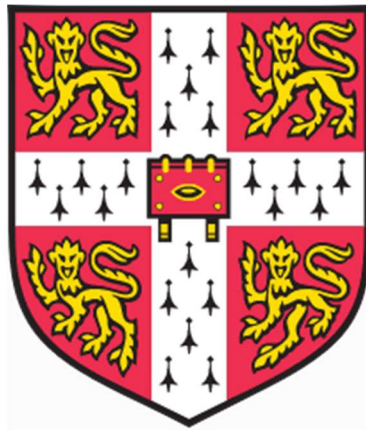


Investigation of mass spectrometry approaches for aiding fragment-based drug discovery



Daniel Shiu-Hin Chan

Christ's College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

February 2020

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit of 60,000 words, excluding the acknowledgements, list of abbreviations, table of contents, diagrams, figure captions, appendix and bibliography.

Summary

Investigation of mass spectrometry approaches for aiding fragment-based drug discovery

The use of mass spectrometry (MS)-based techniques to study protein structure dynamics is a relatively new field, with seminal developments such as nanoelectrospray ionization (nano-ESI) and traveling-wave ion mobility-mass spectrometry (TWIM-MS) MS being introduced only around twenty-five and fifteen years ago, respectively. However, MS methods have not yet found widespread use in drug discovery, particularly in fragment-based drug discovery (FBDD).

The work described in this thesis focuses broadly on the application of MS techniques to aid FBDD campaigns. In Chapter 1, an overview of the strengths and limitations of MS in FBDD is presented. In Chapter 2, native MS and IM-MS were used to explore the effect of DMSO, a common solvent used in fragment screening on protein size and stability in the gas-phase. In Chapter 3, native MS was used to study the structure of the EthR-DNA complex, a tuberculosis target, and to screen fragments against a protein-DNA interaction for the first time. In Chapter 4, native MS and IM-MS were used to study the structure and interactions of the antibacterial target CoaB and its ligands, including potential fragment inhibitors. In Chapter 5, native MS was used to study the protein-protein interaction between Aurora A, an anticancer target, and TPX-2, and its inhibition using chemical compounds. In Chapter 6, the overall conclusions and outlook of this project are presented. Overall, these studies demonstrate the utility of native MS approaches in providing complementary information to other biophysical techniques regarding protein structure and interactions in FBDD campaigns.

Daniel Shiu-Hin Chan

Acknowledgements

Firstly, I would like to express my immense gratitude to my co-supervisors Prof. Chris Abell and Dr. Anthony Coyne. I am grateful to Chris for providing me the opportunity to work in his group and for his support and insight, and to Anthony for his experience and encouragement in the laboratory and office over the course of my PhD. I am also deeply grateful to Dr. Dijana Matak-Vinkovic for training me on the various mass spectrometric instruments used this work and for her help and wisdom in times of need, and to Dr. Katherine Abell for her guidance in dealing with administrative and departmental issues. I would also like to thank the various members of the Abell group who have assisted me with this project, including Dr. Christina Spry, Dr. Jeannine Hess, Dr. Duncan Scott, Wei-guang Seetoh, Brendan McConnell, Andrew Whitehouse and Madeline Kavanagh, and to all other group members for their friendship and camaraderie. Moreover, I would like to Prof. Sir Tom Blundell and his group members Dr. Vitor Mendes, Dr. Michal Blaszczyk and Sherine Thomas, and Dr. Marko Hyvönen and his group members Dr. Mathieu Rocaboy, Dr. Maxim Rossman and Dr. Tomasso Moschetti for their collaboration.

I would like to acknowledge the support of the Croucher Foundation and the Cambridge Commonwealth, European and International Trust for receipt of a Croucher Cambridge International Scholarship.

Finally, I would like to thank my wife Natalie, who has accompanied me throughout this journey through thick and thin, and to my parents for their unwavering support and encouragement.

Abbreviations

1D	one-dimensional
2D	two-dimensional
ADH	alcohol dehydrogenase
ADP	adenosine-5'-diphosphate
ATD	arrival time distribution
ATP	adenosine-5'-triphosphate
AurA	Aurora A
AV	avidin
BEI	binding efficiency index
CCA	concanavalin A
CCS	collision cross-section
CD	circular dichroism
CID	collision-induced dissociation
CID50	half-maximal collision-induced dissociation energy
CIU	collision-induced unfolding
CIU50	half-maximal collision-induced unfolding energy
CoA	coenzyme A
CPMG	Carr-Purcell-Meiboom-Gill
CSP	chemical shift perturbations

CTP	cytidine-5'-triphosphate
CV	collision voltage
Da	dalton
DMSO	dimethyl sulfoxide
DSF	differential scanning fluorimetry
<i>Ec</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
FBDD	fragment-based drug discovery
HDX	hydrogen-deuterium exchange
HTS	high-throughput screening
IC ₅₀	half-maximal inhibitory concentration
IM-MS	ion mobility-mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_a	equilibrium association constant
K_d	equilibrium dissociation constant
LC-MS	liquid chromatography-mass spectrometry
LE	ligand efficiency
MDR	multidrug resistant
MS	mass spectrometry

<i>Mtb</i>	<i>Mycobacterium tuberculosis</i>
MWCO	molecular-weight cut-off
<i>m/z</i>	mass/charge
nESI	nanoflow-electrospray ionization
NH ₄ OAc	ammonium acetate
NMR	nuclear magnetic resonance
Pan	pantothenate
PDB	Protein Data Bank
PEG	polyethylene glycol
PK	pyruvate kinase
PPan	4'-phosphopantothenate
PPC	phosphopantothenoylcysteine
PPCDC	phosphopantothenoylcysteine decarboxylase
PPCS	phosphopantothenoylcysteine synthetase
qTOF	quadrupole time-of-flight
rpm	revolutions per minute
SAR	structure-activity relationship
SPR	surface plasmon resonance
STD	saturation transfer difference
TB	tuberculosis
TEV	tobacco etch virus

T_m	melting temperature
TPX2	targeting protein for Xklp2
TWIM	traveling wave ion mobility
WAC	weak affinity chromatography
WaterLOGSY	water-ligand observed via gradient spectroscopy
XDR	extensively drug-resistant

Contents

Declaration.....	i
Summary	ii
Acknowledgements.....	iii
Abbreviations	iv
Contents.....	viii
1. Introduction.....	1
1.1. The search for new drug discovery approaches	1
1.2. Fragment-based drug discovery (FBDD)	2
1.3. A biophysical cascade for FBDD	4
1.4. Mass spectrometry (MS) methods for drug discovery	7
1.4.1. Mechanism of electrospray ionization (ESI)	7
1.4.2. Native MS.....	10
1.4.3. Ion mobility-mass spectrometry (IM-MS).....	16
1.4.4. Collision-induced dissociation (CID) and collision-induced unfolding (CIU).....	18
1.4.5. Hydrogen-deuterium exchange (HDX)-MS	19
1.4.6. Ligand-based MS techniques	21
1.4.7. Other MS-based techniques	24
1.5. Project aims.....	25
2. Insight into effects of DMSO on protein charging, conformation, stability and interactions as assessed by native ion mobility mass spectrometry	27
2.1. Introduction	27
2.2. Results and discussion	29
2.2.1. Effect of DMSO on protein charge and conformation.....	29
2.2.1.1. DMSO exerts opposing trends on average protein charge.....	29
2.2.1.2. DMSO exerts opposing trends on average protein size.....	33
2.2.1.3. DMSO affects protein size at the level of individual charge states	34
2.2.1.4. Error analysis.....	40
2.2.1.5. Towards a consistent understanding of protein compaction by DMSO.....	44
2.2.1.6. Summary	46
2.2.2. Effect of DMSO on protein stability and interactions.....	47

2.2.2.1.	Charging behavior of avidin and CYP142A1 at high levels of DMSO	47
2.2.2.2.	Effect of high levels of DMSO on avidin and CYP142A1 size	50
2.2.2.3.	DMSO modulates the CID stability of avidin.....	52
2.2.2.4.	DMSO influences the stability and heme dissociation pathway of CYP142A1	57
2.2.2.5.	DMSO modulates the CIU stability of proteins.....	64
2.2.2.6.	Monitoring the protective effect of the heme group using CIU	71
2.2.2.7.	Summary	72
2.3.	Experimental	73
2.3.1.	Materials and methods.....	73
2.3.2.	Mass spectrometry	73
2.3.3.	Data processing.....	74
3.	Structural insights and fragment screening against the EthR-DNA interaction using native mass spectrometry	76
3.1.	Introduction	76
3.2.	Results and discussion	78
3.2.1.	Structural insights into the EthR-DNA interaction	78
3.2.1.1.	Optimization of MS conditions for detecting EthR-DNA complexes	78
3.2.1.2.	Investigating the stoichiometry of the EthR-DNA interaction	83
3.2.1.3.	EthR dimers bind to separate sites on DNA.....	90
3.2.1.4.	Isothermal titration calorimetry of the EthR-DNA interaction	92
3.2.1.5.	Summary	95
3.2.2.	Fragment screening against the EthR-DNA interaction using native MS	96
3.2.2.1.	Validation of native MS assay to screen EthR-DNA interaction disruptors	96
3.2.2.2.	Fragment screening of EthR-DNA disruptors by native MS.....	98
3.2.2.3.	Validation of hit fragments using X-ray crystallography.....	106
3.2.2.4.	Summary	108
3.3.	Experimental	108
3.3.1.	Materials and methods.....	108
3.3.2.	Protein expression and purification.....	109
3.3.3.	Mass spectrometry	110
3.3.4.	Isothermal titration calorimetry (ITC).....	111
3.3.5.	Differential scanning fluorimetry (DSF)	111

3.3.6.	Surface plasmon resonance (SPR).....	111
3.3.7.	X-ray crystallography	112
4.	Application of native mass spectrometry to the characterization and targeting of the bacterial coenzyme A biosynthetic pathway.....	115
4.1.	Introduction	115
4.2.	Results and discussion	118
4.2.1.	Structural insights into the <i>E. coli</i> CoaB domain	118
4.2.1.1.	Native MS.....	118
4.2.1.2.	Determination of the affinity of the EcCoaB dimerization	122
4.2.1.3.	Influence of solution pH on the EcCoaB protein-protein interaction	125
4.2.1.4.	Ion mobility-mass spectrometry of EcCoaB	127
4.2.1.5.	Collision-induced unfolding of EcCoaB	129
4.2.1.6.	Summary	132
4.2.2.	Assessing EcCoaB-ligand interactions by native MS	132
4.2.2.1.	Interaction of EcCoaB with CTP.....	132
4.2.2.2.	Accounting for non-specific binding	135
4.2.2.3.	Screening of EcCoaB inhibitors by native MS	137
4.2.2.4.	Summary	144
4.3.	Experimental	144
4.3.1.	Materials and methods.....	144
4.3.2.	Protein expression and purification.....	145
4.3.2.1.	EcCoaB overexpression and purification	145
4.3.2.2.	EcCoaBC overexpression and purification	146
4.3.3.	Mass spectrometry	147
4.3.4.	K _d determination.....	147
4.3.5.	Ion mobility calibration.....	149
4.3.6.	Inorganic pyrophosphatase-purine nucleoside phosphorylase PNP-PPIase assay.....	151
5.	Investigation of the Aurora A-TPX2 interaction and its targeting by small molecules using native mass spectrometry	152
5.1.	Introduction	152
5.2.	Results and discussion	155
5.2.1.	Expression and purification of AurA and TPX2	155
5.2.2.	Native MS of the AurA-TPX2 interaction	164

5.2.3.	Screening AurA inhibitors by native MS	171
5.2.4.	Summary	175
5.3.	Experimental	176
5.3.1.	Materials and methods	176
5.3.2.	Protein expression and purification.....	176
5.3.2.1.	Aurora A protein and expression and purification	176
5.3.2.2.	TPX2 protein and expression and purification.....	177
5.3.3.	Mass spectrometry	178
6.	Conclusions and outlook	180
	References	186
	List of PhD publications.....	198

1. Introduction

1.1. The search for new drug discovery approaches

The late 1980's to mid-1990's ushered exciting advances in pharmaceutical research and development (R&D) around the world. Novel methods and techniques in molecular biology furnished high-throughput screening (HTS) platforms that were able to screen compounds at rates that were previously unheard of.^[1-4] At the same time, new strategies in combinatorial chemistry promised to generate exponentially-higher numbers of compounds compared to what could be made by a typical bench chemist in a given period of time.^[5-7] Unfortunately, after disappointing results with the new methodology, people came to the realization that the chemical universe was too vast to be empirically explored systematically. Moreover, early combinatorial libraries often sacrificed diversity in order to accommodate facile reaction and purification methodologies, resulting in the generation of libraries of compounds that lacked structural complexity.

Today, HTS is now seldom conducted on purely combinatorial, unfocused compound databases. Instead, both industry and academic researchers have devised a number of new strategies over the past few decades to create "privileged" chemical libraries that aim to sample a more pharmacologically applicable space within the colossal universe of feasible molecular structures. For example, the strategy of drug repurposing or drug repositioning seeks to accelerate drug discovery by restricting the screening set to databases of approved drugs.^[8-10] In diversity-oriented synthesis (DIOS), a collection of diverse central scaffolds are decorated with peripheral functional groups, producing small libraries of molecules for each central scaffold.^[11-13] On the other hand, both biology-orientated synthesis (BIOS)^[14-15] and the reemerging field of natural product-based drug discovery^[16-17] seek to harness knowledge from Mother Nature to generate compounds with an increased propensity to show desirable biological properties.

1.2. Fragment-based drug discovery (FBDD)

Fragment-based drug discovery (FBDD) is another strategy that aims to increase the efficiency of early-stage drug discovery. FBDD broadly describes the identification of small molecules (“fragments”) possessing affinity typically in the 0.1 to 10 mM range and their subsequent elaboration into potent ligands.^[18] The central thesis of fragment-based methods is that while smaller fragment molecules bind only weakly to the target site, they reveal high-quality interactions that can be exploited by assembly of the individual fragments into larger molecules.^[19] Ever since Fesik and co-workers first demonstrated the application of FBDD against the FK506-binding protein over 20 years ago,^[20] the technique has been successfully applied in both academic and industrial medicinal chemistry programmes against a variety of diverse targets.^[21] The utility of FBDD has been further exemplified by the approval of vemurafenib (Zelboraf)^[22] targeting the oncogenic kinase BRAF for the treatment of metastatic melanoma, and venetoclax (Venclexta) targeting the anti-apoptotic protein Bcl-2 for the treatment of chronic lymphocytic leukemia (CLL).^[23] Very recently, ribociclib (Kisqali),^[23] an orally bioavailable, highly selective inhibitor of CDK4/6 developed by an Astex/Novartis collaboration, received approval from the United States Food and Drug Administration and the European Commission for the treatment of hormone receptor-positive, human epidermal growth factor receptor 2-negative (HR+/HER2-) advanced or metastatic breast cancer.^[24]

At its core, the fragment-based approach aims to tackle the challenge of the immense chemical space of potential drug molecules that has stymied classical combinatorial approaches. While Lipinski’s “rule-of-five” (molecular weight < 500 Da, calculated logP < 5, hydrogen bond donors < 5 and hydrogen bond acceptors < 10)^[25] aimed to narrow the universe of small molecules to those possessing drug-like characteristics, there still remains an estimated 10^{60-200} drug-like molecules of HTS size (300–500 Da).^[26] In contrast, fragments typically obey a “rule-of-three”, with molecular weight < 300 Da, calculated logP of < 3, hydrogen bond donors < 3 and hydrogen bond acceptors < 3 (**Figure 1.1**).^[27] As there are only about 10^7 possible molecules composed of up to 11 atoms of C, N, O, and F that follow the rule-of-three,^[28] a significantly larger percentage of the

chemical space can be sampled with a smaller fragment library (typically *ca.* 10^3 fragments) than with a much larger library of *ca.* 10^5 – 10^6 larger molecules in an HTS campaign.

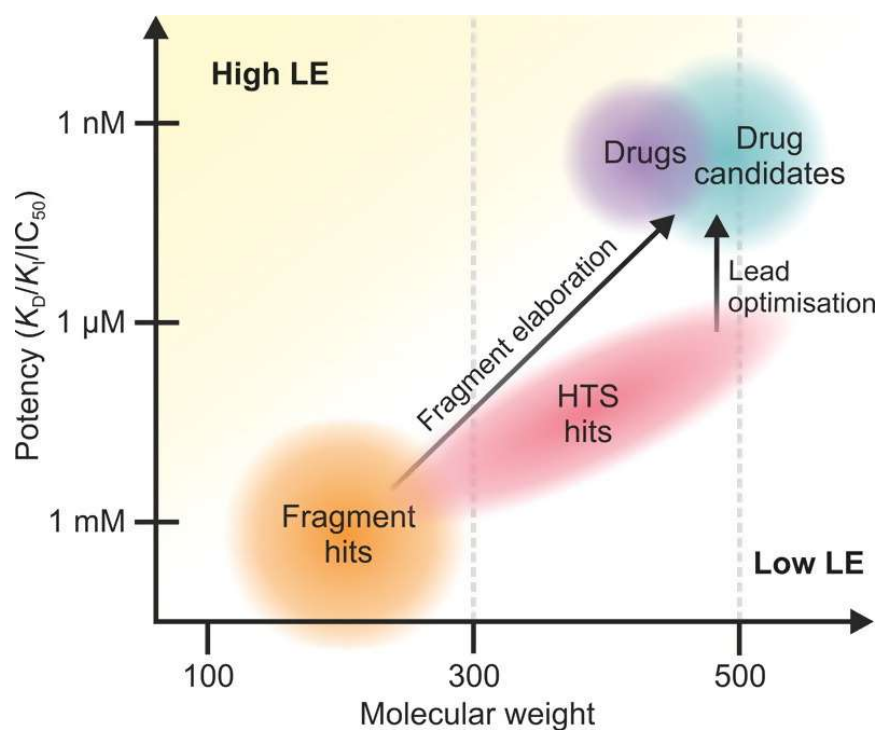


Figure 1.1 Comparison of potency of leads vs. molecular mass developed from conventional high-throughput screening and fragment-based drug discovery approaches. Reproduced with permission from Scott et al.^[18]

A second fundamental advantage of fragment-based screening is that since the fragment hits are individually only weakly binding, they must form very high-quality interactions with the target to be detected. Thermodynamically, fragments identified as weakly-binding hits must possess high intrinsic binding enthalpy in order to overcome the entropic penalty of binding. Thus, fragments are very “atom-efficient” ligands with high ligand efficiency (LE) values, which is calculated as the binding energy per mole divided by the number of heavy atoms. Importantly, the minimal structural complexity of fragments allows them to effectively probe binding sites and to establish geometrically unencumbered, high-quality, and enthalpically-biased interactions, making them ideal starting points for further chemical elaboration.^[18]

1.3. A biophysical cascade for FBDD

A number of biophysical tools are available for the identification and characterization of weakly-binding fragments. Due to the low binding affinities generally seen with fragments, the screening stage of FBDD requires the use of biophysical techniques significantly more sensitive than the methods utilized in HTS.^[29] Differential scanning fluorimetry (DSF), also known as the thermal shift assay (TSA), is a low-cost and high-throughput technique that can rapidly screen fragments for validation. The principle of DSF is based upon measuring the changing fluorescence of an exogenous dye as the protein to which it is bound unfolds (or alternatively, the change of fluorescence has been attributed to the onset of aggregation).^[30] Ligand binding usually stabilizes a protein, thereby increasing the melting temperature (T_m) of the protein and producing in a positive thermal shift of typically ± 0.5 to 2 °C for a weakly binding fragment. However, one drawback of DSF is that it is susceptible to false negatives. Many strong binders have been observed to cause insignificant thermal shifts, or even negative thermal shifts.

One-dimensional (1D) ligand-observed ^1H nuclear magnetic resonance (NMR) spectroscopy methods, such as saturation transfer difference (STD), water-ligand observed via gradient spectroscopy (WaterLOGSY) and Carr-Purcell-Meiboom-Gill (CPMG) techniques, are also commonly used for fragment screening.^[31] When performed in the presence of known ligands, ligand-observed NMR experiments can also provide preliminary binding site information as well as relative affinity data. Moreover, fragments can be screened in cocktails to expedite the validation process, followed by singleton screening to confirm hits.

More detailed structural information can be obtained using protein-observed two-dimensional (2D) NMR experiments, which analyze chemical shift perturbations (CSPs) of protein amide signals induced by the ligand.^[32] Another method that is commonly utilized for the screening of fragment libraries is surface plasmon resonance (SPR).^[33] The change of SPR signal when a solution of ligand is passed over proteins immobilized on the biosensor surface allows the determination of both kinetic and thermodynamic parameters of binding.

The fragment binding characterization stage involves determining the energetic and structural parameters of ligand binding. Isothermal titration calorimetry (ITC) can be used to determine the enthalpy (ΔH), entropy (ΔS), and Gibb's free energy (ΔG) changes associated with binding, as well as the equilibrium dissociation constant (K_D) and the binding stoichiometry (n).^[34] Meanwhile, X-ray crystallography provides detailed structural information on the fragment-target interaction, by confirming the binding location of the fragment and determining the orientation of the fragment within the binding site of the protein.^[35-36] Although ITC and X-ray crystallography are traditionally considered to be relatively low-throughput methods, it is also possible to use them as preliminary screening techniques provided that sufficient automation is available.^[34-35] In particular, high-throughput fragment screening using X-ray crystallography has been performed since 1999 by Astex, who pioneered this approach by integrating computational tools such as AutoSolve to efficiently interpret electron density maps for fragment binding, with robotic systems that are able to store and automatically manipulate protein crystals.^[37] More recently (since 2015), the XChem beamline at the Diamond Light Source has also been available for fragment screening via X-ray crystallography. A comparison of commonly used fragment screening techniques is presented in **Table 1.1**.

Table 1.1. A comparison of commonly used fragment screening techniques.

Technique	Principle	Advantages	Limitations
DSF	Fragment increases stability of protein against unfolding	High-throughput, low-cost, easy to perform, low protein consumption	Low information content, susceptibility to false positives and false negatives
Ligand-observed NMR	Change in T_2 relaxation rate (CPMG) or NOE (STD, waterLOGSY) upon ligand binding	Medium-throughput, protein can be unlabeled, allows verification of protein and ligand purity	Prone to false positives from aggregation, limited to fragments that transiently bind in the fast exchange regime, high protein consumption

Protein-observed NMR	Ligands induce CSPs of amide signals of protein	Determination of ligand binding site and K_D	Low-throughput, requires large amounts of labeled protein, limitations on protein size
SPR	Change in refractive index at interface of an optical biosensor induced by ligand binding	Medium-throughput, provides both kinetic and affinity data, low protein consumption	Requires that protein be stable to immobilization on surface
ITC	Detection of heat change upon ligand titration	Provides thermodynamic parameters of binding, binding affinity and stoichiometry	Low-throughput, high protein consumption
X-ray crystallography	Diffraction of X-rays by crystals of protein-ligand complexes	High-resolution structural detail of ligand binding mode	Protein must form well-diffracting crystals, high ligand concentration required, reliant on significant technical instrumentation and expertise

In a typical fragment screening “cascade”, the initial primary screening is performed using a high-throughput and low-cost technique, such as DSF, to rapidly generate a shortlist of hits for further testing. The hits are then validated using a secondary screening technique, such as 1D NMR, which can additionally provide information on relative binding affinity and binding site information if conducted in competition mode. The most promising hits from the validation stage are then finally characterized using a combination of ITC and X-ray crystallography. An important benefit of this three-stage biophysical cascade is that the binding of the fragments is validated with a combination of orthogonal biophysical techniques, which enables confidence in using the

fragment hit for further lead optimization.^[29] The validated fragment hits can then be synthetically elaborated into lead compounds by iterative cycles of synthesis, guided by a synergy of structural information, bioaffinity measurement, and rational chemical design (**Figure 1.2**).^[18]

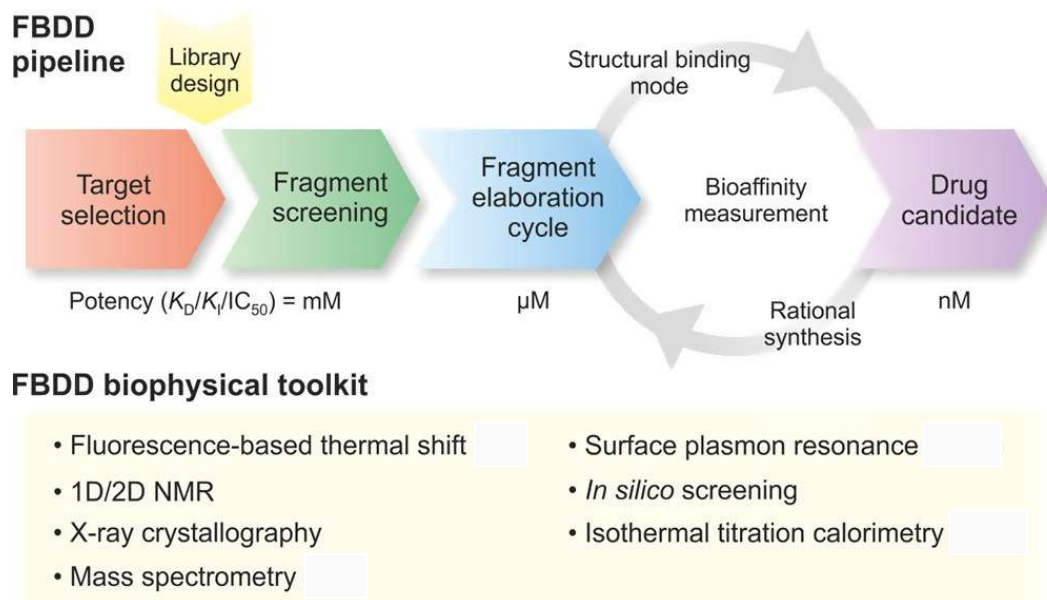


Figure 1.2 Schematic representation of the general workflow of FBDD. Reproduced with permission from Scott et al.^[18]

1.4. Mass spectrometry (MS) methods for drug discovery

While mass spectrometry (MS) is a well-established technique for characterizing biomolecular structure and interactions,^[38-41] it has been less widely applied for drug discovery.^[42-43] Only recently have MS approaches been utilized for FBDD.^[44-46] MS-based methods have the important advantages of high sensitivity, low sample consumption and being label-free. The following sections provide an introduction to how MS-based methods can be used to elucidate protein-ligand interactions, with a view towards applications in drug discovery.

1.4.1. Mechanism of electrospray ionization (ESI)

The mechanism of electrospray ionization (ESI) is well-known.^[40] Briefly, a potential difference is applied between the aperture of the mass spectrometer and a conducting capillary containing

the analyte solution, which causes the analyte solution to be dispersed into fine, charged droplets at the tip of the capillary. The droplets evaporate and shrink in size as they pass into the vacuum of the mass spectrometer. Droplet fission occurs when the droplet size reaches the Rayleigh limit, which is the point where the surface tension holding the droplets together is equal to the electrostatic repulsion between the charges on the surface. Successive rounds of evaporation and fission occur until the analyte ions are set free in the gaseous state.

This technique was further refined by the introduction of a miniaturized version of the ESI process by Wilm and Mann in 1994.^[47] Termed nanoflow-ESI (nanoESI), the use of much narrower spray apertures enhances the dispersion of the liquid into nanodroplets with a favorable surface-to-volume ratio even under very mild combinations of applied voltage, pressure and temperature. Importantly, the lower flow rates afforded by nanoESI greatly reduces the consumption of protein material (as little as 1 μL of a 1 μM protein solution can suffice for a single experiment), which is a very important consideration when protein supplies are scarce and/or expensive. A schematic overview of the nanoESI process is depicted in **Figure 1.3**.

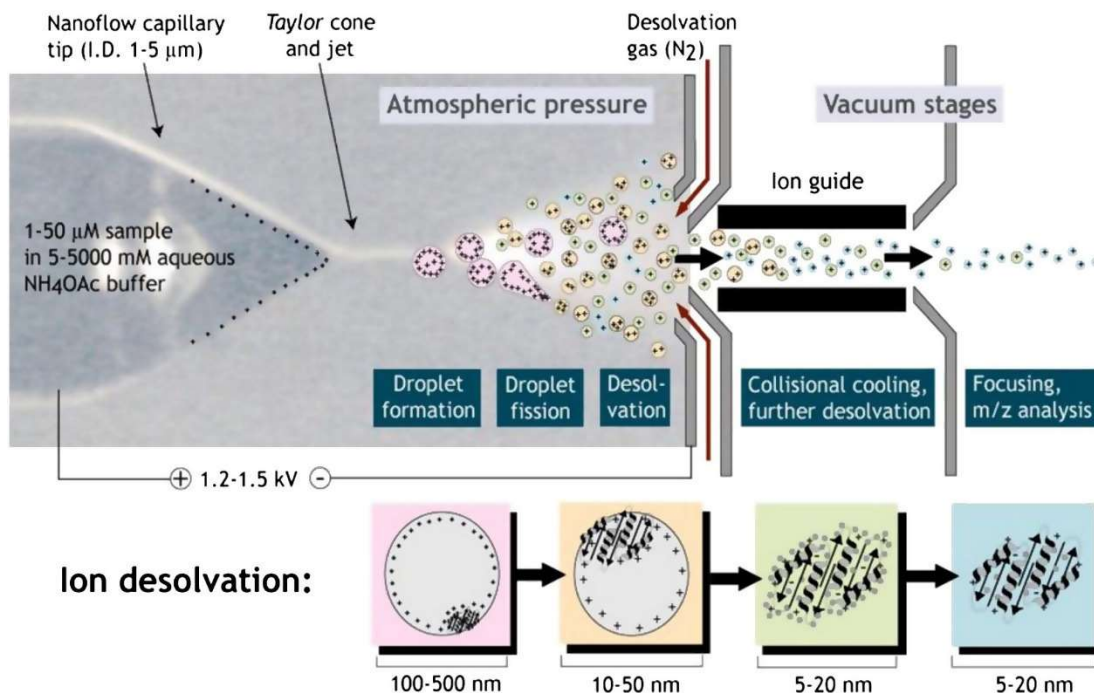


Figure 1.3 A schematic representation of the nanoESI process. Reproduced with permission from Konijnenberg et al.^[41]

The precise mechanisms by which gaseous ions are formed from charged droplets are a source of active investigation. The three main mechanisms that have been proposed for explaining ion formation in ESI are the ion evaporation model (IEM), the charge residue model (CRM) and the chain ejection model (CEM) (**Figure 1.4**).^[48] In the IEM, which is considered to be the primary mechanism by which low molecular-weight ions are generated, a charged analyte ion “evaporates” from the surface of a droplet as a result of internal repulsion (**Figure 1.4a**). A transient solvent bridge is thought to initially tether the departing ion to the parent droplet, until it eventually ruptures to release a small gas-phase cluster containing the analyte ion and a few solvent molecules. The solvent shell is subsequently lost as the cluster travels through the various stages of the mass spectrometer, allowing the free analyte ion to be detected. Globular protein ions are widely accepted to be generated via the CRM (**Figure 1.4b**). In this model, repeated fission of larger droplets occurs until the final droplets contain a single analyte ion each. As the final solvent shell disappears, the remaining charge of the droplet is transferred entirely to the analyte ion. It is important to note that in the initial stages, larger droplets can shed excess charge via the ejection of protons or other cations, which accounts for why proteins of different shapes and/or sizes can end up with different charge state distributions. Finally, unfolded or disordered proteins are thought to yield gaseous ions through the CEM (**Figure 1.4c**). The exposure of hydrophobic residues in unfolded or disordered proteins means that it is unfavorable for them to remain within the aqueous interior of the droplet. Thus, unfolded chains migrate to the droplet surface and become expelled into the gaseous environment. Stepwise ejection of the remaining protein eventually leads to complete release of the unfolded protein ion from the droplet.

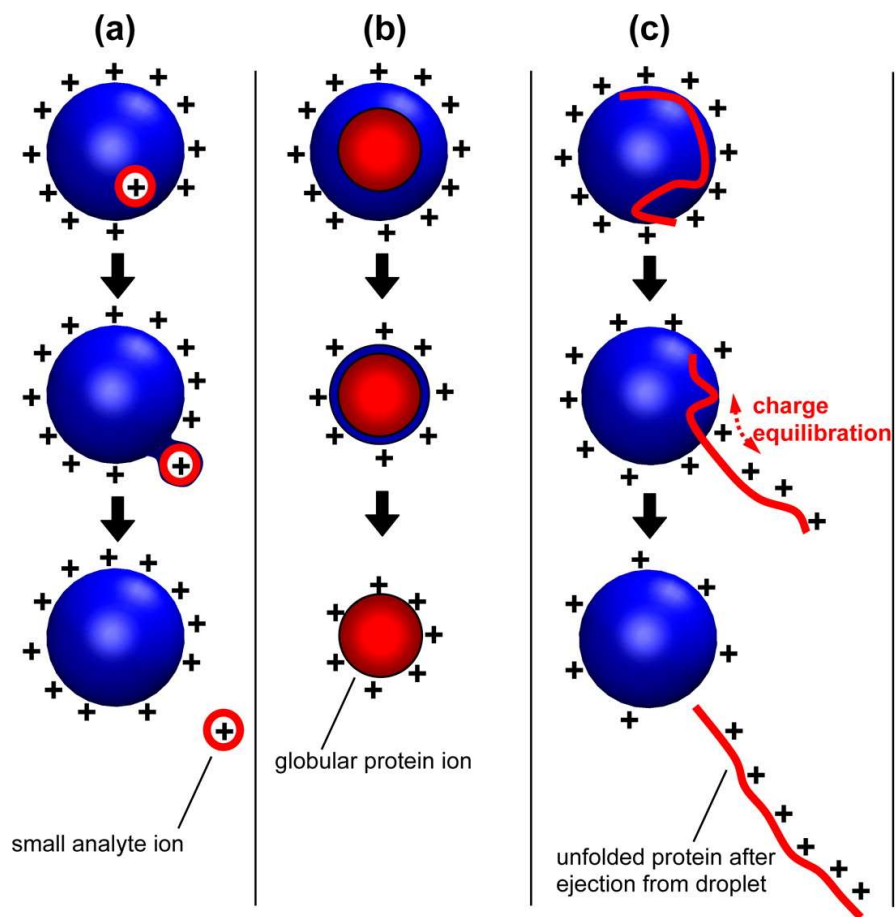


Figure 1.4 Schematic representation of (a) ion evaporation model, (b) charge residue model and (c) chain ejection model for electrospray ionization. Reproduced with permission from Konermann et al.^[48]

1.4.2. Native MS

Native MS, also known as native state MS or non-denaturing MS, aims to observe biomolecules and their complexes in as natural state as possible (*i.e.*, with protein structures and interactions preserved) given the constraints of mass spectrometric detection in the gas phase.^[55] Due to its advantages of low sample consumption and tolerance to salts and buffers, ESI, or more specifically nanoESI, is by far the most commonly used ionization technique for native MS. As ESI-MS is a relatively soft ionization technique, the native conformation of biomolecules as well as any biomolecular interactions can be maintained even as analytes are transferred into the gas phase.^[41]

Given that the ESI process produces analyte ions in the gas phase, how relevant are detected gas-phase interactions to drug discovery? Enthalpic interactions, which are driven primarily by electrostatic or polar interactions, survive or are even strengthened upon transfer from solution to the gas phase.^[56-57] In contrast, hydrophobic effect interactions are weakened and may not survive transfer into the gas phase. This implies that native MS has an inherent bias towards identifying ligands that bind predominantly *via* electrostatic interactions rather than through hydrophobic interactions, which is an advantage for drug discovery. Enthalpically-driven leads are associated with superior clinical outcomes, whereas leads optimized through entropic or hydrophobic interactions often have poorer prospects and suffer an increased probability of attrition.^[58]

Despite the fact that interactions that are important for protein folding in solution (such as hydrogen bonds with water and the hydrophobic effect) are lost in the gas phase, protein ions are able to retain their solution-like protein conformations during ESI because of kinetic trapping.^[59] In other words, the detected protein conformation might not necessarily be the most thermodynamically stable conformation of the gaseous protein ion. However, these other conformations are inaccessible by the protein ion in the gas phase (where water and other protein refolding catalysts are absent), as the solution-like conformation that the protein ion initially adopts after electrospray is located in a local energy minimum that cannot be escaped from on the time scale of typical ESI-MS experiments.

Additionally, and in contrast to structural techniques such as X-ray crystallography and NMR spectroscopy, ESI-MS also offers the capability to investigate the interactions that are inherent to a particular protein-ligand or protein-protein system in solvent-free or independent manner.^[42] This is because the gas-phase complex is detected in an isolated environment that is free from the effects of buffers, solvent or other known or unknown chemical additives in the analyte solution. Coupled with the selectivity and specificity of mass spectrometry analysis, ESI-MS provides a reproducible environment for reproducible studies on protein structure and interactions. However, one limitation is that there are limits to how well an isolated, gas-phase

system can mimic the real cellular environment where proteins constantly interacting with the surrounding water molecules and other chemical entities in an aqueous environment.

Native MS has been widely applied to study the stoichiometry and interactions of macromolecular assemblies.^[60-61] A natural extension of the concept is to use native MS to screen for protein-ligand interactions. Hits are readily detected with native MS because the presence of a bound fragment will cause the mass/charge (m/z) signal of the protein-ligand complex to be higher compared to the free protein. As an example, **Figure 1.5a** shows the native mass spectrum of carbonic anhydrase recorded in negative ion mode.^[44] The narrow charge state envelope together with the low absolute charge of the ions (from 8^- to 10^-) indicates that the protein remained compacted in the gas phase. In the presence of a cocktail of fragments, shifts to higher m/z values are observed due to the formation of protein-ligand complexes (**Figure 1.5b**). These signals are readily deconvoluted because multiplying the m/z shift ($\Delta m/z$) and the charge state directly gives the molecular weight of the bound ligand. Additionally, evaluating the proportion of bound and unbound protein species allows determination of K_D values.^[62] It is important to note that in native MS, the voltages at various stages within the mass spectrometer should be minimized as far as possible in order to preserve non-covalent interactions, while still being high enough to maintain sufficient desolvation and ion transmission. This also has consequences for the resolution of the technique, as the ability to increase m/z resolution (narrower spectral peaks) by raising the accelerating voltages must be balanced against the possibility of disrupting non-covalent interactions, including both those that are responsible for maintaining the tertiary structure of proteins, as well as any quaternary protein-protein interactions or protein-ligand interactions. On the other hand, insufficient voltages lead to broader peak shapes due to adduct formation with water molecules and salt ions, which hampers the detection of protein-ligand binding particularly for small fragments. Besides voltage, gas pressures can also be used to control the degree of desolvation and transmission of ions.

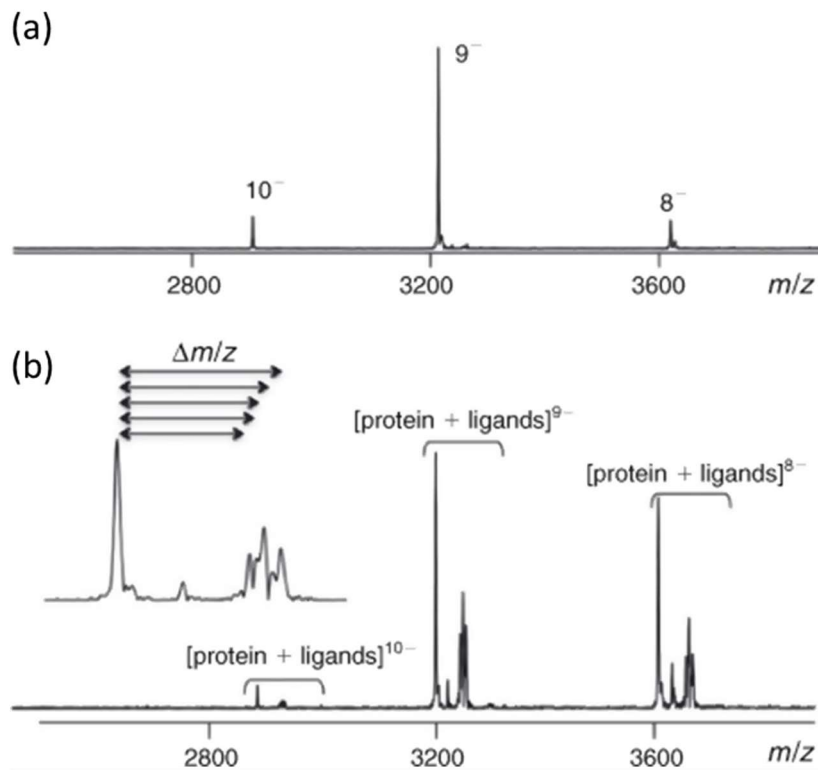


Figure 1.5 Native mass spectra of carbonic anhydrase in the (a) absence and (b) presence of a fragment cocktail. The formation of protein-ligand complexes causes a shift ($\Delta m/z$) in the protein signals, with the magnitude of the m/z shift being equal to the m/z of the bound fragment. Reproduced with permission from Poulsen.^[44]

A number of studies applying native MS to fragment screening have been described in recent years. In 2012, Maple *et al.* provided one of the first demonstrations of automated fragment screening using native MS.^[63] To increase throughput, experiments were performed with the Triversa NanoMate (Advion),^[64-65] a microfluidic chip-based sample injection and liquid handling robot. The experimental setup was first validated using lysozyme as a model enzyme, which revealed good agreement between K_D values for acetylglucosamine oligomers determined by native MS and those determined by ITC. Subsequently, the automated native MS screening platform was applied to screen a focused phenylpyrazole fragment library against Bcl-x_L, an anti-apoptotic protein that has been implicated in tumorigenesis.^[66] Of the fragments screened, 84

out of 157 were revealed as hits ($K_D < 300 \mu\text{M}$), and a subset of these were further subjected to validation using other biophysical techniques, including STD NMR, 2D NMR and ITC.

Woods *et al.* recently described a fragment screening campaign against human carbonic anhydrase using native MS in conjunction with SPR and X-ray crystallography.^[67] In the first stage of this study, a library of 720 fragments were screened against carbonic anhydrase by SPR, and the 7 hits that emerged were validated by native MS and X-ray crystallography. Subsequently, a focused library of 70 compounds selected based on structural similarity to the fragment hits from the first stage, were independently screened using both native MS and SPR. The second stage revealed 37 hit compounds, 24 of which were detected as positive by both native MS and SPR, and there was agreement between the two techniques for 58 out of the 70 compounds in the library.

In a recent study, Göth *et al.* applied native MS to four protein targets, 7,8-dihydro-8-oxoguanine triphosphatase (MTH1), lysine-specific demethylase 5B (KDM5B), bromodomain and PHD finger-containing protein 1 (BRPF1), and E3 ubiquitin-protein ligase (UHRF1), and compared these results with those obtained with DSF.^[68] For MTH1 and KDM5B, there was good overlap between native MS and DSF, with 16 out of 24 stabilizing fragments in DSF also showing as MS hits for MTH1, and 8 out of 9 for KDM5B. However, for BRPF1 and UHRF1 the agreement between the two techniques was poor. For BRPF1, none of the 21 tested fragments qualified as a native MS hit despite 11 of these showing thermal stabilization by DSF, while for UHRF1, the only native MS hit showed as a non-binder by DSF. Additionally, the researchers observed that the complicating issues of charge state distribution shifts and non-specific binding were largely protein-dependent, highlighting the need to carefully assess the suitability of a protein target (*e.g.* in terms of size or ionization efficiency) before embarking on a fragment-based screening campaign using native MS.

Fragments can also be screened as cocktails in native MS. Vu *et al.*^[69] screened 331 natural product-like fragments against *Plasmodium falciparum* deoxyuridine 5'-triphosphate nucleotidohydrolase (*Pfd*UTPase), a potential antimalarial target.^[70] Fragments were assessed in

cocktails of eight compounds each, revealing securinine and 4- α -hydroxy-allosecurinine as hits. The two hits together with five other securinine compounds were subsequently evaluated for their ability to inhibit *Pfd*UTPase activity as well as block *P. falciparum* replication. There was good correlation between binding affinities determined by native MS and inhibition of *P. falciparum* growth, although surprisingly, the molecules promoted rather than suppressed *Pfd*UTPase activity. In a later study,^[71] the same group of researchers applied native MS to screen 659 natural product-like fragments against *P. falciparum* Rab11a, a Rab GTPase that regulates parasite replication.^[72] Using a similar cocktail method as before, 11 fragments emerged as hits against *Pf*Rab11a.

Even when native MS is not employed as the primary screen, the technique can still be used to aid FBDD campaigns. Our group has recently developed low nanomolar inhibitors of *Mycobacterium tuberculosis* (*Mtb*) CYP121, a cytochrome P450 enzyme essential for *Mtb* viability,^[73] via a fragment-merging strategy. Previous work had shown that azole antifungal compounds bind tightly to CYP121, while also inducing significant dissociation of dimeric CYP121 into its monomeric form, whereas other small molecule inhibitors and the natural substrate (cYY) did not affect the oligomeric status of the enzyme.^[74] In our study, we tested the binding affinity of several lead compounds (250–373 Da) derived from an initial triazol-1-yl phenol fragment (161 Da) against CYP121 by native MS.^[75] The lead compounds bound up to twice per CYP121 dimer, while having no effect on the dimer-monomer equilibrium, which was attributed to differences in the mode of heme binding between the tested compounds and previous antifungal azoles. Additionally, the double-occupancy observed for one of the compounds in X-ray crystallography was not detected by native MS, which could indicate that one of the two binding sites in the X-ray crystal structure of CYP121 is a low-affinity site. Finally, the rank order of binding affinities determined by native MS correlated well with affinity data recorded with ITC and UV-Vis optical titration. This study highlighted the utility of native MS in confirming ligand binding stoichiometry as well as to test for potential changes in protein quaternary structure.

1.4.3. Ion mobility-mass spectrometry (IM-MS)

The separation of ions by ion mobility (IM) is based on the principle that the movement of ions through an electric field containing an inert background gas is dependent on the charge, size and shape of the ion, as well as other parameters of the drift chamber that are usually considered constant for a given set of experimental conditions (see Mason-Schamp equation below). Larger ions collide more frequently with the inert gas, and thus move through the ion mobility chamber more slowly. Additionally, ions bearing a larger charge will experience a greater electric force and move more quickly through the chamber. By measuring the drift time (t_D) of an ion of charge z , it is possible to determine the collision cross-section (CCS) (Ω) of a protein ion by use of the Mason-Schamp equation:^[76]

$$\Omega = \frac{(18\pi)^{1/2}}{16} \frac{ze}{(k_b T)^{1/2}} \left[\frac{1}{m_i} + \frac{1}{m_N} \right]^{1/2} \frac{t_D E 760}{L P} \frac{T}{273.2 N}$$

where k_b is the Boltzmann constant, T is the temperature, e is the elementary charge, m_i is the mass of the ion, m_N is the mass of the inert gas, E is the electric field strength, L is the length of the drift region, P is the pressure in mmHg, and N is the number density of the neutral gas at 760 mmHg. Note that in the IM measurement of a protein ion, the mass of the protein has little effect on the CCS because the reduced mass of the protein and neutral gas molecule is calculated.

Coupling IM with MS therefore allows for the three-dimensional evaluation of protein ions in the gas phase, with the first dimension being the mass-to-charge ratio of the protein or protein complex, the second being the CCS of the ion, and the third being the relative abundance for all the ions observed.

The use of IM-MS to study proteins or protein complexes is a relatively new field. The first commercially available mass spectrometer capable of performing IM-MS measurements, the SYNAPT HDMS® (Waters), was introduced in 2006.^[77] The traveling wave (TW)-type ion mobility separator used in this instrument, which employs time-varying electric fields to propel the ions through the drift chamber (**Figure 1.6**), was developed only two years prior in 2004.^[78] Compared

to the older drift tube (DT)-type ion mobility separator, TW separation improves the transmission efficiency of ions through the drift cell, however, direct measurement of CCS is not possible. Instead, the drift times measured for the protein ion of interest are calibrated against those measured for ions of known CCS values under the same conditions.^[42] A calibration dataset for a range of denatured peptides and proteins, and native proteins and protein complexes was presented by the Robinson and Ruotolo groups in 2010.^[79]

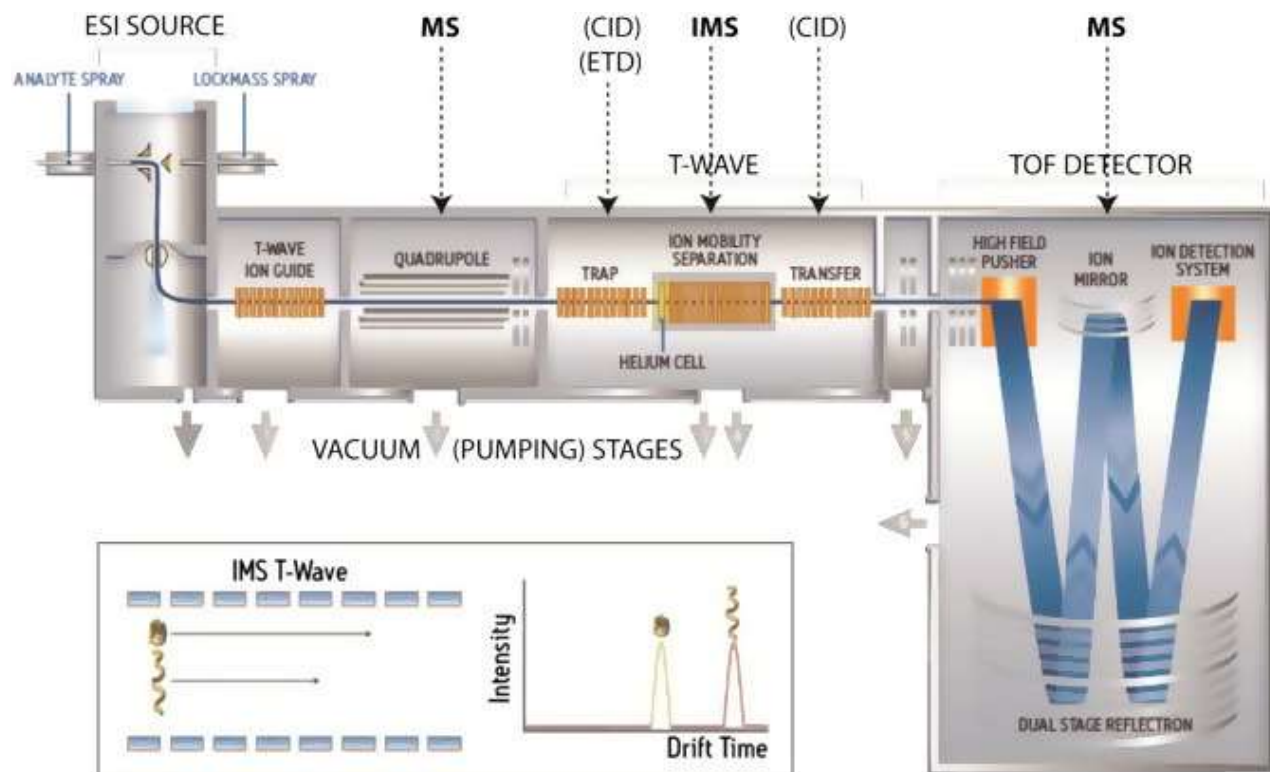


Figure 1.6 A schematic representation of the SYNAPT HDMS mass spectrometer. IM-MS is performed in the ion mobility separation cell. Reproduced with permission from Konijnenberg et al.^[41]

IM-MS can be used to study the conformational changes of proteins or protein complexes upon ligand binding. For example, Hopper and Oldham have used IM-MS to study the effect of ligand binding on the conformational stability of proteins such as FK-binding protein, lysozyme and myoglobin.^[51] Barran and co-workers have shown using IM-MS that the metal-bound form of a zinc finger peptide motif is more stable with respect to unfolding compared to the apo-form.^[80] More recently, Cianférani, Giglione and co-workers have used IM-MS to investigate very small

conformational changes in peptide deformylase (PDF1B), a potential antimicrobial target, induced by different ligands.^[81] Due to the low-throughput nature of this technique, IM-MS is best suited for the characterization of hit ligands rather than as a primary screening technique. Apart from ligand screening however, IM-MS has found much greater use for the characterization of protein conformation and stability, as described in the next section.

1.4.4. Collision-induced dissociation (CID) and collision-induced unfolding (CIU)

The internal energy of gaseous protein ions can be increased by raising the acceleration voltages at different stages of the mass spectrometer, as a result of more energetic collisions with neutral gas molecules. Activated multiprotein complexes can undergo collision-induced dissociation (CID), which commonly takes place through the asymmetric charge partitioning mechanism.^[49-50] In this mechanism, an unfolded monomeric subunit is ejected which carries away a disproportionately large amount of charge for its size, leaving behind a “stripped” oligomer of relatively lower charge. The exact process of dissociation is thought to depend on various factors including the internal energy, charge state, and conformation of the complex, as well as the conformational flexibility of the monomers in the complex.^[49] In one model, the asymmetry is induced by the unfolding of one of the monomers in the complex, presumably the subunit with the higher conformational flexibility, upon activation. The activated subunit overcomes the dissociation barrier so that further unfolding of the protein is promoted by the repulsive Coulomb potential.^[50] As the departing subunit increases in surface-to-mass ratio during the separation process, protons are driven to the unfolding subunit to reduce overall Coulomb energy, which accounts for why the ejected monomer is able to carry away a disproportionate amount of charge.

Moreover, activated protein ions can undergo collision-induced unfolding (CIU).^[51-53] CIU can be considered to be the gas-phase equivalent of differential scanning calorimetry experiments that are performed in solution. The unfolding of protein ions can be tracked by monitoring their sizes as a function of collision energy using IM-MS. The collisional activation of proteins or protein complexes often yields a collection of partially folded intermediates that are stable on the

millisecond time scale, generating a “CIU fingerprint” that is unique for a protein ion under a given set of conditions.^[54] Although the data collected by CIU are only indicative of the relative and not absolute thermodynamic properties of a protein, they can still provide useful insight into the structure, stability and conformations of proteins and their complexes.

1.4.5. Hydrogen-deuterium exchange (HDX)-MS

The amide hydrogens of a peptide or protein are labile and exchange with solvent hydrogen or deuterium through acid-, base- or water-catalyzed reactions. The rate of exchange is highly dependent on the structural environment of the amide as well as the solution conditions. For example, an amide hydrogen in a random coil conformation will exchange on the order of milliseconds to seconds at pH 7.0 and room temperature, but may exhibit a reduction in exchange rate by a factor of as high as 10^8 if found within the interior of a folded protein.^[82] Thus, amide hydrogen exchange rates give an indication of the local structure and solvent accessibility of different regions of a protein, allowing changes in the structure and dynamics of a protein to be monitored indirectly.

A schematic overview of the HDX-MS process as applied to ligand screening is depicted in **Figure 1.7**.^[83] In the first stage, “on-exchange” of hydrogen for deuterium takes place when the protein is incubated in D_2O in either the presence or absence of the ligand. Aliquots of this solution are withdrawn at pre-specified time intervals and then quenched by the addition of cold and acidic solution. This quench step is essential to preserve the level of deuterium incorporation until MS analysis is performed. The partially-deuterated protein is then digested, typically by passing through a pepsin column, to generate a collection of small peptides, which are then separated chromatographically before MS analysis. The percentage incorporation of deuterium for each peptide at each time point can be determined, and a graph of % deuterium uptake against time for the given peptide can then be plotted. Further analysis of overlapping proteolytic fragments can generate more localized HDX data, sometimes down to single-residue resolution (although this is complicated by H/D scrambling during gas-phase fragmentation of peptides).

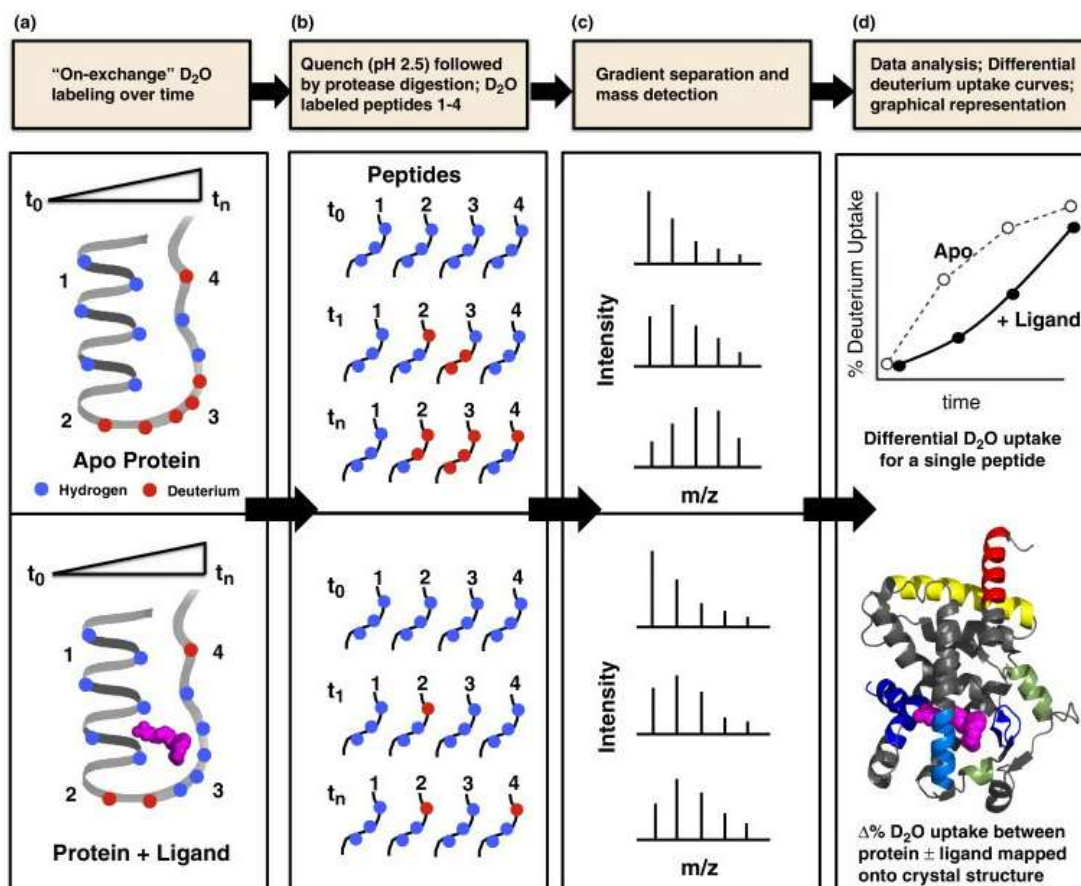


Figure 1.7. A schematic representation of the HDX-MS process as applied to ligand screening. a) Proteins are labelled with D_2O either in the absence (upper panel) or presence (lower panel) of ligand. (b) Reaction aliquots are quenched at pre-specified time intervals, followed by protease digestion. (c) Peptides are separated chromatographically and analyzed by MS. (d) The rate of deuterium uptake for the apo and ligand-bound forms are compared for each peptide and differences are mapped onto the crystal structure. Reproduced with permission from Marciano et al.^[83]

Comparing the differential deuterium uptake between the apo and ligand-bound forms of a protein therefore gives information on the effects of ligand binding on the structure of the protein. At the simplest level, the binding of a ligand to a protein should shield the amide protons of the ligand-binding site from the deuterated solvent, thereby reducing the % deuterium uptake of the binding site residues to them. This can be used to generate an ‘HDX fingerprint’ for each ligand, and analysis of these fingerprints can allow for the clustering of ligands into groups that

share a comparable binding mode.^[84] However, differences in HDX patterns upon ligand binding can also result from changes to the protein structure or dynamics in regions that are distant from the binding site (*i.e.*, allosteric effects).

While HDX-MS has been widely employed for studying the structure and dynamics of biopharmaceuticals,^[85-87] its use for small-molecule ligand screening has received comparatively less attention. The primary drawback of HDX-MS is that it is a low-throughput technique, and the screening of even as few as one thousand compounds will incur significant expense in both time and money. In a recent study, Carson and co-workers used HDX-MS as a secondary screen to validate the binding of fragment molecules to vitamin D receptor (VDR), a regulator of calcium and phosphorus homeostasis.^[88] Interestingly, significant protection of H/D exchange was observed for some fragments in helices 3, 7, and 8 of the ligand-binding domain of VDR, regions which were similar to those stabilized by the natural hormone vitamin D3. This study, which was the first and to-date only report on the application of HDX-MS to FBDD, showed how this technique can be used to identify suitable fragments as starting points for further synthetic elaboration.

1.4.6. Ligand-based MS techniques

Ligand-based MS approaches have also been employed as a fragment screening technique for drug discovery. The general workflow of a ligand-observed MS experiment is shown in **Figure 1.8**.^[89] In the first step, the protein target of interest is incubated with a cocktail of fragments, while in a separate control experiment, the same fragment cocktail is incubated in the absence of protein. Next, the target and control samples are subjected to ultrafiltration using a centrifugal concentrator. This retains the protein-ligand complexes in the supernatant, while the unbound ligands are driven into the filtrate. Note that this step presumes that the protein-ligand complexes are stable to centrifugation. The protein-ligand complexes are subsequently dissociated by the addition of 90% aqueous methanol, and the released fragments are analyzed by liquid chromatography-mass spectrometry (LC-MS). Fragments that are enriched in the target sample compared to the control sample are deemed to be binders, as these compounds were

retained in the supernatant presumably through binding to the target protein. Measurement of the bound fragment concentration moreover enables the calculation of binding dissociation constants (K_D) for the target-ligand interaction. It is also possible to perform LC-MS analysis on the unbound ligand fraction, which is obtained as the filtrate after the ultrafiltration step. In this case, fragments that are reduced in intensity in the unbound fraction are deemed to be binders.

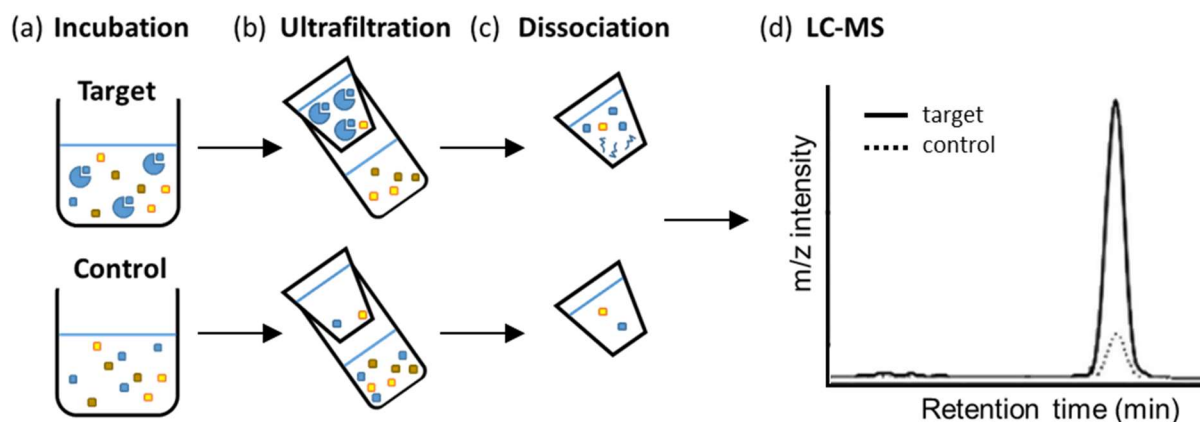


Figure 1.8 General workflow of a ligand-observed MS experiment for fragment screening. (a) The target protein is incubated with a mixture of fragments. The negative control experiment is performed without protein. (b) Ultrafiltration separates the free ligand from the protein-ligand complexes, which are retained in the supernatant. (c) The protein-ligand complexes are dissociated using 90% aqueous methanol to liberate the ligands. (d) The released ligands are quantified by LC-MS. Fragments that are significantly enriched in the target sample compared to the control sample are deemed to be binders. Adapted from Chen et al.^[89]

In 2015, Shui and co-workers explored the use of ligand-observed MS for fragment screening.^[90] Their methodology was first validated using the model enzyme human carbonic anhydrase I (hCAI) and reported inhibitors of the enzyme, which showed a close correlation between K_D values determined from ligand-observed MS with those determined by ITC. The researchers then screened 50 randomly selected fragments against hepatitis C virus RNA polymerase NS5B, an enzyme that is essential for viral replication.^[91] Nine fragment hits emerged from the ligand-observed MS screening campaign, with K_D values determined to be in the range of 0.7 to 9.6 mM. Seven of those hits were also shown to bind to NS5B using SPR. In a follow-up study, the

researchers integrated ligand-observed MS into a fragment-based lead discovery pipeline.^[89] A library of 384 fragments was screened against NS5B either all at once, generating 20 fragment hits, or in cocktails of about 50 fragments, which yielded 12 fragment hits. Screening of the mixture of the latter 12 compounds using a further round of ligand-observed MS validated 10 of those hits, which bound to NS5B with K_D values of between 1 to 20 mM. Although binding affinities were too weak to be determined by SPR, the rank order of SPR response showed reasonable correlation with the order determined by ligand-observed MS.

Instead of ultrafiltration, fragment binders can be separated from non-binders by the use of weak affinity chromatography (WAC). In this approach, fragments are passed through a column upon which the protein of interest is immobilized. Compounds that bind to the protein are retarded within the column and elute at a longer retention time compared to inactive compounds (**Figure 1.9a**). In 2011, Ohlson and co-workers first applied WAC to fragment screening.^[92] They initially validated their methodology using a small collection of 20 known binders of trypsin and thrombin as model target proteases. Screening was then performed by passing fragments through the column using HPLC, either as singletons or as a mixture. The apparent K_D of compounds were calculated by dividing the total number of binding sites on the column by the product of the adjusted retention time and the flow rate. The results showed that fragment binding could be detected over a K_D range of 10 to 1000 μ M, with good correlation between the K_D of compounds calculated by WAC with IC_{50} values determined by an enzymatic assay. Additionally, the issue of non-specific binding could be addressed by pre-treating the column with highly-affinity protease inhibitors. As these inhibitors saturate the active site, any retardation observed in the presence of the protease inhibitor would represent non-specific binding of the compounds to other parts of the protein molecule. The net retention time (t_R'), which is the difference in retention time for a particular compound in the presence and absence of the protease inhibitor, therefore gives an indication of the specific component of binding (**Figure 1.9b**). The Ohlson group later applied WAC to high-throughput screening of a commercial fragment library.^[93] 590 fragments randomly chosen from the TimTec fragment collection were passed through a thrombin column in cocktails of 35 to 65 fragments, resulting in 30 initial hits. However, screening of the compounds using an

inhibitor-blocked thrombin column revealed that only 1 of the 30 compounds bound specifically to the active site. Fragment screening campaigns using WAC have also been conducted by the Ohlson group against cyclin G-associated kinase^[94] and HSP90^[95] In the case of HSP90, good overlap was observed between hits revealed by WAC and those identified by NMR and SPR experiments.

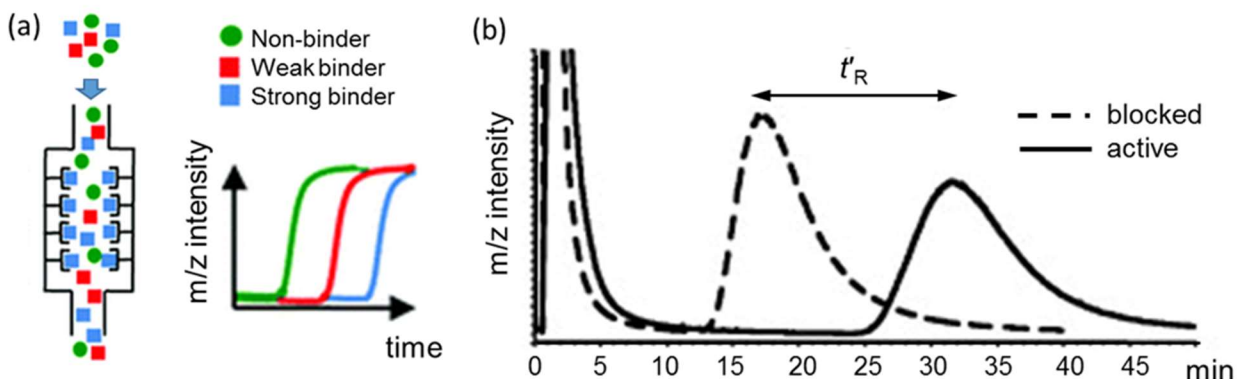


Figure 1.9. (a) Schematic of WAC. Compounds elute in order of their affinity to the column, with non-binders eluting first and strong binders eluting last. Adapted with permission from Potterat et al.^[96] (b) The net retention time (t'_R) is calculated by subtracting the retention time of the fragment on the inhibitor-blocked column (dashed line) from the corresponding retention time on the active column (solid line). Adapted with permission from Duong-Thi et al.^[92]

1.4.7. Other MS-based techniques

The entire field of MS is vast and cannot be satisfactorily reviewed here. Other techniques, such as protein footprinting or chemical cross-linking, can be used in conjunction with MS to aid the study of protein-ligand interactions.^[42] Additionally, matrix-assisted laser desorption/ionization (MALDI) is an alternative ionization technique that has also been commonly used in protein MS as well as in the emerging field of MS imaging.^[97] A summary of the MS-based techniques utilized for fragment screening is present in **Table 1.2**.

