Characterizing the Structure-Function Relationship of Pathological Proteins

Xiaojing Huang

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CHEMISTRY

YORK UNIVERSITY

TORONTO, ONTARIO

April 2019

Abstract

Protein structure and dynamics dictate the fundamentals of protein function. For example, a structured protein must change its conformation when interacting with various ligands. In addition, misfolded or intrinsically disordered proteins (IDPs) oligomerize into more structured aggregates or amyloid fibrils. Therefore, the study of protein conformations becomes necessary to further understand their cellular functions. By monitoring the exchange of backbone amide hydrogens as a measurement of protein dynamics, time-resolved hydrogen-deuterium exchange (TR HDX) provides advantages in capturing the subtle and transient structural differences over other equilibrium state techniques, such as X-ray crystallography and NMR. Using TR HDX together with other biophysical techniques, we demonstrate the unfolding of lipocalin 2 (Lcn2) in accommodating siderophores. In the second part, we attempt to understand the change in aggregation rate corresponding to conformational changes of polyglutamine expansion protein, huntingtin exon1 (Httex1).

Acknowledgements

First, I thank my supervisor, Derek Wilson, for his support and guidance. I would also extend my appreciation to my committee members, Mark Bayfield, Demian Ifa, and Yi Sheng, for their critiques and advice. I would like to thank Gary Sweeney for all helpful discussions as well. I also appreciate the advice and input from Philip Johnson, Sladjana Slavkovic, and Erfei Song.

My deepest gratitude goes toward my fellow lab mates. I would like to give my thanks to Cristina Lento for always listening, to Irina Oganesyan for suffering in the lab with me, to Lucienne Nouchikian for being my rock climbing and aerial silk partner, to Kerene Brown for teaching me everything when I knew nothing, to Shaolong Zhu for showing me proper LC, to Lisa Szymkowicz for helping me survive the first week at Sanofi Pasteur, to Shenbaga Moorthy for having all the answers about HDX, and to Peter Luini for being the godfather of the lab. Naturally, I must thank my parents, Jian Huang and Ping Han, for expecting nothing special of me so that everything I achieved is exceptional. I appreciate Youcheng Huang for being such an annoying little brother that always looks up to me, so I must be the better sibling. My most special thanks to the horrible Toronto weather, for being my endless motivation to finish this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN -- acetonitrile

AMP -- ampicillin

ATD -- arrival time distribution

CCS -- collisional cross section

CE -- collisional energy

CIX -- cation exchange

CID -- collisional induced unfolding

CIU -- collisional induced dissociation

CBP -- CREB binding protein

CREB -- cAMP response element-binding protien

CV -- column volume

CytC -- cytochrome C

DHBA -- dihydrobenzoic acid

DMSO -- dimethyl sulfoxide

E. coli -- *Escherichia coli*

EDTA -- ethylenediaminetetraacetic acid

EK -- enterokinase

Ent -- enterobactin

ESI -- electrospray ionization

FA -- formic acid

F/T -- flow-through

HDX -- hydrogen-deuterium exchange

Httex1 -- huntingtin exon 1

IDP -- intrinsic disordered protein

IDR -- intrinsic disordered region

ITC -- isothermal titration calorimetry

IPTG -- isopropyl-B-D-thiogalactopyranoside

HD -- Huntington's disease

HEAT -- **h**untingtin, **e**longation factor 3, protein phosphatase 2**A**, and yeast kinase **T**OR1

HD16Q -- huntingtin exon 1 with 16 glutamine

HD25Q -- huntingtin exon 1 with 25 glutamine

HD39Q -- huntingtin exon 1 with 39 glutamine

HD46Q -- huntingtin exon 1 with 46 glutamine

LB -- Luria Bertani

Lcn2 -- lipocalin 2, siderocalin

MES -- 2-(N-morpholino)ethanesulfonic acid

MW -- molecular weight

m/z -- mass to charge ratio

Ni-IMAC -- nickel immobilized metal ion affinity chromatography

NMR -- nuclear magnetic resonance

RT -- room temperature

OD_{280} -- optical density at 280nm wavelength

O/N -- overnight

pI -- isoelectric point

PMMA -- poly-methyl methacrylate

PolyQ -- polyglutamine

PRR -- proline-rich region

SDS-PAGE -- sodium dodecyl sulphate polyacrylamide gel

S/N -- supernatant

SOC -- super optimal broth with catabolite repression

TR-HDX -- time-resolved hydrogen-deuterium exchange

TWIMS -- travelling wave ion mobility separations

LIST OF PUBLICATIONS

1. Song, E. *et al.* Holo-lipocalin-2-derived siderophores increase mitochondrial ROS and impair oxidative phosphorylation in rat cardiomyocytes. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1576–1581 (2018).

