. D. Rodriguez and S. Vahidi, *Anal. Chem.*, 2012, **85**, 2–9.

52 Z. Hall, H. Hernández, J. A. Marsh, S. A. Teichmann and C. V. Robinson, *Structure*, 2013, **21**, 1325–1337.

53 J. T. S. Hopper and N. J. Oldham, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1851–1858.

54 L. G. Migas, A. P. France, B. Bellina and P. E. Barran, *Int. J. Mass Spectrom.*, 2018, **427**, 20–28.

55 J. D. Eschweiler, J. N. Rabuck-Gibbons, Y. Tian and B. T. Ruotolo, *Anal. Chem.*, 2015, **87**, 11516–11522.

56 M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, *Anal. Chem.*, 2015, **87**, 4370–4376.

57 B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung and C. V Robinson, *Nat. Protoc.*, 2008, **3**, 1139–1152.

58 B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon and P. E. Barran, *Analyst*, 2014, **139**, 6348–51.

59 A. Theisen, B. Yan, J. M. Brown, M. Morris, B. Bellina and P. E. Barran, *Anal. Chem.*, 2016, **88**, 9964–9971.

60 A. Theisen, R. Black, D. Corinti, J. M. Brown, B. Bellina and P. E. Barran, *J. Am. Soc. Mass Spectrom.*, 2018, **30**, 24–33.

61 R. Black, A. Barkhanskiy, L. A. I. Ramakers, A. Theisen, J. M. Brown, B. Bellina, D. K. Trivedi and P. E. Barran, *Int. J. Mass Spectrom.*, 2021, **464**, 116588.

62 X. Zeng, Z. Ren, Q. Wu, J. Fan, P.-P. Peng, K. Tang, R. Zhang, K.-H. Zhao and X. Yang, *Nat. Plants*, , DOI:10.1038/nplants.2014.6.

63 A. F. M. Gavriilidou, B. Gülbakan and R. Zenobi, *Anal. Chem.*, 2015, **87**, 10378–10384.

64 P. Dugourd, R. R. Hudgins, D. E. Clemmer and M. F. Jarrold, *Rev. Sci. Instrum.*, 1997, **68**, 1122–1129.

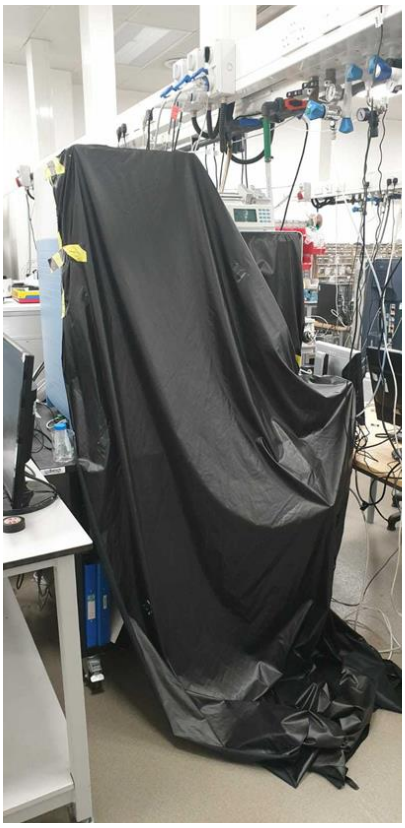
65 R. Salbo, M. F. Bush, H. Naver, I. Campuzano, C. V. Robinson, I. Pettersson, T. J. D. Jorgensen and K. F. Haselmann, *Rapid Commun. Mass Spectrom.*, 2012, **26**, 1181–1193.

66 S. J. Allen, K. Giles, T. Gilbert and M. F. Bush, *Analyst*, 2016, **141**, SI.

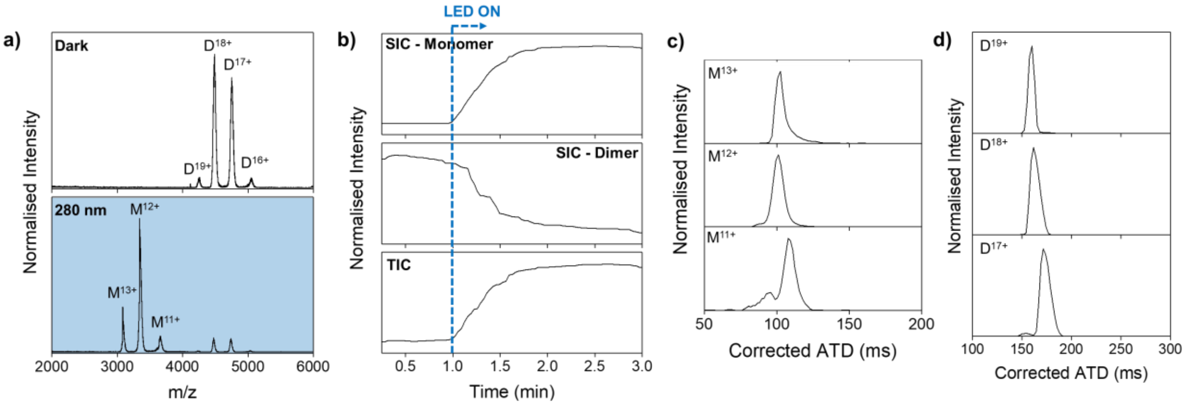
67 M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson and B. T. Ruotolo, *Anal. Chem.*, 2010, **82**, 1–9.

68 H. Hernández and C. V. Robinson, *Nat. Protoc.*, 2007, **2**, 715–726.

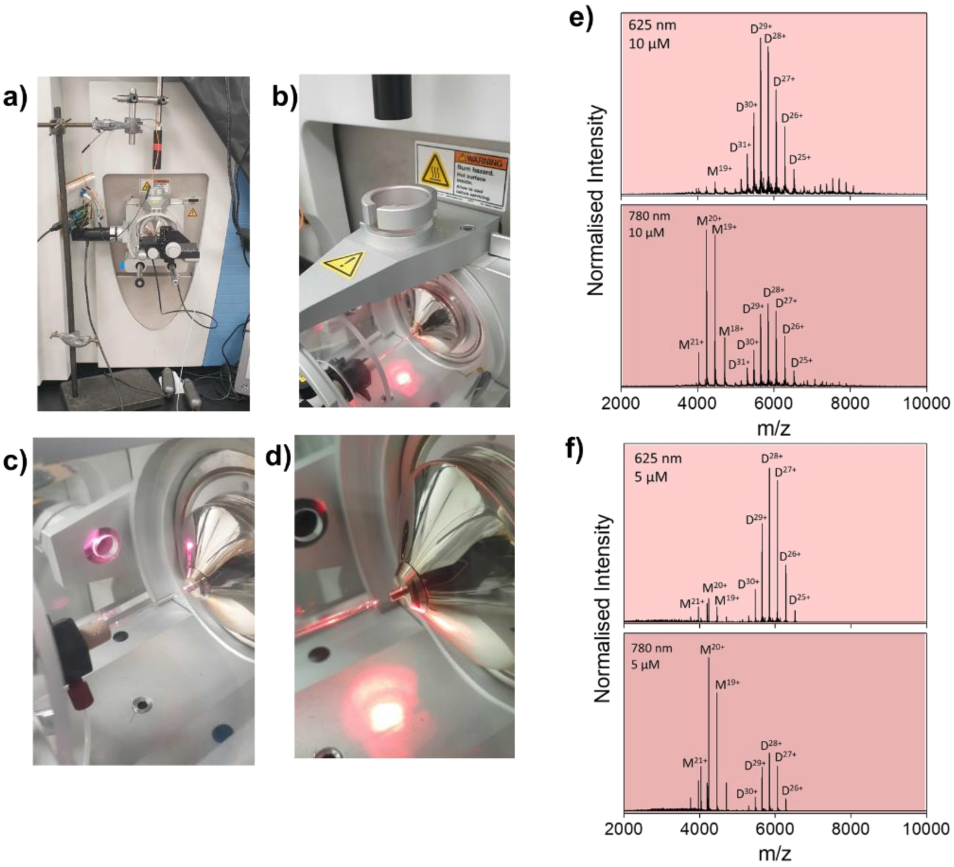
3.12 Supporting information



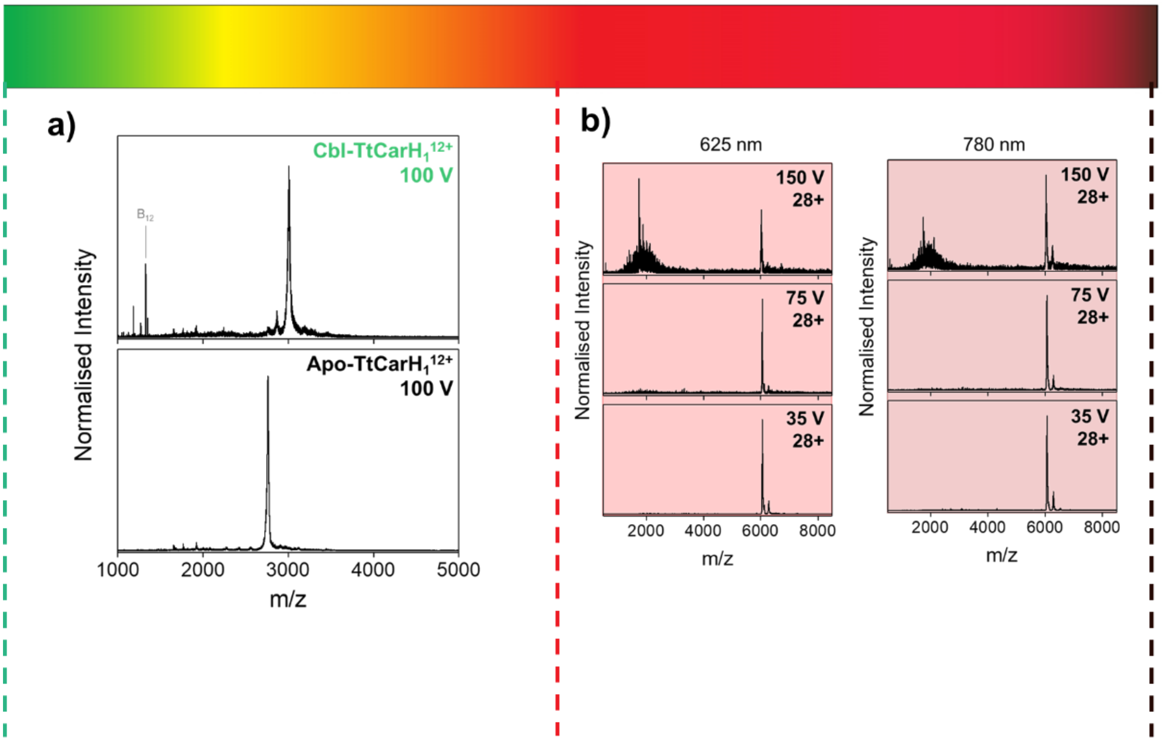
SI Figure 1: Black out curtain around the Q-exactive UHMR UHMR mass spectrometer to block out any laboratory light.



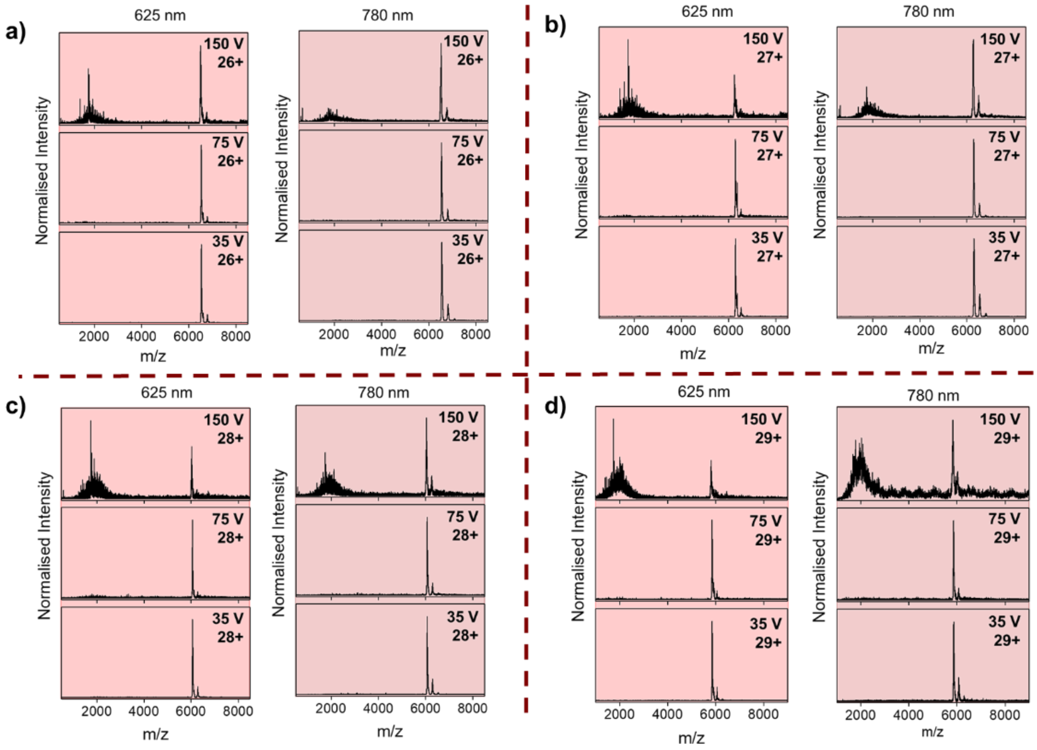
SI Figure 2: Core UVR8 illumination measurements, full mass spectrum under both dark and 280 nm UV illumination (a), the corresponding chromatogram showing the selected ion count for the monomer and dimer along with the total ion count (b), the corrected ATD for the monomer and dimer species (c and d respectively).



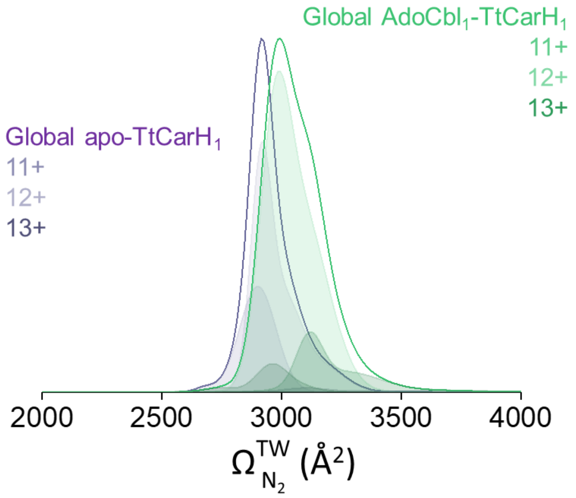
SI Figure 3: Installing the LEDs onto the Q-exactive UHMR UHMR, a) Installation of both a mounted and unmounted LED onto the source block, the mounted LED fits directly into the nESI where the cameras would usually fit, the unmounted is then mounted using a clamp stand to maintain the flexibility and removal of the LED, this is then focussed on other nESI tip within the source. b and d) the unmounted LED directed nESI tip and c) the mounted LED directed onto the nESI tip from the left camera insert. The measurements were performed at (e) 10 µM and (f) 5µM under both 625 nm and 780 nm on the Q-exactive UHMR UHMR mass spectrometer.



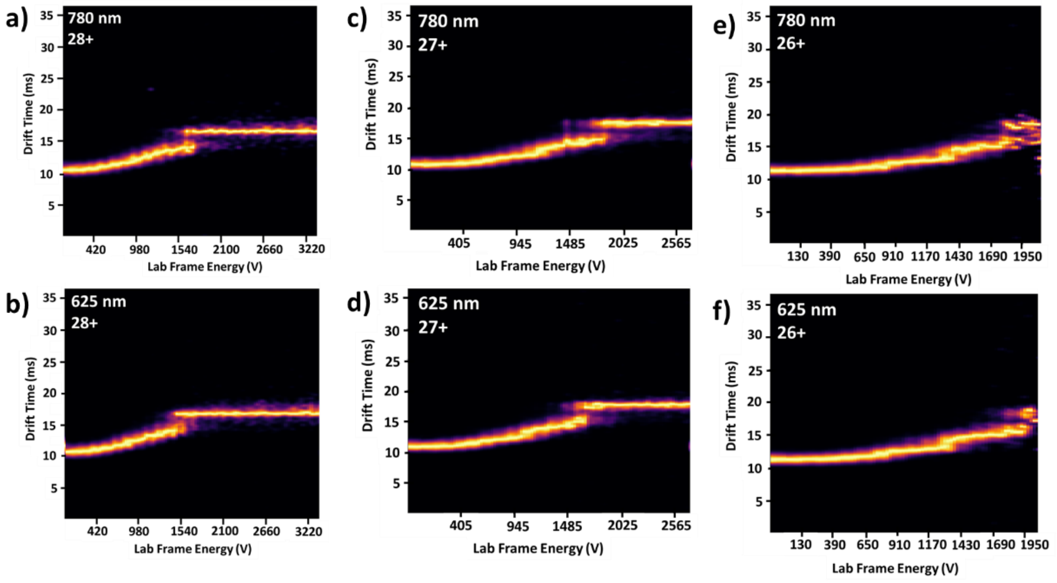
SI Figure 4: Collisional activation of selected ions, WT Tt-CarH 100 V was applied to the monomer species with apo-TtCarH and AdoCb-TtCarH post illumination (a) and R. Palustris under 625 nm and 780 nm, collisional activation was applied at 35 V, 75 V and 150 V to the 28+ dimer species. (b).



SI Figure 5: The R. Palustris phytochrome collision induced dissociation under trap voltage of 35 V, 75 V and 150 V. This is performed on the most prominent charge states that are observed, a) $[M+26H]^{26+}$, b) $[M+27H]^{27+}$, c) $[M+28H]^{28+}$ and d) $[M+29H]^{29+}$ under both 625 nm and 780 nm illumination.



SI Figure 6: The CCSD plot for the WT TtCarH monomer both apo (solid purple) and with the 5'-deoxyadenosyl group bound (solid green) after illumination with 530 nm. Each displayed charge state is shown underneath the global CCSD plot and have been scaled to take the MS peak intensity and peak area.



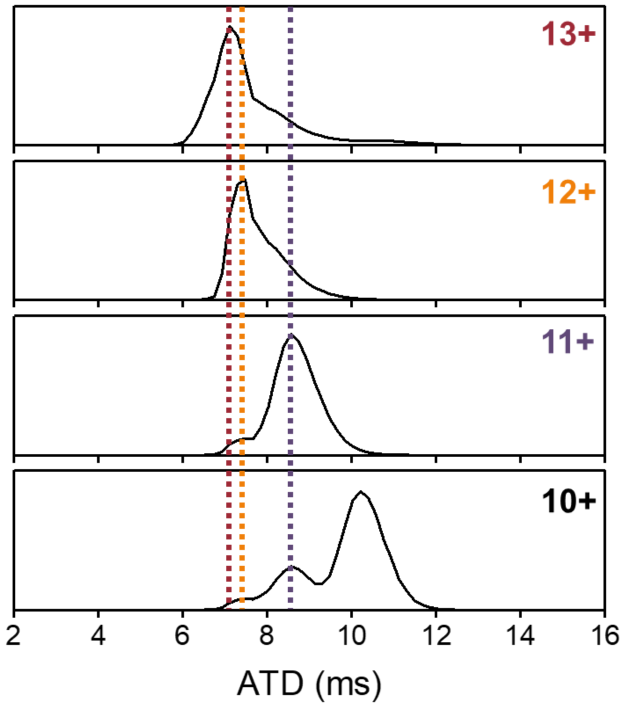
SI Figure 7: Activated ion mobility measurements of the R. Palustris phytochrome, selected for the dimer charge states, 28+ (a and b), 27+ (c and d) and 26+ (e and f) under both red and far red illumination conditions.

Source Voltages and gas flows	
Capillary voltage (kV)	<1.5
Source temperature (^{o}C)	131
Drying gas flow (l/min)	11.6
Fragmentor	350
DC Voltages	
HP Funnel delta	177.1
Trap entrance grid delta	2
Trap entrance grid low	97.4
Trap entrance grid high	99.4
Trap entrance	91.0
Trap exit	89.9
Trap exit grid 1 delta	5
Trap exit grid 1 low	88.3
Trap exit grid 1 high	93.3
Trap exit grid 2 delta	8
Trap exit grid 2 low	87.7
Trap exit grid 2 high	95.7
Trap Funnel delta	165.0
IM Hexapole entrance	40.7
Rear Funnel entrance	239.4
Rear Funnel exit	150.0
RF Voltages (V)	
Octapole	750
High Pressure Funnel	149.0
Trap Funnel	179.0
IM Hexapole	599.0
Rear Funnel	150.0
Pressures (Torr)	
HP Funnel	4.960
Drift tube	3.996
Trap Funnel	3.872

SI Table 1: Agilent 6560 IMS settings for acquisition of UVR8.

Ion Source	
Spray Voltage (kV)	<1.5
Spray Current (µA)	0.15
Capillary Temperature (˚C)	350
Ion Optics	
Bent Flatapole DC (V)	2.2
Inj. Flatapole A DC (V)	5.3
Inj. Flatapole B DC (V)	5.1
Trans Multipole DC (V)	0.1
HCD Multipole DC (V)	-0.6
Inj. Flat. RF Amp (V)	150
Inj. Flar. RF Freq (kHz)	550
Bent Flat. RF Amp (V)	940
Bent Flar. RF Freq (kHz)	1100
RF2 and RF3 Amp (V)	250
RF2 and RF3 Freq (kHz)	770
Inter Flatapole DC (V)	4
Quad Exit DC (V)	-20
C-Trap Entrance Lens DC (V)	1.4
C-Trap RF Amp (V)	2300
C-Trap RF Freq (kHz)	2.2
C-Trap RF Curr (A)	0.4
C-Trap Exit Lens DC (V)	15.6
HCD Exit Lens DC (V)	35
Vacuum	
Fore Vacuum Sensor (mbar)	1.45
High Vacuum Sensor (mbar)	3.39E-09
UHV Sensor (mbar)	5.51E-10
Source TMP Speed	1000
UHV TMP Speed	1200

SI Table 2: Q Exactive settings for the acquisition of photoreceptor proteins.



SI Figure 8: Effect of charge stripping on the ATD of the WT TtCarH apo- monomer when there is no m/z selection. Each of the higher charge states are colour coded and a dotted line represent the peak apex for the main ATD of that charge state, burgundy for 13+, yellow for 12+ purple for 11+. From dotted lines matching with the smaller peak apexes also present in the lower charge states, it is clear that the 12+ ion is also being detected in the 11+ data and both the 12+ and 11+ are being detected in the 10+ data. Here the ions have travelled through the mobility cell as the higher charge state only to lose a charge prior to detection.

4 Interplay between chromophore-binding and domain assembly by the B_{12}-dependent photoreceptor protein, CarH

4.1 Declaration

This chapter consists of one published article:

This article has been reproduced in an unchanged format except for minor adjustments to incorporate it in to this thesis.

As joint first author on this publication I performed all MS and IMS experiments on both WT Tt-CarH and G192Q variant. I compiled and prepared Figures 4.2 and 4.3 alongside SI Figure 4-7 and 9-11. I also drafted an initial MS data manuscript and contributed to the preparation and editing of the final manuscript. IC expressed and purified G192Q and WT Tt-CarH. DJH and SH performed the stopped-flow measurements and LOJ performed all MD simulations and produced corresponding figures.

MS project planning, experimental work and result discussion was performed by myself. This included the experimental approach taken which was developed by myself and BB, experimental approach includes the development of monitoring the increase in concentration of cofactor and the concentration ratios chosen.

Supporting Information

Additional Figures are provided in the Supporting Information (SI) in section 4.12 of this thesis

Chemical Science, DOI: 10.1039/D1SC00522G

Interplay Between Chromophore-Binding and Domain Assembly by the B_{12}-Dependent Photoreceptor Protein, CarH

Inês S. Camacho[1,2,†], Rachelle Black[1], Derren J. Heyes[1], Linus O. Johannissen[1], Lennart A. I. Ramakers[1], Bruno Bellina[1,2], Perdita E. Barran*[1,2], Sam Hay*[1] and Alex R. Jones*[1,2,†]

1 Manchester Institute of Biotechnology and Department of Chemistry, The University of Manchester, 131 Princess Street, Manchester, M1 7DN

2 Photon Science Institute, The University of Manchester, Oxford Road, Manchester, M13 9PL

4.2 Abstract

Organisms across the natural world respond to their environment through the action of photoreceptor proteins. The vitamin B_{12}-dependent photoreceptor, CarH, is a bacterial transcriptional regulator that controls the biosynthesis of carotenoids to protect against photo-oxidative stress. Binding of B_{12} to CarH monomers in the dark results in formation of a homo-tetramer that complexes with DNA; B_{12} photochemistry results in tetramer dissociation, releasing DNA for transcription. Although details of the response of CarH to light are beginning to emerge, the biophysical mechanism of B_{12}-binding in the dark and how this drives domain assembly is poorly understood. Here – using a combination of molecular dynamics simulations, native IMMS and time-resolved spectroscopy – we reveal a complex picture that varies depending on the availability of B_{12}. When B_{12} is in excess, its binding drives structural changes in CarH monomers that result in the formation of head-to-tail dimers. The structural changes that accompany these steps mean they are rate-limiting. The dimers then rapidly combine to form tetramers. Strikingly, when B_{12} is scarcer, as is likely in nature, tetramers with native-like structures can form without a B_{12} complement to each monomer, with only one apparently required per head-to-tail dimer. We thus show how a bulky chromophore such as B_{12} shapes protein/protein interactions and in turn function, and how a protein can adapt to a sub-optimal availability of resources. This nuanced picture should help

guide the engineering of B_{12}-dependent photoreceptors as light-activated tools for biomedical applications.

4.3 Introduction

Vitamin B_{12} is the largest and most structurally complex vitamin in nature.[1] At its center is a highly conjugated cobalamin macrocycle, which enables one active derivative, 5'-deoxyadenosylcobalamin (AdoCbl, Figure S1a), to act as a latent source of radicals for numerous mutase[2] and eliminase[3] enzymes. This extensive conjugation also means it absorbs light from across much of the UV and visible regions of the spectrum, making it an ideal chromophore for environmental light-sensing functions in biology. Although the photochemistry of B_{12} species has been studied for decades, its role in photobiology is only now becoming apparent.[4] The discovery of the bacterial transcriptional regulator, CarH,[5, 6] has brought forth a new area of photobiology based on B_{12} as photoactive chromophore.[7] AdoCbl photochemistry in CarH results in transcriptional activation in bacteria that leads to the biosynthesis of carotenoids in response to photo-oxidative stress. Mechanistic details about its function are now beginning to emerge[4, 6-11] and it is already showing great promise and versatility as the basis of photoactivated, biomolecular tools.[12-16]

The size and structural complexity of AdoCbl reflects a long and expensive biosynthetic pathway,[17] and means its uptake[18] and subsequent binding to its dependent enzymes[19] is often tightly regulated. The importance of these pathways is highlighted by the genetic disorders that are caused by mutations to the regulatory proteins.[20] Similarly, the binding of AdoCbl to riboswitches and other proteins is a crucial aspect of its role as regulator of genetic control elements.[21] In CarH, this is not only because it serves as photoactive chromophore, but also because binding triggers the formation of oligomeric forms that bind to, and thus block, operator DNA. Binding of AdoCbl to CarH from Thermus thermophilus (TtCarH), for example, converts apo-monomers into holo-tetramers that bind DNA.[6] Light absorption by AdoCbl then results in disassembly of this protein / DNA complex and transcriptional activation. We now have some insight into the structural[10] and mechanistic[11, 22] basis of this light-dependent activation. By contrast, to date there has been no biophysical investigation into how binding of AdoCbl drives the assembly of CarH oligomers. This not only limits our understanding of how it achieves transcriptional regulation in nature but also holds back tool development and optimization.

Here, we have probed the binding of AdoCbl to TtCarH and subsequent oligomer assembly using native IMMS, time-resolved spectroscopy and molecular dynamics (MD) simulations. This powerful combination of biophysical techniques has afforded new insights that detail the dynamic interplay between the binding of the chromophore and subunit assembly.

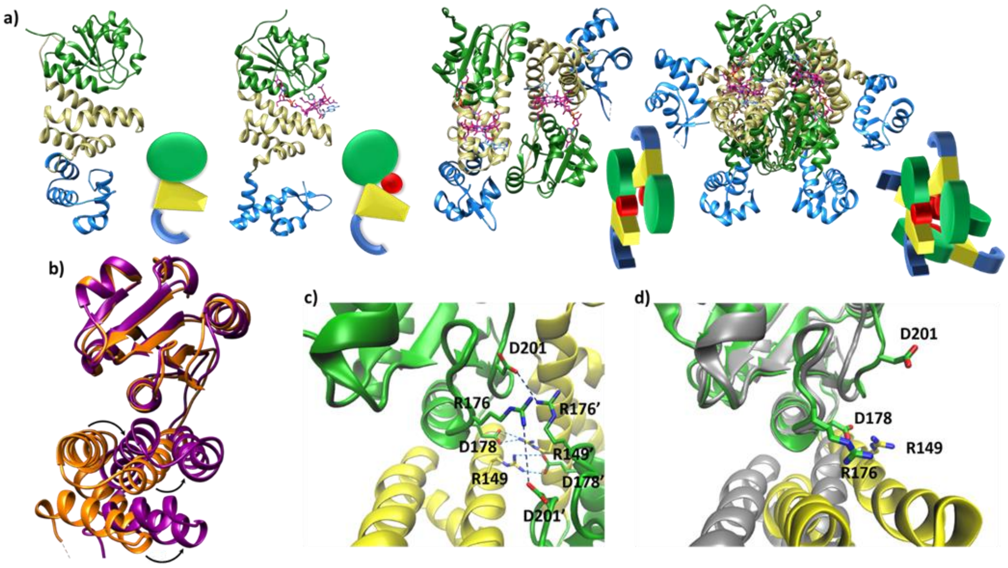


Figure 4.1: a) TtCarH structures and their cartoon representations, colored by domain: B_{12}-binding (green), 4-helix bundle (yellow), DNA-binding (blue), and AdoCbl (magenta). L-R: apo-monomer (representative from MD simulations), holo-monomer, holo-dimer, and holo-tetramer (all PDB: 5C8D). b) 4-helix bundle (bottom) positions in each monomer aligned to the B_{12}-binding domain (top): holo-TtCarH (orange, PDB: 5C8D) and apo-TtCarH (purple, simulated). A different color scheme to panel (a) is used to distinguish between different structures. Arrows indicate helical displacement in apo-TtCarH relative to holo-TtCarH (see also Figures S3a&b). c) Close-up of the interface between monomer units in the holo-TtCarH head-to-tail dimers with salt-bridges identified here (D178—R149) and previously10 (D201—R176) highlighted. d) The same interfacial region illustrated in panel c, but now of the simulated apoTtCarH monomer. The change in 4-helix bundle conformation relative to holo-TtCarH (in grey) results in the key salt-bridge residues moving into positions where they are no longer able to stabilize the head-to-tail dimers (see also Figure S3c).

4.4 Results

4.4.1 AdoCbl binding drives the structural changes in TtCarH that facilitate oligomerization

Published structural data have revealed much about the tertiary and quaternary structure of TtCarH.[10] Each monomer comprises a N-terminal DNA-binding domain, a four helix bundle and a Rossman fold (B_{12}-binding) domain at the C-terminus (Figure 4.1a). Although the quaternary structure is formally a homo-tetramer when AdoCbl is bound, it is perhaps better described as a dimer of head-to-tail dimers (Figure 4.1a). AdoCbl binds to the Rossman fold in a conformation where the lower axial 5,6-dimethylbenzimidazole base is displaced and the Co is instead coordinated by H177 (Figure S1b).[6, 10] The position of AdoCbl between the Rossman fold and the helix bundle places the upper axial 5'-deoxyadenosyl ligand (Ado) into steric contact with W131 from the helix bundle (Figure S1b).10 This is thought to force each holo-TtCarH monomer to adopt an upright conformation that facilitates the formation of the head-to-tail dimers. In the absence of structural data for apo-TtCarH, however, this hypothesis has not been confirmed.