Table 1.2. A comparison of MS techniques used for fragment screening.

Technique	Principle	Advantages	Limitations
-----------	-----------	------------	-------------

Native MS	Detection of gas-phase protein-ligand complexes generated by ESI	Low sample consumption, medium-throughput, provides binding affinity and stoichiometry, allows verification of protein and ligand purity	Gas-phase species may not reflect distribution in solution
IM-MS	Gaseous ions are retarded by drift gas molecules based on their collision cross section	Can reveal conformational changes	Low-throughput
HDX-MS	Ligand changes deuteration rate of protein amide residues in solution	Can reveal ligand binding site and allostery	Low-throughput
Ligand-observed MS	Bound and unbound ligands separated by ultrafiltration or affinity chromatography before MS detection	Medium to high-throughput, easy to perform	Indirect detection technique, non-equilibrium method, requires that protein be stable to centrifugation or immobilization on column

1.5. Project aims

The use of MS-based techniques to study protein structure and interactions is a relatively new field and moreover, MS methods have not yet found widespread use in drug discovery, particularly in FBDD. At the time this project was initiated, MS methods were not routinely part of the biophysical cascade used in FBDD by the Abell research group. With the above

considerations in mind, this project broadly aimed to explore whether MS methods, particularly native MS, can become a valuable complement to the arsenal of existing biophysical tools employed within a FBDD campaign.

DMSO is common co-solvent used in fragment screening, yet literature studies on its effects on protein structure have been conflicting. In **Chapter 2**, native MS and IM-MS were used to explore the effect of DMSO on protein size and stability in the gas-phase, revealing trends that were both protein and concentration-dependent. This chapter was published in *ChemistrySelect* (*ChemistrySelect*, **2016**, *1*, 5686) and in *Analytical Chemistry* (*Anal. Chem.*, **2017**, *89*, 9976).

CoaBC, an integral member of the bacterial coenzyme A biosynthetic pathway, is a newly validated antibacterial target. In **Chapter 3**, native MS and IM-MS were used to study the structure and interactions of CoaBC and its ligands. Additionally, the application of native MS to screen fragments against the CoaB subunit was demonstrated. This chapter was published in *Chemical Communications* (*Chem. Commun.*, **2017**, *53*, 3527) and in *Angewandte Chemie International Edition* (*Angew. Chem. Int. Ed.*, **2017**, *56*, 7488).

EthR increases the resistance of *Mycobacterium tuberculosis* to ethionamide, a second-line anti-tubercular drug. In **Chapter 4**, native MS was used to study the structure of the EthR-DNA complex, revealing a hitherto unknown 6:1 stoichiometry. Additionally, the use of native MS to screen fragments against a protein-DNA interaction was demonstrated for the first time. This chapter was published in *Biochemical Journal* (*Biochem. J.* **2019**, *476*, 3125).

The aberrant activity of Aurora A kinase is implicated in the development of several kinds of cancers. In **Chapter 5**, native MS was used to study the interaction between different Aurora A mutants and TPX-2, a cognate binding protein that increases the activity and stability of the enzyme.

In **Chapter 6**, conclusions of the work and the future outlook of the use of native MS in FBDD are presented.

2. Insight into effects of DMSO on protein charging, conformation, stability and interactions as assessed by native ion mobility mass spectrometry

2.1. Introduction

Over the last two decades, electrospray ionization-mass spectroscopy (ESI-MS) has emerged as a versatile technique to probe protein structure and function.^[40, 98-99] The ability of ESI-MS to generate macromolecular ions from aqueous solution under relatively soft ionization energies has enabled the study of both non-covalent protein-protein and protein-ligand complexes under native-like conditions.^[40, 60] The interfacing of ESI-MS with ion-mobility (IM) spectrometry, termed ion-mobility mass spectrometry (IM-MS), provides an additional dimension of resolution whereby ions are separated by their collision cross-section (CCS) as they traverse a chamber of neutral gas particles under the influence of an electric field.^[42, 61, 100-101]

Biological screening systems often contain significant quantities (up to 10%) of dimethyl sulfoxide (DMSO), particularly where high concentrations of weakly-binding ligands such as fragments are used.^[29, 102] Interestingly, low DMSO concentrations (<10%) decrease the average charge of protein ions generated by ESI, whereas higher concentrations (>10%) of DMSO lead to an increase in average charge and also induce a broadening of charge state distributions.^[103-107] However, the mechanism by which low concentrations of DMSO decrease charge is still under debate.^[106, 108] Williams and co-workers have attributed the charge-reducing effects of DMSO to a “global compaction” of protein structure, as supported by both circular dichroism (CD) and hydrogen-deuterium exchange (HDX) experiments of hen egg white lysozyme and equine myoglobin in solution.^[105] Their results suggested that the proteins underwent compaction at concentrations of DMSO up to about 50%, followed by unfolding at higher concentrations.^[105] In contrast, Tjernberg and co-workers have reported that DMSO concentrations as low as 3% induced destabilization, degradation and aggregation of the phosphatase domain of PFKFB1 (BPase) as revealed by both ESI-MS and solution assays.^[103] Other additives that have been reported to reduce charge in ESI include DMF,^[108] imidazole,^[109] and various solvent vapors.^[110]

In terms of protein interactions, Zenobi and co-workers reported that DMSO (0.5 to 8%) reduced the binding affinities of three different protein-ligand complexes, which was attributed to the ability of DMSO to destabilize the protein and perturb the ligand-binding pocket, interfere with the ligand, or act as a competing inhibitor.^[111] In contrast, Landreh and co-workers found that 3% DMSO stabilized the tetrameric state of transthyretin (TTR) and its non-covalent interaction with the thyroid hormone thyroxine (T4), as well as the trimeric state of the C-terminal domain of lung surfactant protein C (CTC) and its non-covalent interaction with a trivalent peptide.^[108] This protective effect of DMSO was attributed to a combination of charge reduction and a cooling effect from adduct dissociation.^[108]

The resistance of protein complexes to dissociation and unfolding in the gas phase can be studied by collision-induced dissociation (CID) and collision-induced unfolding (CIU), respectively. Previous studies mainly by the Ruotolo group have explored the effect of various cations and ions on the CID and CIU stabilities of model proteins.^[112-115] However, to our knowledge, no systematic study on the effect of DMSO on protein CID or CIU behavior has yet been reported.

This chapter aims to clarify whether the supposed compaction of protein structure at low concentrations of DMSO is manifested as a change in CCS that can be detected by traveling-wave ion-mobility mass spectrometry (TWIM-MS) for four different proteins: avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase. These four tetrameric proteins were chosen due to their different molecular weights and roughly globular structures and well-characterized behavior in IM-MS. In the second phase of this study, the effect of DMSO on protein stability and interactions was investigated by CID and CIU. Avidin and CYP142A1 were chosen as representative examples of proteins with quaternary structure or prosthetic groups, respectively. Avidin is a well-characterized glycoprotein found in the egg-white of birds and reptiles. The protein forms a multiprotein complex that consists of four identical subunits assembled together in a dimer of dimers arrangement.^[116] In contrast, the *Mycobacterium tuberculosis* (*Mtb*) cytochrome P450 enzyme CYP142A1 is predominantly monomeric and binds a heme-*b* prosthetic

group. CYP142A1 functions as a cholesterol oxidase and contributes to the ability of *Mtb* to establish a chronic infection.^[117-119]

To our knowledge, this is the first time that native IM-MS has been applied to study the effect of DMSO on protein conformation and stability. It was found that DMSO influences the size of protein ions not only by its effect on charge state distribution, but also at the level of individual charge states. Our work also provides evidence that DMSO influences CID and CIU stability via a dual mechanism of charge modulation, as well as via charge state-independent effects. Finally, tandem MS/MS experiments showed that DMSO can influence the heme dissociation pathway of CYP142A1, and also demonstrate that the protective effects of the heme group on CYP142A1 stability can be tracked using CIU, a technique that should be broadly applicable to other heme-containing proteins. Overall, these results provide insight into the charge-reducing mechanism of DMSO and may help to reconcile previously conflicting observations regarding this phenomenon.

2.2. Results and discussion

2.2.1. Effect of DMSO on protein charge and conformation

2.2.1.1. DMSO exerts opposing trends on average protein charge

Native mass spectra of avidin (**Figure 2.1a**), concanavalin A (**Figure 2.1b**), alcohol dehydrogenase (**Figure 2.1c**) and pyruvate kinase (**Figure 2.1d**) in the absence of DMSO indicated that their quaternary structures were well-preserved, with the tetramer presenting as the predominant species in all cases. For concanavalin A, a significant amount of the dimeric protein and a small quantity of the monomeric protein were also observed, consistent with previous reports.^[104, 120] Meanwhile, some non-specific dimerization of tetramers was observed for avidin, while in the case of pyruvate kinase, an unknown impurity of around 155 kDa was observed at a lower m/z region relative to the expected tetramer peaks.

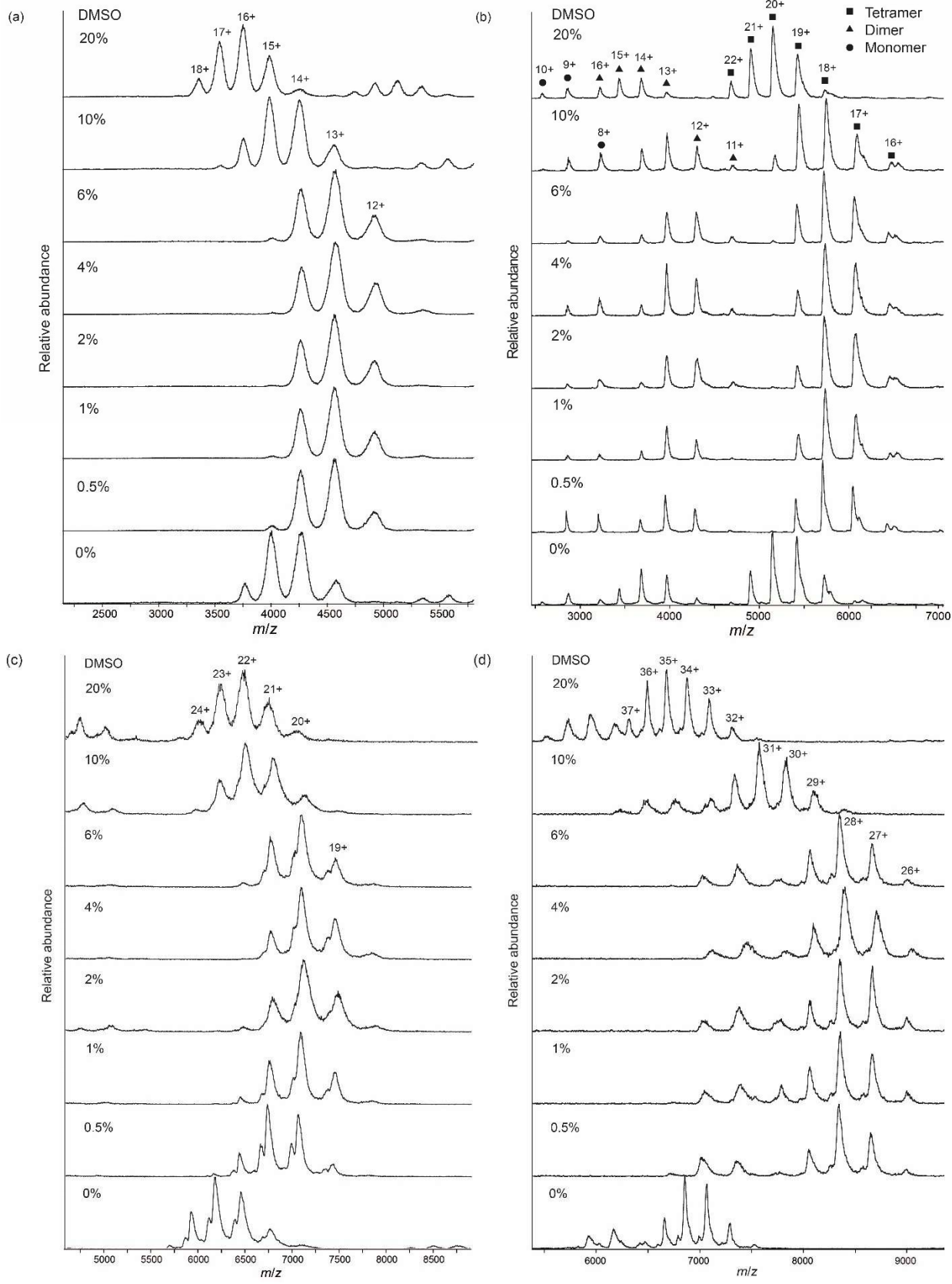


Figure 2.1 nESI-MS spectra of (a) avidin (20 μ M), (b) concanavalin A, (c) alcohol dehydrogenase, and (d) pyruvate kinase in NH_4OAc (200 mM, pH 7.0) containing 0, 0.5, 1, 2, 4, 6, 10 or 20% DMSO. For all spectra except concanavalin A (b), only the charge states of the tetramer are indicated. For avidin (a), some non-specific dimerization of tetramers was observed. For pyruvate kinase (d), an unknown impurity of around 155 kDa was observed at a lower m/z region relative to the tetramer signals.

As the concentration of DMSO was increased from 0 to 20%, the charge state distributions of the four protein complexes were observed to shift to higher m/z values (lower charge) at low DMSO concentrations, followed by a reversal of the trend towards lower m/z values (higher charge) at higher DMSO concentrations (**Figures 2.2a-d**). For example, the abundance-weighted average charge of avidin decreased monotonically from 15.5+ to 14.1+ as the DMSO concentration was increased from 0% to 4%, followed by an increase to 17.3+ as the DMSO concentration was increased to 20% (**Figure 2.2a**). These opposing trends of DMSO on protein charging are consistent with previous reports,^[103, 105-106] suggesting that at least two factors are at play here. Additionally, peak broadening was observed at higher DMSO concentrations in some cases, which might be attributed to formation of DMSO adducts of the protein.

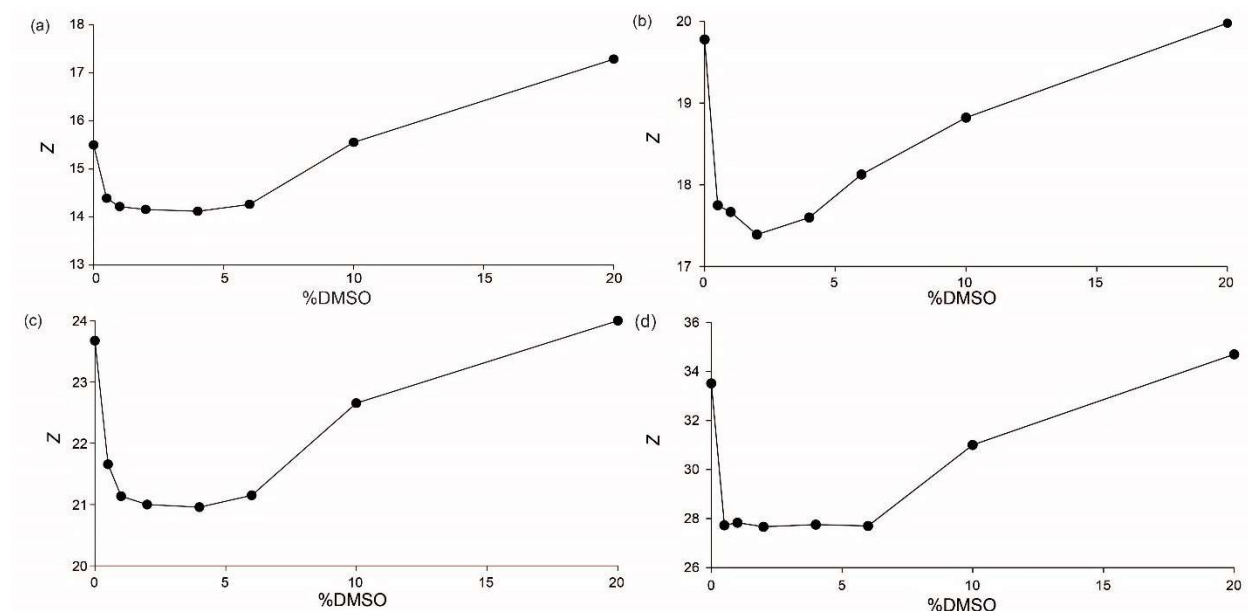


Figure 2.2 Weighted-average charge of (a) avidin tetramers, (b) concanavalin A tetramers, (c) alcohol dehydrogenase tetramers, and (d) pyruvate kinase tetramers as a function of DMSO concentration.

The minimum average charge for the four proteins occurred at 2 or 4% concentration of DMSO (**Figures 2.2a-d**). Furthermore, concanavalin A dimers and monomers showed similar charging behavior in response to DMSO concentration as compared to the tetramers (**Figure 2.3a**). For all of the protein complexes, significant charge state distribution broadening was not apparent until DMSO concentrations reached high values ($\geq 10\%$). At 20% DMSO, the average charge of the proteins became greater than their initial values in the absence of DMSO, suggesting the onset of global protein unfolding. This could account for the increase in the number of occupied charge states at higher DMSO concentrations,^[121] as well as the increase of the abundance of dimerized avidin tetramers that are presumed to be non-specific (**Figure 2.1a**). It should be noted that because of the higher boiling point of DMSO compared to water, ESI droplets become enriched in DMSO (estimated to be 3- to 5-fold in Williams et al.^[105]) during the desolvation process, and hence the effective DMSO concentration experienced by the gaseous protein ions is likely to be higher than in solution.

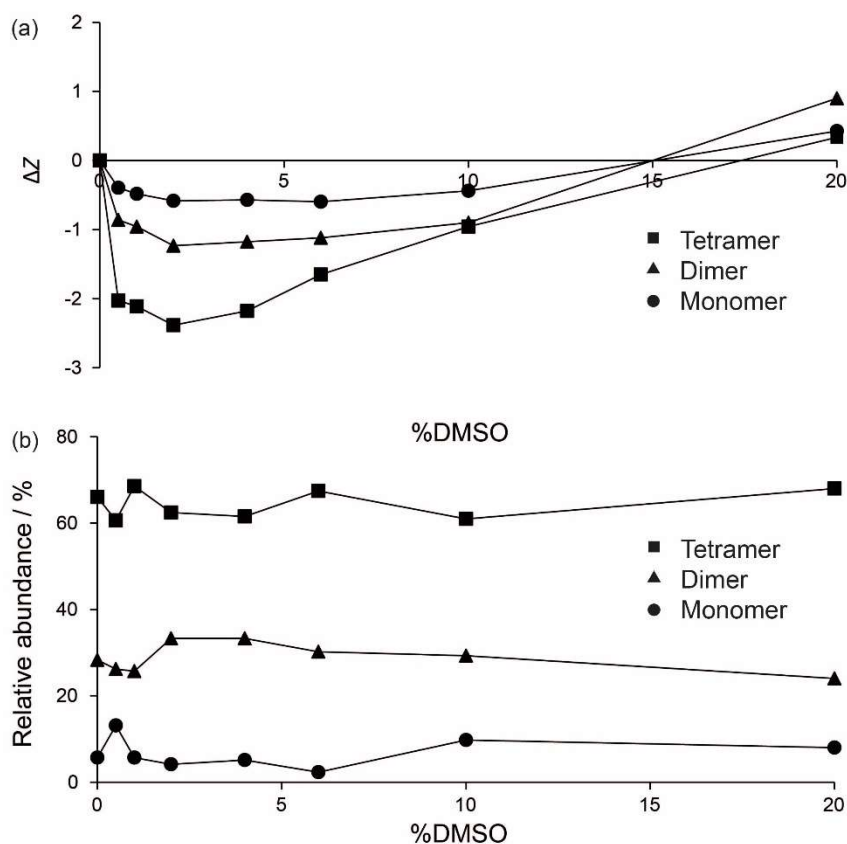


Figure 2.3 (a) Relative charge reduction of concanavalin A tetramers, dimers and monomers as a function of DMSO concentration. (b) Relative abundance of concanavalin A tetramers, dimers and monomers as a function of DMSO concentration.

A previous report found that 2.5% DMSO induced significant dissociation of the bacterial NAD⁺ synthase from dimeric into monomeric form, possibly via weakening of the hydrophobic effect.^[122] In comparison, no dissociation of any of the four protein tetramers was observed by native MS at any DMSO concentration. Moreover, the tetramer-dimer-monomer equilibrium for concanavalin A was not significantly perturbed by the addition of up to 20% of DMSO (**Figure 2.3b**). This also contrasts with observations from supercharging experiments, in which the addition of as little as 0.5% of *m*-nitrobenzyl alcohol resulted in significant dissociation of concanavalin A dimers into monomers.^[104] This suggests that the charge-reducing effects of DMSO protect against, or at least do not promote, subunit dissociation in ESI in this case. It has also been previously reported that charge reduction of protein complexes using crown ethers protected against dissociation.^[123]

2.2.1.2. DMSO exerts opposing trends on average protein size

CCS values for the protein complexes at various DMSO concentrations were obtained in N₂ and calibrated with CCS^{N₂} reference values for the proteins under DMSO-free conditions following the procedure of Ruotolo and Robinson *et al.*^[76] Experiments were performed at four different wave heights (7, 8, 9 and 10 V) to rule out the effect of electric field strength on gaseous ion separation. Intriguingly, the abundance-weighted average CCS of the four proteins underwent a decrease in value at low DMSO concentrations followed by an increase at higher DMSO levels (**Figures 2.4a-d**), paralleling the trend observed for average charge. For example, the average CCS of the avidin tetramer decreased from $4130 \pm 10 \text{ \AA}^2$ at 0% DMSO (DT CCS^{N₂} = 4150 \AA^2)^[79] to $4090 \pm 10 \text{ \AA}^2$ at 2% DMSO, followed by an increase to $4200 \pm 30 \text{ \AA}^2$ at 20% DMSO (**Figure 2.4a**). While these differences were small (*ca.* 1–2%), the overall trends were repeatable in independently-performed experiments.

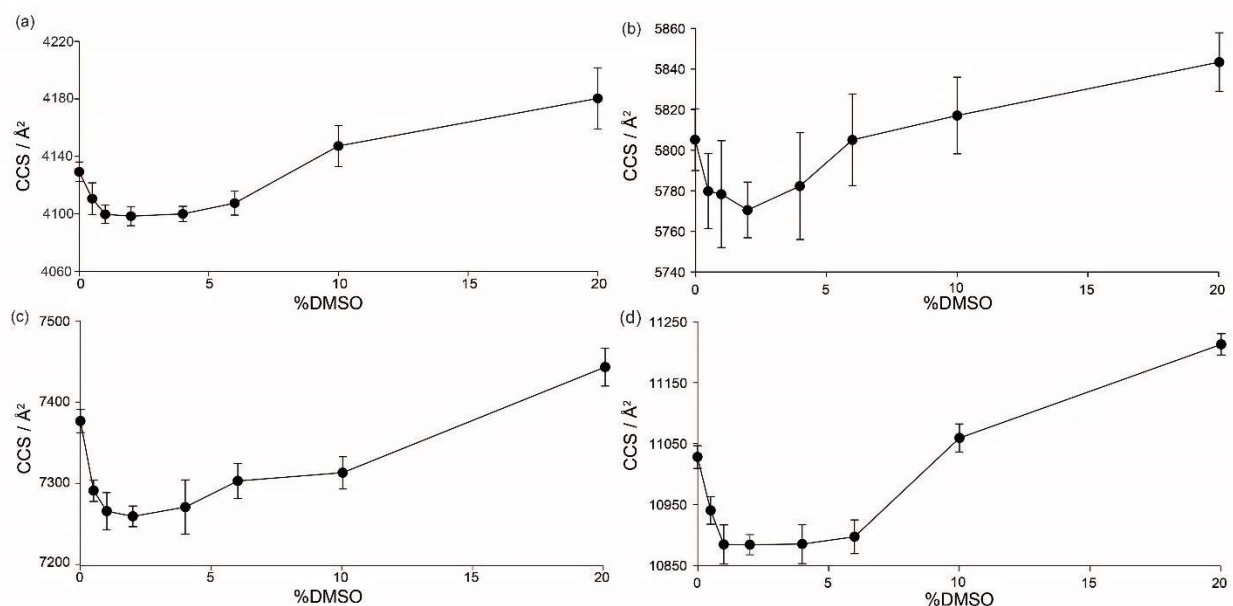


Figure 2.4 Weighted-average CCS of (a) avidin tetramers, (b) concanavalin A tetramers, (c) alcohol dehydrogenase tetramers, and (d) pyruvate kinase tetramers as a function of DMSO concentration.

The minimum CCS values for the four proteins occurred at 1 or 2% DMSO (**Figures 2.4a-d**), which are similar to the DMSO concentrations at which minimum charge was reached. At 20% DMSO, the average CCS of the proteins became higher than their initial values under DMSO-free conditions, which is consistent with preliminary unfolding and the onset of protein supercharging at this DMSO concentration. These data would therefore support the idea of global protein compaction at low DMSO levels,^[105] followed by the onset of protein unraveling at higher concentrations.

2.2.1.3. DMSO affects protein size at the level of individual charge states

Higher charge states of a protein generally have larger CCSs than lower charge states of the same protein, which has been attributed to increased Coulombic repulsion leading to expansion of ions in the gas phase, or alternatively, to the transfer of less compact proteins from solution that provide greater surface area for charge development in ESI.^[124] This poses an interesting “chicken-or-egg” question: is the apparent compaction of protein structure at low DMSO concentrations a direct consequence of the charge-reducing capability of DMSO in ESI, allowing

protein ions to access lower charge states that have decreased internal repulsion and hence smaller size (Mechanism A), or is protein compaction an *ab initio* effect by DMSO in solution that in turn decreases the charge developed in ESI (Mechanism B) (**Figure 2.5**)?

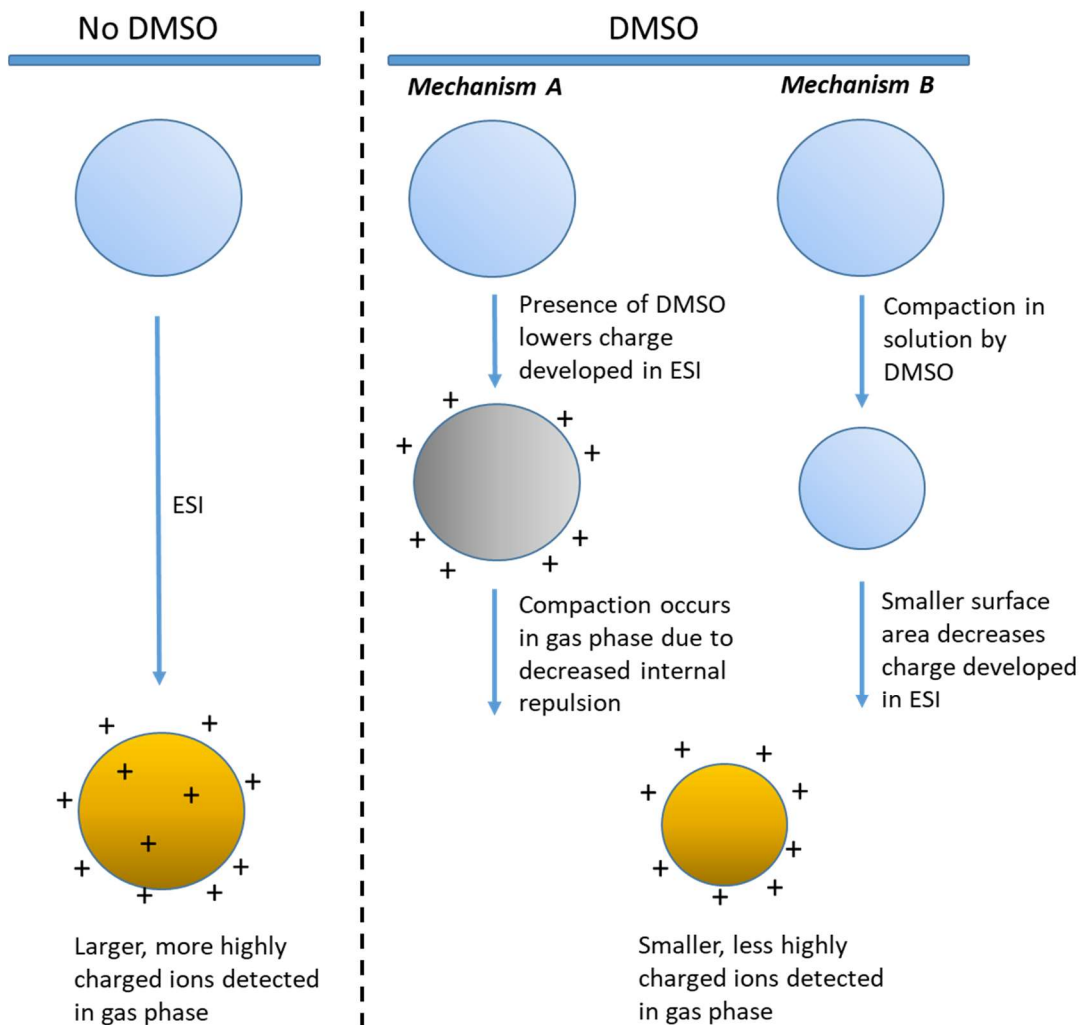


Figure 2.5 Two possible mechanisms for size reduction of protein ions by DMSO. In Mechanism A, DMSO decreases charge development during ESI and compaction occurs in the gas phase due to lower internal repulsion of ions. In Mechanism B, compaction by DMSO occurs in solution, decreasing the area for charge development during ESI.

To investigate this issue, the relationship between the CCSs for single charge states and DMSO concentration was analyzed. If the observed compaction of protein structure was solely due to

the ability of DMSO to shift the protein charge state distribution to ions of lower charge, then the CCSs of specific charge states should not be appreciably different from each other regardless of DMSO concentration. Interestingly, for several of the charge states that are present at a range of different DMSO concentrations, their CCSs exhibited the characteristic opposing trends in response to DMSO concentration (**Figures 2.6a-d**) as for average charge and average CCS values.

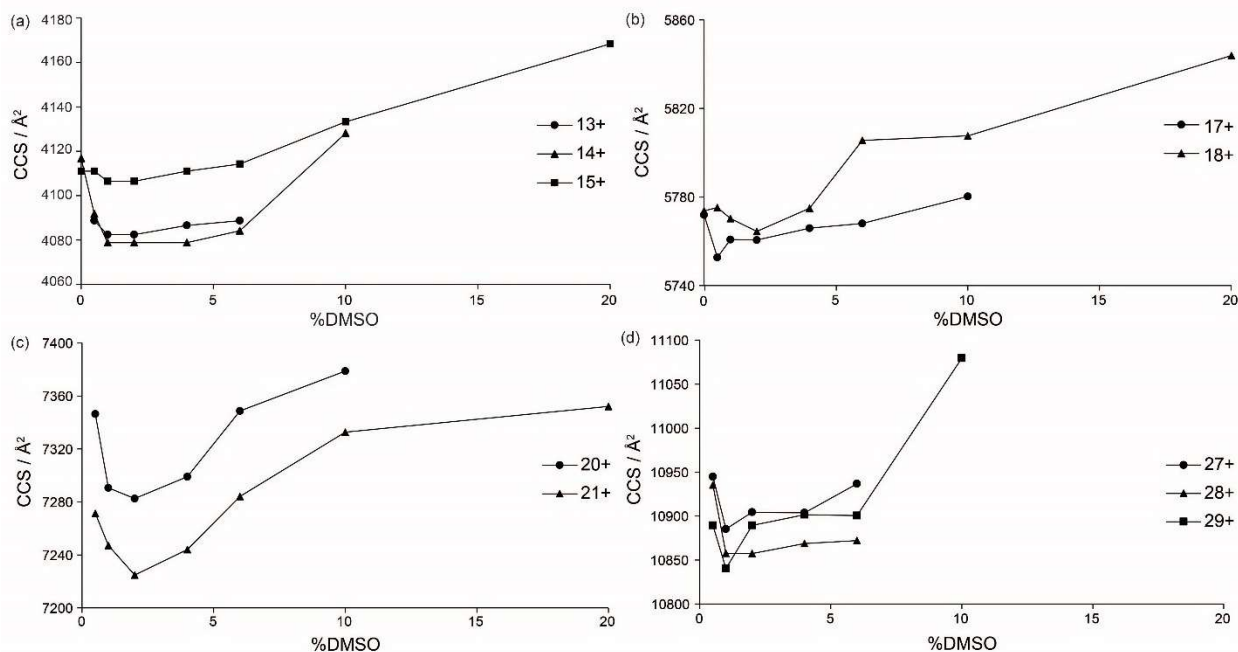
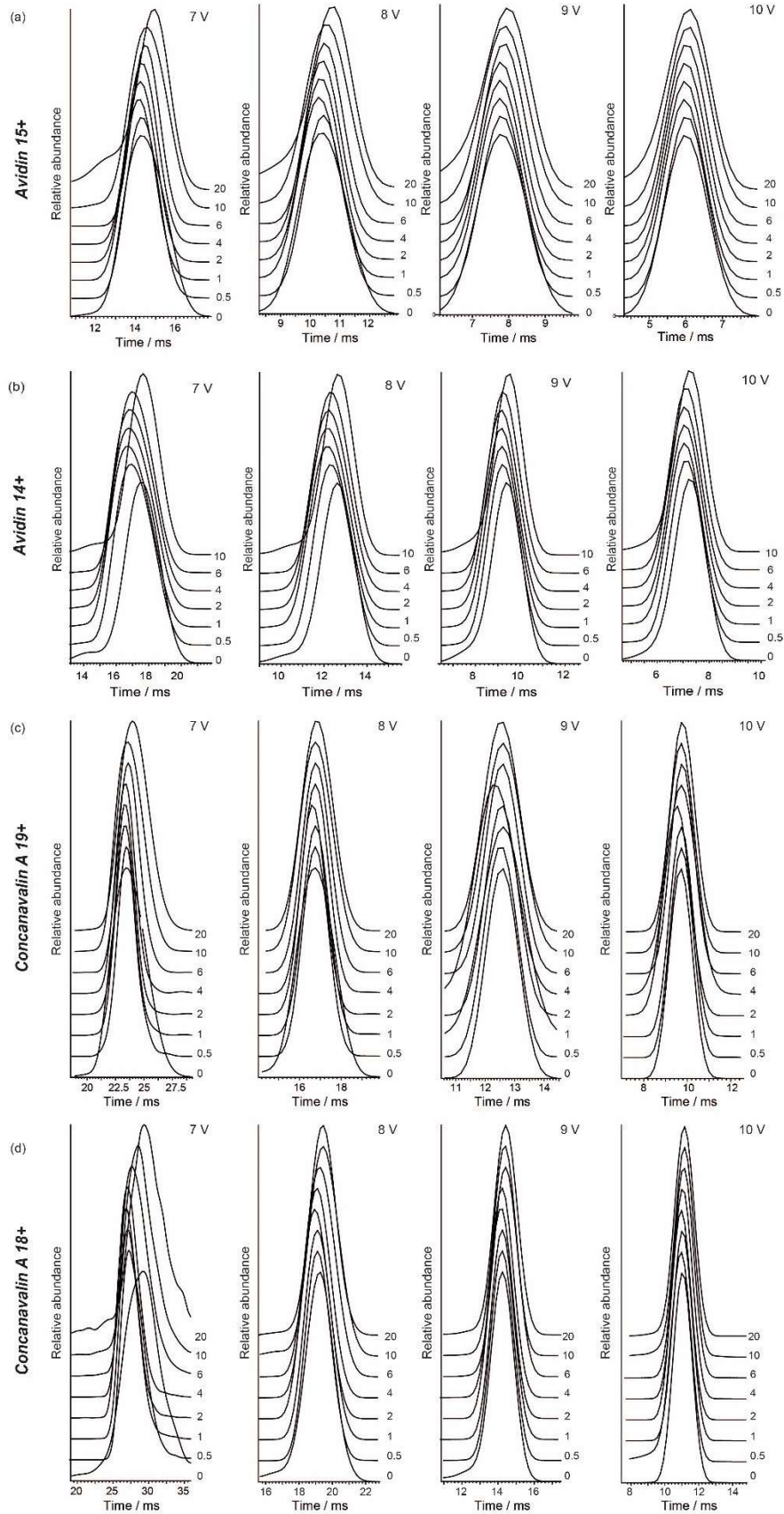


Figure 2.6 (a) CCS of 13+, 14+ and 15+ charge states of avidin tetramers as a function of DMSO concentration. (b) CCS of 17+ and 18+ charge states of concanavalin A tetramers as a function of DMSO concentration. (c) CCS of 20+ and 21+ charge states of alcohol dehydrogenase tetramers as a function of DMSO concentration. (d) CCS of 27+, 28+ and 29+ charge states of pyruvate kinase tetramers as a function of DMSO concentration.

This behavior was also evident from consideration of the arrival time distributions (ATDs) for individual charge states (**Figures 2.7a-h**), which shift to lower drift times at lower DMSO concentrations (indicating protein compaction), followed by a shift to larger drift times at higher DMSO concentrations. These observations argue against the hypothesis that the decreased Coulombic repulsion afforded by DMSO-induced charge reduction in ESI is the sole factor driving the overall compaction of the proteins. This is because if the sole effect of DMSO was to reduce

protein charge, then identical charge states of a protein generated at different DMSO concentrations should show the same drift time, however our data instead show that DMSO affects drift time even at the level of individual charge states.



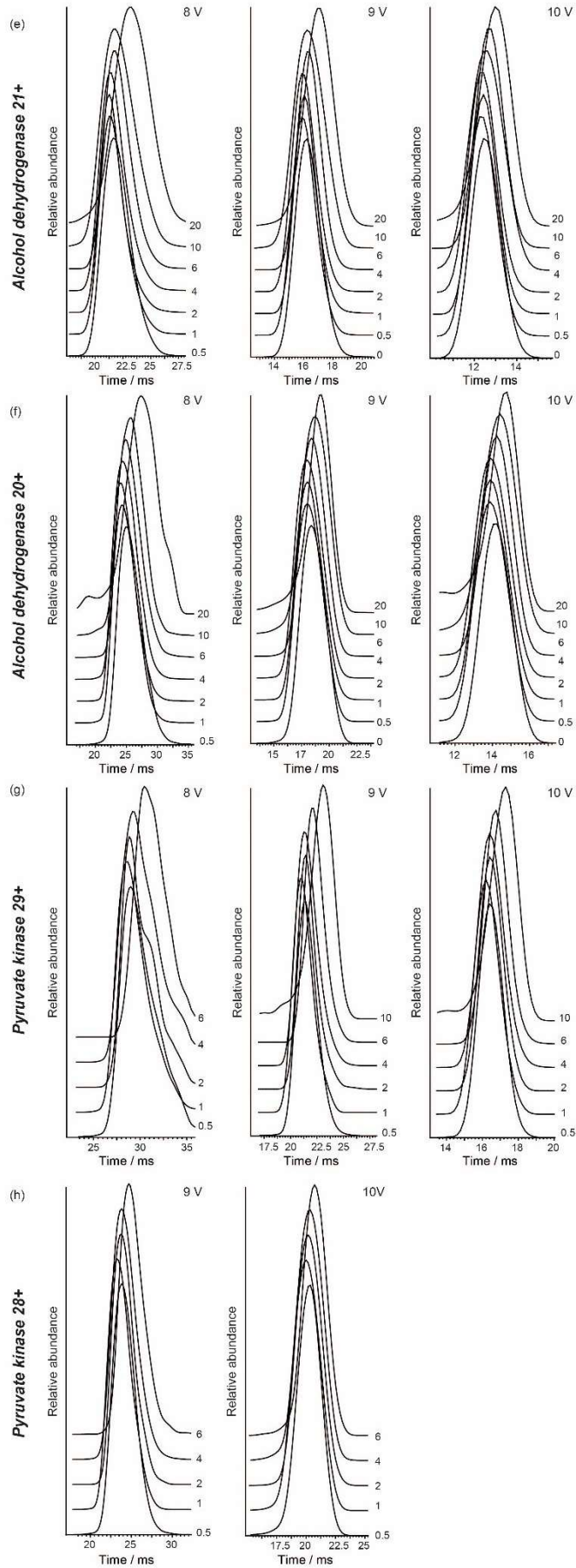


Figure 2.7 Arrival time distributions for different wave heights as a function of DMSO concentration for (a) avidin 15+, (b) avidin 14+, (c) concanavalin A 19+, (d) concanavalin A 18+, (e) aldehyde dehydrogenase 21+, (f) alcohol dehydrogenase 20+, (g) pyruvate kinase 29+, and (h) pyruvate kinase 28+ charge states. Wave heights are indicated at the top and DMSO concentrations (%) are indicated at the side of the plots.

It was also observed that in the subcharging regime, lower charge states do not necessarily exhibit smaller CCSs than higher charge states at a given DMSO concentration (**Figures 2.6a-d**). This is similar to observations from the Robinson group showing that the lower charge states of serum amyloid P component (SAP) and aldehyde dehydrogenase complexes subjected to charge reduction by triethylammonium acetate (TEAA) do not show significant changes in CCS compared to higher, non-reduced charge states.^[125] Moreover, drift tube-ion mobility-MS measurements of native-like, non-supercharged proteins suggest that the positive correlation between charge and CCS is present only for smaller proteins, whereas for larger proteins or protein complexes, this relationship is not apparent.^[79, 126] This further suggests that internal repulsion should not have a significant effect on the CCSs of the different charge states of the proteins in this study, which was performed under non-activating conditions.

2.2.1.4. Error analysis

Due to the relatively small deviations in CCS observed in this study, error analysis was performed to evaluate the significance of the findings. According to Ruotolo and Robinson *et al.*,^[76] the total error (E_T) of the calibrated CCSs is given by the sum of E_R , E_{Cal} and E_S , where E_R is the standard deviation of replicate measurements, E_{Cal} is the average error of the calibration curve, and E_S is the error carried by the protein standards that were used to calibrate the drift times of the unknown ions. Each of these errors for this experiment are considered separately below.

E_R : The reproducibility error (E_R), as estimated by the relative standard deviation of replicate measurements across four different wave heights, was found to be low for each charge state of the four proteins across all DMSO concentrations (**Table 2.1**). The majority of E_R values were less than 1%, indicating good reproducibility between replicate measurements. The average E_R for

avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase across all DMSO concentrations were 0.33%, 0.48%, 0.40% and 0.50%, respectively (average 0.43%).

E_{cal} : The calibration curve R^2 values were ≥ 0.978 for all four wave heights, indicating good linearity (**Table 2.2**). The average calibration curve-based error (E_{cal}) at wave heights 7 V, 8 V, 9 V and 10 V were estimated to be 1.41%, 1.24%, 1.21% and 1.19%, respectively (average 1.26%).

E_s : The error in mobilities of the protein calibrants (E_s) has been assumed to be *ca.* 1% by Ruotolo and Robinson *et al.*,^[76] or in a more recent paper by the same group, to be <3%.^[79]

Adding all three uncertainty components in quadrature suggests that the total error E_T associated with the calibrated CCSs in this study is 1.67% (when $E_s = 1\%$) or 3.28% (when $E_s = 3\%$). While the magnitude of these errors are significant in relation to the DMSO-induced CCS changes observed, it is argued that two of the three error contributors, namely the calibration-related errors, E_{cal} and E_s , do not impact the validity of the conclusions of this study. This is because the main findings of this study do not rely on comparisons of experimental CCS values with calculated CCS values of model structures, for which calibration of measurements is an essential step. Instead, the context of this study is such that only the determination of relative changes in CCSs are necessary for assessing the effect of DMSO on protein conformation. It is therefore assumed that the two calibration-based uncertainties, E_{cal} and E_s , are systematic in nature, and thus should have no effect on the relative differences in CCS observed in this study. As such, the CCS errors in this chapter incorporate only the reproducibility error (E_R).

This argument is bolstered by the arrival time distribution data for selected charge states (**Figures 2.7a-h**), which clearly show the opposite trends of DMSO on drift time even at the level of individual charge states. Moreover, the trends were broadly consistent for each protein ion across the different wave heights. Importantly, these relative differences in arrival time distribution, being antecedent to the calibration process, are unadulterated by the calibration-related uncertainties E_{cal} and E_s .

Finally, the overall trends in DMSO-induced conformational change were repeatable from independently-performed repeat experiments conducted on different days. Taken together, this data provides sufficient evidence for the significance of the small, but repeatable changes in protein conformation induced by DMSO.

Table 2.1 Table showing the reproducibility error (E_R) as estimated by the standard deviation (%SD) for four replicate measurements for each charge state of avidin (AV), concanavalin A (CCA), alcohol dehydrogenase (ADH) and pyruvate kinase (PK) at different DMSO concentrations.

[DMSO] / %	AV		CCA		ADH		PK	
	z	%SD	z	%SD	z	%SD	z	%SD
0	14	0.21	18	0.30	22	0.44	32	0.35
	15	0.18	19	0.21	23	0.24	33	0.44
	16	0.25	20	0.40	24	0.22	34	0.60
	17	0.13	21	0.38	25	0.26	35	0.53
0.5	13	0.26	16	0.78	20	0.38	26	0.56
	14	0.46	17	0.26	21	0.28	27	0.60
	15	0.18	18	0.30	22	0.17	28	0.76
	16	0.58	19	0.67	23	0.27	29	0.21
1	13	0.31	16	1.13	20	0.68	26	1.45
	14	0.16	17	0.91	21	0.40	27	0.54
	15	0.19	18	0.34	22	0.26	28	0.81
	16	0.22	19	0.39	23	0.23	29	0.26
2	13	0.31	16	0.76	20	0.45	26	0.46
	14	0.16	17	0.01	21	0.06	27	0.49

	15	0.19	18	0.45	22	0.31	28	0.28
	16	0.56	19	0.19	23	0.50	29	0.21
4	13	0.11	16	0.97	20	1.09	26	0.74
	14	0.16	17	0.83	21	0.44	27	0.66
	15	0.18	18	0.34	22	0.40	28	0.46
	16	0.56	19	0.34	23	0.44	29	0.80
6	13	0.26	16	0.92	20	0.62	26	1.10
	14	0.22	17	0.36	21	0.34	27	0.43
	15	0.25	18	0.47	22	0.26	28	0.28
	16	0.55	19	0.49	23	0.44	29	0.47
10	14	0.26	17	1.11	21	0.76	29	0.53
	15	0.36	18	0.23	22	0.48	30	0.60
	16	0.60	19	0.43	23	0.22	31	0.28
	17	0.34	20	0.28	24	0.23	32	0.13
20	16	0.33	18	0.73	22	1.14	33	0.25
	17	0.76	19	0.13	23	0.34	32	0.47
	18	0.64	20	0.36	24	0.39	33	0.25
	19	0.77	21	0.29	25	0.32	34	0.19
			22	0.24	26	0.30	35	0.04
							36	0.86
							37	0.29

Average	0.33	0.48	0.40	0.50
---------	------	------	------	------

Table 2.2 Table showing the curve R^2 and calibration curve-based error (E_{cal}) at four different wave heights.

Wave height / V	7	8	9	10	Average
Curve R^2	0.978	0.981	0.981	0.981	0.980
E_{cal} / %	1.41	1.24	1.21	1.19	1.26

2.2.1.5. Towards a consistent understanding of protein compaction by DMSO

The observations above suggest that protein compaction by DMSO may be occurring, at least in part, in solution phase (*i.e.*, Mechanism B, **Figure 2.5**). The compaction effect by DMSO in solution has been previously documented for lysozyme by the groups of Voets^[127] and Williams^[105] on the basis of CD, small-angle neutron scattering, Rayleigh scattering and HDX-MS results. This compaction of protein structure in solution could then be transferred into the gas-phase for it to be detected by TWIM-MS, as was shown in this work. However, while this could potentially account for the charge reduction of protein ions at low DMSO concentrations,^[105] it is considered that this is unlikely to be the only operative mechanism, as the degree of CCS reduction (*ca.* 1–2%) was small in comparison to the magnitude of charge reduction (*ca.* 10%).

Quantitatively, Marijnissen and co-workers have suggested based on theoretical considerations that the charge of a progeny droplet generated during ESI is proportional to droplet surface area raised to the power of 3/4.^[128] According to the CRM model, the final droplet is not much larger than the protein itself. Therefore, as the droplet finally evaporates, the entirety of this charge should be transferred to the encapsulated protein.^[129] By correlating the average charge of proteins observed in ESI-MS with their predicted surface areas from crystal structures, Kaltashov and co-workers determined that the charge–surface area relationship followed a power function of 0.69 ± 0.02 .^[129] In our study, plotting a log-log graph of average charge against average CCS for

the four proteins under DMSO-free conditions revealed an excellent linear fit ($R^2 = 0.997$), with a power constant of 0.78, which was close to the expected value of 3/4 (**Figure 2.8a**). However, log-log plots for the individual complexes under varying DMSO concentrations showed both worse fits ($R^2 = 0.82$ to 0.97) as well as much higher power constants of 4.5 to 10.6 (**Figure 2.8b**), suggesting that the DMSO-induced change in CCS, and by extension surface area, is itself insufficient to account for the entirety of the change in charge.

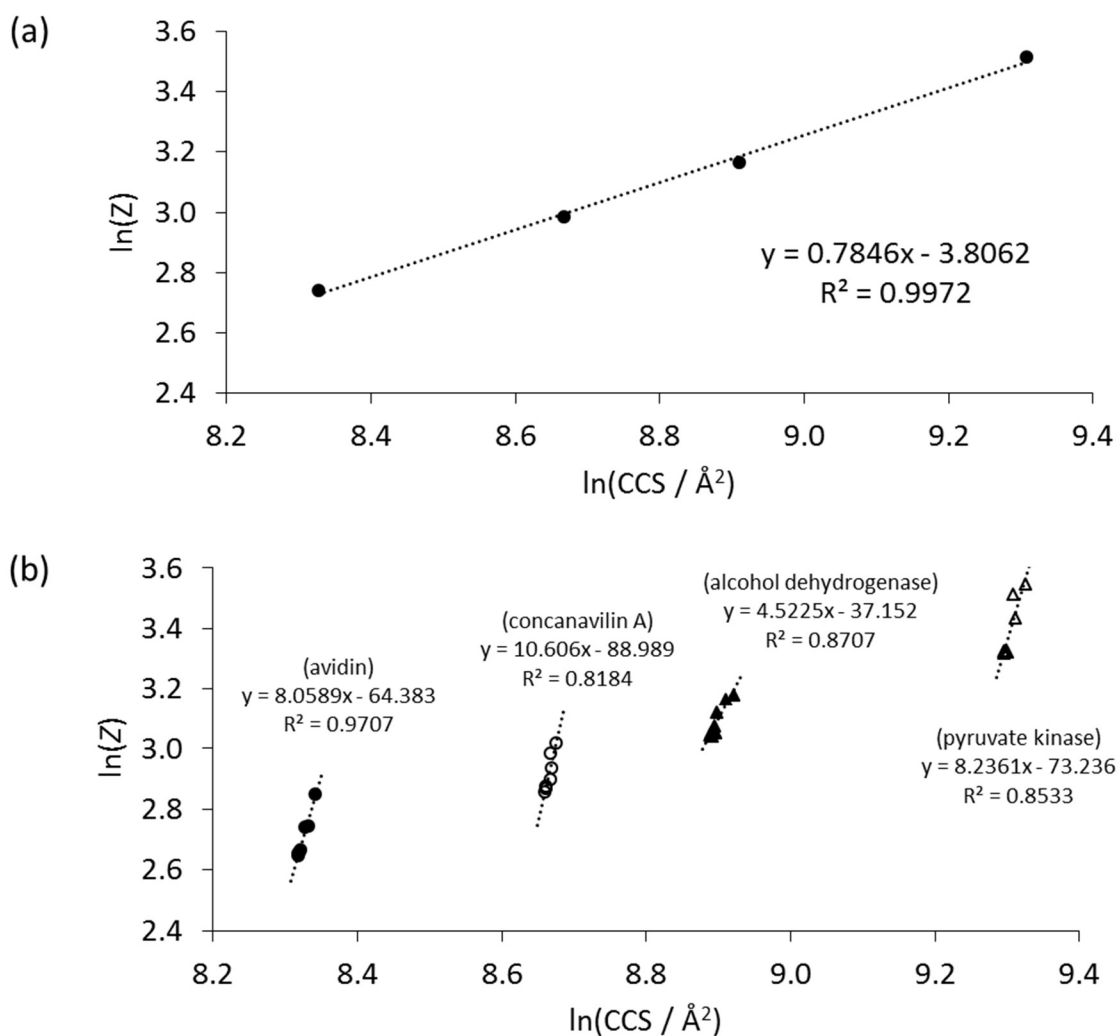


Figure 2.8. (a) Relationship between average charge and CCS plotted on a log–log scale for the four protein complexes at 0% DMSO, showing the equation and correlation coefficient for a linear trendline. (b) Relationship between average charge and CCS plotted on a log–log scale for the four protein complexes at

various DMSO concentrations (0 to 20%). Values for avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase are shown as closed circles, open circles, closed triangles and open triangles, respectively. Linear trendlines are plotted for each individual protein.

Instead, other chemical or physical mechanisms, such as gas-phase proton transfer^[106] or dissociative cooling,^[108-109] may account for the observed subcharging of protein ions at low DMSO concentrations. Additionally, this does not preclude the possibility that the observed differences in CCS might be due to other factors; for example, gaseous ions might, depending on the DMSO concentration, vary in their internal energy as a result of different desolvation, leading to small changes in drift time. In the specific case of ring-like protein assemblies, initial compaction can also be achieved through collision activation,^[130] although this mechanism is unlikely to be operative for the approximately globular multiprotein complexes considered in this work.

2.2.1.6. Summary

The precise mechanisms governing the relationships between electrospray conditions, protein charge and CCS values are still not completely understood. In particular, a consistent framework that accounts for the charge-reducing effects of DMSO in ESI has not yet been fully developed. Against this backdrop, this study has provided evidence for the first time of a modest, but repeatable, decrease of protein size at low DMSO concentrations. Our data also suggests that this compaction is unlikely to be solely attributable to the shift of charge state distribution by DMSO, as different charge states of the same protein were also individually compacted by DMSO, and also because internal Coulombic repulsion was unlikely to be a dominant factor for these systems based on consideration of protein sizes in the subcharging regime. Instead, our results lead us to speculate that the previously reported compaction of protein structure by DMSO in solution can be adequately maintained upon transfer to the gas-phase, allowing for this phenomenon to be detected by TWIM-MS. However, the observed CCS reduction is relatively small and is thus unlikely to account for the entirety of charge reduction by DMSO, suggesting that alternative subcharging mechanisms are operative. Finally, and in contrast to previous

studies with other proteins, no evidence was observed of gross protein denaturation, degradation or aggregation (except for some dimerization of avidin) of proteins under study at 4 °C over the timescale of days even at 20% DMSO. This suggests that the deleterious effects of DMSO on protein structure and stability are highly protein-dependent, and further reinforces the notion that the DMSO concentration for each biological assay has to be carefully optimized in order to ensure that the results are not compromised by unintended effects of DMSO on protein binding or activity.

2.2.2. Effect of DMSO on protein stability and interactions

2.2.2.1. Charging behavior of avidin and CYP142A1 at high levels of DMSO

The first part of this chapter presented above investigated the charging behavior of four tetrameric proteins under low to moderate (0.5 to 20% v/v) DMSO concentrations. In the next phase of the study, avidin and CYP142A1 were subjected to CID and CIU experiments to assess the effect of DMSO on protein stability and interactions. In the CID experiments, the effect of DMSO on maintaining the tetrameric structure of avidin and the protein-heme interaction of CYP142A1 was investigated by monitoring complex dissociation. Meanwhile, the CIU experiments investigated the effect of DMSO on the resistance of avidin and CYP142A1 to unfolding, as tracked with IM-MS. However, as much higher concentrations of DMSO (up to 60%) would be used for these experiments compared to the first phase of the study, the charging behavior under higher levels of DMSO first had to be investigated under minimal activation conditions.

For avidin, concentrations of DMSO higher than 60% could not be examined due to protein aggregation, whereas for CYP142A1 the maximum DMSO concentration that could be tested was 40%. The tetrameric state of avidin was well-preserved at concentrations up to 60% DMSO (**Figure 2.9a**). At 50 and 60% (v/v) DMSO, a small amount of monomeric avidin ions (16 kDa) could be detected centered around the 9+ charge state. This result suggests that high concentrations of DMSO could induce the dissociation of avidin even under minimal activation conditions.

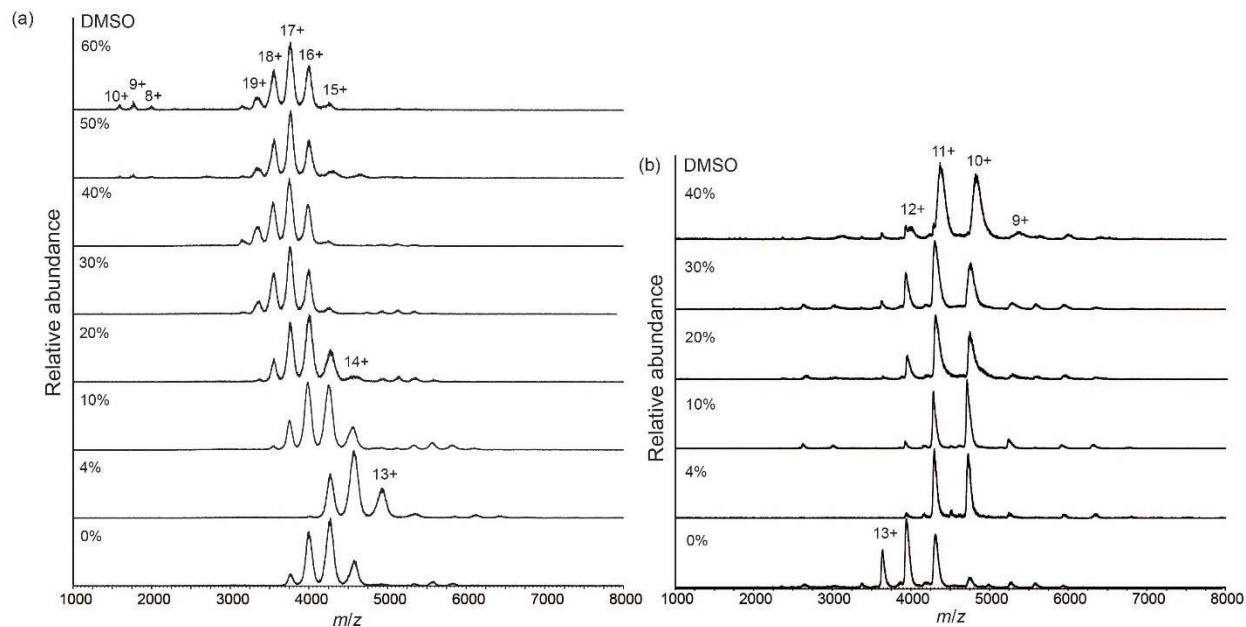


Figure 2.9 (a) Native mass spectra of avidin in the presence of 0, 4, 10, 20, 30, 40, 50 or 60% DMSO. (b) Native mass spectra of CYP142A1 in the presence of 0, 4, 10, 20, 30, or 40% DMSO.

For CYP142A1, the native mass spectra showed that CYP142A1 exists predominantly in monomeric form (**Figure 2.9b**), consistent with the reported X-ray crystal structure,^[117] and was centered around the 12+ charge state in the absence of DMSO. The observed mass of 47.2 kDa matched well with the theoretical mass of the construct (47178 kDa) used in this experiment. The protein remained monomeric as the DMSO concentration was increased to 40%. However, signals of individual charge states were broadened, which could be attributed to the adduction of small molecules (*e.g.* solvent or ions) to the protein at higher DMSO concentrations. Moreover, CYP142A1 fully retained its heme moiety even at 40% DMSO, in contrast to a previous ESI-MS study of myoglobin in which 90% of heme was already lost at 20% DMSO.^[105] To our knowledge, this is the first reported study of CYP142A1 by native MS.

The opposing trends of DMSO on average protein charge was also evident (**Figure 2.10**). However, very high DMSO concentrations (50% for avidin and 40% for CYP142A1) decreased average charge slightly, similar to previously observed for lysozyme and myoglobin.^[105] Intriguingly, the average charge of CYP142A1 reached at 30% or 40% DMSO was still lower than under DMSO-free

conditions (**Figure 2.10b**), in contrast to avidin where supercharging was observed from 10% of DMSO (**Figure 2.10a**). To our knowledge, CYP142A1 is the first protein that has been reported to remain charge-reduced at such high DMSO concentrations. In the first part of this chapter, avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase were supercharged at DMSO concentrations greater than 10 or 20%. Meanwhile, the onset of supercharging was around 20% DMSO for lysozyme and myoglobin,^[105] while for BPase both supercharging and protein denaturation were evident at as little as 3% of DMSO.^[103]

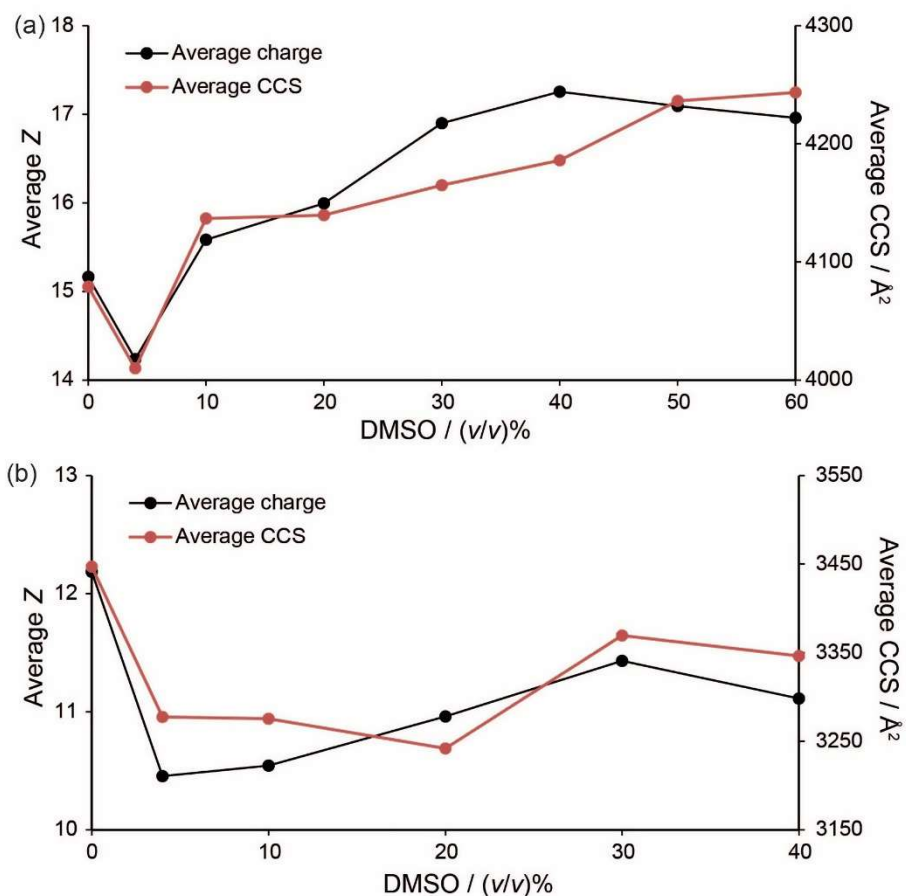


Figure 2.10 (a) Average charge and CCS of avidin as a function of DMSO concentration. (b) Average charge and CCS of CYP142A1 as a function of DMSO concentration.

2.2.2.2. Effect of high levels of DMSO on avidin and CYP142A1 size

IM-MS was performed to assess the effect of DMSO concentration on the size of avidin and CYP142A1 ions. The weighted-average CCS of CYP142A1 at 0% DMSO was within 2% (see Section 2.3.3.) of the theoretical CCS predicted from the X-ray crystal structure (PDB: 2XKR),^[117] suggesting that CYP142A1 retained its native structure in the gas phase. The abundance-weighted average CCS of avidin tetramers decreased from 4080 Å² without DMSO to 4010 Å² at 4% DMSO (**Figure 2.10a**), while for CYP142A1 the CCS decreased from 3450 Å² without DMSO to 3240 Å² at 20% DMSO (**Figure 2.10b**). This is consistent with the compaction of protein structure at low DMSO concentrations as shown previously by IM-MS (*vide infra*) or solution-phase techniques.^[105, 127]

Increasing the DMSO concentration beyond 4% led to an increase in average CCS for avidin, initially mirroring the trend for average charge (**Figure 2.10a**). However, the two trends deviated at higher DMSO concentrations, with CCS continuing to rise as the DMSO concentration was increased beyond 40%, even while average charge decreased. The maximal CCS of avidin tetramers (4240 Å²) was reached at 60% DMSO. This CCS was 4% larger than the average size of ions in the absence of DMSO (4080 Å²), and 6% larger than the compacted avidin ions at 4% DMSO (4010 Å²). For CYP142A1, the change in average CCS of CYP142A1 ions generally paralleled the trend for charge, except that the minimum CCS (3240 Å²) was reached at 20% DMSO, instead of at 4% DMSO for minimum charge (**Figure 2.10b**). Moreover, the average CCS at 40% DMSO (3350 Å²) was still 3% lower than that in the absence of DMSO, suggesting that the overall structure of CYP142A1 was still compacted even at the highest DMSO concentration used. Since electrospray droplets become enriched (estimated to be 3 to 5-fold^[105]) in DMSO as they shrink, this suggests that CYP142A1 may remain compacted at DMSO concentrations even higher than 40% during transfer from solution phase to gas phase. The reason for the remarkable gas-phase stability of CYP142A1 under such high DMSO concentrations is not known.

To investigate whether the change in CCS at higher DMSO concentrations could be related to the conformational changes of individual ions, the CCS of individual charge states were plotted

against DMSO concentration. For avidin, the CCS of individual avidin charge states decreased to a minimum in the presence of 4% DMSO, then became larger as the concentration of DMSO was increased (**Figure 2.11a**). In contrast, individual charge states of CYP142A1 were compacted by DMSO, with minimum CCS values being reached at 20% DMSO for all charge states (**Figure 2.11b**). These data therefore suggest that DMSO could induce expansion or compaction of avidin and CYP142A1 ions, respectively, even at the level of individual charge states. We propose that the increase in average CCS of avidin induced by high levels of DMSO (**Figure 2.10a**) could arise from a combination of two factors. Firstly, the supercharging effect of DMSO allows avidin ions to access higher charge states that are also larger due to increased Coulombic repulsion. Secondly, DMSO causes expansion in the size of individual charge states, which is presumably related to its denaturing ability, since DMSO is known to stabilize unfolded conformations of proteins via binding to hydrophobic and aromatic residues.^[131] Conversely, DMSO is proposed to reduce the average CCS of CYP142A1 (**Figure 2.10b**) both as consequence of its ability to shift the charge state distribution of CYP142A1 to lower charge, as well as its ability to compact protein structure at the level of individual charge states.