Chapter One: Introduction

1.1 Hydrogen/Deuterium exchange mass spectrometry (HDX MS), a tool for studying protein structural dynamics

1.1.1 Hydrogen/Deuterium exchange (HDX)

HDX is an isotopic exchange reaction in which heavier deuterium replaces labile hydrogens. Within a protein, the most common labile hydrogens are located on the backbone amide, however side chain S-H and O-H also participate in this process.[1] To understand the correlation between the rate of deuterium uptake and the structural dynamics of a protein, the contributing factors of the HDX rate are critical.[2]

The primary sequence effects the reaction rate at the individual amino acid level. For example, prolines are less meaningful in an HDX reaction since they do not contain labile hydrogens. Additionally, a bulky side chain hinders HDX of its neighbouring amino acids.[1] The secondary structure and hydrogen bond network alter the isotopic exchange rate as well. A solvent accessible surface will generally have more deuterium displacement, while a rigid hydrogen bond network reduces the exchange event.[3,4]

In addition to these, pH and temperature are the other main factors effecting HDX rates.[5] Calculated from the average activation energy (17kJ) of HDX, deuterium uptake is faster at higher temperatures following a 3-fold-increase per 10˚C rule at room temperature.[6] Furthermore, because of the nature of the catalytic mechanism, HDX is heavily pH-dependent.[7] Although the HDX reaction can be acid or base catalyzed, base catalysis is dominant at pH 7.[8] Initiated by deprotonation at the peptide bond amide, deuterium displaces hydrogen via an imine intermediate (Figure 1.1 a). Acid catalysis has two possible mechanisms, N-protonation or O-protonation. Tautomerization rearranges

the deuterium to the amide position in an O-protonation situation (Figure 1.1 b, c). The water-mediated HDX is slow and usually ignored in the analysis.