The lack of structural data for apo-TtCarH is presumably because it is prone to precipitation at higher concentrations. We have therefore performed MD simulations (see Experimental Section and Figure S2) of apo-TtCarH that confirm the above hypothesis. From what is known about the photoconversion mechanism,[10, 11] when the Ado is photo-dissociated, W131 moves into the vacated space, resulting in a displacement of the four helix bundle (Figure S3a) and formation of a bis-histidine adduct with H132 (Figure S1b).[10, 11] It appears that movement of the 4-helix bundle then disrupts the head-to-tail dimer interface, thus driving tetramer dissociation. Our MD simulations of apo-TtCarH confirm that in the absence of the entire AdoCbl the 4-helix bundle is displaced relative to the dark structure (8.14 ± 1.33 Å) akin to when Ado is photo-dissociated (9.7 Å) but to a slightly different position (Figure 4.1b, Figure 4.3a & b).

It might seem surprising that such a significant structural change occurs on the relatively short timescale of these MD simulations (400 ns). This would be a reasonable concern if the simulations started from a near-equilibrium structure, which is not the case here. Briefly, the crystal structure of photo-converted holo-TtCarH (i.e., with the Ado missing; PDB: 5C8F)[10] was taken and the rest of the cobalamin was then removed in silico. Three 400 ns simulations were run in parallel (Figure S2b) after energy minimization. As we have demonstrated in previous MD simulations,[11] removing just the 5’-deoxyadenosyl ligand from the holo-TtCarH causes a rapid conformational change due to the steric strain of the protein pushing against the Ado. Removing the entire cobalamin therefore produces a starting structure that is far from equilibrium, which results in a conformational change on the simulation timescale.

In the published structure of holo-TtCarH,[10] two charged residues, R176 and D201, were identified at the surface of each monomer unit of holo-TtCarH, which form salt-bridges that stabilize the head-to-tail dimers. We have identified two further residues, R149 and D178, which could feasibly fulfill a similar role (Figure 4.1c). In the photoconverted state these residues are no longer in a position to readily form salt-bridges. Our MD simulations structures suggest a similar situation for apo-TtCarH where the salt-bridge residues are displaced away from the dimer interface relative to the holo-TtCarH structure (Figure 4.1d and S3c). Our simulations therefore provide strong additional evidence in support of the hypothesis that binding of AdoCbl to TtCarH triggers the structural change that facilitates oligomer formation.

4.4.2 For TtCarH tetramers to form, it is not necessary for an AdoCbl to be bound to each monomer

Our native IMMS data indicate that, although AdoCbl-binding drives tetramer formation, each tetramer does not need to comprise four holo-monomers. Data were acquired for wild-type (WT) TtCarH samples (10 µM apo-protein) containing AdoCbl ranging from sub-stoichiometric quantities to a two-fold excess (Figure 4.2a). In the absence of AdoCbl, the spectrum is dominated by signals from the apo-monomer with a low population of apo-dimer, both presenting over narrow charge state distributions. Ion mobility data show little variance in the collision cross sections ($^{TW}CCS_{N2}$) across the monomer charge states, implying these are compact and homogenous forms (Figure S4a). With increasing concentrations of AdoCbl, the apo-monomer is gradually replaced by signals predominantly from tetrameric species (Figure 4.2a). The holo-tetramer with a full complement of AdoCbl has a mass of ~ 140.5 kDa. Strikingly, at lower TtCarH:AdoCbl ratios, tetramer populations are observed with sub-stoichiometric

AdoCbl; i.e., two (~137.3 kDa) or three (~138.9 kDa) AdoCbl per tetramer (Figure 4.2b). Each form presents with a charge state distribution similar to the tetramer with four AdoCbl bound (i.e., 23 to 26+; Figure 4.2c, S4 and S5). This observation indicates that, when the B_{12} chromophore is scarce, tetramers form without a full complement of AdoCbl.

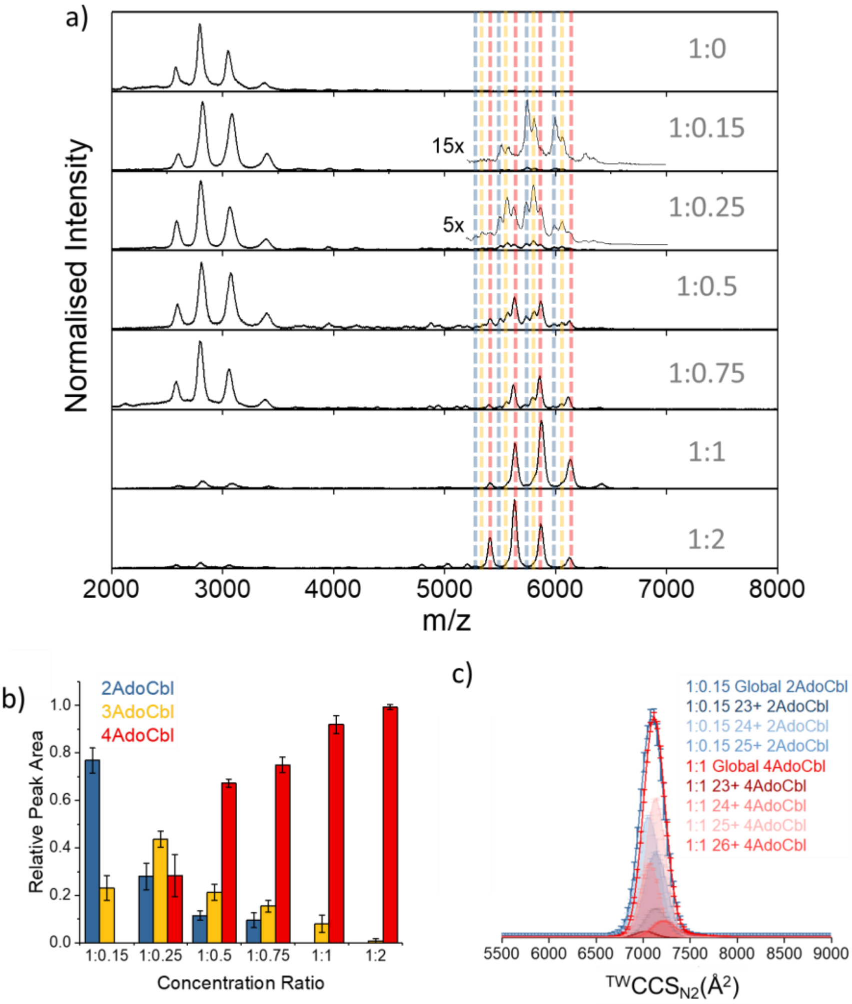


Figure 4.2: a) Native mass spectra of WT TtCarH with increasing (top to bottom) molar ratios of AdoCbl. Vertical dashed lines highlight peaks from different tetramer species and are color-coded as indicated in panel (b). b) Relative peak area as a function of WT TtCarH:AdoCbl ratio for the mass spectral signals highlighted in panel (a) that correspond to different WT TtCarH tetramer species. c) $^{TW}CCS_{N2}$ distributions from ion mobility data for the major tetramer species present at both 1:0.15 (blue, $AdoCbl_2$-$TtCarH_4$) and 1:1 (red, $AdoCbl_4$-$TtCarH_4$). Each global $^{TW}CCS_{N2}$ is the summation of the various charge states, corrected for their peak area and height.

For the sake of clarity, we will now use a systematic nomenclature to define different TtCarH oligomers with different B_{12} stoichiometries. Protein with no B_{12} bound will be referred to as apo-TtCarHx, where x signifies the oligomeric state (e.g., apo-TtCarH$_2$ for dimers). For the holo-protein we will also indicate the type and number of B_{12} species bound (i.e. either AdoCbl for adenosylcobalamin or MeCbl for methylcobalamin): B_{12}y-TtCarHx (e.g., AdoCbl$_2$-TtCarH$_4$ for tetramers with two AdoCbl bound).

As the concentration of AdoCbl is increased, AdoCbl$_4$-TtCarH$_4$ (~140.5 kDa) dominates, suggesting this is the thermodynamically favored form. Perhaps surprisingly, the narrow, invariant charge state distributions (Figure 4.2a) and ion mobility data (Figure 4.2c, S4 and S5) suggest there is little conformational variation between AdoCbl$_2$-TtCarH$_4$, AdoCbl$_3$-TtCarH$_4$ and AdoCbl$_4$-TtCarH$_4$. Each has a comparable $^{TW}CCS_{N2}$ distribution, consistent with all tetrameric forms adopting a similar quaternary arrangement. This fact, alongside the absence of signal from AdoCbl$_1$-TtCarH$_4$, suggests that one AdoCbl$_1$-TtCarH$_1$ is enough to provide a structural ‘template’ for a partnering apo-TtCarH$_1$ to adopt the correct conformation in each head-to-tail dimer for tetramers to form.

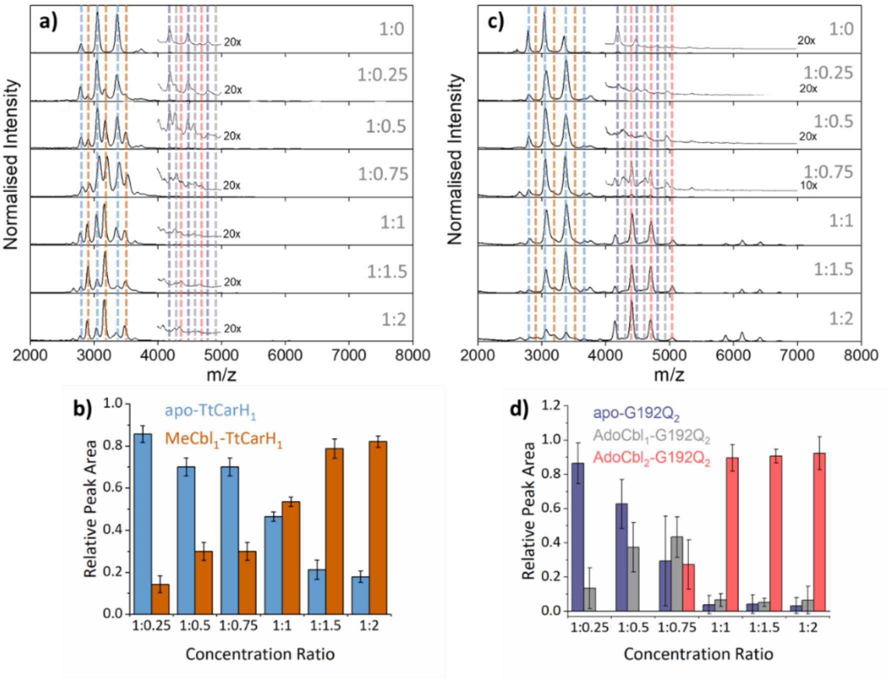


Figure 4.3: a) Native mass spectra of WT TtCarH with increasing (top to bottom) molar ratios of MeCbl. Vertical dashed lines highlight peaks from different monomer and dimer species, which are color-coded as indicated in panel (b) and Figure S9, respectively. b) Relative peak area as a function of WT TtCarH:MeCbl ratio for the mass spectral signals highlighted in panel (a) that correspond to different WT TtCarH monomer species. c) Native mass spectra of G192Q with increasing (top to bottom) molar ratios of AdoCbl. Vertical dashed lines highlight peaks from different monomer and dimer species, which are color-coded as indicated in panel (d) and Figure S10, respectively. d) Relative peak area as a function of G192Q:AdoCbl ratio for the mass spectral signals highlighted in panel (c) that correspond to different G192Q dimer species.

It is also possible that AdoCbl binds directly to apo-TtCarH$_2$. The low intensity dimer signals in Figure 4.2a are from apo-TtCarH$_2$, AdoCbl$_1$-TtCarH$_2$ and AdoCbl$_2$-TtCarH$_2$ (Figure S6). Moreover, there is only the slightest suggestion of signal from AdoCbl$_1$-TtCarH$_1$, and only when AdoCbl is in a two-fold excess (Figure S7). This last observation is either consistent with preferential binding to apo-TtCarH$_2$ or that AdoCbl binding to the apo-TtCarH$_1$ simply drives the position of equilibrium overwhelmingly towards AdoCbl$_2$-TtCarH$_2$. Either way, the low population of dimers suggests that when holo-dimers do form, they rapidly combine to form tetrameric species. This is supported by the fact that the relative populations of the dimer species follow a different pattern with increasing AdoCbl concentration to the various tetramer species (Figure S6). Here, apo-TtCarH$_2$ remains a significant sub-population, including when AdoCbl is in excess.

4.4.3 B$_{12}$ binds to both monomeric and dimeric TtCarH

To probe the role of dimer intermediates further, we investigated the binding of methylcobalamin (MeCbl), again using native IMMS with samples containing the same range of TtCarH:B$_{12}$ ratios. Although MeCbl is known to bind to TtCarH, size exclusion chromatography (SEC) and isothermal calorimetry suggest the protein remains in the monomeric form.[6] This is almost certainly because MeCbl lacks both the steric bulk of AdoCbl in the upper axial position (Figures S1a and S8) to cause the structural changes that drive TtCarH oligomerization and the capacity to form any necessary stabilizing interactions made by the Ado group. As before, in the absence of MeCbl the mass spectrum of TtCarH contains signals mainly from apo-TtCarH1, with a low population of apo-TtCarH$_2$ (Figure 4.3a). As the MeCbl concentration increases, the apo-TtCarH$_1$ population is displaced by signals from MeCbl$_1$-TtCarH$_1$ (Figure 4.3a and 3b) and apo-TtCarH$_2$ evolves to MeCbl$_1$-TtCarH$_2$ and MeCbl$_2$-TtCarH$_2$ (Figure 4.3a and S9). These data confirm that B$_{12}$ species are able to bind directly to both apo-TtCarH$_1$ and apo-TtCarH$_2$. Across the range of ratios, however, the overall dimer populations remain very low. The implications of this are two-fold. First, direct binding to apo-TtCarH$_2$ is a subsidiary as opposed to preferential route. Second, unlike for AdoCbl, MeCbl binding to apo-TtCarH$_1$ does not shift the position of equilibrium towards the dimeric form. The fact that MeCbl-bound dimers have a more significant population relative to apo-TtCarH$_2$ than the equivalent AdoCbl-bound species is consistent with their populations not being depleted to form tetramers.

We also investigated the binding of the native chromophore, AdoCbl, to a TtCarH variant that predominantly forms dimers rather than tetramers. G192Q places bulky glutamine residues at the interface between head-to-tail dimers, which sterically hinder formation of tetramers following chromophore binding.[10] The mass spectra (Figure 4.3c) reveal that the apo-G192Q$_1$ (main population) and apo-G192Q$_2$ (minor population) signals are replaced upon increased AdoCbl concentration with signals not only from AdoCbl$_2$-G192Q$_2$ as expected (the dominant species at a 1:1 ratio or in excess) but also AdoCbl$_1$-G192Q$_1$ and AdoCbl$_1$-G192Q$_2$. Again, these data are consistent with B$_{12}$ species binding directly to both apo-TtCarH$_1$ and apo-TtCarH$_2$. In a similar way to the WT tetramers with fewer than four AdoCbl bound (Figure 4.2a), AdoCbl$_1$-G192Q$_2$ makes up a significant population of dimers when AdoCbl is at low concentration (Figure 4.3d). There is a more significant sub-population of AdoCbl$_1$-G192Q$_1$ (Figure S10a) when compared to WT (Figure S7), which becomes increasingly apparent at higher AdoCbl concentrations. This indicates a dynamic equilibrium between oligomeric states, which is highly dependent both on AdoCbl concentration and on the stability of the tetramer. This would explain why there is little evidence of AdoCbl$_1$-TtCarH$_1$ in the WT protein – where the tetramer is presumably highly stable – than in the G192Q 'dimer' variant. Tetramers with a slightly smaller ^{TW}CCS$_{N2}$ than the WT protein are observed for G192Q, suggesting a stable, compact form, but only when AdoCbl is around stoichiometric concentrations relative to G192Q, or in excess (Figure 4.3c, S10b and S11b). These are therefore likely to be non-specific in their formation and only form when AdoCbl$_2$-G192Q$_2$ is the dominant species.

Taken together, our MS data reveal that TtCarH tetramers can form with fewer than four AdoCbl bound, but that at least one AdoCbl appears to be required per head-to-tail dimer. The ion mobility data suggest these forms adopt structures similar to AdoCbl$_4$-TtCarH$_4$. They are also in significant populations when AdoCbl is at sub-stoichiometric concentrations relative to the protein, which is not an unlikely scenario in vivo considering the expense of AdoCbl biosynthesis.[17] We can also conclude that the head-to-tail dimers are able to pre-assemble in a form that can subsequently bind AdoCbl as an additional route to formation of the active complex.

4.4.4 AdoCbl binding to TtCarH triggers oligomerization

The mass spectral data presented above have provided good evidence for the arrangement of protein units and chromophores that are necessary for AdoCbl binding and domain assembly. These are equilibrium measurements, however, and questions remain about the pre-equilibrium mechanism. To examine this, we conducted time-resolved fluorescence measurements using stopped-flow spectroscopy. When AdoCbl is titrated into a sample of TtCarH, the protein emission following excitation at 280 nm (predominantly from the five tryptophan residues in each TtCarH monomer, Figure S12a) is significantly quenched (Figures S12b-c and S13a and Table S1). Because of the spectra overlap (Figure S12b) between the tryptophan emission and AdoCbl absorption when bound to TtCarH, we predicted that Förster resonance energy transfer (FRET) between the two chromophores will be the dominant quenching mechanism and that the emission from tryptophans within ~ 40 Å of each AdoCbl is likely to be quenched with reasonable efficiency (Figure S12c and Table S1). As will become apparent, this quenching provides a useful means of probing the kinetics and mechanism of chromophore binding and oligomerization following rapid mixing of B_{12} species and apo-TtCarH in a stopped-flow (Figure S13b).

When WT TtCarH is rapidly mixed with a ≥ ten-fold excess of AdoCbl (*i.e.*, pseudo-first order conditions) the fluorescence signal is quenched over the course of ~ 30 s (Figure 4.4a, green trace). The data reveal two kinetic phases, the first of which has an apparent rate (k_{app}) that is linearly dependent on the concentration of AdoCbl (Figure 4.4b). It thus represents a bimolecular reaction involving AdoCbl – *i.e.*, its binding to apo-TtCarH – which occurs with a second-order rate coefficient of 34.3 ± 1.4 $\mathrm{s}^{-1}\ \mathrm{mM}^{-1}$. By contrast, the k_{app} of the second, slower phase is independent of AdoCbl concentration (Figure 4.4b) and presumably therefore corresponds to protein domain assembly steps. If so, this clearly shows that AdoCbl binding drives oligomerization of TtCarH.

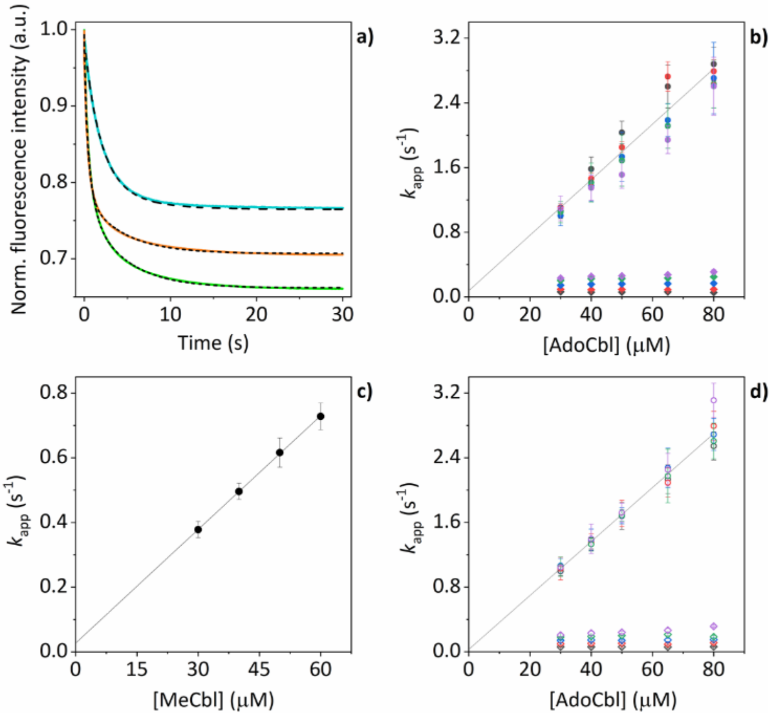


Figure 4.4: a) Example stopped-flow traces (solid lines) and corresponding fits (dashed lines) of the fluorescence quenching that follows the rapid mixing of: WT TtCarH vs. AdoCbl (green); WT vs. MeCbl (blue); G192Q vs. AdoCbl (orange). In each case: [protein] = 5 µM; $[B_{12}]$ = 60 µM. b) Apparent rate (k_{app}) as a function of [AdoCbl] for a range of [WT TtCarH] (see Table S2). k1 is [AdoCbl]-dependent and the data from all [TtCarH] were concatenated and fit linearly. k2 is not [AdoCbl]-dependent. c) k_{app} as a function of [MeCbl] for WT TtCarH and fit linearly. d) As panel (b) in all aspects but data are from the G192Q variant.

This proposal is supported by the equivalent data from mixing of apo-TtCarH with excess MeCbl. Here, the fluorescence quenching only follows a single phase (Figure 4.4a, blue trace), the k_{app} of which is dependent on MeCbl concentration (Figure 4.4c). The amplitude of this single phase is also smaller than that following AdoCbl binding. These observations are consistent with the data only corresponding to binding of MeCbl to apo-TtCarH with little subsequent protein domain assembly (and corresponding quenching), as is expected. In fact, the quenching amplitude for MeCbl is smaller than the first phase from the AdoCbl data that corresponds to chromophore binding. This is almost certainly because the close steric contact between Ado and W131 (Figure S1b) and any stabilizing interactions are absent when MeCbl is bound. One might expect the lesser steric bulk of the upper axial ligand of MeCbl would make binding more rapid, but the second order rate (11.7 ± 0.1 s^{-1}) is three-fold slower than for AdoCbl binding. This result instead suggests that the polar Ado aids with binding to the protein, perhaps helping anchor it in the correct position for favorable binding of the corrin macrocycle. The stopped-flow data for the WT protein and AdoCbl resolve only a single kinetic phase for what we assume corresponds to protein assembly steps. One might expect, however, two or more phases for tetramer assembly if it happens in a stepwise manner. This could indicate one of several things: i) that AdoCbl binds preferentially to apo-TtCarH$_2$; ii) that, following binding of AdoCbl to apo-TtCarH$_1$, the tetramer assembles in a concerted manner; iii) that the dimer to tetramer step is spectrally silent or cannot be kinetically resolved. These possibilities will be explored below.

4.4.5 Protein domain assembly is rate-limited by dimerization

We next conducted equivalent stopped-flow measurements with the G192Q variant, which, as we have seen, predominantly forms dimers following AdoCbl binding. Upon mixing with AdoCbl, the fluorescence quenching again follows two kinetic phases with many features in common with the data from the WT protein (Figure 4.4a, orange trace). The first phase is again dependent on AdoCbl concentration (Figure 4.4d), with a second order rate (33.4 ± 0.8 s^{-1} mM^{-1}) that is the same within error as that of the WT protein. The initial binding of AdoCbl is therefore kinetically equivalent for both WT and G192Q variants (as is binding to MeCbl, Figure S14). The second phase is also independent of AdoCbl concentration (Figure 4.4d), with very similar k_{app} to those measured for the WT and is therefore likely to represent protein domain assembly as before.

Interestingly, however, the amplitude of the fluorescence quenching for G192Q is reduced compared to when AdoCbl binds to WT TtCarH but is greater than when MeCbl binds to the WT (Figure 4.4a). Looking at the amplitudes of each phase more closely (Figure S15), one can see that the amplitudes for the AdoCbl-binding step are very similar between WT and G192Q. The same is not true for the second phase; here, the amplitude for G192Q – where for the vast majority of the population oligomerization stops at the dimer – is roughly half the amplitude for the WT (Figures S15 and S16). This strongly suggests that the data can, in part, resolve the protein monomer to dimer step on the one hand and the dimer to tetramer step on the other, based on differences in spectral amplitude. The similarity between the k_{app} of the second phase for each variant is consistent with the monomer to dimer step being rate-limiting and the dimer to tetramer step being relatively rapid; so much so that, although spectrally resolved, it is not kinetically resolved.

4.4.6 The TtCarH tetramer assembles predominantly via a 1 → 2 → 4, stepwise mechanism

Before coming to any firm conclusions about the kinetics and mechanism of protein domain assembly, we first need to confirm that the second kinetic phase arises from the TtCarH oligomerization process. One can only infer this indirectly from the stopped-flow data in Figure 4.4 because they are presented as a function of AdoCbl concentration. Because domain assembly is a multi-molecular event involving the protein, the kinetics should therefore be dependent on the concentration of TtCarH. Stopped-flow data as a function of protein concentration show this to be the case (Figure 4.5). The first kinetic phase for both the WT (Figure 4.5a) and G192Q (Figure 4.5c) variants is independent of protein concentration. Although AdoCbl binding is a bimolecular process involving the protein, the B_{12} species is at saturating concentrations ($\geq 10\ \times$) so the independence of the kinetics on the much more dilute protein concentration is to be expected. The kinetics of the second phase by contrast show a strong dependence on protein concentration for each variant (Figure 4.5b and d). In both cases, this dependence is not perfectly linear, which is probably caused by a small but significant inner-filter effect at higher protein concentrations (Figure S17). Despite this slight artefact, linear fits give a second order rate for the WT (50.4 ± 2.3 s^{-1} mM^{-1}) that is slightly faster than for G192Q (45.1 ± 1.9 s^{-1} mM^{-1}). If this marginal difference is significant, it might be because the WT data do not reflect a truly second order process and that the non-linearity here is to some extent also owing to ‘contamination’ from a higher order process; i.e., dimer to tetramer transition.

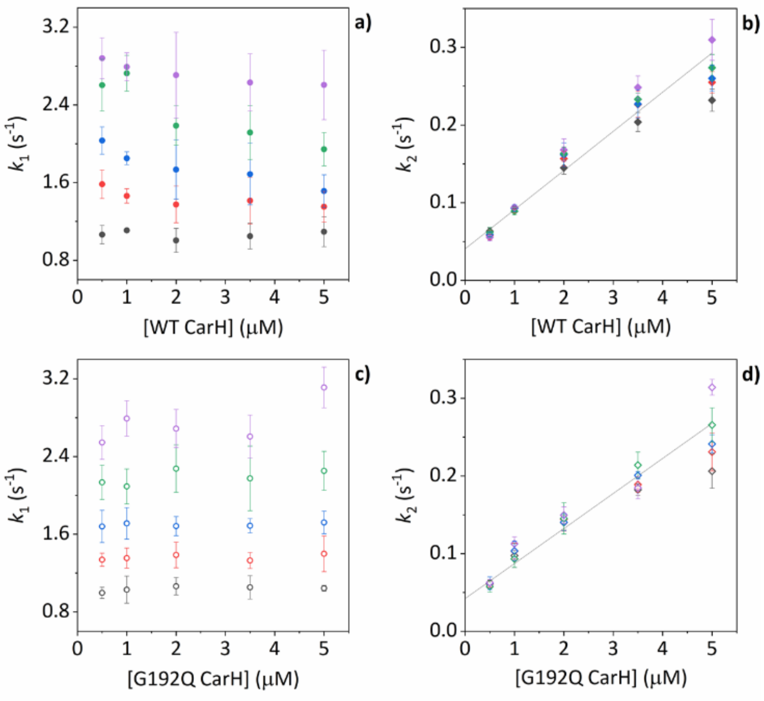


Figure 4.5: a-b) k_{app} as a function of [WT TtCarH] for a range of [AdoCbl] (see Table S2). k1 is not [WT]-dependent (a). k2 is [WT]-dependent and the data from all [AdoCbl] were concatenated and fit linearly (b). c-d) As panel (a-b) in all aspects but data are from the G192Q variant.

Taken together, these stopped-flow data are consistent with a stepwise mechanism, where AdoCbl predominantly binds to apo-TtCarH$_1$, which drives the formation of the AdoCbl$_2$-TtCarH$_2$. Owing to the structural changes involved in each of these steps they are both to some extent rate-limiting. The subsequent association of two sets of AdoCbl$_2$-TtCarH$_2$ to form AdoCbl$_4$-TtCarH$_4$ is then relatively rapid because no further structural changes are necessary.

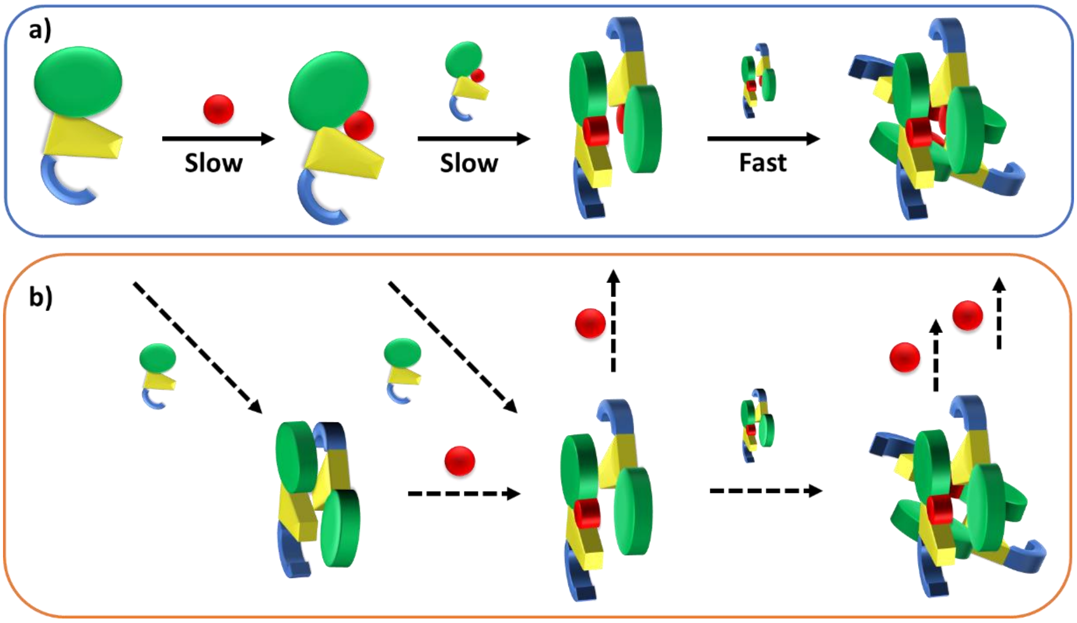


Figure 4.6: Cartoons illustrating the B_{12}-binding and domain assembly events for TtCarH. The domains are color-coded in the same way as Figure 4.1a, B_{12}-binding (green), helix bundle (yellow), DNA-binding (blue), and B_{12} (magenta/red). a) Proposed, stepwise mechanism under 'ideal' conditions; i.e., when AdoCbl is in excess. 'Slow' and 'Fast' indicate whether a step is rate-limiting or kinetically unresolved, respectively. b) Additional binding and assembly steps that provide alternative, more convoluted pathways from (a) under conditions where AdoCbl is scarcer. Dashed arrows indicate alternative pathways to the 'ideal' scheme illustrated in panel (a).

4.5 Discussion

There is mounting evidence that AdoCbl binds to various genetic control elements to achieve transcriptional regulation and that this is widespread in prokaryotes.[21] In both Gram-positive and Gram-negative bacteria, AdoCbl negatively regulates its own biosynthesis independently of light by directly binding to regions of messenger RNA known as riboswitches.[23] A light-dependent role for AdoCbl became apparent from studies into the regulatory circuits in the Gram-negative Myxococcus xanthus that control the stimulation of carotenoid biosynthesis by blue light.[24] A combination of AdoCbl and CarH from M. xanthus (MxCarH) was found to be necessary for the down-regulation of a light-inducible promoter.[5, 6] This repression is relieved under white light in M. xanthus but not by red light alone,[6] consistent with the absorption spectrum of AdoCbl when bound to TtCarH. Elsewhere, it has been demonstrated that B_{12} is an essential element in the control of bacteriochlorophyll biosynthesis in the photosynthetic purple bacteria, Rhodobacter capsulatus.[25] Here, aerobic photochemistry of AdoCbl results in its binding to the ‘aerobic repressor’ protein AerR, which together act as an anti-repressor to the tetrapyrrole regulator CrtJ.