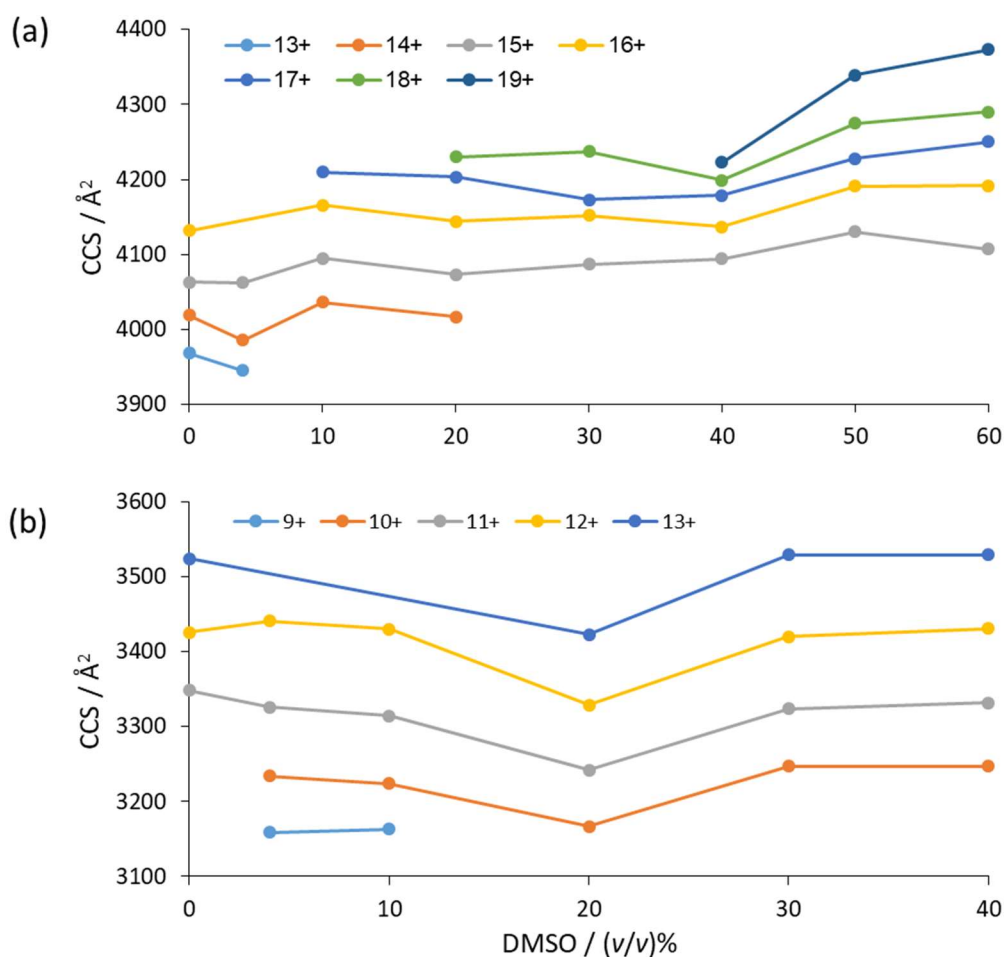


Figure 2.11 DMSO modulates the CCS of individual charge states in the gas phase. CCS of (a) avidin tetramers and (b) CYP142A1 as a function of DMSO concentration for each individual charge state.

2.2.2.3. DMSO modulates the CID stability of avidin

While several research groups have studied the effect of DMSO on the stability of non-covalent interactions by ESI-MS,^[108, 111] no systematic study had been performed on the effect of high concentrations of DMSO on the CID stability of protein structure and interactions in the gas phase. In this study, avidin tetramers underwent CID at higher trap collision voltages (CV) to generate highly-charged, unfolded monomers together with “stripped” trimers via the asymmetric charge partitioning mechanism (**Figure 2.12a**). As the CV was increased, the proportion of tetramer dissociation also increased. Moreover, higher charge states of avidin

became depleted first compared to lower charge states, which could be attributed to the greater Coulombic repulsion experienced by higher charge states, as well as the fact that at any given voltage, higher charge states will be accelerated to a greater energy than lower charge states.

When DMSO was added, the overall CID pathway did not change and similar distributions of monomers and trimers were produced. However, dissociation was facilitated by higher DMSO concentrations. For instance, in the absence of DMSO, 77% of the tetramer was preserved at a CV of 60 V, whereas with 60% DMSO, only 17% of the tetramer remained intact at the same CV (**Figure 2.12b**). Plotting the abundance of the tetramer as a function of CV (**Figure 2.12c**) reveals that DMSO concentrations higher than 4% promoted tetramer dissociation. The **Figure 2.12c inset** presents CID50 values (the CV needed to dissociate 50% of tetramers) as a function of DMSO concentration. The CID50 of avidin tetramers increased from 66.5 to 78.1 V (+11.6 V) as the DMSO concentration was raised from 0 to 4%, indicating a protective effect against dissociation and is consistent with the observed compaction. However, further increases in the DMSO concentration led to a decrease in CID50, with the minimum CID50 value of 53.6 V being reached at 50% DMSO, corresponding to a decrease of 12.9 V compared to in the absence of DMSO. Taken together, our data show that high concentrations of DMSO have a significant destabilizing effect on the CID of avidin. Only at 4% DMSO, where the average charge of the avidin tetramer was lowest, did DMSO have a protective effect against subunit dissociation.

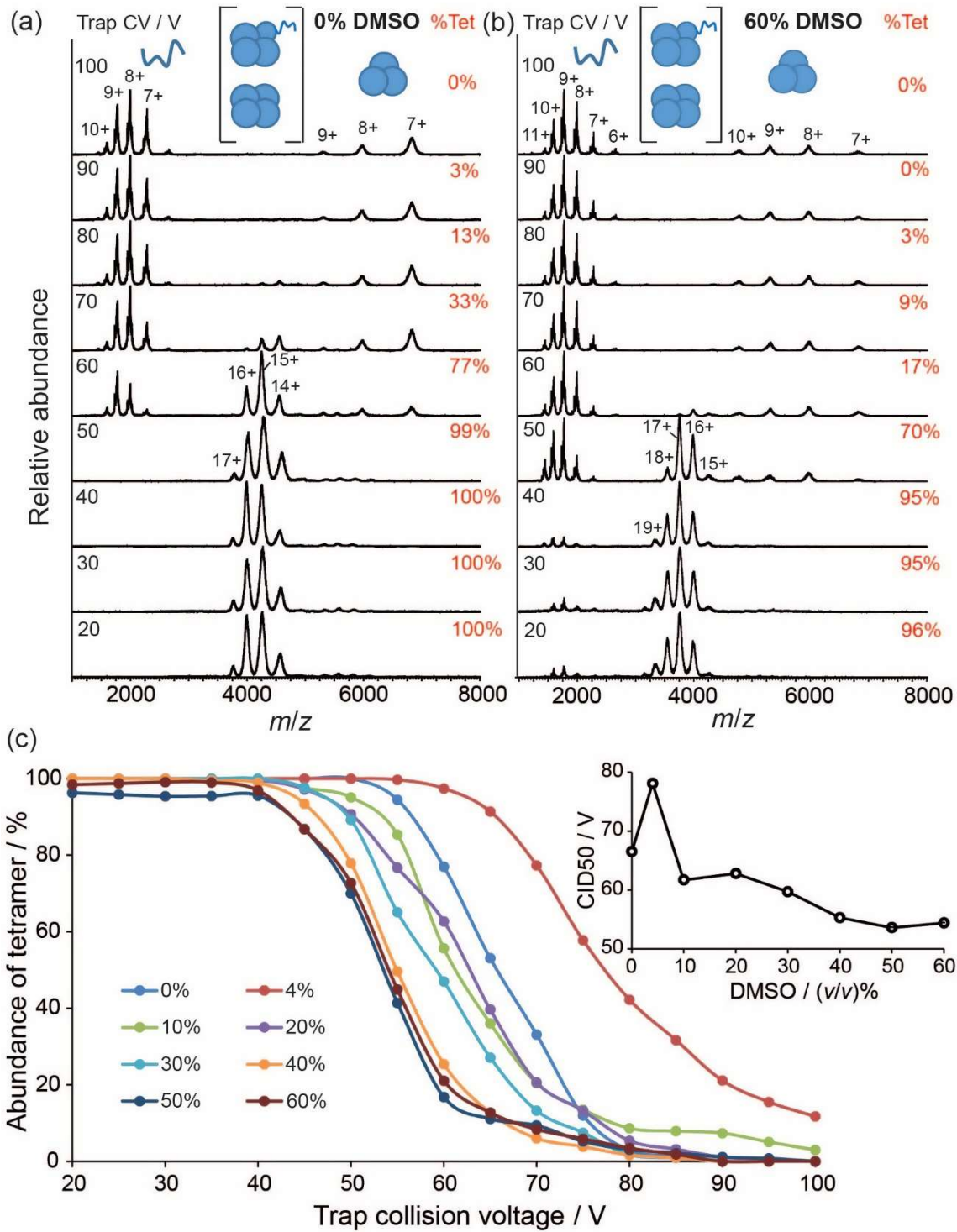


Figure 2.12 DMSO influences the CID stability of avidin. Native MS spectra of avidin in the presence of (a) 0% or (b) 60% DMSO at different trap CVs. The abundance of the tetramer (%Tet) at each CV is indicated.

(c) Abundance of tetramer as a function of trap CV at each DMSO concentration. The inset shows the CID50 as a function of DMSO concentration.

Generally, the CID50 values of avidin were inversely related to average protein charge. That is, DMSO concentrations that produced higher average charges of the avidin tetramer also gave lower overall CID50 values. This was expected as higher charge state ions possess more collision energy at any given CV, and are also inherently less stable due to increased Coulombic repulsion compared to lower charge state ions. Therefore, it was also of interest to determine whether the destabilization of avidin tetramers by DMSO was due solely to the ability of DMSO to shift the charge state distribution of avidin tetramers towards higher charge, or whether DMSO also exerted destabilizing effects on individual ions that are independent of its effect on charge state distribution. To address this, the CID50 values for the individual charge states were plotted as a function of DMSO concentration (**Figure 2.13a**). Interestingly, DMSO also promoted the dissociation of individual charge states. Note that when comparing between individual charge states, the concept of collision energy can be used which is calculated by multiplying the charge of the ion by the collision voltage. The collision energy more accurately reflects the energy available for dissociation and unfolding of a particular charge state compared to collision voltage. For example, the CID50 of the 16+ charge state was lowered from 1043 to 954 eV (−89 eV) when the DMSO concentration was raised from 0 to 60%, while for the 17+ charge state a reduction from 1020 to 923 eV (−97 eV) in CID50 was observed upon increasing DMSO concentration from 20 to 60%. This is in contrast to previously reported results from Landreh and co-workers, where 3% DMSO had no effect on the CID stability of the 16+ charge state of TTR.^[108] However, our study used relatively high concentrations of DMSO, compared to the 3% DMSO used for TTR. Indeed, the protective effect of 4% DMSO against dissociation could not be detected at the level of individual charge states in our study (**Figure 2.13a**). Taken together, the overall reduction in the CID stability of avidin at high DMSO concentrations could be attributed to two, possibly interrelated, mechanisms. First, DMSO shifts the charge state distribution of avidin tetramers towards ions of higher charge, which have lower intrinsic stability against dissociation. Second,

DMSO may also exert a destabilizing effect against individual charge states, presumably as a consequence of its denaturing effect.

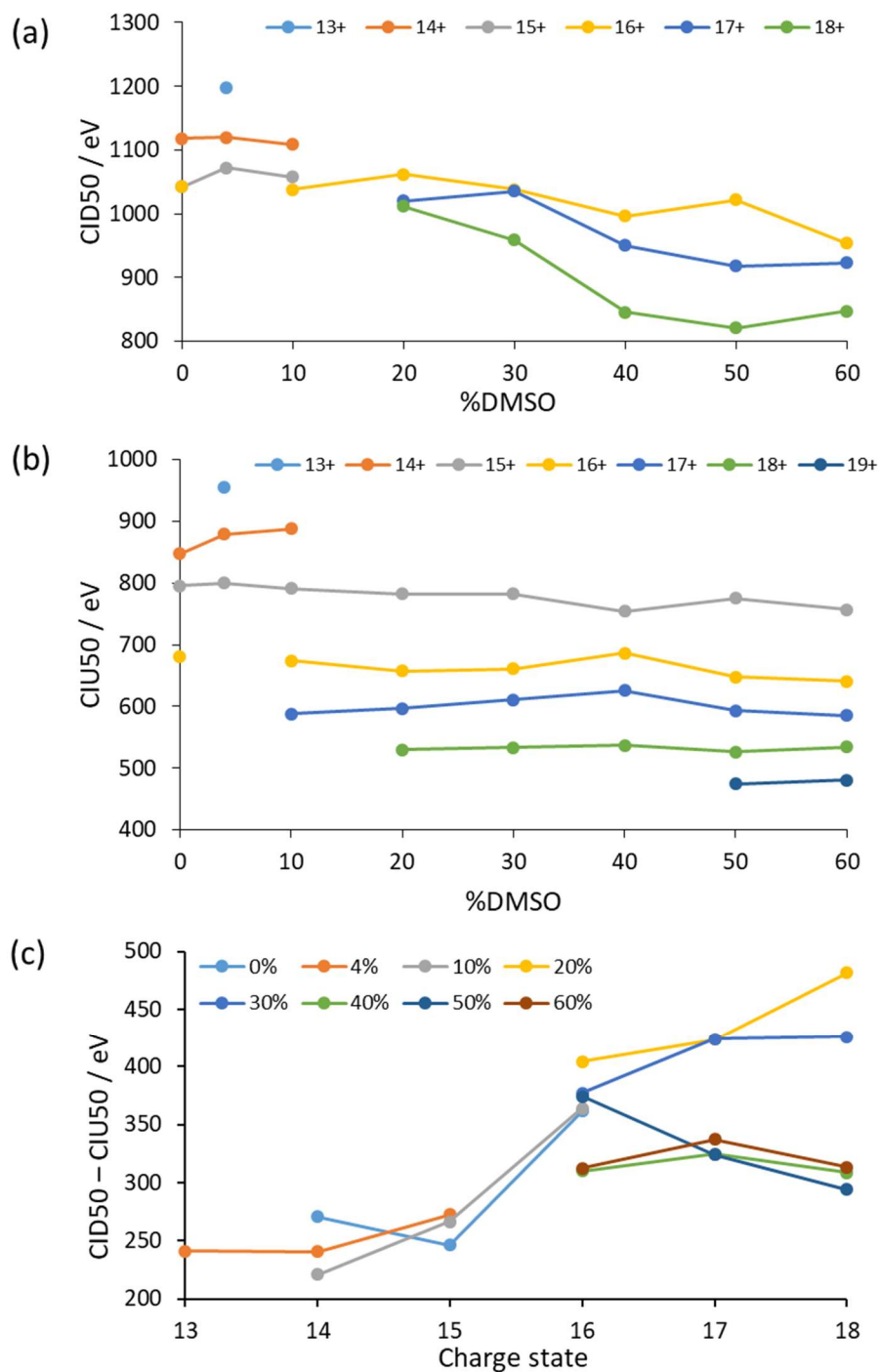


Figure 2.13 (a) CID50 and (b) CIU50 values of avidin tetramers as a function of DMSO concentration for individual charge states. (c) Collision energy difference (CID50 – CIU50) of avidin tetramers as a function of charge state at different DMSO concentrations. Note that (b) and (c) are discussed in Section 2.2.2.5.

2.2.2.4. DMSO influences the stability and heme dissociation pathway of CYP142A1

The influence of DMSO on the CID behavior of CYP142A1 was also investigated. **Figures 2.14a** and **2.14b** show CID spectra (without selection) of CYP142A1 at 0 and 30% DMSO, respectively. CYP142A1 holoenzyme underwent CID to release the apoenzyme and also the free heme-*b* ion at m/z 616 Da ($[M]^+$) together with an ammonia adduct of heme-*b* at m/z 633 Da ($[M+NH_3]^+$). The CYP142A1 peaks also generally became sharper at higher CV, consistent with the dissociation of buffer and/or DMSO molecules from the protein. Interestingly, DMSO had a protective effect against heme loss. For instance, only 20% of CYP142A1 ions retained heme at 140 V in the absence of DMSO (**Figure 2.14a**), whereas in the presence of 30% DMSO, the proportion of the holoenzyme was 37% (**Figure 2.14b**). Among all DMSO concentrations tested, heme dissociation occurred earliest at 0% DMSO (**Figure 2.14c**), with a CID50 value (the CV needed to dissociate 50% of holoenzyme) of 118 V (**Figure 2.14c inset**). The highest stability of the holoenzyme was observed at 4% DMSO (CID50 = 138 V), with stability decreasing as the DMSO concentration was increased further.

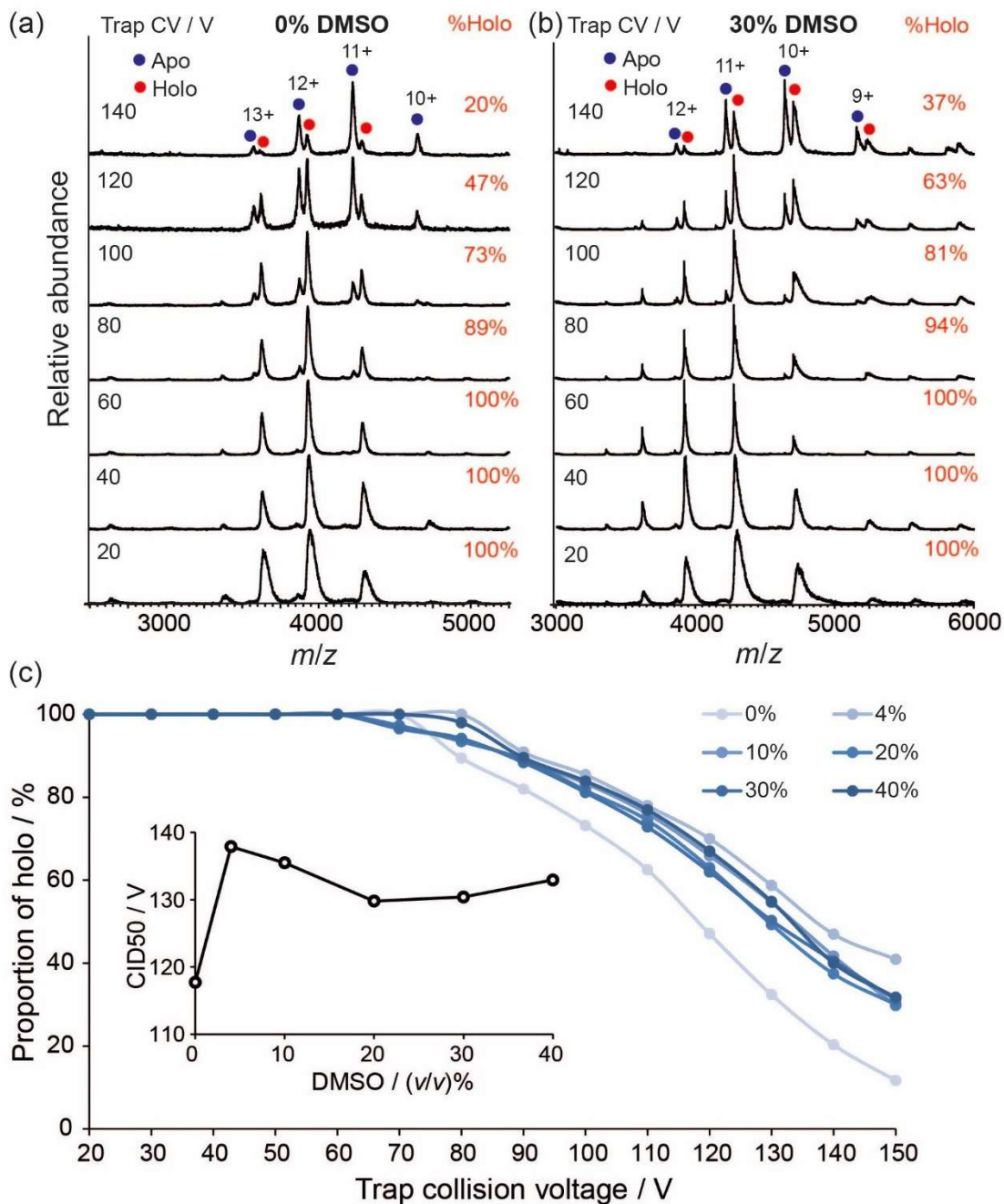


Figure 2.14 DMSO influences the CID stability of CYP142A1. Native MS spectra of CYP142A1 in the presence of a) 0% or b) 30% DMSO at different trap CVs. The proportion of holo-CYP142A1 (%Holo) at each CV is indicated. c) Proportion of holo-CYP142A1 as a function of trap CV at each DMSO concentration. The inset shows the CID50 as a function of DMSO concentration.

As with avidin dissociation, it was also of interest to explore whether DMSO could protect against dissociation of the heme group in a manner that is independent of its effects on charge state distribution. To study this issue further, tandem MS/MS (precursor ion selection) was performed to study the dissociation pathways of individual charge states. In the absence of DMSO, CYP142A1 ions exhibited two heme dissociation pathways: charge stripping and charge retention (**Figure 2.15a**). In the charge stripping pathway, the holo-CYP142A1 N^+ ion dissociates into the heme monocation and the $(N - 1)^+$ ion of the apoenzyme. This is the anticipated dissociation pathway for CYP142A1 because the protein is isolated in the ferric (3^+) oxidation state, and therefore would be expected to dissociate the $(\text{Fe(III)heme})^+$ ion.^[132] In contrast, in the charge retention pathway, the liberated apoenzyme has the same charge as the initial holoenzyme, which is presumed to occur via the dissociation of a neutral heme species, possibly a heme-acetate adduct.^[133] However, at 30% DMSO, the charge retention pathway is eliminated and only the charge stripping pathway is observed (**Figure 2.15b**). It was reasoned that the much higher concentration of DMSO (*ca.* 4.2 M) relative to acetate (200 mM) causes the anionic acetate to be displaced from heme in the initial solution through the formation of a cationic $(\text{heme-DMSO})^+$ adduct,^[134-135] thereby eliminating the charge retention pathway which requires the release of a neutral heme adduct.

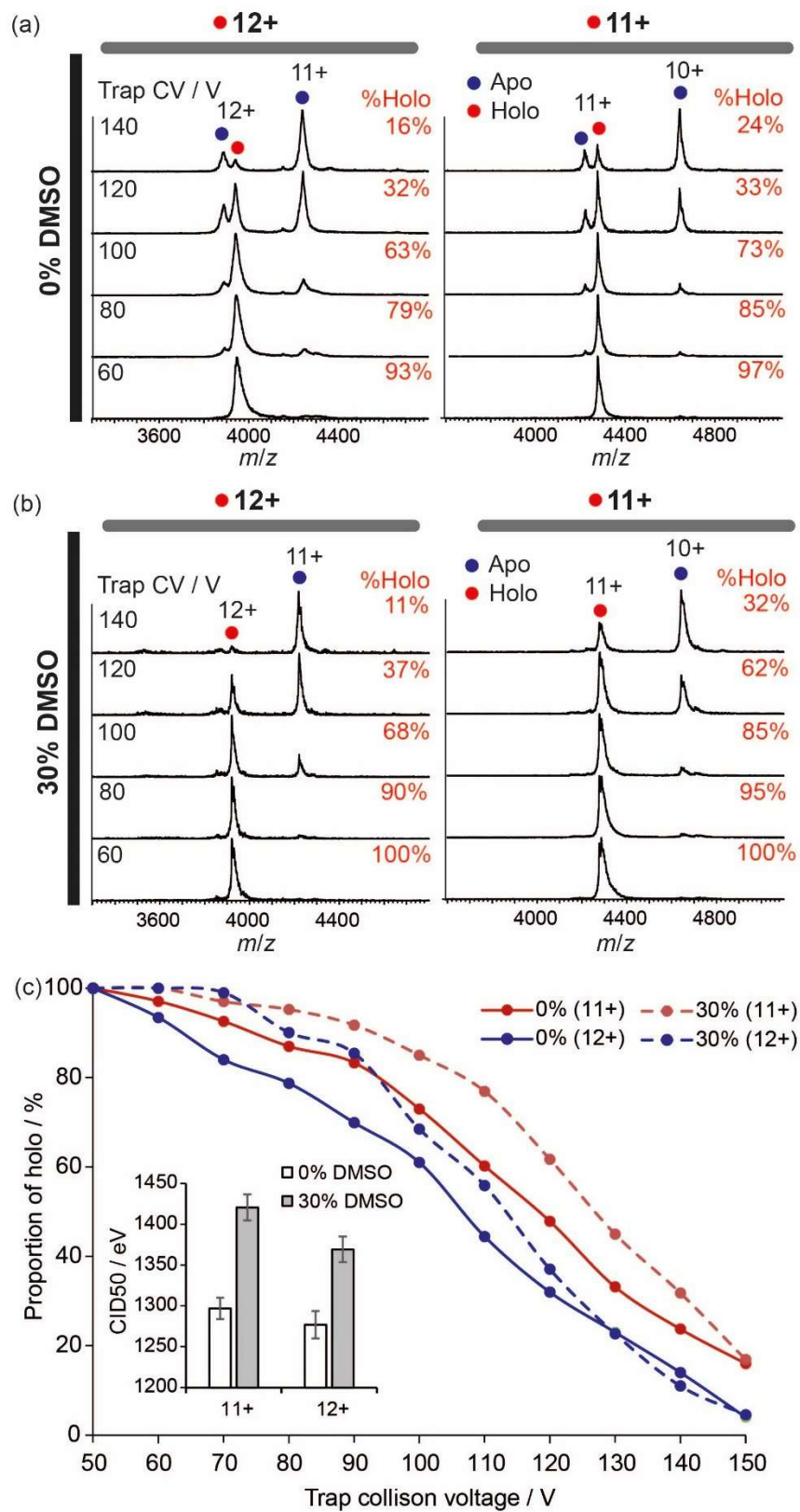


Figure 2.15 DMSO influences the dissociation pathway of CYP142A1. Tandem MS/MS of CYP142A1 at (a) 0% DMSO and (b) 30% DMSO with selection of the 12+ holoenzyme (left panel) or 11+ holoenzyme (right panel) as the precursor ion (indicated by a filled red circle). The trap CV and proportion of holo-CYP142A1 (%Holo) at each CV is indicated. (c) Proportion of holo-CYP142A1 as a function of trap CV at 0 or 30% DMSO for 11+ and 12+ charge states. The inset shows the CID50 for the 11+ and 12+ charge states at 0 and 30% DMSO.

In contrast, avidin tetramers exhibited only a minor amount of charge stripping before dissociation (**Figure 2.16**), as indicated by the weak intensity of the $(N - 1)^+$ when the N^+ ion was selected as the precursor ion.

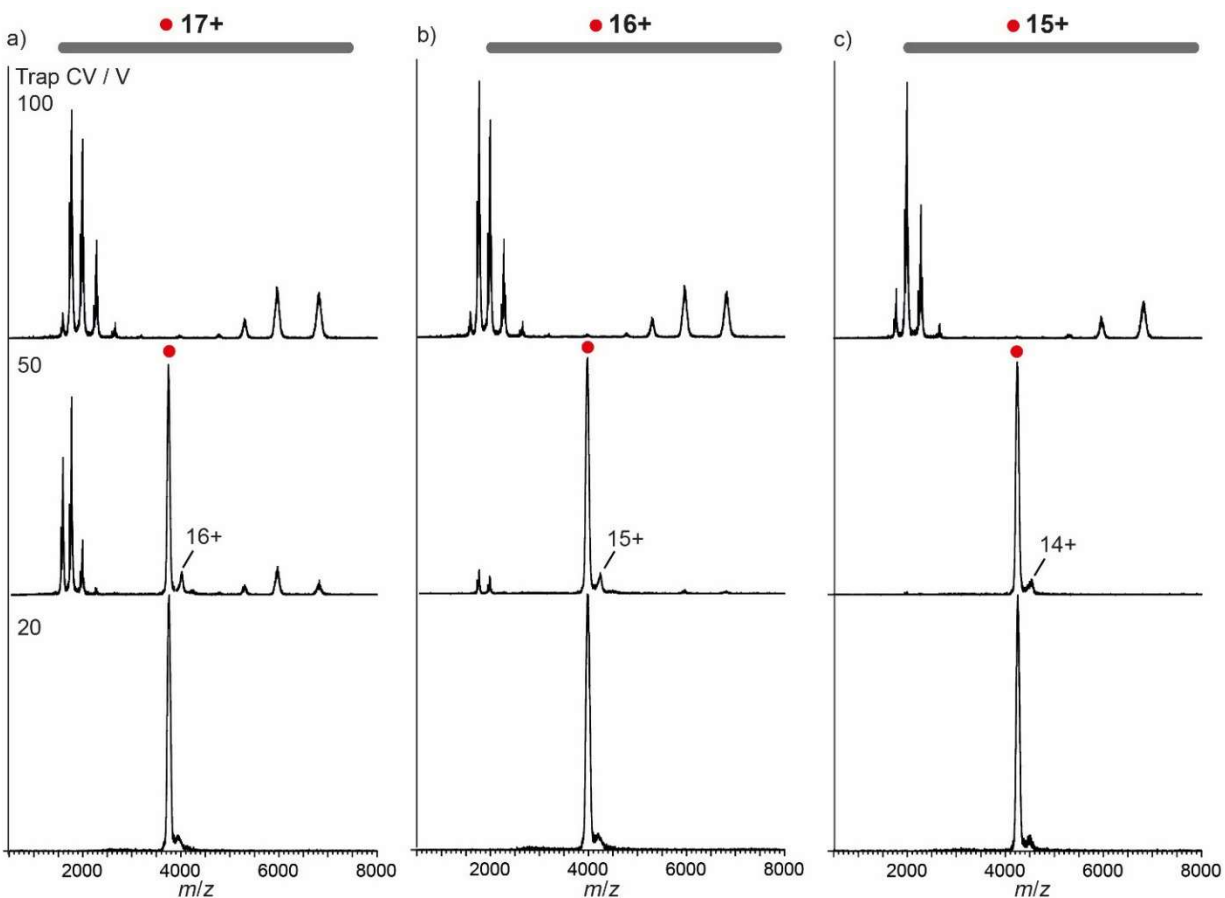


Figure 2.16. Tandem MS/MS of avidin (a) 17+, (b) 16+ and (c) 15+ tetramer charge states at 0% DMSO showing only minimal charge stripping. The precursor ion is indicated by a red circle.

The MS/MS data also confirmed that higher charge states of CYP142A1 dissociated heme more easily than lower charge states, as expected (**Figures 2.14c** and **2.17a**). Moreover, DMSO conferred a protective effect against heme-*b* loss even at the level of individual charge states. For instance, the CID50 values of the 11+ and 12+ charge states increased from 1297 ± 13 eV to 1421 ± 18 V (+124 eV) and from 1277 ± 17 eV to 1354 ± 16 eV (+77 eV) respectively as the DMSO concentration was raised from 0 to 30% (**Figure 2.14c inset**). Taken together, these data suggest that DMSO protects against heme loss both by shifting the charge state distribution of CYP142A1 towards ions of lower charge, which have higher intrinsic stability to dissociation, as well as by enhancing the stability of individual charge states. The latter effect could be linked to the compaction of CYP142A1 ions by higher DMSO concentrations, that is, more compact holoenzymes have a greater resistance against heme loss.

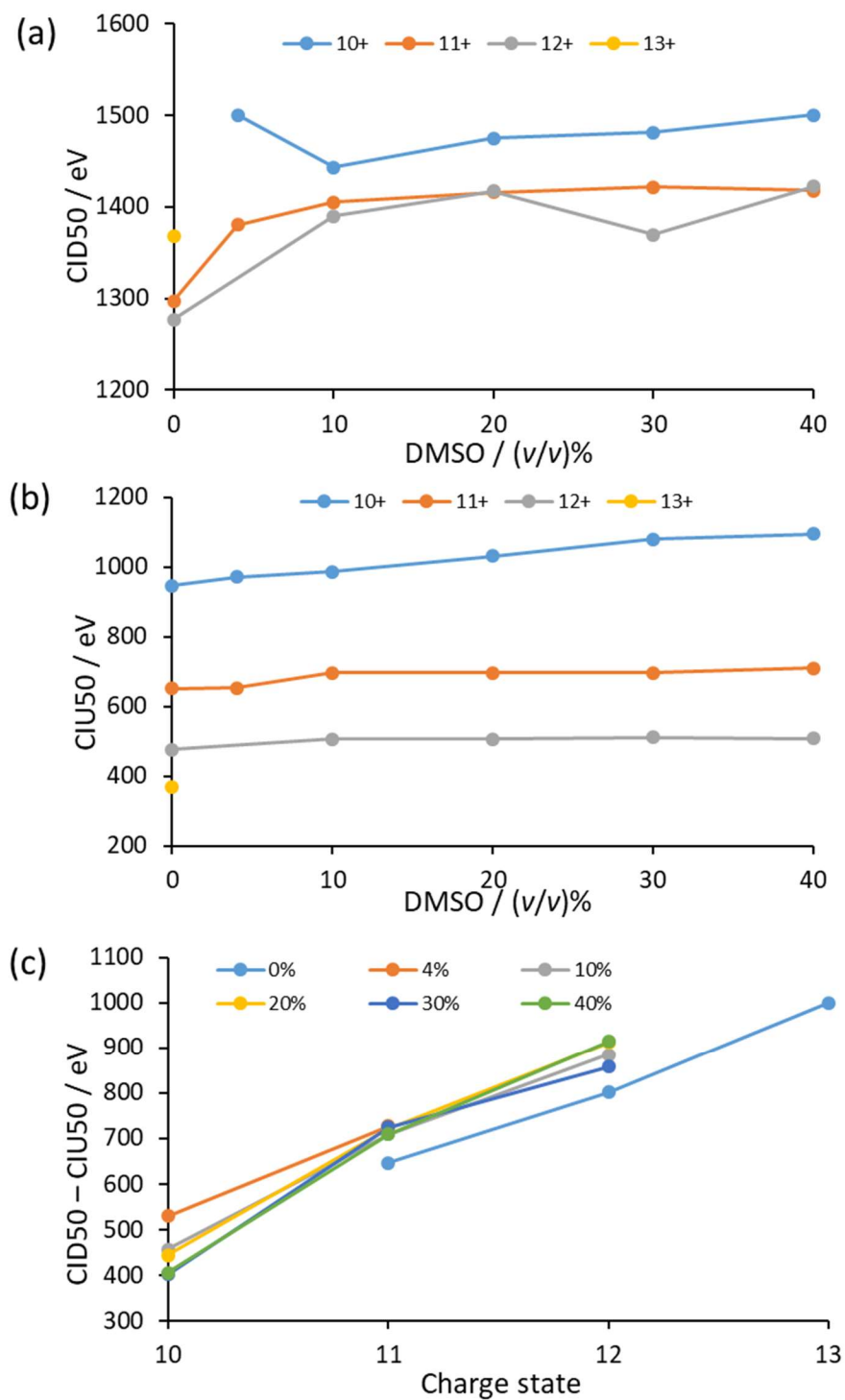


Figure 2.17 (a) CID50 and (b) CIU50 values of the CYP142A1 holoenzyme as a function of DMSO concentration for individual charge states. (c) Collision energy difference (CID50 – CIU50) of the CYP142A1

holoenzyme as a function of charge state at different DMSO concentrations. Note that (b) and (c) are discussed in Section 2.2.2.5.

2.2.2.5. DMSO modulates the CIU stability of proteins

A number of studies have investigated the effect of various solution additives on the CIU stability of different proteins. The Ruotolo group have shown that both anions^[112] and cations^[113] can have dramatic effects on both CIU and CID stabilities. It was postulated that anions stabilize proteins through dissociative cooling-type mechanisms, whereas cations increase stability by forming multidentate interactions within proteins to tether together non-contiguous protein regions.^[114] Robinson, Ruotolo and co-workers have found that TrisH⁺ also increased CIU stability, which was attributed to an evaporative cooling mechanism or alternatively to the binding of TrisH⁺ to critical regions of the protein so as to replace solvent contacts that are needed for maintaining native-like conformations.^[115] However, to our knowledge, the effect of DMSO on the CIU stability of proteins has not yet been reported in the literature.

Figure 2.18a shows CIU fingerprints for the 14+ to 17+ charge states of the avidin tetramer in the absence of DMSO. At least three less compact conformations in addition to the native conformation were observed at 0% DMSO for the 15+ charge state of avidin, which was similar to previously reported.^[114-115] Higher charge states of the avidin tetramer unfolded at lower CV, which can again be attributed to increased Coulombic destabilization. Unfolding pathways for the avidin tetramer were broadly similar in the presence of 50% DMSO (**Figure 2.18b**). However, lower voltages were needed to unfold avidin ions of a given charge state at 50% DMSO compared to in the absence of DMSO. For instance, the CIU50 (energy needed to deplete 50% of the most compact conformation) value of the 15+ ion of the avidin tetramer decreased from 786 ± 11 eV in the absence of DMSO to 744 ± 17 eV (-42 eV) with 50% DMSO (**Figures 2.18c** and **2.13b**). However, the decreases in CIU50 induced by DMSO are comparatively smaller than the reductions of CID50 described previously (shown for representative charge states in **Table 2.3**), indicating that DMSO has a greater destabilizing effect on tetramer dissociation than unfolding.

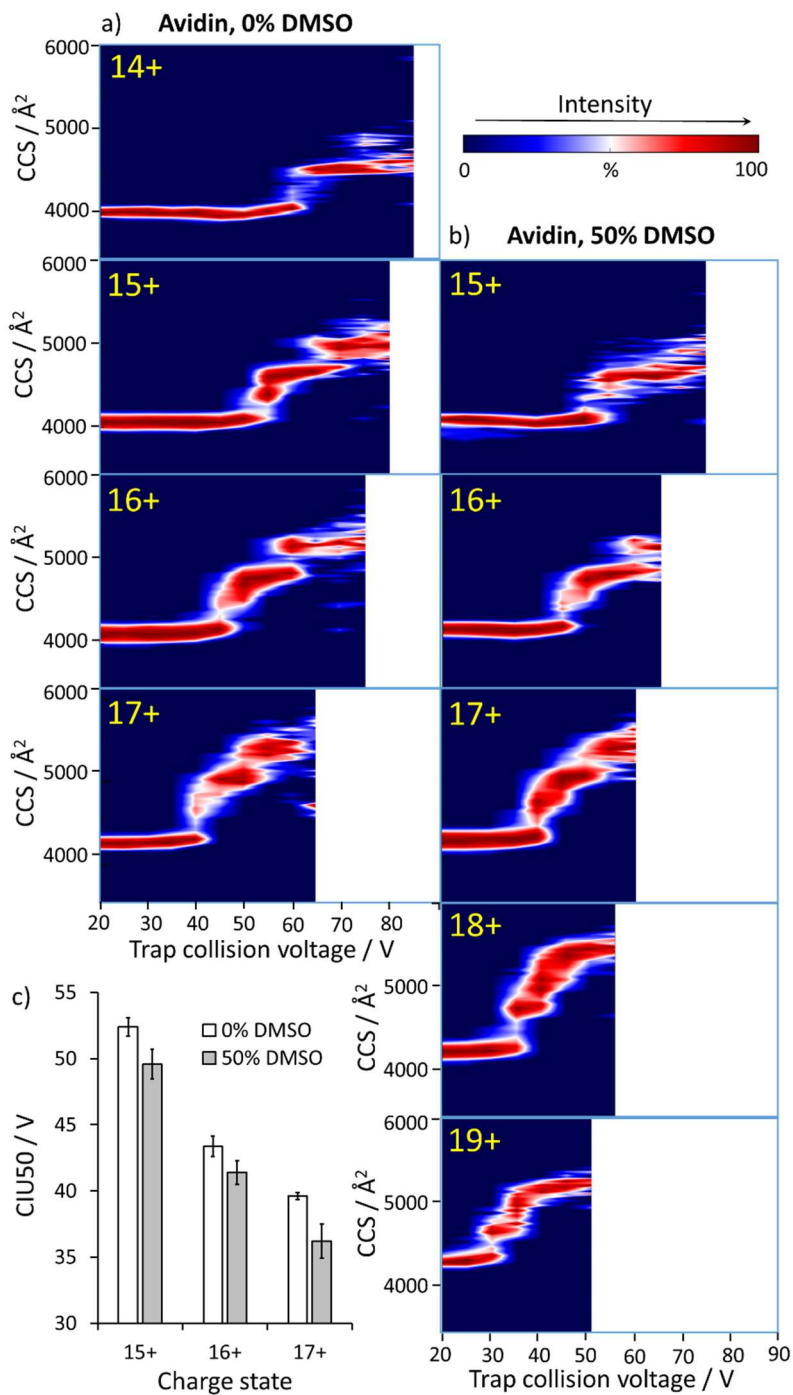


Figure 2.18 DMSO decreases the CIU stability of avidin, as indicated by an increased ease of unfolding as collision energy is increased. CIU fingerprints of avidin tetramer charge states in the presence of (a) 0% or (b) 50% DMSO. (c) CIU50 values for the 15+, 16+ and 17+ charge states at 0 and 50% DMSO. Error bars represent the standard deviation of duplicate results.

Table 2.3. CID50 and CIU50 values of selected avidin and CYP142A1 charge states. Errors represent the standard deviation of duplicate results.

Avidin 16+			
[DMSO] / %	CID50 / eV	CIU50 / eV	CID50 – CIU50 / eV
0	1043 ± 12	682 ± 17	361 ± 29
10	1038 ± 17	675 ± 14	363 ± 31
20	1062 ± 28	658 ± 15	404 ± 40
30	1038 ± 20	661 ± 8	377 ± 28
40	997 ± 22	687 ± 20	290 ± 42
50	1022 ± 15	648 ± 14	372 ± 29
60	954 ± 17	641 ± 13	313 ± 31
CYP142A1 11+			
[DMSO] / %	CID50 / eV	CIU50 / eV	CID50 – CIU50 / eV
0	1297 ± 22	650 ± 15	647 ± 37
4	1381 ± 15	654 ± 11	727 ± 26
10	1405 ± 18	697 ± 17	708 ± 35
20	1416 ± 27	694 ± 15	722 ± 42
30	1421 ± 19	696 ± 9	725 ± 28
40	1418 ± 18	709 ± 14	709 ± 32

The collision energy differences between CID50 and CIU50 (CID50 – CIU50) were invariably positive (**Figure 2.13c**), indicating that all charge states of the avidin tetramer undergo unfolding

before dissociation, consistent with the asymmetric protein dissociation hypothesis. The magnitudes of these differences were comparable to a previous CID/CIU study on avidin by Ruotolo and co-workers.^[136] Moreover, at DMSO concentrations between 0 and 30%, the collision energy difference generally increased with charge state, similar to previous observations.^[136] As both CID and CIU values decrease with charge, this suggests that Coulombic repulsion has a larger destabilizing effect on unfolding compared to dissociation (*i.e.*, unfolding is more charge-driven compared to dissociation). However, the collision energy differences were generally reduced at high DMSO concentrations ($\geq 40\%$). For example, the collision energy difference for the 16+ charge state of avidin was decreased from 404 ± 40 eV at 20% DMSO to 313 ± 31 eV at 60% DMSO. This observation is consistent with the greater ability of DMSO to promote dissociation compared to unfolding of individual charge states, as noted above.

CIU fingerprints for the 10+ to 13+ charge states of the CYP142A1 holoenzyme at 0 and 30% DMSO are shown in **Figures 2.19a** and **2.19b**, respectively. As with avidin, CYP142A1 holoenzymes transitioned through progressively more extended conformers as the CV was increased, with the exception of the 10+ ion which exhibited only two conformational states (compact and extended) over the range of CV tested. Unlike with avidin, however, DMSO increased the stability of CYP142A1 ions against unfolding. For instance, the CIU50 value of the 11+ holoenzyme increased from 650 ± 15 eV at 0% DMSO to 696 ± 9 eV (+46 eV) at 30% DMSO (**Figures 2.19c** and **2.17b**). Thus, and in similar fashion to its effect on CID stability, it is proposed that DMSO increases the stability of CYP142A1 holoenzymes against unfolding both by reducing the average charge of the ions, as well as by increasing the CIU stability at the level of individual charge states. However, as for avidin the increases in CIU50 induced by DMSO are relatively small compared to the increases of CID50 described above (**Table 2.3**). **Figure 2.17c** plots the collision energy differences between CID50 and CIU50 for different charge states of the CYP142A1 holoenzyme. As with avidin, CYP142A1 holoenzymes generally unfold before dissociating, and the collision energy differences at lower charge states were also reduced. This is consistent with the observation that DMSO has a greater protective effect on CID compared to CIU, as noted above.

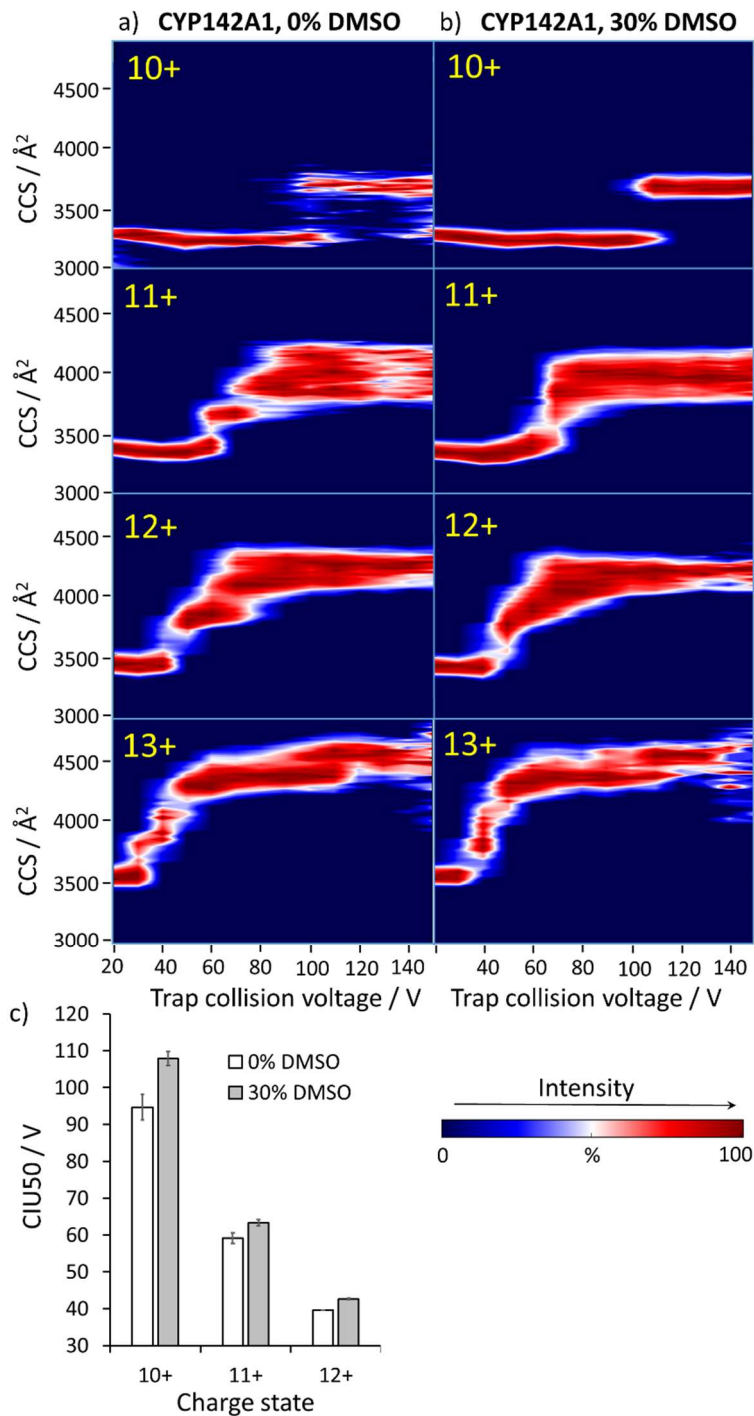


Figure 2.19 DMSO increases the CIU stability of CYP142A1, as indicated by an increased resistance to unfolding as collision energy is increased. CIU fingerprints of CYP142A1 holoenzyme charge states in the presence of (a) 0% or (b) 30% DMSO. (c) CIU50 values for the 10+, 11+ and 12+ charge states at 0 and 30% DMSO. Error bars represent the standard deviation of duplicate results.

The unfolding behavior of the ejected CYP142A1 apoenzyme was also tracked. CIU fingerprints of the apoenzymes were similar in the absence (**Figure 2.20a**) or presence of 30% DMSO (**Figure 2.20b**). At 30% DMSO, the 9+ apoenzyme was ejected in a compact state (CCS = 3060 Å²) that did not unfold even at the highest CV tested. The ejected 10+ apoenzyme was also initially compact (CCS = 3180 Å²), but transitioned to an extended conformation (CCS = 3620 Å²) at voltages beyond 90 V. These CCS values are similar to those for the compact and extended states of the 10+ holoenzyme at the same CV (CCS = 3200 Å² and 3640 Å² respectively), indicating that the loss of the heme group did not have a major effect on the size of either the compact or extended conformations of CYP142A1. The 11+ to 13+ charge states of the apoenzyme were ejected in extended states.

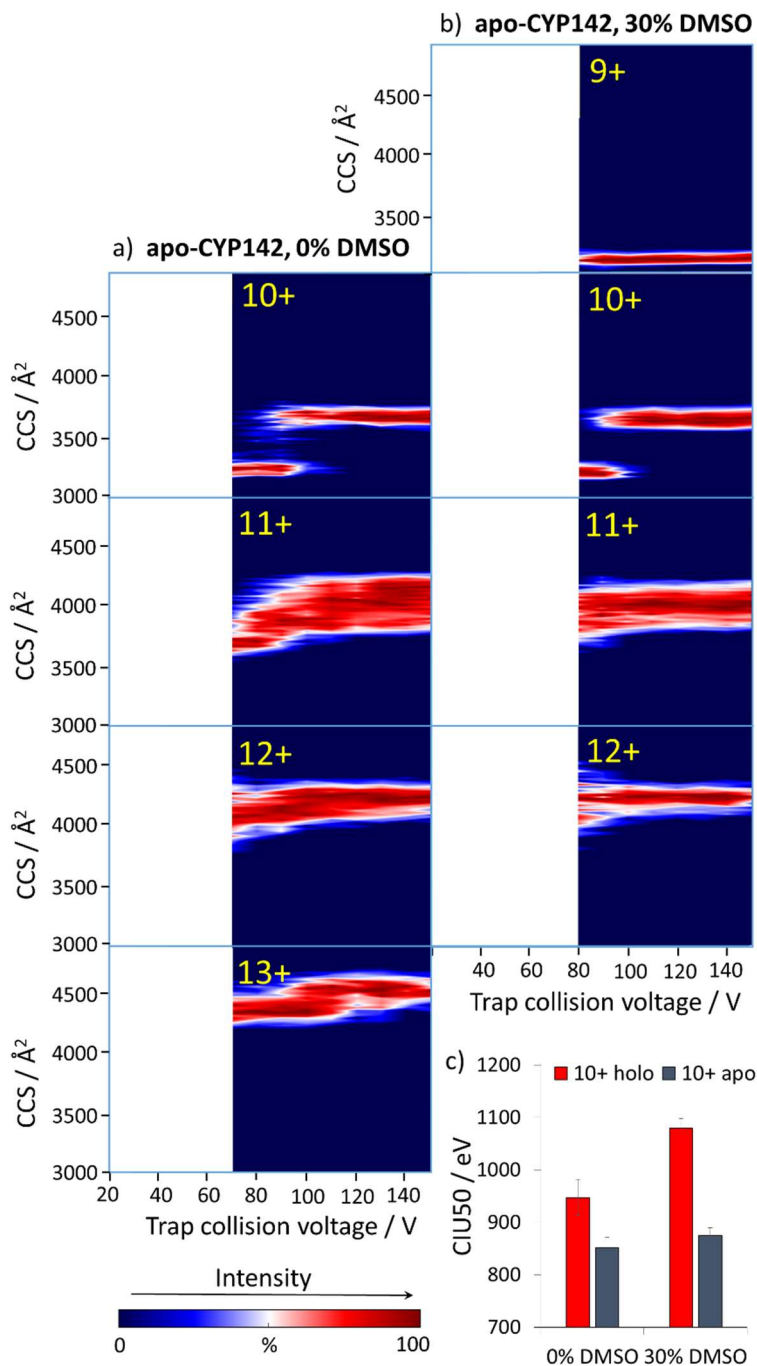


Figure 2.20 CIU fingerprints of CYP142A1 apoenzyme charge states in the presence of (a) 0% or (b) 30% DMSO. (c) CIU50 values for the 10+ holoenzyme and apoenzyme at 0% and 30% DMSO. Error bars represent the standard deviation of duplicate results.

2.2.2.6. Monitoring the protective effect of the heme group using CIU

The heme prosthetic group is known to stabilize the structure of heme-containing proteins, including cytochrome P450s.^[137] Using ultraviolet photodissociation (UVPD) mass spectrometry, Brodbelt and Cammarata showed that the cleavage of holo-myoglobin was suppressed in regions that interact with heme, indicating that the heme had a protective effect against fragmentation.^[138] Williams and co-workers found that the heme group protected against the DMSO-induced unfolding of myoglobin by solution-phase HDX-MS experiments.^[105] Intriguingly, the stabilizing effect of the heme group on the unfolding stability of CYP142A1 could also be monitored by CIU. For instance, at 0% DMSO, the CIU50 value for the 10+ holoenzyme was 947 ± 3 eV, but this decreased to 852 ± 2 eV (-95 eV) for the 10+ apoenzyme (**Figure 2.20c**). Similarly, at 30% DMSO the CIU50 value for the 10+ holoenzyme (1079 ± 20 eV) was 204 eV higher than for the 10+ apoenzyme (875 ± 14 eV). These results indicate that the heme group substantially stabilizes the CYP142A1 holo-enzyme in the gas phase. However, while high DMSO concentrations protected against the unfolding of the holoenzyme, it had little effect on the unfolding stability of the corresponding apoenzyme (**Figure 2.21**).

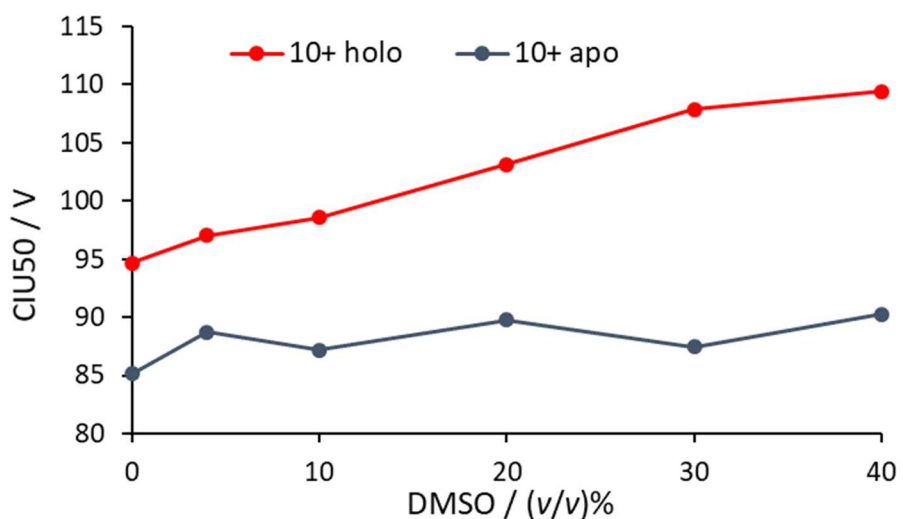


Figure 2.21 CIU50 values of the 10+ charge state of the CYP142A1 holoenzyme and apoenzyme as a function of DMSO concentration.

2.2.2.7. Summary

Native nESI-IM-MS has been used to study the effect of DMSO on the dissociative and unfolding stability of avidin and *Mtb* CYP142A1. To our knowledge, this is the first systematic study of the effect of DMSO on protein CID and CIU in the literature. There are a number of general conclusions from our work.

Evidence has been provided that suggests a dual mechanism for the ability of DMSO to modulate the CID and CIU stability of proteins. DMSO not only shifts the charge state distributions of proteins to higher or lower charge states, which vary in intrinsic stability due to differences in electrostatic repulsion, but also modulates stability via effects at the level of individual charge states.

The data also show that the effects of DMSO on protein structure and interactions are highly protein-dependent, which has implications for the design of biological assays where DMSO is frequently employed as a co-solvent. While DMSO concentrations higher than 4% facilitated the dissociation and unfolding of the avidin tetramer, CYP142A1 exhibited remarkable stability in the presence of up to 40% DMSO, which protected the enzyme from both heme loss and unfolding. To our knowledge, CYP142A1 is the first protein that has been reported to remain charge-reduced at such high DMSO concentrations in native MS. However, compared to previous work with ions,^[112-114] the stabilizing effects of DMSO on CYP142A1 are comparatively weaker. This could be due to the fact that as DMSO is neutral, it interacts relatively weakly with protein, and hence less energy is removed when DMSO adducts dissociate from the protein.

Additionally, it was established using tandem MS/MS experiments that DMSO could modify the heme dissociation pathway of CYP142A1. In the absence of DMSO, CYP142A1 can dissociate either the positive (Fe(III)heme)⁺ ion or the neutral (Fe(III)heme)-acetate adduct, resulting in charge stripping and charge retention, respectively. However, DMSO is hypothesized to displace acetate, such that only the charge stripping pathway is observed in the presence of DMSO. Finally, we have demonstrated that the protective effects of the heme group on the gas-phase stability

of CYP142A1 can be assessed using CIU. It is envisaged that this approach could be broadly applied to study the stabilizing effect of heme or other prosthetic groups on other proteins in the gas phase.

2.3. Experimental

2.3.1. Materials and methods

Avidin (egg white), concanavalin A (*Canavalia ensiformis*), alcohol dehydrogenase (*Saccharomyces cerevisiae*) and pyruvate kinase (rabbit heart) were obtained from Sigma-Aldrich (UK). *Mtb* His₆-tagged CYP142A1 was provided by our collaborator Kirsty McLean in the Munro group.^[117] Protein samples were exchanged into 200 mM ammonium acetate (pH 7.0) solution using Micro Bio-Spin 6 chromatography columns (Bio-Rad, UK) and diluted to a final concentration of 20 μM before analysis.

2.3.2. Mass spectrometry

Nano electrospray ionization-ion mobility-mass spectrometry (nESI-IM-MS) was performed on a Synapt HDMS mass spectrometer (Waters, UK) modified for studying high masses (32k quadrupole) and equipped with a traveling-wave (TW) IM device. 2.5 μL of protein solution was injected into a gold-coated borosilicate emitter (Thermo Scientific, UK) for sampling. Typical conditions were: capillary voltage 1.5 kV, cone voltage 50 V, trap collision voltage 20 V, trap DC bias 35 V, transfer collision voltage 12 V, source temperature 20 °C, backing pressure 3–4 mbar, trap pressure 3–4 × 10⁻² mbar, IM (N₂) pressure 5–6 × 10⁻¹ mbar and TOF pressure 7–8 × 10⁻⁷ mbar.

IM-MS spectra were acquired as above except that the instrument was run in mobility TOF mode. The traveling wave velocity was 250 ms⁻¹ and four wave heights (7, 8, 9, and 10 V) were used. These wave heights were selected to ensure that the ions arrived within the time course of these IM experiments. The reported CCS values were an average of the data recorded over all of the wave heights.

To promote CID and CIU, the trap collision voltage was raised in 5 V increments from 20 to 100 V for avidin and in 10 V increments from 20 to 150 V for CYP142A1. To record CIU, the instrument was operated in mobility TOF mode using an IM wave velocity of 250 ms^{-1} and a wave height of 10 V without precursor ion selection. Avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase were used as calibrant ions for CCS determination.

2.3.3. Data processing

Mass spectra were calibrated externally using a solution of cesium iodide (100 mg mL^{-1}). Data acquisition and processing were performed using MassLynx 4.1 (Waters) and DriftScope 2.5 (Waters). CCS calibration was performed as previously described.^[76]

For CID analysis, the measured peak heights were used and all charge states were taken into account. CID50 values were calculated from non-linear curve fitting using Origin 9.1 (OriginLab). Unless otherwise stated, all CCS values are quoted as N_2 values. CID50 and CIU50 values are quoted as mean \pm standard deviation of at least two independent experiments. For the CIU experiments, PULSAR^[53] was used to calibrate CCS values, visualize CIU fingerprints and calculate CIU50 values. Tim Allison (University of Oxford) is acknowledged for assistance with PULSAR.

As algorithms for CCS calculation from X-ray crystal structures are parameterized for helium collision gas,^[139] the corresponding CCS(He) of CYP142A1 was also determined by calibration of the drift times with helium CCS values in order to enable comparison between the experimental result and the theoretical CCS. The experimental CCS(He) of CYP142A1 was 3140 \AA^2 , which is within 2% of the theoretical CCS(He)_{calc} value of 3090 \AA^2 , where $\text{CCS(He)}_{\text{calc}} = 1.14 \times \text{CCS(He)}_{\text{PA}} \times (M_{\text{exp}}/M_{\text{PDB}})^{2/3}$ and $\text{CCS(He)}_{\text{PA}}$ is the CCS calculated from the X-ray crystal structure of CYP142A1 (PDB: 2XKR)^[117] using the projection approximation (PA) algorithm implemented in DriftScope, M_{exp} is the mass of His₆-tagged CYP142A1 (47178 Da) used in the experiment and M_{PDB} is the mass of the CYP142A1 construct (45015 Da) in the reported crystal structure. The coefficient of 1.14 in the above calculation is an empirically determined scaling factor that takes into account

the underestimation of CCS by the PA algorithm due to its masking of the cavity slow-down effect.^[61, 139]

3. Structural insights and fragment screening against the EthR-DNA interaction using native mass spectrometry

3.1. Introduction

Tuberculosis (TB) is a contagious disease caused by *Mycobacterium tuberculosis* (*Mtb*) that exerts an enormous burden on human health and wellbeing worldwide. The World Health Organization (WHO) has estimated that in 2017, TB killed 1.3 million people, while another 10 million people became infected.^[140] Moreover, around 23% of the world population, or about 1.7 billion people, are estimated to be latently infected.^[140] Progress against TB has been challenged by the rise of multidrug resistant (MDR) and extensively drug-resistant (XDR) *Mtb* strains. 3.5% of new cases and 18% of previously treated cases have MDR-TB, and of those, an estimated 8.5% are XDR-TB.^[140]

Ethionamide is a second-line drug used for the treatment of MDR-TB. Mechanistically, ethionamide is a prodrug that is activated *in vivo* by EthA, a flavin-containing monooxygenase enzyme in *Mtb*, to form an ethionamide-NAD adduct.^[141] This adduct inhibits the 2-*trans*-enoyl reductase enzyme InhA, which in turn leads to the inhibition of the *Mtb* type II fatty acid synthase system (FAS II).^[142] However, the potency of ethionamide is reduced by EthR, which is a transcriptional repressor of *ethA* expression.^[143] This suggests that inhibitors of EthR activity could function as ethionamide boosters,^[144] allowing for lower dosages of the drug to be used. A number of EthR inhibitors have shown ethionamide potentiating activity in cellular or even *in vivo* models of TB infection.^[144-147]

EthR belongs to the TetR/CamR repressor protein family, whose members show high sequence homology between their N-terminal DNA-binding domains, and is expected to bind to the DNA major groove via its helix-turn-helix (HTH) motif.^[148] Using DNase footprinting assays, Baulard and co-workers showed that EthR recognizes a 55-bp operator sequence within the *ethA*-R intergenic region.^[143] Subsequent surface plasmon resonance (SPR) analysis suggested that up to eight units of EthR could bind cooperatively to a 62-bp sequence (DNA₆₂) encompassing the

operator site (**Figure 3.1**).^[143] There are numerous X-ray crystal structures of the EthR dimer in complex with various small-molecule ligands,^[149-150] but the structure of the EthR-DNA complex has not yet been solved.

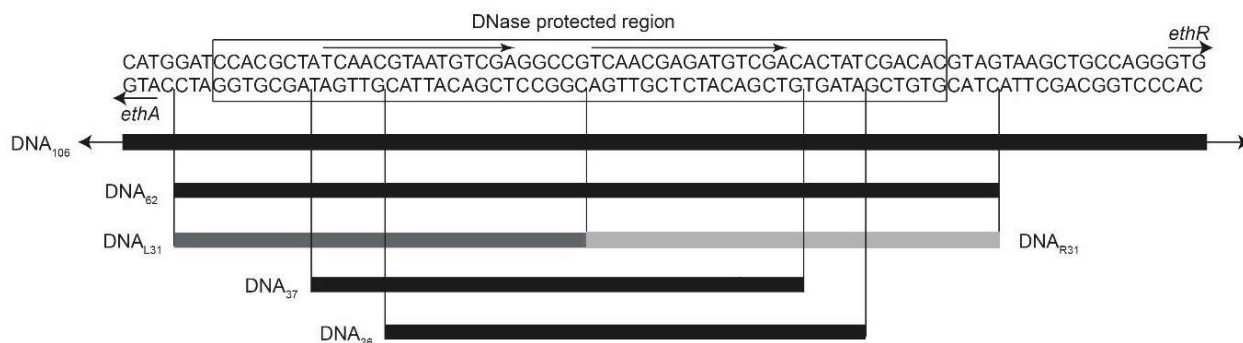


Figure 3.1 Sequence of the *ethA-R* intergenic region. The positions of the DNA sequences used in this work are shown. The two arrows in the DNase protected region indicate direct repeats.

Native ESI-MS is an ideal technique for the study of the interactions and stoichiometries of macromolecular complexes.^[40, 151] A large body of work has established that the native structure and composition of biomolecular complexes can be adequately maintained as solution species are transferred into the gas phase during the electrospray process. However, compared to the application of ESI-MS for multi-protein complexes, fewer studies on protein-DNA complexes have been reported.^[152-155] Analysis of protein-DNA complexes containing large DNA sequences using positive-ion native MS is complicated by the heterogeneity of cation adduction and as well as difficulties with achieving a stable electrospray.^[153, 155]

Meanwhile, fragment-based approaches have been applied to develop EthR inhibitors. Willand and co-workers started from a 4-iodo-*N*-prop-2-ynylbenzenesulfonamide fragment that was initially identified using an *in situ* click chemistry approach.^[156] Subsequent fragment elaboration furnished analogues that disrupted the EthR-DNA interaction and boosted ethionamide activity in *Mtb*-infected macrophages with sub-micromolar potency.^[157] Our group has previously employed DSF (also known as the thermal shift assay) to screen a 1250-member fragment library against EthR, with fragments showing $\Delta T_m \geq 1^\circ \text{C}$ at 10 mM concentration being classed as a

hit.^[158] Fragment merging and linking strategies subsequently generated improved analogues with low-micromolar IC₅₀ values against the EthR-DNA interaction.^[158-159] While the use of nanoelectrospray ionization-mass spectrometry (nESI-MS) to study biomolecular assemblies has been well-documented,^[40, 151] native MS is a relatively underutilized technique especially in fragment screening for drug discovery.^[44-45, 63, 67, 160]

In this chapter, native MS was used to provide structural insights into the EthR-DNA interaction. Our results indicate that up to six subunits of EthR are able to bind to its operator, instead of the eight subunits that had been previously reported. Moreover, the first demonstration is provided of the use of native MS for screening fragments against a protein-DNA interaction. Hits were validated using surface plasmon resonance and X-ray crystallography, resulting in the identification of two new fragments that disrupt the EthR-DNA interaction *in vitro* and that bind to the hydrophobic channel of the EthR dimer.

3.2. Results and discussion

3.2.1. Structural insights into the EthR-DNA interaction

3.2.1.1. Optimization of MS conditions for detecting EthR-DNA complexes

Histidine-tagged EthR was expressed in this project in excellent yield and purity following a literature procedure.^[158] Mass spectra were obtained by nano-electrospray ionization (nESI) from a hybrid quadrupole time-of-flight (qTOF) SYNAPT HDMS (Waters) instrument. The native MS spectrum of EthR alone confirmed the dimeric nature of EthR in solution, with the charge state distribution being centered around the 13+ state (**Figure 3.2a**). The observed mass of dimeric EthR (50,475 ± 97 Da) was consistent with the theoretical mass of the dimer of the construct (50,456 Da). A small amount of monomeric EthR was also observed, at about 5% of the total protein content.