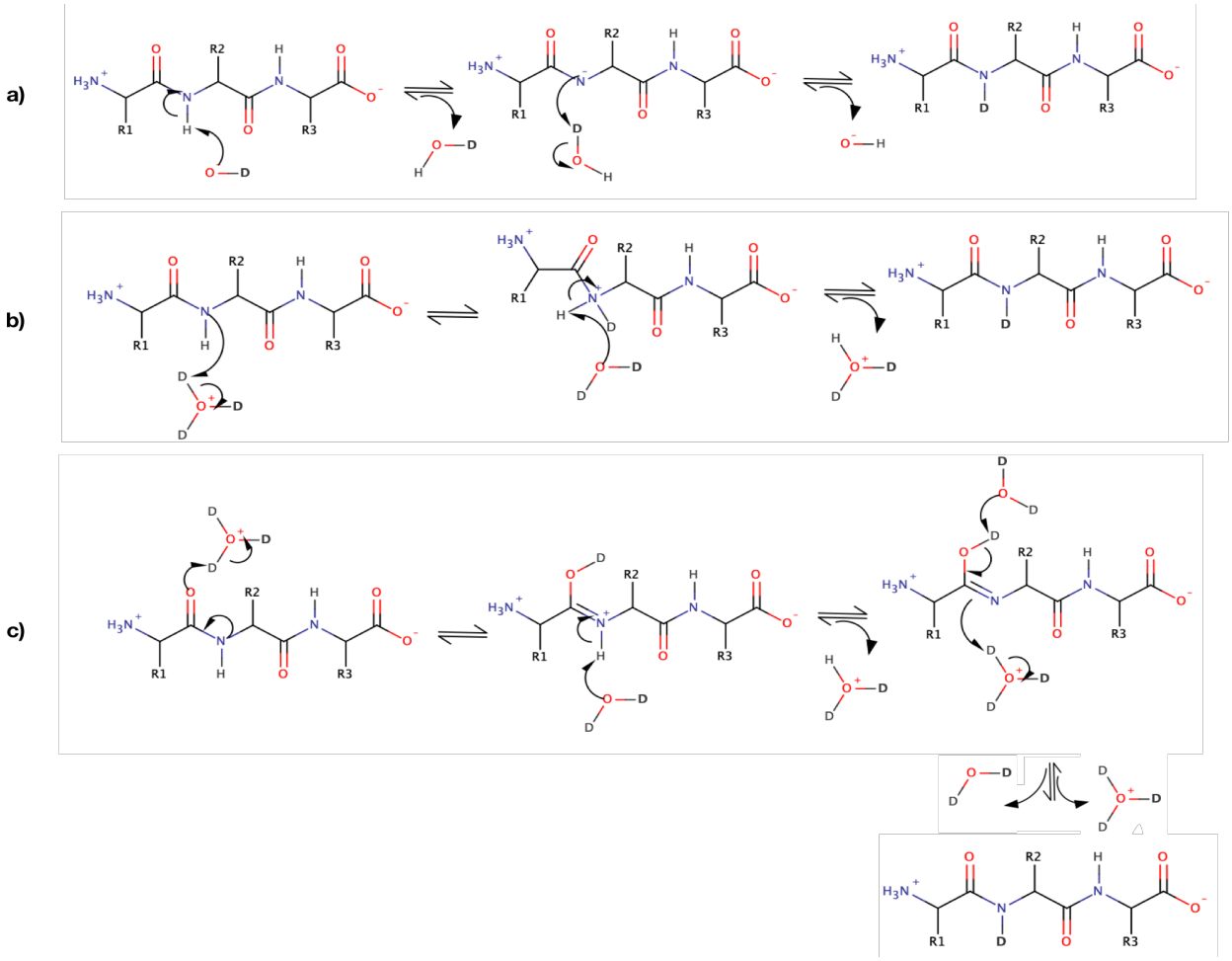


Figure 1.1. The mechanisms of acid and base catalyzed HDX reaction.

The electron transfer of backbone amide HDX via a) base catalysis, b) N-protonated acid catalysis, and c) O-protonated acid catalysis.

Specific HDX rates for single amino acids are measured using unstructured peptides. The results demonstrate that the slowest exchange rate of backbone amides occur at pH 2.5 while other side chain hydrogens have much higher rates.[9] Therefore, HDX reactions are quenched at pH 2.5 to prevent further deuteration at backbone amides while all side

chains back-exchange into hydrogens. The quenching makes downstream application possible by controlling specific deuterium uptake time.

HDX has long been used as a bioanalytical tool for probing protein dynamics with deuterium uptake levels.[5] A two-step mechanism of the HDX reaction explains the correlation (Equation 1):

$$\mathbf{NH_{close}} \rightleftharpoons \mathbf{NH_{open}} \longrightarrow \mathbf{ND_{open}} \rightleftharpoons \mathbf{ND_{close}} \quad (1)$$

Proteins naturally exhibit a folding and unfolding equilibrium, however, HDX can only occur during the open state of a protein. Therefore, a fast first step makes the exchange reaction the rate-determining step. Such kinetics, referred to as EX2, is the most commonly observed.[10] In EX2 kinetics, the observed HDX rate is the measurement of dynamics in the hydrogen bonded network. Nonetheless, protein unfolding ranges from milliseconds to minutes to a degree which protein unfolding may be the slowest step in the mechanism. In the latter case, referred to as EX1 kinetics, the observed HDX rate is limited by the open-close equilibrium of a protein.[11] By comparing multiple protein states at the same pH, temperature and buffer conditions, the change of protein conformational dynamics can be described by HDX rates.