It is now clear that the function of both CarH and AerR is mediated through changes to protein-protein interactions, which are mediated by both the binding of B_{12} and its photochemistry. The anti-repressor activity of AerR appears to result from the light-triggered formation of an AerR/B_{12}/CrtJ complex, which precludes the binding of CtrtJ to the bchC promoter.[25] In contrast, yeast two-hybrid analysis indicates that MxCarH exists as an oligomer in the dark in vivo,[5] and E. coli two-hybrid data are consistent with the C-terminal domain of TtCarH (TtCarHCt) self-interacting only in the presence of AdoCbl. SEC reveals that, whereas TtCarHCt elutes as the holo-tetramer, the H177A variant, which cannot bind any B_{12}, elutes as an apo-monomer, even in the presence of excess AdoCbl.[6] These in vitro data confirmed that oligomerization is facilitated via direct interaction between AdoCbl and the CarHCt. Both in vivo and in vitro, formation of CarH tetramers is impaired by exposure to light of wavelengths < 600 nm,6 consistent with a role for AdoCbl photochemistry.[11]

Here, we present the first detailed, biophysical investigation into the molecular mechanism of how the binding of AdoCbl drives TtCarH oligomerization. Our MD simulations give the first direct evidence that apo-TtCarH does not adopt the correct configuration to form head-to-tail dimers, which was only previously inferred indirectly from structural data of $AdoCbl_4$-$TtCarH_4$.[10] A key mutation of a salt-bridge residue (D201R) at the head-to-tail dimer interface, is known to completely abolish tetramer formation in the presence of AdoCbl (i.e., it elutes in SEC as a holo-monomer).[10] Our simulations reveal that this and other possible stabilizing interactions are also unlikely to form in the apo-$TtCarH_1$.

Published SEC and analytical ultracentrifugation (AUT) data are consistent with apo-TtCarH being only a monomer, with no evidence of higher molecular weight species.[8] Native MS has a significantly higher resolution, however, and our data suggest a more nuanced picture (Figure 4.2 and Figure 4.3). We reveal a low population of apo-$TtCarH_2$ that appears to be receptive to AdoCbl-binding. Interestingly, apo-CarH from Bacillus megaterium has also been shown to self-interact independently of AdoCbl, but to a significantly greater extent than for TtCarH.[9] apo-BmCarH elutes in SEC as a molten globule (i.e., loosely packed) oligomeric species with a molecular weight consistent with a trimer, although these oligomeric states do not bind DNA. AdoCbl-binding to BmCarH induces tetramer formation, which is much more rigidly structured (and binds DNA in vivo) and dissociates into dimers instead of monomers upon illumination. To some extent it appears that TtCarH hedges its bets (Figure 4.6): a majority population binds to AdoCbl as apo-$TtCarH_1$, but we propose that a minority partially pre-assembles as apo-$TtCarH_2$ providing an alternative channel to oligomerization. This inclination to self-associate either in the absence of AdoCbl or at low AdoCbl concentrations is further illustrated by the fact that TtCarH oligomers form with only sub-stoichiometric levels of AdoCbl bound (Figure 4.2, Figure 4.3, S6 and S10). Although we do not know with precision what AdoCbl concentrations are found in bacteria, it seems unlikely that AdoCbl would be in a significant excess over TtCarH and its concentration might be close to its K_D for binding to the protein (255 nM).[11] It is therefore possible that a more fragmentary assembly process is likely under natural conditions, where lower order oligomers form initially with less than a full complement of AdoCbl, before the fully active $AdoCbl_4$-$TtCarH_4$ state is reached (Figure 4.6b).

Published structures[10] of $AdoCbl_4$-$TtCarH_4$ in the dark reveal a quaternary structure that might suggest a stepwise assembly process. This is borne out by our stopped-flow data (Figure 4.4 and Figure 4.5). We clearly resolved kinetic phases that are consistent first with AdoCbl binding predominantly to apo-$TtCarH_1$, which then drives formation of the head-to-tail $AdoCbl_2$-$TtCarH_2$. To similar extents both steps are rate-limiting to the overall process. It appears that once the corresponding structural changes have happened, the assembly of the thermodynamically stable $AdoCbl_4$-$TtCarH_4$ is relatively fast; presumably the energy barrier to this final step is low. Our proposed mechanism for chromophore binding and domain assembly for TtCarH is in Figure 4.6. We present two schemes: Figure 4.6a is a simple mechanism that is likely to take place under ‘ideal’ conditions, where AdoCbl is in a large excess. It shows a 1 → 2 → 4, stepwise assembly mechanism that is consistent with the stopped-flow data and intuitive from the point of view of what is known about the quaternary structure of TtCarH. Figure 4.6b shows the more fragmented picture with intermediates supported by the mass spectral data, which might better reflect what occurs under natural conditions when the TtCarH and AdoCbl concentrations are likely to be both similar to K_D.

4.6 Experimental section

Experimental details are contained in the Supporting Information alongside Figure S16 and Tables S1 & S2.

4.7 Associated content

Supporting Information. Containing supplementary fig-ures & tables and experimental methods. This material is available free of charge via the Internet at http://pubs.acs.org.

4.8 Author information

Corresponding author

*Alex R. Jones., alex.jones@npl.co.uk; *Sam Hay, sam.hay@manchester.ac.uk; *Perdita Barran, perdi-ta.barran@manchester.ac.uk

Present addresses

†Biometrology Group, Department of Chemical and Biological Sciences, National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW.

Author contributions

The manuscript was written by ARJ, ISC and RB with contributions from all authors. All authors have given approval to the final version of the manuscript.

Funding sources

ARJ thanks The University of Manchester and the National

Measurement System of the Department for Business, Energy and Industrial Strategy for funding. ISC was supported by a PhD studentship from the Engineering and Physical Sciences Research Council. RB is supported by a Biotechnology and Biological Sciences Research Council grant BB/M011208/1. This work was supported by BBSRC grants BB/L002655/1, BB/L016486/1 and BB/M01108/1 and by Waters Corp. PB and LAIR are grateful to the MS SPIDOC project funded by the European

Union’s Horizon 2020 FET-OPEN Research and Innovation Programme, grant no. 801406.

Notes

Any additional relevant notes should be placed here.

4.9 Acknowledgements

The authors thank S. Padmanabhan and Montserrat Elías-Arnanz for sharing the plasmid for the G192Q TtCarH variant and for helpful discussions.

4.10 Abbreviations

AdoCbl, 5’-deoxyadenosylcobalamin; MeCbl, methylcobala-min; TtCarH, CarH from Thermus thermophilus; WT, wild-type; G192Q, G192Q variant of TtCarH; TtCarHCt, C-terminal domain of TtCarH; MxCarH, CarH from Myxococcus xan-thus; BmCarH, CarH from Bacillus megaterium; $^{TW}CCS_{N2}$, T-wave collision cross section measured in N2 drift gas; MD, molecular dynamics; SEC, size exclusion chromatography.

4.11 References

1. K. L. Brown, Chem. Rev., 2005, 105, 2075-2149.

2. R. Banerjee, Chem. Rev., 2003, 103, 2083-2094.

3. T. Toraya, Chem. Rev., 2003, 103, 2095-2127.

4. A. R. Jones, Photochem. Photobiol. Sci., 2017, 16, 820-834.

5. M. C. Pérez-Marín, S. Padmanabhan, M. C. Polanco, F. J. Murillo and M. Elías-Arnanz, Mol. Microbiol., 2008, 67, 804-819.

6. J. M. Ortiz-Guerrero, M. C. Polanco, F. J. Murillo, S. Padmanabhan and M. Elías-Arnanz, Proc. Natl. Acad. Sci. USA, 2011, 108, 7565-7570.

7. S. Padmanabhan, M. Jost, C. L. Drennan and M. Elías-Arnanz, Annu. Rev. Biochem., 2017, 86, 485-514.

8. A. Díez, J. Ortiz-Guerrero, A. Ortega, M. Elías-Arnanz, S. Padmanabhan and J. García de la Torre, Eur. Biophys. J., 2013, 42, 463-476.

9. J. Fernández-Zapata, R. Pérez-Castaño, J. Aranda, F. Colizzi, M. C. Polanco, M. Orozco, S. Padmanabhan and M. Elías-Arnanz, J. Biol. Chem., 2018, 293, 17888-17905.

10. M. Jost, J. Fernández-Zapata, M. C. Polanco, J. M. Ortiz-Guerrero, P. Y.-T. Chen, G. Kang, S. Padmanabhan, M. Elías-Arnanz and C. L. Drennan, Nature, 2015, 526, 536-541.

11. R. J. Kutta, S. J. O. Hardman, L. O. Johannissen, B. Bellina, H. L. Messiha, J. M. Ortiz-Guerrero, M. Elias-Arnanz, S. Padmanabhan, P. Barran, N. S. Scrutton and A. R. Jones, Nat Commun, 2015, 6, 7907.

12. S. Kainrath, M. Stadler, E. Reichhart, M. Distel and H. Janovjak, Angew. Chem. Int. Ed., 2017, 56, 4608-4611.

13. C. Chatelle, R. Ochoa-Fernandez, R. Engesser, N. Schneider, H. M. Beyer, A. R. Jones, J. Timmer, M. D. Zurbriggen and W. Weber, ACS Synth. Biol., 2018, 7, 1349-1358.

14. R. Wang, Z. Yang, J. Luo, I.-M. Hsing and F. Sun, Proc. Natl. Acad. Sci. USA, 2017, 114, 5912-5917.

15. D. Xu, J. Ricken and S. V. Wegner, Chem. Eur. J., 2020, 26, 9859-9863.

16. B. Jiang, X. Liu, C. Yang, Z. Yang, J. Luo, S. Kou, K. Liu and F. Sun, Sci. Adv., 2020, 6, eabc4824.

17. M. J. Warren, E. Raux, H. L. Schubert and J. C. Escalante-Semerena, Nat. Prod. Rep., 2002, 19, 390-412.

18. P. J. Bassford and R. J. Kadner, J. Bacteriol., 1977, 132, 796-805.

19. M. Jost, V. Cracan, P. A. Hubbard, R. Banerjee and C. L. Drennan, Proc. Natl. Acad. Sci. USA, 2015, 112, 2419-2424.

20. D. S. Froese and R. A. Gravel, Expert reviews in molecular medicine, 2010, 12, e37.

21. G. Klug, Molecular Microbiology, 2014, 91, 635-640.

22. M. Jost, J. H. Simpson and C. L. Drennan, Biochemistry, 2015, 54, 3231-3234.

23. A. Nahvi, J. E. Barrick and R. R. Breaker, Nucleic Acids Res, 2004, 32, 143-150.

24. M. Elias-Arnanz, M. Fontes and S. Padmanabhan, in Myxobacteria: Multicellularity and Differentiation, ed. D. E. Whitworth, ASM Press, Washington DC, 2008, pp. 211–225.

25. Z. Cheng, K. Li, L. A. Hammad, J. A. Karty and C. E. Bauer, Mol. Microbiol., 2014, 91, 649-664.

4.12 Supporting information

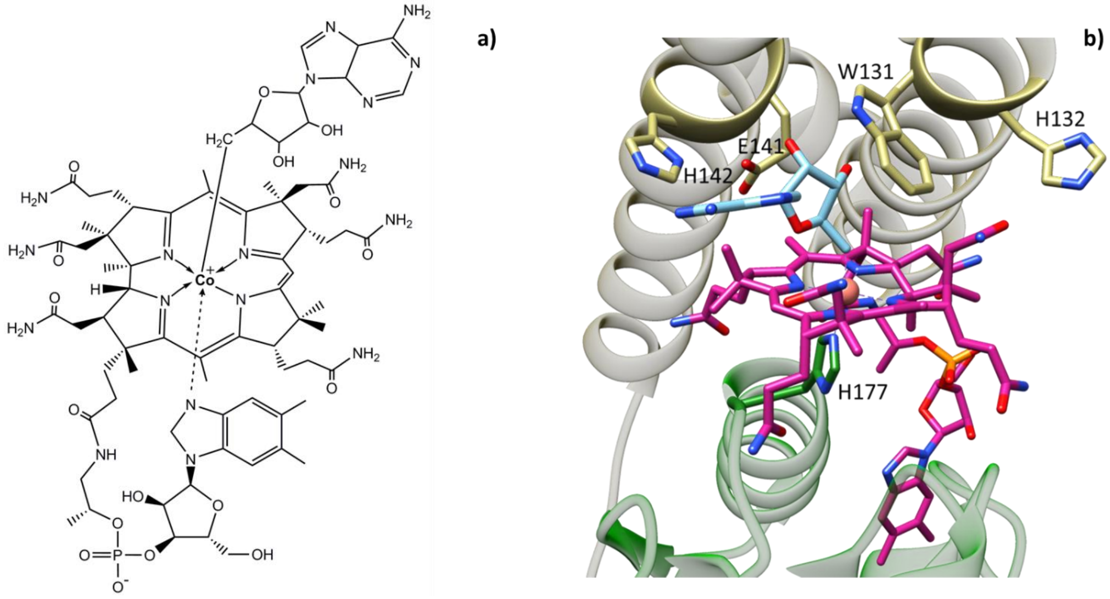


Figure S1a) Chemical structure of AdoCbl in the base-on conformation (with the lower axial coordination site of the central cobalt occupied by the 5,6-dimethylbenzimidazole (DMB) base). b) Close-up of the B_{12}*-binding site of TtCarH showing AdoCbl in its base-off/His-on conformation, with the residues that contact the upper axial ligand of AdoCbl, 5'dAdo (cyan) highlighted in yellow, and the cobalt-coordinating His (H177) replacing the DMB base in the lower axial coordination site. The domain colors are the same as in Figure 4.1a.*

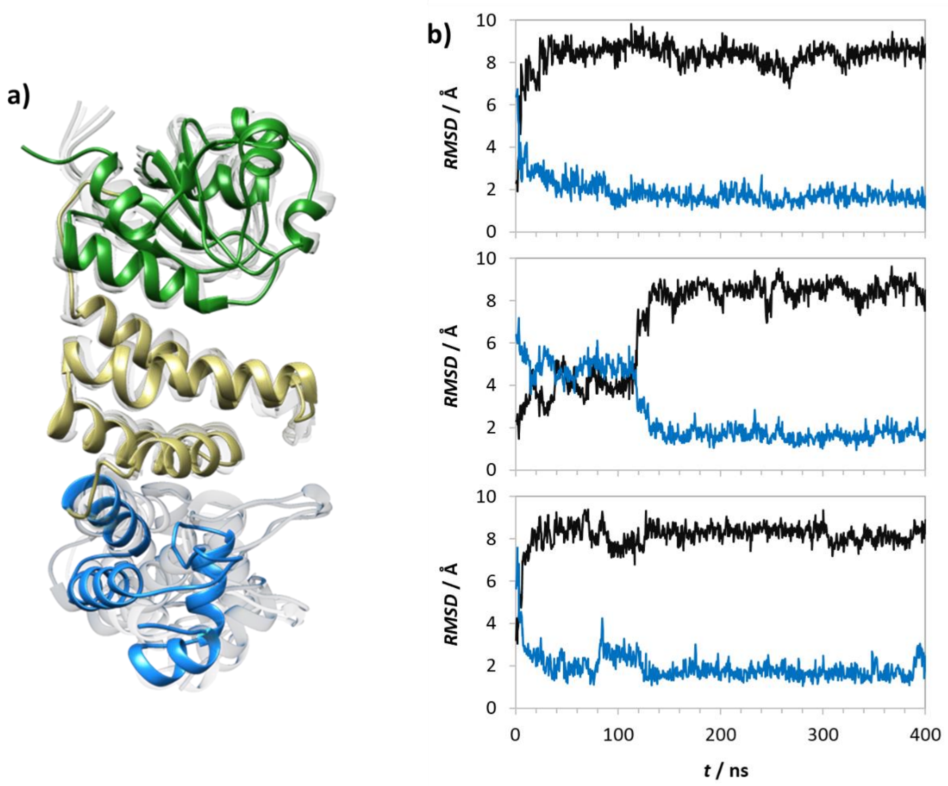


Figure S2a) Representative structures of apo-TtCarH from MD simulations chosen using a single linkage clustering method (see Experimental section). The structures shown cover 95.3% of the total conformational space, with the most populated structure shown in color and the next four most populated clusters represented in grey to illustrate variation. b) Black traces: root mean squared deviation (RMSD) of the protein backbone atoms of the simulated apoTtCarH monomer structure relative to the starting, photoconverted holo-TtCarH structure (PDB: 5C8F) after alignment to the photoconverted holo-TtCarH structure. Blue traces: RMSD of the protein backbone atoms of the simulated apoTtCarH monomer structure relative to the average structure across all three simulations after alignment to the photoconverted holo-TtCarH structure. These data suggest that the system reaches equilibration in less than 15 ns in each simulation.

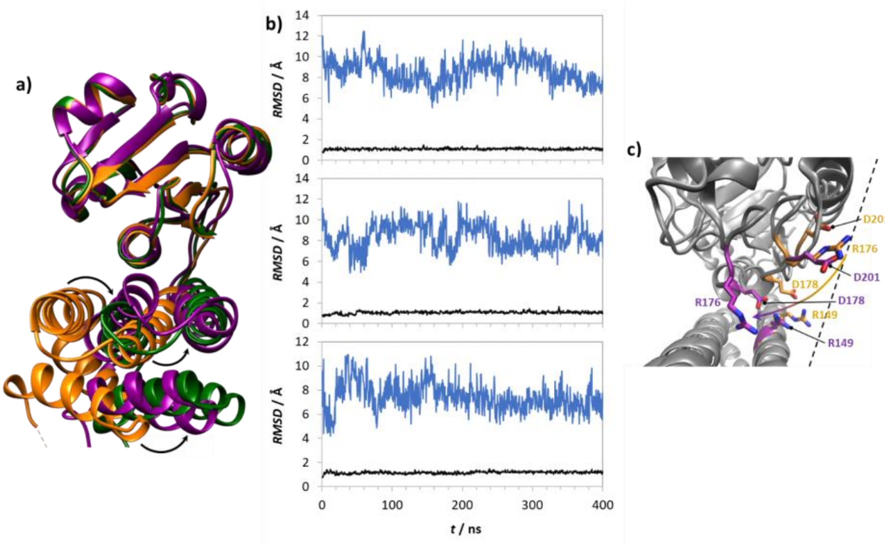


Figure S3a) 4-helix bundle (bottom) positions in each monomer aligned to the B_{12}-binding domain (top): holo-TtCarH (orange, PDB: 5C8D), photo-converted holo-TtCarH (dark green, PDB: 5C8F; i.e., with the upper axial Ado missing) and apo-TtCarH (purple, simulated). Arrows indicate displacement of the 4-helix bundle in photo-converted holo-TtCarH and apo-TtCarH relative to holo-TtCarH. b) RMSD of the protein backbone atoms of the apo-TtCarH B_{12}-binding domain (black) and 4-helix bundle (blue) relative to the holo-TtCarH structure after alignment to the B_{12}-binding domain. The average of the blue RMSD traces gives a displacement of the apo-TtCarH 4-helix bundle of 8.14 ± 1.33 Å relative to the holo-TtCarH structure. For comparison, the RMSD of the 4-helix bundle in the photo-converted holo-TtCarH structure relative to holo-TtCarH after alignment to the B_{12}-binding domain is 9.7 Å, and the RMSD for the photo-converted holo-TtCarH structure relative to the apo-TtCarH is 3.7 ± 1.2 Å. The 4-helix bundle displacement for both apo-TtCarH and photo-converted holo-TtCarH is therefore of a similar magnitude, but they end up in slightly different positions c) Overlay of monomers from holo-TtCarH and apo-TtCarH but aligned to the 4-helix bundle instead of the B_{12}-binding domain as previously. Salt-bridge residues identified here (D178 & R149) and previously[1] (D201 & R176) are highlighted for each monomer in orange and purple, respectively. In combination with Figure 4.1d, this image serves to illustrate that, whatever domain the structures are aligned to, the residues in the apo-TtCarH monomer are no longer in a position relative to the dimer interface (dashed line) to form stabilising salt-bridges. This shifting away from the dimer interface between the holo-TtCarH and apo-TtCarH structures is highlighted by the curved arrow

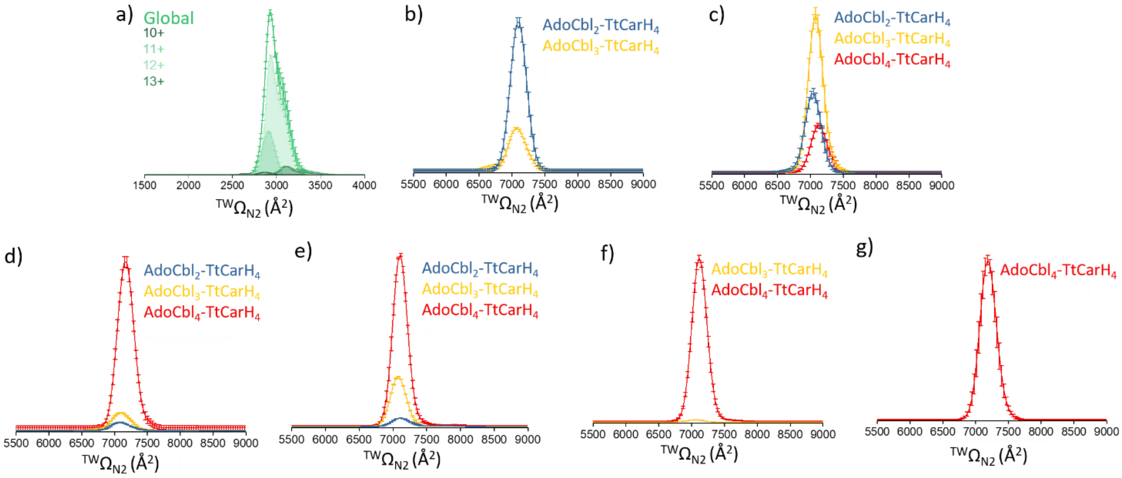


Figure S4. The collision cross section ($^{TW}CCS_{N2}$, $^{TW}\Omega_{N2}$) distributions from ion mobility data of the apo-monomer species (a) and of all tetramer species at the different concentration ratios (b-g) observed at charge states 23+ to 26+ for WT TtCarH:AdoCbl ratios: 1:0.15, 1:0.25, 1:0.5, 1:0.75, 1:1 and 1:2, respectively. All were performed in triplicate. a) The apo-monomer distrubution remains unchanged with increasing proportions of AdoCbl. The charge states ranging from 10+ to 12+ all sit within a very similar range whereas $^{TW}CCS_{N2}$ distribution of 13+ appears to indicate a slightly more unfolded state. b-g) Three different tetramer species are evident – AdoCbl$_2$-TtCarH$_4$, AdoCbl$_3$-TtCarH$_4$ and AdoCbl$_4$-TtCarH$_4$ – with relative populations that vary across the range of ratios. Each tetramer returns the same $^{TW}CCS_{N2}$ distribution within error, however, regardless of the number of AdoCbl bound. This suggests that all tetramer structures are highly similar.

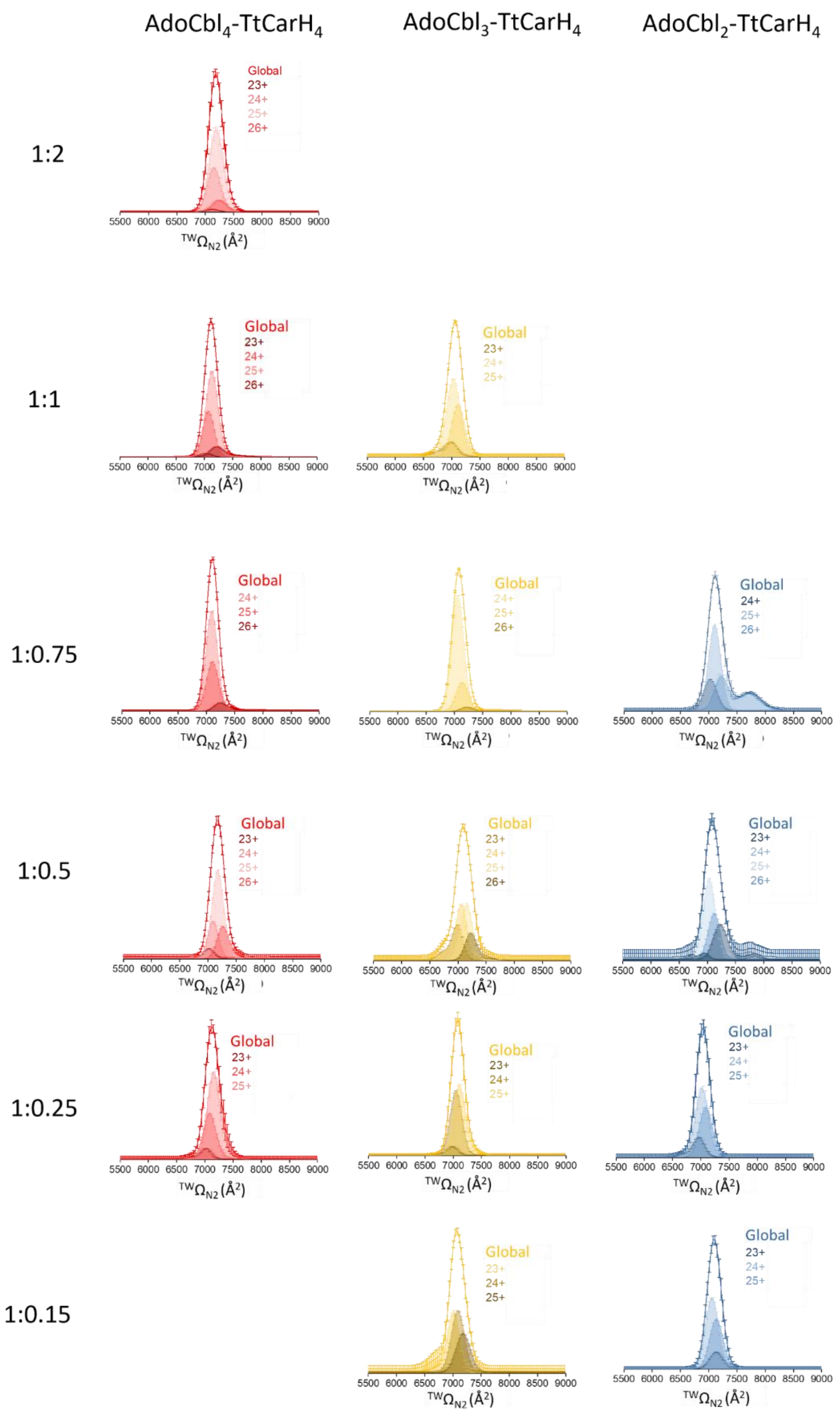


Figure S5. $^{TW}CCS_{N2}$ ($^{TW}\Omega_{N2}$) *distributions for all charge states of each tetrameric species (AdoCbl*$_4$*-TtCarH*$_4$ *– red; AdoCbl*$_3$*-TtCarH*$_4$ *– yellow; AdoCbl*$_2$*-TtCarH*$_4$ *– blue) at the different WT TtCarH:AdoCbl ratios (indicated on the left). For each species at each ratio, the global* $^{TW}CCS_{N2}$ *is shown as a colored line with the different charge states displayed beneath in gradient shades. Data for each charge state have been* *corrected for their m/z peak height and peak area and were then summed together to give the global* $^{TW}CCS_{N2}$*. There is little change in the distribution of* $AdoCbl_4$*-*$TtCarH_4$ *with increasing AdoCbl content implying that the global 3D structure remains similar throughout. There is slightly greater fluctuation between charge states for* $AdoCbl_3$*-*$TtCarH_4$ *and* $AdoCbl_2$*-*$TtCarH_4$*. This could be because these forms adopt slightly less rigid structures than* $AdoCbl_4$*-*$TtCarH_4$*, but the reduction in the species signal intensity these sub-populations undergo with increasing AdoCbl might also contribute.*

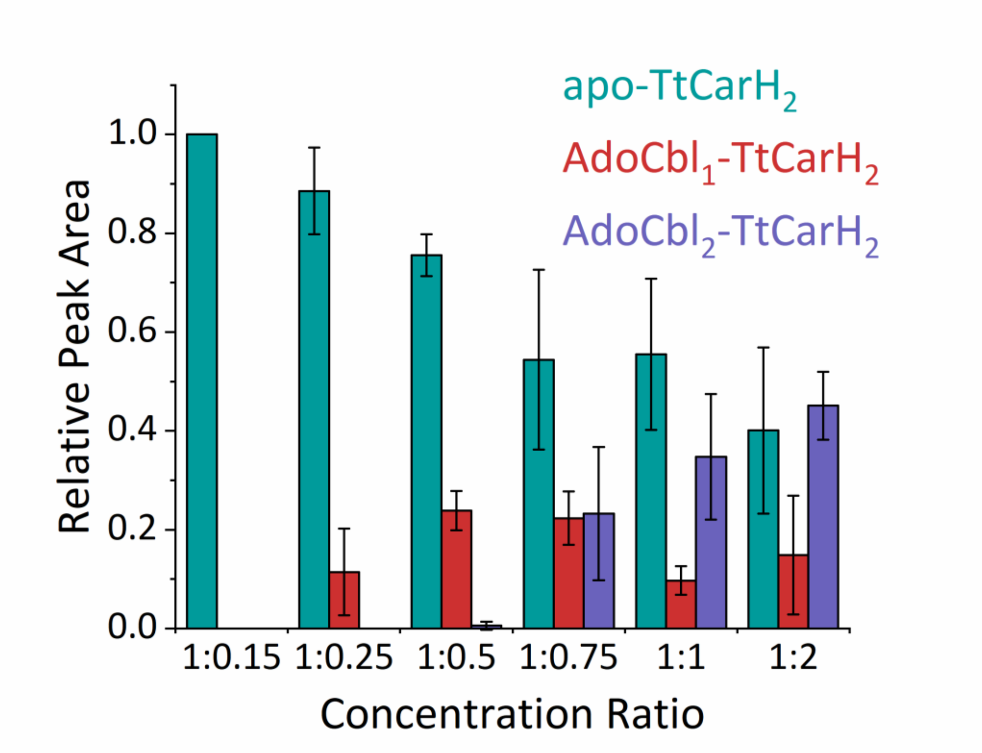


Figure S6. The relative peak area as a function of WT TtCarH:AdoCbl ratio for the low amplitude mass spectral signals between 3500 - 4500 m/z in Figure 4.2a that correspond to WT TtCarH dimer species. For the sake of simplicity, these peaks are not highlighted on the mass spectrum in Figure 4.2a with colored, dashed lines as for the tetramer species. Three dimer species are observed – apo-$TtCarH_2$*, *$AdoCbl_1$*-*$TtCarH_2$* and *$AdoCbl_2$*-*$TtCarH_2$* – the populations of which vary across the range of ratios. At a ratio of 1:0.15, there is only apo-*$TtCarH_2$* present. When compared to the equivalent plots for the tetramers (Figure 4.2a), these data suggest that any *$AdoCbl_1$*-*$TtCarH_2$* and *$AdoCbl_2$*-*$TtCarH_2$* that form at this ratio rapidly combine to form tetrameric species. This trend appears to continue. Although increasing populations of *$AdoCbl_1$*-*$TtCarH_2$* and *$AdoCbl_2$*-*$TtCarH_2$* are evident at higher AdoCbl concentrations, they always have a lower relative population than the equivalent tetramers (*$AdoCbl_2$*-*$TtCarH_4$*, *$AdoCbl_3$*-*$TtCarH_4$* and *$AdoCbl_4$*-*$TtCarH_4$*) at the same ratios.*

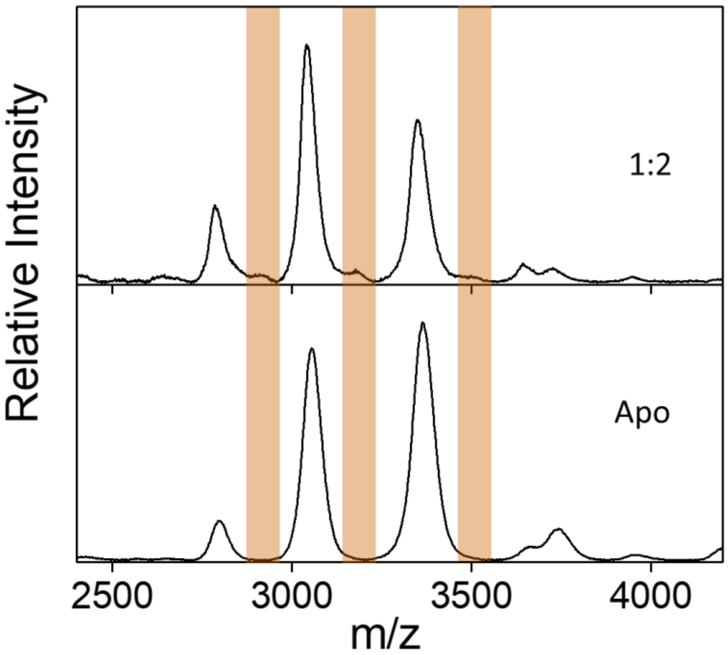


Figure S7. An expansion of the region of the mass spectrum that corresponds to the TtCarH monomer species. Data are compared for a sample of apo-TtCarH (bottom) to data when AdoCbl is in excess (top, i.e., a WT TtCarH:AdoCbl ratio of 1:2) where additional, low amplitude signals (highlighted) are evident. The mass difference is equivalent to AdoCbl (~1550 Da) and these signals therefore correspond to AdoCbl$_1$*-TtCarH*$_1$*. There is no evidence of this signal at WT TtCarH:AdoCbl ratios less than 1:2.*

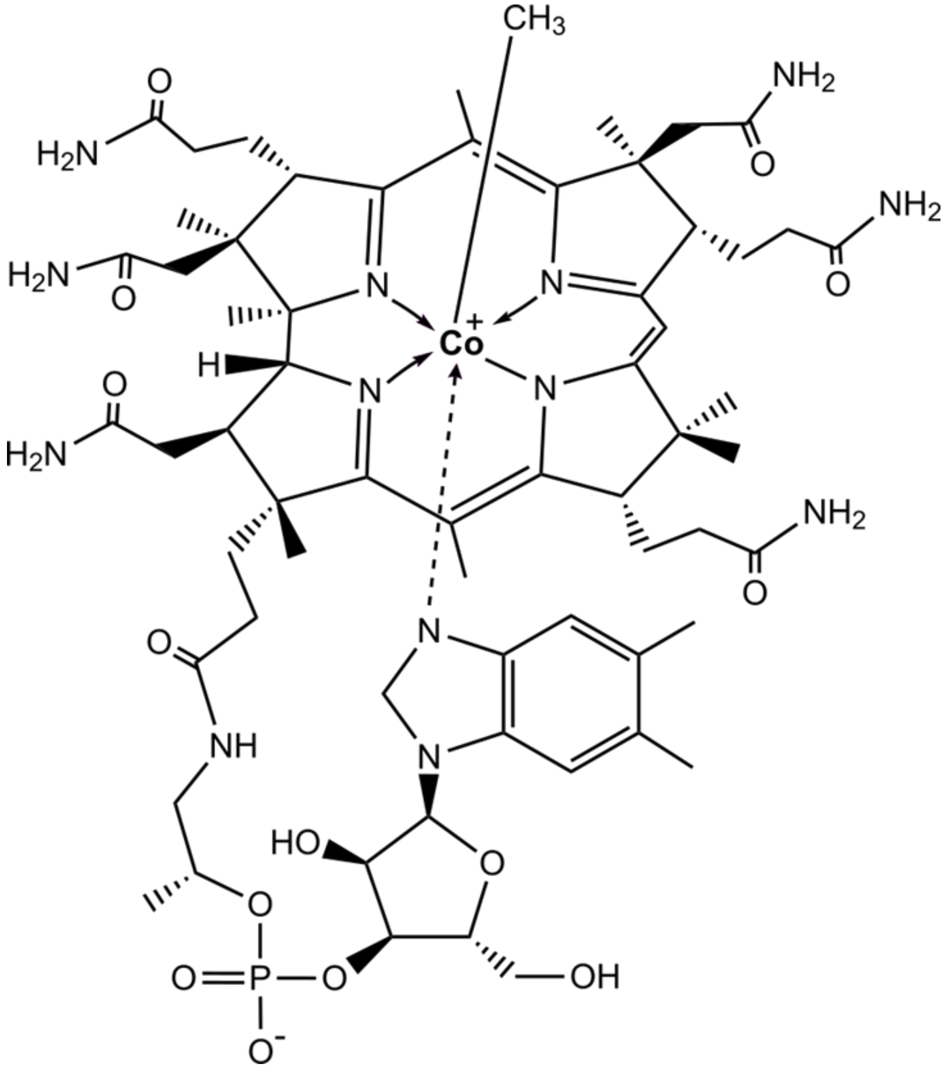


Figure S8. Chemical structure of MeCbl in the base-on conformation.