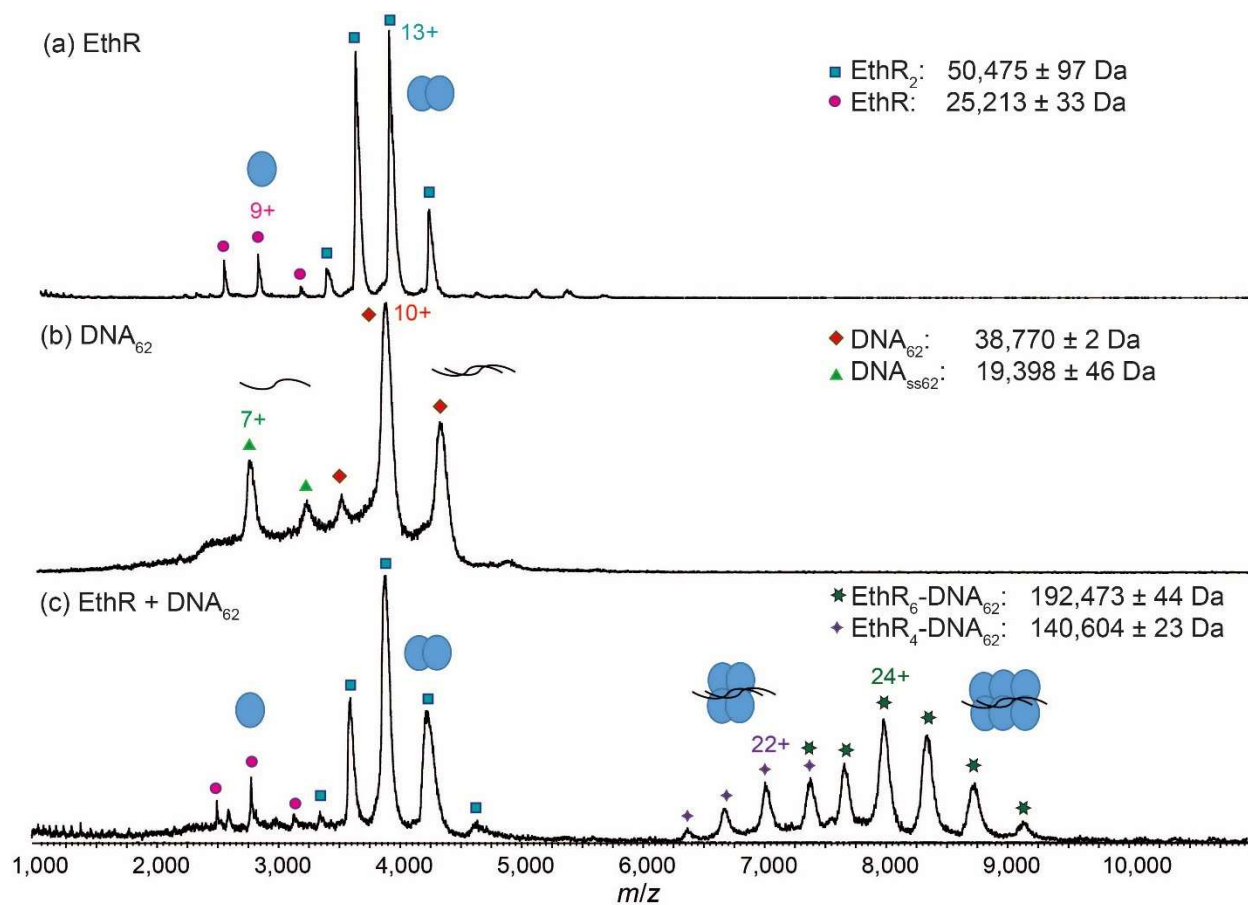


Figure 3.2 Native MS reveals the formation of EthR₄-DNA and EthR₆-DNA complexes. MS spectra of (a) EthR alone showing that the protein exists predominantly as the dimer, (b) DNA₆₂ alone, and (c) a mixture of EthR and DNA₆₂ in a 8:1 ratio, showing the formation of both EthR₄-DNA₆₂ and EthR₆-DNA₆₂ complexes.

The native mass spectrum of DNA₆₂ alone showed that while DNA₆₂ existed primarily in its expected duplex state centered around the 10+ charge state, a fraction of the DNA was single-stranded (DNA_{ss62}), which could be due to an excess of one of the two complementary oligonucleotides (**Figure 3.2b**). DNA *m/z* signals were generally broad, which could be a result of the large amount of cations that must be adducted to the polyanionic DNA in order for the nucleic acids to be detected in positive-ion mode.^[161]

Previous reports had described difficulties with achieving a stable electrospray with samples containing protein-DNA complexes.^[153, 155] Optimization experiments were therefore carried out

to investigate the effect of various parameters on the EthR-DNA interaction. Due to the sensitivity of protein-DNA interactions to ionic strength,^[162] the effect of NH_4OAc concentration was first investigated. A concentration of 200 mM of NH_4OAc was found to provide the most optimal conditions for detecting the EthR-DNA complex by native MS (**Figure 3.3**). At 20 or 50 mM of NH_4OAc , only a small proportion of EthR-DNA complexes were formed and a significant fraction of free protein and DNA remained. On the other hand, at 500 mM of NH_4OAc , the signals of the EthR-DNA complex broadened and became were hard to resolve. Higher concentrations of ions are also expected to weaken protein-DNA interactions by shielding electrostatic interactions. Therefore, a concentration of 200 mM NH_4OAc was used for further investigation.

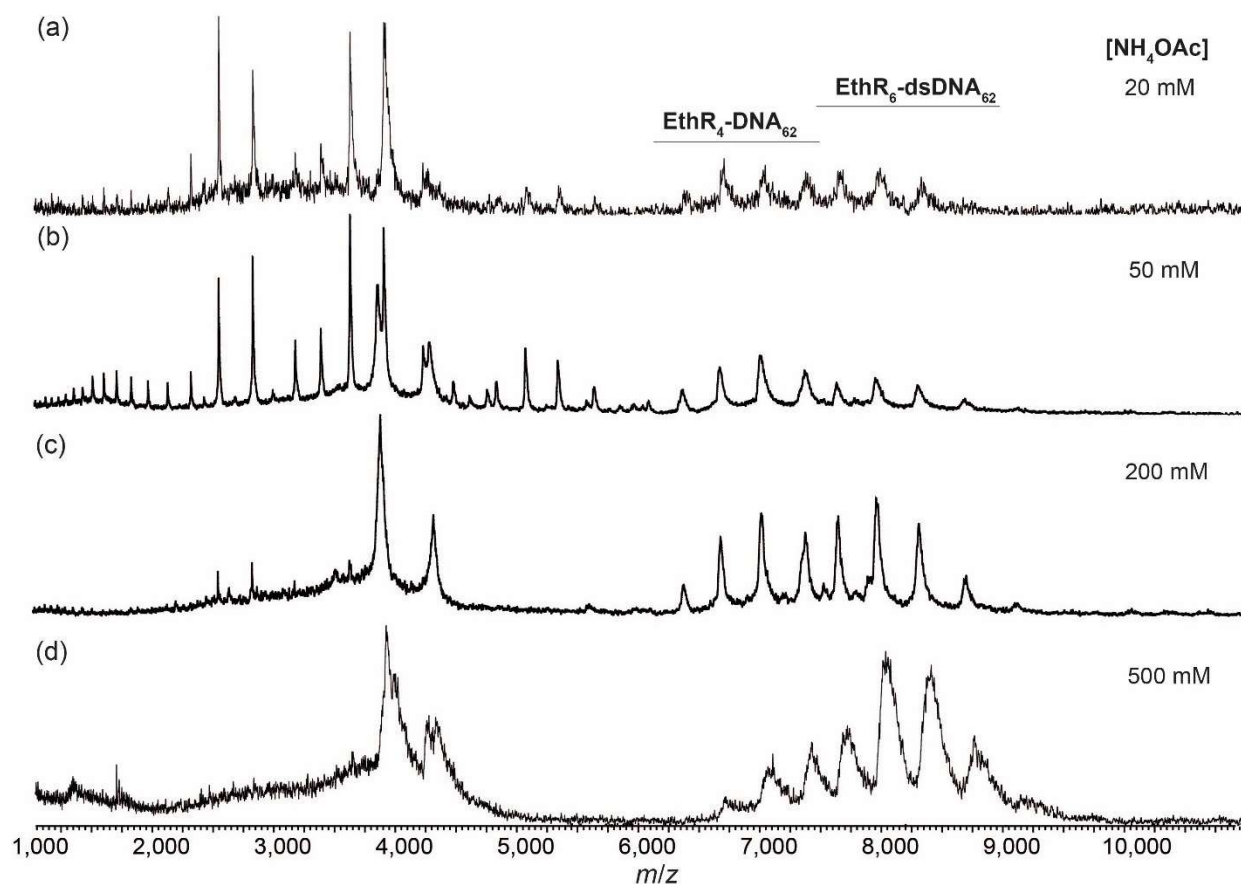


Figure 3.3 Native mass spectra of EthR (20 μM) and DNA₆₂ (2.5 μM) in NH_4OAc (pH 7.0) of concentration (a) 20 mM, (b) 50 mM, (c) 200 mM, (d) 500 mM.

Next, the effect of the addition of Mg^{2+} ions was investigated. Mg^{2+} is thought to play an important role in the structure and function of nucleic acids, including regulating the binding of proteins to DNA. For example, a citrate- Mg^{2+} moiety aided the binding of the dimeric aconitase repressor AcnR, which like EthR is a transcriptional repressor of the TetR/CamR family, to DNA.^[163] However, the addition of 500 μM of Mg^{2+} (as magnesium acetate) did not significantly increase the proportion of protein-DNA complexes (**Figure 3.4**). Higher Mg^{2+} concentrations were not examined because non-volatile ions exert a detrimental effect on the quality of native mass spectra. The presence or absence of Mg^{2+} ions also did not influence the extent of EthR-DNA formation in an SPR assay performed by our collaborator Sherine Thomas in the Blundell group (data not shown).

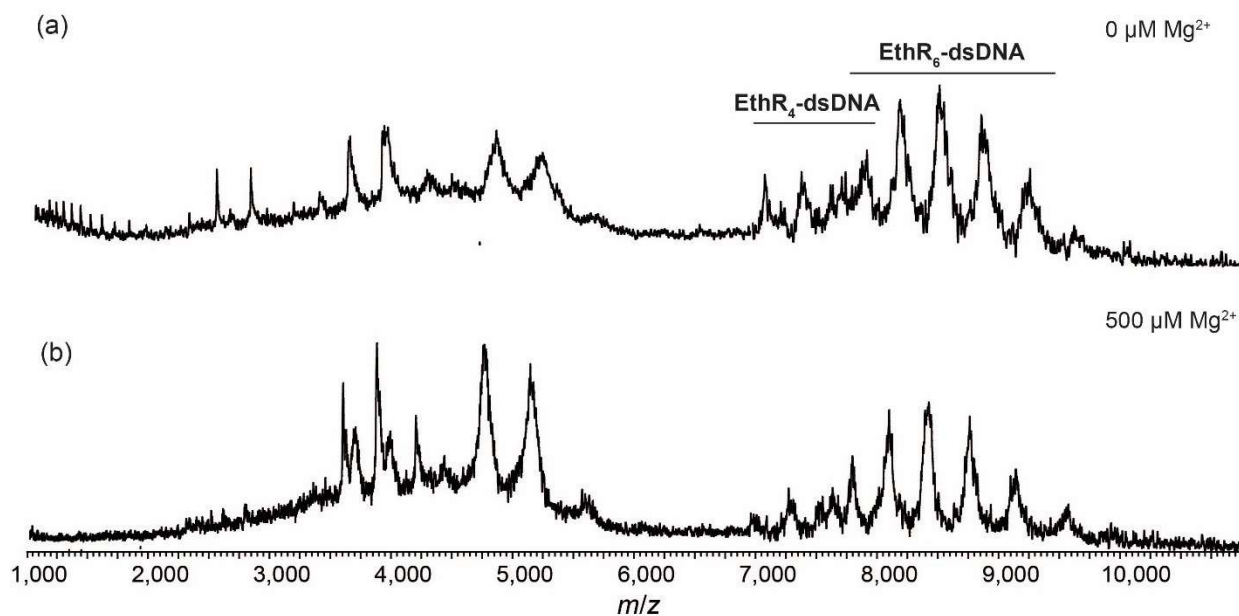


Figure 3.4 Native mass spectra of EthR (20 μM) and DNA₁₀₆ (2.5 μM) in NH_4OAc (200 mM, pH 7.0) in the (a) absence and (b) presence of Mg^{2+} ions (500 μM).

The time course of EthR-DNA complex formation was also monitored by native MS. The data showed that equilibrium was reached within 5 min, as no significant change in the proportion of EthR-DNA complexes were observed at 10 min and 40 min (**Figure 3.5**).

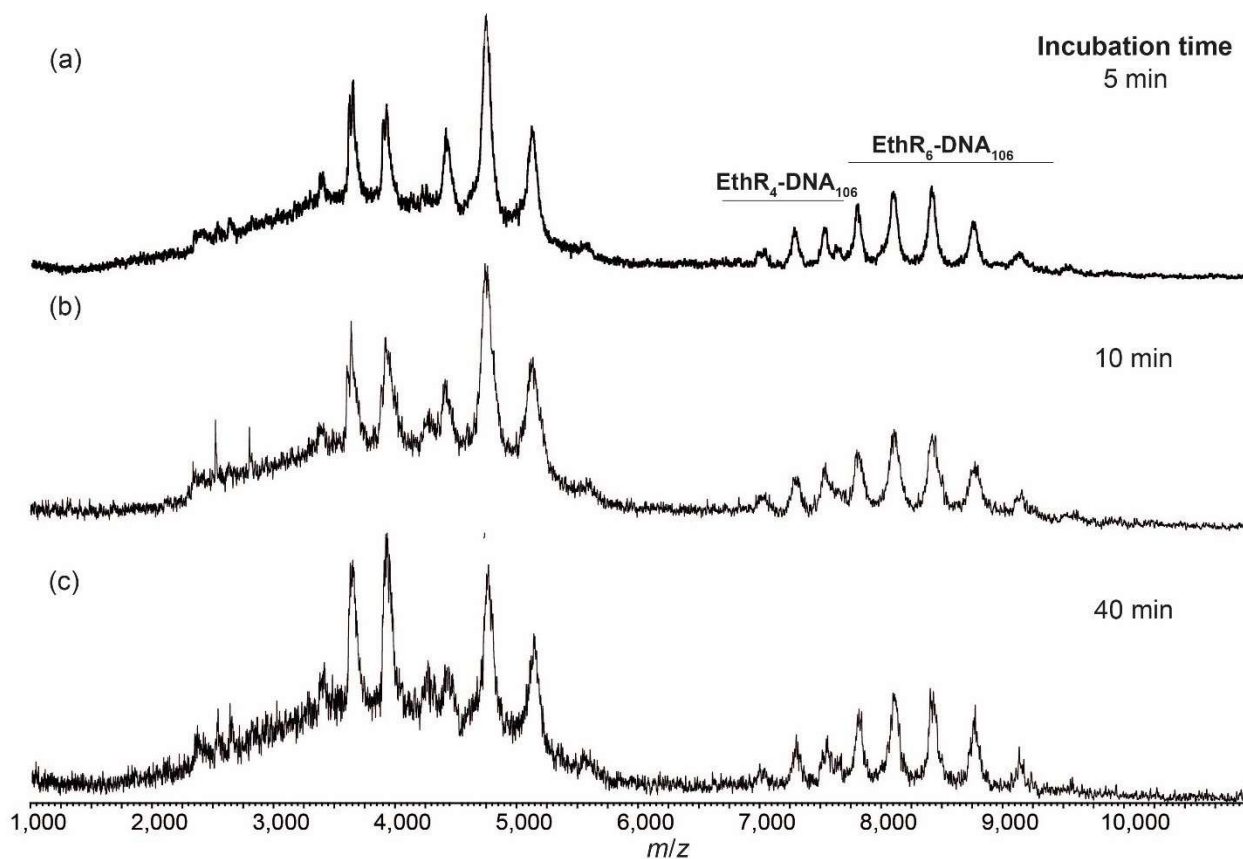


Figure 3.5 Native mass spectra of EthR (20 μM) and DNA₁₀₆ (2.5 μM) in NH₄OAc (200 mM, pH 7.0) after incubation for (a) 5 min, (b) 10 min, and (c) 40 min.

For native MS, the protein is required to be buffer-exchanged into a volatile buffer such as NH₄OAc before analysis. However, there was no significant change in the spectra when EthR and DNA were incubated and then buffer-exchanged together using centrifugal concentrators, versus mixing together separately buffer-exchanged protein with DNA (**Figure 3.6**).

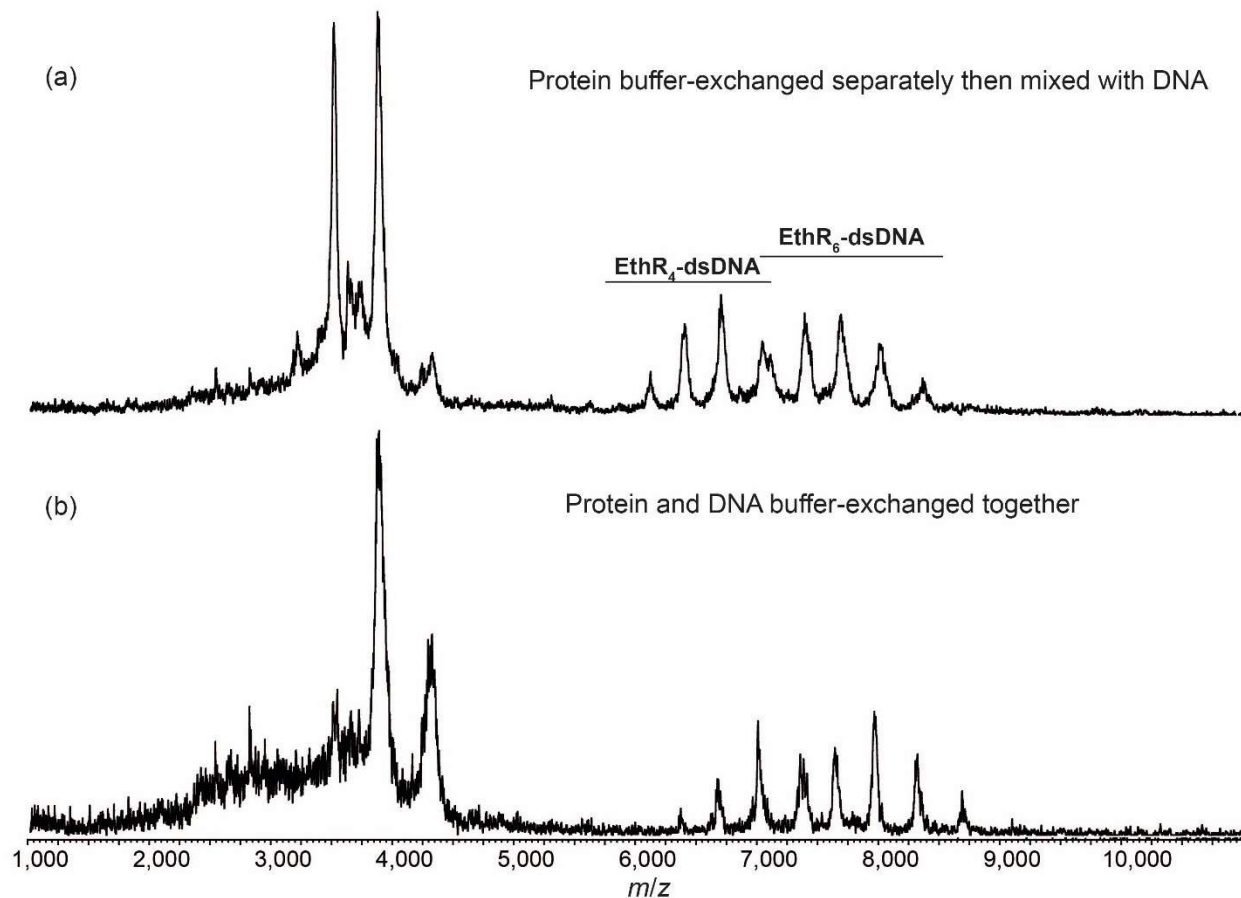


Figure 3.6. Native mass spectra of EthR (20 μ M) and DNA₆₂ (2.5 μ M) in NH₄OAc (200 mM, pH 7.0). In (a), buffer-exchanged protein was incubated with DNA before measurement. In (b), protein was incubated with DNA, and the mixture was buffer-exchanged before measurement.

3.2.1.2. Investigating the stoichiometry of the EthR-DNA interaction

Under the optimized conditions, EthR (20 μ M) and DNA₆₂ (2.5 μ M) were mixed together in an 8:1 ratio and subjected to nESI-MS. Surprisingly, instead of the expected EthR₈-DNA₆₂ complex, only the EthR₆-DNA₆₂ complex ($192,473 \pm 44$ Da), centered around the 24+ state, and the EthR₄-DNA₆₂ complex ($140,604 \pm 23$ Da), centered around the 22+ state, were detected (**Figure 3.2c**). The observed masses are about 1% higher than the theoretical masses of 189,555 and 139,099 Da for the EthR₆-DNA₆₂ and EthR₄-DNA₆₂ species, respectively, which can be attributed to the adduction of weakly-bound molecules or ions. The EthR₆-DNA₆₂ and EthR₄-DNA₆₂ complexes are presumably formed by the assembly of three or two EthR dimers, respectively, onto the DNA. In contrast, no

protein-DNA complexes were observed by native MS when EthR was mixed with a random 55-bp sequence (DNA_{R55}) in a 8:1 or 12:1 ratio (**Figure 3.7**), indicating that the complexes formed between EthR with its operator DNA₆₂ were specific.

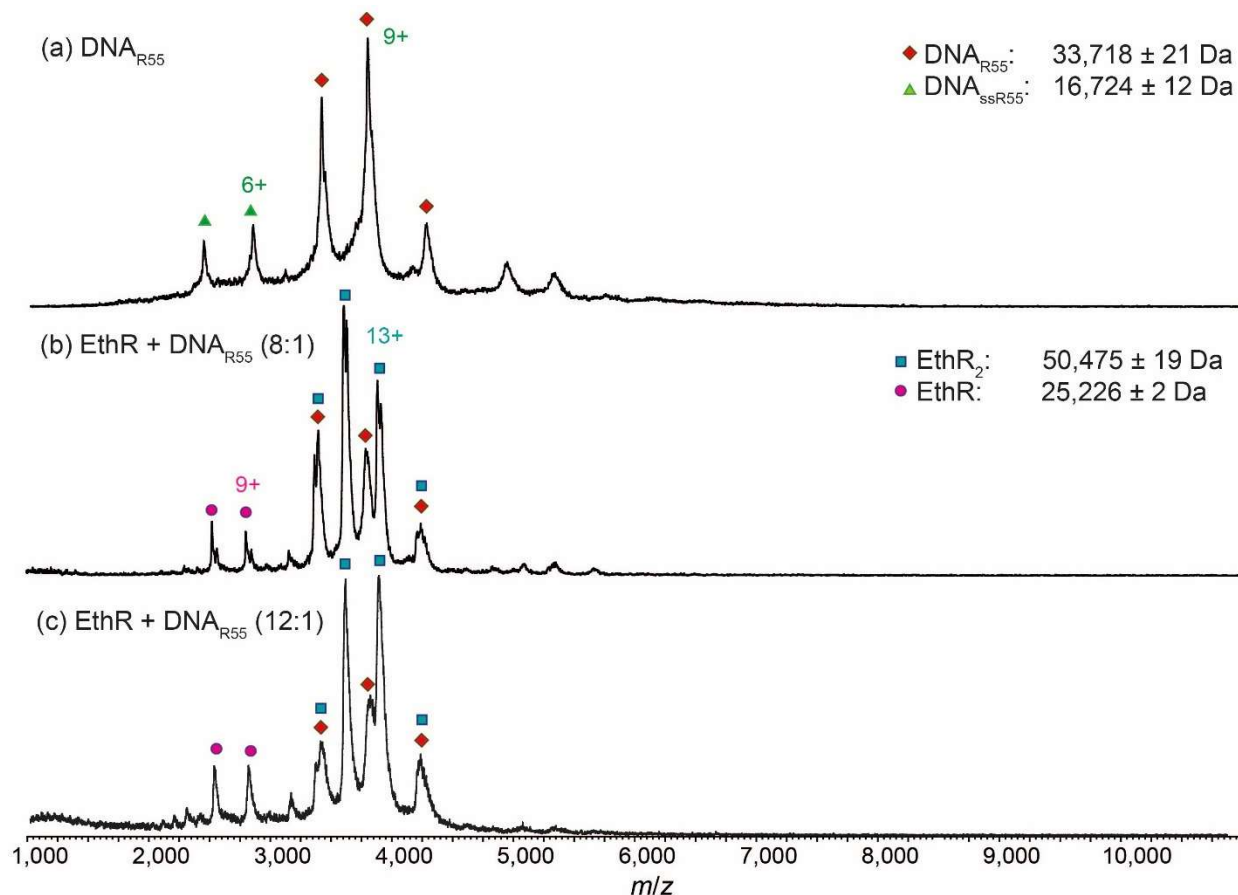


Figure 3.7 Native mass spectra (a) DNA_{R55} (2.5 μM) and (b) a mixture of EthR (20 μM) and DNA_{R55} (2.5 μM) in an 8:1 ratio, and (c) a mixture of EthR (30 μM) and DNA_{R55} (2.5 μM) in a 12:1 ratio. All spectra were recorded in NH₄OAc (200 mM, pH 7.0).

The proportion of EthR and DNA₆₂ was varied in order to investigate whether alternative protein-DNA complexes could be formed at different protein to DNA ratios. At a 6:1 or 7:1 ratio of EthR to DNA₆₂, the EthR₄-DNA₆₂ and EthR₆-DNA₆₂ complexes were detected as the major and minor species, respectively, and some free DNA could also be observed (**Figure 3.8a,b**). The relative intensity of free DNA₆₂ in the native mass spectra was much higher than expected for a slight

excess of DNA over protein, which could be due to the greater ionization efficiency of DNA compared to the protein-DNA complex, as has been previously reported.^[152] Additionally, there could be uncertainty in the protein-DNA ratio due to variability in the absorbance measurements of DNA on the NanoDrop spectrophotometer. As the proportion of EthR was increased to 8:1, the EthR₆-DNA₆₂ complex became the predominant protein-DNA species, and only free protein could be detected (**Figure 3.8c**). Even at a 12:1 ratio of EthR to DNA₆₂, the EthR₆-DNA complex was the largest species detected and no evidence of non-specific binding giving rise to higher oligomeric entities could be observed (**Figure 3.8d**).

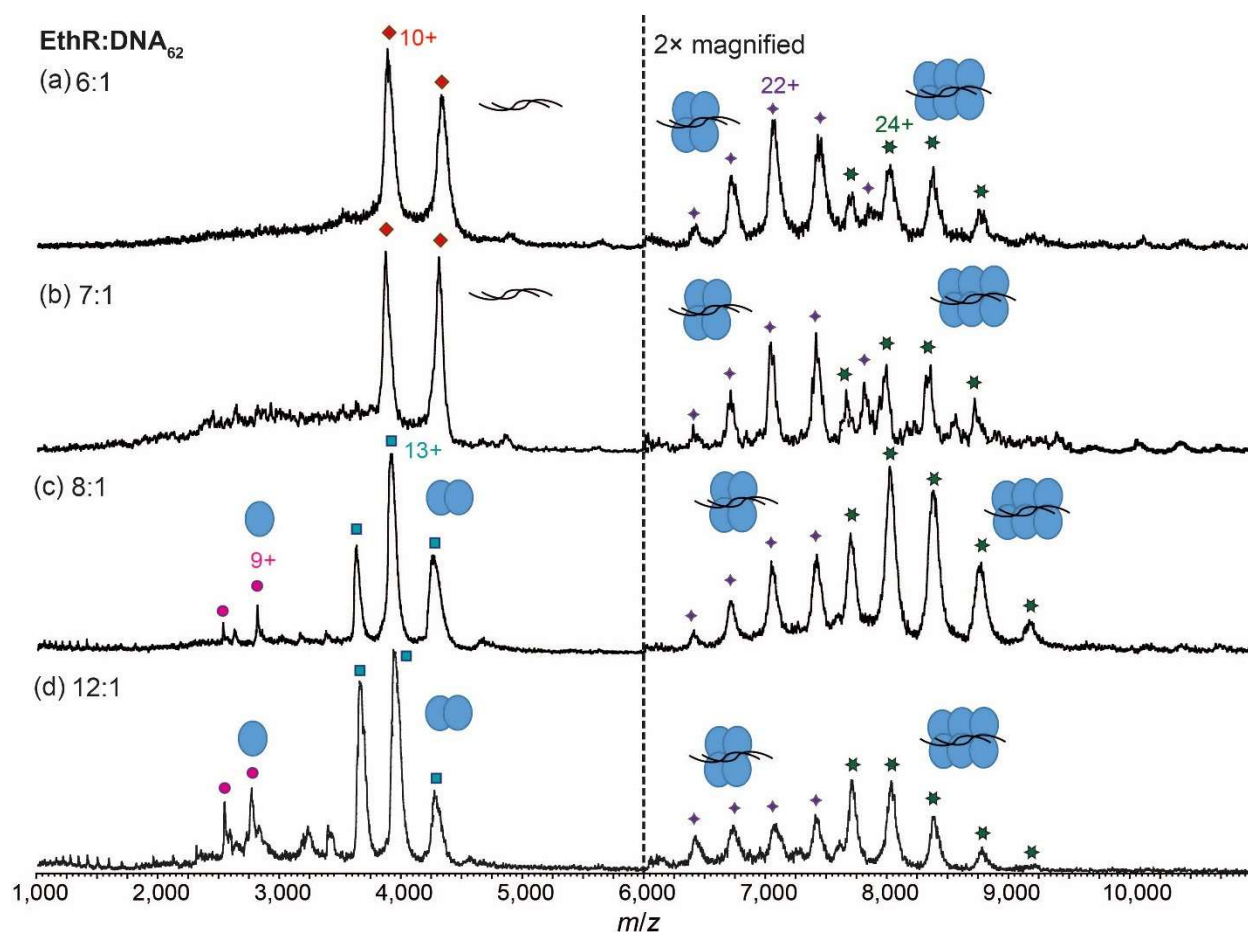


Figure 3.8 Native mass spectra of a mixture of EthR and DNA₆₂ in a (a) 6:1 ratio, (b) 7:1 ratio, (c) 8:1 ratio or (d) 12:1 ratio. The concentration of DNA₆₂ was held at 2.5 μM. All spectra were recorded in NH₄OAc (200 mM, pH 7.0). The protein-DNA peaks have been magnified by 2× for clarity.

A small amount of the EthR₅-DNA₆₂ complex was also sometimes observed (**Figure 3.9**), which could presumably be formed from the association of a monomeric EthR subunit with a EthR₄-DNA₆₂ complex. However, the relatively low abundance of the EthR₅-DNA₆₂ complex and the EthR₃-DNA₃₇ complex (see below) compared to complexes containing an even number of EthR subunits suggests that their biological relevance may be relatively minor. An alternative interpretation is that these odd-numbered species may represent transient intermediates during protein-DNA complex formation.

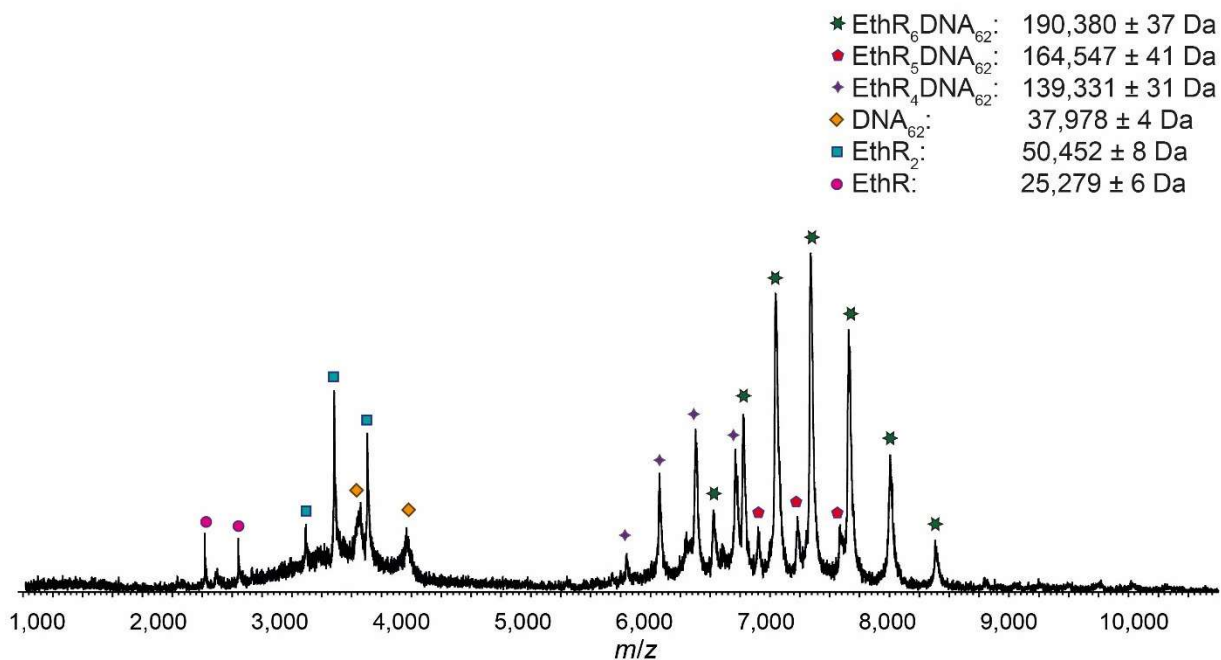


Figure 3.9 Native mass spectra of a mixture of EthR (20 μM) and DNA₆₂ (2.5 μM) in NH₄OAc (200 mM, pH 7.0).

The stoichiometric heterogeneity of the protein-DNA species observed by native MS could possibly account for the difficulties in obtaining an X-ray crystal structure of the EthR-DNA complex thus far. A longer 106-bp DNA sequence (DNA₁₀₆) encompassing the entire ethA-R

intergenic region also gave rise to EthR₆-DNA₁₀₆ and EthR₄-DNA₁₀₆ complexes when incubated with EthR (**Figure 3.10**), however the quality of the native mass spectra was reduced. Generally, it was found in this investigation that longer DNA sequences gave lower-quality spectra in native MS compared to shorter DNA sequences.

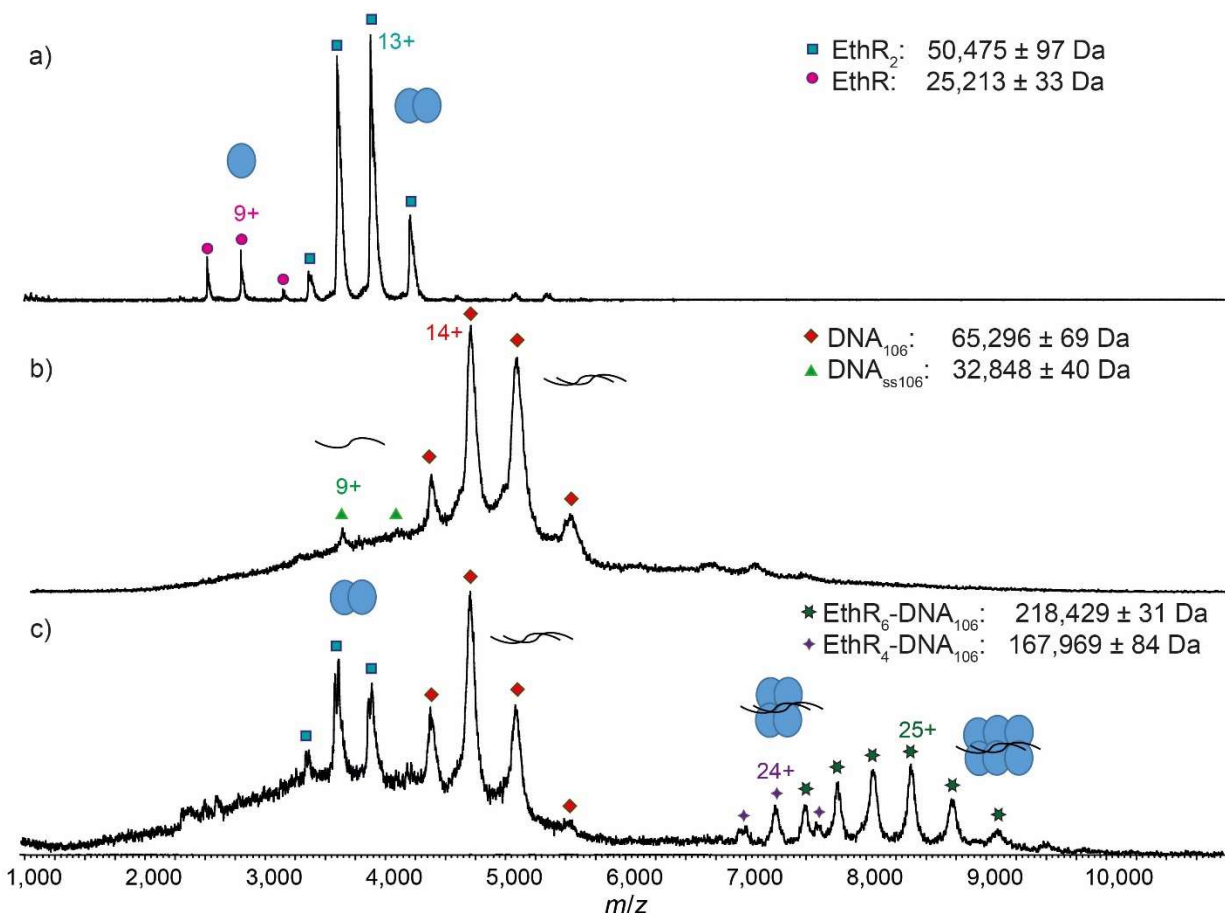


Figure 3.10 Native mass spectra (a) EthR (20 μM), (b) DNA₁₀₆ (2.5 μM) and (c) a mixture of EthR (20 μM) and DNA₁₀₆ (2.5 μM). All spectra were recorded in NH₄OAc (200 mM, pH 7.0).

Next, EthR was incubated separately with the DNA duplexes DNA₃₇ and DNA₃₆ (**Figure 3.1**), and native mass spectra of the resulting complexes were recorded. In the SPR experiments reported previously,^[143] each DNA₃₇ duplex bound to an average of 4.5 EthR molecules, whereas with DNA₃₆, protein-DNA binding was greatly decreased. In native MS, a mixture of EthR (15 μM) and DNA₃₇ (2.5 μM) produced mostly the EthR₄-DNA₃₇ complex, centered around the 19+ charge

state, along with some of the EthR₂-DNA₃₇ complex and a small amount of the EthR₃-DNA₃₇ complex (**Figure 3.11a**). With DNA₃₆, mainly EthR₂-DNA₃₆ complexes were formed, along with a small amount of the EthR₄-DNA₃₆ complex (**Figure 3.11b**). Furthermore, significant quantities of both unbound protein and DNA were detected, which is consistent with the greatly diminished binding capacity of EthR to DNA₃₆ as observed by SPR.

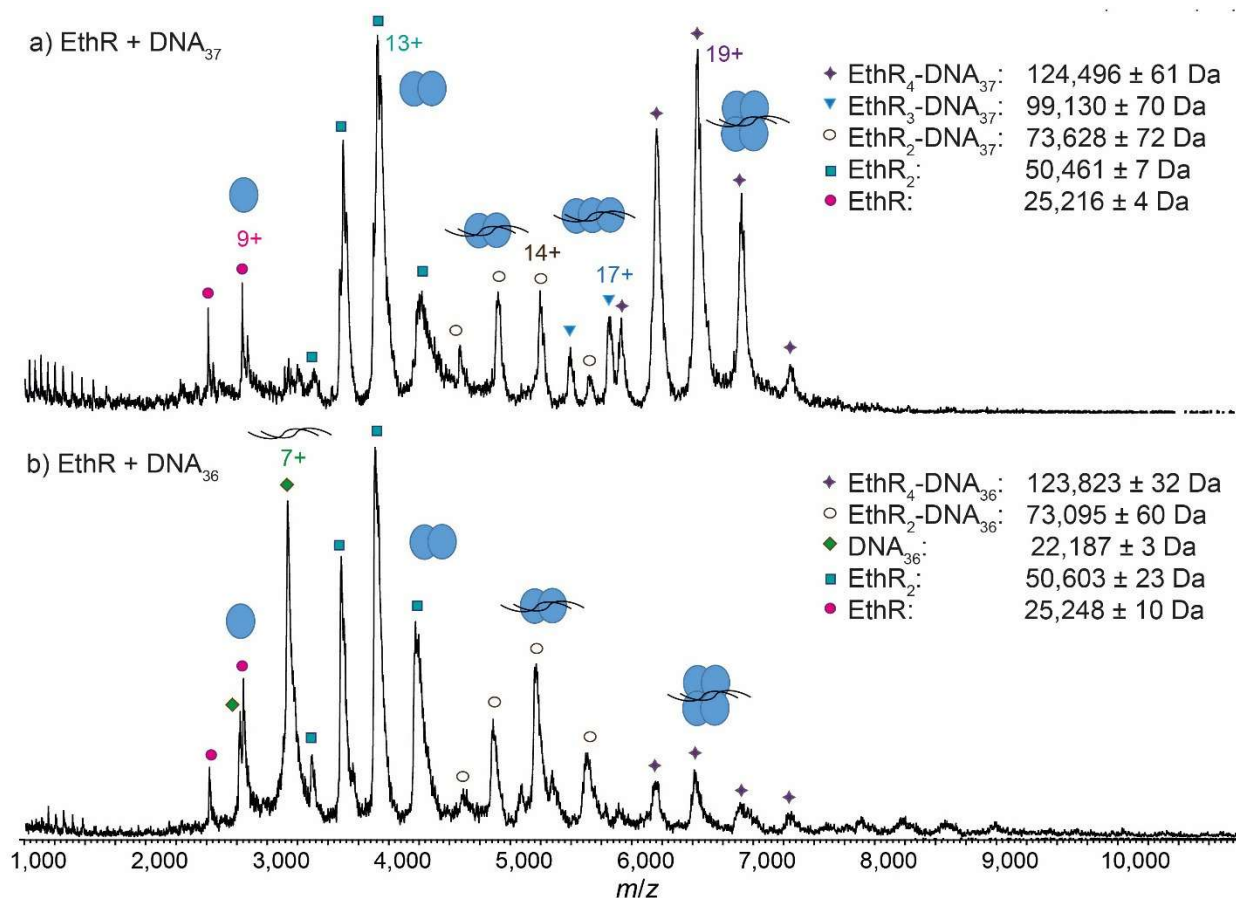


Figure 3.11 Native MS showing reduced stoichiometry of EthR with the shorter DNA₃₇ and DNA₃₆ sequences. MS spectra of (a) EthR and DNA₃₇ in a 6:1 ratio, showing the formation of mainly EthR₄-DNA₃₇ complex, some EthR₂-DNA₃₇ complex, and a trace of EthR₃-DNA₃₇ complex, and (b) EthR and DNA₃₆ in a 6:1 ratio, showing formation of mainly EthR₂-DNA₃₆ complex and some EthR₄-DNA₃₆ complex.

A collision-induced dissociation (CID) experiment was also performed to investigate whether the EthR-DNA complex could be dissociated in the gas phase. Increasing the collision energy of ions through raising the trap/transfer voltage led initially to sharpening of the protein-DNA signals

accompanied by degradation of free DNA species (**Figure 3.12**). As collision energy was raised further, degradation of protein-DNA complexes could also be observed. However, at no point did any new protein-DNA stoichiometries appear that could be assigned to dissociation products of the initial protein-DNA complexes. When the experiment was performed in MS/MS mode with precursor ion selection, the precursor protein-DNA ion decreased in intensity as the collision energy was increased but no recognizable daughter species could be observed (data not shown). This suggests that the energy needed to dissociate the EthR subunits from the complex is greater than the energy needed to degrade the EthR-DNA complex, so that only degradation is observed in this experiment.

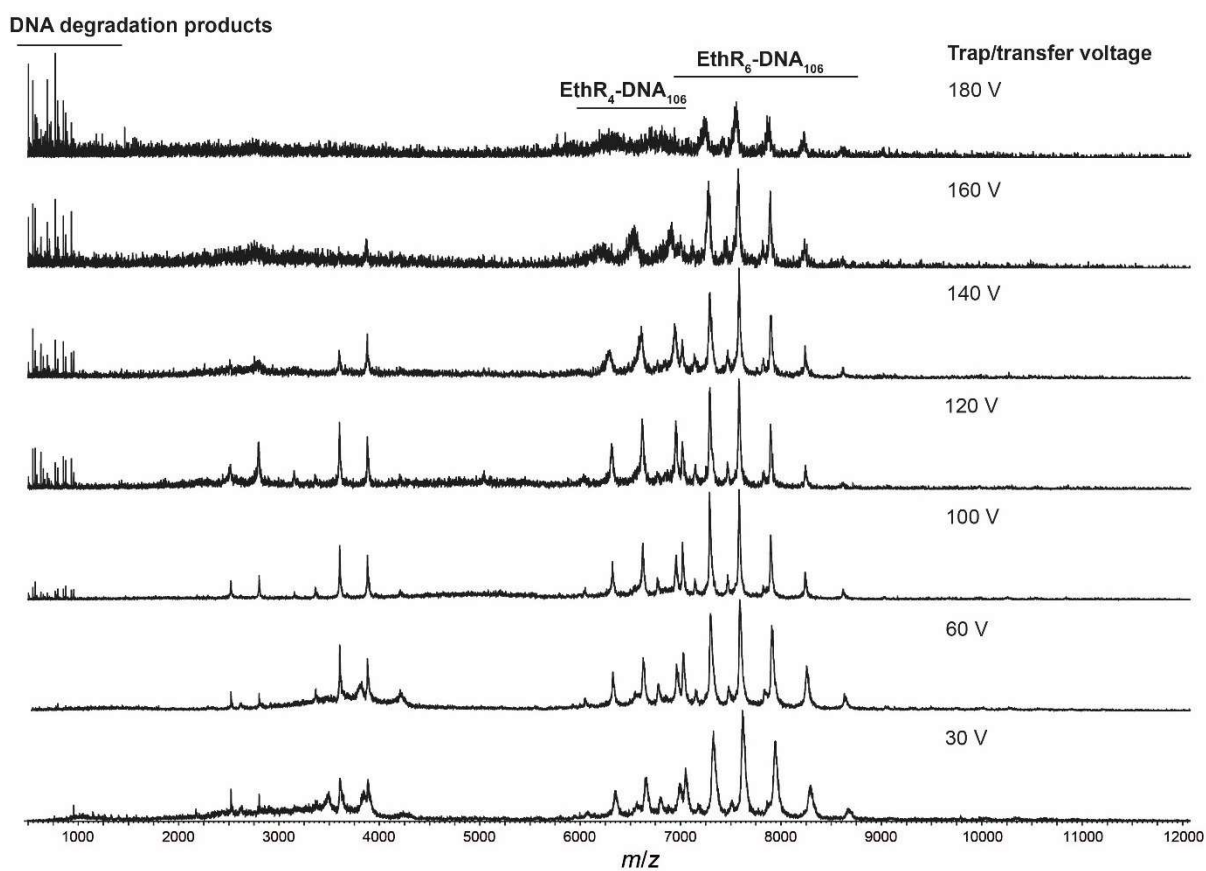


Figure 3.12 CID spectra of EthR-DNA₁₀₆. Collision energy was raised from 30 V to 180 V by adjusting both the trap and transfer voltages.

3.2.1.3. EthR dimers bind to separate sites on DNA

To further investigate the EthR-DNA interaction, DNA₆₂ was split into two half-sites, the left-hand site (DNA_{L31}) and the right-hand site (DNA_{R31}) (**Figure 3.1**). Each site contained one copy of an imperfect direct repeat that was thought to be responsible for the EthR-DNA binding interaction.^[143] When EthR (15 μM) was incubated with DNA_{L31} (2.5 μM), only EthR₂-DNA_{L31} complexes were formed (**Figure 3.13a**). On the other hand, mainly EthR₄-DNA_{R31} complexes were observed with DNA_{R31}, along with a small amount of EthR₂-DNA_{R31} and EthR₃-DNA_{R31} species (**Figure 3.13b**). As before, the significant quantities of both unbound protein and DNA that are observed suggests that the interactions between EthR and the individual half-sites are relatively weak.

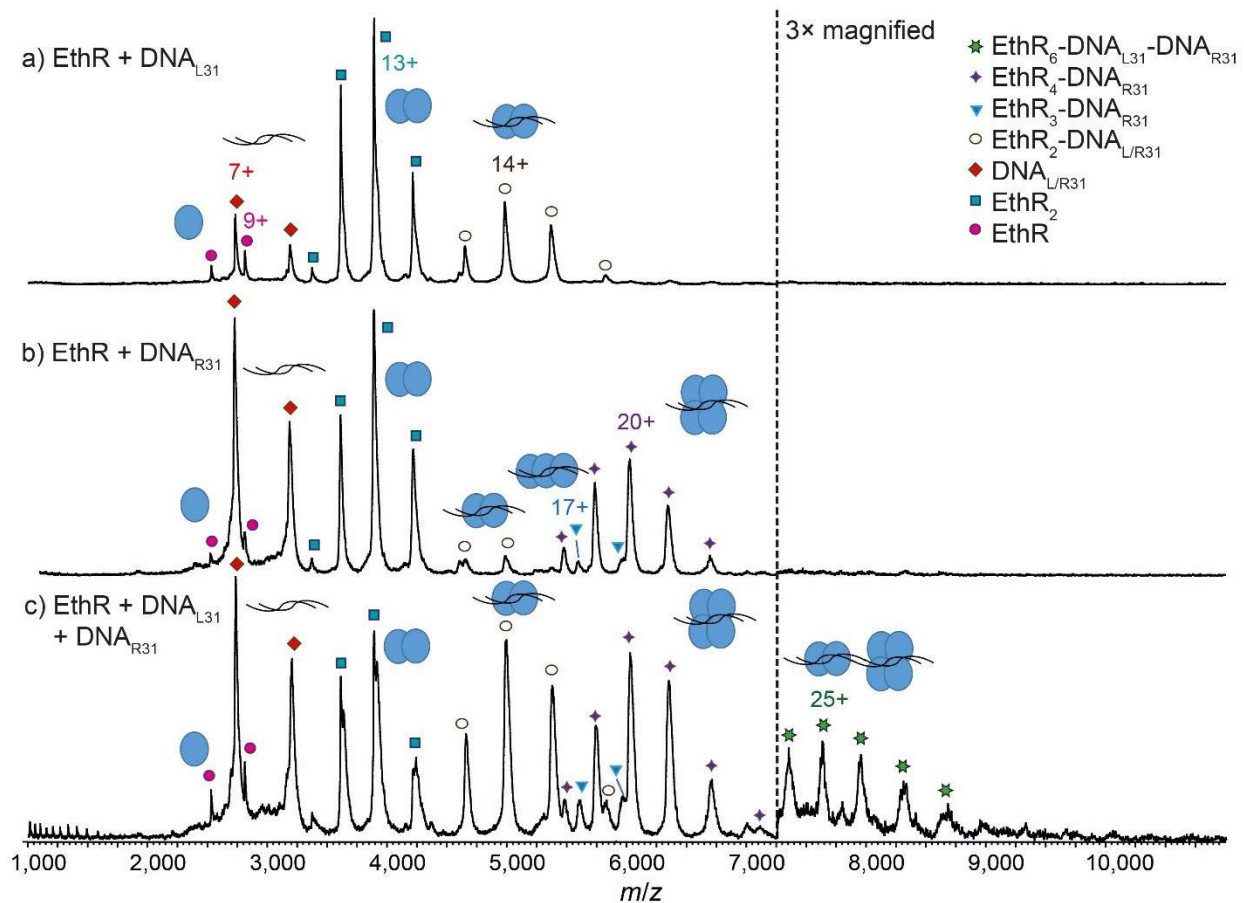


Figure 3.13 Native MS showing recapitulation of the EthR6-DNA complex from two independent half-sites. MS spectra of (a) EthR and DNA_{L31} in a 6:1 ratio, showing formation of mainly EthR₂-DNA complex, and (b) EthR and DNA_{R31} in a 6:1 ratio, showing formation of mainly EthR₄-DNA complex, along with a small amount of EthR₂-DNA and EthR₃-DNA complexes. (c) When EthR was mixed with both DNA_{L31} and DNA_{R31}, a small amount of the putative EthR₆-DNA_{L31}-DNA_{R31} complex was formed.

Intriguingly, when EthR was incubated with both DNA_{L31} and DNA_{R31} at the same time, a small amount of higher-order species could be detected that could be putatively assigned as the EthR₆-DNA_{L31}-DNA_{R31} complex ($189,872 \pm 23$ Da) (**Figure 3.13c**). This complex could be formed from the association of the EthR₂-DNA_{L31} complex with the EthR₄-DNA_{R31} complex in solution. This interaction appears to be asymmetric, as homodimeric (EthR₂-DNA_{L31})₂ or (EthR₄-DNA_{R31})₂ species were not detected when EthR was treated with DNA_{L31} or DNA_{R31} separately. Additionally, DNA_{L31} and DNA_{R31} do not directly associate with each other in the absence of EthR (**Figure 3.14**), indicating that some kind of communication must exist between EthR and the DNA in order to form the putative EthR₆-DNA_{L31}-DNA_{R31} complex.

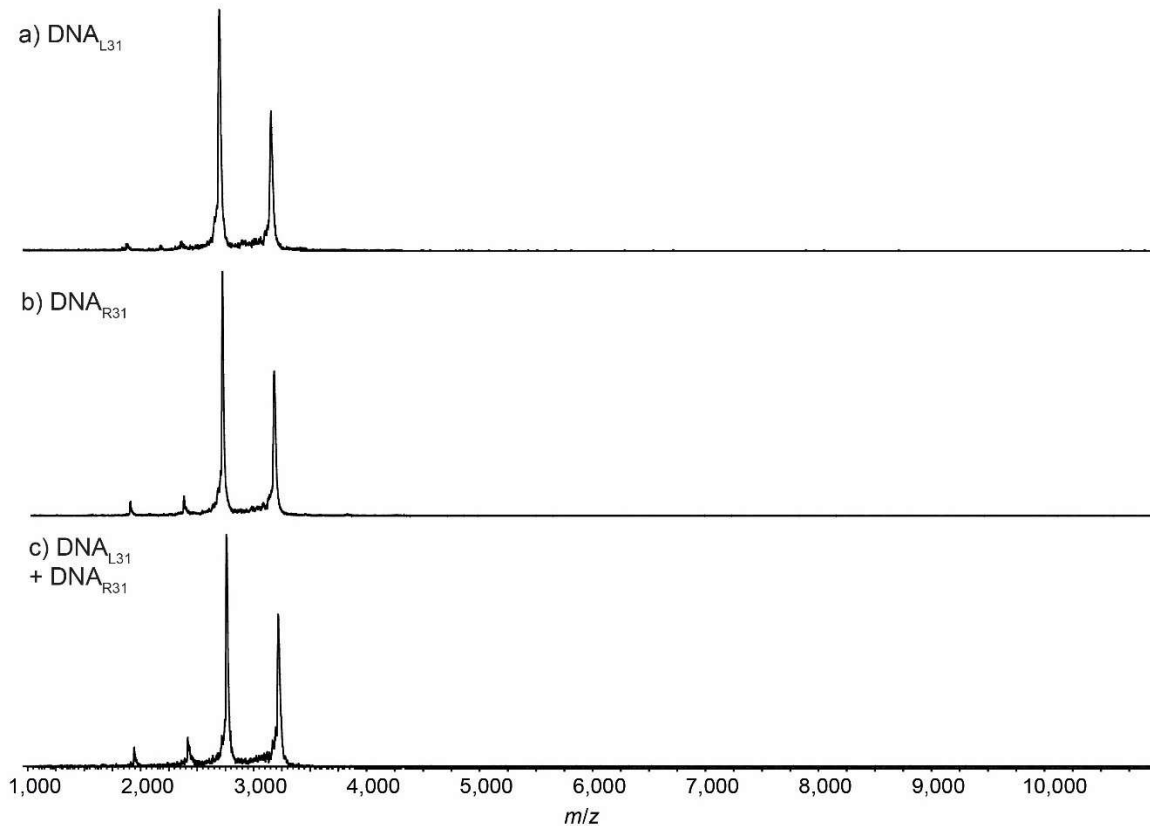


Figure 3.14. Native mass spectra of (a) DNA_{L31} (2.5 μ M), b) DNA_{R31} (2.5 μ M) and (c) a mixture of DNA_{L31} (2.5 μ M) and DNA_{R31} (2.5 μ M). All spectra were recorded in NH₄OAc (200 mM, pH 7.0). The theoretical masses of DNA_{L31} and DNA_{R31} are 19032 Da and 19031 Da respectively.

3.2.1.4. Isothermal titration calorimetry of the EthR-DNA interaction

The EthR-DNA interaction was also investigated using isothermal titration calorimetry (ITC). EthR was titrated into DNA₆₂, generating a complex, non-sigmoidal binding isotherm (**Figure 3.15**) that was analyzed using AFFINImeter software (S4SD). The experiment was performed twice. In the first experiment, the concentration of EthR dimer was 418 μ M and the concentration of DNA₆₂ was 10.7 μ M. In the second experiment, the concentration of EthR dimer was 548 μ M and the concentration of DNA₆₂ was 15.7 μ M. A stoichiometric equilibrium binding model was designed that assumed the stepwise formation of the EthR₆-DNA complex from free EthR dimers and DNA₆₂, proceeding through EthR₂-DNA and EthR₄-DNA intermediates (*i.e.*, EthR₂ + DNA₆₂ \rightleftharpoons EthR₂-DNA₆₂ \rightleftharpoons EthR₄-DNA₆₂ \rightleftharpoons EthR₆-DNA₆₂). Fitting of the data to this model revealed a

stoichiometry of 2.9 ± 0.1 EthR dimers per duplex of DNA₆₂, which is consistent with the native MS data. The binding affinities of the first and second dimers to DNA₆₂ were identical within experimental error ($K_{D(1)} = 3.8 \pm 0.8 \mu\text{M}$, $K_{D(2)} = 3.6 \pm 0.6 \mu\text{M}$). The third EthR dimer binds to DNA₆₂ with weaker affinity ($K_{D(3)} = 10 \pm 3 \mu\text{M}$). The EthR-DNA₆₂ binding affinities derived from ITC in this work are about an order of magnitude weaker than those previously determined by SPR for the EthR-DNA₁₀₆ interaction (average $K_D = 146 \text{ nM}$).^[143]

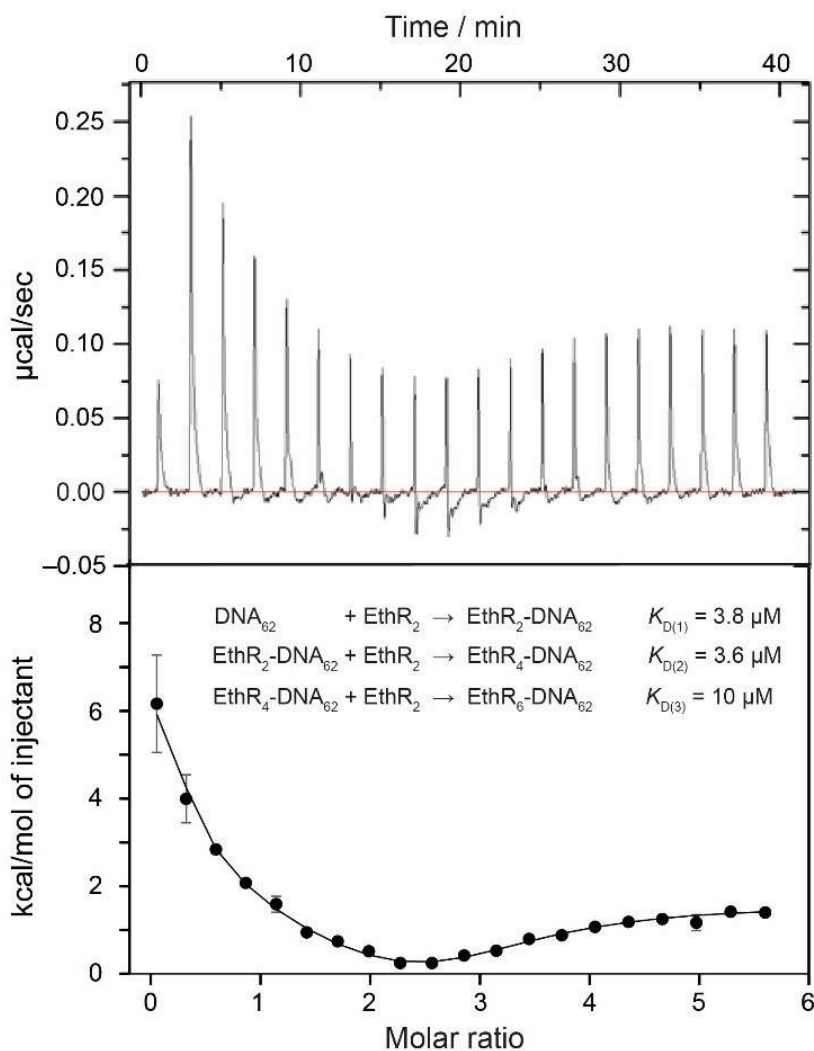


Figure 3.15 ITC data for the titration of EthR into DNA₆₂ showing the thermogram (upper panel) and binding isotherm (lower panel). Integrated data were fit with a stoichiometric equilibrium binding model. Moles of EthR are given as the dimer. ITC data were fit to an equilibrium binding model that assumes the

stepwise formation of the EthR₆-DNA complex from EthR dimers, as indicated in the lower panel along with the associated equilibrium dissociation constants for the binding of the first ($K_{D(1)}$), second ($K_{D(2)}$) and third ($K_{D(3)}$) EthR dimer to the DNA.

The calculated thermodynamic parameters indicate that the binding of the first EthR dimer to DNA is entropically favorable ($T\Delta S = +12.1 \pm 0.4 \text{ kcal mol}^{-1}$) but enthalpically unfavorable ($\Delta H = +4.7 \pm 0.4 \text{ kcal mol}^{-1}$) (**Figure 3.16**). The favorable increase in entropy may be due to the release of counter cations or solvent molecules that were associated with the DNA.^[162] However, the enthalpic penalty of binding stands in contrast to that observed for most major groove-binding proteins.^[164] This suggests that significant structural rearrangement takes place in EthR and/or the DNA to accommodate binding of the first EthR dimer. Similar to what has been observed with the related TetR/CamR family repressor QacR,^[165] which binds to DNA as a dimer of dimers, this structural rearrangement could then present the DNA in a conformation that readily accepts the second dimer. Indeed, binding of the second EthR dimer is driven almost entirely by enthalpy ($\Delta H = -6.2 \pm 0.6 \text{ kcal mol}^{-1}$) with only a minor entropic component ($T\Delta S = -1.3 \pm 0.6 \text{ kcal mol}^{-1}$). The third dimer binds via a combination of favorable enthalpic ($\Delta H = -2.1 \pm 0.6 \text{ kcal mol}^{-1}$) and entropic ($T\Delta S = -4.7 \pm 0.6 \text{ kcal mol}^{-1}$) terms.

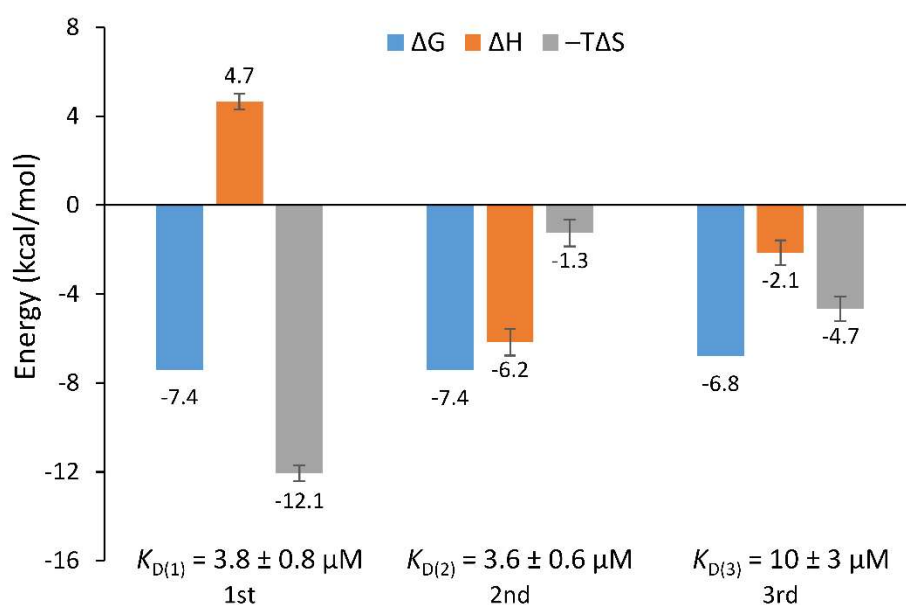


Figure 3.16 Thermodynamic parameters of the EthR-DNA₆₂ interaction as determined by ITC. ΔG , ΔH and $-T\Delta S$ values are shown for the binding of the first, second and third EthR dimers with DNA. Moles of EthR are given as the dimer. Error bars represent the standard deviation of duplicate results.

The discrepancy in the stoichiometry of the EthR-DNA complex as determined by native MS and ITC versus the previous SPR analysis^[166] may be a consequence of the correction that was applied to the SPR data. In that experiment, a correction factor of 0.73 was applied to the response of DNA, reflecting the different molar refractive indices of DNA versus proteins.^[167] However, more recent work has shown that proteins and nucleic acids behave similarly in SPR, and so there may have been no need for this correction factor.^[168] When this is taken into account, the previous SPR data instead suggest that each DNA₆₂ duplex binds to 6.3 His₆-EthR molecules, while each DNA₃₇ duplex binds to 3.3 His₆-EthR molecules, both numbers being consistent with the present study.

The reverse titration of DNA₆₂ into EthR was also attempted, but this gave a poorly defined isotherm that reached completion very quickly. This could be due to the fact that the titration of DNA₆₂ into EthR would create a very large excess of protein to DNA in the sample cell, strongly favoring the formation of 6:1 EthR-DNA complexes. Instead, the forward titration of EthR into DNA₆₂ allows the system to transition progressively through intermediate complexes (e.g. 2:1 or 4:1), generating more gradual isotherms for analysis.

3.2.1.5. Summary

In conclusion, structural insights into the interaction between EthR and its operator have been obtained by native MS. While EthR was observed to exist as a dimer in solution as expected, the interaction of EthR with the full-length operator produced EthR₆-DNA and EthR₄-DNA complexes. The stoichiometry of the EthR-DNA complex was confirmed by ITC, which also revealed thermodynamic parameters that were consistent with a cooperative mode of binding. This study also highlights the capability of native MS to provide structural information on macromolecular assemblies, including where heterogeneous mixtures of complexes exist that are intractable to crystallization.

3.2.2. Fragment screening against the EthR-DNA interaction using native MS

3.2.2.1. Validation of native MS assay to screen EthR-DNA interaction disruptors

Previous results have suggested that six subunits of EthR bind to its DNA operator sequence with micromolar affinity.^[169] As fragments are weakly binding, they would not be able to significantly disrupt the EthR-DNA interaction at concentrations that are compatible with native MS (typically ≤ 1 mM). Therefore, screening was performed against a partial protein-DNA complex that was formed between four subunits of EthR and DNA_{R31}, that encompasses only about half of the full-length DNA operator sequence (DNA₆₂) (**Figure 3.1**).^[143] It was reasoned that the use of this lower-affinity EthR₄-DNA_{R31} complex would increase the sensitivity of the assay towards detecting weak fragments.

To validate this hypothesis, dose-response experiments were performed with the thienyloxadiazole **BDM31381** ($IC_{50} = 0.52 \mu\text{M}$ by SPR)^[144] against both the partial EthR₄-DNA_{R31} complex (**Figure 3.17a**) and the native EthR₆-DNA₆₂ complex (**Figure 3.17**) by native MS. The results showed that **BDM31381** disrupted the partial EthR₄-DNA_{R31} complex with about 4-fold higher potency compared to the native EthR₆-DNA₆₂ complex (IC_{50} values of $0.4 \mu\text{M}$ and $1.6 \mu\text{M}$, respectively) (**Figure 3.17**). This result suggests that the EthR₄-DNA_{R31} complex is more susceptible to dissociation by small molecules, which should improve the sensitivity of the assay for the screening of fragments.