1.1.2 Electrospray Ionization mass spectrometry (ESI MS) coupled HDX

HDX is historically coupled with NMR, where single amino acid HDX rates can be accurately measured.[12] However, in the protein realm, NMR poses challenges in the large amount of material required, size restriction and long scanning times lasting more than 10 hours. Because of the mass increase of isotopic reactions, the amount of deuterium incorporation is easily detected by mass spectrometry (MS). Development of the

electrospray ionization (ESI) technique enables MS coupled HDX.[5] ESI is a soft-ionization technique that applies a high voltage to the analyte in a metal capillary, generating multiply charged ions.[13] Although single amino acid resolution is difficult to obtain, HDX MS is a more protein-friendly method for monitoring HDX rates. HDX-ESI-MS has been widely adopted in the bioanalytical field as a rapid method to study protein-ligand interactions.[14–16]

1.1.3 Conventional HDX

In a typical HDX experiment, a protein sample is labelled with deuterium oxide at pH 7 for various time points. The reaction is then quenched at 0˚C and pH 2.5. In addition to the low pH, the low temperature of the quenching condition further prevents back-exchange at backbone amides.[10] Depending on the application, the downstream processes vary. Intact protein can be detected directly by ESI MS for global deuterium level analysis. The results are plotted as HDX time versus mass increases. In a global HDX set-up, the dynamics at different regions of a protein are averaged so that no site-specific information is provided (Figure 1.2 top).

Instead, the quenched protein can be digested into peptides before introduction into the mass spectrometer (Figure 1.2 bottom). The peptides are usually separated based on hydrophobicity on a C18 column by liquid chromatography (LC). Since only a fraction of the peptides are sent into the mass spectrometer at a specific retention time, LC can effectively reduce peptide overlap. This bottom-up HDX workflow provides information localizing dynamics to different areas of a protein. The representation of local HDX data can vary, but they generally focus on the difference of deuterium level between various protein states and use 3D structures with a heat map to aid the eye.

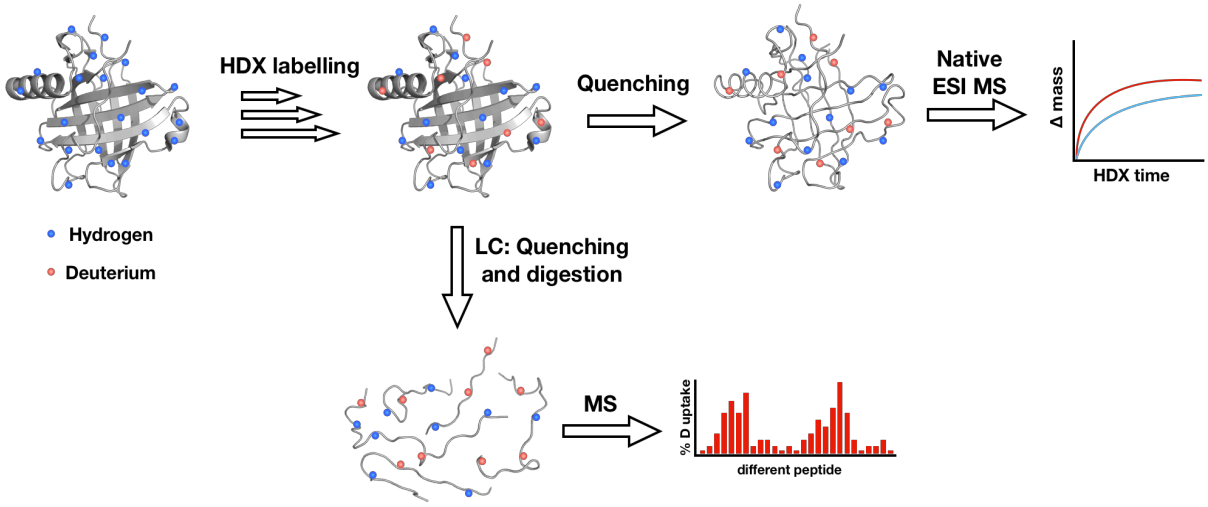


Figure 1.2. Schematic representation of typical HDX workflows.

After different HDX labelling times and quenching, detection by ESI MS can be performed on intact protein (top) or peptides (bottom).