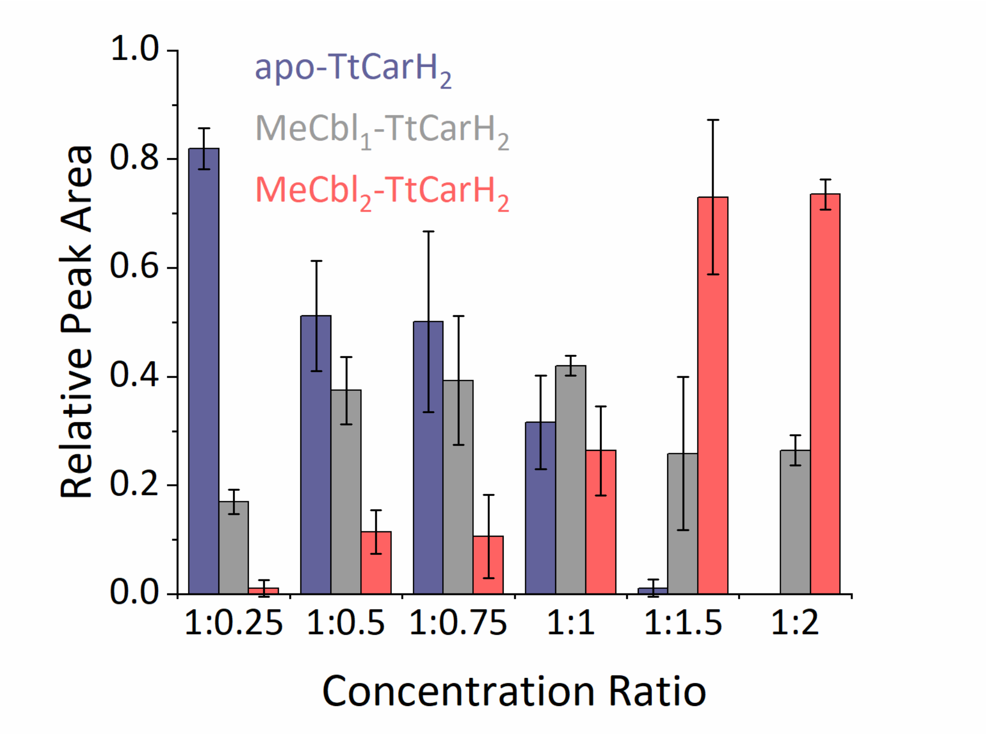


Figure S9. The relative peak area as a function of WT TtCarH:MeCbl ratio for the low amplitude mass spectral signals between 4000 - 5000 m/z in Figure 4.3a that correspond to dimer species (color matched in Figure 4.3a with the dashed lines). Much like for AdoCbl (Figure S6), three dimer species are observed – apo-TtCarH$_2$*, MeCbl*$_1$*-TtCarH*$_2$* and MeCbl*$_2$*-TtCarH*$_2$* – and again their populations vary across the range of ratios. The lowest MeCbl concentration is also predominantly apo-TtCarH*$_2$*, but unlike AdoCbl, small but significant populations of the other dimer species are clearly evident. Indeed, relative to apo-TtCarH*$_2$*, MeCbl*$_1$*-TtCarH*$_2$* and MeCbl*$_2$*-TtCarH*$_2$* become far more prominent with increasing concentration of MeCbl. This is almost certainly because, unlike those binding AdoCbl, the dimers here do not combine to form tetramers and therefore their populations are not depleted in the same way. Nevertheless, all dimer species remain very much minor sub-populations in a spectrum dominated by monomer species (Figure 4.3a&b).*

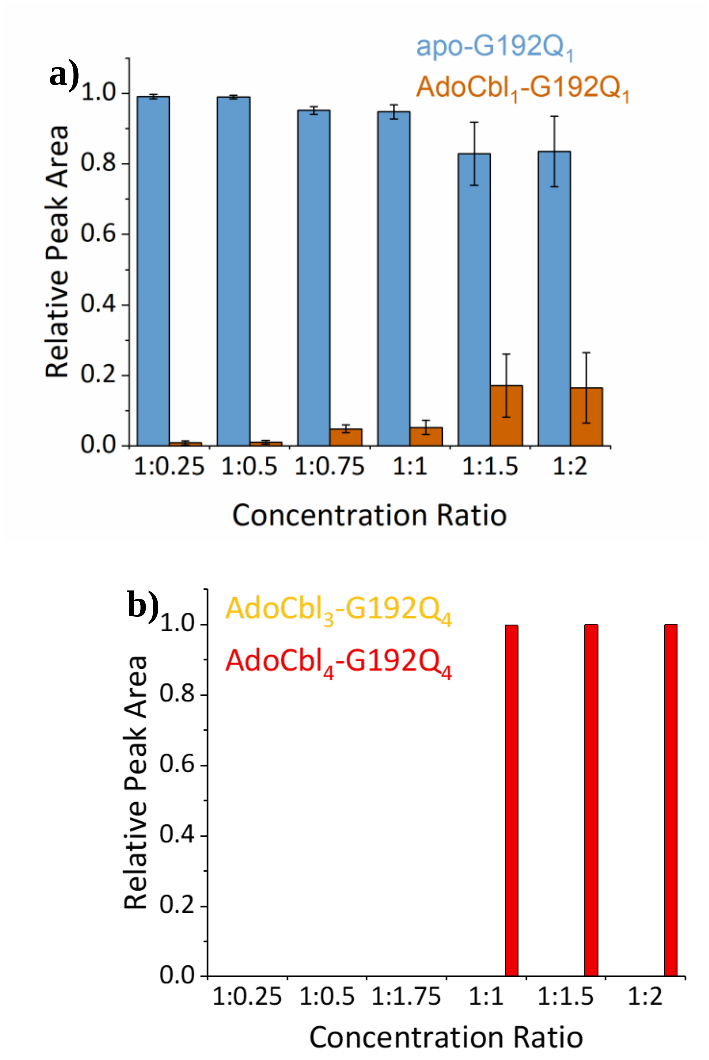


Figure S10. The relative peak area as a function of G192Q:AdoCbl ratio for the mass spectral signals in Figure 4.3c that corresponds to monomer (a) (colour matched in Figure 4.3c with the dashed lines) and tetramer species (b) (not highlighted in Figure 4.3c, for the sake of simplicity, but the low amplitude signals are visible between 5500 – 6500 m/z). a) Two monomer species are observed – apo-G192Q$_1$ and AdoCbl$_1$-G192Q$_1$ – the populations of which vary across the range of ratios. Like for WT TtCarH (Figure S6), apo-G192Q$_1$ is by far the most significant monomer species across all ratios. This is likely to reflect the dominant dimer species (Figure 4.3c&d) being more stable and that most AdoCbl$_1$-G192Q$_1$ that does form combines to form AdoCbl$_1$-G192Q$_2$ or AdoCbl$_2$-G192Q$_2$. Unlike WT TtCarH, however, where there is only the slightest hint of AdoCbl$_1$-TtCarH$_1$ when the AdoCbl is in excess (Figure S7), AdoCbl$_1$-G192Q$_1$ is evident even at the lowest AdoCbl concentrations. The dominant G192Q dimers are therefore not as stable as the corresponding dominant tetramer for WT TtCarH, which draws the position of equilibrium further away from the holo-monomer. b) Although *published SEC data suggest G192Q does not form tetramers,*[1] *a small sub-population is clearly evident in Figure 4.3c at ratios >1:1. Unlike when AdoCbl is bound to WT TtCarH, there is only one tetramer population, AdoCbl*$_4$*-G192Q*$_4$*, which only appears at AdoCbl concentrations where AdoCbl*$_2$*-G192Q*$_2$ *becomes the predominant species (Figure 4.3d). The G → Q substitution is thought to sterically hinder the native association between pairs of head-to-tail dimer, which means these tetramers could be the result of non-specific association between AdoCbl*$_2$*-G192Q*$_2$*. This is supported by the fact that for the WT, tetramers are observed even at the low [AdoCbl] concentrations. The* $^{TW}CCS_{N2}$ *for both the AdoCbl*$_4$*-TtCarH*$_4$ *and the AdoCbl*$_4$*-G192Q*$_4$ *are compared in Figure S11b.*

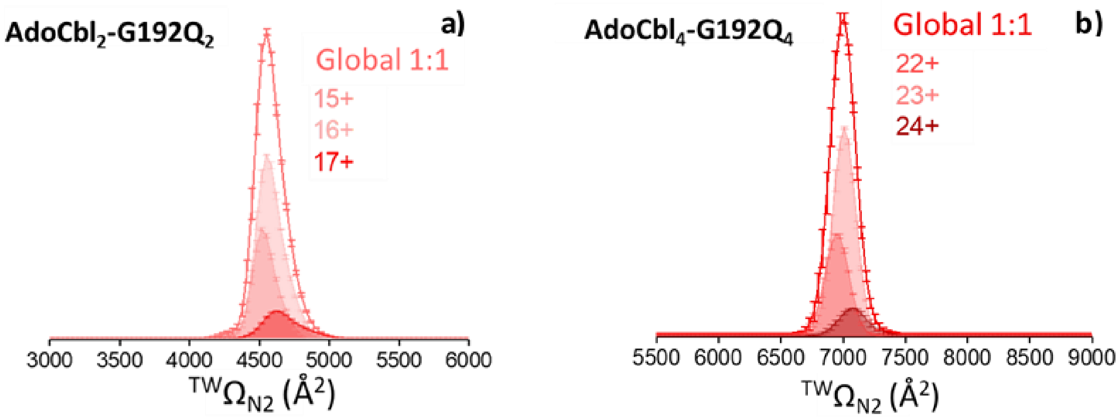


Figure S11. $^{TW}CCS_{N2}$ *(*$^{TW}\Omega_{N2}$*) distributions from ion mobility data of AdoCbl*$_2$*-G192Q*$_2$ *(a) and AdoCbl*$_4$*-G192Q*$_4$ *(b) for a G192Q:AdoCbl ratio of 1:1. For each species, the global* $^{TW}CCS_{N2}$ *is shown as a colored line with the different charge states displayed beneath in gradient shades. Data for each charge state have been corrected for their m/z peak height and peak area and were then summed together to give the global* $^{TW}CCS_{N2}$*. a) The 15+ and 16+ states for AdoCbl*$_2$*-G192Q*$_2$ *each give similar* $^{TW}CCS_{N2}$ *and are thus likely be the native-like conformations. The minor 17+ state is slightly broader and hence has more conformation flexibility. b) The AdoCbl*$_4$*-G192Q*$_4$ *distribution follows a very similar pattern, also suggesting a predominantly native-like conformation. AdoCbl*$_4$*-G192Q*$_4$ *adopts a slightly smaller global* $^{TW}CCS_{N2}$ *than AdoCbl*$_4$*-TtCarH*$_4$ *(Figure 4.2c), perhaps indicative of non-specific association, this marginal (1.6%) difference is within instrumental error.*

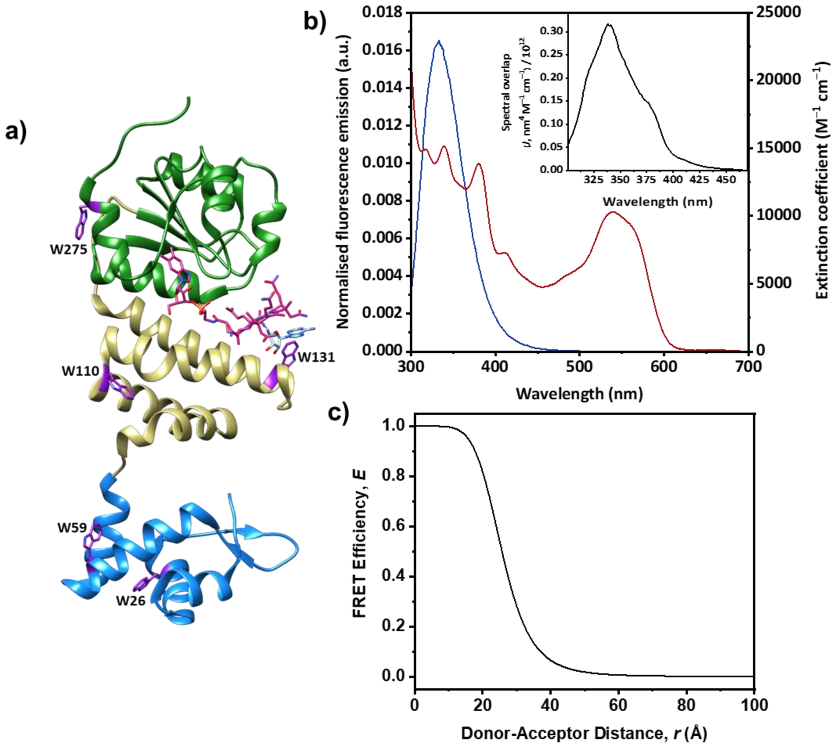


Figure S12a) A monomer unit from holo-TtCarH (PDB: 5C8D) with the five tryptophan residues (purple) and bound AdoCbl (magenta) highlighted. b) Absorbance spectrum of the AdoCbl in holo-TtCarH (red) and the normalized fluorescence emission spectrum of apo-TtCarH (blue, excitation at 280 nm); inset: spectral overlap (J) calculated from the spectra in the main panel using the Förster equation.[2] A Förster distance, R0 = 26 Å, was calculated from the spectral overlap assuming κ 2 = 2/3 and the refractive index, n = 1.4. A florescence quantum yield, Q = 10%, was used for the tryptophan emission. If Q = 1% or 30% - reasonable lower and upper bounds for tryptophan fluorescence3 - the resulting R0 = 18 and 31 Å, respectively. c) FRET efficiency (E) as a function of donor (tryptophan) - acceptor (AdoCbl) distance (r), calculated[2] using R0 = 26 Å. One would therefore expect some extent of quenching by AdoCbl of tryptophans within ~ 40 Å (see Table S1).

Table S1. Distances between the Cα in each of the 20 tryptophans (W) in the entire holo-TtCarH tetramer (i.e., 5 in each monomer, Figure S12a) and the Co ion of AdoCbl in monomer A. Distances were measured using the published[1] crystal structure (PDB: 5C8D) and each monomer of the tetramer is given a letter designation A-D, with A/B and C/D being the two head-to-tail dimers. There will, of course, be equivalent and corresponding distances between each W and the AdoCbl in monomers B-D. From Figure S12, we can estimate that fluorescence from W within ~ 40 Å is likely to be quenched with non-negligible efficiency by AdoCbl (highlighted in green below). This includes W from each monomer, which means the fluorescence quenching experiments conducted by stopped-flow spectroscopy (Figure S13) are highly likely to be sensitive to both AdoCbl-binding and the subsequent tetramer domain assembly steps

Tryptophan and Monomer	**Distance (Å)**
W131.A	9.63
W59.D	25.58
W110.A	26.22
W131.B	28.10
W275.A	32.53
W110.C	34.15
W275.B	35.69
W110.D	35.94
W26.D	36.70
W110.B	38.51
W275.C	42.09
W59.C	44.40
W26.A	44.83
W59.A	46.57
W26.C	46.78
W275.D	49.73
W131.D	55.28

Tryptophan and Monomer	Distance (Å)
W26.B	57.05
W131.C	57.31
W59.B	59.42

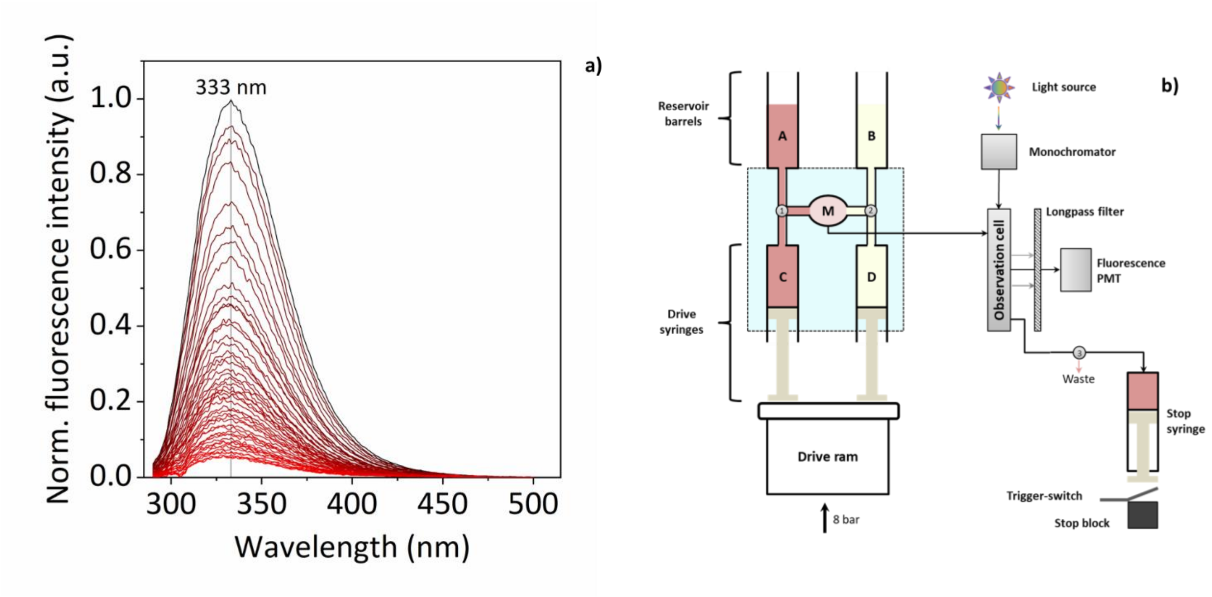


Figure S13a) Spectral evolution of the fluorescence (λ_{exc} = 280 nm) from TtCarH during titration with increasing [AdoCbl]. This fluorescence is significantly quenched by AdoCbl binding and thus changes to the fluorescence signal that follow rapid mixing of TtCarH and AdoCbl enables us to monitor the binding mechanism. Small increments (10 µL) of AdoCbl (up to 30 µM) were sequentially added to a sample of 6 µM WT apo-CarH, with fluorescence emission spectra acquired following each addition. Each spectrum is corrected for dilution. b) Schematic representation of the SX20 stopped-flow spectrometer (Applied Photophysics) used in all transient measurements. A and B, reservoir barrels; C and D, drive syringes; M, mixing chamber; PMT, photomultiplier tube. The system valves are represented by greyed numbered circles. TtCarH (syringes A/C) and AdoCbl (syringes B/D) were rapidly mixed together in the 20 µL observation cell.

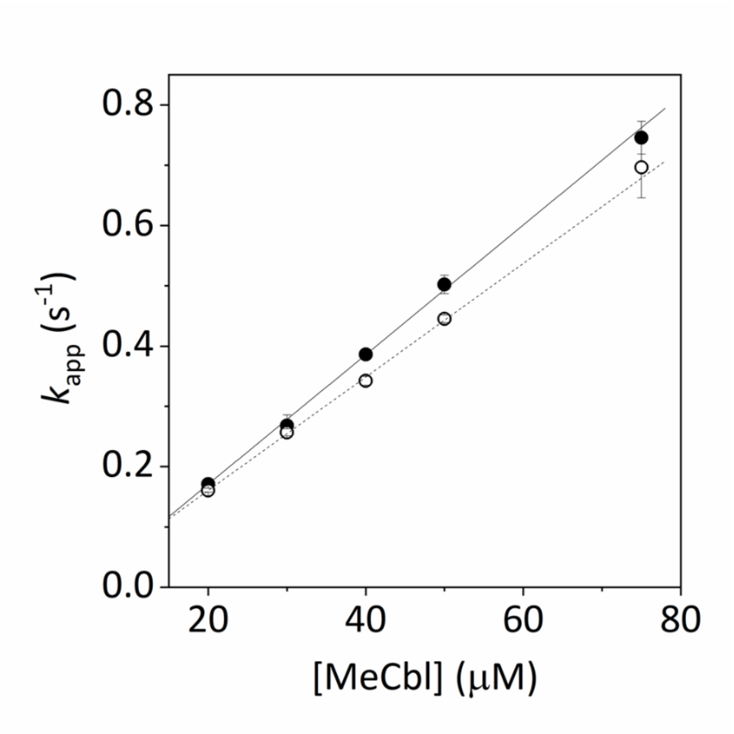


Figure S14. Comparison of the apparent rate constant (k_{app}) of WT (solid circles) and G192Q (open circles) TtCarH variants as a function of MeCbl concentration, at a fixed protein concentration of 5 μM. In both cases, the fluorescence quenching that follows rapid mixing of protein and MeCbl contains a single kinetic phase that is linearly-dependent on [MeCbl]. The data therefore correspond to the bimolecular binding of MeCbl to protein monomers, with no significant quenching contributions from protein domain assembly steps, and the second order rates for WT (10.8 ± 0.1 s^{-1} mM^{-1}) and G192Q (9.4 ± 0.1 s^{-1} mM^{-1}) are very similar.

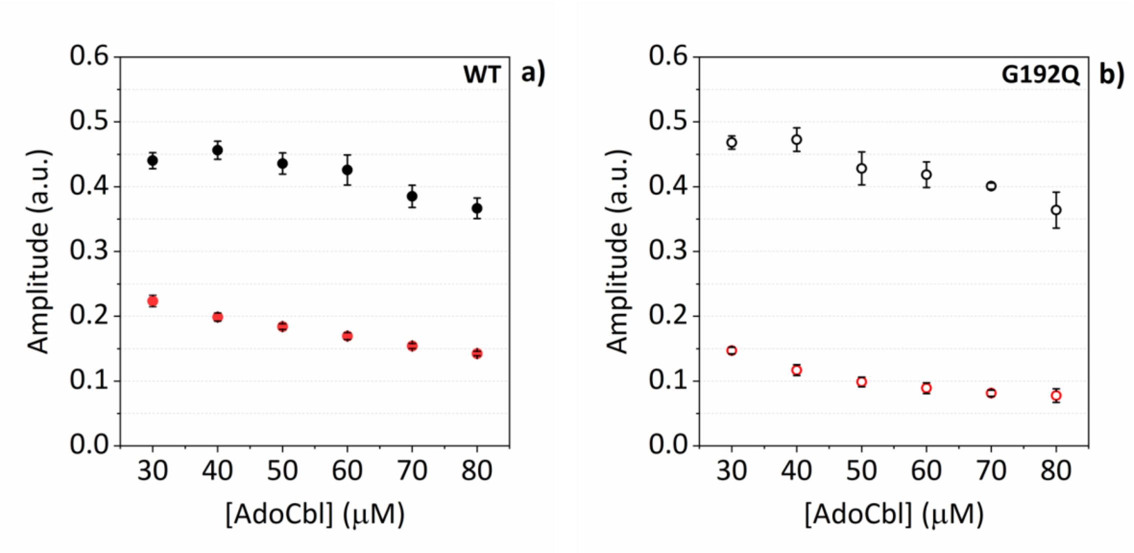


Figure S15 Amplitudes of the different kinetic phases of the fluorescence quenching that follows rapid mixing of AdoCbl with WT TtCarH (filled circles) (a) and G192Q (open circles) (b) as a function of [AdoCbl]. The amplitudes for the faster kinetic phase (black) are the same within error for each variant, indicating that they represent an equivalent process (i.e., binding of AdoCbl to protein monomers). The amplitudes for the slower kinetic phase (red) are consistently smaller for G192Q than WT TtCarH. Because this phase is linearly-dependent on [protein] (Figure 4.4b&d) for both variants, it corresponds to quenching from protein domain assembly. The smaller amplitude for G192Q therefore indicates that there is less quenching than for WT TtCarH, consistent with domain assembly predominantly stopping at the dimer rather than going on to form a significant proportion of tetramers.

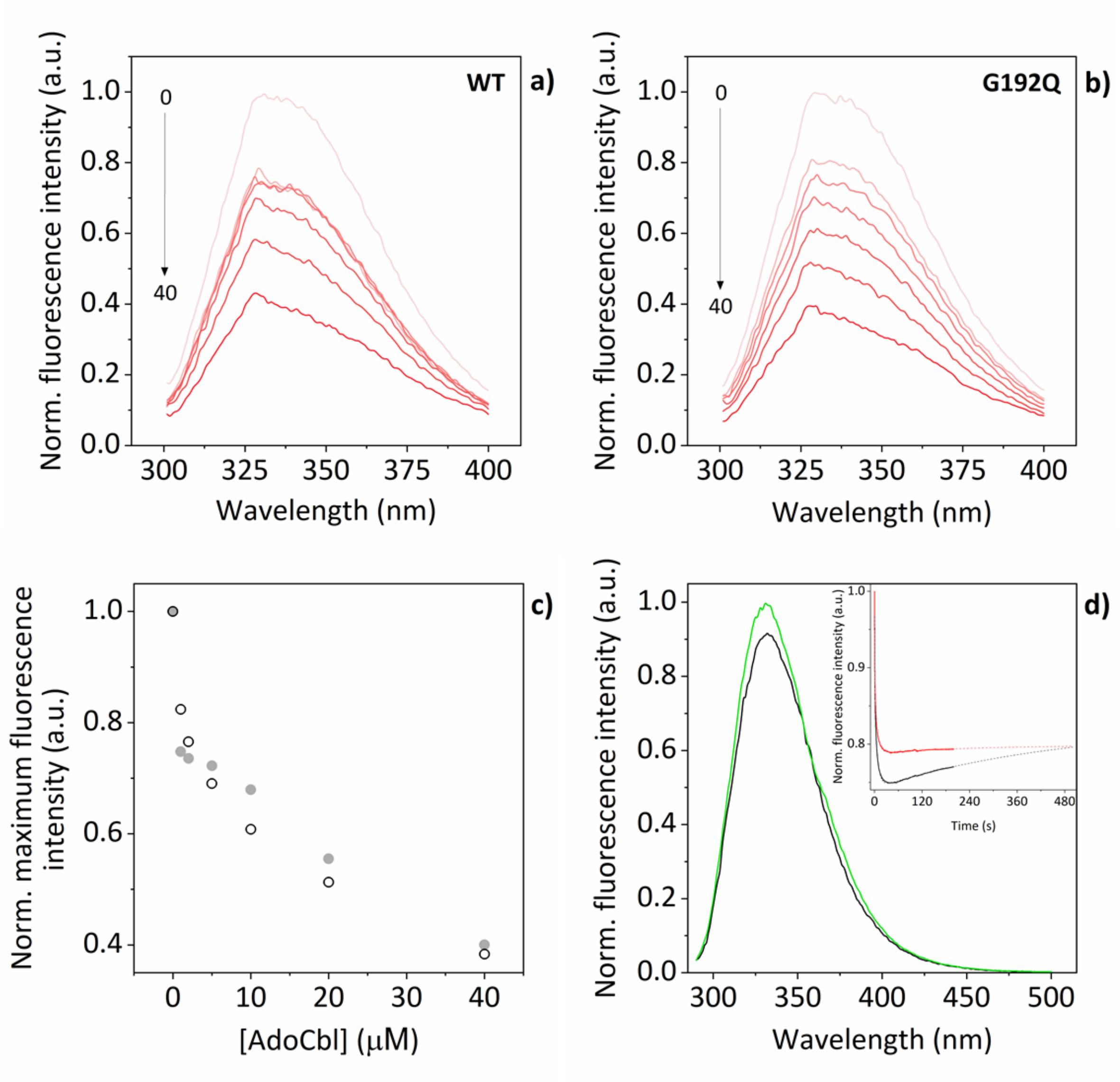


Figure S16. Comparison of fluorescence quenching between TtCarH variants. To further investigate the difference in quenching amplitude observed in the stopped flow data (Figure 4.4a & S15), static fluorescence spectra (λ_{exc} = 280 nm) from WT TtCarH (a) and G192Q (b) were acquired during titration with increasing concentrations of AdoCbl (up to 40 µM). c) Maximum fluorescence intensity (at 335 nm) as a function of AdoCbl concentration for WT TtCarH (solid circles) and G192Q (open circles). In contrast to the stopped flow data, panels (a-c) show no difference in the fluorescence quenching between the two variants. d) Emission spectra of WT TtCarH before (black) and after (green) exposure to a 530 nm LED for 5 s. The photoconverted monomer has a larger fluorescence magnitude than the 'dark' state tetramer. Inset: long time-base *fluorescence stopped flow traces following rapid mixing of WT TtCarH (solid black line) and G192Q (solid red line) with AdoCbl. In each case, the data were fit to the sum of three exponentials (two negative phases and one positive phase) and extrapolated to longer times (dashed lines) to illustrate convergence. Although the quenching amplitudes in panels (a-c) are ostensibly the same, the data in (d) reveal this to be an artefact of the measurement. In each case the static spectra where acquired from the same protein sample with sequential additions of AdoCbl, with each TtCarH sample therefore exposed to the excitation light for several minutes. This light was also absorbed by bound AdoCbl over the same period, which thus slowly and irreversibly activated the TtCarH photoreceptor. The static fluorescence spectra for each variant following the titration were the same and dominated by the photoconverted protein.*

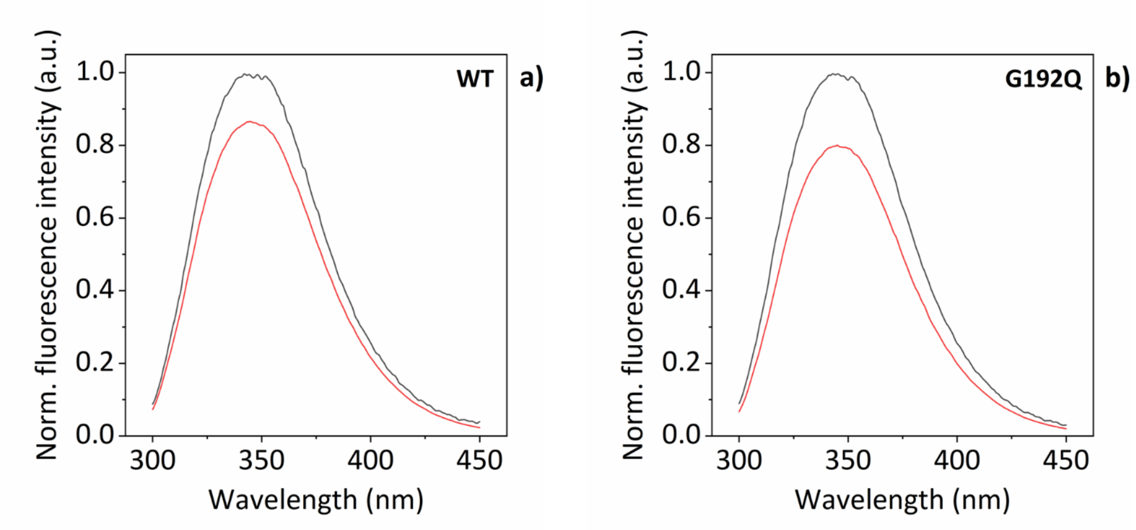


Figure S17. Fluorescence spectra of 0.5 µM (black) and 5 µM (red) apo-TtCarH for the WT (a) and G192Q variants (b). To normalise, the 0.5 µM spectra have been multiplied by 10. For both variants, the normalised fluorescence amplitude for the 5 µM protein sample is lower than for the 0.5 µM sample, presumably owing to the inner filter effect.