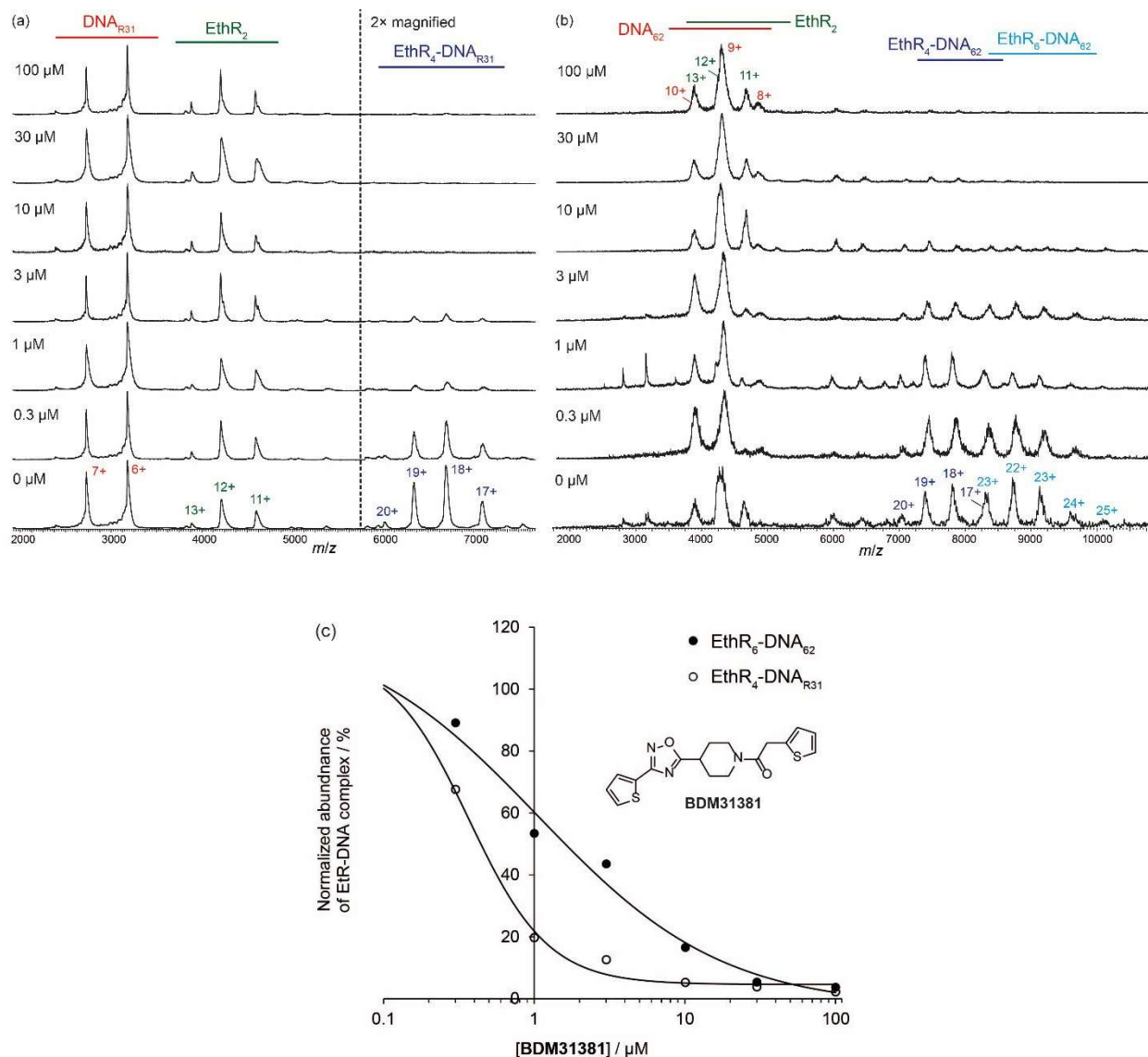


Figure 3.17 Inhibition of (a) EthR-DNA_{R31} and (b) EthR-DNA₆₂ complexes by **BDM31381** (0–100 μM) as determined by native MS. Charge states corresponding to the EthR dimer (EthR₂), DNA duplex (DNA_{R31} or DNA₆₂), tetramer-DNA complex (EthR₄-DNA_{R31} or EthR₄-DNA₆₂) and hexamer-DNA complex (EthR₆-DNA₆₂) are indicated on the mass spectra. (c) Plot of normalized abundance of the EthR-DNA complex against **BDM31381** concentration. Estimated IC₅₀ values: 0.4 μM against EthR₄-DNA_{R31} and 1.6 μM against EthR₆-DNA₆₂.

EthR possesses a long hydrophobic channel that is known to bind a wide variety of molecules, including even adventitious components of crystallization buffers such as hexyl octanoate^[149] or

1,4-dioxane.^[150] The binding of molecules to this site can trigger a conformational change of the EthR homodimer into a conformation that is incompatible with DNA binding.^[149] Given that the screening of fragment molecules often entails the use of significant quantities of DMSO at the high concentrations of compounds that are used, DMSO might be able to enter the EthR ligand binding site and interfere with DNA binding. Our results showed that the EthR₄-DNA_{R31} complex was disrupted by >50% at DMSO concentrations at or above 4%(v/v) (**Figure 3.18**). Therefore, a concentration of 2% DMSO was selected as a compromise between preserving the stability of the EthR-DNA_{R31} complex and raising the upper limit of fragment concentration that could be used. At 2% DMSO, the abundance of the EthR₄-DNA_{R31} species was decreased by about 12%.

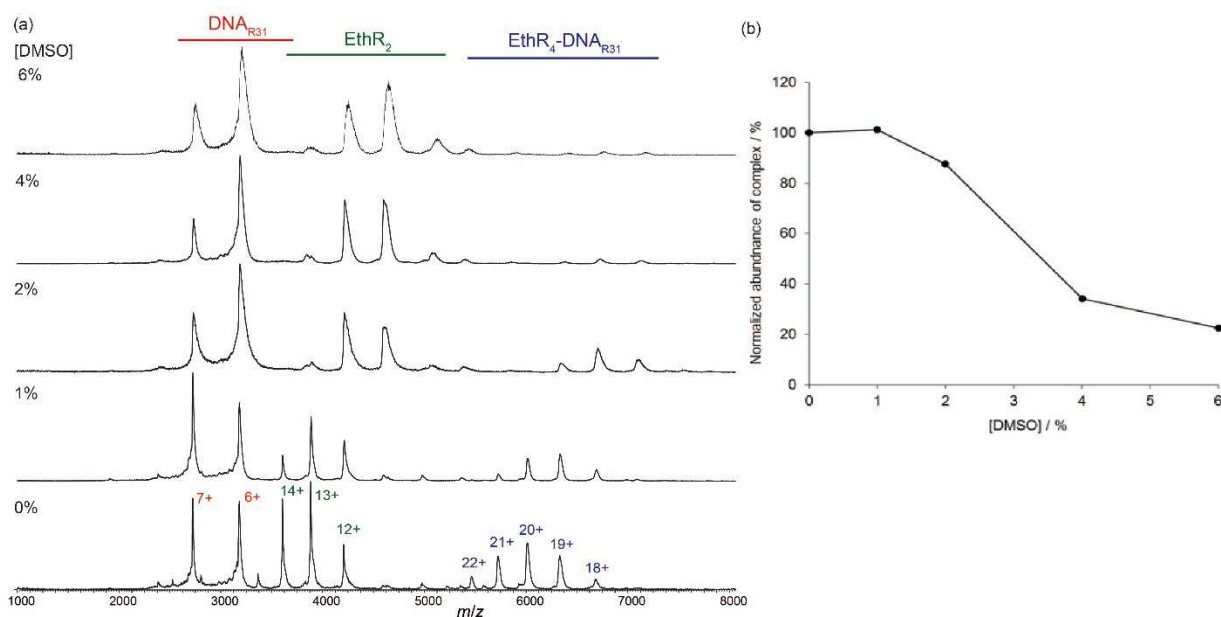


Figure 3.18 Effect of DMSO concentration (0–6% v/v) on EthR-DNA_{R31} complex formation as determined by native MS. (a) Charge states corresponding to the EthR dimer (EthR₂), DNA duplex (DNA_{R31}) and tetramer-DNA complex (EthR₄-DNA_{R31} or EthR₄-DNA₆₂) are indicated on the mass spectra. (b) Plot of normalized abundance of the EthR-DNA complex against DMSO concentration.

3.2.2.2. Fragment screening of EthR-DNA disruptors by native MS

The automated nESI-MS screening of a fragment library has been reported by several groups, which is typically achieved with the aid of a liquid handling and sample infusion robot.^[63, 67, 160]

However, without access to such automation, a smaller library of 80 fragments was employed to demonstrate the proof-of-concept of the use of native MS for screening fragments against a protein-DNA interaction. Of the 80 fragments, 7 could not be screened because they were not sufficiently soluble at 500 μM in 200 mM NH_4OAc buffer containing 2% DMSO. The 73 remaining compounds were screened at 500 μM against the $\text{EthR}_4\text{-DNA}_{\text{R31}}$ complex by native MS. **Figure 3.19a** and **Figure 3.19b** respectively show the histogram and bar chart of $\text{EthR}_4\text{-DNA}_{\text{R31}}$ disruption values for the screened compounds.

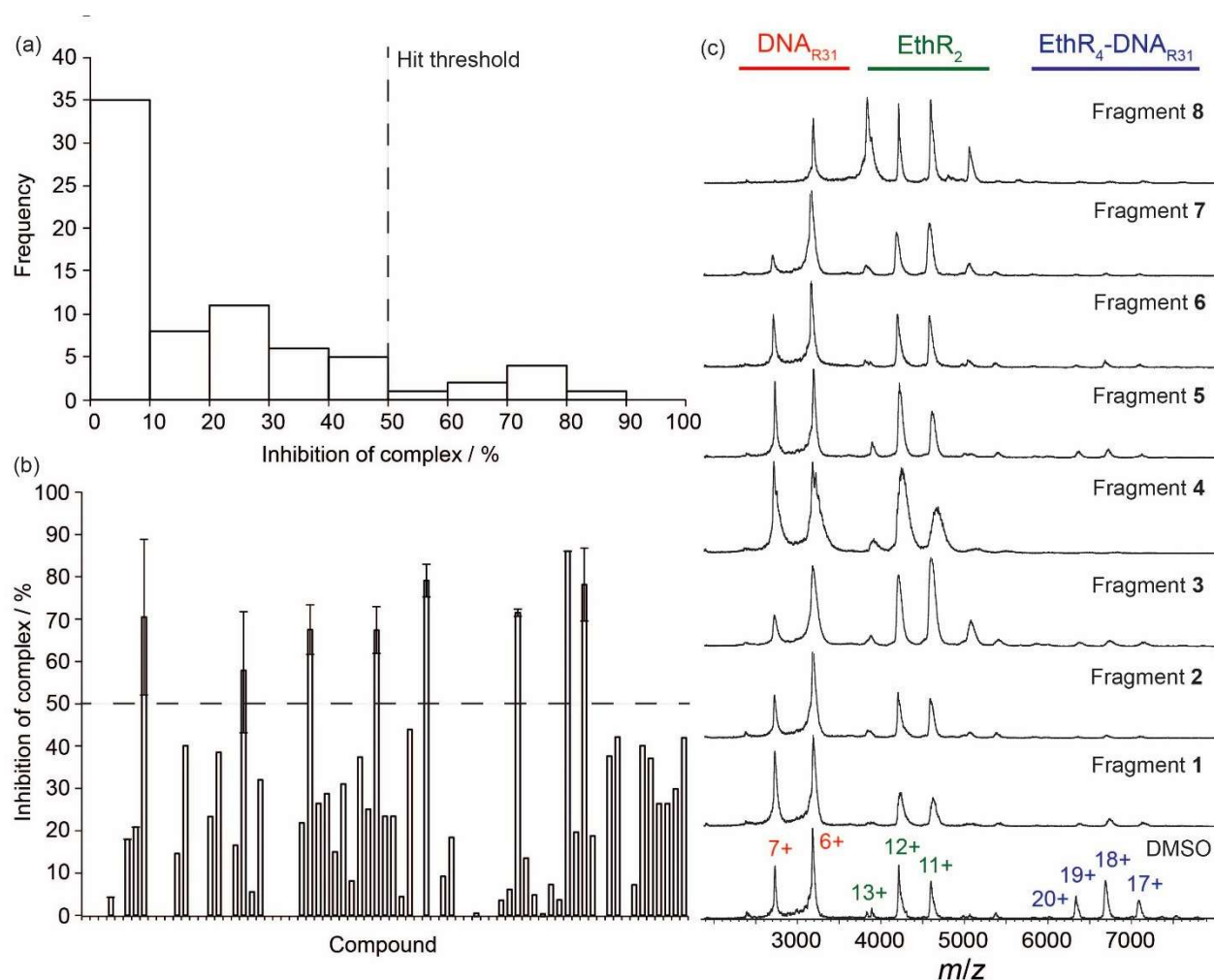


Figure 3.19 Fragment screening against the EthR -DNA interaction by native MS. (a) Histogram and (b) bar chart of disruption values of the $\text{EthR}_4\text{-DNA}_{\text{R31}}$ complex by fragments. The hit threshold (>50% disruption) is indicated by a dashed line. Error bars represent the standard deviation of duplicate results. (c) Native

mass spectra of the EthR₄-DNA_{R31} complex in the presence of the fragment hits (1–8). Fragments were tested at 500 μM in NH₄OAc buffer (200 mM) containing 2% DMSO.

Eight fragments (compounds 1–8, **Figure 3.20**) disrupted the EthR₄-DNA_{R31} interaction by more than 50% and were deemed as hits, giving an overall hit rate of 11%. The mass spectra for the EthR-DNA complexes in the presence of hit fragments are displayed in **Figure 3.19c**. Hits were retested once for confirmation. The entire native MS screening campaign for the 73 fragments utilized only about 0.2 mg of protein and 0.04 mg of total compound, highlighting the important advantage of low sample consumption requirement for this technique.

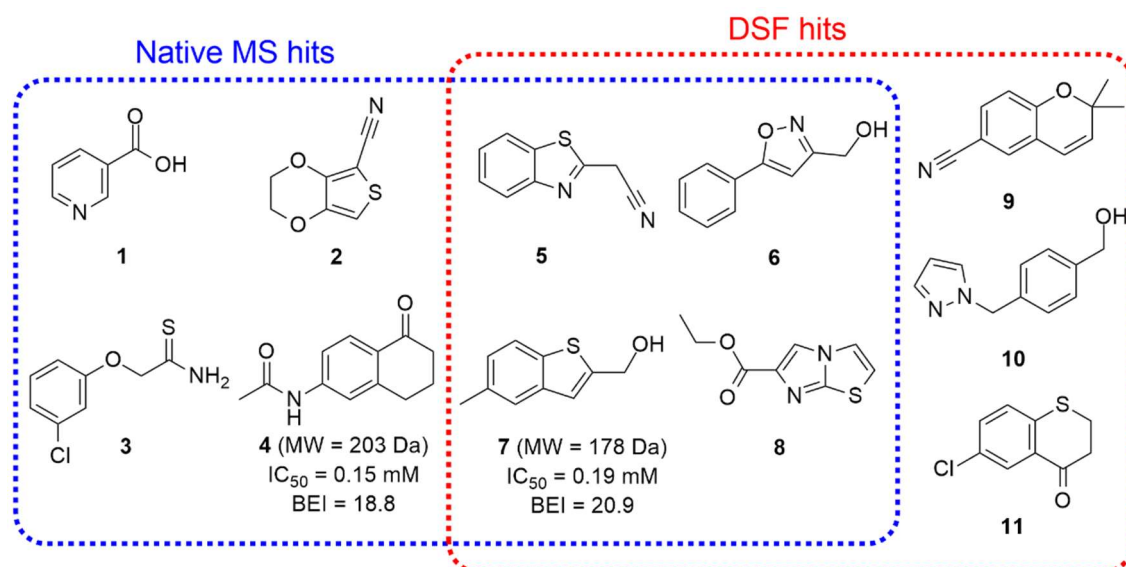


Figure 3.20 Chemical structures of fragment hits identified by native MS (1–8) and DSF (5–11). Native MS hits were defined as compounds that disrupted the EthR-DNA_{R31} complex by >50% at 500 μM while DSF hits increased T_m by an average of at least 0.75 °C at 500 μM. Molecular weight (MW) as well as IC₅₀ and binding efficiency index (BEI) values calculated from the native MS assay are presented for the two most potent fragments, 4 and 7.

A DSF assay was also conducted on the 73 fragments in parallel to allow for a comparison between native MS and DSF. The same concentration of fragment (500 μM) and DMSO concentration (2%) was used. Each fragment was tested twice and fragments that were able to raise the melting temperature of EthR by an average of at least 0.75 °C (1.5 times the standard deviation of vehicle

control T_m values) were classed as hits (**Figure 3.21a**). **Figure 3.20b** shows the individual ΔT_m values of EthR in the presence of the fragments. Seven (compounds **5–11**, **Figure 3.20**) passed this threshold, resulting in a hit rate of 9.6%, with four of these (compounds **5–8**) also appearing as hits in the native MS assay. Screening the entire set of 73 fragments once by DSF consumed about 2 mg of protein and 0.4 mg of total compound, which are about 10 times the amounts used for the native MS assay.

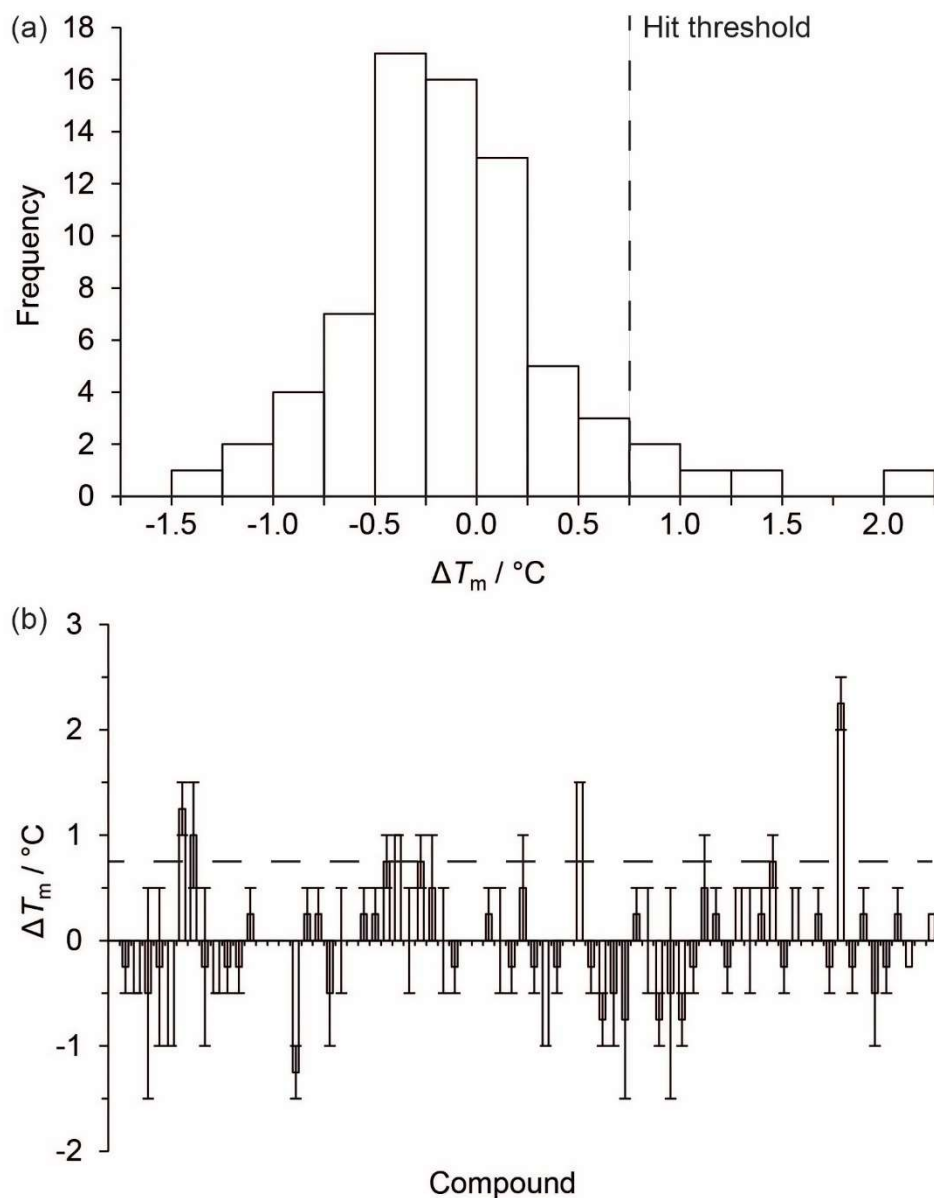


Figure 3.21 Fragment screening against EthR by DSF. (a) Histogram and (b) bar chart of EthR ΔT_m values in the presence of the fragments. The hit threshold ($\Delta T_m \geq 0.75$ °C) is indicated by a dashed line. Error bars indicate the standard deviation of duplicate results. Fragments were tested at 500 μM in Tris buffer (200 mM Tris-HCl pH 8.5, 150 mM NaCl) containing 2% DMSO.

The 11 fragments that appeared as hits in at least one of the primary assays were passed to the fragment validation stage, which was performed by our collaborator Sherine Thomas in the

Blundell group using SPR as previously described.^[158-159] In the SPR assay, the ability of compounds to disrupt the binding of EthR to a biotinylated 106-bp DNA sequence (DNA₁₀₆) encompassing the operator region is measured. The same fragment (500 μ M) and DMSO (2%) concentrations were used to maintain consistency between the different techniques. The most potent compound in the SPR assay was fragment **4**, showing 49% disruption of the EthR-DNA₁₀₆ interaction at 500 μ M, followed by fragment **7** with 47% disruption (**Figure 3.22a**). In general, fragments that appeared as hits in both assays (**5–8**) or in native MS only (**1–4**) showed higher disruption of the EthR-DNA₁₀₆ interaction in SPR compared to fragments that appeared as hits in DSF only (**9–11**). This suggests that native MS was a better predictor of activity in the SPR experiment compared to the DSF assay, which is not surprising because the native MS assay tests for disruption of the EthR-DNA interaction, whereas the DSF assay tests only for binding to EthR. A reasonably high rank correlation ($R = 0.75$) was observed between native MS and SPR disruption values, indicating good correlation between these two techniques (**Figure 3.22b**).

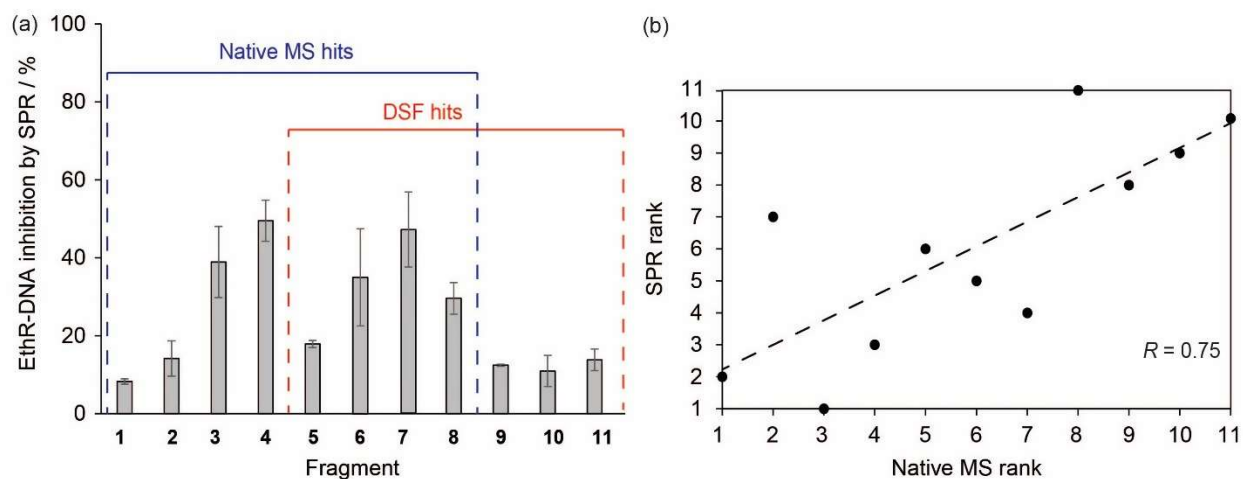


Figure 3.22 Fragment validation by SPR and native MS. (a) Disruption of the EthR-DNA interaction for the fragment hits (500 μ M) as measured by SPR. (b) Scatterplot showing rank correlation between native MS and SPR disruption values.

Further validation for the two most potent fragments was provided by both SPR and native MS dose-response experiments. In the SPR experiment performed by our collaborators, **4** and **7**

showed IC_{50} values of 460 and 610 μM respectively against the EthR-DNA₁₀₆ interaction (**Figure 3.23a**).

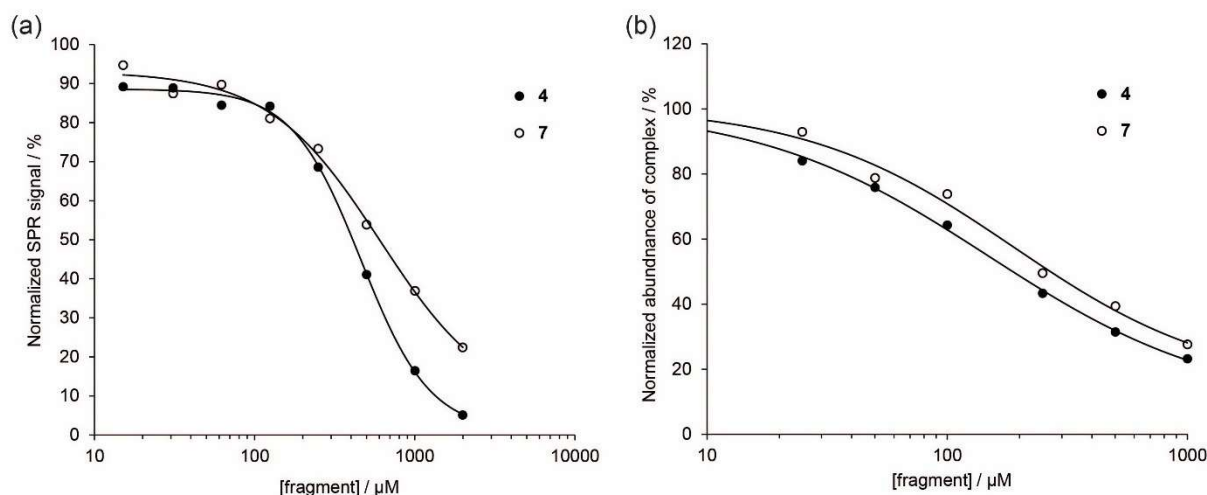


Figure 3.23 (a) Dose-response curves for fragments **4** and **7** against the EthR-DNA₁₀₆ interaction by SPR. Estimated IC_{50} values: 460 μM for **4** and 610 μM for **7**. (b) Dose-response curves for fragments **4** and **7** against the EthR-DNA_{R31} complex by native MS. Estimated IC_{50} values: 150 μM for **4** and 190 μM for **7**.

Meanwhile, in the native MS assay, **4** and **7** disrupted the EthR₄-DNA_{R31} complex with IC_{50} values of 150 and 190 μM respectively (**Figure 3.23b**). Note that these IC_{50} values are only estimates because the peak intensities were not adjusted for differences in response between protein and protein-DNA complexes. Native mass spectra for the EthR₄-DNA_{R31} complex with different concentrations of fragment **4** and **7** are shown in **Figure 3.24a** and **Figure 3.24b** respectively, showing increased disruption of the active complex when the concentration of fragment was increased.

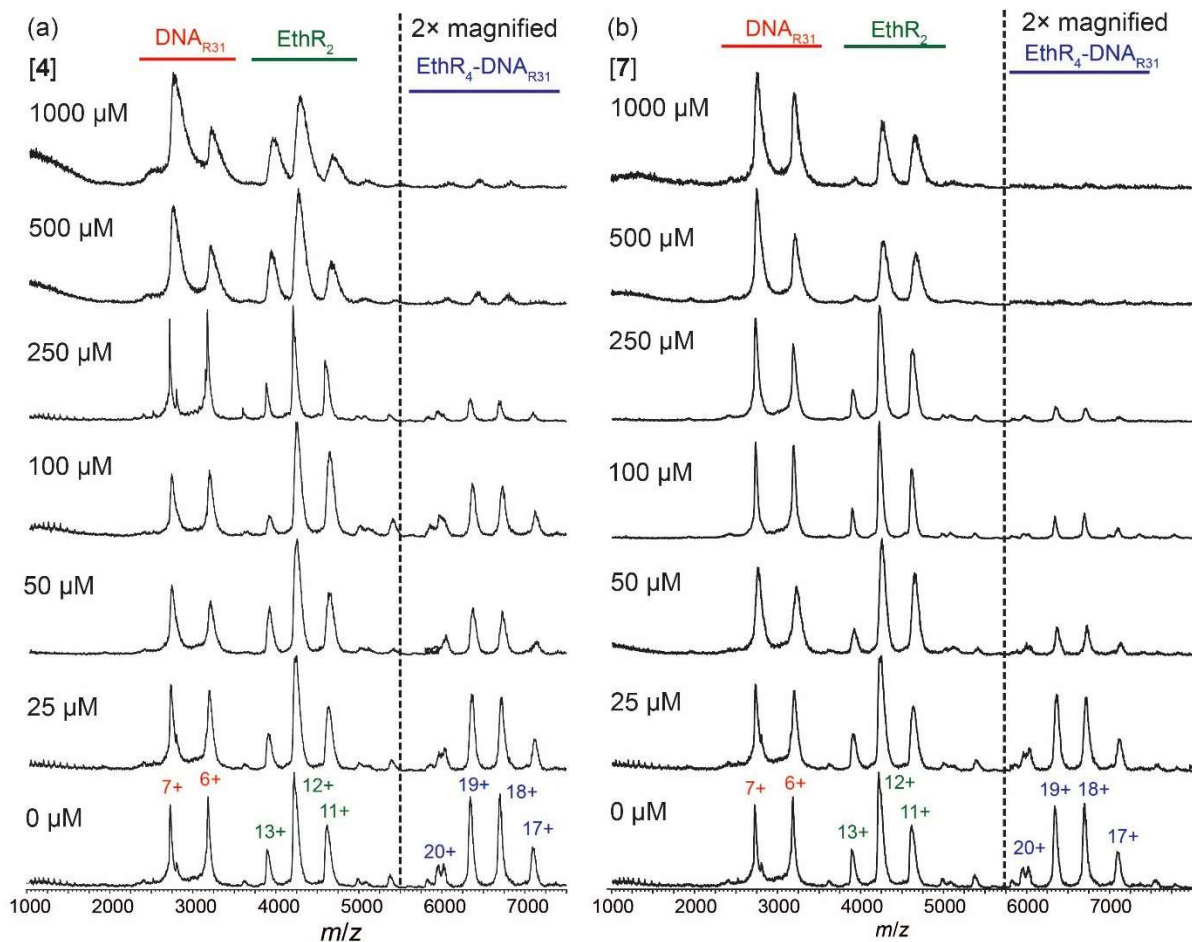


Figure 3.24 Dose-response experiment for (a) fragment **4** and (b) fragment **7** against the EthR-DNA_{R31} complex by native MS. b) Dose-response curves for fragments **4** and **7** against the EthR-DNA_{R31} complex by native MS. Estimated IC_{50} values: 150 μ M for **4** and 190 μ M for **7**.

Binding efficiency index (BEI) values (where $BEI = pIC_{50} / MW$ in kDa)^[170] for **4** (203 Da) and **7** (178 Da) were calculated to be 18.8 and 20.9 using the native MS data, suggesting that they are more ligand-efficient than the optimized lead compound **BDM31381** (359 Da, BEI = 17.7) and might represent fertile starting points for fragment elaboration. To our knowledge, the dihydronaphthalen-1-one core of **4** and the benzothiophene core of **7** have not previously been encountered in reported inhibitors of EthR.

3.2.2.3. Validation of hit fragments using X-ray crystallography

EthR exists as a homodimer in solution, and each monomer is comprised of nine α -helices. The first three α -helices of each monomer comprise the DNA-binding “head”, and the two heads of the dimer must be in close proximity in order for EthR to bind to DNA. However, the binding of inhibitors to a lipophilic channel within the protein forces the DNA-binding heads to move apart by around 18 Å, resulting in a conformation that is incompetent for binding to DNA. EthR inhibitors are therefore allosteric inhibitors of EthR activity.

The binding of the hit fragments to EthR was further validated using X-ray crystallography by our collaborator Dr. Vitor Mendes in the Blundell group (**Table 3.1**). As expected, both fragments bind to the EthR hydrophobic channel. Fragment **4** (PDB: 5MXK) forms hydrogen bonds with side chains of Asn176 and Asn179 but also has π -interactions with Phe110 and Trp207 (**Figure 3.25a**). In contrast, fragment **7** (PDB: 5MWO) binds to two different sub-sites in the EthR channel (**Figure 3.25b**), a characteristic also previously observed with other fragments in EthR.^[157-159] Fragment **7** forms an extended network of π -interactions with Phe114, Phe110, Trp138 and Trp145 in binding mode “**A**” and with Trp103, Trp145, Trp207 and Tyr148 in binding mode “**B**”. Binding mode **B** also forms a hydrogen bond with a water molecule that interacts with the carbonyl group of Met102.

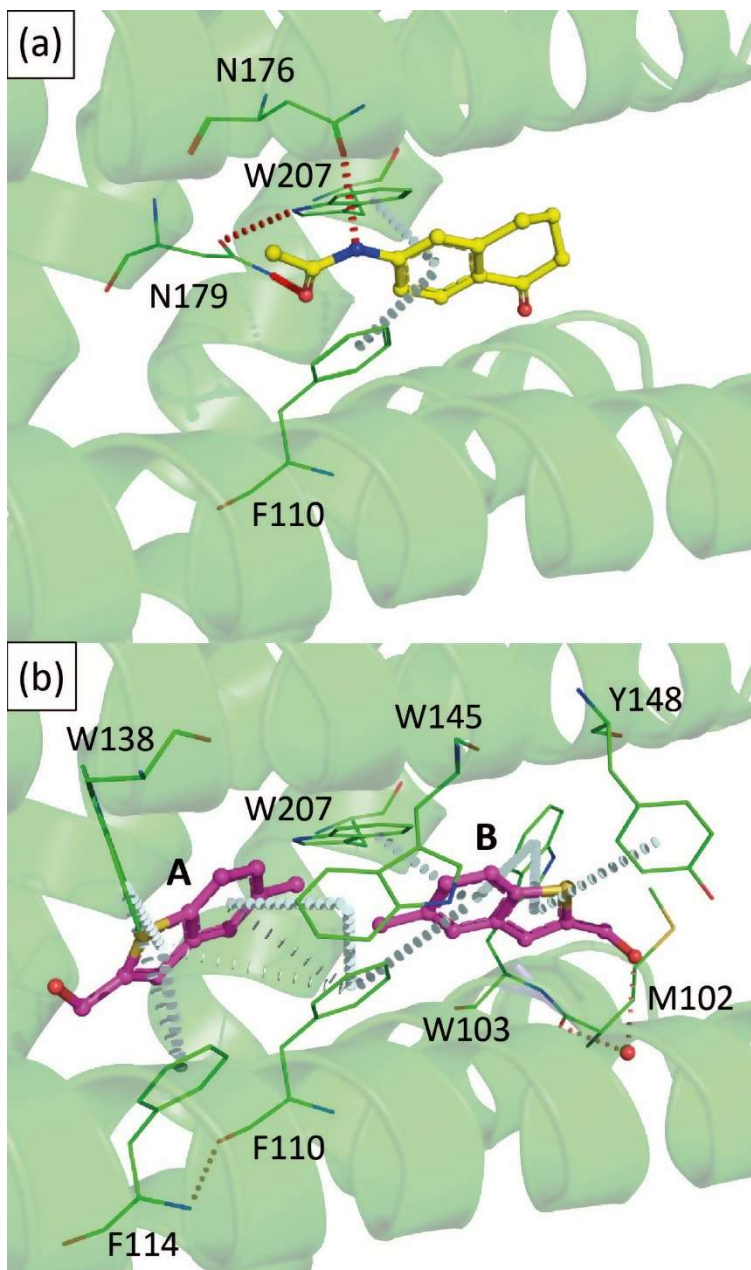


Figure 3.25 X-ray crystal structures of (a) EthR:4 (PDB: 5MXK) and (b) EthR:7 (PDB: 5MWO) complexes. Fragment 7 binds EthR with two different binding modes (“A” and “B”) occupying distinct parts of the binding pocket. Red discs represent hydrogen bonds while gray discs represent π -interactions.

3.2.2.4. Summary

In conclusion, the first application of native MS for fragment screening against a protein-DNA complex has been demonstrated. The sensitivity of the native MS assay was enhanced by screening the fragments against a partial, weaker-affinity EthR-DNA complex. The native MS screening campaign consumed about 10 times less protein and compound compared to DSF, a frequently used primary screen for EthR inhibitors. Fragment hits were validated using SPR, and the results showed that native MS was a better predictor of *in vitro* EthR-DNA disruption activity compared to DSF. Two new fragment scaffolds, a dihydronaphthalen-1-one and a benzothiophene, have been identified that disrupted the EthR-DNA interaction with sub-millimolar potency. X-ray crystallography revealed that both fragments bind to the hydrophobic channel of EthR, thus providing a structural basis for their disruption of the EthR-DNA interaction.

3.3. Experimental

3.3.1. Materials and methods

BDM31381 was synthesized by group member Brendan McConnell following a literature method.^[144] HPLC or PAGE-purified oligonucleotides were purchased from Sigma-Aldrich. Duplex DNA samples were prepared by reconstituting oligonucleotides in NH₄OAc (200 mM, pH 7.0), heating complementary oligonucleotides at 95 °C for 10 min followed by slow cooling to room temperature, and then buffer-exchanging once more into NH₄OAc (200 mM, pH 7.0). The sequences of the oligonucleotides used were (F and R denote forward and reverse complementary sequences, respectively):

- DNA₁₀₆F: 5'- CGG TCA TGG ATC CAC GCT ATC AAC GTA ATG TCG AGG CCG TCA ACG AGA TGT CGA CAC TAT CGA CAC GTA GTA AGC TGC CAG GGT GAC CAC CTC CGC GGC CAG TCA G -3'
- DNA₁₀₆R: 5'- CTG ACT GGC CGC GGA GGT GGT CAC CCT GGC AGC TTA CTA CGT GTC GAT AGT GTC GAC ATC TCG TTG ACG GCC TCG ACA TTA CGT TGA TAG CGT GGA TCC ATG ACC G -3'

- DNA₆₂F: 5' - GAT CCA CGC TAT CAA CGT AAT GTC GAG GCC GTC AAC GAG ATG TCG ACA CTA TCG ACA CGT AG - 3'
- DNA₆₂R: 5' - CTA CGT GTC GAT AGT GTC GAC ATC TCG TTG ACG GCC TCG ACA TTA CGT TGA TAG CGT GGA TC - 3'
- DNA₃₇F: 5'- ATC AAC GTA ATG TCG AGG CCG TCA ACG AGA TGT CGA C-3'
- DNA₃₇R: 5'- GTC GAC ATC TCG TTG ACG GCC TCG ACA TTA CGT TGA T -3'
- DNA₃₆F: 5' - GTA ATG TCG AGG CCG TCA ACG AGA TGT CGA CAC TAT - 3'
- DNA₃₆R: 5' - ATA GTG TCG ACA TCT CGT TGA CGG CCT CGA CAT TAC - 3'
- DNA_{L31}F: 5' - GAT CCA CGC TAT CAA CGT AAT GTC GAG GCC G - 3'
- DNA_{L31}R: 5' - CGG CCT CGA CAT TAC GTT GAT AGC GTG GAT C - 3'
- DNA_{R31}F: 5' - TCA ACG AGA TGT CGA CAC TAT CGA CAC GTA G - 3'
- DNA_{R31}R: 5' - CTA CGT GTC GAT AGT GTC GAC ATC TCG TTG A - 3'

3.3.2. Protein expression and purification

The *Mycobacterium tuberculosis ethR* gene was cloned into a PHAT5 expression vector^[171] having BamHI and EcoRI restriction sites by our collaborator Dr. Vitor Mendes in the Blundell group.

Expression and purification of His₆-EthR protein was performed in this project by following the protocol described in our previous work.^[158] *Escherichia* BL21 (DE3) strain (Novagen) containing the above plasmids were grown overnight at 37 °C in LB media for the primary culture. 25 mL of this culture was used to inoculate 1 L each of fresh LB media containing ampicillin (100 µg/mL) and was grown at 37 °C at 220 rpm until the culture reached an optical density (A_{600nm}) of 0.6-0.8. The expression of the recombinant construct was induced by IPTG (0.5–1 mM) and the culture was further allowed to incubate at 37 °C for 3 h. The cells were harvested by centrifugation (4200 g, 20 min at 4 °C) and resuspended in lysis buffer [50 mM Hepes (pH 7.5), 150 mM NaCl] supplemented with EDTA-free protease inhibitor cocktail (Roche). The resuspended cells were lysed by sonication (Brason) using 10 pulses of 35% amplitude at intervals of 30 s each and the lysate was clarified by centrifugation (26,000 g for 1 h at 4 °C). The supernatant was passed-through a pre-equilibrated (with lysis Buffer) 5 mL HiTrap IMAC Fast

Flow column (GE Healthcare), charged with Ni²⁺, at a flow rate of 3 mL/min. The column was washed with 50ml of wash buffer [50 mM Hepes (pH 7.5), 150 mM NaCl and 20 mM Imidazole] and the bound protein was eluted using elution buffer [50 mM Hepes (pH 7.5), 150 mM NaCl and 250 mM imidazole]. The protein thus obtained was further subjected to size-exclusion chromatography (Superdex 200 16/60, GE Healthcare) and fractions containing pure, homogeneous EthR protein were pooled and concentrated (4500 g at 4 °C) using 10 kDa Amicon Ultra® centrifugal concentrators.

3.3.3. Mass spectrometry

Spectra were recorded on a Synapt HDMS mass spectrometer (Waters) modified for studying high masses (32k quadrupole). Mass spectra were calibrated externally using a solution of cesium iodide (100 mg mL⁻¹). EthR and DNA samples were exchanged into NH₄OAc (200 mM, pH 7.0) solution using Micro Bio-Spin 6 chromatography columns (Bio-Rad). For complexes, buffer-exchanged EthR and DNA were incubated together at the indicated concentrations for at least 15 min at room temperature before measurement. 2.5 µL of sample solution was injected into a borosilicate emitter (Thermo Scientific) for sampling. Instrument conditions were optimized to enhance ion desolvation while minimizing dissociation of macromolecular complexes. Typical conditions were capillary voltage 1.8–2.0 kV, sample cone voltage 100 V, extractor cone voltage 1 V, trap collision voltage 30 V, trap DC bias 35 V, transfer collision voltage 30 V, source temperature 20 °C, backing pressure 5 mbar, trap pressure 3–4 × 10⁻² mbar, IM (N₂) pressure 5–6 × 10⁻¹ mbar and TOF pressure 7–8 × 10⁻⁷ mbar. Data acquisition and processing were performed using MassLynx 4.1 (Waters).

In the fragment screening campaign, the vehicle control experiment (2% DMSO) was performed eight times to establish the baseline for the assay. The responses (percentage of intact EthR₄-DNA_{R31} complex) showed a standard deviation of 12%, indicating acceptable precision. To calculate IC₅₀ values, the native MS data were fit using nonlinear regression to a dose-response curve with variable slope on GraphPad Prism v5.00.

3.3.4. Isothermal titration calorimetry (ITC)

EthR and DNA₆₂ were separately buffer-exchanged into a system of Tris-HCl (10 mM, pH 7.5), NaCl (200 mM), MgCl₂ (2 mM) and EDTA (0.1 mM) using Vivaspin 500 centrifugal concentrators with a 10K MWCO (Sartorius). EthR was titrated into DNA₆₂ in the reaction cell using a MicroCal Auto-iTC200 isothermal titration calorimeter (Malvern). The titration experiment was performed at 25 °C, and consisted of 20 injections of 2 µL every 120 s, with stirring at 1000 rpm. Data were fitted to a bespoke model ($\text{EthR}_2 + \text{DNA}_{62} \rightleftharpoons \text{EthR}_2\text{-DNA}_{62} \rightleftharpoons \text{EthR}_4\text{-DNA}_{62} \rightleftharpoons \text{EthR}_6\text{-DNA}_{62}$) and analysed using the AFFINImeter software (S4SD). Global fitting of independently conducted ITC experiments was performed to improve the robustness of the fit. The experiment was performed twice. In the first experiment, the concentration of EthR₂ was 418 µM and the concentration of DNA₆₂ was 10.7 µM. In the second experiment, the concentration of EthR₂ was 548 µM and the concentration of DNA₆₂ was 15.7 µM. Eva Muñoz (AFFINImeter) is acknowledged for helpful discussion on the ITC analysis.

3.3.5. Differential scanning fluorimetry (DSF)

DSF measurements were carried out using a Bio-Rad CFX Connect machine with a 96-well reaction module. Samples (50 µL each) containing 20 µM EthR, 150 mM NaCl, 20 mM Tris-HCl pH = 8.5, 2.5× SYPRO orange, 2%v/v DMSO and 500 µM fragment were prepared in 96-well plates. The 96-well plates were heated linearly from 25 °C to 95 °C using a temperature increment of 0.5 °C every 30 seconds. Melting curves represent plots of the fluorescence emission intensity at λ_{max} 490/575 nm of each sample against temperature. Melting temperatures (T_m) were determined as the temperatures at which the minima on the negative first derivatives of the melting curves occurred.

3.3.6. Surface plasmon resonance (SPR)

The SPR assay was carried out by our collaborator Sherine Thomas in the Blundell group using a BIAcore T100 instrument as described previously.^[158] The assay was designed to measure the interaction of EthR with the experimental 106-bp ethA promoter DNA sequence (DNA₁₀₆),

immobilized via biotin-streptavidin linkage onto an SA Series S Sensor Chip (BIAcore). DNA from pUC19 (113-bp) was used as a control against non-specific binding. The experimental and control DNA fragments were produced as described previously.^[143-144] Biotinylated experimental and control DNA were immobilized to the chip surface to achieve stable fixation levels of 152 and 150 resonance units (RU) respectively.

For screening, the EthR/fragment solution (1.5 μ M EthR and 0.5 mM concentration of fragment made up in running buffer (2 mM $MgCl_2$, 10 mM Tris-Cl pH 7.5, 0.1 mM EDTA, 200 mM NaCl, 2%(v/v) DMSO) was flowed over the chip at 20 μ L/min for 90 s. The dissociation time was 120 s. To determine binding levels, the response of the control channel at steady state was subtracted from that of the experiment channel. The chip was regenerated between samples using 0.03% w/v SDS in running buffer (passed at a flow rate of 20 μ L/min for 30 s). For IC_{50} calculations, the SPR response of EthR binding to the immobilized DNA was measured at various concentrations of compound. The resulting RUs were used to fit the data using nonlinear regression to a dose-response curve with variable slope using GraphPad Prism v5.00. IC_{50} values were calculated as the compound concentrations necessary to inhibit 50% of the maximal interaction between EthR and its DNA operator sequence.

3.3.7. X-ray crystallography

Crystallization of EthR was performed by our collaborator Dr. Vitor Mendes in the Blundell group using the sitting-drop vapor diffusion method at 25 °C. A drop consisted of 1.0 μ L of reservoir solution (1.7–2.1 M ammonium sulfate, 0.1 M MES-Na (pH 6–7), 5–15%(v/v) glycerol and 7–12%(v/v) 1,4-dioxane) and 0.5–1.0 μ L of protein solution (20 mg/ml EthR, 0.5 M NaCl, 15 mM Tris/HCl pH 8.0 and 10%(v/v) glycerol).^[158] Compounds (100 mM in DMSO) were mixed with mother liquor (1.9 M ammonium sulfate, 0.1 M MES-Na pH 6.5 and 12.5%(v/v) glycerol) to a final concentration of 10 mM. EthR crystals were washed free from 1,4-dioxane by placing them in 1,4-dioxane free mother liquor for a few hours. The washed EthR crystals were then transferred to the fragment-containing solutions and incubated for 16 h. Crystals were cryoprotected by passing them briefly through mother liquor containing 20%(v/v) of ethylene glycol and then flash-

frozen in liquid nitrogen. X-ray crystallographic datasets were collected at the Diamond Light Source (Harwell, UK). X-ray datasets were indexed and integrated using autoPROC,^[172] XDS^[173] and Mosfilm. The scaling of datasets was carried out using SCALA/AIMLESS software.^[174] Structures were solved using the molecular replacement method with PHASER,^[175] included in the PHENIX software package^[176] (PDB: 1T56 was used as search probe). Structures were further refined with PHENIX^[176] to satisfactory level of R/Rfree using maximum-likelihood restrained refinement. Ligand restrain files were prepared with PHENIX elbow software. Every structure was modelled manually in Coot^[177] (including ligand and essential water molecules). Images of X-ray crystal structures in figures were prepared using PyMOL (<http://www.pymol.org>).

Table 3.1 X-ray crystallography data collection and final refinement statistics.

Ligand#	4	7
PDB ID	5MXK	5MWO
Data collection*		
Space group	<i>P4₁2₁2</i>	<i>P4₁2₁2</i>
Cell parameters:		
Resolution range [Å]	121.2 – 1.93	121.93 – 1.96
No. of observations		
R _{merge}	0.148 (1.415)	0.193 (1.479)
CC _{1/2}	0.999 (0.668)	0.998 (0.389)
I/σ(I)	13.7 (1.8)	11.9 (2.0)
Completeness [%]	95.0 (100.0)	98.2 (100.0)
Multiplicity	12.6 (12.4)	12.8 (12.9)
Refinement		

Refinement program	PHENIX	PHENIX
Resolution [Å]	60.59 – 1.93	86.22 – 1.96
No. reflections	18531	18489
R _{work} /R _{free} [%]	18.6/22.3	20.1/25.9
RMS deviations		
Bonds [Å]	0.006	0.007
Angles [°]	0.736	0.795
Ramachandran		
Favoured [%]	98	98
Outliers [%]	0	0

** Parameters shown in brackets are for the highest resolution shell*

4. Application of native mass spectrometry to the characterization and targeting of the bacterial coenzyme A biosynthetic pathway

4.1. Introduction

Coenzyme A (CoA) is a ubiquitous and essential cofactor that is synthesized from pantothenate (Pan, also known as vitamin B₅) through a conserved pathway consisting of five enzymatic steps.^[178] In the second step, phosphopantothenoylcysteine synthetase (PPCS) catalyzes the Mg²⁺-dependent conversion of 4'-phosphopantothenate (PPan) and L-cysteine into 4'-phosphopantothenoylcysteine (PPC) using either CTP (in bacteria) or ATP (in eukaryotes) for activation.^[179] In bacteria, PPan is first coupled with the α -phosphate of CTP to give an acyl cytidylate intermediate, 4'-phosphopantothenoyl cytidylate, with the concomitant release of pyrophosphate (PP_i) (**Figure 4.1a**). The CMP-activated acyl group is then transferred to the amino group of L-cysteine, and PPC and CMP are released.^[179]

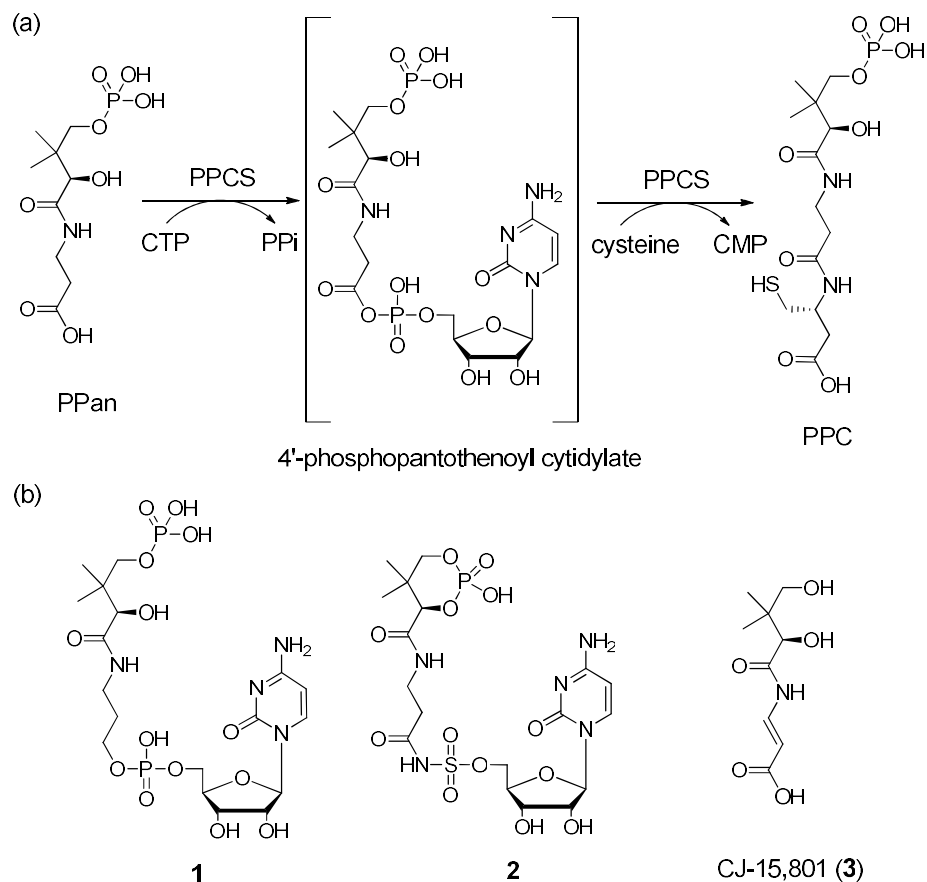


Figure 4.1 (a) Conversion of PPan into PPC catalyzed by PPCS and proceeding through an acyl cytidylate intermediate. (b) Structures of CoaB inhibitors previously reported.

Bacterial PPCS, also referred to as CoaB, is usually fused to the next enzyme of the CoA biosynthetic pathway, phosphopantothenoylcysteine decarboxylase (PPCDC, also referred to as CoaC in bacteria) to form the bifunctional CoaBC protein (*coaBC* gene product, previously known as *dfp*).^[180] The bacterial CoaBC enzyme has been proposed to be a dodecamer based on gel filtration experiments^[181] and by analogy with the related flavoprotein EpiD, for which the X-ray crystal structure has been solved.^[182]

The CoaB domain crystallizes as a homodimer and the residues responsible for dimerization are conserved in all domains of life.^[183] Dimerization is largely mediated by a dimerization domain comprising two antiparallel β -sheets (Thr¹⁹⁸ to Asn²¹⁰ and Ala²⁷⁸ to Lys³⁰⁶ in *Escherichia coli*).^[184]

The latter region, which contains the highly conserved Lys²⁹⁸-X-Lys-Lys (X = Ile in *E. coli*) motif, reaches out to the other protomer as a partially disordered loop that is situated near the entry of the neighboring active site, with Lys²⁹⁸ contacting the γ -phosphate of CTP.^[184]

Kinetic and structural studies have indicated that the nucleobase binding sites of bacterial CoaB enzymes are substantially different from the corresponding human enzyme,^[185] rendering CoaB as a promising antimicrobial target.^[178] The bacterial CoaB enzyme has a strict requirement for CTP, in contrast to human PPCS which uses ATP about four times more efficiently than it does CTP (based on k_{cat}/K_m).^[186] In *E. coli* CoaB (*EcCoaB*), the nucleotide binding pocket is quite rigid and shows strong shape complementarity to the cytosine moiety.^[184] In contrast, human PPCS has an enlarged and also more flexible nucleotide binding pocket that confers preference for ATP, but that still can accommodate the smaller CTP molecule.^[185] Recently, Mizrahi and co-workers have established the essentiality of CoaBC in a murine model of *Mycobacterium tuberculosis* infection.^[187] Silencing of *coaBC* was bactericidal regardless of whether knockdown was initiated at infection or during either the acute or chronic stages of infection, demonstrating that CoaBC is needed *for Mtb* to grow and persist in mice.

The first CoaB inhibitors were reported by Dotson and co-workers in 2009.^[188] These compounds are structural mimics of the acyl cytidylate intermediate that forms during catalysis, except with the reactive acyl-phosphate moiety of the natural intermediate replaced by either a stable phosphodiester or sulfamate isostere (**Figure 4.1b**). These compounds are non-competitive, slow-onset and tight-binding inhibitors that exhibit nanomolar IC₅₀ and K_i values against the CoaB enzymes of *E. coli*, *Enterococcus faecalis*, and *Streptococcus pneumoniae*, as well as up to 1000-fold selectivity over human PPCS. More recently, Burkart, Strauss and co-workers have characterized the natural product CJ-15,801, an antibiotic effective against drug-resistant strains of *Staphylococcus aureus*, as the precursor to a tight-binding inhibitor of CoaB activity.^[189] CJ-15,801 structurally resembles Pan (differing only because it has a *trans*-substituted double bond in the β -alanine moiety), and it can be phosphorylated by the *S. aureus* Pan kinase (PanK) to furnish a PPan-like compound (**Figure 4.1b**) that binds to the CoaB domain. Subsequent

cytidylation by CoaB generates a tight-binding structural mimic of the native acyl cytidylate intermediate with nanomolar potency against the *E. coli* CoaB domain as well as against *S. aureus* CoaBC.^[189] However, despite the CoaB enzyme being characterized over ten years ago, no non-mimetic inhibitors of CoaB have yet been reported in the literature. A better understanding of the structure and interactions of CoaB could aid the development of therapeutic compounds against this promising antimicrobial target.

In this work, native nESI-IM-MS experiments were conducted to provide structural insights into the *E. coli* PPCS (CoaB) domain. Our results indicate that CoaB exists in an equilibrium of monomers and dimers in solution, with dimer formation being favoured by the presence of CTP. Equilibrium titration experiments suggest that the binding of CTP to CoaB is cooperative, and also confirm the preference for *E. coli* CoaB for CTP over ATP. IM-MS analysis characterized the size and unfolding profiles of the EcCoaB monomer and dimer in the gas phase. Finally, native MS was used to screen a small library of sulfonamide compounds against *E. coli* CoaB, revealing a new class of small-molecule, non-mimetic inhibitors of CoaBC.

4.2. Results and discussion

4.2.1. Structural insights into the *E. coli* CoaB domain

4.2.1.1. Native MS

Although the bifunctional *EcCoaBC* enzyme has been suggested by size exclusion chromatography to exist as a dodecamer,^[181] the isolated CoaB domain crystallizes as a homodimer.^[183] Strauss and Begley have shown that the product of the PPCS (CoaB) reaction is released from *EcCoaBC* before it returns to a different active site for transformation by PPCDC (CoaC).^[190] Therefore, the functional significance of the CoaB-CoaC fusion is not currently apparent. As the isolated CoaB domain of *E. coli* is fully functional,^[183] in this study we have used the isolated CoaB domain as a surrogate for CoaBC in binding and mechanistic studies. The dimeric *EcCoaB* enzyme (56 kDa), being far smaller than the dodecameric *EcCoaBC* protein (ca. 500 kDa), is much more amenable to native MS analysis, and easier to express and purify. Doubly-

His₆-tagged *EcCoaB*, containing residues 182 to 406 of the *EcCoaBC*, was overexpressed and purified in a high yield from *E. coli* culture. Protein identity and purity were assessed by denaturing LC-MS (**Figure 4.2**) and SDS-PAGE analysis.

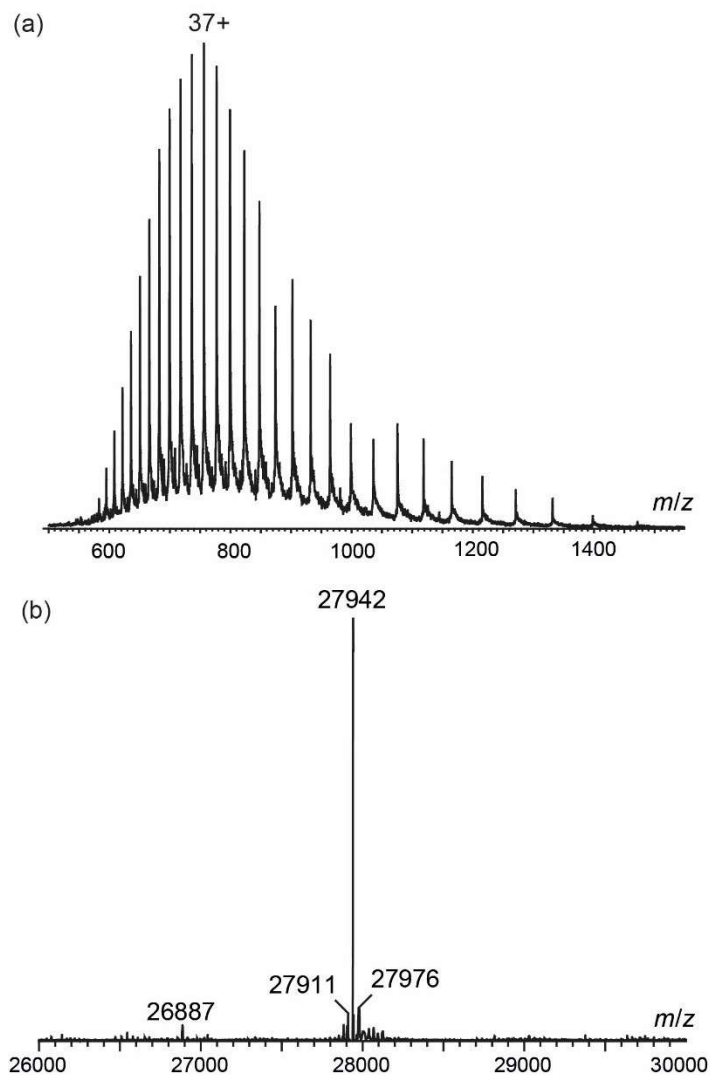


Figure 4.2 (a) Denaturing LC-MS mass spectrum of *EcCoaB*. (b) Deconvoluted LC-MS mass spectrum of *EcCoaB*.

Typical buffers used for protein expression and purification are incompatible with ESI-MS owing to ion suppression as well as the formation of extensive protein-ion adducts. Therefore, *EcCoaB* was first buffer-exchanged into 200 mM ammonium acetate (NH₄OAc) before native MS analysis.

The native mass spectrum of *E. coli* EcCoaB alone showed that EcCoaB existed as a mixture of monomeric and dimeric species (**Figure 4.3a**). The EcCoaB monomer, centered around the 10+ charge state, had an observed mass of 28002 ± 24 Da which was close to the theoretical mass of the protein (without its N-terminal methionine, 27942 Da) and the mass observed by denaturing LC-MS (27942 Da, **Figure 4.2**). Meanwhile, the EcCoaB dimer, centered around the 15+ charge state, had an observed mass of 55999 ± 42 Da, which was also close to the theoretical mass of 55883 Da for two molecules of CoaB. The masses of the EcCoaB monomer and dimer observed by native MS are slightly higher than the theoretical masses presumably due to the adduction of weakly-bound buffer or water molecules to the protein under the soft ionization conditions employed.

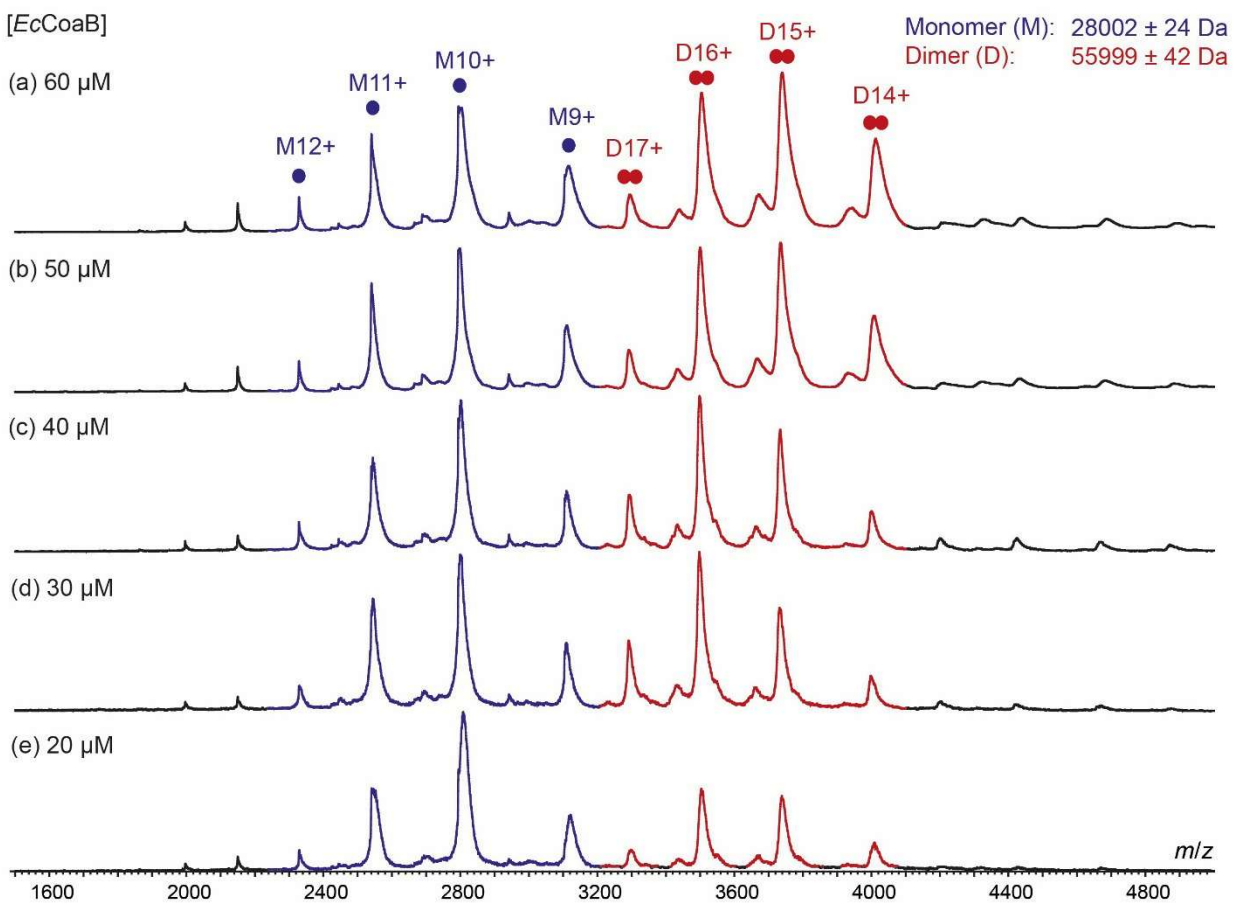


Figure 4.3 The dimerization of EcCoaB is concentration-dependent. Native mass spectra of EcCoaB in 200 mM NH_4OAc at (a) 60 μ M, (b) 50 μ M, (c) 40 μ M, (d) 30 μ M and (e) 20 μ M, with monomer (M) and dimer

(D) charge states indicated. The theoretical mass of the monomer (without its N-terminal methionine) is 27942 Da.

Non-covalent interactions in native MS can be affected by the concentration of NH_4OAc buffer used.^[191] In this study, the effect of NH_4OAc concentration on the appearance of the *EcCoaB* native mass spectrum was investigated. At 50 mM NH_4OAc , higher-order species corresponding to trimers, tetramers, pentamers and hexamers of *EcCoaB* were observed (**Figure 4.4a**). However, the abundance of the higher-order species (trimers and above) was greatly diminished at 200 mM NH_4OAc (**Figure 4.4b**) or 500 mM NH_4OAc (**Figure 4.4c**), suggesting that the oligomers were caused by the non-specific aggregation of *EcCoaB*. This finding is similar to a recent native MS study on CK2 kinase in which the oligomerization state of the protein was also found to be dependent on ionic strength.^[192] The proportion of monomer and dimer species was similar at 200 and 500 mM NH_4OAc . However, the electrospray became more difficult to maintain at 500 mM NH_4OAc , therefore a concentration of 200 mM NH_4OAc was chosen for subsequent experiments.