1.1.4 Time-resolved HDX

The conventional bottom-up workflow is the most widely employed HDX MS technique. However, the deuterium exchange time is usually limited to 10 seconds or longer. To study subtle or transient changes in protein conformational dynamics, time-resolved (TR) HDX with millisecond to second deuterium labelling time scales is more desirable. The first TR HDX MS experiment was performed on cytochrome C in 1996. The short reaction time was achieved by a quench-flow system, where the deuterium contact time is determined by the diameter of the mixing tube and flow rates.[17] A more elaborate continuous-flow set-up was constructed in the Konermann lab to study partially folded myoglobin. In this set-up, the isotopic reaction takes places in a fused-silica capillary, where the reaction time is controlled by the volume of the mixing capillary and flow rates.[18]

The Wilson lab further modified TR HDX as an online labelling and digestion system on a microfluidic device.[19] The critical component, a capillary based kinetic mixer,

replaces the mixing capillary in controlling the exchange time (Figure 1.3). The inner capillary of the mixer is sealed at the end. A notch, 2 mm away from the end of the inner capillary allows protein solution to flow into the outer capillary. The protein solution contacts and exchanges with deuterium in the inter-capillary space following the notch. Because the HDX time is controlled directly by the volume after the notch, the time of the analytes staying in this region determines the reaction time. This can be modified by tuning the flow rates and the inner capillary position. After labelling, deuterated protein is quenched at pH 2.5 by excess acid. Although the acid quenching happens at room temperature, the relatively short time between acid contact and detection is unlikely to introduce significant back-exchange. TR HDX can provide global or peptide level information depending on whether a protease chamber is connected. Proteases are immobilized to agarose beads so that they stay in the chamber.

Although TR HDX is an excellent tool for probing weak changes in a protein structure at short time scales, the lack of liquid handling from LC poses challenges in digestion efficiency and limitations on the buffer choices. In terms of obtaining spatial resolution of peptides, ion mobility separation (IMS) fractions peptides based on size and charge instead of C18 and hydrophobicity.[20] More detailed experimental parameters are described in section 2.3.7.

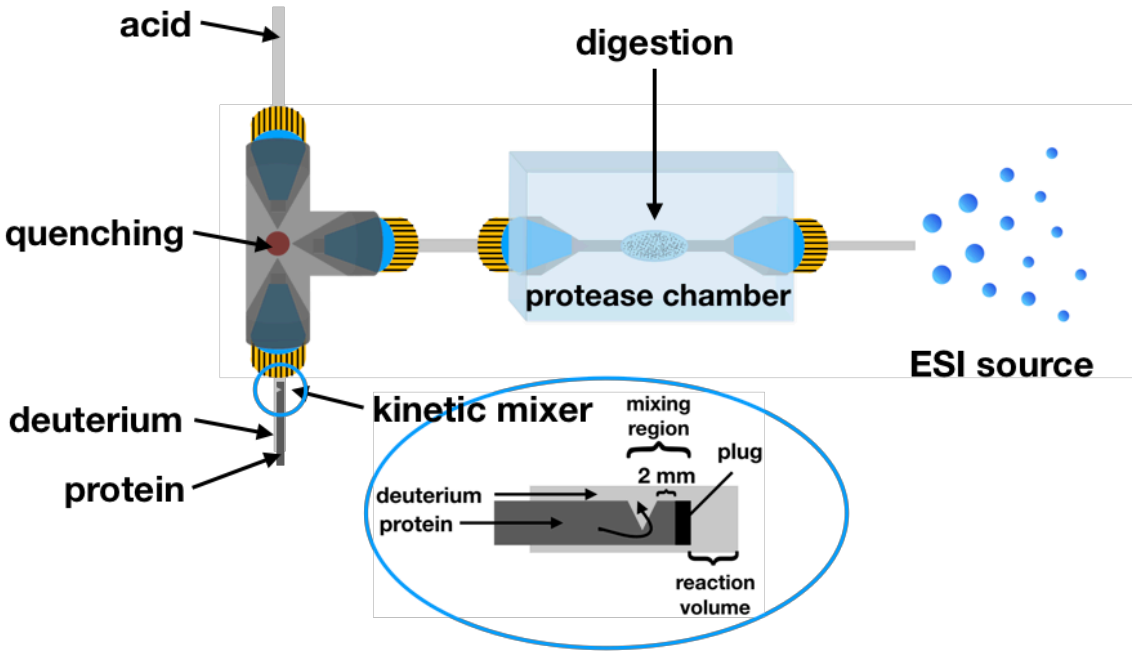


Figure 1.3. Schematic overview of TR HDX coupled to ESI MS.