Experimental Section

Molecular Dynamics Simulations

Molecular dynamics simulations of apo-TtCarH were performed using the crystal structure[1] of light-exposed full-length TtCarH (PDB: 5C8F) as a starting point, using the Gromacs package[3] with the Amber03 force field,[4] a solvation box of minimum 10 Å around the protein and periodic boundary conditions. Three MD simulations were run in parallel after energy minimization using the following protocol: the system was initially thermalized to 300 K for 100 ps using the *NVT* ensemble (constant volume), and the pressure was then equilibrated for 100 ps using the *NPT* ensemble (constant pressure) with harmonic constraints of force constant 10 kJ mol^{-1} $Å^2$ applied to the protein; constraints and pressure couplings were then switched off and the system relaxed using *NPT* at 250, 280, 290 and 300 K for 1 ns each. Finally, 400 ns of *NPT* dynamics were run at 300 K. Representative structures (Figure S2a) were chosen using a single linkage clustering on the whole protein (not including hydrogen atoms) for all three simulations together, with a cut-off of 0.17 nm (the average RMSD between any two structures being 0.35 nm), which resulted in 95.3% of structures present in the top 5 clusters, which have the following populations: 78.7%, 8.89%, 4.56%, 2.38% and 0.833%.

Materials and Protein Production

Unless otherwise stated, all commercial reagents were obtained from Sigma-Aldrich and used without further purification. Plasmids containing the genes encoding WT and G192Q variants of TtCarH were kindly provided by S. Padmanabhan and Montserrat Elías-Arnanz. The genes had been cloned into modified pET15b (*Novagen*) expression vectors using NdeI and BamHI restriction enzymes, to provide a N-terminal 6xHis affinity tag as previously described.[5]

The following overexpression and purification protocols were used for both WT and G192Q variants of TtCarH. The plasmids were transformed into *E. coli* BL21 (DE3) cells (*Novagen*), and a single colony inoculated into a small volume of selective (Amp^R) LB medium. The cell cultures were grown to an OD_{600} ~ 0.8 at 37 °C and 200 rpm, at which time they were inoculated into a larger volume of fresh LB. The fresh cell cultures were grown to an OD_{600} ~ 0.7 at 37 °C and 200 rpm, at which time they were cooled down to the induction temperature of 25 °C. Protein expression was then induced with 0.5 mM IPTG and the cells were left to incubate overnight at 25 °C and 200 rpm. Cell were then harvested by centrifugation at 5,000 rpm and 4 °C for 20 mins, collected, rapidly frozen in liquid nitrogen, and stored at − 80 °C. For purification, cells were thawed and resuspended in buffer A (50 mM sodium phosphate + 300 mM NaCl, pH 7.5) supplemented with 2.5 mM imidazole, lysozyme, DNAse, $MgCl_2$, and protease inhibitors, and lysed by cell disruption. Cell debris was removed by centrifugation at 20,000 rpm and 4 ^oC for 1 h. The collected supernatant was filtered through a 0.22 μm membrane and incubated with 5 mL TALON metal affinity resin (*Clontech*) for 2 h at 4 ^oC with rolling. Protein-bound resin was washed with buffer A supplemented with 5 mM (wash 1) and 10 mM imidazole (wash 2). Protein was eluted with buffer A supplemented with 150 mM imidazole following 1 h of incubation at 4 ^oC with rolling. Fractions containing purified protein were further purified by size-exclusion chromatography using a Superdex200 high-performance liquid chromatography column (*Cytiva*), equilibrated with 50 mM phosphate buffer + 150 mM NaCl, pH 7.5. Purified sample was brough to the desired concentration using 10K molecular weight cut off Vivaspin centrifugal filter devices (*Sartorius*).

Native Mass Spectrometry and Ion Mobility Mass Spectrometry CCS Measurements

Mass spectrometry experiments were performed on a modified traveling-wave ion mobility enabled Synapt G2-S (Waters), described previously.[6] The n-ESI tips were pulled in-house from borosilicate capillaries (outer diameter 1.2 mm, inner diameter 0.69 mm, length 10 cm, *Science Products GmbH*) using a laser-based P-2000 micropipette puller (*Sutter Instrument Company*). A positive voltage was applied to the solution *via* a platinum wire (*Goodfellow Cambridge Ltd*). Data were analysed using MassLynx v4.1 (*Waters Corporation*), OriginPro 9.1 (*OriginLab Corporation*), and Microsoft Excel 2010 (*Microsoft*). Native mass spectra were recorded on a modified Synapt G2-S employing gentle source conditions with a capillary voltage between 1.1-1.4 kV, cone voltage of 10 V and all radio frequencies set to zero. $^{TW}CCS_{N2}$ measurements were performed on the same modified mass spectrometer following the standard calibration procedure utilising the Mason-Schamp equation.[7-9] Measurements were made in nitrogen and spraying conditions were again kept as gentle as possible with an applied capillary voltage range of 1.1 to 1.4 kV, cone voltage maintained at 10 V and a source temperature of 333.13 K.

Table S2. Standard proteins and their ion mobility $^{TW}CCS_{N2}$ values determined by DTIMS. These were used to calibrate the TtCarH TWIMS ions mobility data.

Protein	**Mass**	**Charge State**	**Literature CCS (Ω_{He}, Å^2)^3**	**Literature CCS (Ω_{N2}, Å^2)^5**
Alcohol Dehydrogenase	~148 kDa	26+	6720	7473
		25+	6830	7465
		24+	6940	7447
		23+	6940	7422
BSA	~66 kDa	17+	4040	4523
		16+	4060	4514
		15+	4100	4478
		14+	4090	4425
B-Lactoglobulin (monomer)	~18 kDa	9+	1780	NA
		8+	1690	
		7+	1660	
B-Lactoglobulin (dimer)	~36 kDa	13+	2850	NA
		12+	2900	
		11+	2960	
Concanavalin A	~103 kDa	22+	5550	5930
		21+	5550	5921
		20+	5480	5902

Mass spectrometry titration experiments

WT TtCarH and G192Q variants were buffer exchanged into 250 mM Ammonium Acetate, pH 6.8 using Slide-a-Lyzer Dialysis Cassettes (*ThermoFisher*). The concentration of this new stock solution was then checked by absorption at 280 nm using a DS-11 Spectrophotometer (*Denovix*). The AdoCbl and MeCbl were diluted to the desired concentration in ultrapure water that was obtained from a Milli-Q Advantage

A10 ultrapure water filtration system (Merck Millipore, Darmstadt, Germany). Data were acquired for a range of TtCarH:B_{12} ratios, from excess proteins to excess B_{12}. Mixtures at each ratio were made up under red light to avoid spurious photoactivation, vortexed and inserted into the nESI tip, which was then covered and kept in the dark during acquisition. The final monomeric concentration for the mass spectrometry was 12 µM for the apo-protein and an overall concentration of 10 µM for the titration mixtures. Fresh titration mix was prepared for all experimental MS runs and all experiments were performed in triplicate.

Fluorescence stopped-flow measurements

Fluorescence stopped-flow measurements were carried out in a SX20 stopped-flow spectrometer (*Applied Photophysics*), with Pro-Data SX and Pro-Data viewer operating software. All samples were prepared in 150 mM NaCl, 50 mM phosphate, pH 7.5. Protein samples at each working concentration (see Figure 4.4 & Figure 4.5) were prepared under white light, whereas the cobalamin samples were prepared under red light and stored in black vials due to their photosensitivity. Stopped flow measurements were performed in a 20 µL quartz cell at ~ 25 ^{o}C, using an excitation wavelength of 280 nm, 2 mm monochromator slit width (entrance and exit), a 320 nm long-pass filter, and 400 V in the fluorescence channel. All measurements were performed under red light to stop spurious photoactivation.

After loading into the drive syringes, all samples were thermal equilibrated with the water bath for ~ 10 minutes. Fluorescence emission data were acquired following rapid mixing of protein and B_{12} with 1000 data points acquired over the acquisition period. The first three shots were discarded per syringe fill to account for the dead volume, and the kinetic data acquisition repeated for each of the following 5 shots. Each repeat trace was fitted from 0.1 ms to 30 s to an appropriate sum of exponentials using the fitting tool on OriginPro 2020 (*OriginLab Corporation*). Each fitting yielded decay constants that were converted to pseudo first-order rate constants; these were then averaged, and standard deviations and standard errors calculated. Concentration dependences of each rate constants was assessed by plotting the averaged values as a function of both cobalamin and TtCarH concentrations. Second order rates were derived from the gradient of linear fits (Figure 4.4, Figure 4.5 and S18, and Table S3).

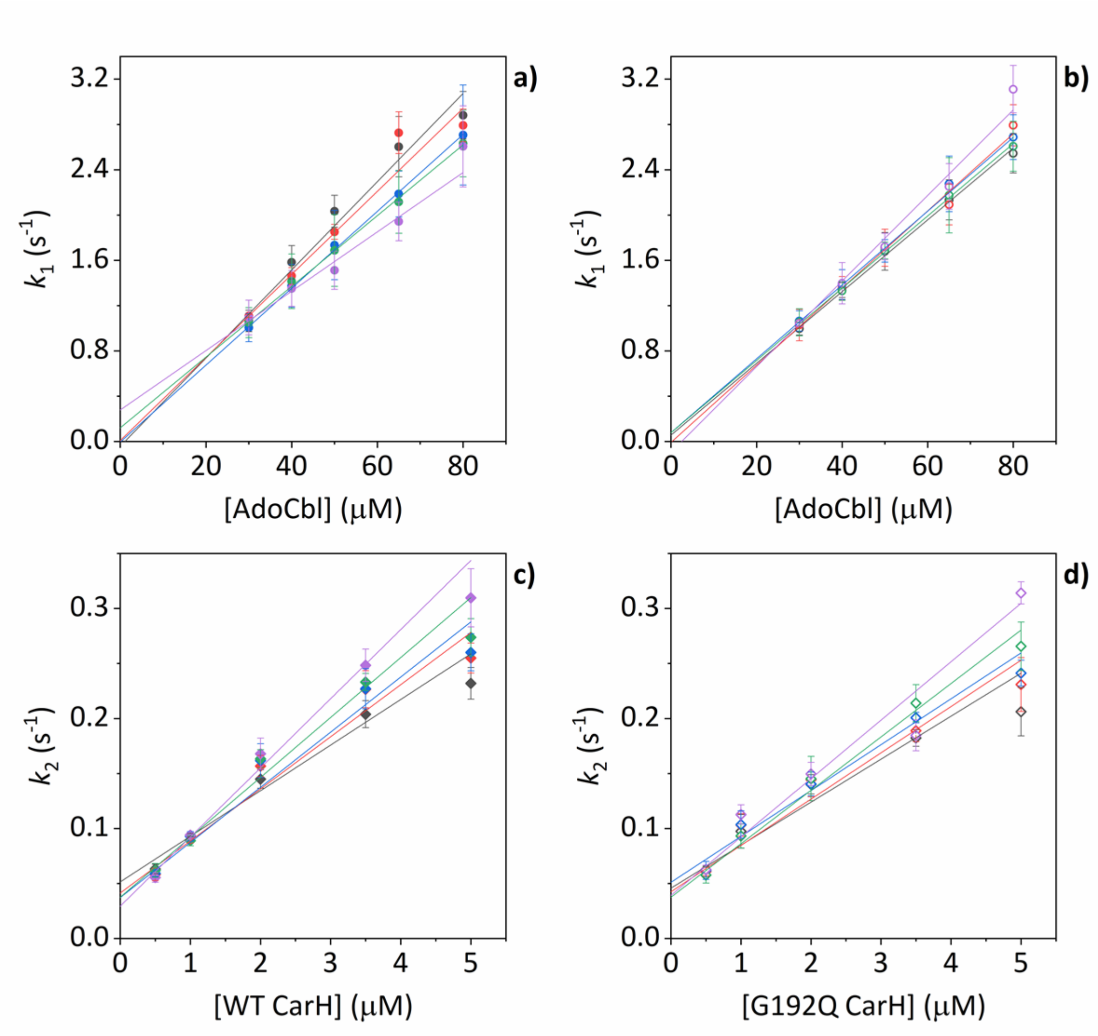
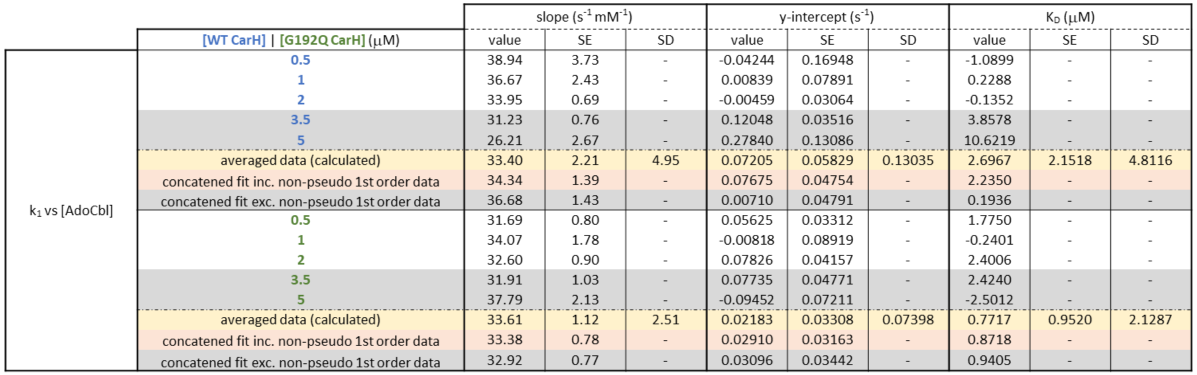
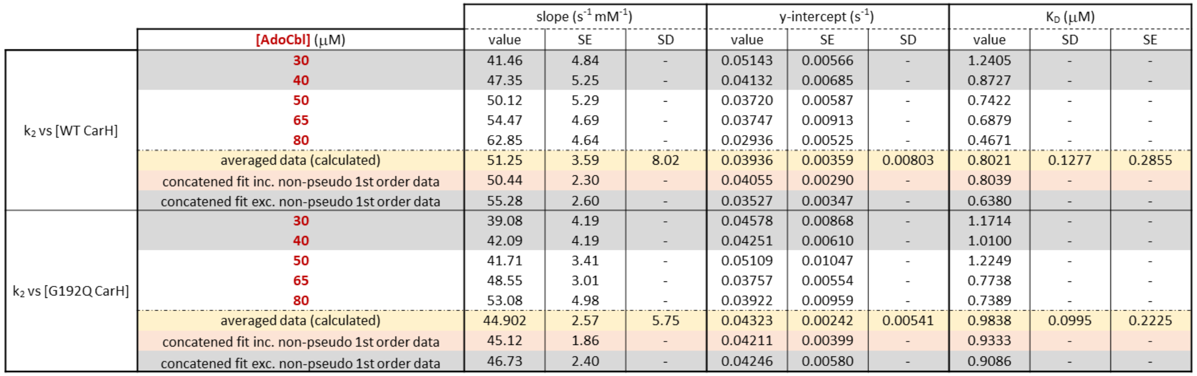


Figure S18. Linear fits of the concentration-dependent kinetic data acquired from the fluorescence stopped-flow experiments. (a-b) Linear fits to the variation of pseudo first order rates as a function of [AdoCbl] for each [TtCarH] for WT (a) and G192Q (b). Data points and fit lines are colored according to the [TtCarH]: black, 0.5 μM; red, 1 μM; blue, 2 μM; green, 3.5 μM; purple, 5 μM. (c-d) linear fits to the variation of pseudo first order rates as a function of [TtCarH] for WT (c) and G192Q (d) and each [AdoCbl]. Data points and fits lines are colored according to the [AdoCbl]: black, 30 μM; red, 40 μM; blue, 50 μM; green, 65 μM; purple, 80 μM. The parameters from each fit are in Table S2.

Table S3. Second order rate (slope) and dissociation coefficients (y-intercept) for both B_{12} binding to WT and G192Q TtCarH variants and for their domain assembly. Fitting was conducted in two ways. First, for the [AdoCbl]-dependences, the variation of the pseudo first order rate was fit linearly at each concentration of the two TtCarH variants. Similarly, for the [protein]-dependences, data were fit for each [AdoCbl]. For these fits, the parameters were averaged and errors calculated (highlighted yellow; c.f., Figure S18). Second, concatenated fits were performed (c.f., Figure 4.4 & Figure 4.5), where a single linear fit was made for the [AdoCbl]-dependence to all of the data from every [protein] (separately for WT and G192Q). The same was done for the [protein]-dependences for all [AdoCbl]. This was either conducted including (highlighted orange) or excluding (highlighted grey) the non-pseudo first order data.



		slope (s^{-1} mM^{-1})			y-intercept (s^{-1})			K$_D$ (μM)		
	[WT CarH] | **[G192Q CarH]** (μM)	value	SE	SD	value	SE	SD	value	SE	SD
k$_1$ vs [AdoCbl]	**0.5**	38.94	3.73	-	-0.04244	0.16948	-	-1.0899	-	-
	1	36.67	2.43	-	0.00839	0.07891	-	0.2288	-	-
	2	33.95	0.69	-	-0.00459	0.03064	-	-0.1352	-	-
	3.5	31.23	0.76	-	0.12048	0.03516	-	3.8578	-	-
	5	26.21	2.67	-	0.27840	0.13086	-	10.6219	-	-
	averaged data (calculated)	33.40	2.21	4.95	0.07205	0.05829	0.13035	2.6967	2.1518	4.8116
	concatened fit inc. non-pseudo 1st order data	34.34	1.39	-	0.07675	0.04754	-	2.2350	-	-
	concatened fit exc. non-pseudo 1st order data	36.68	1.43	-	0.00710	0.04791	-	0.1936	-	-
	0.5	31.69	0.80	-	0.05625	0.03312	-	1.7750	-	-
	1	34.07	1.78	-	-0.00818	0.08919	-	-0.2401	-	-
	2	32.60	0.90	-	0.07826	0.04157	-	2.4006	-	-
	3.5	31.91	1.03	-	0.07735	0.04771	-	2.4240	-	-
	5	37.79	2.13	-	-0.09452	0.07211	-	-2.5012	-	-
	averaged data (calculated)	33.61	1.12	2.51	0.02183	0.03308	0.07398	0.7717	0.9520	2.1287
	concatened fit inc. non-pseudo 1st order data	33.38	0.78	-	0.02910	0.03163	-	0.8718	-	-
	concatened fit exc. non-pseudo 1st order data	32.92	0.77	-	0.03096	0.03442	-	0.9405	-	-



		slope (s^{-1} mM^{-1})			y-intercept (s^{-1})			K$_D$ (μM)		
	[AdoCbl] (μM)	value	SE	SD	value	SE	SD	value	SD	SE
k$_2$ vs [WT CarH]	**30**	41.46	4.84	-	0.05143	0.00566	-	1.2405	-	-
	40	47.35	5.25	-	0.04132	0.00685	-	0.8727	-	-
	50	50.12	5.29	-	0.03720	0.00587	-	0.7422	-	-
	65	54.47	4.69	-	0.03747	0.00913	-	0.6879	-	-
	80	62.85	4.64	-	0.02936	0.00525	-	0.4671	-	-
	averaged data (calculated)	51.25	3.59	8.02	0.03936	0.00359	0.00803	0.8021	0.1277	0.2855
	concatened fit inc. non-pseudo 1st order data	50.44	2.30	-	0.04055	0.00290	-	0.8039	-	-
	concatened fit exc. non-pseudo 1st order data	55.28	2.60	-	0.03527	0.00347	-	0.6380	-	-
k$_2$ vs [G192Q CarH]	**30**	39.08	4.19	-	0.04578	0.00868	-	1.1714	-	-
	40	42.09	4.19	-	0.04251	0.00610	-	1.0100	-	-
	50	41.71	3.41	-	0.05109	0.01047	-	1.2249	-	-
	65	48.55	3.01	-	0.03757	0.00554	-	0.7738	-	-
	80	53.08	4.98	-	0.03922	0.00959	-	0.7389	-	-
	averaged data (calculated)	44.902	2.57	5.75	0.04323	0.00242	0.00541	0.9838	0.0995	0.2225
	concatened fit inc. non-pseudo 1st order data	45.12	1.86	-	0.04211	0.00399	-	0.9333	-	-
	concatened fit exc. non-pseudo 1st order data	46.73	2.40	-	0.04246	0.00580	-	0.9086	-	-

Static fluorescence measurements

Static fluorescence measurements were carried out in a FLS920 spectrofluorometer (*Edinburgh Instruments*), with the F980 spectrometer operating software. Samples were diluted to ~ 0.1 OD at the wavelength of interest with the same buffer used in the stopped-flow measurements. Quartz cuvettes with a fully masked standard exterior size, 10 mm emission aperture window pathlength and 2 mm sample chamber were used. For the photoconversion experiment, a holo-TtCarH sample was illuminated using a high-power light-emitting diode (LED; *Thorlabs*) with $\lambda_{\max}$ = 530 nm.

References

(1) Jost, M.; Fernández-Zapata, J.; Polanco, M. C.; Ortiz-Guerrero, J. M.; Chen, P. Y.-T.; Kang, G.; Padmanabhan, S.; Elías-Arnanz, M.; Drennan, C. L. Structural basis for gene regulation by a B12- dependent photoreceptor. Nature 2015, 526, 536.

(2) Főrster, T. 10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation. Discuss. Faraday Soc. 1959, 27 (0), 7.

(3) Eftink, M. R. In Methods of Biochemical Analysis, 1991, DOI:https://doi.org/10.1002/9780470110560.ch3
https://doi.org/10.1002/9780470110560.ch3.

(4) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, LoadBalanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. 2008, 4 (3), 435.

(5) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003, 24 (16), 1999.

(6) Ortiz-Guerrero, J. M.; Polanco, M. C.; Murillo, F. J.; Padmanabhan, S.; Elías-Arnanz, M. Lightdependent gene regulation by a coenzyme B12-based photoreceptor. Proc. Natl. Acad. Sci. USA 2011, 108 (18), 7565.

(7) Bellina, B.; Brown, J. M.; Ujma, J.; Murray, P.; Giles, K.; Morris, M.; Compagnon, I.; Barran, P. E. UV photodissociation of trapped ions following ion mobility separation in a Q-ToF mass spectrometer. Analyst 2014, 139 (24), 6348.

(8) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion mobility–mass spectrometry analysis of large protein complexes. Nat. Prot. 2008, 3 (7), 1139.

(9) Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. Anal. Chem. 2008, 80 (24), 9689.

(10) France, A. P.; Migas, L. G.; Sinclair, E.; Bellina, B.; Barran, P. E. Using Collision Cross Section Distributions to Assess the Distribution of Collision Cross Section Values. Anal. Chem. 2020, 92 (6), 4340.

5 Conclusion and outlook

Mass spectrometry began as a method to investigate atomic ions and electrons,[1,2] and progressed to the study of complex hydrocarbons and other small molecules, this limit in scope and upper molecular weight in part was limited by the destructive nature of early ionisation sources. The 1980s saw the development of soft-ionisation techniques that allowed for the ionisation of larger biological molecules. These ionisation methods included ESI and MALDI which stemmed from the work on fast atom bombardment (FAB) that was developed at the University of Manchester.[3–5] This has allowed MS to prove itself as a beneficial technique for structural characterisation of large, intact proteins and their complexes.

Once in the gas phase there are several tools at our disposal with which to interrogate the structure of ions, *via* activation and dissociation. Traditionally, protein sequencing is carried out *via* bottom-up methodology, whereby the protein is first exposed to an enzyme and its digestion products can then go on to be fragmented *via* CID. This provides significantly improved fragmentation yield compared to the use of CID/CAD on native proteins, however in both cases, any conformational information is lost in the process.

Photons being implemented into mass spectrometers as a tool to induce fragmentation *via* methods such as UVPD and IRMPD, has seen in increase in use since the 1990s due to the increased availability of laser systems.[6–9] However, there is also a growing application of photons being utilised to investigate photoactive proteins and their chromophores. The latter can be exploited for the investigation of photoreceptors and cofactors *via* MS and IMMS, as demonstrated in studies by Bieske[10–14] and Barran.[15–17]

In this thesis, both applications of photons have been demonstrated and expanded upon, where previous UVPD investigations were performed on conformationally diverse commercially available proteins, this work has been extended to larger oligomeric proteins and PLS-DA has been incorporated for data processing. We selected conformationally diverse commercially available proteins that we then exposed to UVPD, travelled through the mobility cell and were then further exposed to CID that would bestow enough energy to break any remaining non-covalent interactions. We observed more fragmentation from larger more extended conformations, this gave further evidence to the effect of protein conformation on the UVPD fragmentation yield. The additional CID that was applied to the protein conformation allowed for the dissociation of UVPD produced fragments that had been held to the precursor *via* non-covalent interactions. Larger proteins were also investigated *via* 213 nm UVPD. Two multimeric systems were chosen, concanavalin A and haemoglobin for which much prior structural data exists. UVPD performed on these multimeric proteins gave information on the native-fold *via* observed fragments from intact tetramers in both cases that could not be found with CID. The usual method to interpret fragmentation data uses tools that predict fragment ion m/z values that arise solely from backbone cleavages and does not account for the conformation of the initial precursor protein. Here, UVPD has also been combined with PLS-DA, adapting methods commonly applied to metabolomics data. This revealed mass resolved features that could be targeted to determine protein conformation based on IM data, regardless of where they come from in the sequence of the protein. This method would allow protein structure to be probed in tandem mass spectrometers where ion mobility is not available. This procedure was also performed on CID fragmentation data and showed these exhibited a dependence on the precursor charge state.

PLS-DA applied here leads to greater understanding of “dark” regions of the MS, where traditional fragment assignment fails. The future for this work lies within the ability of the MVA processing gaining a further understanding of the dark regions of the MS, traditionally, MS fragmentation yields the fragments and the McLucky sequencing is used to assign the results to a defined sequence position. This system is limited to proteins where the sequence is either known or the system can give good sequence coverage. With the application of MVA a much larger picture can be monitored and focussed on to determine where the difference in the fragmentation are occurring. These can then be further focussed on to determine their position and the implication of the modification or change can be determined. In this case, this thesis shows the ability of MVA to distinguish between different change states or conformations however this could be branched out to monitor the effect of co-factor binding, PTMs and the directed evolution of enzymes.

Alongside the application of UVPD to different proteins and other large biomolecues across multiple academic institutes and commercial instrument developments there are other UVPD applications which includes incorporating other well developed techniques and applying UVPD as another fragmentation technique to these. An example of this is its application into trapped ion-mobility mass spectrometry instrumentation, allowing for conformer selection and irradiation *via* 213 nm UV laser.[18] Other approaches include 2D MS which will correlate the product ions to their precursor ions without the requirement of chromatography separation, again this approach has been applied to the fragment ions produced by UVPD.[19] These sorts of developments push the understanding of biomolecules and demonstrate the value that UVPD can add.

Photoactivation-MS has been extended in this thesis to investigate photoreceptor proteins which absorb at different wavelengths. The LED set-up has been utilised in previous studies and here further developed to a more flexible arrangement with minimal instrument alteration. This approach was applied to different photoreceptor proteins that span the visible spectrum as well as a small molecule as a standard to test the arrangement. Shown here, UVR8, TtCarH and R. *Palustris* phytochrome all exhibit conformational changes when exposed to a specific wavelength, this will then go on to result in a change in the mass spectrum. With this arrangement, an LED will illuminate the analyte solution in the nESI tip and any photoactivated change can be monitored *via* mass spectrometry and where available, ion mobility. The work herein showcases the capability of this methodology to those working with photoreceptor proteins. It can also be combined with activated IMMS methods, including UVPD, to gain more insights into the protein. MS is readily available in most laboratories and the modification demonstrated here requires little set-up and will not disrupt current MS-users.

The future of this work lies within the range of biomolecules that could be investigated using the methodology described in this thesis, firstly described here the implementation of LEDs has been shown on a number of difference systems. Previous photoisomerisation work spoken about here focussed on the chromophore retinal, with the instrumentation described here, this can be extended through to the opsin family of biomolecules and extend this previous research *via* the utilisation of this LED-IMMS system. Current limitations in this method is the implementation of the LED is a single wavelength with no option to scan across wavelengths. This could be improved with different rig whereby several LEDs can be installed and potentially their light can be fed into a fibre optic cable to be directed into the source, with a different LED then switched to when the next wavelength is to be monitored. This methodology cannot be compared to the likes of a laser system due to the limitations when it comes to wavelength generation and the fixed nature of the LEDs, meaning the switching from one LED to the next will not be the slow scan through but rather a fixed wavelength to the next wavelength.

Research into WT TtCarH has used MS alongside stop-flow experiments and computational modelling to determine the mechanism of tetramer formation. The MS data was invaluable in the determination of the mechanism, unlike stop-flow investigations, the concentration of AdoCbl can be kept at a fraction of the protein concentration. This allows the monitoring of the tetramer formation, with the MS showing the formation of partially-filled tetramer species in environments of AdoCbl deficiency. As the concentration of AdoCbl is increased the tetramer species present from the mainly $AdoCbl_2$-$TtCarH_4$ through to the full $AdoCbl_4$-$TtCarH_4$ tetramer. $AdoCbl_4$-$TtCarH_4$ becomes the thermodynamically stable form and significantly more prominent feature of the MS spectrum. Combined with stop-flow data, which showed tetramer formation followed a two-step process when in the presence of excess AdoCbl, an overall tetramer formation mechanism could be determined which incorporated both the two-step process and the results from the MS data. This work will provide greater understanding to the function of CarH and we predict this greater understanding will influence its use as an optogenetic tool.