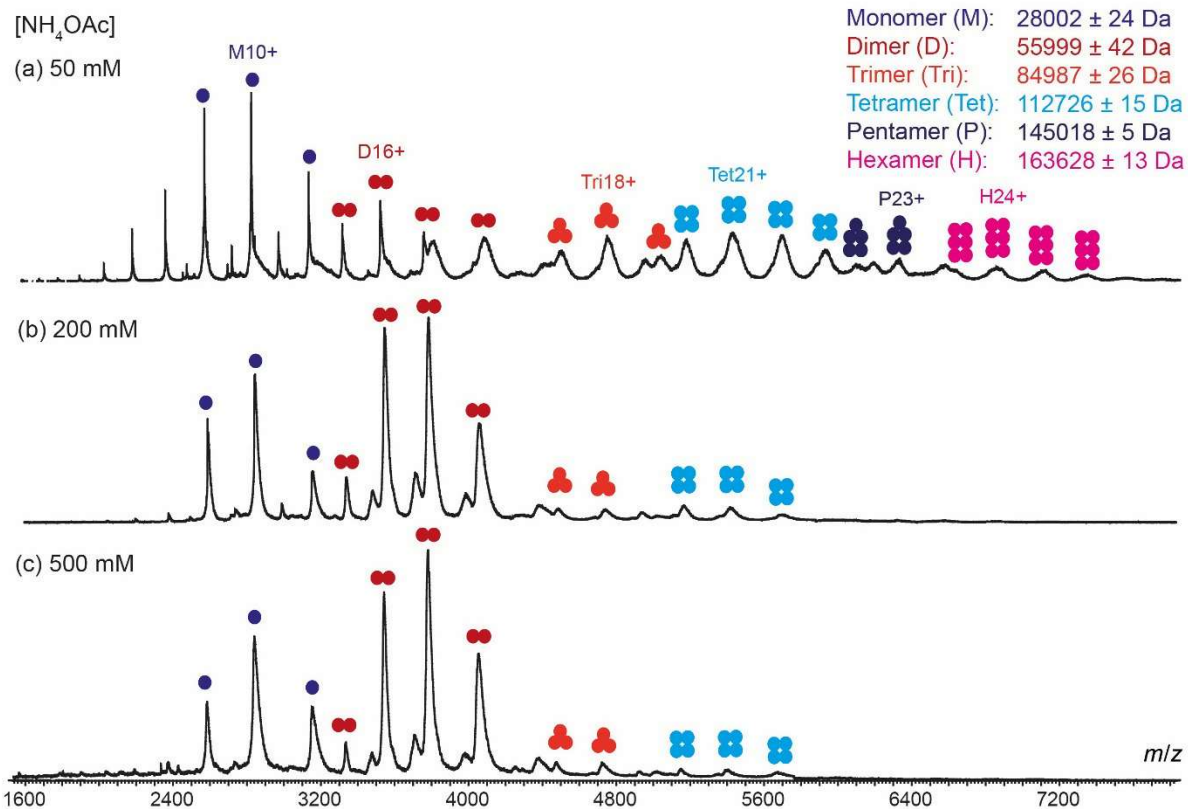


Figure 4.4 The ionic strength affects the oligomerization status of *EcCoaB*. Native mass spectra of *EcCoaB* (30 μ M) in 200 mM NH₄OAc at (a) 50 mM, (b) 200 mM, (c) 500 mM.

4.2.1.2. Determination of the affinity of the *EcCoaB* dimerization

The peak intensities of the different species in the mass spectrum are often considered to be proportional to the concentration of those species in solution. However, this assumes that the probability of ions being ionized, transmitted, and detected is the same for all species, which may not be true for the monomer and dimer of *EcCoaB* given their different sizes. A more accurate equilibrium dissociation constant can be determined by taking into account the fact that the response factors, which are the products of efficiencies of all the processes affecting the signal intensities of the ions, may be different for different species. In this study, the method of Zenobi and co-workers^[193] (based on an earlier work by Gabelica *et al.*^[194]) was used to fit the equilibrium association constant (K_a) of the *EcCoaB* monomer-dimer equilibrium (equation 1) as well as the

ratio (R) between the individual dimer (R_D) and monomer (R_M) response factors (equation 4) from the measured peak ion intensities of the dimer (I_D) and monomer (I_M) according to equation 5 (see Section 4.3.4. for equations).

To derive R and K_a , the native mass spectrum of *EcCoaB* was recorded at different concentrations of the enzyme (**Figure 4.3a-e**). The results show that as the concentration of protein was decreased, the relative abundance of the *EcCoaB* dimer declined in proportion to the monomer, consistent with the monomer and dimer being in equilibrium with each other and not just co-existing. Using nonlinear regression on equation 5, R and K_a were determined to be 1.8 and $3.5 \times 10^4 \text{ M}^{-1}$ (corresponding to $K_d = 29 \text{ }\mu\text{M}$), respectively (**Table 4.1**). **Figure 4.5** shows a graph of experimental I_M/I_D values against total *EcCoaB* monomer concentration (M_0) as well as the fitting curve derived from nonlinear regression of equation 5. The value of R (1.8) is within the range of 0.5 to 6.7 for biomolecules reported in the literature.^[193] It should also be noted that the decrease in apparent dimer formation could also partially be attributed to the reduction of non-specific artefactual oligomers that arise due to droplet crowding.

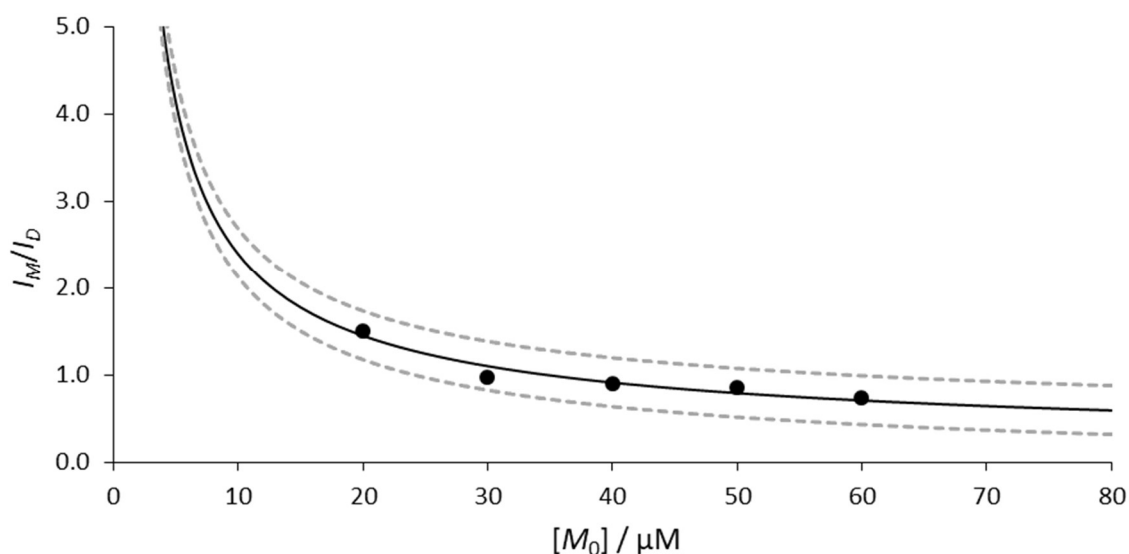


Figure 4.5 Plot of the ratio (I_M/I_D) of the monomer (I_M) and dimer (I_D) peak intensities (filled circles) against the total concentration of *EcCoaB* monomer (M_0) (see **Figure 4.3** for titration experiment). The solid line

represents the fitting curve obtained by non-linear regression of the data using equation 5, from which R and K_a were determined to be 1.8 and $3.5 \times 10^4 \text{ M}^{-1}$, respectively. The dashed lines represent the 95% confidence interval of the regression curve.

Gabelica and co-workers emphasized that a very high quality data set from the titration experiment is required to obtain precise fitting values.^[194] However, similar to observed by Zenobi and co-workers,^[193] the I_M/I_D values showed significant scatter around the fitting curve. Therefore, the confidence intervals that were obtained by fitting the data using unconstrained values of R and K_a were relatively broad (**Table 4.1**, Entry 1). More precise K_a values may be obtained by first constraining the choice of R . With reference to Zenobi and co-workers,^[193] the fitted K_a values when R was restricted to 1, 2 or 4, were determined to 8.6×10^4 , 2.9×10^4 and $1.1 \times 10^4 \text{ M}^{-1}$ respectively (**Table 4.1**, Entries 2–4). The K_a values decrease with R as expected because increasing the value of R (the ratio of response factors between the dimer and the monomer) decreases the estimated dimer to monomer concentration ratio.

Table 4.1 R and K_a values obtained by nonlinear regression of equation 5, using data obtained from the *EcCoaB* titration experiment (Fig. 2). The corresponding K_d values (where $K_d = K_a^{-1}$) are shown to aid interpretation. 95% confidence intervals determined by the Monte Carlo method are shown in parentheses.

Entry	R	$K_a / \times 10^4 \text{ M}^{-1}$	$K_d / \mu\text{M}$
1 ^a	1.8 (0.9–3.0)	3.5 (1.6–5.4)	29 (19–63)
2	1	8.6 (7.7–9.3)	12 (11–13)
3	2	2.9 (2.7–3.1)	34 (32–37)
4	4	1.1 (1.0–1.2)	91 (83–100)

^a R was unconstrained. This gave broader confidence intervals for both R and K_a . In entries 2–4, R was constrained to 1, 2 or 4 respectively to generate more precise estimates for K_a .

There is evidence that the interconversion between the monomeric and dimeric forms of *EcCoaB* is fairly rapid, occurring within the timescale of minutes. First, passage of an *EcCoaB* solution through a Vivaspin 50 kDa molecular-weight cut-off (MWCO) centrifugal concentrator did not lead to significant changes in monomer-to-dimer ratio in either the concentrator compartment or in the filtrate (data not shown). This indicates that the *EcCoaB* dimer (56 kDa) retained in the concentrator could readily dissociate into monomers, while the monomer (28 kDa) that passed through the membrane into the filtrate could easily reassemble into dimers. Secondly, a gel filtration experiment performed by Kupke using a different *EcCoaB* construct (*EcCoaB'*, 26 kDa) indicated a single peak with an apparent molecular weight of 42 kDa.^[183] Although gel filtration is a comparatively low-resolution technique for determining molecular masses, the substantial difference between the observed mass and the expected mass (52 kDa) of the *EcCoaB'* dimer might indicate that *EcCoaB'* was in fact eluting as a mixture of monomers and dimers in rapid equilibrium.

4.2.1.3. Influence of solution pH on the *EcCoaB* protein-protein interaction

Since pH governs the strength of many protein-protein interactions, it was of interest to determine the influence of solution pH on the dimerization of the *EcCoaB* domain. The native mass spectra of *EcCoaB* at pH values of 7.0, 5.0 and 3.0 are shown in **Figure 4.6**.

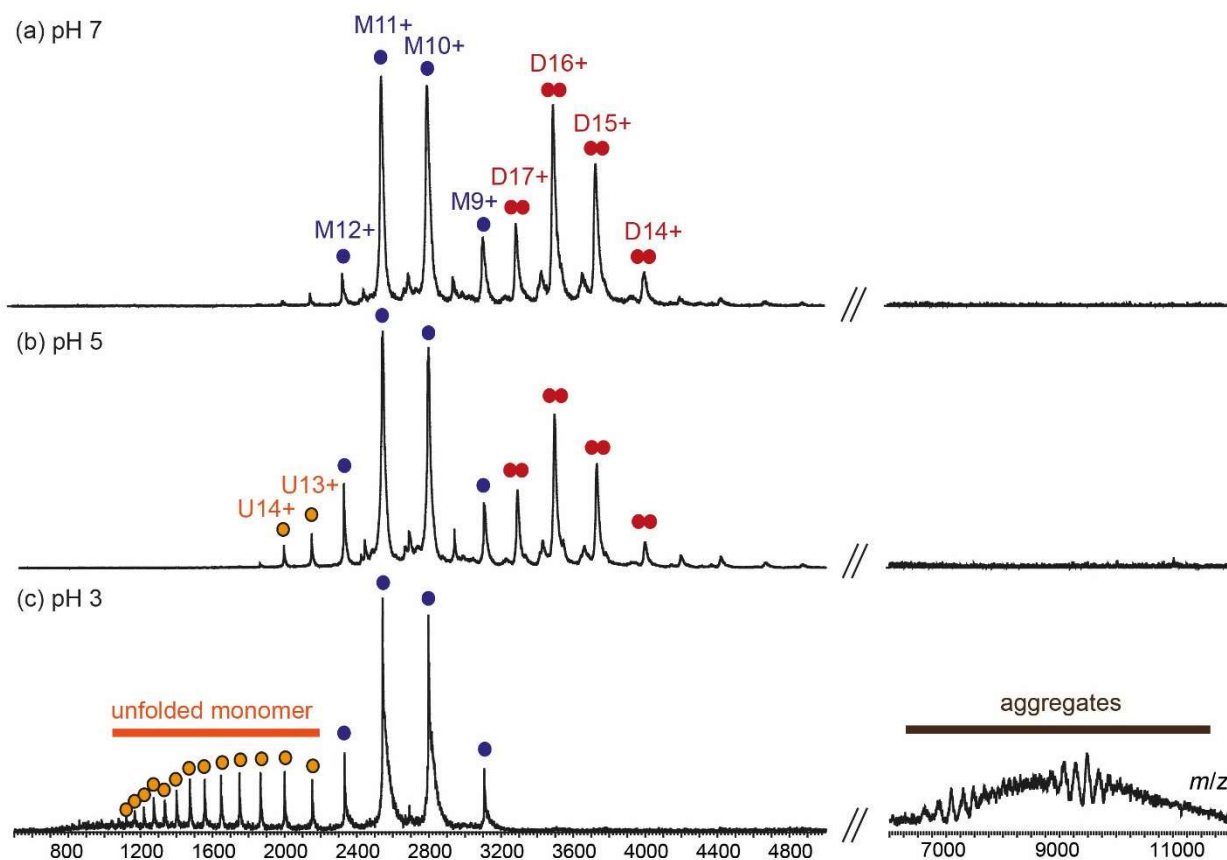


Figure 4.6 Lowering of pH disrupts *EcCoaB* dimerization and partially unfolds the protein. Native mass spectra of *EcCoaB* (30 μ M) in 200 mM NH_4OAc at (a) pH 7, (b) pH 5, and (c) pH 3, with monomer (M), dimer (D) and unfolded monomer (U) charge states indicated. For clarity, charge states for the unfolded monomer are not assigned in (c). The molecular masses of the aggregates in (c) were unable to be assigned. Note the change in x-axis scale in the higher m/z region.

The results show that the native mass spectra of *EcCoaB* at pH 7 (**Figure 4.6a**) and pH 5 (**Figure 4.6b**) were broadly similar, except for a slight decrease in *EcCoaB* dimer abundance and the appearance of higher charge states of the monomer (13+ and 14+), suggesting the partial unfolding of the protein. However, the appearance of the spectrum was dramatically changed at pH 3 (**Figure 4.6c**). The dimer completely disappeared, and instead, a series of highly-charged monomeric *EcCoaB* ions were observed accompanied by the appearance of aggregates in the higher m/z region. All of these features are indicative of the chemical denaturation of *EcCoaB*. As the protein unfolds, its surface area increases which allows for a greater accumulation of positive

charge on the molecule during the electrospray process. The exposure of interior lipophilic residues could also allow the protein to interact via hydrophobic interactions to form higher-order aggregates. Additionally, the lowering of the pH is likely to disrupt polar interactions that are important for dimerization,^[184] leading to the disruption of the *EcCoaB* protein-protein interface even while some of the monomer remains natively folded (as indicated by the intact charge envelope encompassing the 9+ to 12+ charge states of the monomer in **Figure 4.6c**). This suggests that the tertiary structure of the *EcCoaB* monomer is relatively more resistant to acid-induced denaturation than the quaternary interactions between the protomers in the dimer.

4.2.1.4. Ion mobility-mass spectrometry of *EcCoaB*

IM-MS experiments were performed to investigate the structural characteristics of *EcCoaB* in the gas phase. CCS values for individual charge states of the *EcCoaB* monomer and dimer at either 0 or 5%(v/v) DMSO are presented in **Table 4.2**. To ensure that the CCS values reflect the dimensions of the folded state, IM-MS experiments were recorded at a collision voltage (20 V) that precedes the onset of unfolding (see below) in order to minimize protein unfolding, which would impact CCS values. CCS values generally increase with charge state, which could be due to the increased internal electrostatic repulsion of protein ions at higher charge. At 0% DMSO, the average CCS values of the two lowest (most native-like) charge states of the *EcCoaB* monomer and dimer ions are 2430 and 3985 Å², respectively. For comparison, the theoretical CCS_{calc} of the *EcCoaB* dimer calculated from the X-ray crystal structure (PDB: 1U7U)^[184] was 4258 Å². This suggests that the *EcCoaB* dimer remains natively folded in the gas-phase.

Table 4.2 CCS values of *EcCoaB* monomer and dimer ions as determined by IM-MS at 0 and 5%(v/v) DMSO. The average CCS values of the two lowest (most native-like) charge states of the monomer and dimer are also presented. Errors in CCS are estimated to be 3%.^[61, 79]

Monomer			Dimer		
	0% DMSO	5% DMSO		0% DMSO	5% DMSO
Charge	CCS / Å ²		Charge	CCS / Å ²	
8+	-	2307	11+	-	3674
9+	2358	2407	12+	-	3747
10+	2502	2542	13+	-	3841
11+	2755	-	14+	3938	3945
12+	3028	-	15+	4031	-
			16+	4199	
			17+	4442	
Average	2430	2357		3985	3711

In the presence of 5%(v/v) DMSO, the charge state distributions of both the *EcCoaB* monomer and dimer shifted to lower charge states (**Figure 4.7**). This is accompanied with a compaction of the protein structure (**Table 4.2**), and is consistent with previous studies on the effects of DMSO on protein size both in solution and in the gas phase.^[105, 127, 195] However, the relative abundance of monomers and dimers was not substantially changed by the presence of DMSO.

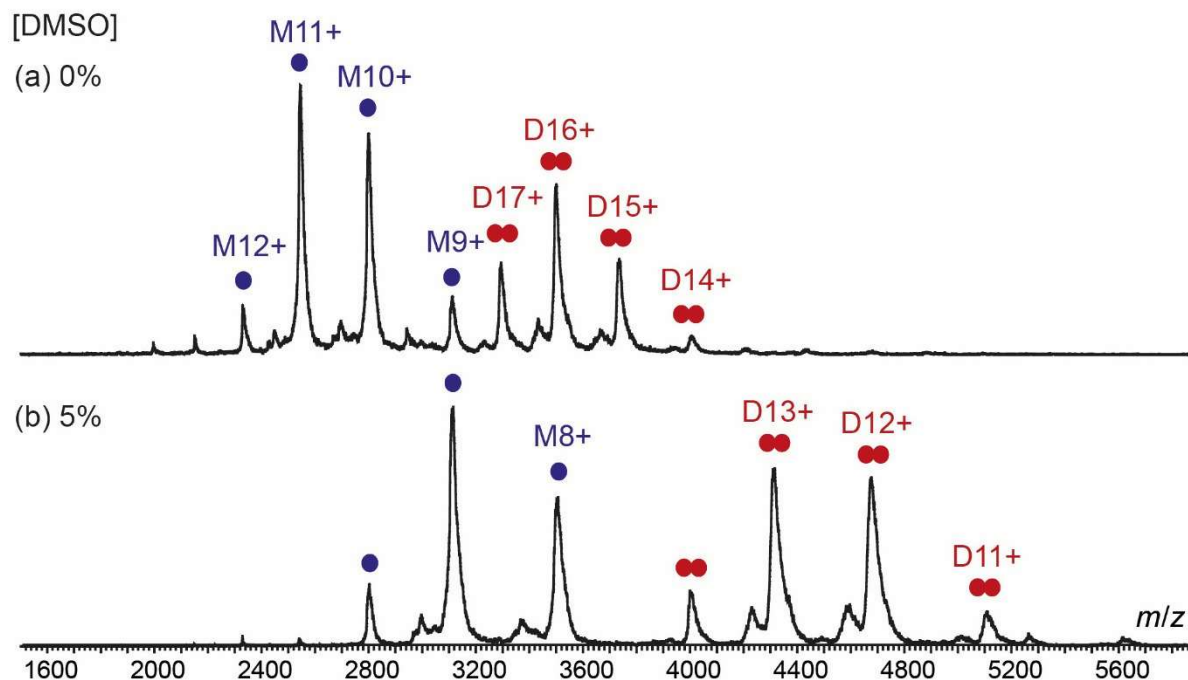


Figure 4.7 DMSO shifts charge state distributions of *EcCoaB* to lower charge states. Native mass spectra of *EcCoaB* (30 μ M) in 200 mM NH_4OAc in the presence of (a) 0% or (b) 5% DMSO(v/v).

4.2.1.5. Collision-induced unfolding of *EcCoaB*

The use of collisional activation to generate unfolding plots (or “fingerprints”) is a recently developed technique that has been used to study the gas-phase stability of proteins.^[52-53] Collision-induced unfolding (CIU) fingerprints are obtained by recording the native mass spectra of a protein at incrementally higher acceleration voltages within the collision cell of a mass spectrometer. As protein ions unfold through collisions with neutral gas molecules, changes in their CCS can be tracked as a function of collision voltage, allowing the quantitation of a protein’s resistance to unfolding in the gas phase. CIU IM-MS experiments have been used, for example, to evaluate the stabilizing effects of ligands on protein stability,^[53] to differentiate between Type I and Type II kinase inhibitors of Abl kinase,^[196] or to characterize differences between complexes formed between Aurora kinase C and different mutants of INCENP.^[197]

The CIU fingerprints for *EcCoaB* monomeric (**Figure 4.8a**) and dimeric (**Figure 4.8b**) ions are complex and reveal multiple unfolding transitions over the range of collision voltages tested. For

instance, for the 9+ state of the monomer, the native-like protein (**N**, $\text{CCS} \approx 2400 \text{ \AA}^2$) initially transitions into a partially unfolded intermediate form (**I**, $\text{CCS} \approx 2800 \text{ \AA}^2$) at moderate collision voltages, then shifts into a more extended form (**E**, $\text{CCS} \approx 3100 \text{ \AA}^2$) as the collision energy is increased further. It is assumed that **E** represents the most unfolded form of the monomeric 9+ ion as populations with larger CCS values were not observed even as the collision voltage was increased to 150 V, at which point protein degradation also began to occur (data not shown). The CIU50 values of the native-like (**N**) and intermediate (**I**) forms of the monomeric 9+ ion (*i.e.* the collision voltage needed to deplete 50% of the population of the conformation) were determined to be 34 and 62 V, respectively. For the other monomeric and dimeric *EcCoaB* charge states, multiple intermediate species were observed in the CIU plots, indicative of complicated unfolding pathways for those ions. The observation that the CIU unfolding profiles for monomeric and dimeric *EcCoaB* are broadly similar suggests that an asymmetric unfolding mechanism might be operative for the dimer (*i.e.*, one subunit of the dimer unfolds first).

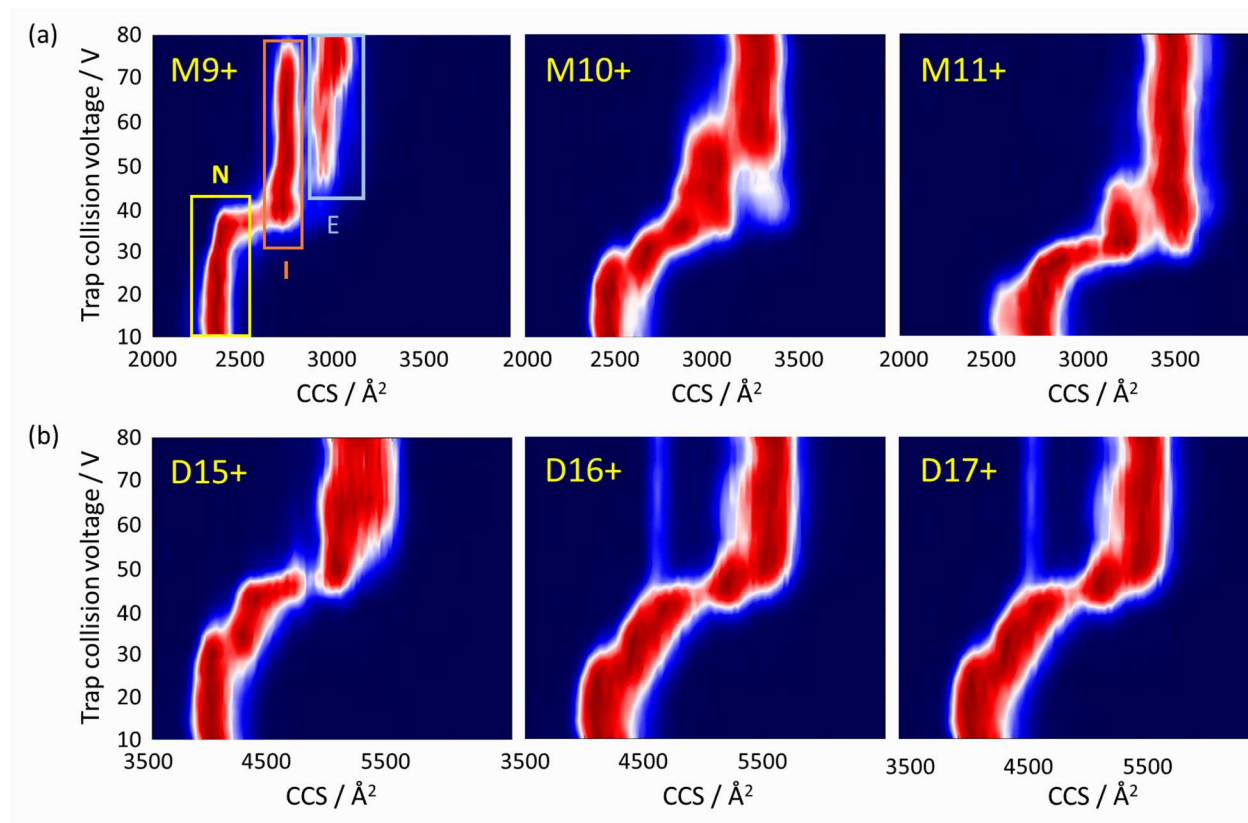


Figure 4.8 CIU IM-MS reveals gas-phase unfolding transitions of *EcCoaB*. CIU fingerprints of indicated charge states of the *EcCoaB* (a) monomer (M) and (b) dimer (D). For the 9+ state of the monomer, native-like (N), intermediate (I) and extended (E) forms of the protein are labeled.

CIU plots for *EcCoaB* monomeric 9+ and 10+ states are broadly similar in the presence of 5% DMSO(v/v) (**Figure 4.9a**) compared to the respective plots in the absence of DMSO. The 8+ state of the monomer, accessible only via charge reduction conditions with DMSO, did not unfold even at the highest collision voltage of 80 V tested (although we do not preclude the possibility that the monomer unfolds at an even greater voltage). The greater CIU stability of lower charge states is known^[136, 195] and can be attributed to the lesser internal electrostatic repulsion experienced by proteins of lower charge. For the *EcCoaB* dimer in the presence of DMSO, the 11+ state did not unfold, while the 12+ and 13+ states underwent one and three unfolding transitions respectively over the collision voltage range tested (**Figure 4.9b**).

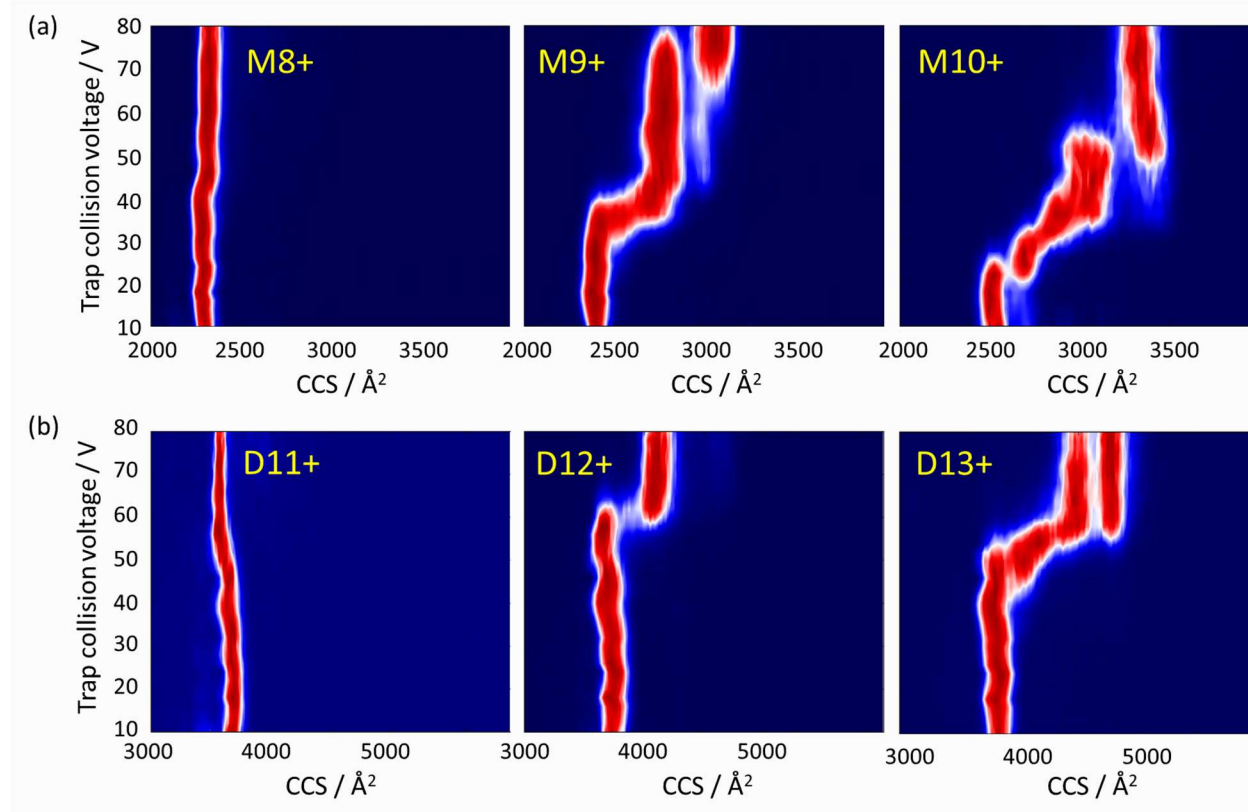


Figure 4.9 CIU IM-MS of *EcCoaB* in the presence of 5% DMSO. CIU plots of indicated charge states of the *EcCoaB* (a) monomer (M) (8+ to 10+) and (b) dimer (D) (11+ to 13+) with 5% DMSO(v/v).

4.2.1.6. Summary

In conclusion, native nESI-IM-MS has been conducted to provide structural insights into the *EcCoaB* domain. While *EcCoaB* crystallizes as a dimer, our experiments demonstrate that in solution, an equilibrium exists between the monomer and dimer. The gas-phase unfolding behavior of monomeric and dimeric *EcCoaB* ions was studied by IM-MS, revealing complex CIU fingerprints for most charge states.

4.2.2. Assessing *EcCoaB*-ligand interactions by native MS

4.2.2.1. Interaction of *EcCoaB* with CTP

Since *EcCoaB* utilizes CTP to activate the carboxylate group of PPan, the interaction between CTP and *EcCoaB* was studied by native MS. Interestingly, the addition of 50 μM CTP to *EcCoaB* (25 μM) (**Figure 4.10b**) increased the proportion of dimeric *EcCoaB* compared to in the absence of CTP (**Figure 4.10a**). With CTP present, the K_D value for the monomer-dimer equilibrium was estimated to be 4 μM , which is substantially lower than corresponding K_D value of 29 μM determined in the absence of CTP. This result indicates that CTP stabilizes the dimeric form of *EcCoaB*, and is consistent with the observation that CTP contacts both subunits of the dimer in the X-ray co-crystal structure of the *EcCoaB* domain.^[184]

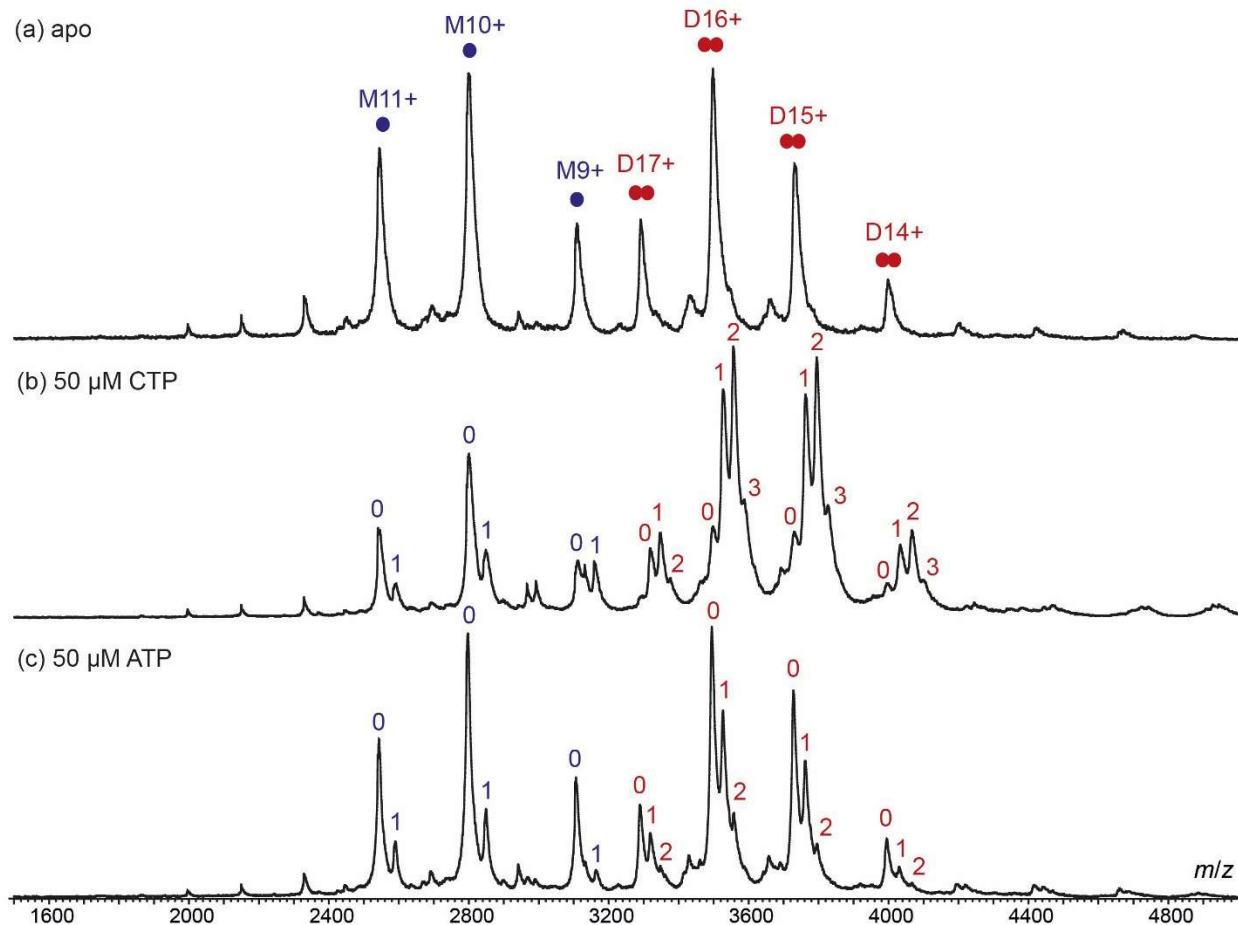


Figure 4.10 *EcCoaB* selectively binds CTP over ATP, and CTP stabilizes the dimer. Native MS of *EcCoaB* (25 μM) in the (a) absence, or presence of (b) CTP (50 μM) or (c) ATP (50 μM). In (a), the monomer (M) and dimer (D) charge states are indicated. In (b) and (c), the number of ligands bound to each monomer or dimer species is indicated.

A titration experiment was performed to assess the effect of adding different concentrations of CTP. The results showed that the proportion of dimer was further increased at higher concentrations of CTP (**Figure 4.11a-e**). At 200 μM CTP, only a small amount of the monomer could be detected (**Figure 4.11e**).

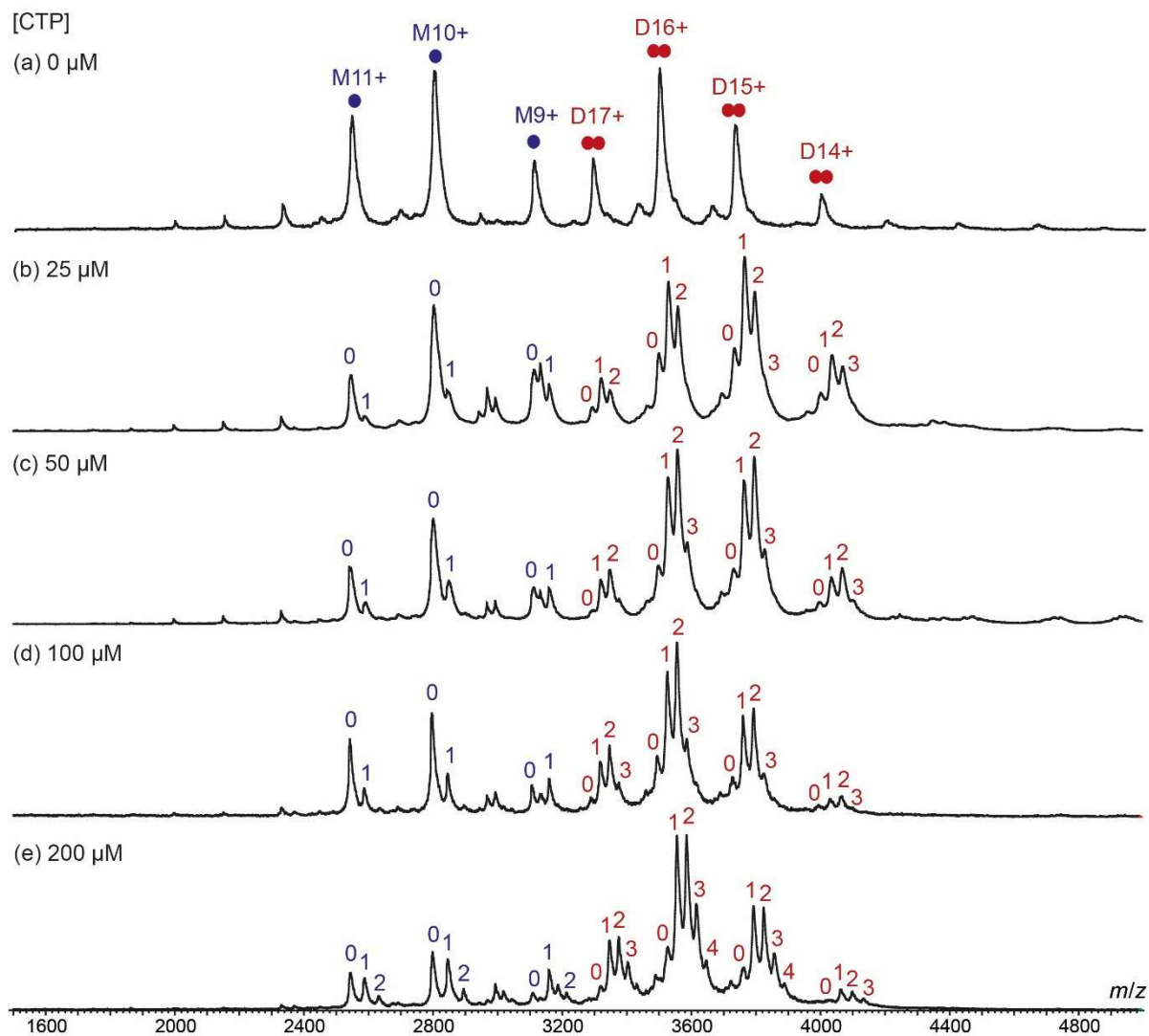


Figure 4.11 Titration of CTP into EcCoaB reveals non-specific binding at higher CTP concentrations. Native mass spectra of EcCoaB (30 μM) in the presence of (a) 0 μM , (b) 30 μM , (c) 50 μM , (d) 100 μM and (e) 200 μM of CTP. In (a), the monomer (M) and dimer (D) charge states are indicated. In (b)–(e), the number of CTP molecules bound to each monomer or dimer species is indicated.

At 50 μM CTP, 66% of the EcCoaB monomer was unbound while 34% was singly-bound by CTP (**Figure 4.10b**). In contrast, the majority of the dimer was observed in singly-bound (32%) or doubly-bound (38%) states. This can be rationalized on the basis of the fact that the CTP-binding site of EcCoaB is formed from residues from both subunits of the dimer (**Figure 4.12**).^[184]

Therefore, the binding affinity of CTP for the monomer would be expected to be weaker than to the dimer.

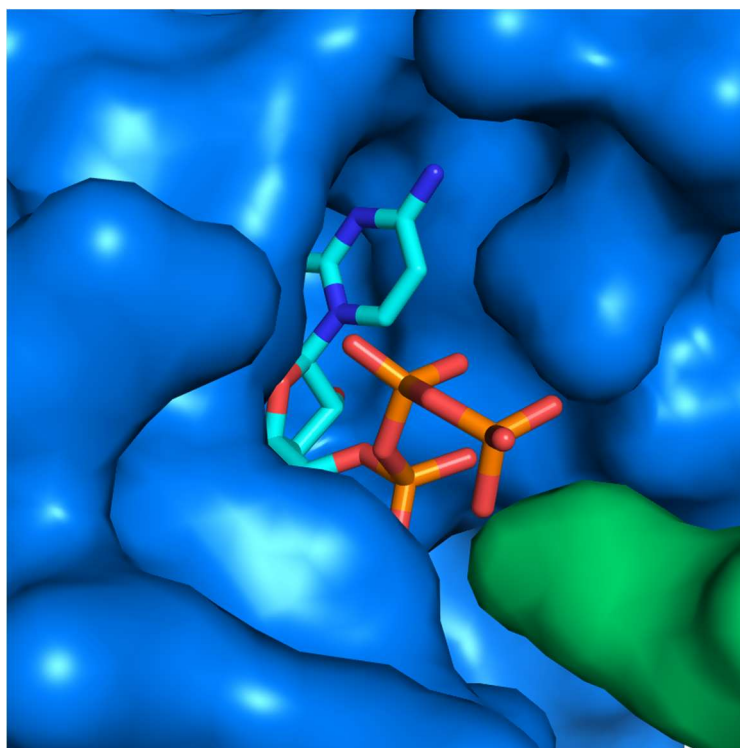


Figure 4.12 The CTP binding pockets of the *EcCoaB* dimer are formed from residues of both protomers. Protomer A and B of the CTP-bound *EcCoaB* dimer (PDB ID: 1U7W)^[184] are shown in blue and green, respectively, with CTP shown with its carbons in cyan.

4.2.2.2. Accounting for non-specific binding

A small amount (17%) of the *EcCoaB* dimer was bound to three molecules of CTP (**Figure 4.10b**). As each dimer only has two CTP-binding sites, this suggests that the third molecule of CTP is bound non-specifically to the protein. The observation that non-specific binding is observed at only a mild excess of CTP over protein (50 μM of CTP vs. 25 μM of *EcCoaB*) could possibly be accounted for by the fact that electrostatic interactions, which are presumed significant between the doubly-anionic CTP molecule and the positively-charged *EcCoaB* protein, are generally thought to be strengthened in the gas phase. In previous studies, a high degree of non-specific binding has also been observed in native MS experiments between creatine kinase and ATP,^[198]

or between GroEL and ATP.^[199] With *EcCoaB*, non-specific binding was even more pronounced at higher CTP concentrations, with up to four CTP molecules bound per dimer at 200 μM of CTP, and two molecules of CTP bound per monomer (**Figure 4.11e**).

A number of methods have been developed to tackle the issue of non-specific binding in native ESI-MS experiments. van der Rest and co-workers^[198] developed a deconvolution method that modeled the specific binding as a binomial distribution and the non-specific binding as a Poisson distribution (originally proposed by Klassen and co-workers).^[200] Alternatively, Guan and co-workers have modeled the non-specific component as a power-law distribution,^[201] while in yet another approach, the group of Sharon and Horovitz have modeled all non-specific sites with a single binding constant.^[202]

In this study, to isolate the effect of non-specific binding in this system, the interaction between *EcCoaB* and ATP was monitored using native MS. The native mass spectrum of *EcCoaB* with ATP (50 μM) is shown in **Figure 4.10c**. As expected, ATP showed substantially reduced ability to increase the proportion of *EcCoaB* dimer compared to with CTP. While the presence of 50 μM CTP lowered the K_d value for the monomer-dimer equilibrium from 29 μM to 4 μM (see above), the presence of an equivalent amount of ATP only reduced the K_d value to 19 μM , which is within the 95% confidence interval for the K_d value for the apo protein (19 to 63 μM). This indicates that ATP does not stabilize the *EcCoaB* dimer as effectively as CTP, presumably because it is too large to enter into the nucleobase binding pocket of the enzyme. Moreover, ATP showed weaker binding to the dimer compared to CTP, as reflected by the lower proportion of ligand-bound protein for ATP compared to CTP at the same concentration. Based on the findings of Strauss *et al.*,^[180] which showed that *EcCoaB* enzyme activity was not promoted even at ATP concentrations as high as 10 mM, we assumed that *EcCoaB*-ATP complexes observed by native MS were non-specific. With this assumption, the specific binding component of CTP can be estimated by subtracting the occupancy of ATP from the occupancy of CTP at each site of the protein. Applying this method to the data shown in **Figure 4.10** revealed that the specific binding constant of CTP to the monomer is 380 μM , while the average specific binding constant of CTP to the two sites of

the dimer was 70 μM . The latter affinity is composed of a specific binding constant of 580 μM for the first CTP molecule to the dimer, and a specific binding constant of 9 μM for the second CTP molecule to the dimer, which indicates some degree of positive cooperativity in the binding. The average specific binding constant (70 μM) is also consistent with the K_M value of *EcCoaB* for CTP, which has been reported to be 48 μM .^[189]

4.2.2.3. Screening of *EcCoaB* inhibitors by native MS

Due to its ability to detect protein-ligand interactions, native MS is finding increasing use as a drug screening technique, including for fragment-based drug discovery (FBDD).^[44-46] In line with the continuing interest within the Abell research group in the application of FBDD methods to develop new antimicrobial compounds, native MS was applied to screen a small library of fragment-sized sulfonamide compounds against *EcCoaB*. These analogues were designed based on the structure of the sulfonamide fragment **1**, which was identified as a hit against *Mtb* CoaBC using a combination of differential scanning fluorimetry and ligand-based NMR studies. The library analogues **2-11** (synthesized by Dr. Jeannine Hess and Robert Starley, Abell research group), retain the sulfonamide core of fragment **1**, but have simpler aromatic motifs in place of the synthetically challenging aniline and pyrimidine moieties of **1** (**Figure 4.13**).

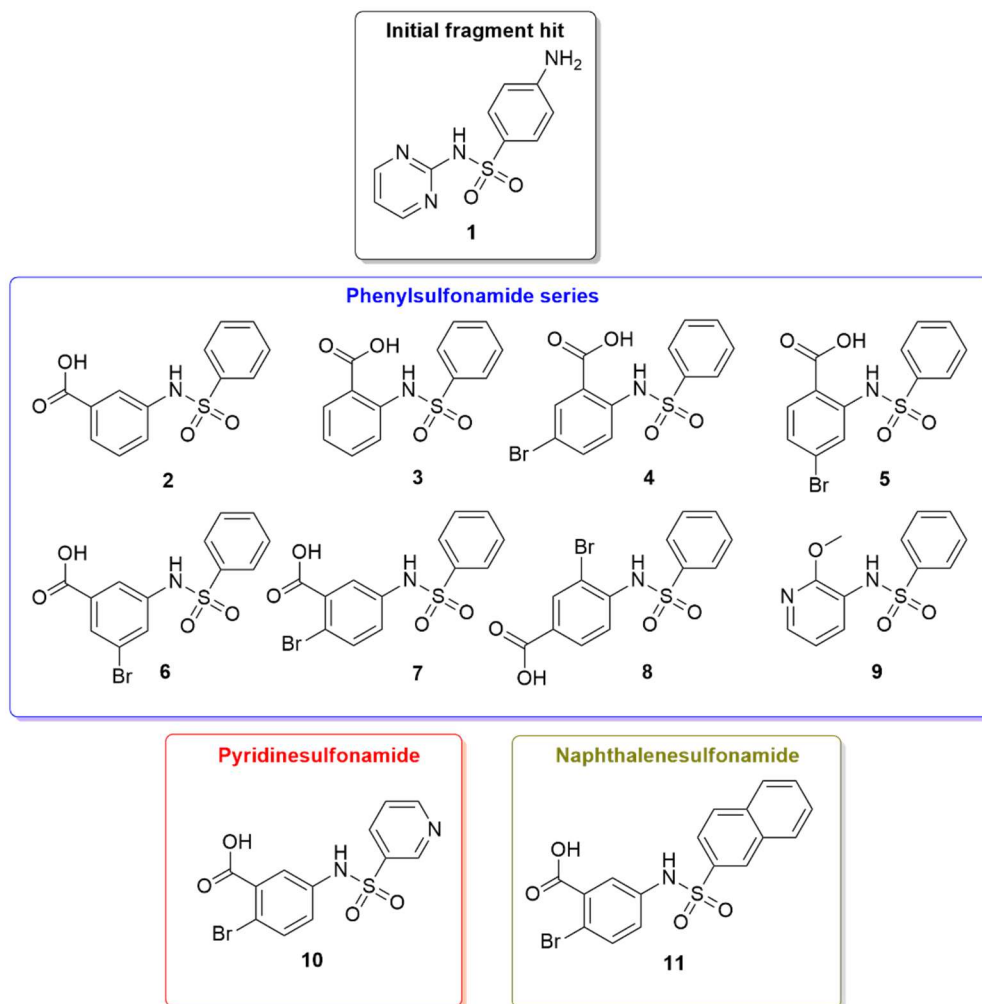


Figure 4.13 Structures of sulfonamide compounds **1-11** examined in this work.

Additionally, the structure-activity relationships (SAR) around the phenyl substituent attached to the sulfonamide nitrogen atom was investigated by the installation of a carboxyl group and/or a bromine atom at various positions of the ring (except for compound **9**, which carries a 2-methoxypyridin-3-yl substituent in place of the phenyl ring). All of the library compounds (with the exception of naphthalene compound **11**) have either 19 and 20 heavy atoms, putting them close to the upper limit in size of typical fragment-sized molecules.

To minimize non-specific binding, the compounds were tested at a concentration of 30 μM with *Ec*CoaBC at 15 μM , giving only a 2:1 excess of ligand to protein. The native mass spectra of

EcCoaBC incubated with the compounds **1-11** in the presence of 1% (v/v) DMSO are shown in **Figure 4.14**. In general, the compounds did not increase the ratio of the dimer to the monomer, indicating that unlike CTP they do not stabilize the *EcCoaB* dimer. This result is not surprising as the compounds are significantly smaller than CTP and thus less likely to be able to simultaneously contact both subunits of the dimer, even if they were binding to the CTP-binding site.

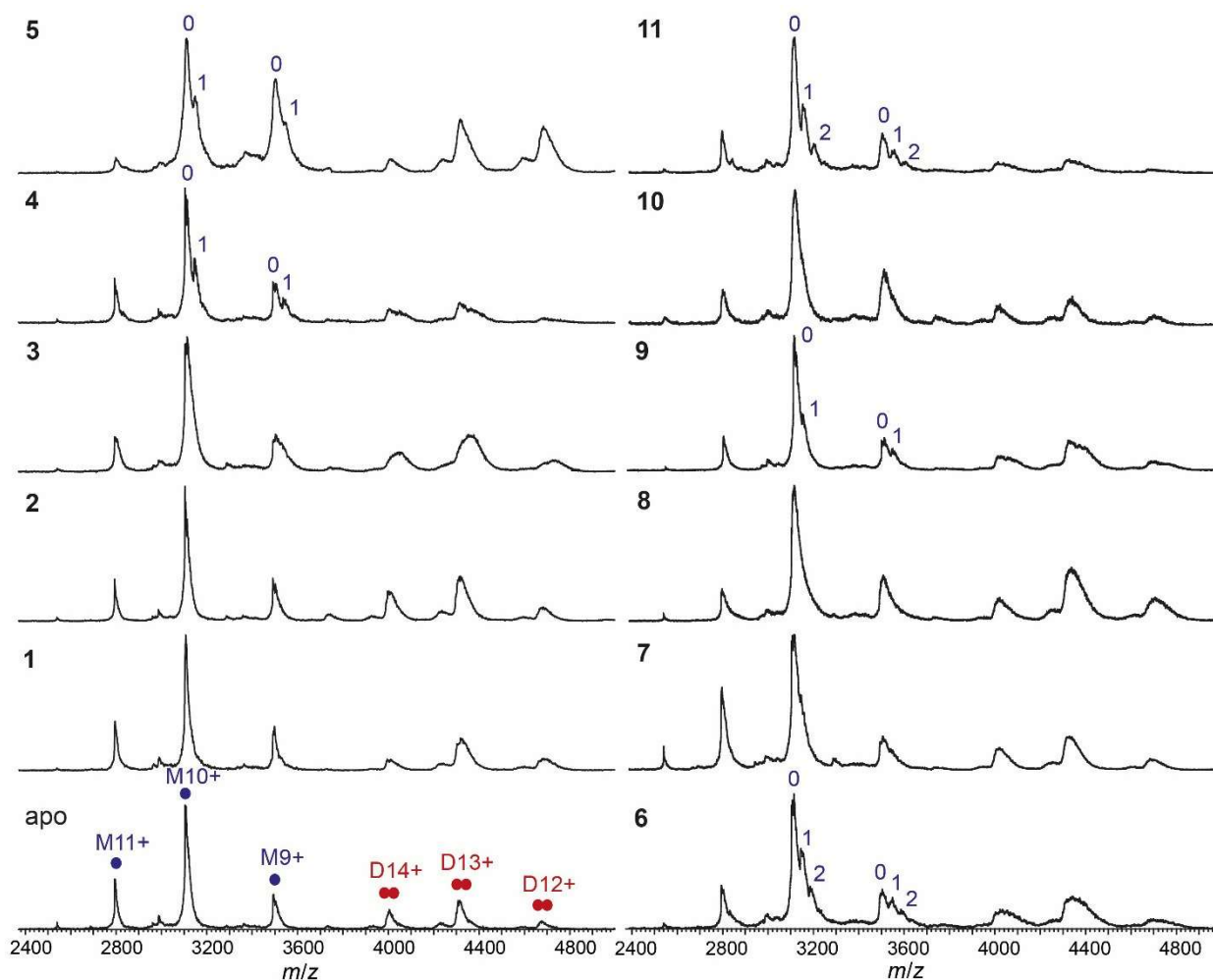


Figure 4.14 Screening of sulfonamide compounds **1-11** against *EcCoaB* by native MS. In each experiment, the protein concentration was 15 μM and the ligand concentration was 30 μM . For the screening hits (compounds **4, 5, 6, 9** and **11**), the number of ligands bound to each monomer species is indicated.

The native MS screen revealed five hits (compounds **4**, **5**, **6**, **9** and **11**) that bound to *EcCoaB*, as indicated by the appearance of additional peaks corresponding to the mass of the protein-ligand complex in the spectra. Due to the relatively small size of the ligands, only ligands bound to the monomer could be discretely resolved. For compounds **6** and **11**, a small degree of non-specific binding was observed, as revealed by the presence of doubly-bound *EcCoaB* monomer species, which could either represent non-specific binding, or possibly two fragments binding at the same time to the active site. The mass shifts observed in the spectra were consistent with the molecular weights of the ligands (Table S1). For example, binding of ligand **11** (406 Da) to the *EcCoaB* monomer shifted the 9+ state by $m/z = 48$ (expected m/z shift = 45) and the 10+ state by $m/z = 44$ (expected m/z shift = 41).

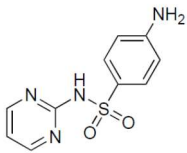
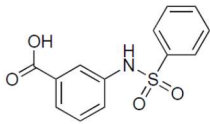
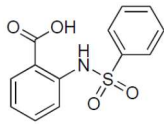
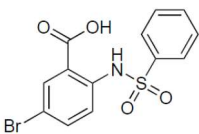
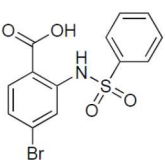
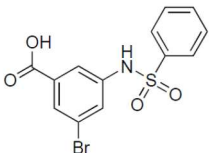
Table 4.3 Expected and observed m/z shifts of ligands binding to the *EcCoaB* monomer by native MS.

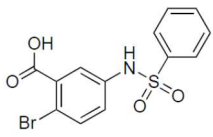
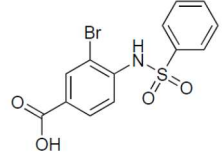
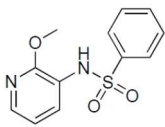
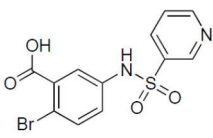
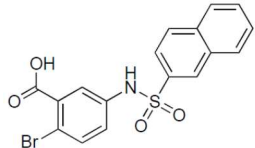
Ligand	Molecular weight (Da)	9+ (observed)	9+ (expected)	10+ (observed)	10+ (expected)
4	355	42	39	38	36
5	355	43	39	38	36
6	355	43	39	37	36
9	264	32	29	28	26
11	406	48	45	44	41

To validate the screening hits, they were tested for *EcCoaBC* inhibitory activity using an enzymatic assay. In this assay (performed by Dr. Jeannine Hess, Abell research group) the production of pyrophosphate that occurs concomitantly with the first half-reaction catalyzed by *EcCoaB* (the formation of 4'-phosphopantothenoyl-CMP) was monitored. As *EcCoaB* was used as a surrogate for the bifunctional *EcCoaBC* in the MS studies, *EcCoaBC* (16 nM) was used in the enzymatic assay to ensure that the inhibitors identified were biologically relevant. Ligands were tested at a

concentration of 1 mM, in the presence of CTP, PPan and L-cysteine at 125, 125 and 500 μ M, respectively. The initial sulfonamide fragment **1** showed the weakest *Ec*CoaBC inhibitory activity of $6 \pm 3\%$, while the most potent compound was the naphthalene analogue **11**, which inhibited *Ec*CoaBC activity by $92\% \pm 5$ (**Table 4.4**).

Table 4.4 *Ec*CoaBC inhibition activity of compounds at 1 mM as measured by an enzymatic assay.

Compound	Structure	% inhibition of <i>Ec</i> CoaBC at 1 mM	Native MS hit
1		6 ± 3	No
2		15 ± 4	No
3		21 ± 4	No
4		34 ± 4	Yes
5		83 ± 5	Yes
6		46 ± 4	Yes

7		28 ± 8	No
8		13 ± 2	No
9		29 ± 5	Yes
10		25 ± 5	No
11		92 ± 5	Yes

Pleasingly, there was good agreement between the results of the native MS assay and the biochemical assay. The five native MS hits (compounds **4**, **5**, **6**, **9** and **11**) were also the five most active compounds in the biochemical assay, with an average *EcCoaBC* inhibition of 57% at 1 mM (range of 29 to 92%) (**Figure 4.15**). Conversely, the compounds that failed to bind in native MS (compounds **1**, **2**, **3**, **7**, **8** and **10**) showed the lowest activity in the biochemical assay, on average inhibiting activity by 18% (range of 6 to 28%). Taken together, these results validate the use of native MS as a simple and rapid screening technique for identifying inhibitors of *EcCoaBC*.

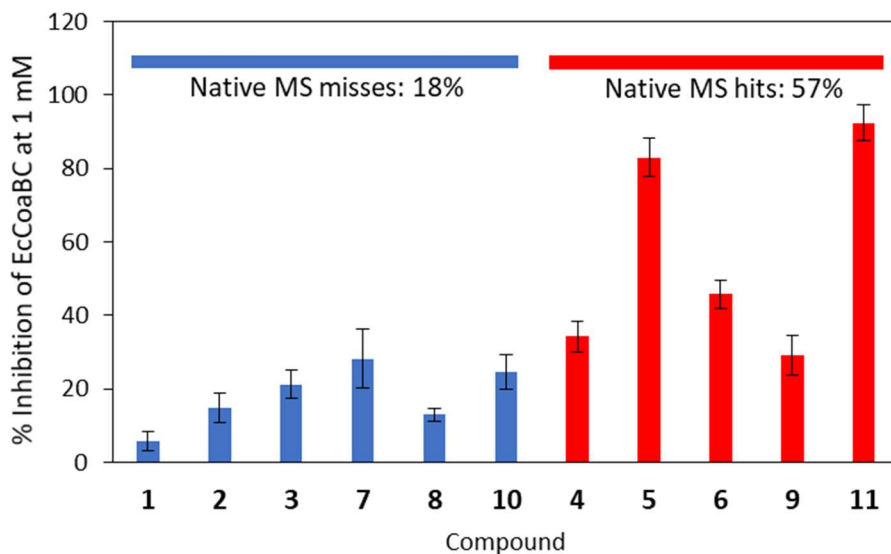


Figure 4.15 Inhibition of the PPCS activity of EcCoaBC by compounds **1–11** at 1 mM as measured by an enzymatic assay. EcCoaBC, CTP, phosphopantothenate and L-cysteine were present in the assay at concentrations of 16 nM, 125 μ M, 125 μ M and 500 μ M, respectively. Native MS hits (**4, 5, 6, 9** and **11**) inhibited enzymatic activity by an average of 57%. Native MS misses (**1, 2, 3, 7, 8** and **10**) inhibited enzymatic activity by an average of 18%. The error bars represent the standard deviation of three experiments.

However, it should be noted that the native MS screening was performed on the EcCoaB domain, while the enzyme inhibition studies were performed with the more biologically relevant EcCoaBC protein. This difference could account for why the correlation between MS screening hits and inhibition potency is not clear in all cases. For example, compound **7** and **9** inhibited the PPCS activity of EcCoaBC to similar extents (28 ± 8 and 29 ± 5 % inhibition, respectively), but only compound **9** was a hit in the native MS assay.

In future work, native MS could be used to derive binding affinities for the inhibitors identified here (as was done for CTP) and the correlation between MS-determined binding affinities for EcCoaB and inhibitory activity against EcCoaBC explored. Moreover, CIU IM-MS could be used to investigate the effect of the inhibitors and other known ligands on EcCoaB stability.

4.2.2.4. Summary

Native MS was used to study the interactions by between the *EcCoaB* domain and its ligands. Our data show that *EcCoaB* dimer formation is promoted by CTP, but not ATP, which is consistent with previous structural and biochemical data. A similar native nESI-IM-MS-based analysis of *EcCoaBC* will be important for understanding how the *EcCoaB* interactions observed contribute to the oligomeric interactions of dodecameric *EcCoaBC*, to which the CoaC domain (which CoaB is fused with) will also contribute.

Additionally in this study, native MS was applied for screening a small library of fragment-sized sulfonamide compounds against *E. coli* CoaB, and the hit compounds were validated using a biochemical assay. Good agreement was observed between the native MS binding studies with *EcCoaB* and the enzyme inhibition assays performed with *EcCoaBC*. The active sulfonamide compounds represent a new class of small-molecule, non-mimetic inhibitors of CoaBC, a recently validated antimicrobial target. This study thus highlights the utility and versatility of native MS both for structural biology and for drug discovery.

4.3. Experimental

4.3.1. Materials and methods

All chemicals were of reagent grade quality or better, obtained from commercial suppliers and used without further purification. Phosphopantothenate^[203] and the fragment-sized sulfonamide **2-11** compounds were synthesized by group members Dr. Jeannine Hess and Robert Starley. Unless otherwise stated, reactions were conducted under positive pressure of a dry nitrogen atmosphere. Anhydrous solvents were purchased directly from commercial sources and used without further purification. Analytical thin layer chromatography (TLC) was performed using Merck glass-backed silica (Kieselgel 60 F254 0.25 mm) plates. Compounds were visualized using shortwave (254 nm) ultraviolet light. Flash column chromatography was performed using an Isolera Spektra One/Four purification system and the appropriately sized Biotage SNAP column containing KP-silica gel (50 μ m). Solvents are reported as volume/volume eluent mixtures.

Reactions were monitored by TLC and LCMS to determine consumption of starting materials. Nuclear magnetic resonance (NMR) spectra were recorded at 300 K unless otherwise stated, using either a Bruker 400 MHz AVANCE III HD Smart Probe, 400 MHz QNP cryoprobe, or 500 MHz DCH cryoprobe spectrometer. All spectra were recorded in the deuterated solvent indicated. Data are reported as chemical shift in parts per million (δ ppm) relative to the residual protonated solvent resonance peak. The relative integral, multiplicity and coupling constants (J Hz) of peaks has been provided where possible. Liquid chromatography mass spectrometry (LCMS) was carried out using an AQUITY UPLC H-class system (Waters, Manchester UK). Samples were run using a gradient of water (1–5%) (+0.1% formic acid) in acetonitrile over a period of 4 min. High resolution mass spectra (HRMS) were recorded using a Waters Quadrapole-Time-of-Flight (Q-Tof) spectrometer or an Orbitrap LCMS spectrometer attached to a Dionex Ultimate 3000 HPLC. The mass to charge ratio (m/z) of the molecular ion and difference from calculated mass (δ ppm) are quoted.

4.3.2. Protein expression and purification

4.3.2.1. EcCoaB overexpression and purification

EcCoaB overexpression and purification was performed by group member Dr. Christina Spry. A pET-28b plasmid carrying the coding sequence for the CoaB domain of *E. coli* CoaBC was used. The plasmid encodes *N*- and *C*-terminally hexa-histidine-tagged *EcCoaB* (MGSSHHHHHSSGLVPRGSHM-Pro¹⁸²-Arg⁴⁰⁶-LEHHHHHH, CoaBC residues indicated). *E. coli* BL21(DE3) cells transformed with the construct were grown in LB supplemented with 30 $\mu\text{g}/\text{mL}$ kanamycin until an OD₆₀₀ of approximately 0.6 was reached. Immediately thereafter overexpression was induced by the addition of 1 mM IPTG. After overnight incubation at 37 °C, the bacteria were harvested by centrifugation (4000 *g*, 15 min) and the cell pellet was stored at –20 °C. The harvested cells were lysed in 20 mM Tris, 500 mM NaCl, pH 8.0 containing one EDTA-free cComplete protease inhibitor cocktail tablet (Roche) per 50 mL lysate, using an EmulsiFlex-C5 High Pressure Homogenizer (Avestin). The cells were passaged through the pressure cell three times at a pressure of 40 000 psi, after which the lysate was centrifuged at 35 000 *g* for 30 min.

For nickel-affinity chromatography, the clarified lysate was applied to a 5 mL HisTrap FF column (GE Healthcare) and, following washing with lysis buffer, the His-tagged protein was eluted using a linear gradient of imidazole (0–0.5 M). Elution was monitored at A_{280} and by SDS-PAGE analysis. Fractions containing *EcCoaB* were pooled and concentrated using a Vivaspin 20 centrifugal concentrator with a 10 kDa molecular weight cut-off before being subjected to gel filtration using a Superdex75 26/60 column (GE Healthcare) previously equilibrated with 20 mM Tris, 150 mM NaCl, pH 8.0. Gel filtration was performed at a flow rate of 2.5 mL/min and elution was again monitored at A_{280} and by SDS-PAGE analysis. Fractions containing purified His-tagged *EcCoaB* were pooled and concentrated as before, and aliquots of the concentrated protein were flash-frozen in liquid N_2 and stored at $-80\text{ }^\circ\text{C}$. The identify and purity of *EcCoaB* was checked by denaturing LC-MS (**Figure 4.2**), which showed a mass for the protein that was identical with the theoretical mass of the protein without the N-terminal methionine (27942 Da).

4.3.2.2. *EcCoaBC* overexpression and purification

EcCoaBC overexpression and purification was performed by group member Dr. Christina Spry. *E. coli* CoaBC was overexpressed in *E. coli* BL21(DE3) cells from a codon-optimised pET-28a construct. Transformed cells were grown at $37\text{ }^\circ\text{C}$ in LB culture medium containing $50\text{ }\mu\text{g/mL}$ kanamycin to an OD_{600} of approximately 0.5, then expression was induced with $500\text{ }\mu\text{M}$ IPTG. The culture was incubated for a further 20 h at $20\text{ }^\circ\text{C}$, and the cells were pelleted by centrifugation ($4000\text{ }g$, 20 min). Cell pellets are frozen and stored at $-80\text{ }^\circ\text{C}$. Cells were resuspended in lysis buffer [50 mM Hepes/Na (pH 7.4), 1 M NaCl, 25 mM imidazole, 1 tablet of Pierce Protease Inhibitor Tablet], and lysed by sonication for 3.5 min (10 s on/20 s off, amplitude 50%). Insoluble material was removed from the lysate by centrifugation at $30\text{ }000\text{ }g$ for 30 min at $4\text{ }^\circ\text{C}$. The supernatant was then loaded onto a 5 mL HisTrap FF column (GE Healthcare), which had been pre-equilibrated with lysis buffer. The column was further washed with lysis buffer, and then *E. coli* CoaBC was eluted from the column by increasing the imidazole concentration to 500 mM . The eluate was concentrated to $\leq 10\text{ mL}$ and injected onto a Superdex-200 Hiload 16/60 FPLC column (GE Healthcare) previously equilibrated with 50 mM Tris (pH 8.0), 250 mM NaCl, and 5

mM mercapto-ethanol. The purified protein was concentrated and aliquots were flash-frozen in liquid N₂ and stored at -80 °C.