The protein flows in the inner capillary of the kinetic mixer, and only begins contact with deuterium after the notch. The reaction volume can be altered by changing flow rates and inner capillary position. The protease chamber is optional for bottom-up peptide level analysis only.

1.2 Protein-ligand interactions

As the end products of most gene expressions, proteins execute diverse functions via physical interaction with other molecules, either proteins or small molecules. In this work, the term ligands refers to binding partners which form a noncovalent but specific complex with their target protein. In any protein-ligand system, the forming and dissociating of the complex exists as an equilibrium.[21] Equation (2) describes the simplest expression of the complex formation of one protein (P) and one ligand (L). The dissociation constant (K_d) derived from the equilibrium state is usually used to express the affinity because a low K_d (high affinity) requires either a high K_{on} or a low K_{off} (Equation 3).[22] However, instead of a

low K_d, a decrease in the system's free energy is the true driving force. Based on the second law of thermodynamics, enthalpy (H) and entropy (S) collectively decides the spontaneity of the reaction (Equation 4). The enthalpy drop is mainly contributed by the formation of noncovalent forces, including charge-charge, dipole-dipole, charge-dipole interactions, hydrogen bonds, and hydrophobic effects.[23] Since the interaction happens in a solution, both the water and solvent molecules contribute to the interaction as well. Therefore, the system entropy changes are determined by hydration network disturbance at the binding surface (ΔS_{solv}), protein conformational rigidity (ΔS_{conf}) and ligand rotational translational freedom ($\Delta S_{r/t}$) (Equation 5).[22]

$$\mathbf{P + L \underset{K_{off}}{\overset{K_{on}}{\rightleftharpoons}} PL} \tag{2}$$

$$\mathbf{K_a = \frac{1}{K_d} = \frac{[PL]}{[P][L]} = \frac{K_{on}}{K_{off}}} \tag{3}$$

$$\mathbf{\Delta G = \Delta H - T\Delta S} \tag{4}$$

$$\mathbf{\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t}} \tag{5}$$

Theories explaining protein-ligand interactions have evolved over a century following the first lock-and-key model. Together with two other popular models, induced-fit and conformational selection, the true mechanism governing this interaction is still under debate. The original lock-and-key model assumes a rigid protein structure so that only a ligand with the correct shape will fit in.[24] This theory has been gradually disproven because many proteins change their shapes in order to accommodate their ligands.[25] The induced-fit model hypothesizes that the initial engagement of the ligand is strong enough to change the protein into a favourable conformation to form the complex.[26] In this model, the protein only has one fixed structure before the ligand changes it into a more desirable

conformation. Finally, originating from the protein folding energy landscape theory, conformational selection believes that the protein has many pre-existing conformers at individual local energy minima (Figure 1.4b).[27,28] The association of ligands stabilizes one of these conformations and shifts the population towards a single dominant species. Most recently, induced-fit has been considered as an extreme case of conformational selection where a smooth energy funnel produces only a single dominant conformer (Figure 1.4a).[28]

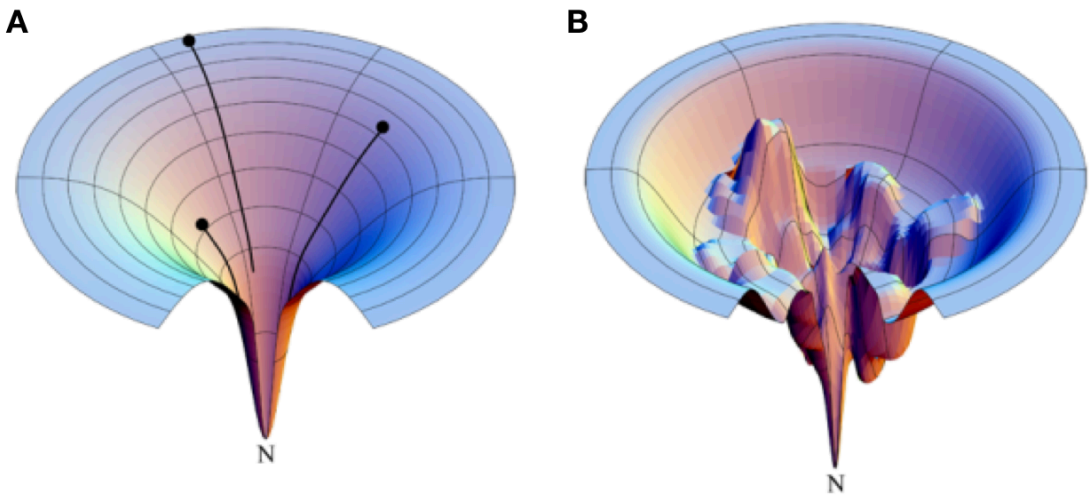


Figure 1.4. The free energy funnel of protein folding.