This methodology also provides the simplest method for monitoring the next step in the CarH and coenzyme B_{12} pathway, this being the monitoring of interaction and binding with DNA. Described here, the under occupied tetramer species have been speculated to behave the same way as the filled tetramer in their binding to DNA, IMMS provide a simple method to monitor and test this hypothesis. The procedure applied to CarH shows the added benefit of implementing IMMS techniques in the investigation of photoreceptor proteins, the data filled gaps in knowledge that were previously left unresolved with standard approaches that are usually implemented for the investigation of photoreceptors. This method described is not limited to photoreceptor proteins but can also be implemented to man-made supramolecular systems, these are routinely being developed to respond to changing light environments and perform tasks similar to that in nature. This type of technique will aid in the development of these as well as the understanding of the action that needs to be performed.

Aside from LEDs and UVPD, further instrument modifications are in development, such as the introduction of an IR laser following the current UV laser pathway. This has been successfully integrated onto the instrument previously (data not shown in this thesis) prior to instrument relocation and would provide an added benefit in the investigation and identification of small molecules especially in the research of Parkinson’s disease. Designs have also been made to incorporate action-FRET into the trap cell of the photo-Synapt instrument to further increase the scope of the instrument and unique capabilities. Recently, a coupled YAG-OPO laser system has been obtained with the plan to perform action spectroscopy by implementing this into the source and illuminating the nESI as performed with LEDs. The benefit of utilising this nanosecond laser system is its capability to scan across wavelengths and obtain an action spectroscopy plot of the light reactive proteins. There are also further designs for the modification of a Waters Cyclic instrument to incorporate the 213 nm laser, this will allow for increased resolution and separation of the ion mobility measurements.

Upgrading to the Waters Cyclic instrument would also benefit in the LED investigations; the flexibility of LED design means this is a straightforward modification. Again, the increased mobility resolution would answer any question of conformational change when under LED illumination for many of the photoreceptors displayed in this thesis. Another approach to test for this is combining illumination with HDX in source. Phytochromes exist as an open or closed structure depending on the illumination in solution, however this is only indicated in the MS by the appearance of subunit fragmentation. HDX might be used to probe this conformational change by revealing more easily accessible exchange sites, it would also potentially give access to kinetics with the MS2min fluidics.

1 J. J. Thomson, *Proc. R. Soc. A*, 1913, **89**, 1–20.

2 F. W. Aston, *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, 1920, **39**, 449–455.

3 H. R. Morris, M. Panico, M. Barber, R. S. Bordoli, R. D. Sedgwick and A. Tyler, *Biochem. Biophys. Res. Commun.*, 1981, **101**, 623–631.

4 J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.

5 K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida and T. Matsuo, *Rapid Commun. Mass Spectrom.*, 1988, **2**, 151–153.

6 M. R. Robinson, J. M. Taliaferro, K. N. Dalby and J. S. Brodbelt, *J. Proteome Res.*, 2016, **15**, 2739–2748.

7 J. A. Madsen and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 349–358.

8 B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon and P. E. Barran, *Analyst*, 2014, **139**, 6348–51.

9 J. Oomens, N. Polfer, D. T. Moore, L. van der Meer, A. G. Marshall, J. R. Eyler, G. Meijer and G. Von Helden, *PCCP*, 2005, **7**, 1345–1348.

10 N. J. A. Coughlan, B. D. Adamson, K. J. Catani, U. Wille and E. J. Bieske, *J. Phys. Chem. Lett.*, 2014, **5**, 3195–3199.

11 B. D. Adamson, N. J. A. Coughlan, R. E. Continetti and E. J. Bieske, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9540–9548.

12 B. D. Adamson, N. J. A. Coughlan, P. B. Markworth, R. E. Continetti and E. J. Bieske, *Rev. Sci. Instrum.*, 2014, **85**, 123109-1-123109–8.

13 N. J. A. Coughlan, M. S. Scholz, C. S. Hansen, A. J. Trevitt, B. D. Adamson and E. J. Bieske, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 1483–1490.

14 P. B. Markworth, B. D. Adamson, N. J. A. Coughlan, L. Goerigk and E. J.

Bieske, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25676–25688.

15 I. S. Camacho, A. Theisen, L. O. Johannissen, L. A. Díaz-ramos, J. M. Christie, G. I. Jenkins, B. Bellina, P. E. Barran and A. R. Jones, *Proc. Natl. Acad. Sci.*, 2019, **116**, 1116–1125.

16 R. J. Kutta, S. J. O. Hardman, L. O. Johannissen, B. Bellina, H. L. Messiha, J. M. Ortiz-Guerrero, M. Elías-Arnanz, S. Padmanabhan, P. Barran, N. S. Scrutton and A. R. Jones, *Nat. Commun.*, 2015, **6**, 7907.

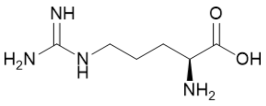
17 I. S. Camacho, R. Black, D. J. Heyes, L. O. Johannissen, L. A. I. I. Ramakers, B. Bellina, P. E. Barran, S. Hay and A. R. Jones, *Chem. Sci.*, 2021, **12**, 8333–8341.

18 F. C. Liu, M. E. Ridgeway, J. S. R. V. Winfred, N. C. Polfer, J. Lee, A. Theisen, C. A. Wootton, M. A. Park and C. Bleiholder, *Rapid Commun. Mass Spectrom.*, 2021, **35**, 1–12.

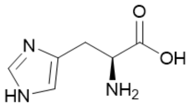
19 B. P. Marzullo, T. E. Morgan, A. Theisen, A. Haris, C. A. Wootton, S. J. Perry, M. Saeed, M. P. Barrow and P. B. O’Connor, *Anal. Chem.*, 2021, **93**, 9462–9470.

Appendix A: Amino acid structures and abbreviations

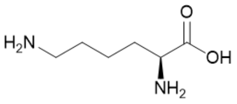
Arginine



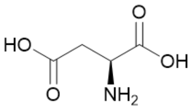
Histidine



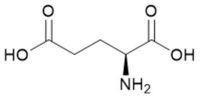
Lysine



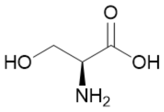
Aspartic Acid



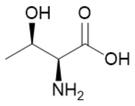
Glutamic Acid



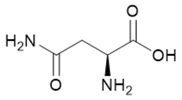
Serine



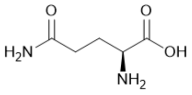
Threonine



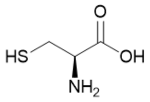
Asparagine



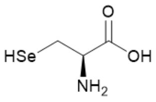
Glutamine



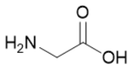
Cysteine



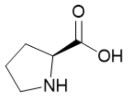
Selenocysteine



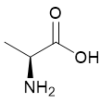
Glycine



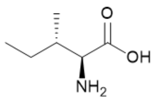
Proline



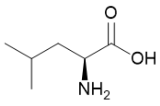
Alanine



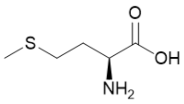
Isoleucine



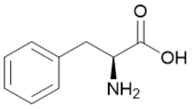
Leucine



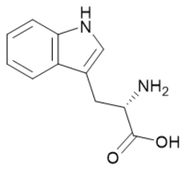
Methionine



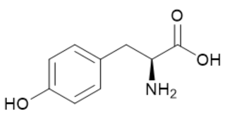
Phenylalanine



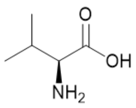
Tryptophan



Tyrosine



Valine



Appendix B: UVPD paper

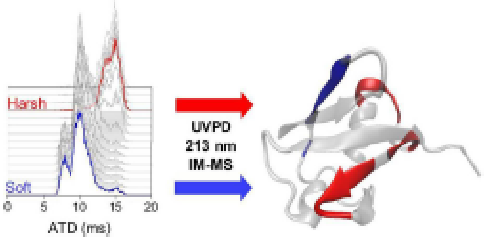
J. Am. Soc. Mass Spectrom. (2018) 30:24–33
DOI: 10.1007/s13361-018-1992-0

FOCUS: HONORING CAROL V. ROBINSON'S ELECTION TO THE NATIONAL ACADEMY OF SCIENCES: RESEARCH ARTICLE

Initial Protein Unfolding Events in Ubiquitin, Cytochrome c and Myoglobin Are Revealed with the Use of 213 nm UVPD Coupled to IM-MS

Alina Theisen,[1] Rachelle Black,[1] Davide Corinti,[2] Jeffery M. Brown,[3] Bruno Bellina,[1] Perdita E. Barran[1]

[1]Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology and Photon Science Institute, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK
[2]Dipartimento di Chimica e Tecnologie del Farmaco, Università di Roma "La Sapienza", 00185, Rome, Italy
[3]Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, UK



Abstract. The initial stages of protein unfolding may reflect the stability of the entire fold and can also reveal which parts of a protein can be perturbed, without restructuring the rest. In this work, we couple UVPD with activated ion mobility mass spectrometry to measure how three model proteins start to unfold. Ubiquitin, cytochrome c and myoglobin ions produced via nESI from salty solutions are subjected to UV irradiation pre-mobility separation; experiments are conducted with a range of source conditions which alter the conformation of the precursor ion as shown by the drift time profiles. For all three proteins, the compact structures result in less fragmentation than more extended structures which emerge following progressive in-source activation. Cleavage sites are found to differ between conformational ensembles, for example, for the dominant charge state of cytochrome c $[M+7H]^{7+}$, cleavage at Phe10, Thr19 and Val20 was only observed in activating conditions whilst cleavage at Ala43 is dramatically enhanced. Mapping the photo-cleaved fragments onto crystallographic structures provides insight into the local structural changes that occur as protein unfolding progresses, which is coupled to global restructuring observed in the drift time profiles.

Keywords: UVPD, IM-MS, Ubiquitin, Myoglobin, Cytochrome c

Received: 5 April 2018/Revised: 8 May 2018/Accepted: 9 May 2018/Published Online: 13 June 2018

Introduction

Protein folding, the process by which an amino acid chain adopts a 3D structure is correlated to protein function and heavily regulated in the cell. Many pathways and intermediate structures form an accessible landscape towards a low-energy structure termed the native-state [1]. Methods such as X-ray crystallography [2], nuclear magnetic resonance (NMR) [3], circular dichroism (CD) [4] and ion mobility-mass spectrometry (IM-MS) [5, 6] are often utilised to characterise protein conformation, and spectroscopic methods are employed to examine dynamics [7, 8].

IM-MS can readily report on conformational ensembles that a given protein can occupy at the same time [9, 10]. In ion mobility, ions progress through a buffer-gas-filled tube pulled by a weak electric field. Collisions with the inert gas molecules retard ions whilst the number and extent of collisions experienced depend on the ion's size and shape, resulting in

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13361-018-1992-0) contains supplementary material, which is available to authorized users.

Correspondence to: Perdita Barran; *e-mail:* perdita.barran@manchester.ac.uk separation that is reflective of the ion's size, shape and charge. Due to the electric field and collisions, ions 'tumble' and rotate as they drift through the tube, and ion mobility drift times can be converted into (or calibrated to obtain) a rotationally averaged collision cross-section, effectively a global structure measurement [11, 12].

The vast majority of studies of the structure of proteins in the gas phase utilise ESI with some modifications from methods used for rapid analysis of mass alone [13]. Nanoelectrospray (nESI) tends to be used more routinely for the analysis of biomolecules with the benefit of low sample consumption and lower spraying voltages, representing an even softer ionisation approach aiding the conservation of native-like conformations [14–16]. In order to retain a protein's native structure in mass spectrometry (native MS), solution conditions have to be used that mimic the native environment. Aqueous salts with high volatility such as ammonium acetate can be used to prevent disruption to the non-covalent interactions present in solution. They act to stabilise the protein as it transfers from solution to the gas phase, whilst not being retained excessively such that a series of ions corresponding to different net charges is resolved.

Of course, the question presents itself whether the observed species in vacuo correspond to structures present in solution, and if they do, how far these similarities go. A number of mass spectrometry experiments have demonstrated that it is possible to at least partially retain solution structure and topology [17–25]. Careful selection of ionisation parameters has allowed analysis of macromolecular complexes keeping non-covalent interactions such as protein-ligand or protein-cofactor interactions [17], complete ribosomes [26] and even viral capsids [27] intact. Following electrospray ionisation, the mass spectrum of a given protein reflects the conformational spread that is present in the electrospray solution, prior to transfer to the gas phase. Typically, higher charge states and a wider charge state distribution are seen for unfolded proteins as the solvent-accessible surface area increases [28–33].

Once in the gas phase, a plethora of activation techniques can be applied to biomolecules to obtain information on structure, sequence and potential interactions that may occur [34–37]. The most widely used is collision-induced activation (CID), which has been used for both 'bottom-up' experiments where proteolytic fragments are analysed and for 'top-down' analysis of the intact protein [38]. The CID process is slow and involves randomisation of energy throughout the molecule, such that is it difficult to obtain topological information from the fragment spectra, even when the fragments are subunits of a given protein complex. In contrast, topological information can be obtained using surface-induced dissociation (SID). This imparts energy to the molecule in a single, fast collision with a surface yielding subunits from complexes. The SID process can be optimised such that dissociation occurs prior to unfolding yielding far better insights to the arrangement of subunit in a given protein complex [39, 40]. Unlike CID, electron capture dissociation (ECD) is a non-ergodic process wherein capture and cleavage occur prior to statistical rearrangement of energy and can be used to probe structure [25]. Unfortunately, ECD is less efficient for low-charge-state ions which are more reflective of the native fold [41, 42]. Electron-transfer dissociation (ETD) is also used for protein structure analysis although some of the caveats for ECD also apply [34]. The efficiency of both ECD and ETD can be increased with the addition of CID however this will perturb the structure in vacuo.

An alternative approach couples laser activation with mass spectrometry termed photodissociation, whereby photons target specific regions within the protein and induce fragmentation [43–47]. Ultraviolet photodissociation (UVPD) utilises high-energy photons causing fragmentation in a single process [48]. The nature of the cleavage and the ensuing photofragment will depend on the wavelength used. For example, at 266 nm, efficient UVPD requires the presence of an aromatic amino acid such as tryptophan, tyrosine or phenylalanine which acts as a chromophore to mediate backbone cleavage [49, 50]. At 'deeper' UV wavelengths, photons are also absorbed by the protein backbone [50]. Brodbelt and co-workers have contributed a large body of work which in the main utilises 193 nm UVPD and have applied this to bottom-up as well as top-down analysis, including phosphate location and investigation of protein topology [51–53]. Stereo-chemical information on cis/trans isomerisation of proline residues in polypeptides has been found by Reilly and co-workers at 157 nm and von Helden and co-workers working at 193 nm [54, 55]. 213 nm UVPD has been used by two independent studies to fragment deuterium-labelled peptides both of which show successful retention of selective sites of deuterium labelling throughout the protein backbone, providing an alternative activation method to electron-mediated approaches [56, 57]. Girod et al. have recently shown that 213 nm UVPD of proline-containing proteins is able to provide sequence information including unusual fragment ions a+2 and b+2, previously seen with 157 nm UVPD, and attributed to prolines [54, 58].

The ability of UVPD at 193 nm to assess the conformation of proteins has been investigated previously by Cammarata and Brodbelt for holo- and apo-horse heart myoglobin on native-like charge states (8+, 9+ and 10+), showing that the fragmentation yield correlates with the rigidity of structural elements [59]. Warnke et al. have incorporated ion mobility separation and selection with 193 nm UVPD, demonstrating that for ubiquitin 7+, irradiation of specific portions of the arrival time distribution produces both similar and distinctive fragmentation patterns that could indicate interconversion of species [60]. Ly and Julian utilised 266 nm UVPD to cleave a carbon-iodide bond incorporated into Tyr59 of ubiquitin and subsequent radical-directed dissociation cleavage sites provided distance constraints for molecular modelling conformational searches, yielding structures that were distinctly different to the crystal structure and in agreement with data from ECD [25, 61]. We have also previously observed differences in the fragmentation pattern of two different, distinct conformational families of 5+ melittin as probed by 266 nm UVPD [62]. Overall, there is increasing evidence that ultraviolet photodissociation, especially in combination with IM-MS, is a promising avenue for protein-structural characterisation.

In this study, we report how UVPD can be used to investigate changes in protein structure that occur during the transfer of proteins into the gas phase, by coupling a 213 nm laser to an ion mobility mass spectrometer. In turn, this shows how this approach could be used to provide topological information on the non-covalent interactions that stabilise protein fold. We use three well-studied proteins ubiquitin, cytochrome c and myoglobin and for each ion mobility data is used to determine ionisation conditions which preserve a compact native-like conformation termed 'gentle' or 'soft' to more elongated states termed 'activating' or 'harsh'. As we alter the conditions, the appearance of distinct photofragments is monitored by mass spectrometry and ion mobility.

Experimental Section

Materials

Methanol (> 99.9% purity) was purchased from Sigma-Aldrich (UK). Ammonium acetate was supplied by Fisher Scientific (Loughborough, UK). Ultrapure water was obtained from a Milli-Q Advantage A10 ultrapure water filtration system (Merck Millipore, Darmstadt, Germany).

Bovine ubiquitin, equine myoglobin and bovine cytochrome c were purchased from Sigma-Aldrich (UK) as lyophilized powders with purities of ≥ 98, 90 and 95% respectively. Proteins were dissolved in 200 mM ammonium acetate. Myoglobin in ammonium acetate was desalted twice using Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA, US).

All samples were diluted to a final protein concentration of 10 μM.

Instrumental Setup for UVPD-IM

Modest modifications have been made to a Waters Synapt G2-S, in addition to those described previously to facilitate better injection of a laser beam in the trap cell region [56, 63]. Essentially, a CaF_2 window was integrated into the source block and a hole was drilled into the stepwave cap allowing the laser beam to enter into the upper part of the ions' pathway to the time-of-flight region where the beam exits via the mirror inserted in the push plate. UVPD is carried out pre-ion mobility separation in the trap cell region of the instrument. The general trapping procedure has been described previously [63] but in brief: a DC potential gate is applied to the trap cell therefore accumulating ions prior to the mobility cell; following this, the t-wave amplitude is set to zero volts and the ion influx is stopped by imposing a stopping DC potential on the trap cell entrance electrode. Subsequently, a mechanical shutter opens, allowing the laser beam to interact with the confined ions. The laser beam (CryLaS, Germany) operates at 1 kHz and delivers 213 nm photons with an average pulse energy of 2 to 3 μJ. After photoactivation, the shutter closes and the ions are pulsed into the ion mobility cell by re-enabling the t-waves in the trap cell and synchronising the fall of the DC potential gate with the IMS cycle. Ion mobility separation occurs and ions proceed to the time-of-flight detector allowing drift time measurement and mass analysis of both precursor and photoproduct ions.

Experimental Workflow

All samples were analysed in positive ion mode using a nanoESI source. Borosilicate glass capillaries (World Precision Instruments, Stevenage, UK) were pulled in house on a Flaming/Brown P-1000 micropipette puller (Sutter Instrument Company, Novato, CA, US). Capillaries were filled with sample using a 10-μl Hamilton syringe (Hamilton Company, Reno, NV, US) and a platinum wire was inserted into the solution to allow application of positive voltage.

Spraying voltages were kept as low as possible to achieve gentle ionisation with a typical capillary voltage of around 1.1 kV, source temperature of 40 °C, and a trap bias settings of 30–35 V to reduce protein activation.

For the UVPD-IM experiments performed on the Synapt G2-S, the species of interest were *m/z* selected in the quadrupole, accumulated in the trap cell for 2 s in order to obtain an ion count of ~ 2e3 and ions were then photoactivated for 2 s before *m/z* and drift-time analysis. This procedure was repeated for each protein at different cone voltages ranging from 10 V up to 150 V. All UVPD experiments were done in triplicate over at least 2 separate days and results averaged between repeats. Normalised fragmentation yield per cleavage site (residue) was calculated according to this relationship:

$$\text{Normalised fragmentation yield} = \frac{\sum \text{ions from specific cleavage site}}{\sum \text{ions from all cleavage sites} + \text{precursor}}$$

aIM-MS experiments were performed on a Synapt G2-Si. The drift time of a mass-selected precursor ion was recorded whilst either the cone voltage or trap collision voltage was ramped in increments of 2 V from 10 to 150 V and 4 to 150 V respectively. Data analysis was done using MassLynx v4.1 (Waters Corporation, USA), ORIGAMI [64], OriginPro 9.1 (OriginLab Corporation, USA) and Microsoft Excel 2010 (Microsoft, USA).

Results and Discussion

Ubiquitin

Bovine ubiquitin was sprayed from a 200 mM ammonium acetate solution and observed as ions $[M+5H]^{5+}$ and $[M+6H]^{6+}$ (SI Figure 1). Upon altering the cone voltage from 10 to 120 V, no major changes were observed in the distribution of charge states in the mass spectrum.

Initially, the $[M+6H]^{6+}$ ion was subjected to either CID or UVPD, and for both activation techniques, the drift times of precursor and all fragment ions were monitored. As demonstrated in Fig. 1a, the use of CID activation (collisions) first

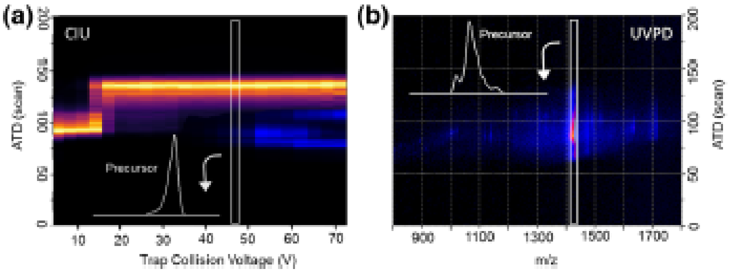


Figure 1. (**a**) aIM-MS and subsequent fragmentation of $[M+6H]^{6+}$ of ubiquitin. The drift time of the precursor ion (yellow) and of selected most intense CID fragments (blue) is plotted as a function of trap collision energy. All obtained fragment ions originate from the unfolded state of the protein as shown in inset ATD of precursor at fragmentation threshold. (**b**) Drift time versus *m/z* of all ions obtained after photoactivation (2 s) of $[M+6H]^{6+}$ using a 213 nm laser. UVPD fragments are obtained from the conformational arrangement shown in inset; no broadening of the precursor ATD is observed

causes the precursor ion to unfold. Fragmentation only starts to occur after initial restructuring and as the precursor ion extends more fragmentation is observed. Past a threshold energy of 210 eV (centre of mass frame), the arrival time no longer changes, and we conclude that energy imparted via collisions to the precursor ion is now only (or predominantly) released through fragmentation. As is commonly reported, CID activation initially breaks non-covalent interactions via internal energy conversion, which can promote restructuring of the protein, prior to bond cleavage. By contrast, when the same ion is subject to UVPD (electronic state activation), this does not alter the arrival time distribution of the precursor and by inference the non-covalent conformational arrangement, despite evidence for bond cleavage as shown in the dispersed low intensity fragment ion signal in Fig. 1b.

The cone voltage (CV) in the source was raised to systematically alter the precursor ion conformation prior to UV activation (Fig. 2b). At CV = 10 V (termed 'soft' conditions), $[M+6H]^{6+}$ presents in predominantly two conformational families at 8 and 10 ms with the later arriving being the most abundant. A minor component was observed at around 15 ms and as CV is raised this increases in intensity. For CV = 120 V (termed 'harsh' conditions), the two initially observed conformers are completely depleted (Fig. 2b). Figure 2a shows the UVPD spectra obtained at cone 10 V for the earliest arriving $[M+6H]^{6+}$ conformer (labelled 'soft') and 120 V for the fully unfolded state (labelled 'harsh'). In both cases, irradiation with 213 nm photons yields a-, b-, c- and y-type fragments.

Whilst fragmentation is already visible by the eye in the spectrum under 'soft' conditions, it increases dramatically with an increase in CV. The most intense fragments in both spectra result from cleavage at residues Glu18, Ile36 and Pro37, which are all on the N-terminal side of a proline residue. The effect of proline enhancing cleavage [66] has been commonly observed in MS/MS spectra of peptides and proteins [67]. No fragmentation was observed in the N-terminal beta strand or in the C-terminal part of the protein after residue Lys48, as well as only few cleavages at the C-terminal end of the alpha helix (see Fig. 2c). Comparing fragmentation yields per residue between 'soft' and 'harsh' (see SI Figure 2), the N-terminal region up to residue Glu18 fragments slightly more abundantly in the 'soft' condition. Between residue Lys33 and Tyr59, yield is significantly enhanced after initial unfolding of the precursor ion.

Generally, the fragments we observe agree with the higher intensity cleavage sites found in previous UVPD studies obtained at 193 nm [68]. The lesser extent of backbone cleavage we obtain is most likely due to the activation wavelength and therefore both a power and absorption cross-section difference.

Using UVPD post-ion mobility, Warnke et al. demonstrated that 193 nm UVPD is sensitive to underlying structural motifs within $[M+7H]^{7+}$ ubiquitin [60]. Our spectra of $[M+6H]^{6+}$ contains six out of the eight fragment motifs which were used to assign related conformers (y39, y40, y58, a36, b36 and a37). The two absent ones are higher charged species of two of these fragment ions most likely absent due to the lower charge state of the precursor ion here. Whilst we observe a general increase in the abundance of the mentioned fragmentation motifs when the ATD is shifted towards higher drift times, this does not appear to have been observed for $[M+7H]^{7+}$ when selectively irradiating portions of the ATD [60].

To ensure that the increase in fragmentation yield between 'soft' and 'harsh' conditions was a result of a change in conformation and not a result of increased internal energy from raising the cone voltage, UVPD spectra were obtained for $[M+6H]^{6+}$ at each cone voltage between 10 and 120 V in increments of 10 V (SI Figure 3). Changes in fragmentation yield were only observed where changes in cone voltage resulted in a change in arrival time distribution, indicating that the conformational change is the driving factor for fragmentation yield of $[M+6H]^{6+}$. In further support of this hypothesis, for $[M+8H]^{8+}$ sprayed from 50% methanol, we observed no change in conformation as a function of cone voltage and correspondingly no change in the UVPD fragmentation yield (SI Figure 4). Under the same instrumental conditions, the fragmentation yield does increase as a function of charge state (SI Figure 5) as noted by Brunet et al. for small negatively charged protein ions using photons in the VUV range [69] and by Cammarata and Brodbelt for positively charged myoglobin using 193 nm [59].

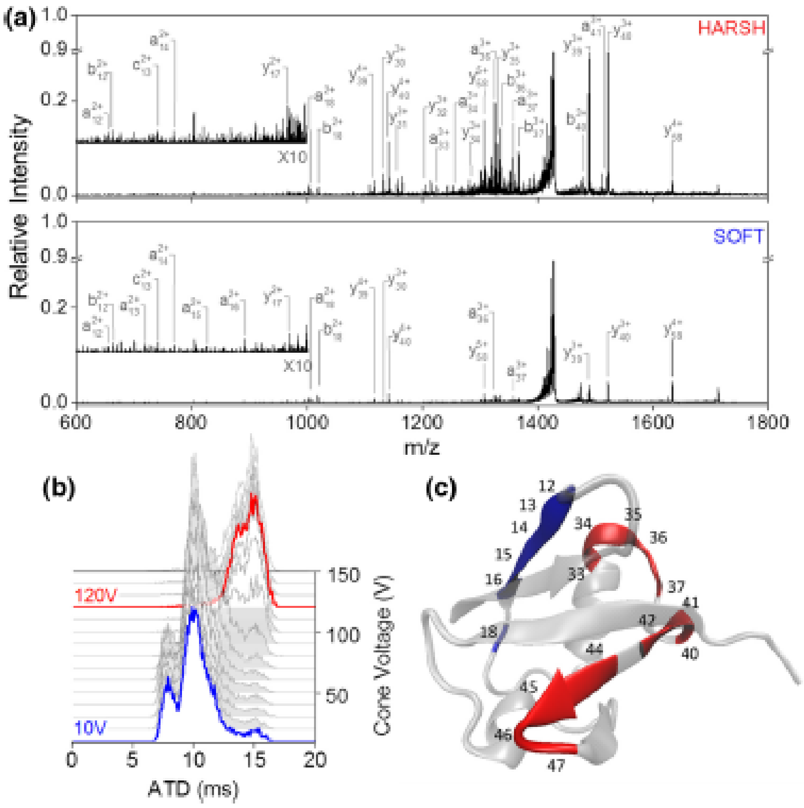


Figure 2. UVPD of two different conformational arrangements of $[M+6H]^{6+}$ ubiquitin. **(a)** UVPD spectra obtained at cone voltage 10 V (blue, bottom spectrum) and cone voltage 120 V (red, top spectrum). **(b)** Arrival time distribution as a function of cone voltage ranging from 10 to 120 V. **(c)** UVPD cleavage sites indicated on the crystal structure (PDB structure 4Z9S [65]) of ubiquitin. Sites with higher relative cleavage abundance in the 'soft' spectra are coloured blue; those with higher abundance in 'harsh' condition are marked red

Since we observe an increase in fragmentation yield as a function of unfolding, the question arises whether the fragments seen under 'soft' conditions come from the initial compact conformational families or the more extended conformational family already present in lower abundance. Comparison of the arrival time distribution of the precursor ion from laser off to laser on (SI Figure 6) shows a proportionally larger decrease in intensity of the more extended conformational families as opposed to a uniform decrease across all species, indicating that upon exposure to 213 nm UV photons, more extended conformations of $[M+6H]^{6+}$ give up fragments more readily. This indicates that the fragments observed at cone voltage 10 V may arise predominantly from direct UVPD from extended structures present under non-activating conditions.

We compared the arrival times of fragment ions which were observed in both 'soft' and 'harsh' conditions (SI Figure 7). There was no difference in arrival times of the investigated fragment ions between cone 10 and 120 V, implying that fragments from both source conditions do indeed arise from the same extended conformational families. It is also possible that fragment ions rearrange into a stable conformation on a timescale faster than we are able to measure. Further to this, some reports have indicated that lack of a specific UVPD fragment does not necessarily indicate strong non-covalent associations and suggests a combination of direct dissociation coupled with internal conversion (CID like) events [70, 71] however in ECD experiments it is well-established that lack of fragments correlates with strong non-covalent associations [24, 25, 36, 72]. For ubiquitin, as shown above, and as shall be shown for cytochrome c and myoglobin (below), we can interpret the UV-induced fragments in terms of regions of the proteins that are less encumbered by non-covalent associations. This indicates that in our experiment the UVPD fragments are observed when that region of the protein is detached from the native fold.

Cytochrome c

When sprayed out of 200 mM ammonium acetate, cytochrome c was observed as ions $[M+zH]^{z+}$ where $z=6$, 7 and 8, with $[M+7H]^{7+}$ as the dominant species (SI Figure 8). The charge state distribution did not change as a function of cone voltage;

however, a slight reduction in the intensity of salt adducts was seen under more activating conditions. $[M + 7H]^{7+}$ was selected for further investigation.