4.3.3. Mass spectrometry

Native nanoelectrospray ionization-ion mobility-mass spectrometry (nESI-IM-MS) spectra were recorded on a Synapt HDMS mass spectrometer (Waters, Manchester, UK) modified for studying high masses (32k quadrupole) and equipped with a traveling wave (TW) IM cell. EcCoaB was exchanged into 200 mM ammonium acetate (pH 7.0 unless otherwise stated) solution using Micro Bio-Spin 6 chromatography columns (Bio-Rad, UK). Protein samples were equilibrated at room temperature in the absence or presence of the indicated concentrations of ligands for at least 15 min before analysis. 2.5 µL of protein solution was injected into a gold-coated borosilicate emitter (Thermo Scientific, UK) for sampling. Typical conditions were capillary voltage 1.5 kV, cone voltage 50 V, trap collision voltage 20 V, trap DC bias 35 V, transfer collision voltage 20 V, source temperature 20 °C, backing pressure 3–4 mbar, trap pressure 3–4 × 10⁻² mbar, IM (N₂) pressure 5–6 × 10⁻¹ mbar and TOF pressure 7–8 × 10⁻⁷ mbar. For ion mobility experiments, the mass spectrometer was operated in mobility TOF mode using a wave velocity of 250 ms⁻¹ and a wave height of 10 V. To effect collision-induced unfolding (CIU), the trap collision voltage was raised in 2 V increments from 10 to 80 V. Avidin, concanavalin A, alcohol dehydrogenase and pyruvate kinase were used as calibrants for CCS determination. All mass spectra were calibrated externally with cesium iodide (10 mg/mL). Data acquisition and processing were performed using MassLynx 4.1 (Waters) and DriftScope 2.5 (Waters).

4.3.4. K_d determination

The determination of dissociation constant (K_d) for the EcCoaB monomer-dimer equilibrium was determined using the procedure of Zenobi and co-workers,^[193] which was based on earlier work by the group of Gabelica.^[194]

Briefly, the equilibrium association constant (K_a) relating the EcCoaB monomer (M) and dimer (D) concentration is described using equation 1,

$$K_a = \frac{[D]}{[M]^2} = \frac{1}{K_d} \quad (1)$$

The response factors (R_M and R_D) governing the relationship between the concentrations of M and D and their relative ion intensities (I_M and I_D respectively) are described using equations 2 and 3,

$$I_M = R_M \times [M] \quad (2)$$

$$I_D = R_D \times [D] \quad (3)$$

Given a total *EcCoaB* monomer concentration of $[M_0]$, and setting the ratio of response factors of the dimer and monomer as R , the equations 3 and 4 can be derived,

$$R = \frac{R_D}{R_M} \quad (4)$$

$$\frac{I_M}{I_D} = \frac{1 + \sqrt{1 + 8K_a M_0}}{2RK_a M_0} \quad (5)$$

After determination of I_M and I_D from the mass spectra, equation 5 can be solved for R and K_a by double-parametric non-linear curve fitting. Non-linear parameter uncertainties were estimated using a Monte Carlo method as described by Si and co-workers.^[204]

The proportion of ligand binding was estimated by considering the relative peak areas of unbound and bound states for each spectrum.^[191] For example, the proportion of singly-bound monomer was calculated by dividing the relative intensity of singly-bound monomer by the total intensities of all monomeric species (unbound, singly-bound and doubly-bound). For this calculation, all charge states were taken into account. Additionally, the experimentally-determined response factor of $R = 1.8$ was used as a parameter to determine the relative amounts of monomeric and dimeric *EcCoaB* species.

For determining the K_d value of the *EcCoaB* monomer binding to CTP, we first determined the proportion of unbound and singly-bound monomer in the presence of 50 μM CTP or ATP by consideration of the relative peak areas of all charge states of the monomer. We further made the assumptions that all ATP binding to the *EcCoaB* monomer is non-specific, and that the degree of non-specific binding by CTP to the protein would be identical to the degree of non-specific binding by ATP to the protein. Operating under that assumption, 24% of the 34% of CTP binding to the monomer was considered to be non-specific, leaving 10% of the monomer being specifically bound and 90% unbound. The corrected concentrations of unbound and singly-bound monomer were then calculated to be 15 μM and 1.6 μM respectively. Finally, the concentration of free CTP was estimated by subtracting the CTP concentration bound to protein (for both monomers and dimers, but corrected for non-specific binding) from the initial CTP concentration, which was 40.5 μM in this case. This gave a K_d value of 380 μM for the monomer-CTP complex. A similar method was used to calculate the K_d value for CTP binding to the *EcCoaB* dimer.

4.3.5. Ion mobility calibration

Calibration with ions of known CCS is required on a TWIM instrument because the analytical equations governing ion mobility are complex owing to the non-linear time- and space-varying dependence of the electric field.^[76, 139]

Briefly, the drift times (t_D) (in ms) are corrected for mass-dependent flight time using equation 6,

$$t'_D = t_D - \frac{C\sqrt{(m/z)}}{1000} \quad (6)$$

where t'_D is the corrected drift time of the calibrant ions, C is the "EDC (Enhanced Duty Cycle) delay coefficient" and is equal to 1.57, and m/z is the mass-to-charge ratio of the observed ion. The calibrant CCS (Ω) values are then corrected for both ion charge state (z) and reduced mass (μ) according to equation 7,

$$\Omega' = \frac{\Omega}{(z \times \sqrt{(1/\mu)})} \quad (7)$$

where Ω' is the corrected CCS value and μ is the reduced mass of the ion- N_2 collision complex. Next, $\ln(t'_D)$ is plotted against $\ln(\Omega')$ and the slope (X) of the graph is determined. A new corrected drift time (t''_D) is then calculated using equation 8,

$$t''_D = t'_D X^X \times Z \times \sqrt{(1/\mu)} \quad (8)$$

Finally, Ω is re-plotted against t''_D to generate a linear calibration plot that can then be used to determine CCS values for analyte ions.

Unless otherwise stated, all CCS values are quoted as N_2 values. PULSAR^[53] was used to perform CCS calibration,^[76] generate CIU unfolding plots and calculate CIU50 values. As algorithms for CCS calculation from X-ray crystal structures are parameterized for helium collision gas,^[139] theoretical CCS(He) values were first determined using the projection approximation (PA) algorithm implemented in DriftScope, and then converted to theoretical CCS(N_2) values as described by Bush *et al.*^[79] However, the PA algorithm, which functions by finding the average projection (or “shadow”) of an ion onto a defined plane, generally underestimates CCS values for biomolecules because it does not take into the account of the greater gas slow-down effect for cavities as well as protein charge which affects ionization of neutral gas molecules.^[139] To account for this, an empirically-determined scaling factor of 1.14 is introduced.^[61] Thus, the “corrected” theoretical CCS value (CCS_{calc}) was calculated using equation 9,

$$CCS_{calc} = 1.14 \times CCS_{PA} \times \left(\frac{M_{exp}}{M_{PDB}}\right)^{\frac{2}{3}} \quad (9)$$

where CCS_{PA} is the CCS calculated from the X-ray crystal structure using the PA algorithm, M_{exp} is the experimental mass of the protein and M_{PDB} is the mass calculated from the crystal structure. In the case of the *EcCoaB* dimer, CCS_{PA} was determined to be 3482 \AA^2 from the reported X-ray crystal structure (PDB: 1U7U),^[184] and M_{exp} and M_{PDB} were 27942 and 24834 Da, respectively. Thus, CCS_{calc} was calculated to be 4258 \AA^2 .

4.3.6. Inorganic pyrophosphatase-purine nucleoside phosphorylase PNP-PPIase assay

The assay was performed by group member Dr. Jeannine Hess. PPCS activity was monitored by the method of Webb^[205] using the commercially available EnzCheck pyrophosphate assay kit (E-6645) (Life Technologies). The assay couples production of pyrophosphate to the phosphorolysis of 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG), which can be monitored by measuring absorbance at 360 nm. Reactions were performed at 25 °C in the wells of a 96-well plate (Greiner), in a final volume of 150 μ L. The final composition of reactions was 16 nM *EcCoaBC*, 125 μ M CTP, 125 μ M 4'-phosphopantothenate, 500 μ M L-cysteine, 0.03 U/mL inorganic pyrophosphatase, 1 U/mL purine nucleoside phosphorylase, 200 μ M MESG, 100 mM Tris, pH 7.6, 1 mM MgCl₂, 1 mM TCEP, and 1 mM of test compound (initially dissolved to 50 mM in DMSO) or an equivalent concentration of DMSO (2%, v/v). Initially, 100 mM Tris, pH 7.6, 1 mM MgCl₂, 1 mM TCEP with CTP, 4'-phosphopantothenate, L-cysteine, MESG and the test compound/DMSO, each at twice the final concentration, in a final volume of 75 μ L, was added in triplicate to the wells of a 96-well plate, and the plate was pre-incubated at 25 °C for 10 min. A 100 mM Tris, pH 7.6, 1 mM MgCl₂, 1 mM TCEP solution containing *EcCoaBC*, inorganic pyrophosphatase and purine nucleoside phosphorylase, each at twice the final concentration, was also prepared and similarly pre-incubated at 25 °C for 10 min. Reactions were subsequently initiated by the addition of 75 μ L of the enzyme solution to each well, and enzyme activity was immediately monitored by measuring absorbance at 360 nm at 20 s intervals over 30 min in a ClarioStar microplate reader (BMG Labtech). Assays were performed in triplicate, and included wells without 4'-phosphopantothenate, which served as "no reaction" controls, as well as wells containing 2% (v/v) DMSO instead of test compound, which served as "uninhibited" controls. Initial reaction rates were determined from the linear portion of the time courses. To convert the reaction rates measured in the presence of each test compound to a percentage inhibition, the average rate determined for the "no reaction" wells was first subtracted from all initial rates. Thereafter the background-corrected rates obtained for the wells containing test compound were subtracted from the average of the background-corrected uninhibited controls, and then the resulting difference in rate divided by the average of the uninhibited control.

5. Investigation of the Aurora A-TPX2 interaction and its targeting by small molecules using native mass spectrometry

5.1. Introduction

Aurora A (AurA) is a Ser/Thr kinase that plays important roles in mitosis and cytokinesis.^[148] During mitosis, Aurora A associates with the centrosome and the spindle microtubules to control centrosome maturation and spindle assembly. However, the overexpression of AurA has been linked with a range of solid and hematological tumors as well as with aneuploidy, defective mitotic spindles, and resistance to apoptosis.^[165] Consequently, AurA has been considered to be an attractive target for the development of anticancer agents.^[206]

Like other Ser/Thr kinases, AurA is activated by phosphorylation of a threonine residue (T288) in its activation loop. AurA can also be activated by binding with its protein binding partner, targeting protein for Xklp2 (TPX2), in multiple ways.^[207] First, TPX2 localizes AurA to the mitotic spindle. Secondly, TPX2 stabilizes the active conformation of AurA by protecting the key phospho-T288 in the kinase activation loop from protein phosphatase 1. Thirdly, TPX2 increases the stability of AurA by protecting it from proteasomal degradation during early mitosis. TPX2 is often overexpressed in many cancers alongside AurA.

Despite being 747 amino acid residues in length, only residues 1–43 of TPX2 (TPX2^{1–43}) are sufficient for the binding and activation of AurA. Interestingly, TPX2 binds to AurA in two distinct stretches, TPX^{7–21} and TPX^{30–43} that are connected by a flexible, non-structured linker that cannot be observed in the X-ray crystal structure (**Figure 5.1**). The “upstream” TPX^{7–21} region adopts an extended conformation that fits within a hydrophobic groove in the N-lobe of AurA. Meanwhile, the “downstream” TPX^{30–43} region forms an α -helix that binds to AurA between the N-lobe and the C-lobes, close to the activation loop.

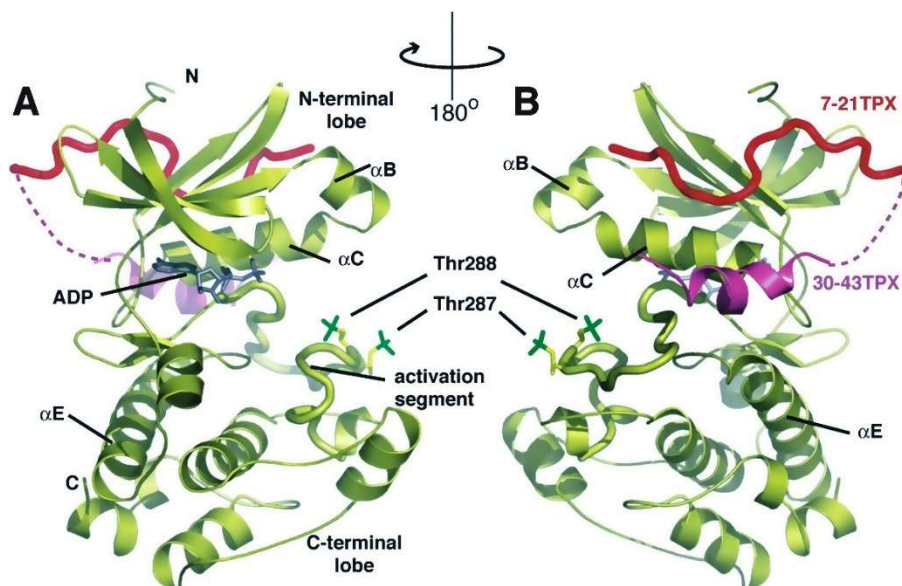


Figure 5.1 X-ray crystal structure of AurA¹²²⁻⁴⁰³ and TPX2¹⁻⁴³ showing the protein-protein interaction binding regions (PDB code: 1OL5). Reproduced with permission from Bayliss et al.^[208]

Kern and co-workers have recently shown that T288 phosphorylation and TPX2 binding results in comparable increases in the catalytic activity of dephosphorylated AurA.^[45] Moreover, they surprisingly showed that dephosphorylated AurA is in the active conformation when bound to TPX2, suggesting that TPX2 is sufficient for AurA activation even in the absence of T288 phosphorylation. The addition of TPX2 to phosphorylated AurA resulted in a further two-fold increase in catalytic activity.

The majority of AurA inhibitors reported in the literature target the ATP-binding site, with the clinically-profiled alisertib (MLN8237, Millennium Pharmaceuticals)^[209] (**Table 5.1**) being among the most well-known. JNJ7706621, a classical “Type 1” ATP-competitive kinase inhibitor, has also been reported to display potent but non-selective activity against AurA.^[210] However, the high degree of structural conservation in the active site of all the members of the human kinome^[211] makes the development of selective ATP-competitive inhibitors of kinases, including AurA, a particularly difficult challenge.

Table 5.1 *AurA* ligands used in this study. K_D values for JF218, JF004 and JF025 were determined by ITC by Abell group members.

Compound	Structure	K_D	Binding region
Alisertib (MLN8237)		40 nM ^[212]	ATP
JNJ7706621		IC ₅₀ = 11 nM ^[210]	ATP (Type 1)
JF218		0.79 μM	Allosteric
JF004		49 μM	Allosteric
JF025		537 μM	Allosteric

Targeting the AurA-TPX2 protein-protein interaction is an alternative approach to blocking AurA activity. Disruption of the AurA-TPX2 interaction has been shown to impair kinase activity,

resulting in the mislocalization of Aurora A, mitotic defects, and cell cycle arrest.^[213] Moreover, since the TPX2-binding site of AurA is much less highly conserved than the ATP-binding site, inhibitors that target TPX2 binding could potentially show improved selectivity for AurA over other kinases, thereby reducing off-target effects.

The Abell research group had a significant interest in developing compounds that target the AurA-TPX2 protein-protein interaction via binding to the TPX2 binding site of AurA. Representative examples of the in-house allosteric AurA inhibitors, **JF218**, **JF004** and **JF025** investigated in this work are shown in **Table 5.1**. The aim of this section of the project was to apply native MS methods to study the AurA-TPX2 interaction and to further validate the mechanism of action of the compounds developed.

5.2. Results and discussion

5.2.1. Expression and purification of AurA and TPX2

In this project, AurA mutants (**Table 5.2**) were expressed in *E. coli* BL21(DE3) cells following a general protocol utilized by the research group of Dr. Marko Hyvönen (Department of Biochemistry). The precise monitoring of these mutants and control of their phosphorylation status had not previously been investigated. The culture medium for AurA protein expression as well as all purification and storage buffers contained 1 mM ATP, as AurA was thought to be unstable without the presence of ATP bound to the active site (Hyvönen group, personal communication).

Table 5.2 List of AurA mutants used in this study.

Mutant	Description
AurA-A	AurA ¹²⁶⁻³⁹⁰ containing kinase domain with T287A mutation and non-cleavable C-terminus His ₆ -tag. Expressed in this study.

AurA-B/AurA-B'	AurA ^{126–391} containing kinase domain with T287A mutation and TEV-cleavable N-terminus His ₆ -tag. AurA-B' is the cleaved version of AurA-B . Expressed in this study.
AurA-C/AurA-C'	AurA ^{123–401} containing kinase domain with T287A/T288D double mutation and TEV-cleavable N-terminal His ₆ -tag. Constitutively active. AurA-C' is the cleaved version of AurA-C . Expressed in this study.
AurA-D	AurA ^{125–403} containing kinase domain with D274N mutation. Catalytically dead. Provided by collaborator Maxim Rossman from the Hyvönen group.

The AurA construct first employed was the AurA mutant containing the kinase domain, a T287A mutation and a non-cleavable C-terminus His₆-tag (**AurA-A**). The T287A mutation was chosen because although phosphorylation at T287 is often observed *in vitro*, this does not appear to have an effect on catalytic activity.^[214] Reducing the number of phosphorylation sites from two to one was thought to be advantageous for MS because then the presence of phosphorylation at the critical T288 residue can be unambiguously identified by a +80 Da *m/z* shift from the unphosphorylated protein in the mass spectrum. The construct also contains the λ-phosphatase gene, which when co-expressed dephosphorylates AurA and reduces its toxicity to the bacterial cells. However, initial attempts at the expression of **AurA-A** were unsuccessful, leading to the isolation of either a defective protein (incorrect mass) or the correct protein but in very low yields. This was attributed to the fact that AurA is highly toxic to *E. coli* cells, which exerts selective pressure on bacteria to develop mutations in the enzyme or to eject the plasmid altogether. To ameliorate the toxicity issue, a number of modifications to the general procedure were investigated (**Table 5.3**).^[215]

Table 5.3 Modifications to the general procedure for AurA expression to improve yield.

Method	Rationale
--------	-----------

The expression culture is inoculated directly from colonies on an agar plate instead of from a starter culture.	The selection of plasmid-containing cells is more stringent on antibiotic-containing agar plates than in liquid culture media.
Induction with a lower IPTG concentration (100 μ M instead of 400 μ M).	Cells induced using low IPTG concentrations have more metabolic control over the toxicity of the expressed protein.
Induce at 37 °C for 3 h instead of 20 °C overnight.	A shorter induction period reduces the time over which the selection pressure of the toxic protein acts.
A second dosage of ampicillin (50 μ g/mL) is added just before induction.	The ampicillin initially present in the medium may be hydrolyzed by β -lactamases by the time the culture reaches $OD_{600} = 0.4$. Adding a second dosage of ampicillin before induction ensures that only cells that still retain the plasmid will survive during the expression period.

Using these modifications, yields of *ca.* 1 mg/L of purified **AurA-A** was obtained. Due to the very low sample consumption requirements of nanoESI-MS, further optimization of the protein expression and purification procedure was not performed.

The **AurA-A** protein was initially purified by affinity chromatography using a nickel Sepharose resin in a gravity flow column, and then further purified by size-exclusion chromatography. However, although sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) analysis of the purified product showed the presence of only a single band, mass determination by high-resolution LC-MS showed that four AurA species were actually present (**Figure 5.2**). The masses of the four species corresponded to phosphorylated **AurA-A** (31834 Da), unphosphorylated **AurA-A** (31754 Da), phosphorylated and demethioninylated **AurA-A** (31702 Da) and unphosphorylated and demethioninylated **AurA-A** (31623 Da). The presence of these four

species was attributed to sporadic cleavage of the N-terminal methionine residue^[216] and/or phosphorylation.

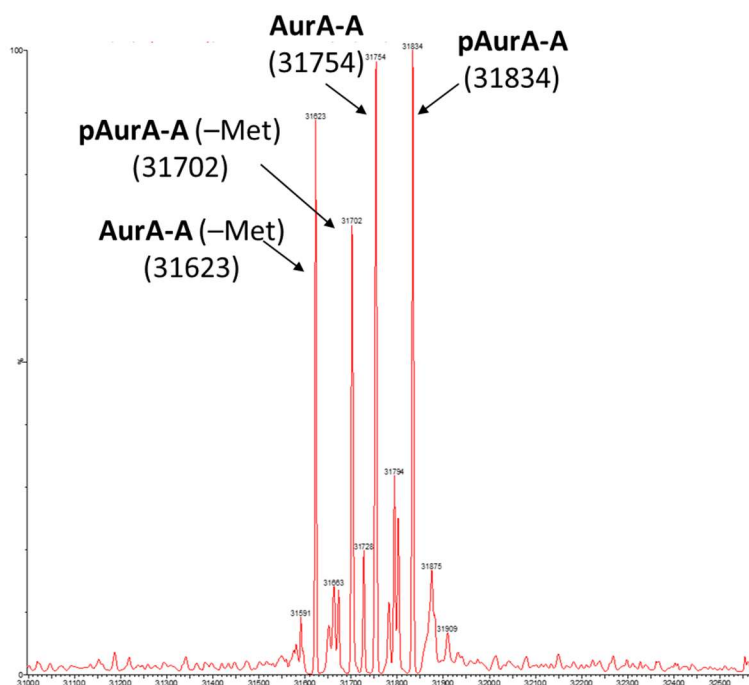


Figure 5.2 Deconvoluted high-resolution LC-MS of **AurA-A**.

This heterogeneity led to difficulties in the subsequent interpretation of native MS data. Consequently, the alternative AurA construct, **AurA-B**, was explored. Like **AurA-A**, this construct also contains the T287A mutation, but differs in the fact that the His₆-tag is located at the N-terminus and is cleavable by the tobacco etch virus (TEV) protease. After cleavage, the N-terminus begins with a glycine residue rather than a methionine residue, so sporadic demethioninylation would be avoided. Under the optimal conditions, **AurA-B** was purified in yields of approximately 1–2 mg/L of culture. LC-MS analysis of the uncleaved product showed the presence of two major species differing by 80 Da, which suggests that one of the two species is phosphorylated while the other is unphosphorylated. However, both masses were 156 Da larger than the expected mass as given by the protein sequence. Nevertheless, the protein sample was subjected to digestion by TEV protease and the LC-MS spectrum of the cleaved product (**AurA-B'**) showed the correct masses of 32,045 and 32,125 Da for the unphosphorylated

and phosphorylated enzyme, respectively. This suggested that the source of the +156 Da shift in the uncleaved protein may have been part of the N-terminal tag that was removed.

While cleaved **AurA-B'** was comprised of fewer protein species than purified **AurA-A** (two instead of four), it was still not a homogeneous sample. Therefore, efforts were made to purify and process **AurA-B'** using various means in order to obtain a homogeneous sample. First, the crude TEV digest containing unphosphorylated and monophosphorylated **AurA-B'** was subjected to nickel column chromatography to remove the excised His₆-tag and TEV enzyme, which also contains a polyhistidine tag. To our surprise, the eluted **AurA-B'** was found to contain a mixture of monophosphorylated and diphosphorylated species. This was attributed to the presence of ATP in the elution and storage buffers, allowing **AurA-B'** to phosphorylate itself *in vitro*. The subsequent addition of λ -phosphatase failed to remove all the phosphate groups from **AurA-B'**, even when the potent AurA inhibitor alisertib was present.

The above results suggested that the presence of 1 mM ATP in the bacterial culture medium as well as in all of the purification and storage buffers for AurA strongly promoted autophosphorylation, making it difficult to isolate the enzyme in the unphosphorylated state. Therefore, ATP was replaced with ADP in the culture medium as well as in all subsequent buffers, which was something that had not previously been attempted. The rationale was that ADP, being similar in size to ATP, could also bind to the active site of AurA and stabilize the enzyme but without promoting its autophosphorylation activity. Gratifyingly, this modification led to the production of **AurA-B'** as nearly all in the unphosphorylated form (**Figure 5.3**).

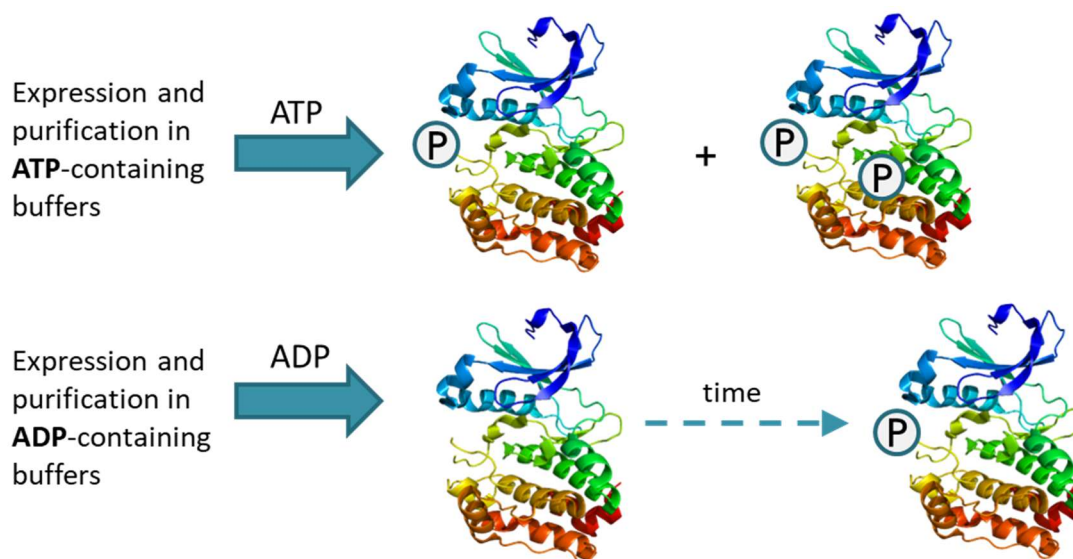


Figure 5.3 Expression of **AurA-B'** in the presence of ADP (lower panel) instead of ATP (upper panel) favors the formation of unphosphorylated enzyme, but which still phosphorylates itself over time. Phosphate groups are represented schematically as purple circles.

However, the unphosphorylated **AurA-B'** still slowly converted into the monophosphorylated form over time. This was attributed to the presence of ATP in commercial ADP stocks,^[217] and was confirmed in this study by native MS (**Figure 5.4**). Even the small amount of ATP in the solution was sufficient to activate the autophosphorylation activity of **AurA-B'**. The contaminating ATP could potentially be removed by incubation with hexokinase and glucose,^[217] but this was not pursued due to time constraints.

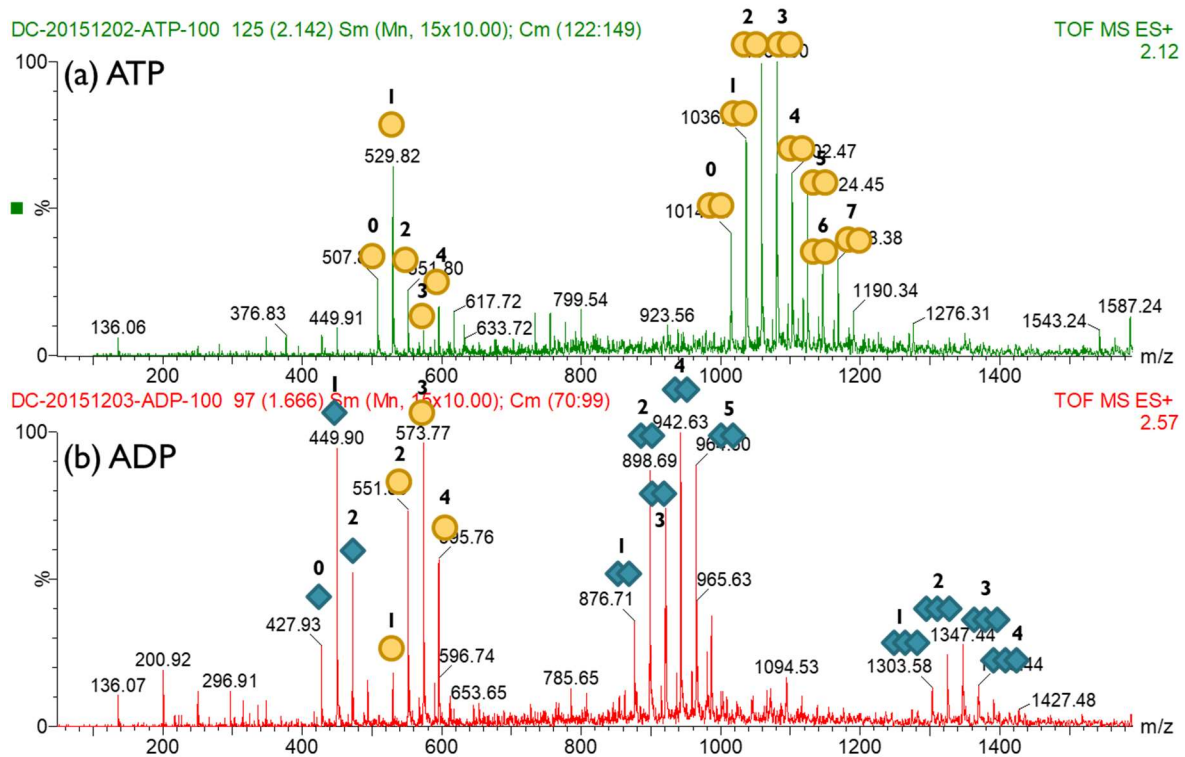


Figure 5.4 Native MS of (a) ATP and (b) ADP solutions. ATP and ADP are indicated by orange circles and purple diamonds respectively (only $z = 1$ states annotated), and the bolded numerals represent the number of Na^+ ions attached. Dimers and trimers of ATP and ADP are presumed to be non-specific.

A sample containing a mixture of monophosphorylated and diphosphorylated **AurA-B'** was subjected to phosphopeptide analysis using LC/MS-MS with tryptic digestion (performed by the Cambridge Centre for Proteomics) to locate the position of the phosphorylation site. As expected, 99% of the phosphate groups were located at the critical T288 residue, while other serine and threonine residues were occasionally phosphorylated (**Figure 5.5**). The presence of other phosphorylation sites could explain the observation of doubly phosphorylated **AurA-B'** when the protein was expressed and purified in an abundance of ATP.

```

GSMGLEVLFGQPRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILALKVLFKAQLEKAGVEHQLRREV
EIQSHLRHPNLRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHR
DIKPENLLLGSAGELKIADFGWSVHAPSSRRATLCGTLDYLPPEMIEGRMHDEKVDLWSLGLVLCYEFLVG
KPPFEANTYQETYKRISRVEFTFPDFVTEGARDLISRLKHNPSQRPMLREVLEHPWITANSSKPS

```

Figure 5.5 Sequence of **AurA-B'** with phosphorylation sites underlined. The critical T288 residue is also shown in bold, and contains 99% of phosphorylation.

Subsequently, a third AurA construct, **AurA-C**, was investigated. This construct contains the T287A/T288D double mutation as well as a N-terminal His₆-tag cleavable by TEV protease. Since the Asp288 residue is a functional mimic of phosphorylated Thr288, this mutant is constitutively active. As a result, **AurA-C** is even more highly toxic to *E. coli* cells, as its catalytic activity is not abrogated in the presence of λ-phosphatase. Moreover, due to its lack of phosphorylation sites at both positions 287 and 288, **AurA-C** is expected to give only one protein species after cleavage of the N-terminal His₆-tag.

AurA-C was expressed in low yields of approximately 1 mg/L of culture. LC-MS analysis of the uncleaved sample showed two major clusters of protein peaks, which was consistent with the two bands that were observed on the SDS-PAGE gel. The higher-mass cluster contained three major peaks corresponding to phosphorylated **AurA-C**, unphosphorylated **AurA-C** and unphosphorylated and demethionylated **AurA-C**. Meanwhile, the lower-mass cluster was thought to be comprised of truncated versions of the protein. Fortunately, cleavage with TEV resulted in a protein sample that formed only a single band on the SDS-PAGE gel, suggesting that the truncations were located in the N-terminal tag. LC-MS analysis revealed the presence of the desired protein species (**AurA-C'**) at 32,278 Da (**Figure 5.6**). Unfortunately, similar to what was observed with **AurA-B'**, **AurA-C'** also autophosphorylated itself over time even when expressed, purified and stored in ADP. As previously discussed, this could be accounted for by the presence of ATP in the ADP stock. Since **AurA-C'** contains both T287A and T288D mutations, this indicates that the protein must be phosphorylating itself at other sites in order to generate the

monophosphorylated and diphosphorylated species observed. At present, the biological significance of phosphorylation of AurA at sites other than T287 and T288 is unknown.

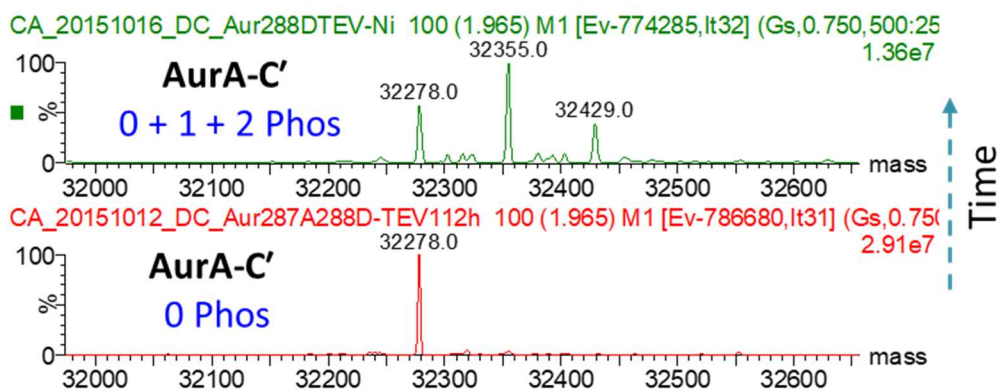


Figure 5.6 Deconvoluted high-resolution LC-MS analysis of unphosphorylated **AurA-C'** after TEV cleavage (lower panel) and a mixture of unphosphorylated, monophosphorylated and diphosphorylated **AurA-C'** after further purification and storage for 72 h (upper panel).

TPX2 was expressed as a GB1 fusion protein in *E. coli* BL21(DE3) cells to maintain solubility during protein expression. Cleavage of the GB1 tag using TEV protease releases a 48-residue fragment containing the active binding region TPX2⁷⁻⁴³, which could be isolated simply by passing through a 10 kDa molecular weight cut-off (MWCO) filter. The purity of TPX2⁷⁻⁴³ was sufficient for use in MS reactions without further purification.

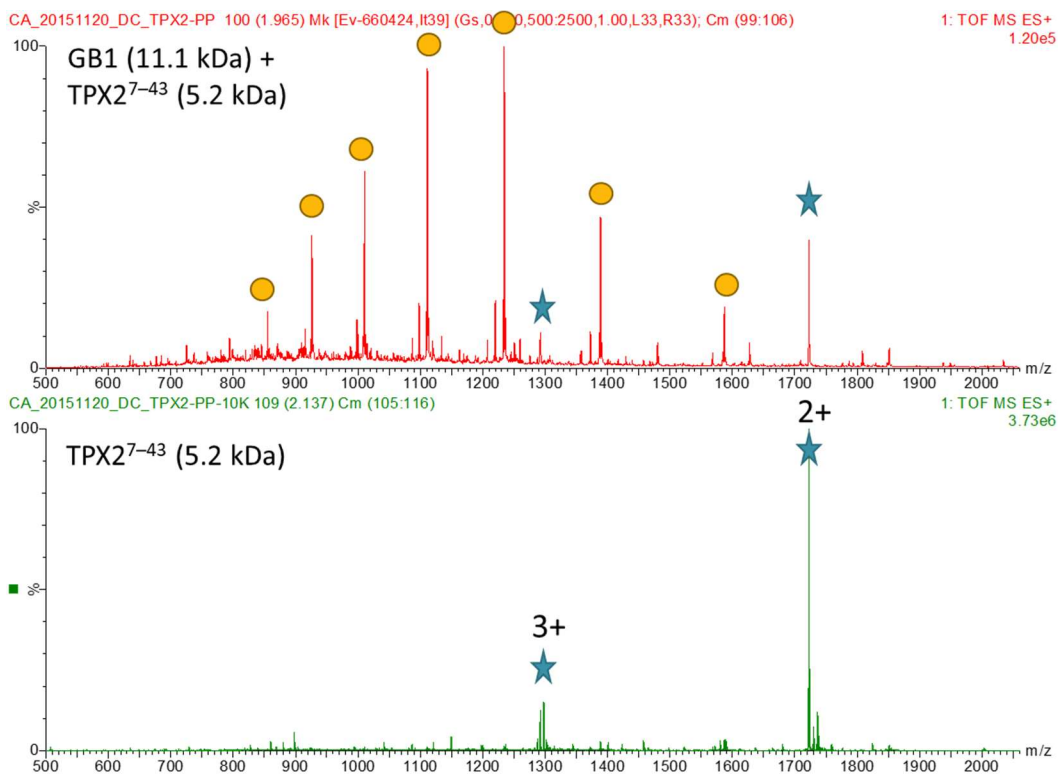


Figure 5.7 High-resolution LC-MS of the mixture of TPX2⁷⁻⁴³ (5.2 kDa) and the cleaved GB1 tag (11.1 kDa) (upper panel), which could be passed through a 10 kDa MWCO filter to isolate TPX2⁷⁻⁴³ (lower panel). Charge states for TPX2⁷⁻⁴³ and the cleaved GB1 tag are indicated with purple stars and orange circles, respectively.

5.2.2. Native MS of the AurA-TPX2 interaction

The interaction between AurA and TPX2 was studied by native MS. In native MS, protein samples must first be exchanged into a volatile buffer such as NH₄OAc, since non-volatile salts can lead to suppression of ionization and extensive adduct formation. However, it was noticed that **AurA-B'** was prone to covalent dimer formation when buffer-exchanged into NH₄OAc (**Figure 5.7**). This was thought to be due to disulfide bridge formation at Cys290, as mutating that position suppressed dimer formation in experiments performed by the Hyvönen group. It was found that dimer formation could also be suppressed by the addition of 1 mM of β-mercaptoethanol in the NH₄OAc solution. This additive was tolerated in native MS and did not impair the quality of the resulting spectrum.

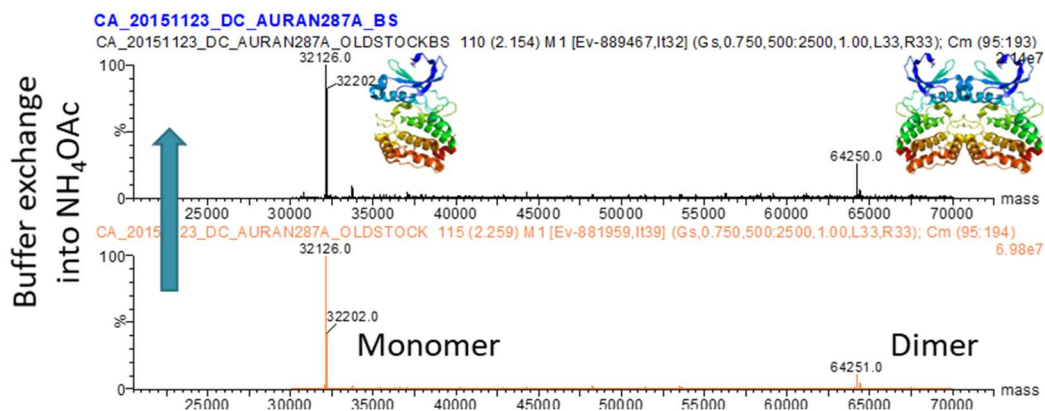


Figure 5.7 Deconvoluted high-resolution LC-MS of **AurA-B'** before (lower panel) and after (upper panel) buffer exchange into ammonium acetate buffer.

Undesirable dimer formation was observed with TPX2⁷⁻⁴³ due to its abundance of Cys residues. Hence, β -mercaptoethanol could also be added to storage buffers for TPX2⁷⁻⁴³ as well as to the NH₄OAc solution just before electrospray.

As described above, **AurA-A**, **AurA-B'** and **AurA-C'** were expressed in a mixture of variously phosphorylated states due to the propensity of the enzyme to phosphorylate itself. This increased the difficulty of peak assignment because peaks of different phosphorylation states, each already broad due to the adduction of weakly-bound solvent or other adducts in native MS, often overlapped with each other. The only exception was **AurA-D**, bearing a D274N mutation that makes it catalytically dead, which was isolated by the Hyvönen group in a completely homogeneous state. This allowed the distinguishing of ADP-bound and ATP-bound ($\Delta m/z = 8$) and even H⁺-bound and Na⁺-bound ($\Delta m/z = 2$) species in native MS (**Figure 5.8**).

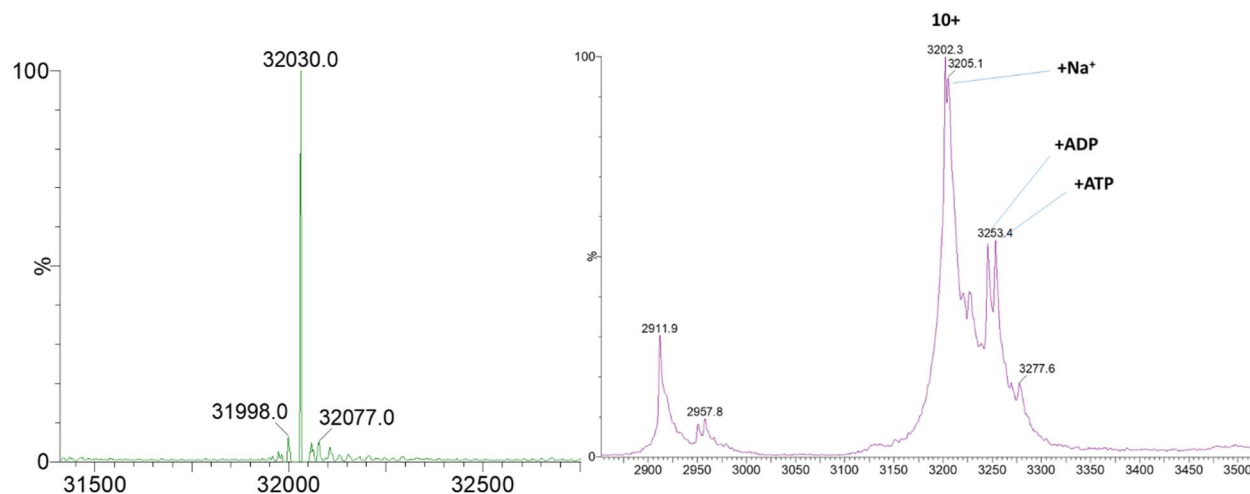


Figure 5.8 Deconvoluted LC-MS spectrum of **AurA-D**, showing near-complete homogeneity (left), and native MS of **AurA-D** showing the discrimination of various small molecule-bound species of the 10+ charge state (right).

Next, **AurA-C'** and TPX2⁷⁻⁴³ were incubated together and their native mass spectrum was recorded (**Figure 5.9**). An obvious shift in the signals of **AurA-C'** (32.2 kDa) was observed upon addition of TPX2⁷⁻⁴³ (5.2 kDa), due to the formation of the non-covalent **AurA-C'**·TPX2⁷⁻⁴³ complex (37.4 kDa). The relatively low charge states of both free **AurA-C'** (9+ and 10+) and the **AurA-C'**·TPX2⁷⁻⁴³ complex (10+ and 11+) suggested that both species were in the folded conformation. Interestingly, despite the strong (*ca.* 6 nM using a fluorescence polarization assay performed by our collaborator Tomasso Moschetti in the Hyvönen group) binding affinity between AurA and TPX2⁷⁻⁴³, a small amount of both unbound **AurA-C'** and unbound TPX2⁷⁻⁴³ were observed in the native mass spectrum. This suggests that one or both of the binding partners were in the incorrect conformation to form the AurA-TPX2 complex in solution. To our knowledge, this is the first time that the AurA-TPX2 interaction has been studied by native MS.

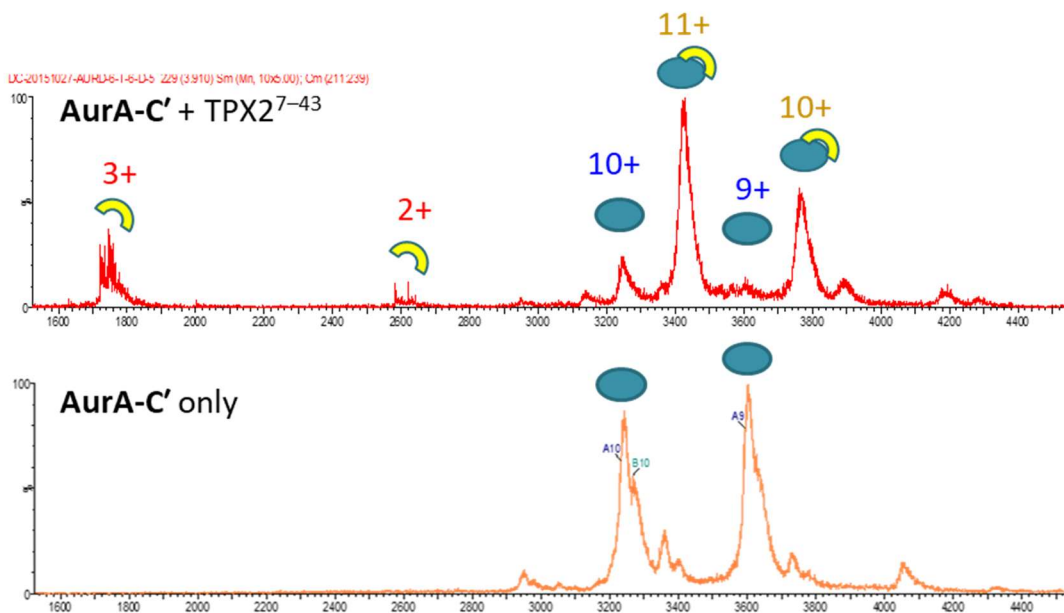


Figure 5.9 Native MS of **AurA-C'** (10 μ M) and TPX2⁷⁻⁴³ (20 μ M) in 200 mM NH₄OAc containing 5% DMSO. **AurA-C'** and TPX2⁷⁻⁴³ are depicted as blue ovals and yellow crescents, respectively.

A collision-induced dissociation experiment was performed on the AurA-TPX2 complex by raising the collision voltage within the trap region of the mass spectrometer (**Figure 5.10**). As the collision energy was increased, the signals for **AurA-B'** and the **AurA-B'**-TPX2⁷⁻⁴³ became sharper due to dissociation of weakly bound solvent or other adduct molecules from the protein. As the collision energy reached 120 V, free TPX2 degraded or otherwise became unobservable. However, even at the highest collision voltage of 150 V, where degradation of the protein backbone was observed to begin and where the protein has already long unfolded (unfolding begins at *ca.* 60 V as observed by CIU-MS, data not shown), the **AurA-B'**-TPX2⁷⁻⁴³ complex still remained robust. This data indicates that the AurA-TPX2 complex possesses high stability in the gas phase.

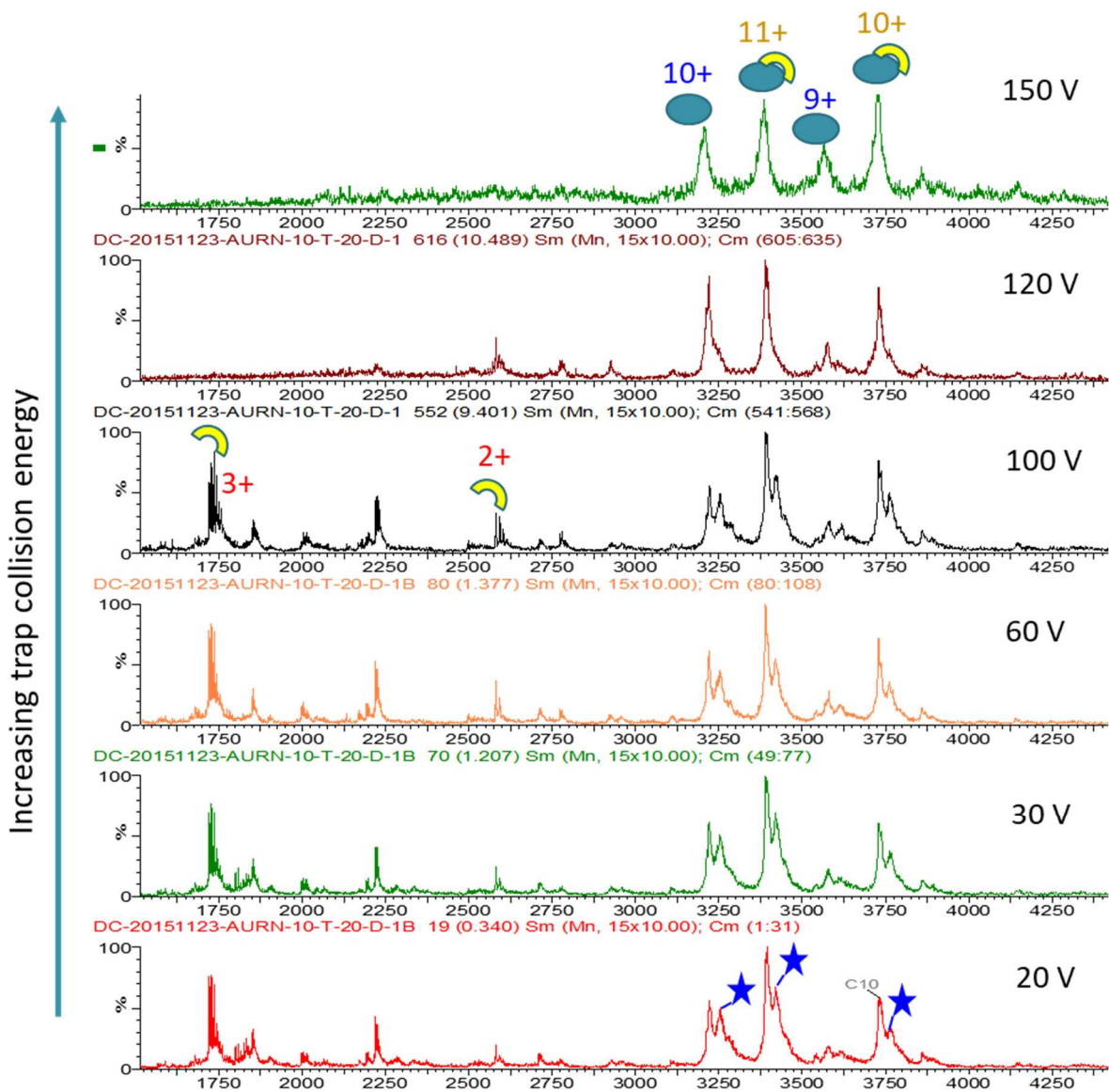


Figure 5.10 Native MS of **AurA-B'** (10 μ M) and **TPX2⁷⁻⁴³** (20 μ M) in 200 mM NH_4OAc containing 1% DMSO. **AurA-B'** and **TPX2⁷⁻⁴³** are depicted as blue ovals and yellow crescents, respectively. An unknown small-molecule impurity is indicated as a blue star (see below).

Because of the strong binding affinity between AurA and TPX2⁷⁻⁴³, it was thought that this protein-protein interaction would be too strong to disrupt using fragment-sized molecules. Therefore, weaker-binding TPX phage display mutants (“M3” and “M6”) (Table 5.4) were obtained from the Hyvönen group and tested for their binding to AurA using native MS.

Table 5.4 Sequence of TPX⁷⁻⁴³ and TPX2 mutants M3 and M6. Residues 7–43 of TPX2 are underlined. Mutations are shown in bold font. Binding affinities (K_D) for AurA were determined by a fluorescence polarization assay by the Hyvönen group.

Construct	K_D	Sequence
TPX ⁷⁻⁴³	6 nM	GPGSYSYDAPSDFINFSSLDDEGDTQNI DSWFEEKANLEN LKGGGCSS
M3 (D11A)	0.3 μ M	MNTIH HHHHH NTSGLNDIFEAQKIEWHEAMSH HHHHHHH NTSSATMAC TIGRISTGSKSLDKLLGGGIETQAITEVFG EF GS GAT QLAHTLAVMVQLPPEE GGLNGSVI WID TENTFRPERIREIAQNRGLDPDEVLKHIYVARAFNSNHQM LLVQQAEDKIKELLNTDRPVKLLIVDSLTS HR SEYIGRGALAERQQKLAKHL ADLHRLANLYDIAVFVTN QVQ AGGGLKSYSYA APSDF IN FSSLDDEGDTQN <u>IDSWFEEKANLEN</u> LKGGGNGHILAHSATLRVYLRKGGKGGKRIARLIDAPHLPE EGEAVFSITEKGIED
M6 (D11A/W34A)	1.1 μ M	MNTIH HHHHH NTSGLNDIFEAQKIEWHEAMSH HHHHHHH NTSSATMAC TIGRISTGSKSLDKLLGGGIETQAITEVFG EF GS GAT QLAHTLAVMVQLPPEE GGLNGSVI WID TENTFRPERIREIAQNRGLDPDEVLKHIYVARAFNSNHQM LLVQQAEDKIKELLNTDRPVKLLIVDSLTS HR SEYIGRGALAERQQKLAKHL ADLHRLANLYDIAVFVTN QVQ AGGGLKSYSYA APSDF IN FSSLDDEGDTQN <u>IDSAFEEKANLEN</u> LKGGGNGHILAHSATLRVYLRKGGKGGKRIARLIDAPHLPE GEAVFSITEKGIED

However, no binding at all was observed between **AurA-B'** with M3 (data not shown) or with M6 by native MS (**Figure 5.11**). Due to time constraints, this avenue of research was not pursued further.

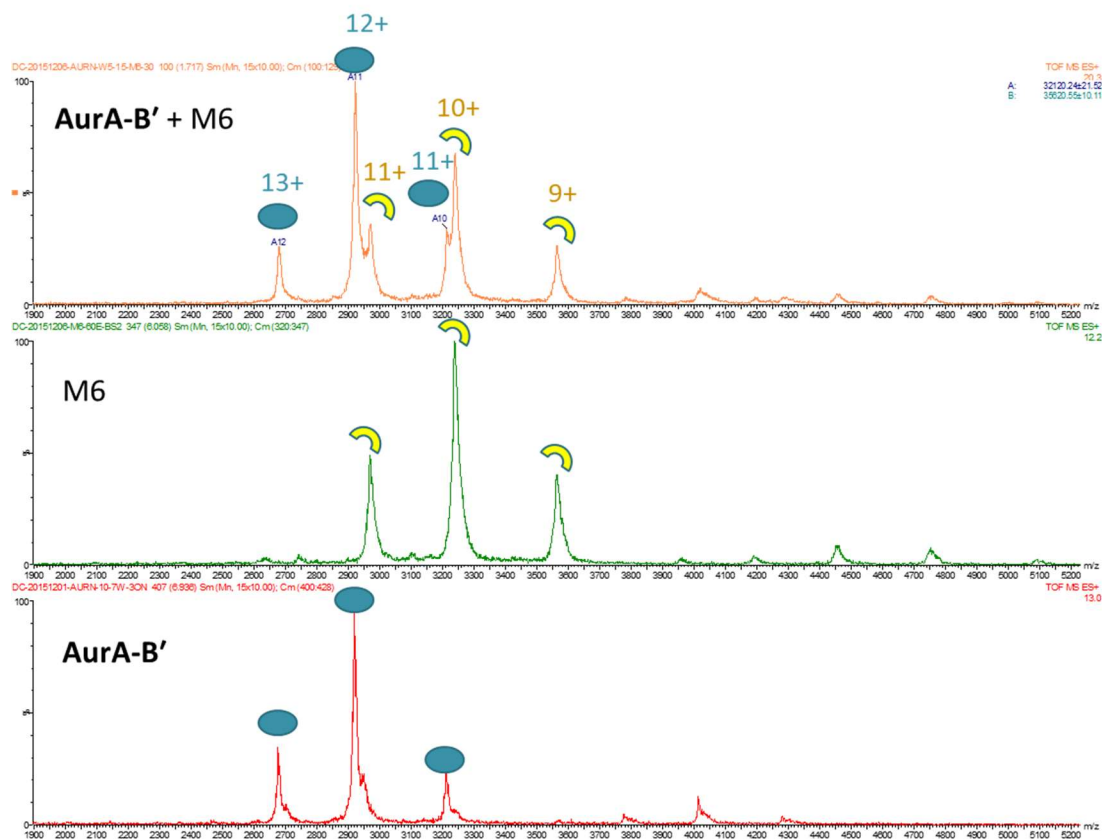


Figure 5.11 Native MS of **AurA-B'** (15 μ M), M6 (30 μ M) and their combination in 200 mM NH_4OAc . No binding between **AurA-B'** and M6 was observed.

It was also observed that there was a small-molecule impurity of around 300 Da that was bound to both of our TPX-2 and AurA protein preparations (lowest panel in **Figure 5.10**). Several efforts were made to identify the nature of this impurity. The impurity showed apparently high binding affinity to AurA, as a sizable proportion of **AurA-B'** remained bound by this impurity even after multiple overnight washes with NH_4OAc buffer. The impurity also dissociated from **AurA-B'** at much higher collision energies (*ca.* 100 V) than required for ADP/ATP or AurA inhibitors (*ca.* 60 V). Denaturing **AurA-B'** or **AurA-C'** with 50% acetonitrile/1% formic acid resulted in the release of a host of small molecules with masses between 200 and 300 Da (**Figure 5.12**), some of which could be assigned to polyethylene glycol (PEG) or plasticizers, which are common contaminants in biochemical laboratories. The Hyvönen group have also observed that AurA is very “sticky” and prone to being bound by hydrophobic molecules (personal communication).

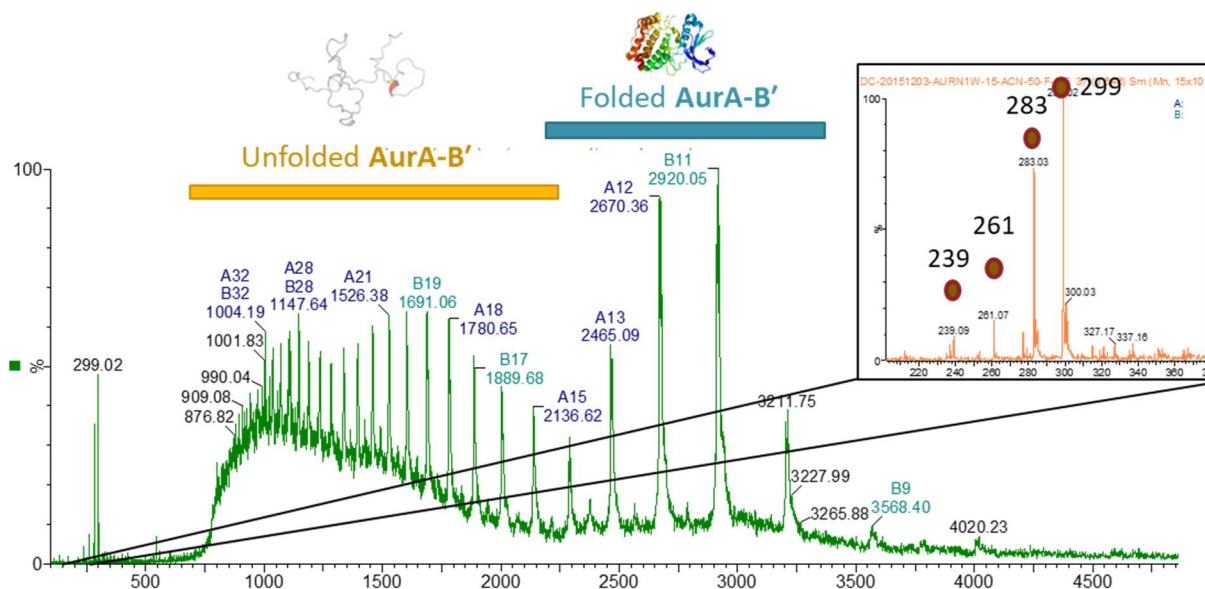


Figure 5.12 Native MS of partially denatured **AurA-B'** in 50% 200 mM NH_4OAc : 49% acetonitrile: 1% formic acid. Both folded and unfolded forms of **AurA-B'** could be seen. Inset: expansion of region between 200 to 400 m/z . The small molecules observed were assigned as follows: 239 (PEG $[\text{A}_5\text{B}+\text{H}]^+$), 261 (PEG $[\text{A}_5\text{B}+\text{Na}]^+$), 283 (PEG $[\text{A}_6\text{B}+\text{H}]^+$) and 299 (unknown).

Because both AurA and TPX2⁷⁻⁴³ are large relative to the small molecule impurities, the disruption of the AurA-TPX2⁷⁻⁴³ complex by small molecules should be obvious in native MS regardless of whether or not impurities were also bound to AurA or TPX2⁷⁻⁴³. Hence, no further efforts were made to isolate and characterize the impurities or to further purify the proteins.

5.2.3. Screening AurA inhibitors by native MS

The in-house and commercial inhibitors were tested against AurA by native MS. The spectra showed that these inhibitors could bind with the various AurA mutants, however, spectra were complicated by the presence of the small-molecule impurity and therefore protein-ligand peaks could not be assigned with certainty (data not shown). We therefore turned to the more important issue of whether or not native MS could be used to monitor the disruption of the AurA-TPX2 protein-protein interaction by small molecules. At the onset of this study, only a few

examples of screening protein-protein interaction disrupters by native MS had been reported in the literature.^[218]

Native MS spectra for **AurA-C'** and TPX2⁷⁻⁴³ were obtained in the absence or presence of the classical Type 1 kinase inhibitor, JNJ7706621 (IC₅₀ = 11 nM^[210]), or our in-house AurA-TPX2 disrupter, **JF218** (K_D = 0.79 μM as determined by ITC) (**Figure 5.13**). The results showed that JNJ7706621 was unable to disrupt the **AurA-C'**-TPX2⁷⁻⁴³ complex as expected. Instead, a ternary **AurA-C'**-TPX2⁷⁻⁴³-JNJ7706621 complex was observed, which could presumably form by the inhibitor binding to the active site of AurA which is unoccluded in the AurA-TPX2 complex. In contrast, **JF218** greatly diminished the intensity of the signals belonging to the **AurA-C'**-TPX2⁷⁻⁴³ complex, while signals for free **AurA-C'** appeared. This suggests that **JF218** was able to disrupt the **AurA-C'**-TPX2⁷⁻⁴³ complex in solution by binding to the TPX2 site of the kinase, thereby displacing TPX2 and generating free AurA species that could be detected in the gas-phase. The native MS data is consistent with the results determined in a fluorescence polarization assay and provide further evidence of the mechanism of action of our in-house inhibitors.

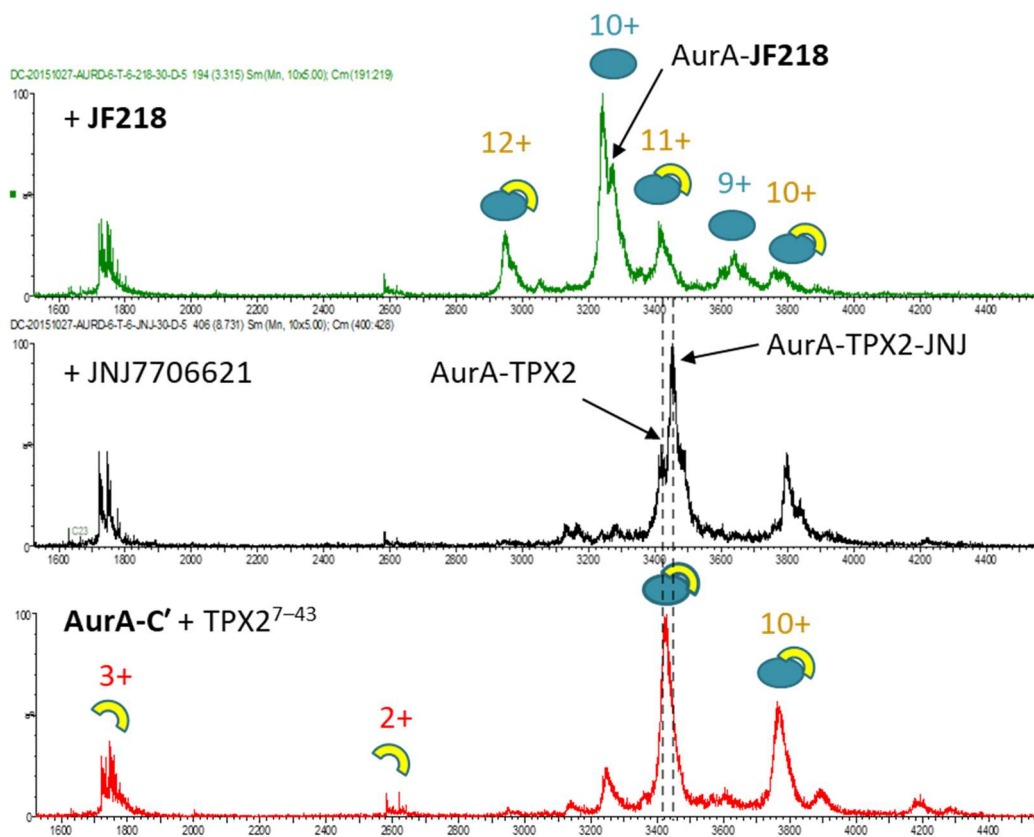


Figure 5.13 Native MS of **AurA-C'** (6 μ M), TPX2⁷⁻⁴³ (6 μ M) in the absence (lower panel) or presence of JNJ7706621 (30 μ M) (middle panel) or JF218 (30 μ M) (upper panel) in 200 mM NH₄OAc containing 5% DMSO.

Subsequently, a panel of in-house allosteric inhibitors with various affinities for AurA were tested for their ability to disrupt the AurA-TPX2 interaction by native MS (**Figure 5.14**). Gratifyingly, the order of disruption of the **AurA-B'**-TPX⁷⁻⁴³ complex in native MS followed the order of binding affinities of the molecules to AurA. The weakest AurA ligand, **JF025** ($K_D = 537 \mu$ M) barely showed any disruption of the protein-protein interaction in native MS, while for the strongest ligand **JF218**, most of the **AurA-B'**-TPX⁷⁻⁴³ complex was disrupted.