The free energy funnels show how unfolded proteins at the top of the funnel fold into their native states (N) at a lower energy. (A) The smooth funnel leads to a single native state, while (B) the landscaped funnel has multiple energy minima near the native state. This figure is adapted from Liu et al.[27]

1.3 Protein folding and aggregation

Other than the physical interaction itself, a correct protein structure is also critical for protein function.[29,30] Protein sequences are selected by evolution to favour correct folding

into a low energy "native state".[31,32] Immediately after proteins are synthesized by ribosomes, they start to fold with the help of isomerases and chaperones. Isomerases facilitate the cis/trans conversion of the peptide bond at prolines and disulphide bond formation at cysteines.[33,34] At the same time, chaperones work in a chain to provide an optimal folding environment and help to sequester the hydrophobic regions.[35]

Although the majority of proteins follow the energy down the folding funnel into its native state, it can get caught in other local minima or even in a misfolded state (Figure 1.5). Under certain circumstances, the misfolded protein can unfold and refold with the aid of chaperones.[36,37] Failure in reaching the native state despite the help from chaperones, misfolded proteins are subject to ubiquitination and proteasome degradation.[38] However, failure in regulation can cause misfolded proteins to fall into the aggregation funnel (Figure 1.5).[39]

Alternatively, intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs) function at the edge of the aggregation funnel due to the absence of stable native structures.[40] In these cases, the exposed hydrophobic regions are aggregation-prone (Figure 1.6). Despite the danger of falling into the aggregation funnel, normal functions of IDPs are essential.[41,42] Additionally, proteins with IDRs play essential roles in cellular interactions, such as DNA recognition by the basic leucine-zipper transcriptional factors.[43,44]