At CV = 10 V, $[M + 7H]^{7+}$ presents predominantly in a single conformer (~ 7 ms in the measured ATD) (Fig. 3b) with two minor species at 10 and 13 ms. Increasing CV to 80 V (termed 'intermediate' condition) causes a shift in the ion intensity from 7 to 10 ms with a slight increase in abundance at 13 ms. Raising the cone voltage further to 120 V causes complete depletion of the conformational family at 7 and 13 ms becomes the most intense peak in the ATD. Figure 3a shows the UVPD spectra obtained at 'soft', 'intermediate' and 'harsh' conditions. Similar to the results obtained for ubiquitin, the fragment types observed in the 213 nm UVPD spectra are a, b and y-type ions and the most intense ions are a result of N-terminal proline cleavage. In the 'soft' condition, fragmentation was only observed at 7 cleavage sites with very low intensity. Shifting the protein into an intermediate conformational arrangement increased the fragmentation yield and cleavage was now observed at 17 sites, and at 24 for CV = 120 V (see SI Figure 9). Interestingly, whilst a uniform increase in fragment intensity is observed in going from 'soft' to 'intermediate', this is not the case when transitioning from 'intermediate' to 'harsh'. Instead, a fragment-dependent increase or decrease occurs as coloured onto the crystal-structure of cytochrome c, with the overall fragmentation yield remaining constant (Fig. 3c). In the 'intermediate' condition, fragmentation is enhanced in the loop region between residue Lys25 and Phe36 whilst in the 'harsh' condition there is cleavage at residue Phe10 within the C-terminal helix as well as enhanced fragmentation after residue Leu35. Cleavage at residue Gly29 is abundant in both 'intermediate' and 'harsh'; however, the fragment identity is

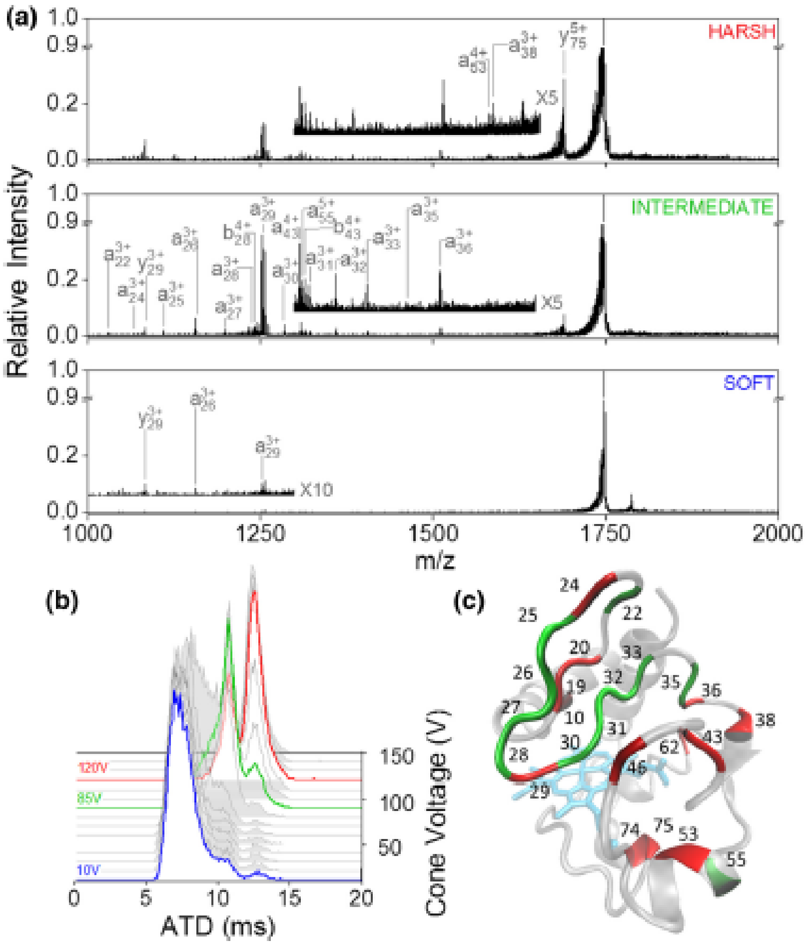


Figure 3. UVPD of 3 different conformational arrangements of $[M + 7H]^{7+}$ cytochrome C. (**a**) UVPD spectra obtained at cone voltage 10 V (bottom trace), 85 V (middle trace) and 120 V (top trace). (**b**) Arrival time distribution as a function of cone voltage. The conformations subjected to UVPD are highlighted in red, green and blue corresponding to 'soft', 'intermediate' and 'harsh 'respectively. (**c**) Backbone cleavage sites overlaid onto the crystal structure (PDB structure 2B4Z [73]) of cytochrome c. Sites with higher cleavage abundance in the 'intermediate' condition are coloured green; those with higher abundance in 'harsh' conditions are marked red

different between conditions. In the ‘intermediate’, a29 and b29 are the highly intense fragments from this cleavage site whilst in ‘harsh’, charge appears to be preferentially retained on the C-terminal side resulting in a dramatic increase in y75. In general, out of all fragments observed, only 3 are within secondary structural motifs other than loops or bends, none of which are found in the C-terminal helix. This is in agreement with solution phase work which showed the terminal helices to be the most stable structural elements [74]. No fragmentation was observed in the heme-binding CXXCH motif (residues Cys14 to His18) and all fragments containing these retained the heme group. This also indicates that the UVPD fragments are due to regions of the protein that are allowed to leave the core native fold, due to unstructuring in source.

Due to the difference in fragmentation pattern between ‘intermediate’ and ‘harsh’, it is likely that UVPD is probing both the intermediate and extended conformational families simultaneously. A comparison of the ATDs at each CV setting between laser off and on (SI Figure 10) shows no change in ATD at ‘soft’ and ‘intermediate’ conditions, indicating that all conformational families are probed equally. However, at CV = 120 V, the extended conformational family shows a disproportionally large decrease in intensity compared to the remainder of the more compact family. Fragment ion ATDs, as in the case of ubiquitin, did not change between conditions (SI Figure 11), indicating that fragments either refold to the same conformation during trapping or were produced from the same conformational family.

Myoglobin

Myoglobin sprayed from 200 mM ammonium acetate was observed in its holo form with the heme group still attached as ions $[M+zH]^{z+}$ where $z = 7, 8$ and 9, with $[M+9H]^{9+}$ as the dominant species (SI Figure 12), which was selected for further analysis.

For $[M+9H]^{9+}$, raising the cone voltage from 10 to 120 V (Fig. 4b) results in a complete depletion of the initial compact conformation centred at 5.5 ms and the appearance of a more extended conformation at 11 ms. Whilst UVPD of the compact conformation does not yield many fragment ions at all and cleavage was only observed at 5 residues (Fig. 4a), irradiation of the more extended species results in a range of intense a-, b- and y-type fragments with cleavage at 21 residues (SI Figure 13). Consistent with our findings so far, fragmentation yield increases

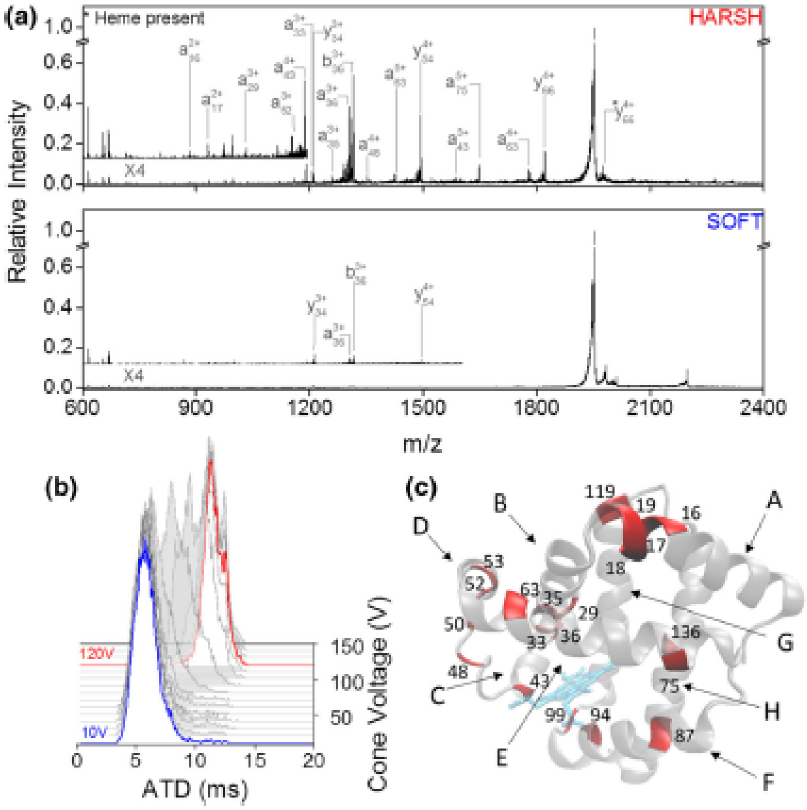


Figure 4. UVPD of 2 different conformational arrangements of $[M+9H]^{9+}$ holo-myoglobin. (a) UVPD spectra at cone voltage 10 V (termed ‘soft’, bottom panel) and at cone voltage 120 V (termed ‘harsh’, top panel). (b) Arrival time distribution as a function of cone voltage. At a cone voltage of 10 V (blue trace), the protein presents in a single conformational family centred around 6 ms whilst at cone voltage 120 V (red trace), the ATD shifts to 12 ms. (c) Backbone cleavage sites highlighted on the myoglobin crystal structure (PDB structure 3LR7 [76]) in red. Helices are labelled A to H

considerably after an initial unfolding event of the precursor ion away from the native compact structure. Whilst we do not obtain cleavage along most of the protein backbone and less fragmentation yield, the cleavage sites we do observe are in good agreement with the high-intensity sites observed by Holden et al. for 193 nm UVPD of the 8+ charge state that was obtained following charge reduction from $[M + 16H]^{16+}$ [75].

No difference was seen when comparing the precursor ATD before and after laser irradiation (SI Figure 14); hence, it is highly likely the absence of UVPD fragments under 'soft' conditions is a result of the compact conformation being impervious to dissociation due to the high degree of folding and strong stabilising non-covalent interactions. Out of the 21 cleavage sites, 7 are C-terminal to a lysine or glutamic acid residue. The most intense fragments were obtained on the N-terminal side of proline residues, namely a36, b36 and y54. In contrast to cytochrome c, residues in proximity to the heme group were cleaved (see Fig. 4c); however, some heme retention was observed on y66 highlighting the ability of UVPD at 213 nm to yield fragments which still possess non-covalent interactions, and the strength of the interaction between the cofactor and the protein.

The unfolding sequence of myoglobin as established by other techniques shows the F-helix to unfold first, followed by C-, D-, E- and B-helix, then lastly A, G and H [59, 77–81]. Consistent with the findings by Cammarata and Brodbelt for 193 nm UVPD of $[M + 9H]^{9+}$ myoglobin [59], the G and H-helix which have been shown to form the core of the protein as the most stable structures exhibit the least amount of fragmentation with only two cleavage sites in total. In contrast, however, the F-helix only fragmented at a single residue and therefore appeared stable in our experiment. Most fragmentation occurred towards the C-terminus of the A-helix, C-terminus of the B-helix, the CD loop and the D-helix.

Reducing the charge state of myoglobin from 9+ to 8+ eliminated fragmentation in the GH-loop, H helix and D helix in our experiment (SI Figures 15 and 16) as well as reducing overall fragmentation yield as found with ubiquitin. The cleavage sites with the most abundant fragmentation remain similar between $[M + 8H]^{8+}$ and $[M + 9H]^{9+}$ (SI Figure 17).

Conclusions

We have shown how the incorporation of UVPD before the ion mobility region can probe initial unfolding events in three different protein systems. Our data shows that, at 213 nm, UVPD is able to discriminate between compact structures and energetically activated more extended forms in terms of fragmentation yield and the intensity distribution within the fragment pattern (Fig. 5). Further to this, we can map the observed fragmentation and the lack of fragments to the crystallographic structure for each protein indicating that there is preservation of aspects of the native fold and that UVPD under these conditions acts as a direct mechanism for bond cleavage and that fragments do not arise due to internal conversion. Extended structures give rise to more observable fragmentation whilst compact structures do not yield as many fragment ions, which we suggest is due to the preponderance of non-covalent interactions tethering the native fold despite probable UV induced cleavage. All changes to protein structure were invoked by raising a single source potential which, whilst detrimental to the compact protein conformations, did not in any case alter the mass spectrum. This highlights the importance of careful tuning of ionisation parameters and in particular when inferring from UVPD data regarding the structure of proteins. We suggest this dataset of conformer-dependent fragments and the approach may be used to establish suitable conditions from which to examine proteins with unknown fold with ab initio UVPD.

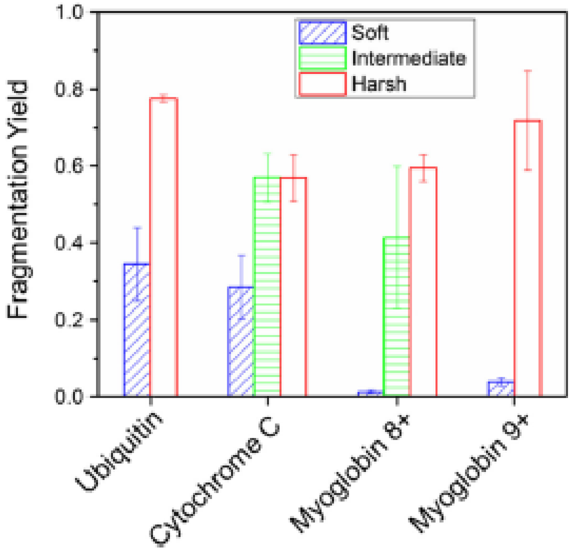


Figure 5. Overall UVPD fragmentation yield per protein compared between instrumental conditions. For all proteins measured, yield increases considerably for more extended conformations. The most dramatic change is observed for myoglobin

Acknowledgements

DC is grateful to the Universita' degli Studi di Roma La Sapienza for funding his stay at the University of Manchester.

Funding Information

This work was supported by BBSRC grants BB/L002655/1, BB/L016486/1 and BB/M011208/1 which is a studentship award to AT and RB with additional financial support from Waters Corp.

Open Access

References

1. Dobson, C.M.: Protein folding and misfolding. Nature. **426**, 884–890 (2003)
2. Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., Deponte, D.P., Weierstall, U., Doak, R.B., Maia, F.R.N.C., Martin, A.V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R.L., Epp, S.W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M.J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Ströder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K.U., Messerschmidt, M., Bozek, J.D., Hau-Riege, S.P., Frank, M., Hampton, C.Y., Sierra, R.G., Starodub, D., Williams, G.J., Hajdu, J., Timneanu, N., Seibert, M.M., Andreasson, J., Rocker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C.D., Krasniqi, F., Bott, M., Schmidt, K.E., Wang, X., Grotjohann, I., Holton, J.M., Barends, T.R.M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Spence, J.C.H.: Femtosecond X-ray protein nanocrystallography. Nature. **470**, 73–78 (2011)
3. Liang, B., Tamm, L.K.: NMR as a tool to investigate the structure, dynamics and function of membrane proteins. Nat. Struct. Mol. Biol. **23**, 568–474 (2016)
4. Roder, H., Colón, W.: Kinetic role of early intermediates in protein folding. Curr. Opin. Struct. Biol. **7**, 15–28 (1997)
5. Clemmer, D.E., Hudgins, R.R., Jarrold, M.F.: Naked protein conformations: cytochrome c in the gas phase. J. Am. Chem. Soc. **117**, 10141–10142 (1995)
6. Wyttenbach, T., Von Helden, G., Bowers, M.T.: Gas-phase conformation of biological molecules: bradykinin. J. Am. Chem. Soc. **118**, 8355–8364 (1996)
7. Kutta, R.J., Hardman, S.J.O., Johannissen, L.O., Bellina, B., Messiha, H.L., Ortiz-Guerrero, J.M., Elías-Arnanz, M., Padmanabhan, S., Barran, P., Scrutton, N.S., Jones, A.R.: The photochemical mechanism of a B12-dependent photoreceptor protein. Nat. Commun. **6**, 7907 (2015)
8. König, I., Zarrine-Afsar, A., Aznauryan, M., Soranno, A., Wunderlich, B., Dingfelder, F., Stüber, J.C., Plückthun, A., Nettels, D., Schuler, B.: Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. Nat. Methods. **12**, 773–779 (2015)
9. Beveridge, R., Migas, L.G., Payne, K.A.P., Scrutton, N.S., Leys, D., Barran, P.E.: Mass spectrometry locates local and allosteric conformational changes that occur on cofactor binding. Nat. Commun. **7**, 12163 (2016)
10. Stuchfield, D., Barran, P.: Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. Curr. Opin. Chem. Biol. **42**, 177–185 (2018)
11. Lanucara, F., Holman, S.W., Gray, C.J., Eyers, C.E.: The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. Nat. Chem. **6**, 281–294 (2014)
12. Giles, K., Pringle, S.D., Worthington, K.R., Little, D., Wildgoose, J.L., Bateman, R.H.: Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. Rapid Commun. Mass Spectrom. **18**, 2401–2414 (2004)
13. Fenn, J.B., Mann, M., Meng, C.K.A.I., Wong, S.F., Whitehouse, C.M.: Electrospray ionisation for mass spectrometry of large biomolecules. Science. **246**, 64–71 (1989)
14. Benesch, J.L.P., Ruotolo, B.T., Simmons, D.A., Robinson, C.V.: Protein complexes in the gas phase: technology for structural genomics and proteomics. Chem. Rev. **107**, 3544–3567 (2007)
15. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M.: Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature. **379**, 466–469 (1996)
16. Karas, M., Bahr, U., Dülcks, T.: Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J. Anal. Chem. **366**, 669–676 (2000)
17. Meyer, T., Gabelica, V., Grubmüller, H., Orozco, M.: Proteins in the gas phase. WIREs Comput. Mol. Sci. **3**, 408–425 (2013)
18. Hall, Z., Politis, A., Robinson, C.V.: Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. Structure. **20**, 1596–1609 (2012)
19. Loo, J.A., He, J.X., Cody, W.L.: Higher order structure in the gas phase reflects solution structure. J. Am. Chem. Soc. **120**, 4542–4543 (1998)
20. Jurneczko, E., Barran, P.E.: How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. Analyst. **136**, 20–28 (2011)
21. Deng, Z., Thontasen, N., Malinowski, N., Rinke, G., Harnau, L., Rauschenbach, S., Kern, K.: A close look at proteins: submolecular resolution of two- and three-dimensionally folded cytochrome c at surfaces. Nano Lett. **12**, 2452–2458 (2012)
22. Hewitt, D., Marklund, E., Scott, D.J., Robinson, C.V., Borysik, A.J.: A hydrodynamic comparison of solution and gas phase proteins and their complexes. J. Phys. Chem. B. **118**, 8489–8495 (2014)
23. Polfer, N.C., Haselmann, K.F., Langridge-Smith, P.R.R., Barran, P.E.: Structural investigation of naturally occurring peptides by electron capture dissociation and AMBER force field modelling. Mol. Phys. **103**, 11–12, 1481–1489 (2007)
24. Harvey, S.R., Porrini, M., Konijnenberg, A., Clarke, D.J., Tyler, R.C., Langridge-Smith, P.R.R., Macphee, C.E., Volkman, B.F., Barran, P.E.: Dissecting the dynamic conformations of the metamorphic protein lymphotactin. J. Phys. Chem. B. **118**, 12348–12359 (2014)
25. Breuker, K., Oh, H., Horn, D.M., Cerda, B.A., McLafferty, F.W.: Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation. J. Am. Chem. Soc. **124**, 6407–6420 (2002)
26. Wyttenbach, T., Bowers, M.T.: Intermolecular interactions in biomolecular systems examined by mass spectrometry. Annu. Rev. Phys. Chem. **58**, 511–533 (2007)
27. Uetrecht, C., Versluis, C., Watts, N.R., Wingfield, P.T., Steven, A.C., Heck, A.J.R.: Stability and shape of hepatitis B virus capsids in vacuo. Angew. Chem. Int. Ed. Engl. **47**, 6247–6251 (2008)
28. Frimpong, A.K., Abzalimov, R.R., Uversky, V.N., Kaltashov, I.A.: Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of alpha-synuclein. Proteins. **78**, 714–722 (2010)
29. Konermann, L., Douglas, D.J.: Acid-induced unfolding of cytochrome c at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure. Biochemistry. **36**, 12296–12302 (1997)
30. Konermann, L.: A minimalist model for exploring conformational effects on the electrospray charge state distribution of proteins. J. Phys. Chem. B. **111**, 6534–6543 (2007)
31. Seo, J., Hoffmann, W., Warnke, S., Bowers, M.T., Pagel, K., von Helden, G.: Retention of native protein structures in the absence of solvent: a coupled ion mobility and spectroscopic study. Angew. Chem. Int. Ed. **55**, 14173–14176 (2016)
32. Zhou, M., Morgner, N., Barrera, N.P., Politis, A., Isaacson, S.C., Matak-Vinković, D., Murata, T., Bernal, R.A., Stock, D., Robinson, C.V.: Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. Science. **334**, 380–385 (2011)
33. Beveridge, R., Covill, S., Pacholarz, K.J., Kalapothakis, J.M.D., Macphee, C.E., Barran, P.E.: A mass-spectrometry-based framework to define the extent of disorder in proteins. Anal. Chem. **86**, 10979–10991 (2014)
34. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F.: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. **101**, 9528–9533 (2004)
35. Zubarev, R.A., Horn, D.M., Fridriksson, E.K., Kelleher, N.L., Kruger, N.A., Lewis, M.A., Carpenter, B.K., McLafferty, F.W.: Electron capture dissociation for structural characterization of multiply charged protein cations. Anal. Chem. **72**, 563–573 (2000)
36. Polfer, N.C., Haselmann, K.F., Langridge-Smith, P.R.R., Barran, P.E.: Structural investigation of naturally occurring peptides by electron capture dissociation and AMBER force field modelling. Mol. Phys. **103**, 1481–1489 (2005)
37. Galhena, A.S., Dagan, S., Jones, C.M., Beardsley, R.L., Wysocki, V.H.: Surface-induced dissociation of peptides and protein complexes in a quadrupole/time-of-flight mass spectrometer. Anal. Chem. **80**, 1425–1436 (2008)

38. Kelleher, N.L.: Peer reviewed: top-down proteomics. Anal. Chem. **76**, 196 A–203 A (2004)
39. Zhou, M., Dagan, S., Wysocki, V.H.: Protein subunits released by surface collisions of noncovalent complexes: nativelike compact structures revealed by ion mobility mass spectrometry. Angew. Chem. Int. Ed. **51**, 4336–4339 (2012)
40. Beardsley, R.L., Jones, C.M., Galhena, A.S., Wysocki, V.H.: Noncovalent protein tetramers and Pentamers with " *n* " charges yield monomers with *n* /4 and *n* /5 charges. Anal. Chem. **81**, 1347–1356 (2009)
41. Breuker, K., Oh, H., Lin, C., Carpenter, B.K., McLafferty, F.W.: Nonergodic and conformational control of the electron capture dissociation of protein cations. Proc. Natl. Acad. Sci. U. S. A. **101**, 14011–14016 (2004)
42. Harvey, S.R., Porrini, M., Konijnenberg, A., Clarke, D.J., Tyler, R.C., Langridge-Smith, P.R.R., MacPhee, C.E., Volkman, B.F., Barran, P.E.: Dissecting the dynamic conformations of the metamorphic protein lymphotactin. J. Phys. Chem. B. **118**, 12348–12359 (2014)
43. Raspopov, S.A., El-Faramawy, A., Thomson, B.A., Siu, K.W.M.: Infrared multiphoton dissociation in quadrupole time-of-flight mass spectrometry: top-down characterization of proteins. Anal. Chem. **78**, 4572–4577 (2006)
44. Gardner, M.W., Smith, S.I., Ledvina, A.R., Madsen, J.A., Coon, J.J., Schwartz, J.C., Stafford Jr., G.C., Brodbelt, J.S.: Infrared multiphoton dissociation of peptide cations in a dual pressure linear ion trap mass spectrometer. Anal. Chem. **81**, 8109–8118 (2009)
45. Brodbelt, J.S.: Photodissociation mass spectrometry: new tools for characterization of biological molecules. Chem. Soc. Rev. **43**, 2757–2783 (2014)
46. Fort, K.L., Dyachenko, A., Potel, C.M., Corradini, E., Marino, F., Barendregt, A., Makarov, A.A., Scheltema, R.A., Heck, A.J.R.: Implementation of ultraviolet photodissociation on a benchtop Q exactive mass spectrometer and its application to phosphoproteomics. Anal. Chem. **88**, 2303–2310 (2016)
47. Zucker, S.M., Lee, S., Webber, N., Valentine, S.J., Reilly, J.P., Clemmer, D.E.: An ion mobility/ion trap/photodissociation instrument for characterization of ion structure. J. Am. Soc. Mass Spectrom. **22**, 1477–1485 (2011)
48. Yeh, G.K., Sun, Q., Meneses, C., Julian, R.R.: Rapid peptide fragmentation without electrons, collisions, infrared radiation, or native chromophores. J. Am. Soc. Mass Spectrom. **20**, 385–393 (2009)
49. Ly, T., Julian, R.R.: Ultraviolet photodissociation: developments towards applications for mass-spectrometry-based proteomics. Angew. Chem. Int. Ed. Engl. **48**, 7130–7137 (2009)
50. Reilly, J.P.: Ultraviolet photofragmentation of biomolecular ions. Mass Spectrom. Rev. **28**, 425–447 (2008)
51. Madsen, J.A., Boutz, D.R., Brodbelt, J.S.: Ultrafast ultraviolet photodissociation at 193 nm and its applicability to proteomic workflows. J. Proteome Res. **9**, 4205–4214 (2010)
52. Morrison, L.J., Brodbelt, J.S.: 193 nm ultraviolet photodissociation mass spectrometry of tetrameric protein complexes provides insight into quaternary and secondary protein topology. J. Am. Chem. Soc. **138**, 10849–10859 (2016)
53. Robinson, M.R., Taliaferro, J.M., Dalby, K.N., Brodbelt, J.S.: 193 nm ultraviolet photodissociation mass spectrometry for phosphopeptide characterization in the positive and negative ion modes. J. Proteome Res. **15**, 2739–2748 (2016)
54. Kim, T.-Y., Valentine, S.J., Clemmer, D.E., Reilly, J.P.: Gas-phase conformation-specific photofragmentation of proline-containing peptide ions. J. Am. Soc. Mass Spectrom. **21**, 1455–1465 (2010)
55. Warnke, S., Baldauf, C., Bowers, M.T., Pagel, K., von Helden, G.: Photodissociation of conformer-selected ubiquitin ions reveals site-specific cis/trans isomerization of proline peptide bonds. J. Am. Chem. Soc. **136**, 10308–10314 (2014)
56. Mistarz, U.H., Bellina, B., Jensen, P.F., Brown, J.M., Barran, P.E., Rand, K.D.: UV photodissociation mass spectrometry accurately localize sites of backbone deuteration in peptides. Anal. Chem. **90**, 1077–1080 (2018)
57. Brodie, N.I., Huguet, R., Zhang, T., Viner, R., Zabrouskov, V., Pan, J., Petrotchenko, E.V., Borchers, C.H.: Top-down hydrogen–deuterium exchange analysis of protein structures using ultraviolet photodissociation. Anal. Chem. **90**, 3079–3082 (2018)
58. Girod, M., Sanader, Z., Vojkovic, M., Antoine, R., MacAleese, L., Lemoine, J., Bonacic-Koutecky, V., Dugourd, P.: UV photodissociation of proline-containing peptide ions: insights from molecular dynamics. J. Am. Soc. Mass Spectrom. **26**, 432–443 (2015)
59. Cammarata, M.B., Brodbelt, J.S.: Structural characterization of holo- and apo-myoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. Chem. Sci. **6**, 1324–1333 (2015)
60. Warnke, S., von Helden, G., Pagel, K.: Analyzing the higher order structure of proteins with conformer-selective ultraviolet photodissociation. Proteomics. **15**, 2804–2812 (2015)
61. Ly, T., Julian, R.R.: Elucidating the tertiary structure of protein ions in vacuo with site specific photoinitiated radical reactions. J. Am. Chem. Soc. **132**, 8602–8609 (2010)
62. Theisen, A., Yan, B., Brown, J.M., Morris, M., Bellina, B., Barran, P.E.: Use of ultraviolet photodissociation coupled with ion mobility mass spectrometry to determine structure and sequence from drift time selected peptides and proteins. Anal. Chem. **88**, 9964–9971 (2016)
63. Bellina, B., Brown, J.M., Ujma, J., Murray, P., Giles, K., Morris, M., Compagnon, I., Barran, P.E.: UV photodissociation of trapped ions following ion mobility separation in a Q-ToF mass spectrometer. Analyst. **139**, 6348–6351 (2014)
64. Migas, L.G., France, A.P., Bellina, B., Barran, P.E.: ORIGAMI: a software suite for activated ion mobility mass spectrometry (aIM-MS) applied to multimeric protein assemblies. Int. J. Mass Spectrom. **427**, 20–28 (2018)
65. Levin-Kravets, O., Shohat, N., Prag, G.: Tetrameric assembly of monoubiquitin accurately mimics the Lys11 polyubiquitin chain structure. Biochemistry. **54**, 4704–4710 (2015)
66. Schwartz, B.L., Bursey, M.M.: Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. Biol. Mass Spectrom. **21**, 92–96 (1992)
67. Reid, G.E., Wu, J., Chrisman, P.A., Wells, J.M., McLuckey, S.A.: Charge-state-dependent sequence analysis of protonated ubiquitin ions via ion trap tandem mass spectrometry. Anal. Chem. **73**, 3274–3281 (2001)
68. Morrison, L.J., Brodbelt, J.S.: Charge site assignment in native proteins by ultraviolet photodissociation (UVPD) mass spectrometry. Analyst. **141**, 166–176 (2016)
69. Brunet, C., Antoine, R., Dugourd, P., Canon, F., Giuliani, A., Nahon, L.: Photo-induced electron detachment of protein polyanions in the VUV range. J. Chem. Phys. **138**, 64301 (2013)
70. Joshi, H.J., Narimatsu, Y., Schjoldager, K.T., Tytgat, H.L.P., Aebi, M., Clausen, H., Halim, A.: SnapShot: O-glycosylation pathways across kingdoms. Cell. **172**, 632–632.e2 (2018)
71. Julian, R.: R.: the mechanism behind top-down UVPD experiments: making sense of apparent contradictions. J. Am. Soc. Mass Spectrom. **28**, 1823–1826 (2017)
72. Breuker, K., Brüschweiler, S., Tollinger, M.: Electrostatic stabilization of a native protein structure in the gas phase. Angew. Chem. Int. Ed. **50**, 873–877 (2011)
73. Mirkin, N., Jaconcic, J., Stojanoff, V., Moreno, A.: High resolution X-ray crystallographic structure of bovine heart cytochrome c and its application to the design of an electron transfer biosensor. Proteins. **70**, 83–92 (2008)
74. Krishna, M.M.G., Maity, H., Rumbley, J.N., Englander, S.W.: Branching in the sequential folding pathway of cytochrome c. Protein Sci. **16**, 1946–1956 (2007)
75. Holden, D.D., McGee, W.M., Brodbelt, J.S.: Integration of ultraviolet photodissociation with proton transfer reactions and ion parking for analysis of intact proteins. Anal. Chem. **88**, 1008–1016 (2016)
76. Yi, J., Orville, A.M., Skinner, J.M., Skinner, M.J., Richter-Addo, G.B.: Synchrotron X-ray-induced photoreduction of ferric myoglobin nitrite crystals gives the ferrous derivative with retention of the O-bonded nitrite ligand. Biochemistry. **49**, 5969–5971 (2010)
77. Schwarzinger, S., Mohana-Borges, R., Kroon, G.J.A., Dyson, H.J., Wright, P.E.: Structural characterization of partially folded intermediates of apomyoglobin H64F. Protein Sci. **17**, 313–321 (2008)
78. Eliezer, D., Yao, J., Dyson, H.J., Wright, P.E.: Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. Nat. Struct. Biol. **5**, 148–155 (1998)
79. Dasmeh, P., Kepp, K.P.: Unfolding simulations of holomyoglobin from four mammals: identification of intermediates and β-sheet formation from partially unfolded states. PLoS One. **8**, e80308 (2013)
80. Lin, X., Zhao, W., Wang, X.: Characterization of conformational changes and noncovalent complexes of myoglobin by electrospray ionization mass spectrometry, circular dichroism and fluorescence spectroscopy. J. Mass Spectrom. **45**, 618–626 (2010)
81. Simmons, D.A., Dunn, S.D., Konermann, L.: Conformational dynamics of partially denatured myoglobin studied by time-resolved electrospray mass spectrometry with online hydrogen–deuterium exchange. Biochemistry. **42**, 5896–5905 (2003)

Supplementary Information

Initial protein unfolding events in Ubiquitin, Cytochrome c and Myoglobin are revealed with the use of 213 nm UVPD coupled to IM-MS

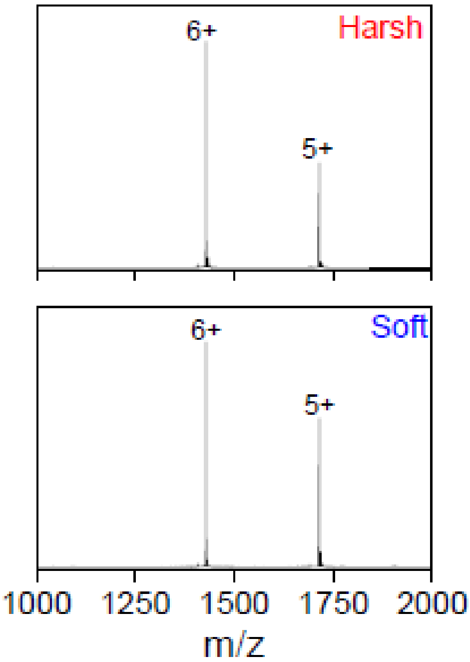
Alina Theisen[1], Rachelle Black[1], Davide Corinti[2], Jeffery M. Brown[3], Bruno Bellina[1], Perdita E. Barran[1]

[1] Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology and Photon Science Institute, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

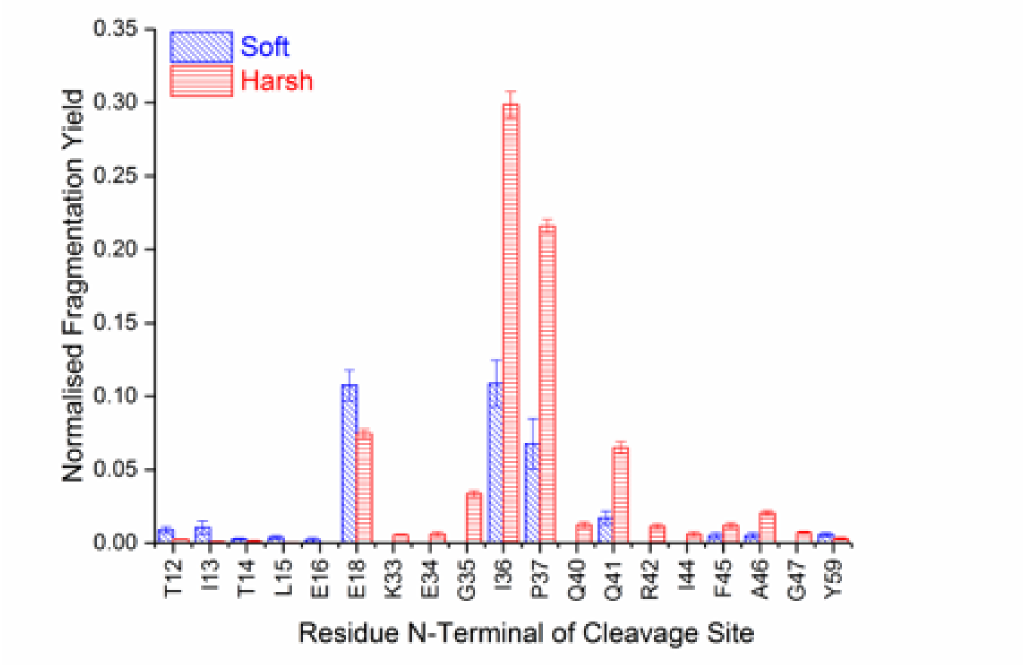
[2] Dipartimento di Chimica e Tecnologie del Farmaco, Università di Roma "La Sapienza", Roma I-00185, Italy

[3] Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, SK9 4AX, United Kingdom

Contents



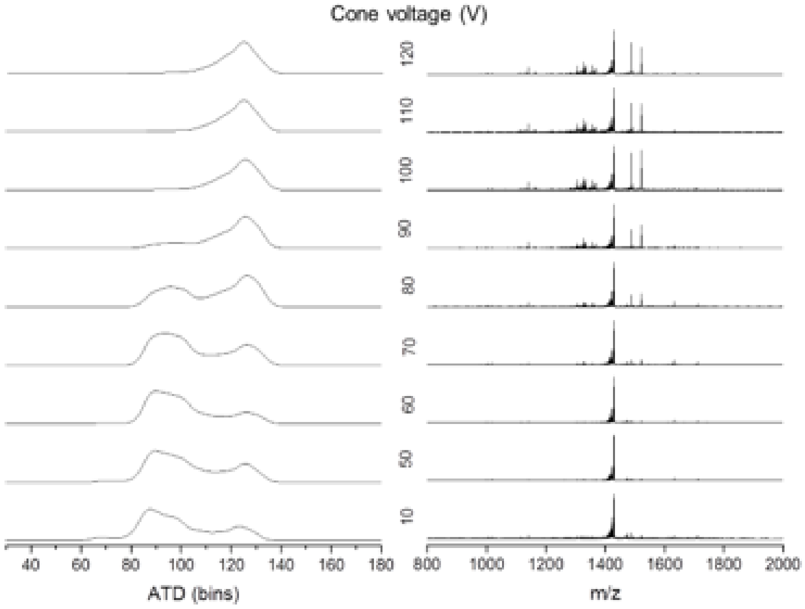
SI Figure 1: Mass spectra of ubiquitin sprayed from 200 mM ammonium acetate at cone voltage 10 V (bottom panel, 'soft') and at cone voltage 120 V (top panel, 'harsh').