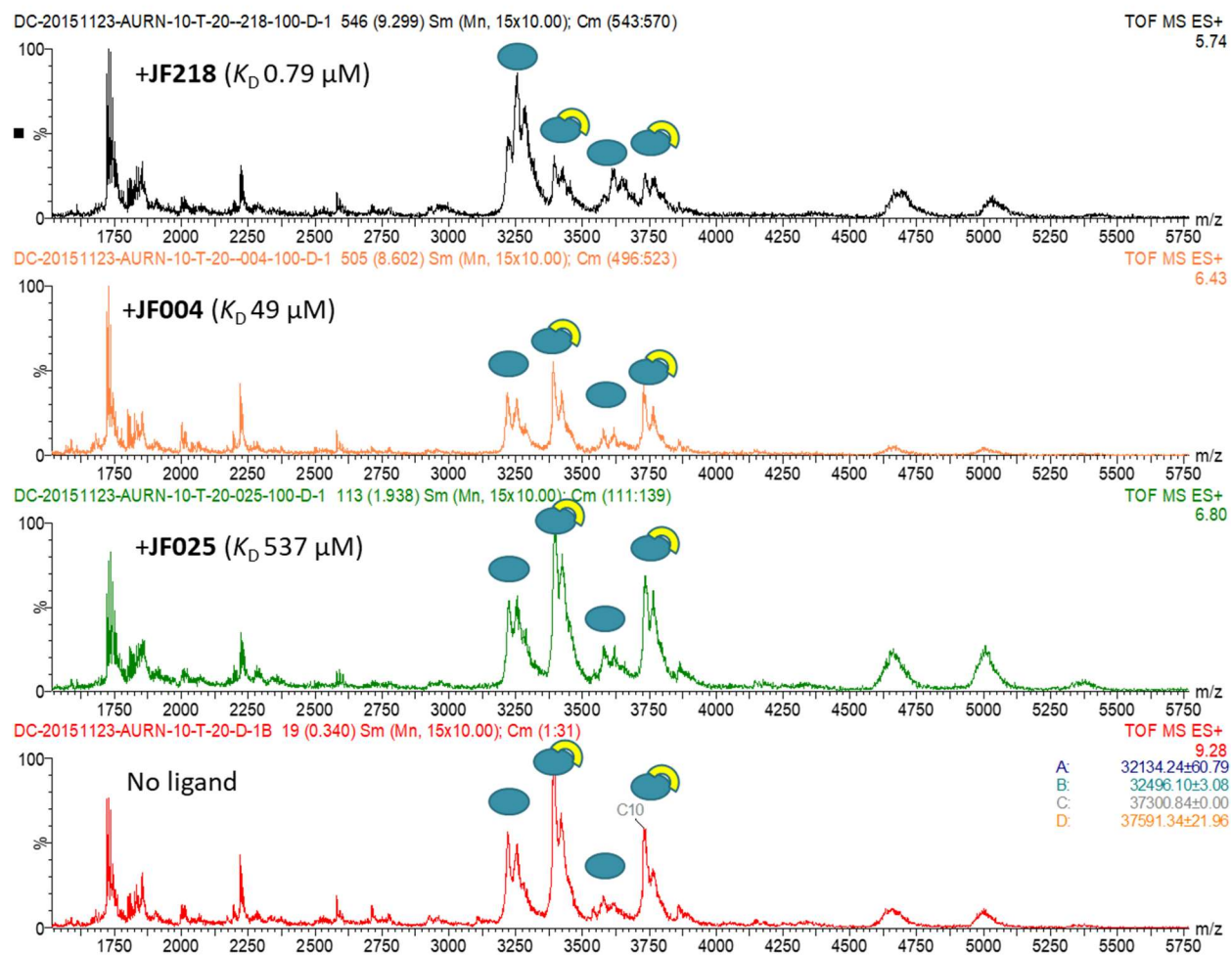


Figure 5.13 Native MS of *AurA-B'* (10 μ M), TPX2⁷⁻⁴³ (20 μ M) in the absence or presence of **JF218**, **JF004** and **JF025** (top to bottom) in 200 mM NH₄OAc containing 1% DMSO.

Since it had not been possible to detect the binding of TPX2 mutants to AurA by native MS, an alternative method was investigated to destabilize the AurA-TPX2 interaction to see if weaker ligands could be tested as protein-protein interaction disruptors by native MS. The collision energy within the trap and transfer regions of the mass spectrometer was raised in order to destabilize the **AurA-B'**-TPX⁷⁻⁴³ complex in the gas phase. However, this did not significantly change the ease at which the ligands dissociated the AurA-TPX2 complex (**Figure 5.14**). This result can be explained by the reasoning that disruption of the AurA-TPX2 complex by small-molecule inhibitors occurs in solution, and that no binding in the gas phase is expected.

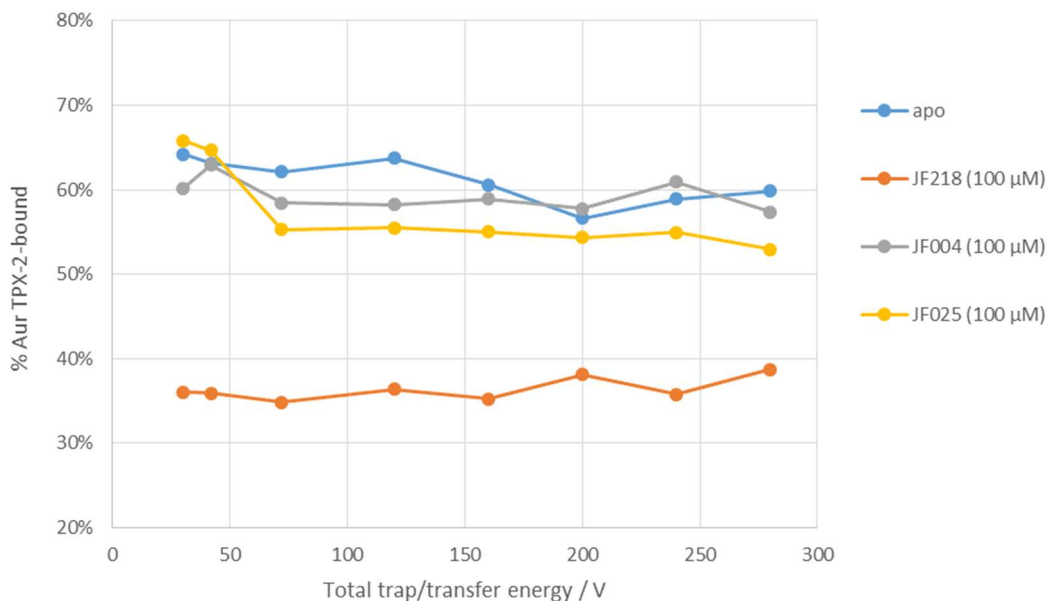


Figure 5.14 The proportion of *AurA-B'* that is bound by TPX-2 as a function of total trap and transfer voltage in the absence of presence of 100 μM of *JF218*, *JF004* or *JF025*.

5.2.4. Summary

The expression of *AurA* was generally low-yielding due to the toxicity of most *AurA* variants to *E. coli*, despite several modifications to the procedure to improve yield. Moreover, isolating *AurA* in a homogeneous form proved to be difficult due to the propensity of *AurA* to phosphorylate itself. Replacing ATP in the buffer with ADP reduced but did not eliminate autophosphorylation, owing to the presence of contaminating ATP in the ADP stock solutions. Additionally, it was discovered that reducing agent needed to be added to the NH_4OAc solution after buffer exchange in order to prevent covalent association of *AurA* or TPX2 via disulfide bridge formation.

The native MS experiments provided important validation of the mechanism of action of the group's in-house *AurA* inhibitors. The results indicated that these inhibitors could disrupt the *AurA*-TPX2 complex, producing free *AurA* species that could be detected in the gas phase. Significantly, the rank order of binding affinities of *AurA* inhibitors determined by ITC was in agreement with the degree of disruption of the *AurA*-TPX2 interaction as observed by native MS. As expected, ATP-competitive inhibitors did not disrupt the *AurA*-TPX2 complex in native MS.

Future work might involve the use of IM-MS (with or without CIU) to investigate the effect that different kinds of inhibitors have on AurA conformation and stability against unfolding.

5.3. Experimental

5.3.1. Materials and methods

All chemicals were of reagent grade quality or better, obtained from commercial suppliers and used without further purification. **JF218**, **JF004** and **JF025** were synthesized by group member Dr Jamy Feng and their binding affinities were measured using ITC by Abell group members. Plasmids, **AurA-D** and TPX2 mutants (M3 and M6) were provided by the Hyvönen group.

5.3.2. Protein expression and purification

5.3.2.1. Aurora A protein and expression and purification

In this project, **AurA-A** and **AurA-B** were expressed in *E. coli* BL21(DE3) cells transformed with the expression vector (pHAT4) carrying the AurA gene and the ampicillin resistance gene using standard molecular biology techniques. The **AurA-C** plasmid contained an additional kanamycin resistance gene. Cells scraped from a fresh transformation were incubated in 2×YT medium containing 100 µg/mL ampicillin (for **AurA-C**, the medium also contained 50 µg/mL kanamycin) until OD₆₀₀ of approximately 0.4–0.5 was reached. Cells were incubated at 4 °C for 30 min and then at 22 °C for 30 min. A second dose of ampicillin (50 µg/mL) was added before expression was induced with 100 µM IPTG. Cells were incubated at 37 °C for 3 h. The bacteria were harvested by centrifugation (4000 *g*, 20 min), washed with 50 mM HEPES pH 7.4, 500 mM NaCl, and then centrifuged again (4000 *g*, 20 min). The pellet was resuspended in lysis buffer containing 50 mM HEPES pH 7.4, 1 M NaCl, 100 mM Mg(OAc)₂, 1mM ATP, 25 mM imidazole, one tablet of protease inhibitors (cOmplete Protease Inhibitor Cocktails, Roche), 500 µL of 2 mg/mL DNase I per 50 mL lysate. Cells were lysed by sonication using 10 s pulses of 50% amplitude at intervals of 30 s each with a total sonication time of 3.5 min. The lysate was centrifuged at 30 000 *g* for 30 min at 4 °C, then clarified by filtration through a 0.45 µm syringe filter.

For nickel-affinity chromatography, the clarified lysate was loaded onto a gravity flow column containing 5 mL Ni Sepharose High Performance (GE Healthcare). The column was washed with 50 mM Hepes pH 7.4, 500 mM NaCl, 100 mM Mg(OAc)₂, 1 mM ATP, 40 mM imidazole, 5 mM β-mercaptoethanol, then eluted with 50 mM Hepes pH 7.4, 500 mM NaCl, 100 mM Mg(OAc)₂, 1 mM ATP, 600 mM imidazole, 5 mM β-mercaptoethanol. Elution was monitored by SDS-PAGE analysis. Fractions containing AurA were pooled and concentrated using a Vivaspin 20 centrifugal concentrator with a 10 kDa molecular weight cut-off before being subjected to gel filtration using a Superdex75 16/60 column (GE Healthcare) previously equilibrated with 50 mM Hepes pH 7.4, 50 mM NaCl, 100 mM Mg(OAc)₂, 1 mM ATP, 5 mM β-mercaptoethanol. Gel filtration was performed at a flow rate of 1 mL/min and elution was again monitored at A₂₈₀ and by SDS-PAGE analysis. Fractions containing purified AurA were pooled and concentrated as before.

To cleave the N-His₆ tag (for **AurA-B** and **AurA-C**), ProTEV (10 U per 100 μg of AurA) (Promega) was added in 1× ProTEV buffer (Promega) containing 1 mM DTT and the mixture was incubated at 4 °C overnight. The cleavage reaction was monitored by SDS-PAGE analysis. After completion of reaction, the solution was passed through a nickel column as before and elution was monitored by SDS-PAGE analysis (the cleaved AurA is eluted in the low-imidazole wash). Fractions containing cleaved AurA were pooled and concentrated before being subjected to gel filtration purification as before. Aliquots of the concentrated protein were flash-frozen in liquid N₂ and stored at -80 °C.

5.3.2.2. TPX2 protein and expression and purification

TPX2⁷⁻⁴³ was expressed in this project as a GB1 fusion protein in *E. coli* BL21(DE3) cells transformed with the expression vector carrying the TPX2 fusion gene and the ampicillin resistance gene using standard molecular biology techniques. Cells scraped from a fresh transformation were incubated in LB medium containing 100 μg/mL ampicillin until OD₆₀₀ of approximately 0.4–0.5 was reached. Expression was induced with 400 μM IPTG, and cells were incubated at 20 °C overnight. The bacteria were harvested by centrifugation (4000 *g*, 20 min), washed with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and then centrifuged again (4000 *g*, 20 min).

The pellet was resuspended in lysis buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM Mg(OAc)₂, one tablet of protease inhibitors (cOmplete Protease Inhibitor Cocktails, Roche), 500 µl of 2 mg/mL DNase I per 50 mL lysate. Cells were lysed by sonication using 10 s pulses of 50% amplitude at intervals of 30 s each with a total sonication time of 3.5 min. The lysate was centrifuged at 30 000 *g* for 30 min at 4 °C, then clarified by filtration through a 0.45 µM syringe filter.

For nickel-affinity chromatography, the clarified lysate was loaded onto a gravity flow column containing 5 mL Ni Sepharose High Performance (GE Healthcare). The column was washed with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, then eluted with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 500 mM imidazole, 5 mM β-mercaptoethanol. Elution was monitored by SDS-PAGE analysis. Fractions containing the TPX2 fusion protein were pooled and concentrated using a Vivaspin 20 centrifugal concentrator with a 10 kDa molecular weight cut-off before being subjected to gel filtration using a Superdex75 16/60 column (GE Healthcare) previously equilibrated with 50 mM Hepes pH 7.4, 50 mM NaCl, 1 mM β-mercaptoethanol. Gel filtration was performed at a flow rate of 1 mL/min and elution was again monitored at A₂₈₀ and by SDS-PAGE analysis. Fractions containing purified TPX2 fusion protein were pooled and concentrated as before.

To cleave the GB1 tag, ProTEV (2U per 100 µg of TPX2 fusion protein) (Promega) was added in 1× ProTEV buffer (Promega) containing 1 mM DTT and the mixture was incubated at 4 °C overnight. The cleavage reaction was monitored by SDS-PAGE analysis. After completion of reaction, the solution was passed through a 10 kDa MWCO filter and the filtrate was aliquoted, flash-frozen in liquid N₂ and stored at –80 °C.

5.3.3. Mass spectrometry

Native nanoelectrospray ionization-ion mobility-mass spectrometry (nESI-IM-MS) spectra were recorded on a Synapt HDMS mass spectrometer (Waters, Manchester, UK) modified for studying high masses (32k quadrupole) and equipped with a traveling wave (TW) IM cell. AurA was

exchanged into 200 mM NH₄OAc (pH 7.0) using Micro Bio-Spin 6 chromatography columns (Bio-Rad, UK). 1 mM β-mercaptoethanol was sometimes added to prevent protein dimerization. 2.5 μL of protein solution was injected into a gold-coated borosilicate emitter (Thermo Scientific, UK) for sampling. Typical conditions were capillary voltage 1.5 kV, cone voltage 50 V, trap collision voltage 20 V, trap DC bias 35 V, transfer collision voltage 20 V, source temperature 20 °C, backing pressure 3–4 mbar, trap pressure 3–4 × 10⁻² mbar, IM (N₂) pressure 5–6 × 10⁻¹ mbar and TOF pressure 7–8 × 10⁻⁷ mbar. All mass spectra were calibrated externally with cesium iodide (10 mg/mL). Data acquisition and processing were performed using MassLynx 4.1 (Waters) and DriftScope 2.5 (Waters).

High-resolution liquid chromatography mass spectrometry (LC-MS) was performed on a Xevo G2-S Q-TOF UPLC instrument (Waters) coupled to an Acquity UPLC system. Samples were eluted through an Acquity UPLC BEH300 C4 column (1.7 μm, 2.1 × 50 mm) using a gradient of 95% Solvent A for 5.21 min, 100% Solvent B for 1 min, and 100% Solvent A for 1 min, where Solvent A = H₂O with 0.1% formic acid and Solvent B = 95% MeCN/5% H₂O containing 0.01% formic acid. A flow rate of 0.2 mL min⁻¹ was used and the total run time was 7.29 minutes. The electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 40 V. Nitrogen was used as the desolvation gas at a total flow of 850 L h⁻¹. Data acquisition and processing was performed using MassLynx 4.1 (Waters) with deconvoluted mass spectra reconstructed from the ion series using the pre-installed MaxEnt algorithm.

6. Conclusions and outlook

At the onset of this project, native mass spectrometry was still an emergent technique in fragment-based drug discovery, and indeed only a handful of studies on the use of native MS for any type of ligand screening had been reported up to that time. Few researchers had employed native MS to investigate the disruption of protein-protein interactions by small molecules, and none had used this technique for the investigation of protein-DNA inhibition. Mass spectrometry also did not feature in our own group's well-established biophysical cascade for fragment screening.

Native MS has now gained increasing acceptance in drug discovery because of the orthogonal information that it can provide regarding biomolecular structure and conformation, although it is still far from mainstream use. One limitation that has historically hindered the wider adoption of native MS for drug screening is its low throughput. Without automation capability, a typical native MS experiment in this project required (after optimization) about 10 min to carry out, from sample loading to the acquisition of a suitable spectrum. An early study (2012) in the literature used the Advion Triversa NanoMate, a microfluidic chip-based sample injection and liquid handling robot, to screen a ligand library at a rate of about 1.5 min per sample.^[63] Today, ultrahigh-throughput workflows have been developed that can analyze multiple samples per second, such as the SCIEX ECHO MS System employing acoustic ejection mass spectrometry (AEMS)^[219] or AstraZeneca's acoustic mist ionisation mass spectrometry (AMI-MS) platform.^[220] However, these setups are expensive and would likely only be of interest to pharmaceutical companies or large academic groups dedicated to drug discovery. The use of automated platforms for fragment screening also requires that the protein target is well-characterized, homogenous, and robust.

This project was the first to use native MS to screen ligands against a protein-DNA interaction (EthR-DNA), and also confirmed the applicability of native MS to investigate inhibition of a protein-protein interaction (AurA-TPX2). There are several advantages to screening ligands against a protein-DNA or a protein-protein complex rather than against the isolated protein

partners. Firstly, measuring the disruption of the complex means that the experiment is a functional assay rather than purely a binding one. Ligands may bind to a protein without disrupting its interactions with its partners, thus resulting in false positives if only the single protein target is screened. In this project, screening of fragments against the EthR-DNA complex by native MS identified hits with higher *in vitro* activity compared to differential scanning fluorimetry (DSF) screening against EthR alone. The second advantage is that the disruption of a biomolecular complex is nearly always unambiguous, whereas binding of a ligand to a single protein target can be complicated by heterogeneity of the protein sample, leading to poorly resolved and broad signals in the native mass spectrum that cannot be definitively assigned as protein-ligand peaks. This was observed in this project where AurA was expressed in a mixture of unphosphorylated and phosphorylated states, as well as sometimes methionylated and demethionylated states as well. Additionally, both AurA and EthR samples that were expressed in this project often contained bound but unidentifiable small molecules even in their supposed apo states, further complicating the analysis of putative protein-ligand interactions. Finally, these issues are also exacerbated by the low molecular weight of fragments, as these produce only small m/z shifts upon binding, especially if the protein is large. However, there are also drawbacks to screening against a protein-protein or protein-DNA complex. Since quaternary interactions can be strong, the screened ligands must be sufficiently potent to disrupt the complex in solution for their activity to be detected by native MS. This issue is especially pronounced for fragments which generally have weak binding affinities due to their small sizes. To overcome this, the quaternary interactions within the complex can be weakened to make them easier to disrupt by fragments, such as through mutagenesis of key interface residues, or by using only a partial (but still functional) complex. The latter approach was successfully employed in this project where we screened fragments against the partial EthR₄-DNA_{R31} complex rather than the full-length EthR₆-DNA₆₂ complex. Finally, it should be noted that not all targets will have pharmacologically relevant protein-protein or protein-DNA interactions that can be screened against, and even for those that do, the complex must be sufficiently stable in the gas-phase to allow for screening by native MS.

Another general consideration for fragment screening by native MS are the concentrations of protein and ligand employed. Since fragments are weakly binding (typically $K_d > 100 \mu\text{M}$), higher concentrations of ligand must be used to ensure that the binding event is detectable. In this project, fragments concentrations of up to 1 mM were used for screening. However, the use of high concentrations of ligand can introduce other concerns. First, since ligands are typically dissolved as stock solutions in DMSO, using high ligand concentrations would necessarily increase the DMSO content in the final sample as well. For example, adding a compound to a final concentration of 1 mM from a standard 100 mM stock solution would result in a 1% (v/v) DMSO solution, which is probably tolerated in most cases. However, for compounds that are only soluble up to 10 mM in DMSO, adding it at 1 mM would create a 10% (v/v) DMSO solution. As shown in this project, such high DMSO concentrations can have significant effects on the conformation and stability of proteins. Therefore, each protein target should be tested for stability and robustness in DMSO before screening with ligands is carried out. In this context, IM-MS (including CIU) can be used to illuminate subtle differences in protein conformation and stability induced by DMSO that may not be detectable using solution techniques. Secondly, high concentrations of ligand can lead to non-specific binding that may complicate the interpretation of ligand binding results, notwithstanding the number of different approaches that have been proposed to allow for this. On the other hand, protein concentrations should be kept as low as possible (typically $<10 \mu\text{M}$) in order to prevent non-specific protein-protein interactions, but still high enough to maintain sufficient signal intensity. Again, the concentration of protein should be optimized for each target before embarking on a fragment screening campaign.

Towards the future, there are several avenues along which this project could be pursued further. Our study has provided evidence that DMSO can modulate protein size and stability through the dual effect of shifting charge state distributions as well as acting at the level of individual charge states. However, the exact mechanisms by which DMSO achieves this are still unclear. More careful control of experimental parameters to ensure stringent reproducibility could allow peak widths to be analyzed for the degree of DMSO adduction under different activating conditions. Using smaller proteins (such as lysozyme or myoglobin) coupled with high-resolution mass

spectrometers (such as an Orbitrap) might even allow such adducts to be resolved down to the single-molecule level. This information could help in understanding the role that DMSO plays in the fine balance between evaporative cooling and collisional heating as the protein ions travel through the various stages of the mass spectrometer. Additionally, the use of mass spectrometry coupled with gas-phase circular dichroism or optical/infrared spectroscopy could allow the effects of DMSO on protein structure and stability to be studied from a different angle, complementing the ion mobility results.

In the EthR study, several open lines of investigation could not be further explored due to time constraints. This study was notable because it provided evidence that six subunits of EthR bound to its operator DNA, instead of eight subunits as had been previously reported based on SPR. Our data also revealed significant heterogeneity in the nature of the complexes formed, which could account for why no X-ray crystal structure of the EthR-DNA complex has yet been reported. The ability to simultaneously resolve and characterize multiple complexes from a heterogeneous mixture is one of the most distinctive advantages of native MS over other bioanalytical techniques. As the exact binding mode of EthR to DNA was (and still is) unknown, preliminary experiments were conducted that aimed to narrow down the precise binding sites on DNA by using oligonucleotides of different length, and/or with specific base sequences scrambled. Although some initial progress was made on this front (e.g. the scrambling of some base sequences weakened EthR binding more than at other sequences), the campaign proved to be quite tedious due to the sheer number of different DNA lengths and sequences that could potentially be tested. IM-MS measurements were also performed that revealed CCS values of 6687 and 8259 Å for the EthR₄-DNA₆₂ and EthR₆-DNA₆₂ complexes respectively. However, the relatively low resolution of this technique meant that these measurements were not too useful in deducing the arrangement of EthR subunits within the complex. It is envisioned that other techniques, such as cryo-electron microscopy or *in silico* molecular modeling, might be needed to further elucidate the structural characteristics of this interaction. Towards the end of this project, the discovery of an EthR analogue (termed “EthR2”) was reported in the literature as a potential anti-tubercular target.^[221] I expressed and purified this protein in high yield from a

plasmid designed by our Biochemistry collaborators, and preliminary native MS results suggested that EthR2 bound to DNA as a mixture of dimers and tetramers. Just like EthR, no X-ray crystal structure of the EthR2-DNA complex has yet been reported, presumably due to this heterogeneity. It would be interesting to use native MS to further characterize this interaction, and to see whether fragments could be screened against this complex in similar fashion to what was achieved with EthR.

On the CoA front, this study has demonstrated that the dimeric *E. coli* CoaB construct could be an effective surrogate for the native CoaBC dodecamer for ligand screening by native MS. In addition to what was described in this thesis, I also successfully recorded the native mass spectra for both *M. tuberculosis* and *M. smegmatis* CoaBC dodecamers after extensive optimization, providing clear evidence for their oligomeric status at a time when the X-ray crystal structure of these complexes had not yet been resolved to a sufficient degree.^[222] CID experiments showed that the dissociation of these dodecamers followed the asymmetric charge partitioning pathway, with the *M. tuberculosis* CoaBC dodecamer (547 kDa) centered around the 59+ charge state dissociating into a “stripped” undecamer (502 kDa) centered around 34+ and a highly-charged monomer (46 kDa) centered around 25+. It would be worthwhile to further explore the interactions and stability of this large multiprotein complex under various conditions, including in the presence of natural substrates or small molecule inhibitors.

Finally, there are several ways that the Aurora A project could be advanced further, notwithstanding the difficulties in obtaining homogeneous protein samples that were also free of contaminants. CIU could be employed to investigate the stability of different AurA mutants, as well as to distinguish between inhibitors with different mechanisms, similar to how CIU was previously used to profile Type I and Type II inhibitors of Abl kinase.^[196] Preliminary results revealed minor differences in the CIU fingerprints of AurA mutants with dasatinib and alisertib (Type I inhibitors), imatinib (Type II inhibitor), or AurkinA and JF218 (TPX2 site binders). Statistical analysis (e.g. χ^2 correlation) would be needed to quantify these smaller differences that are not immediately apparent by eye. Meanwhile, the screening of small molecules against the AurA-

TPX2 protein-protein interaction could be investigated further. This study failed to sufficiently weaken the AurA-TPX2 interaction for fragment screening, but further testing of both AurA and/or TPX2 mutants may yield complexes that are weak enough to be disrupted by fragments.

Overall, native MS deserves more attention in drug discovery workflows, including fragment-based drug discovery. It provides a wealth of information (such as data on protein conformation, stability and stoichiometry) that are orthogonal to other biophysical techniques, and can also tackle, or at least recognize, cases of protein heterogeneity that may not be detectable using other techniques. However, mass spectrometry systems capable of performing more advanced analyses (such as IM-MS) are still expensive and require specialized personnel. It is hopeful that in the future, continual advances in technology and engineering could drive miniaturization and automation of mass spectrometry to such an extent that benchtop mass spectrometers capable of diverse functions (e.g. automated ligand screening and/or recording of CIU fingerprints) could become routinely used by non-specialist groups. At the theoretical level, there is room for improvement in our knowledge of ionization mechanisms, as well as in our understanding of how closely gas-phase observations correlate with solution or *in vivo* conditions. Further demystifying the exact processes and events that occur between ionization and detection would be important for native MS to gain wider acceptance as an essential technique for drug discovery, similar to how mass spectrometry is now an indispensable tool for the study of proteomics today.

References

- [1] G. S. Sittampalam, S. D. Kahl, W. P. Janzen, *Curr. Opin. Chem. Biol.* **1997**, *1*, 384-391.
- [2] J. J. Burbaum, N. H. Sigal, *Curr. Opin. Chem. Biol.* **1997**, *1*, 72-78.
- [3] P. B. Fernandes, *Curr. Opin. Chem. Biol.* **1998**, *2*, 597-603.
- [4] K. H. Bleicher, H.-J. Bohm, K. Muller, A. I. Alanine, *Nat. Rev. Drug Discov.* **2003**, *2*, 369-378.
- [5] J. W. Szostak, *Chem. Rev.* **1997**, *97*, 347-348.
- [6] H. Mario Geysen, F. Schoenen, D. Wagner, R. Wagner, *Nat. Rev. Drug Discov.* **2003**, *2*, 222-230.
- [7] E. M. Gordon, M. A. Gallop, D. V. Patel, *Acc. Chem. Res.* **1996**, *29*, 144-154.
- [8] S. C. Gupta, B. Sung, S. Prasad, L. J. Webb, B. B. Aggarwal, *Trends Pharmacol. Sci.* **2013**, *34*, 508-517.
- [9] G. L. Law, J. Tisoncik-Go, M. J. Korth, M. G. Katze, *Curr. Opin. Immunol.* **2013**, *25*, 588-592.
- [10] T. I. Oprea, J. E. Bauman, C. G. Bologa, T. Buranda, A. Chigaev, B. S. Edwards, J. W. Jarvik, H. D. Gresham, M. K. Haynes, B. Hjelle, R. Hromas, L. Hudson, D. A. Mackenzie, C. Y. Muller, J. C. Reed, P. C. Simons, Y. Smagley, J. Strouse, Z. Surviladze, T. Thompson, O. Ursu, A. Waller, A. Wandinger-Ness, S. S. Winter, Y. Wu, S. M. Young, R. S. Larson, C. Willman, L. A. Sklar, *Drug Discov. Today Ther. Strateg.* **2011**, *8*, 61-69.
- [11] S. L. Schreiber, *Science* **2000**, *287*, 1964-1969.
- [12] W. R. J. D. Galloway, A. Isidro-Llobet, D. R. Spring, *Nat. Commun.* **2010**, *1*, 80.
- [13] C. J. O' Connor, H. S. G. Beckmann, D. R. Spring, *Chem. Soc. Rev.* **2012**, *41*, 4444-4456.
- [14] H. van Hattum, H. Waldmann, *J. Am. Chem. Soc.* **2014**, *136*, 11853-11859.
- [15] S. Wetzel, R. S. Bon, K. Kumar, H. Waldmann, *Angew. Chem. Int. Ed.* **2011**, *50*, 10800-10826.
- [16] H. Lachance, S. Wetzel, K. Kumar, H. Waldmann, *J. Med. Chem.* **2012**, *55*, 5989-6001.
- [17] G. M. Cragg, D. J. Newman, *BBA - Gen. Subjects* **2013**, *1830*, 3670-3695.
- [18] D. E. Scott, A. G. Coyne, S. A. Hudson, C. Abell, *Biochemistry* **2012**, *51*, 4990-5003.
- [19] W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4046-4050.
- [20] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531-1534.

- [21] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, *Nat. Rev. Drug Discov.* **2016**, *15*, 605-619.
- [22] G. Bollag, J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop, P. Hirth, *Nat. Rev. Drug Discov.* **2012**, *11*, 873-886.
- [23] A. Mullard, *Nat. Rev. Drug Discov.* **2016**, *15*, 147-149.
- [24] H. A. Burris, *Expert Rev. Anticancer Ther.* **2018**, *18*, 201-213.
- [25] C. A. Lipinski, *Drug Discov. Today Technol.* **2004**, *1*, 337-341.
- [26] P. Ertl, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374-380.
- [27] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876-877.
- [28] T. Fink, J.-L. Reymond, *J. Chem. Inf. Model.* **2007**, *47*, 342-353.
- [29] E. H. Mashalidis, P. Śledź, S. Lang, C. Abell, *Nat. Protocols* **2013**, *8*, 2309-2324.
- [30] F. H. Niesen, H. Berglund, M. Vedadi, *Nat. Protocols* **2007**, *2*, 2212-2221.
- [31] C. A. Lepre, in *Methods Enzymol., Vol. Volume 493* (Ed.: C. K. Lawrence), Academic Press, **2011**, pp. 219-239.
- [32] M. J. Harner, A. O. Frank, S. W. Fesik, *J. Biomol. NMR* **2013**, *56*, 65-75.
- [33] C. A. Shepherd, A. L. Hopkins, I. Navratilova, *Prog. Biophys. Mol. Biol.* **2014**, *116*, 113-123.
- [34] M. I. Recht, V. Nienaber, F. E. Torres, *Methods Enzymol.* **2016**, *567*, 47-69.
- [35] D. Patel, J. D. Bauman, E. Arnold, *Prog. Biophys. Mol. Biol.* **2014**, *116*, 92-100.
- [36] P. M. Collins, J. T. Ng, R. Talon, K. Nekrosiute, T. Krojer, A. Douangamath, J. Brandao-Neto, N. Wright, N. M. Pearce, F. von Delft, *Acta Cryst. D* **2017**, *73*, 246-255.
- [37] M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle, H. Jhoti, *J. Med. Chem.* **2005**, *48*, 403-413.
- [38] S.-J. Hyung, B. T. Ruotolo, *Proteomics* **2012**, *12*, 1547-1564.
- [39] K. Breuker, M. Jin, X. Han, H. Jiang, F. W. McLafferty, *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1045-1053.
- [40] G. R. Hilton, J. L. P. Benesch, *J. R. Soc. Interface* **2012**, *9*, 801-816.
- [41] A. Konijnenberg, A. Butterer, F. Sobott, *Biochim. Biophys. Acta* **2013**, *1834*, 1239-1256.
- [42] K. J. Pacholarz, R. A. Garlish, R. J. Taylor, P. E. Barran, *Chem. Soc. Rev.* **2012**, *41*, 4335-4355.

- [43] S. S. Rubakhin, J. C. Jurchen, E. B. Monroe, J. V. Sweedler, *Drug Discov. Today* **2005**, *10*, 823-837.
- [44] S.-A. Poulsen, *Aust. J. Chem.* **2013**, *66*, 1495-1501.
- [45] L. Pedro, J. R. Quinn, *Molecules* **2016**, *21*, 984.
- [46] Daniel S.-H. Chan, Andrew J. Whitehouse, Anthony G. Coyne, C. Abell, *Essays Biochem.* **2017**.
- [47] M. Wilm, M. Mann, *Anal. Chem.* **1996**, *68*, 1-8.
- [48] L. Konermann, E. Ahadi, A. D. Rodriguez, S. Vahidi, *Anal. Chem.* **2013**, *85*, 2-9.
- [49] J. C. Jurchen, E. R. Williams, *J. Am. Chem. Soc.* **2003**, *125*, 2817-2826.
- [50] J. C. Jurchen, D. E. Garcia, E. R. Williams, *J. Am. Soc. Mass. Spectrom.* **2004**, *15*, 1408-1415.
- [51] J. T. S. Hopper, N. J. Oldham, *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1851-1858.
- [52] J. D. Eschweiler, J. N. Rabuck-Gibbons, Y. Tian, B. T. Ruotolo, *Anal. Chem.* **2015**, *87*, 11516-11522.
- [53] T. M. Allison, E. Reading, I. Liko, A. J. Baldwin, A. Laganowsky, C. V. Robinson, *Nat. Commun.* **2015**, *6*, 8551.
- [54] S. M. Dixit, D. A. Polasky, B. T. Ruotolo, *Curr. Opin. Chem. Biol.* **2018**, *42*, 93-100.
- [55] A. C. Leney, A. J. R. Heck, *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 5-13.
- [56] S. Yin, Y. Xie, J. Loo, *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1199-1208.
- [57] C. Bich, S. Baer, M. Jecklin, R. Zenobi, *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 286-289.
- [58] E. Freire, *Drug Discov. Today* **2008**, *13*, 869-874.
- [59] M. Bakhtiari, L. Konermann, *J. Phys. Chem. B* **2019**, *123*, 1784-1796.
- [60] S. Mehmood, T. M. Allison, C. V. Robinson, *Annu. Rev. Phys. Chem.* **2015**, *66*, 453-474.
- [61] J. L. P. Benesch, B. T. Ruotolo, *Curr. Opin. Struct. Biol.* **2011**, *21*, 641-649.
- [62] E. N. Kitova, A. El-Hawiet, P. D. Schnier, J. S. Klassen, *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 431-441.
- [63] H. J. Maple, R. A. Garlish, L. Rigau-Roca, J. Porter, I. Whitcombe, C. E. Prosser, J. Kennedy, A. J. Henry, R. J. Taylor, M. P. Crump, J. Crosby, *J. Med. Chem.* **2012**, *55*, 837-851.
- [64] C. A. Keetch, H. Hernández, A. Sterling, M. Baumert, M. H. Allen, C. V. Robinson, *Anal. Chem.* **2003**, *75*, 4937-4941.
- [65] S. Zhang, C. K. Van Pelt, D. B. Wilson, *Anal. Chem.* **2003**, *75*, 3010-3018.

- [66] P. N. Kelly, A. Strasser, *Cell Death Differ.* **2011**, *18*, 1414-1424.
- [67] L. A. Woods, O. Dolezal, B. Ren, J. H. Ryan, T. S. Peat, S.-A. Poulsen, *J. Med. Chem.* **2016**, *59*, 2192-2204.
- [68] M. Göth, V. Badock, J. Weiske, K. Pagel, B. Kuroпка, *ChemMedChem* **2017**, *12*, 1201-1211.
- [69] H. Vu, C. Roullier, M. Campitelli, K. R. Trenholme, D. L. Gardiner, K. T. Andrews, T. Skinner-Adams, G. J. Crowther, W. C. Van Voorhis, R. J. Quinn, *ACS Chem. Biol.* **2013**, *8*, 2654-2659.
- [70] J. L. Whittingham, I. Leal, C. Nguyen, G. Kasinathan, E. Bell, A. F. Jones, C. Berry, A. Benito, J. P. Turkenburg, E. J. Dodson, L. M. R. Perez, A. J. Wilkinson, N. G. Johansson, R. Brun, I. H. Gilbert, D. G. Pacanowska, K. S. Wilson, *Structure* **2005**, *13*, 329-338.
- [71] B. Yang, Y. J. Feng, H. Vu, B. McCormick, J. Rowley, L. Pedro, G. J. Crowther, W. C. Van Voorhis, P. I. Forster, R. J. Quinn, *J. Biomol. Screen.* **2016**, *21*, 194-200.
- [72] C. Agop-Nersesian, B. Naissant, F. Ben Rached, M. Rauch, A. Kretzschmar, S. Thiberge, R. Menard, D. J. P. Ferguson, M. Meissner, G. Langsley, *PLoS Pathog.* **2009**, *5*, e1000270.
- [73] K. J. McLean, P. Carroll, D. G. Lewis, A. J. Dunford, H. E. Seward, R. Neeli, M. R. Cheesman, L. Marsollier, P. Douglas, W. E. Smith, I. Rosenkrands, S. T. Cole, D. Leys, T. Parish, A. W. Munro, *J. Biol. Chem.* **2008**, *283*, 33406-33416.
- [74] K. M. Duffell, S. A. Hudson, K. J. McLean, A. W. Munro, C. Abell, D. Matak-Vinkovic, *Anal. Chem.* **2013**, *85*, 5707-5714.
- [75] M. E. Kavanagh, A. G. Coyne, K. J. McLean, G. G. James, C. W. Levy, L. B. Marino, L. P. S. de Carvalho, D. S. H. Chan, S. A. Hudson, S. Surade, D. Leys, A. W. Munro, C. Abell, *J. Med. Chem.* **2016**, *59*, 3272-3302.
- [76] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V. Robinson, *Nat. Protocols* **2008**, *3*, 1139-1152.
- [77] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinis, R. H. Bateman, M. T. Bowers, J. H. Scrivens, *Int. J. Mass spectrom.* **2007**, *261*, 1-12.
- [78] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, R. H. Bateman, *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401-2414.
- [79] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Anal. Chem.* **2010**, *82*, 9557-9565.
- [80] Y. Berezovskaya, M. Porrini, C. Nortcliffe, P. E. Barran, *Analyst* **2015**, *140*, 2847-2856.
- [81] J. Stojko, S. Fiulaine, S. Petiot-Becard, A. Van Dorselaer, T. Meinel, C. Giglione, S. Cianferani, *Analyst* **2015**, *140*, 7234-7245.

- [82] M. J. Chalmers, S. A. Busby, B. D. Pascal, G. M. West, P. R. Griffin, *Expert Rev. Proteomics* **2011**, *8*, 43-59.
- [83] D. P. Marciano, V. Dharmarajan, P. R. Griffin, *Curr. Opin. Struct. Biol.* **2014**, *28*, 105-111.
- [84] S. Y. Dai, M. J. Chalmers, J. Bruning, K. S. Bramlett, H. E. Osborne, C. Montrose-Rafizadeh, R. J. Barr, Y. Wang, M. Wang, T. P. Burris, J. A. Dodge, P. R. Griffin, *Proc. Natl. Acad. Sci.* **2008**, *105*, 7171-7176.
- [85] H. Wei, J. Mo, L. Tao, R. J. Russell, A. A. Tymiak, G. Chen, R. E. Jacob, J. R. Engen, *Drug Discov. Today* **2014**, *19*, 95-102.
- [86] S. A. Berkowitz, J. R. Engen, J. R. Mazzeo, G. B. Jones, *Nat. Rev. Drug Discov.* **2012**, *11*, 527-540.
- [87] D. Houde, S. A. Berkowitz, J. R. Engen, *J. Pharm. Sci.* **2011**, *100*, 2071-2086.
- [88] M. W. Carson, J. Zhang, M. J. Chalmers, W. P. Bocchinfuso, K. D. Holifield, T. Masquelin, R. E. Stites, K. R. Stayrook, P. R. Griffin, J. A. Dodge, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 3459-3463.
- [89] X. Chen, S. Qin, S. Chen, J. Li, L. Li, Z. Wang, Q. Wang, J. Lin, C. Yang, W. Shui, *Sci. Rep.* **2015**, *5*, 8361.
- [90] S. Qin, Y. Ren, X. Fu, J. Shen, X. Chen, Q. Wang, X. Bi, W. Liu, L. Li, G. Liang, C. Yang, W. Shui, *Anal. Chim. Acta* **2015**, *886*, 98-106.
- [91] M. H. Powdrill, J. A. Bernatchez, M. Götte, *Viruses* **2010**, *2*, 2169-2195.
- [92] M.-D. Duong-Thi, E. Meiby, M. Bergström, T. Fex, R. Isaksson, S. Ohlson, *Anal. Biochem.* **2011**, *414*, 138-146.
- [93] M.-D. Duong-Thi, M. Bergström, T. Fex, R. Isaksson, S. Ohlson, *J. Biomol. Screen.* **2013**, *18*, 160-171.
- [94] E. Meiby, S. Knapp, J. M. Elkins, S. Ohlson, *Anal. Bioanal. Chem.* **2012**, *404*, 2417-2425.
- [95] E. Meiby, H. Simmonite, L. le Strat, B. Davis, N. Matassova, J. D. Moore, M. Mrosek, J. Murray, R. E. Hubbard, S. Ohlson, *Anal. Chem.* **2013**, *85*, 6756-6766.
- [96] O. Potterat, M. Hamburger, *Nat. Prod. Rep.* **2013**, *30*, 546-564.
- [97] A. Römpf, B. Spengler, *Histochem. Cell Biol.* **2013**, *139*, 759-783.
- [98] K. Rajabi, A. E. Ashcroft, S. E. Radford, *Methods* **2015**, *89*, 13-21.
- [99] T. Wyttenbach, M. T. Bowers, *J. Phys. Chem. B* **2011**, *115*, 12266-12275.
- [100] F. Lanucara, S. W. Holman, C. J. Gray, C. E. Eyers, *Nat. Chem.* **2014**, *6*, 281-294.

- [101] M. Sharon, A. Horovitz, *Curr. Opin. Struct. Biol.* **2015**, *34*, 7-16.
- [102] R. Macarrón, R. P. Hertzberg, *Mol. Biotechnol.* **2010**, *47*, 270-285.
- [103] A. Tjernberg, N. Markova, W. J. Griffiths, D. Hallén, *J. Biomol. Screen.* **2006**, *11*, 131-137.
- [104] H. J. Sterling, A. F. Kintzer, G. K. Feld, C. A. Cassou, B. A. Krantz, E. R. Williams, *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 191-200.
- [105] H. J. Sterling, J. S. Prell, C. A. Cassou, E. R. Williams, *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 1178-1186.
- [106] R. R. O. Loo, R. Lakshmanan, J. A. Loo, *J. Am. Soc. Mass. Spectrom.* **2014**, *25*, 1675-1693.
- [107] E. R. Dickinson, E. Jurneczko, J. Nicholson, T. R. Hupp, J. Zawacka-Pankau, G. Selivanova, P. E. Barran, *Front. Mol. Biosci.* **2015**, *2*, 39.
- [108] M. Landreh, G. Alvelius, J. Johansson, H. Jörnvall, *Anal. Chem.* **2014**, *86*, 4135-4139.
- [109] D. Bagal, E. N. Kitova, L. Liu, A. El-Hawiet, P. D. Schnier, J. S. Klassen, *Anal. Chem.* **2009**, *81*, 7801-7806.
- [110] J. T. S. Hopper, K. Sokratous, N. J. Oldham, *Anal. Biochem.* **2012**, *421*, 788-790.
- [111] D. Cubrilovic, R. Zenobi, *Anal. Chem.* **2013**, *85*, 2724-2730.
- [112] L. Han, S.-J. Hyung, J. J. S. Mayers, B. T. Ruotolo, *J. Am. Chem. Soc.* **2011**, *133*, 11358-11367.
- [113] L. Han, S.-J. Hyung, B. T. Ruotolo, *Angew. Chem. Int. Ed.* **2012**, *51*, 5692-5695.
- [114] L. Han, B. T. Ruotolo, *Int. J. Ion Mobil. Spec.* **2013**, *16*, 41-50.
- [115] J. Freeke, M. F. Bush, C. V. Robinson, B. T. Ruotolo, *Chem. Phys. Lett.* **2012**, *524*, 1-9.
- [116] O. Livnah, E. A. Bayer, M. Wilchek, J. L. Sussman, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5076-5080.
- [117] M. D. Driscoll, K. J. McLean, C. Levy, N. Mast, I. A. Pikuleva, P. Lafite, S. E. J. Rigby, D. Leys, A. W. Munro, *J. Biol. Chem.* **2010**, *285*, 38270-38282.
- [118] H. Ouellet, J. B. Johnston, P. R. Ortiz de Montellano, *Trends Microbiol.* **2011**, *19*, 530-539.
- [119] A. K. Pandey, C. M. Sasseti, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4376-4380.
- [120] D. F. Senechal, D. C. Teller, *Biochemistry* **1981**, *20*, 3076-3083.
- [121] I. A. Kaltashov, R. R. Abzalimov, *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1239-1246.
- [122] Z. W. Yang, S. W. Tendian, W. M. Carson, W. J. Brouillette, L. J. Delucas, C. G. Brouillette, *Protein Sci.* **2004**, *13*, 830-841.

- [123] K. Pagel, S. J. Hyung, B. T. Ruotolo, C. V. Robinson, *Anal. Chem.* **2010**, *82*, 5363-5372.
- [124] E. Jurneczko, P. E. Barran, *Analyst* **2011**, *136*, 20-28.
- [125] Z. Hall, C. V. Robinson, *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 1161-1168.
- [126] R. Salbo, M. F. Bush, H. Naver, I. Campuzano, C. V. Robinson, I. Pettersson, T. J. D. Jørgensen, K. F. Haselmann, *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181-1193.
- [127] I. K. Voets, W. A. Cruz, C. Moitzi, P. Lindner, E. P. G. Arêas, P. Schurtenberger, *J. Phys. Chem. B* **2010**, *114*, 11875-11883.
- [128] R. P. A. Hartman, D. J. Brunner, D. M. A. Camelot, J. C. M. Marijnissen, B. Scarlett, *J. Aerosol Sci* **2000**, *31*, 65-95.
- [129] I. A. Kaltashov, A. Mohimen, *Anal. Chem.* **2005**, *77*, 5370-5379.
- [130] Z. Hall, A. Politis, M. F. Bush, L. J. Smith, C. V. Robinson, *J. Am. Chem. Soc.* **2012**, *134*, 3429-3438.
- [131] T. Arakawa, Y. Kita, S. N. Timasheff, *Biophys. Chem.* **2007**, *131*, 62-70.
- [132] P. A. Chrisman, K. A. Newton, G. E. Reid, J. M. Wells, S. A. McLuckey, *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2334-2340.
- [133] J. Qin, R. Perera, L. L. Lovelace, J. H. Dawson, L. Lebioda, *Biochemistry* **2006**, *45*, 3170-3177.
- [134] Q. Li, P. Mabrouk, *J. Biol. Inorg. Chem.* **2003**, *8*, 83-94.
- [135] J. Kuper, T. S. Wong, D. Roccatano, M. Wilmanns, U. Schwaneberg, *J. Am. Chem. Soc.* **2007**, *129*, 5786-5787.
- [136] R. E. Bornschein, S. Niu, J. Eschweiler, B. T. Ruotolo, *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 41-49.
- [137] M. Almira Correia, P. R. Sinclair, F. De Matteis, *Drug Metab. Rev.* **2011**, *43*, 1-26.
- [138] M. B. Cammarata, J. S. Brodbelt, *Chem Sci* **2015**, *6*, 1324-1333.
- [139] V. D'Atri, M. Porrini, F. Rosu, V. Gabelica, *J. Mass Spectrom.* **2015**, *50*, 711-726.
- [140] World Health Organization, **2018**.
- [141] T. A. Vannelli, A. Dykman, P. R. Ortiz de Montellano, *J. Biol. Chem.* **2002**, *277*, 12824-12829.
- [142] X. Duan, X. Xiang, J. Xie, *FEMS Microbiol. Lett.* **2014**, *360*, 87-99.
- [143] J. Engohang-Ndong, D. Baillat, M. Aumercier, F. Bellefontaine, G. S. Besra, C. Locht, A. R. Baulard, *Mol. Microbiol.* **2004**, *51*, 175-188.

- [144] N. Willand, B. Dirie, X. Carette, P. Bifani, A. Singhal, M. Desroses, F. Leroux, E. Willery, V. Mathys, R. Deprez-Poulain, G. Delcroix, F. Frenois, M. Aumercier, C. Loch, V. Villeret, B. Deprez, A. R. Baulard, *Nat. Med.* **2009**, *15*, 537-544.
- [145] M. Flipo, M. Desroses, N. Lecat-Guillet, B. Dirié, X. Carette, F. Leroux, C. Piveteau, F. Demirkaya, Z. Lens, P. Rucktooa, V. Villeret, T. Christophe, H. K. Jeon, C. Loch, P. Brodin, B. Déprez, A. R. Baulard, N. Willand, *J. Med. Chem.* **2011**, *54*, 2994-3010.
- [146] M. Flipo, N. Willand, N. Lecat-Guillet, C. Hounsou, M. Desroses, F. Leroux, Z. Lens, V. Villeret, A. Wohlkönig, R. Wintjens, T. Christophe, H. Kyoung Jeon, C. Loch, P. Brodin, A. R. Baulard, B. Déprez, *J. Med. Chem.* **2012**, *55*, 6391-6402.
- [147] M. Flipo, M. Desroses, N. Lecat-Guillet, B. Villemagne, N. Blondiaux, F. Leroux, C. Piveteau, V. Mathys, M.-P. Flament, J. Siepmann, V. Villeret, A. Wohlkönig, R. Wintjens, S. H. Soror, T. Christophe, H. K. Jeon, C. Loch, P. Brodin, B. Déprez, A. R. Baulard, N. Willand, *J. Med. Chem.* **2012**, *55*, 68-83.
- [148] J. L. Ramos, M. Martínez-Bueno, A. J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan, R. Tobes, *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 326-356.
- [149] F. Frénois, J. Engohang-Ndong, C. Loch, A. R. Baulard, V. Villeret, *Mol. Cell* **2004**, *16*, 301-307.
- [150] L. G. Dover, P. E. Corsino, I. R. Daniels, S. L. Cocklin, V. Tatituri, G. S. Besra, K. Fütterer, *J. Mol. Biol.* **2004**, *340*, 1095-1105.
- [151] H. Hernandez, C. V. Robinson, *Nat. Protocols* **2007**, *2*, 715-726.
- [152] A. Butterer, C. Pernstich, R. M. Smith, F. Sobott, M. D. Szczelkun, J. Tóth, *Nucleic Acids Res.* **2014**, *42*, 5139-5150.
- [153] A. Politis, A. Y. Park, Z. Hall, B. T. Ruotolo, C. V. Robinson, *J. Mol. Biol.* **2013**, *425*, 4790-4801.
- [154] N.-T. Nguyen-Huynh, J. Osz, C. Peluso-Iltis, N. Rochel, N. Potier, E. Leize-Wagner, *Biophys. Chem.* **2016**, *210*, 2-8.
- [155] V. Gabelica, C. Vreuls, P. Filée, V. Duval, B. Joris, E. D. Pauw, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1723-1728.
- [156] N. Willand, M. Desroses, P. Toto, B. Dirié, Z. Lens, V. Villeret, P. Rucktooa, C. Loch, A. Baulard, B. Deprez, *ACS Chemical Biology* **2010**, *5*, 1007-1013.
- [157] B. Villemagne, M. Flipo, N. Blondiaux, C. Crauste, S. Malaquin, F. Leroux, C. Piveteau, V. Villeret, P. Brodin, B. O. Villoutreix, O. Sperandio, S. H. Soror, A. Wohlkönig, R. Wintjens, B. Deprez, A. R. Baulard, N. Willand, *J. Med. Chem.* **2014**, *57*, 4876-4888.
- [158] S. Surade, N. Ty, N. Hengrung, B. Lechartier, Stewart T. Cole, C. Abell, Tom L. Blundell, *Biochem. J* **2014**, *458*, 387-394.

- [159] P. O. Nikiforov, S. Surade, M. Blaszczyk, V. Delorme, P. Brodin, A. R. Baulard, T. L. Blundell, C. Abell, *Org. Biomol. Chem.* **2016**, *14*, 2318-2326.
- [160] J. Schiebel, N. Radeva, H. Köster, A. Metz, T. Krotzky, M. Kuhnert, W. E. Diederich, A. Heine, L. Neumann, C. Atmanene, D. Roecklin, V. Vivat-Hannah, J.-P. Renaud, R. Meinecke, N. Schlinck, A. Sitte, F. Popp, M. Zeeb, G. Klebe, *ChemMedChem* **2015**, *10*, 1511-1521.
- [161] F. Rosu, S. Pirotte, E. D. Pauw, V. Gabelica, *Int. J. Mass spectrom.* **2006**, *253*, 156-171.
- [162] D. C. Mikles, V. Bhat, B. J. Schuchardt, C. B. McDonald, A. Farooq, *J. Mol. Recognit.* **2014**, *27*, 82-91.
- [163] J. García-Nafría, M. Baumgart, J. P. Turkenburg, A. J. Wilkinson, M. Bott, K. S. Wilson, *The Journal of biological chemistry* **2013**, *288*, 15800-15812.
- [164] P. L. Privalov, A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Remeta, C. A. S. A. Minetti, *J. Mol. Biol.* **2007**, *365*, 1-9.
- [165] M. A. Schumacher, M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray, R. G. Brennan, *EMBO J.* **2002**, *21*, 1210-1218.
- [166] A. R. Baulard, J. C. Betts, J. Engohang-Ndong, S. Quan, P. J. Brennan, C. Loch, G. S. Besra, *J. Biol. Chem.* **2000**.
- [167] C. Speck, C. Weigel, W. Messer, *EMBO J.* **1999**, *18*, 6169-6176.
- [168] C. Di Primo, I. Lebars, *Anal. Biochem.* **2007**, *368*, 148-155.
- [169] D. S.-H. Chan, W.-G. Seetoh, B. N. McConnell, D. Matak-Vinkovic, S. E. Thomas, V. Mendes, M. Blaszczyk, A. G. Coyne, T. L. Blundell, C. Abell, *Chem. Commun.* **2017**, *53*, 3527-3530.
- [170] C. Abad-Zapatero, J. T. Metz, *Drug Discov. Today* **2005**, *10*, 464-469.
- [171] J. Peränen, M. Rikonen, M. Hyvönen, L. Kääriäinen, *Anal. Biochem.* **1996**, *236*, 371-373.
- [172] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, G. Bricogne, *Acta Cryst. D* **2011**, *67*, 293-302.
- [173] W. Kabsch, *Acta Cryst. D* **2010**, *66*, 125-132.
- [174] P. R. Evans, *Acta Cryst. D* **2011**, *67*, 282-292.
- [175] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
- [176] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Cryst. D* **2010**, *66*, 213-221.

- [177] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Cryst. D* **2010**, *66*, 486-501.
- [178] Wessel J. A. Moolman, M. de Villiers, E. Strauss, *Biochem. Soc. Trans.* **2014**, *42*, 1080-1086.
- [179] J. Yao, J. D. Patrone, G. D. Dotson, *Biochemistry* **2009**, *48*, 2799-2806.
- [180] E. Strauss, C. Kinsland, Y. Ge, F. W. McLafferty, T. P. Begley, *J. Biol. Chem.* **2001**, *276*, 13513-13516.
- [181] T. Kupke, M. Uebele, D. Schmid, G. Jung, M. Blaesse, S. Steinbacher, *J. Biol. Chem.* **2000**, *275*, 31838-31846.
- [182] M. Blaesse, T. Kupke, R. Huber, S. Steinbacher, *The EMBO Journal* **2000**, *19*, 6299-6310.
- [183] T. Kupke, *J. Biol. Chem.* **2002**, *277*, 36137-36145.
- [184] S. Stanitzek, M. A. Augustin, R. Huber, T. Kupke, S. Steinbacher, *Structure* **2004**, *12*, 1977-1988.
- [185] N. Manoj, E. Strauss, T. P. Begley, S. E. Ealick, *Structure* **2003**, *11*, 927-936.
- [186] M. Daugherty, B. Polanuyer, M. Farrell, M. Scholle, A. Lykidis, V. de Crécy-Lagard, A. Osterman, *J. Biol. Chem.* **2002**, *277*, 21431-21439.
- [187] J. C. Evans, C. Trujillo, Z. Wang, H. Eoh, S. Ehrt, D. Schnappinger, H. I. M. Boshoff, K. Y. Rhee, C. E. Barry, V. Mizrahi, *ACS Infect. Dis.* **2016**, *2*, 958-968.
- [188] J. D. Patrone, J. Yao, N. E. Scott, G. D. Dotson, *J. Am. Chem. Soc.* **2009**, *131*, 16340-16341.
- [189] R. van der Westhuyzen, J. C. Hammons, J. L. Meier, S. Dahesh, W. J. Moolman, S. C. Pelly, V. Nizet, M. D. Burkart, E. Strauss, *Chem. Biol.* **2012**, *19*, 559-571.
- [190] E. Strauss, T. P. Begley, *J. Am. Chem. Soc.* **2001**, *123*, 6449-6450.
- [191] A. F. M. Gavriilidou, B. Gülbakan, R. Zenobi, *Anal. Chem.* **2015**, *87*, 10378-10384.
- [192] W.-G. Seetoh, D. S.-H. Chan, D. Matak-Vinković, C. Abell, *ACS Chem. Biol.* **2016**, *11*, 1511-1517.
- [193] E. Boeri Erba, K. Barylyuk, Y. Yang, R. Zenobi, *Anal. Chem.* **2011**, *83*, 9251-9259.
- [194] V. Gabelica, N. Galic, F. Rosu, C. Houssier, E. De Pauw, *J. Mass Spectrom.* **2003**, *38*, 491-501.
- [195] D. S.-H. Chan, D. Matak-Vinković, A. G. Coyne, C. Abell, *ChemistrySelect* **2016**, *1*, 5686-5690.
- [196] J. N. Rabuck, S.-J. Hyung, K. S. Ko, C. C. Fox, M. B. Soellner, B. T. Ruotolo, *Anal. Chem.* **2013**, *85*, 6995-7002.
- [197] K. R. Abdul Azeez, S. Chatterjee, C. Yu, T. R. Golub, F. Sobott, J. M. Elkins, *Nature Communications* **2019**, *10*, 3166.
- [198] T. Daubenfeld, A.-P. Bouin, G. van der Rest, *J. Am. Soc. Mass. Spectrom.* **2006**, *17*, 1239-1248.

- [199] A. Dyachenko, R. Gruber, L. Shimon, A. Horovitz, M. Sharon, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 7235-7239.
- [200] W. Wang, E. N. Kitova, J. S. Klassen, *Anal. Chem.* **2005**, *77*, 3060-3071.
- [201] S. Guan, M. J. Trnka, D. A. Bushnell, P. J. J. Robinson, J. E. Gestwicki, A. L. Burlingame, *Anal. Chem.* **2015**, *87*, 8541-8546.
- [202] L. Shimon, M. Sharon, A. Horovitz, *Biophys. J.* **2010**, *99*, 1645-1649.
- [203] L. Tautz, J. Rétey, *Eur. J. Org. Chem.* **2010**, *2010*, 1728-1735.
- [204] W. Hu, J. Xie, H. W. Chau, B. C. Si, *Environ. Syst. Res.* **2015**, *4*, 4.
- [205] M. R. Webb, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4884-4887.
- [206] M. Yoshihiko, N. Hitoshi, N. Len, *Curr. Pharm. Des.* **2013**, *19*, 347-365.
- [207] I. A. Asteriti, W. M. Rensen, C. Lindon, P. Lavia, G. Guarguaglini, *Biochim. Biophys. Acta* **2010**, *1806*, 230-239.
- [208] R. Bayliss, T. Sardon, I. Vernos, E. Conti, *Mol. Cell* **2003**, *12*, 851-862.
- [209] J. W. Friedberg, D. Mahadevan, E. Cebula, D. Persky, I. Lossos, A. B. Agarwal, J. Jung, R. Burack, X. Zhou, E. J. Leonard, H. Fingert, H. Danaee, S. H. Bernstein, *J. Clin. Oncol.* **2014**, *32*, 44-50.
- [210] S. Emanuel, C. A. Rugg, R. H. Gruninger, R. Lin, A. Fuentes-Pesquera, P. J. Connolly, S. K. Wetter, B. Hollister, W. W. Kruger, C. Napier, L. Jolliffe, S. A. Middleton, *Cancer Res.* **2005**, *65*, 9038-9046.
- [211] B. Jayaram, T. Jain, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 343-361.
- [212] I. A. Asteriti, F. Daidone, G. Colotti, S. Rinaldo, P. Lavia, G. Guarguaglini, A. Paiardini, *Oncotarget* **2017**, *8*, 32117-32133.
- [213] K. Hoar, A. Chakravarty, C. Rabino, D. Wysong, D. Bowman, N. Roy, J. A. Ecsedy, *Mol. Cell. Biol.* **2007**, *27*, 4513-4525.
- [214] D. H. Williams, C. T. Calderone, D. P. O'Brien, R. Zerella, *Chem. Commun.* **2002**, 1266-1267.
- [215] C. Structural Genomics, S. Gräslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H.-w. Park, A. Savchenko, A. Yee, A. Edwards, B. Architecture et Fonction des Macromolécules, R. Vincentelli, C. Cambillau, C. Berkeley Structural Genomics, R. Kim, S.-H. Kim, C. China Structural Genomics, Z. Rao, Y. Shi, S. Integrated Center for, I. Function, T. C. Terwilliger, C.-Y. Kim, L.-W. Hung, G. S. Waldo, C. Israel Structural Proteomics, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, G. Joint Center for Structural, R. C. Stevens, S. A. Lesley, I. A. Wilson, G. Midwest Center for Structural, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, X. R. C. f. S. G. New York Structural Genomi, S. K. Burley, J. S. Emtage, J. M. Sauder, D.

- Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, C. Northeast Structural Genomics, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, F. Oxford Protein Production, D. I. Stuart, R. J. Owens, S. Daenke, M. D. C. f. M. M. Protein Sample Production Facility, A. Schütz, U. Heinemann, R. S. G. P. Initiative, S. Yokoyama, S. Complexes, K. Büssow, K. C. Gunsalus, *Nat. Methods* **2008**, *5*, 135.
- [216] P. T. Wingfield, *Curr. Protoc. Protein Sci.* **2017**, *88*, 6.14.11-16.14.13.
- [217] R. A. Nicholas, W. C. Watt, E. R. Lazarowski, Q. Li, K. Harden, *Mol. Pharmacol.* **1996**, *50*, 224-229.
- [218] D. Cubrilovic, K. Barylyuk, D. Hofmann, M. J. Walczak, M. Graber, T. Berg, G. Wider, R. Zenobi, *Chem. Sci.* **2014**, *5*, 2794-2803.
- [219] K. J. DiRico, W. Hua, C. Liu, J. W. Tucker, A. S. Ratnayake, M. E. Flanagan, M. D. Troutman, M. C. Noe, H. Zhang, *ACS Medicinal Chemistry Letters* **2020**, *11*, 1101-1110.
- [220] I. Sinclair, G. Davies, H. Semple, *Expert opinion on drug discovery* **2019**, *14*, 609-617.
- [221] N. Blondiaux, M. Moune, M. Desroses, R. Frita, M. Flipo, V. Mathys, K. Soetaert, M. Kiass, V. Delorme, K. Djaout, V. Trebosc, C. Kemmer, R. Wintjens, A. Wohlkönig, R. Antoine, L. Huot, D. Hot, M. Coscolla, J. Feldmann, S. Gagneux, C. Locht, P. Brodin, M. Gitzinger, B. Déprez, N. Willand, A. R. Baulard, *Science* **2017**, *355*, 1206-1211.
- [222] V. Mendes, S. R. Green, J. C. Evans, J. Hess, M. Blaszczyk, C. Spry, O. Bryant, J. Cory-Wright, D. S. H. Chan, P. H. M. Torres, Z. Wang, S. O'Neill, S. Damerow, J. Post, T. Bayliss, S. L. Lynch, A. G. Coyne, P. C. Ray, C. Abell, K. Y. Rhee, H. I. M. Boshoff, C. E. Barry, V. Mizrahi, P. G. Wyatt, T. L. Blundell, *bioRxiv* **2019**, 870154.

List of PhD publications

1. **D. S.-H. Chan (first author)**, J. Hess, E. Shaw, C. Spry, R. Starley, C. Dagostin, M. V. B. Dias, R. Kale, V. Mendes, T. L. Blundell, A. G. Coyne, C. Abell. Structural insights into *Escherichia coli* phosphopantothienoylcysteine synthetase by native ion mobility–mass spectrometry. *Biochem. J.* **2019**, *476*, 3125.
2. **D. S.-H. Chan (first author)**, M. E. Kavanagh, K. McLean, A. Munro, D. M.-V. Vinković, A. G. Coyne, C. Abell. Effect of DMSO on protein structure and interactions assessed by collision-induced Dissociation and Unfolding. *Anal. Chem.*, **2017**, *89*, 9976.
3. **D. S.-H. Chan (first author)**, V. Mendes, S. E. Thomas, B. N. McConnell, D. M.-V. Vinković, A. G. Coyne, T. L. Blundell, C. Abell. Fragment screening against the EthR-DNA interaction by native mass spectrometry. *Angew. Chem. Int. Ed.*, **2017**, *56*, 7488 (**chosen as a “Hot Paper” by the Editors for its importance in a rapidly evolving field of high current interest; rated as “Highly/Very Important” by all three referees; highlighted by ChemViews Magazine, chemistryviews.org/details/ezone/10560326/.html; highlighted by Practical Fragments, <http://practicalfragments.blogspot.co.uk/2017/07/native-mass-spectrometry-revisited.html>).**
4. **D. S.-H. Chan (first author)**, W.-G. Seetoh, B. N. McConnell, D. M.-V. Vinković, S. E. Thomas, V. Mendes, M. Blaszczyk, A. G. Coyne, T. L. Blundell, C. Abell. Structural insights into the EthR–DNA interaction using native mass spectrometry. *Chem. Commun.*, **2017**, *53*, 3527.
5. **D. S.-H. Chan (first author)**, A. J. Whitehouse, A. G. Coyne, C. Abell. Mass spectrometry for fragment screening. *Essays Biochem.*, **2017**, *61*, 465.
6. **D. S.-H. Chan (first author)**, D. M.-V. Vinković, A. G. Coyne, C. Abell. Insight into protein conformation and subcharging by DMSO from native ion mobility mass spectrometry. *ChemistrySelect*, **2016**, *1*, 5686.
7. A. Trapero, A. Pacitto, **D. S.-H. Chan**, C. Abell, T. L. Blundell, D. B. Ascher, A. G. Coyne. Covalent inactivation of *Mycobacterium thermoresistibile* inosine-5′-monophosphate dehydrogenase (IMPDH). *Bioorg. Med. Chem. Lett.* **2020**, *30*, 126792.
8. A. J. Whitehouse, S. E. Thomas, K. P. Brown, A. Faourakis, **D. S.-H. Chan**, M. D. J. Libardo, V. Mendes, H. I. M. Boshoff, R. A. Floto, C. Abell, T. L. Blundell, A. G. Coyne. Development of Inhibitors against *Mycobacterium abscessus* tRNA (m¹G37) Methyltransferase (TrmD) Using Fragment-Based Approaches. *J. Med. Chem.* **2019**, *62*, 15, 7210.
9. C. Marchetti, **D. S.-H. Chan**, A. G. Coyne, C. Abell. Fragment-based approaches to TB drugs. *J. Parasitol.*, **2018**, *145*, 184.

10. W.-G. Seetoh, **D. S.-H. Chan**, D. M.-V. Vinković, C. Abell. Mass spectrometry reveals protein kinase CK2 high-order oligomerization via the circular and linear assembly. *ACS Chem. Biol.*, **2016**, *11*, 1511.
11. M. E. Kavanagh, A. G. Coyne, K. J. McLean, G. G. James, C. W. Levy, L. B. Marino, L. P. S. de Carvalho, **D. S.-H. Chan**, S. A. Hudson, S. Surade, D. Leys, A. W. Munro, C. Abell. Fragment-based approaches to the development of *Mycobacterium tuberculosis* CYP121 inhibitors. *J. Med. Chem.*, **2016**, *59*, 3272.