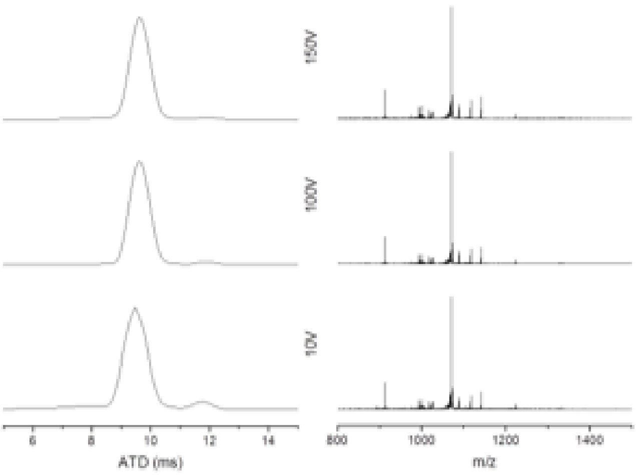
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG

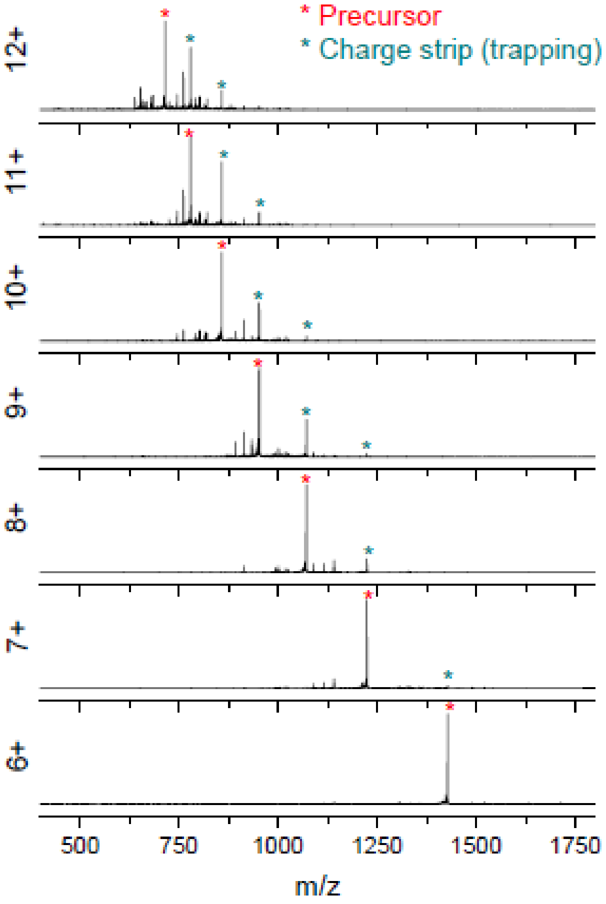
SI Figure 2: Fragmentation yield per cleavage site of [M+6H]$^{6+}$ ubiquitin. Normalised fragmentation yield was calculated by dividing the sum of ions cleaved at a specific residue by the sum of ions from all cleavage sites including the precursor.



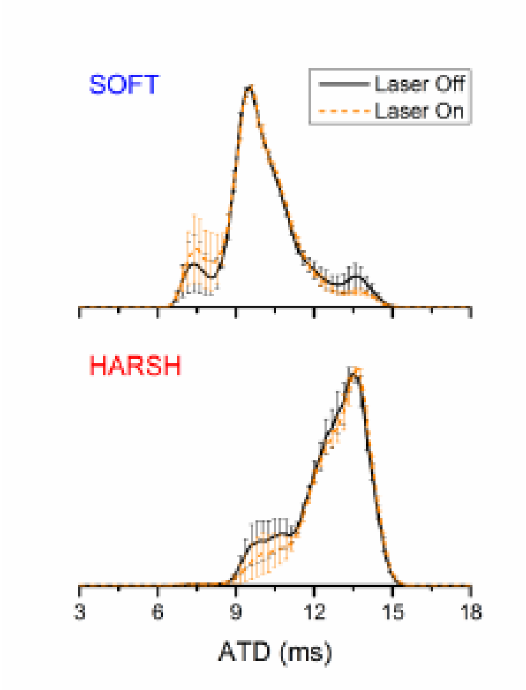
SI Figure 3: UVPD of $[M+6H]^{6+}$ ubiquitin as a function of cone voltage. Change in UVPD spectra is only observed when the cone voltage raise results in a change in ATD.



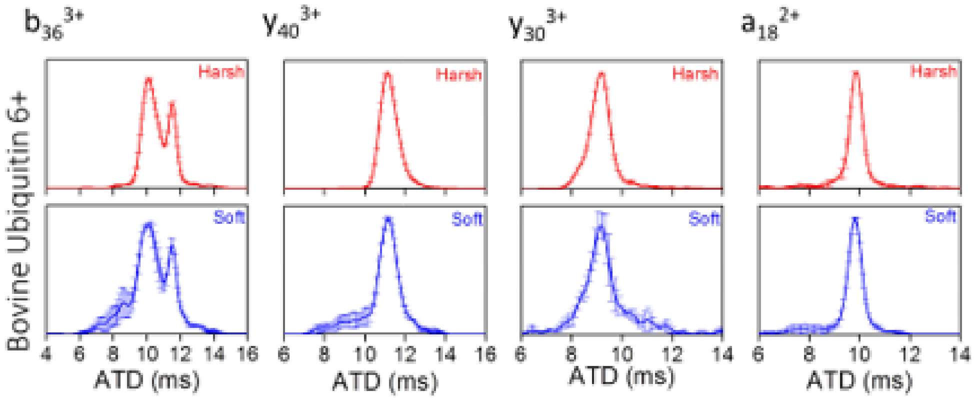
SI Figure 4: UVPD of $[M+8H]^{8+}$ ubiquitin as a function of cone voltage. No change is observed in conformation when the cone voltage is raised; the fragmentation yield stays constant throughout.



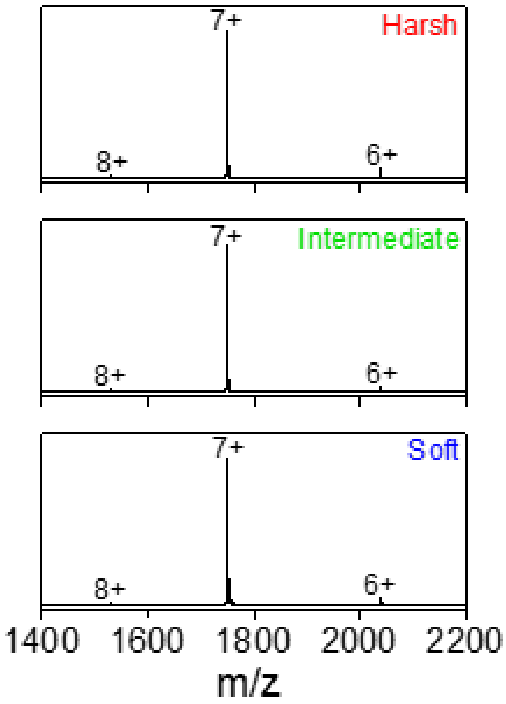
SI Figure 5: UVPD of Ubiquitin charge states 6+ to 12+. Ubiquitin was sprayed from 50% methanol with 0.1% formic acid. Instrumental conditions were kept the same for all charge states with a cone voltage of 10 V and trap bias of 35 V. Irradiation time was 2 seconds. Loss of charge (cyan asterix) from the precursor ion (red asterix) is a result of trapping.



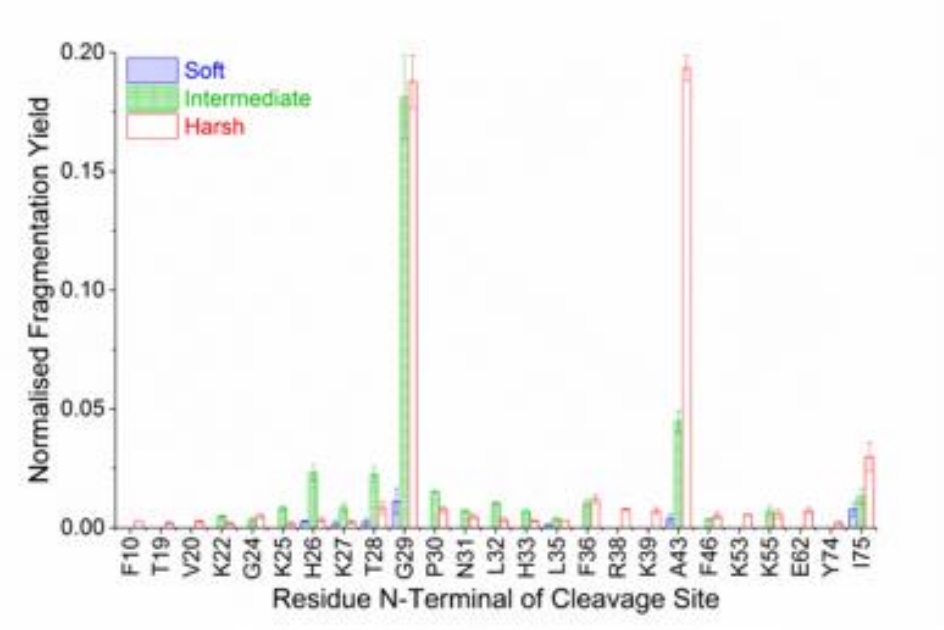
SI Figure 6: Comparison of ATDs of [M+6H]$^{6+}$ ubiquitin between laser off and laser on. Each ATD was normalised to the highest peak for comparison purposes as absolute signal intensity generally drops when the laser is switched on, even if no fragmentation occurs. In soft conditions (cone 10 V), the most compact family centred at 7.3 ms experienced the least decrease intensity while the most extended family at 13.5 ms experienced the most. In harsh conditions, the intermediate family at 9.6 ms was observed to decrease most.



SI Figure 7: Arrival time distributions of UVPD fragments of $[M+6H]^{6+}$ ubiquitin compared between soft (cone voltage 10) and harsh (cone voltage 120) conditions. ATDs did not change between conditions. Fragment intensity in soft conditions was much less than in harsh, resulting in much larger error bars.



SI Figure 8: Mass spectra of cytochrome c sprayed out of 200mM ammonium acetate recorded at cone voltage 10 V (bottom panel, 'soft'), 85 V (middle panel, 'intermediate') and 120 V (top panel, 'harsh'). Charge state distribution does not change between ionisation conditions, a slight reduction in salt adducts is observed as cone voltage is raised.

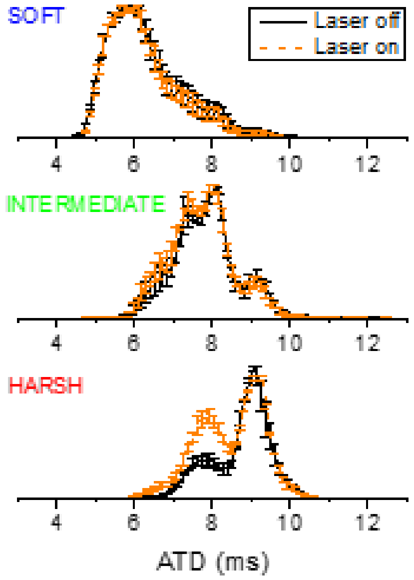


GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFA
GIKKKGEREDLIAYLKKATNE

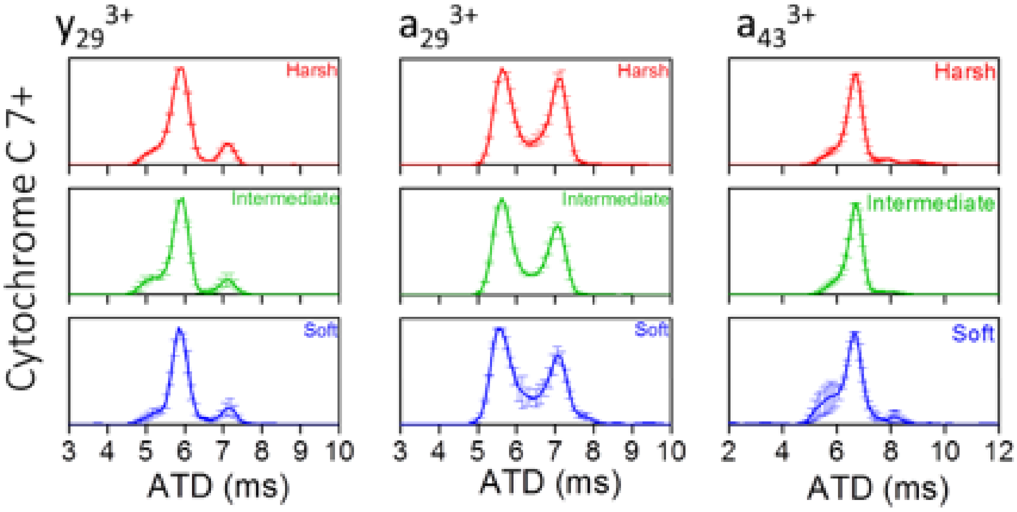
GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFA
GIKKKGEREDLIAYLKKATNE

GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFA
GIKKKGEREDLIAYLKKATNE

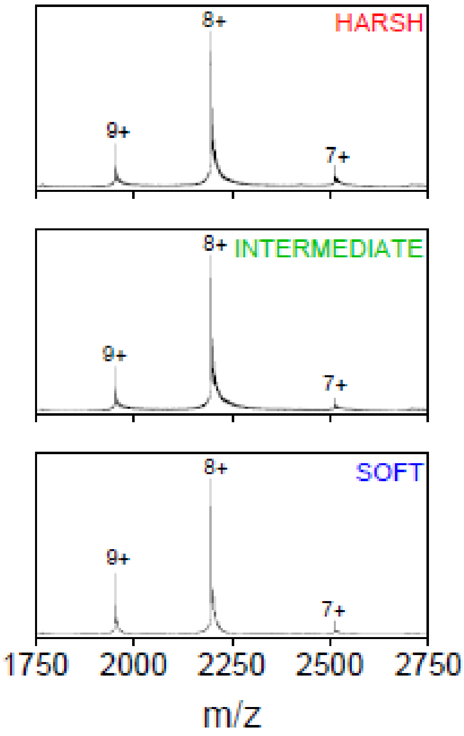
SI Figure 9: Fragmentation yield per residue of $[M+7H]^{7+}$ cytochrome c. Normalised fragmentation yield was calculated by dividing the sum of ions cleaved at a specific residue by the sum of ions from all cleavage sites including the precursor.



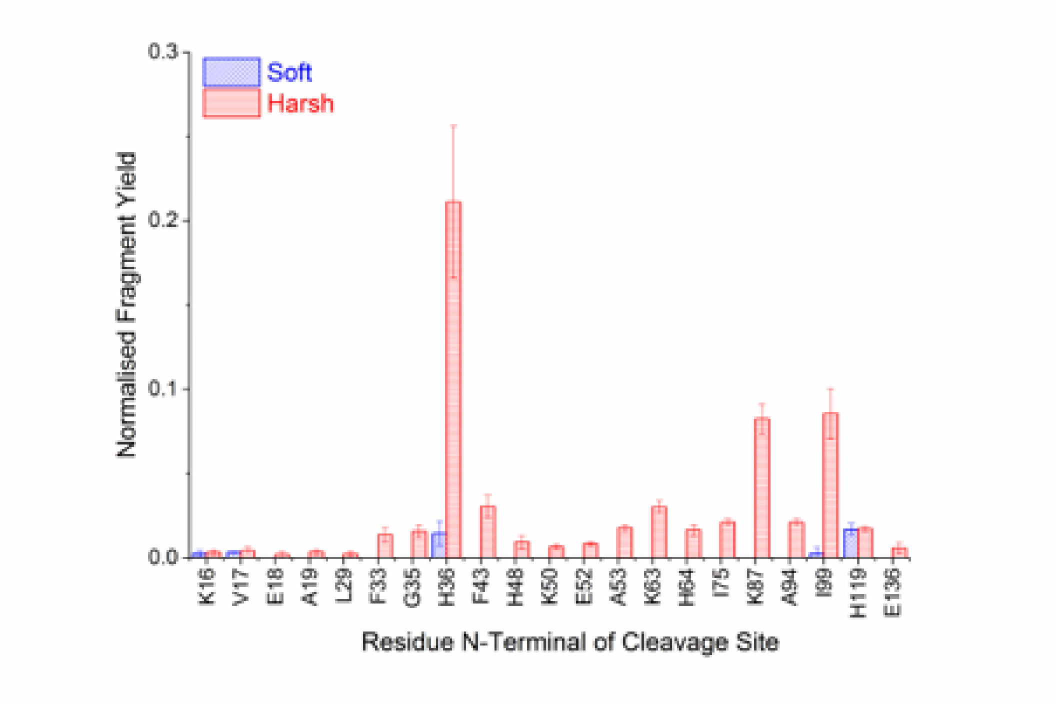
SI Figure 10: Comparison of arrival time distribution of cytochrome c $[M+7H]^{7+}$ between laser off and laser on. Each ATD was normalised to the highest peak for comparison purposes as absolute signal intensity generally drops when the laser is switched on, even if no fragmentation occurs. In soft and intermediate conditions, no portion of the ATD was found to differ between laser off and on, however, in harsh conditions the larger conformational family's intensity decreased significantly more than the more compact one.



SI Figure 11: Arrival time distribution of UVPD fragments of $[M+7H]^{7+}$ cytochrome compared between soft (cone voltage 10), intermediate (cone voltage 85) and harsh (cone voltage 120) conditions. No change in ATDs was observed. Fragment intensity in soft conditions was much less than in harsh, resulting in larger error bars.



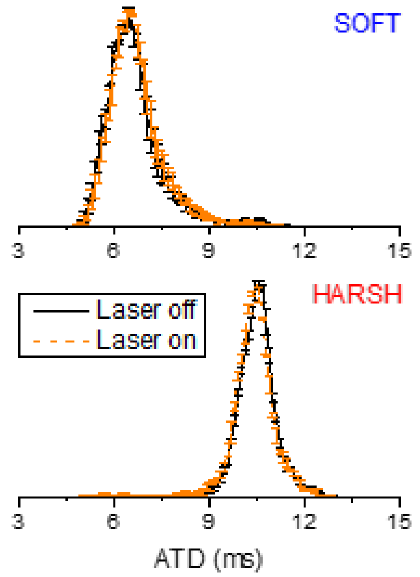
SI Figure 12: Mass spectra of equine myoglobin sprayed out of 200 mM ammonium acetate in soft (cone voltage 10), intermediate (cone voltage 105) and harsh (cone voltage 130) instrumental conditions. The mass spectrum does not differ between cone voltages.



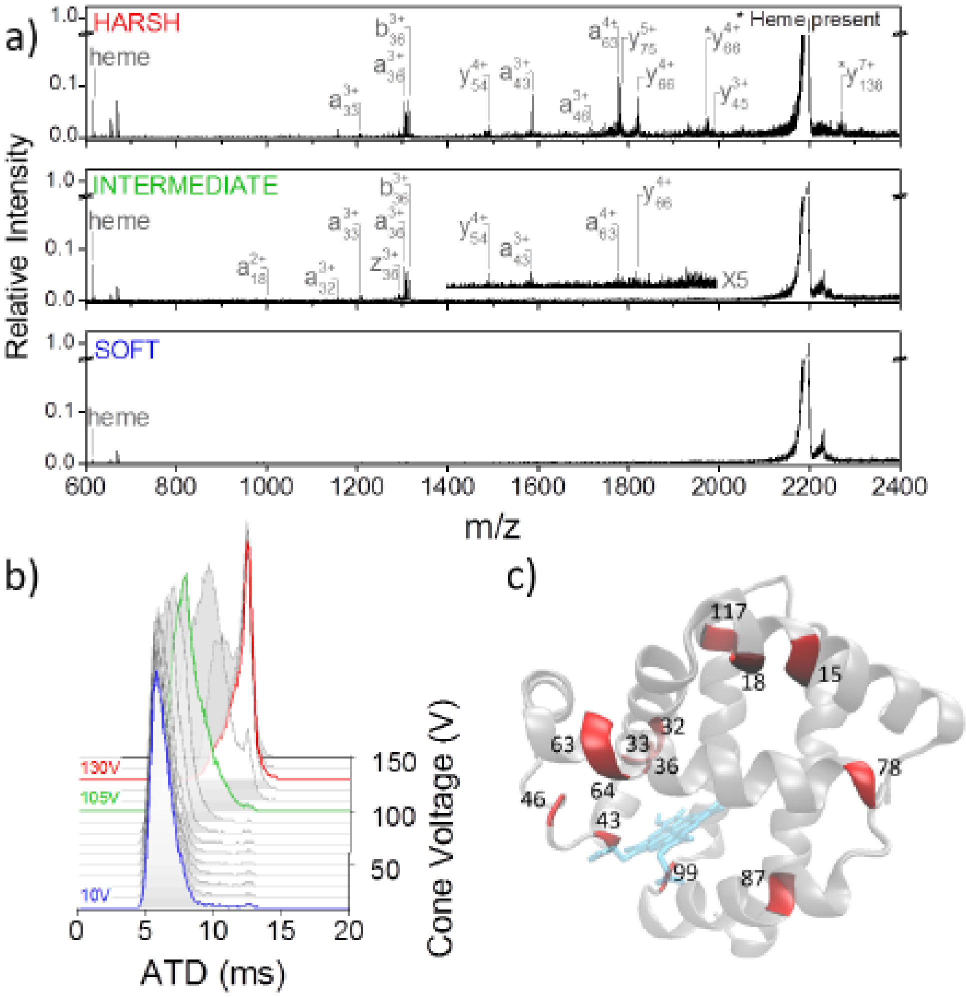
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEA
ELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEA
ELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

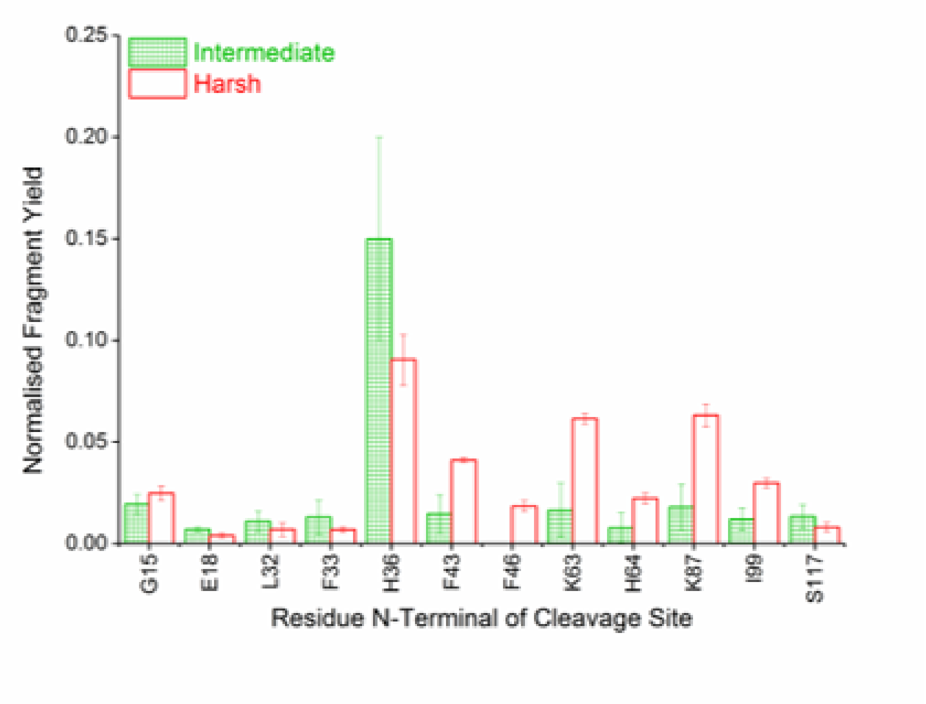
SI Figure 13: Normalised fragmentation yield per residue for $[M+9H]^{9+}$ myoglobin. Normalised fragmentation yield was calculated by dividing the sum of ions cleaved at a specific residue by the sum of ions from all cleavage sites including the precursor.



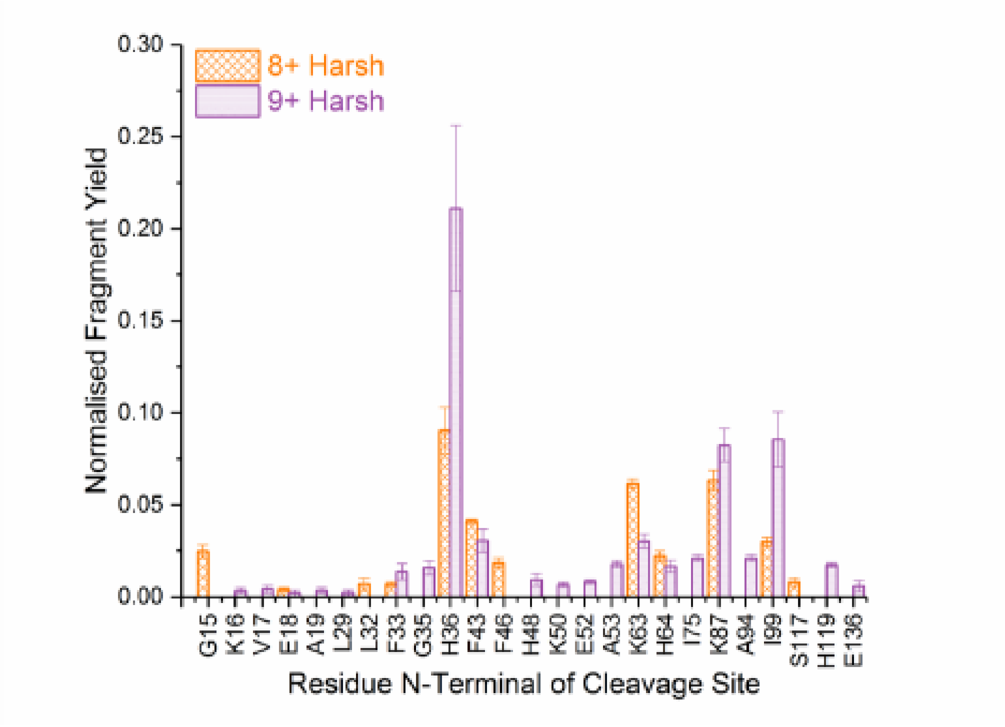
SI Figure 14: Comparison of ATDs of $[M+9H]^{9+}$ myoglobin between laser off and laser on. No change in ATD was detected in either soft or harsh conditions. Each ATD was normalised to the highest peak for comparison purposes as absolute signal intensity generally drops when the laser is switched on, even if no fragmentation occurs.



SI Figure 15: UVPD of three different conformational arrangements of $[M+8H]^{8+}$ holo-myoglobin. a) UVPD spectra obtained at cone voltage 10 V (blue, bottom spectrum), 105 V (green, middle spectrum) and 130 V (red, top spectrum). b) Arrival time distribution of $[M+8H]^{8+}$ myoglobin as a function of cone voltage. c) UVPD cleavage sites indicated in red on the crystal structure (PDB structure 3LR7).



SI Figure 16: Normalised fragmentation yield per residue for $[M+8H]^{8+}$ myoglobin. Normalised fragmentation yield was calculated by dividing the sum of ions cleaved at a specific residue by the sum of ions from all cleavage sites including the precursor.



SI Figure 17: Normalised fragmentation yield per residue compared between $[M+8H]^{8+}$ and $[M+9H]^{9+}$ myoglobin at harsh conditions. Normalised fragmentation yield was calculated by dividing the sum of ions cleaved at a specific residue by the sum of ions from all cleavage sites including the precursor.