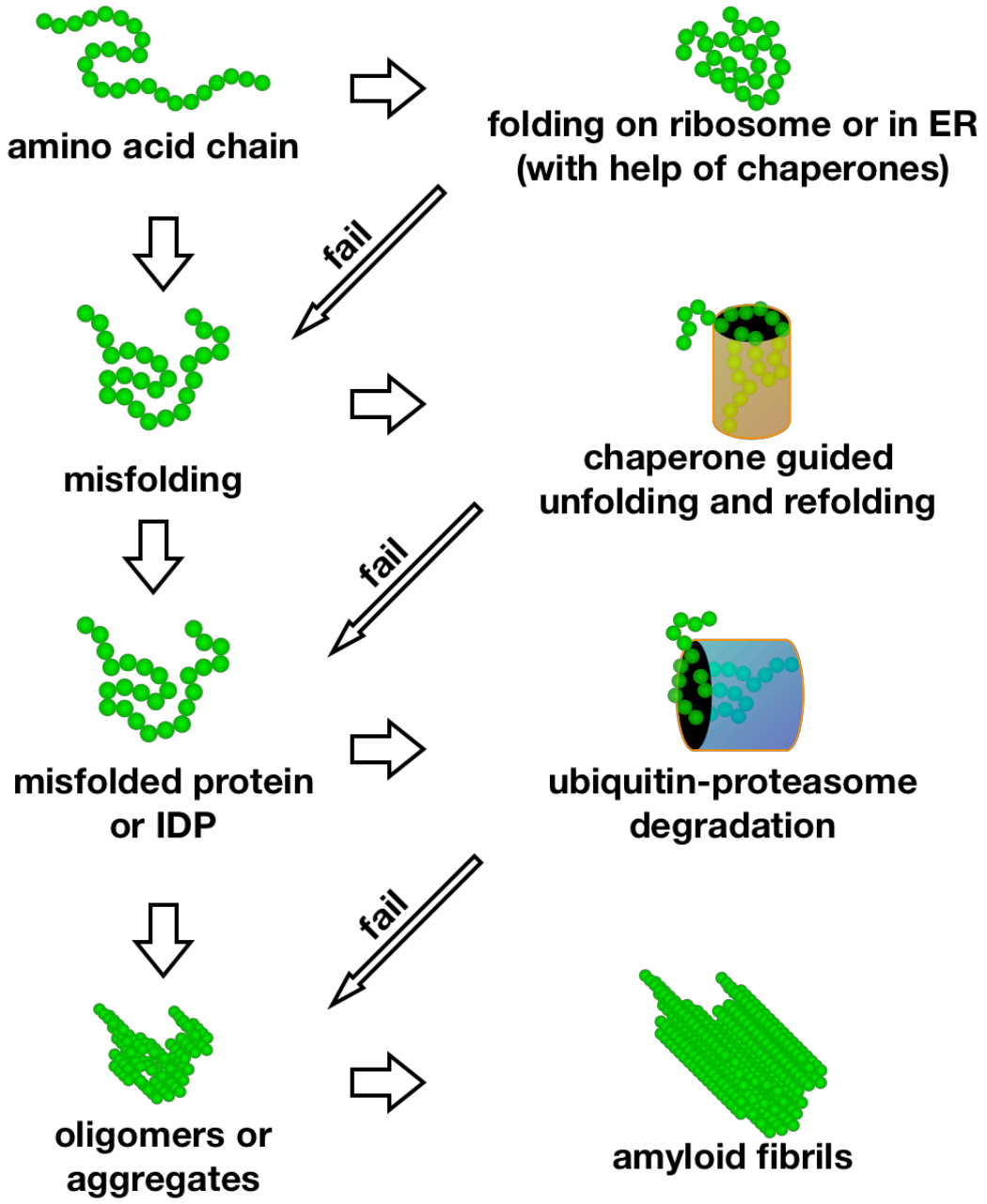


Figure 1.5. Protein folding, misfolding and aggregation.

Most proteins fold into their native states. Occasionally generated misfolded proteins are regulated by chaperone refolding or ubiquitin-proteasome degradation. The dysregulation of aberrant folding leads to oligomerization and aggregation. Intrinsically disordered proteins (IDPs), at a similar energy regime as misfolded proteins, head to a fibrilization pathway when controlled degradation fails.

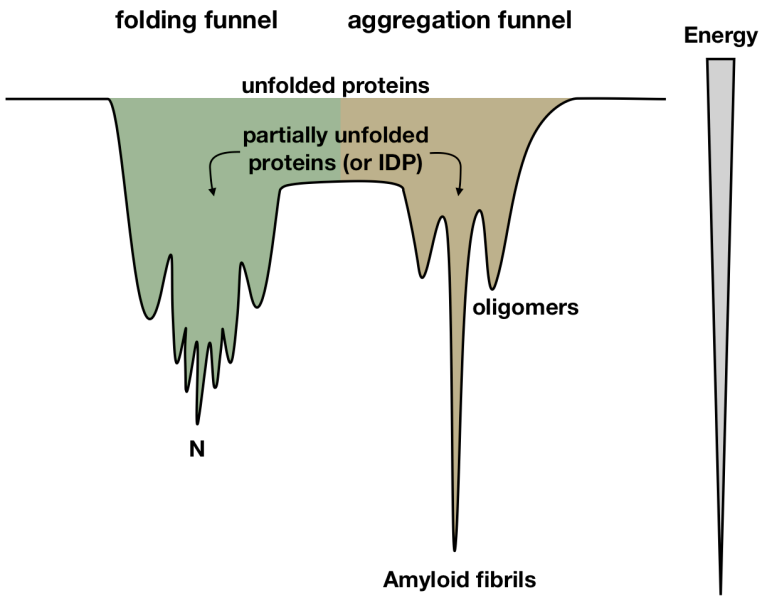


Figure 1.6. The interconnection between folding and aggregation funnels of a partially unfolded protein or IDP.

Proteins partially unfold to carry out part of their function. However, when hydrophobic sequences are exposed, IDPs and proteins with IDRs are prone to aggregate. The folding funnel has a rough landscape and multiple low energy native states (left), while the aggregation funnel has a single metastable fibril state (right).

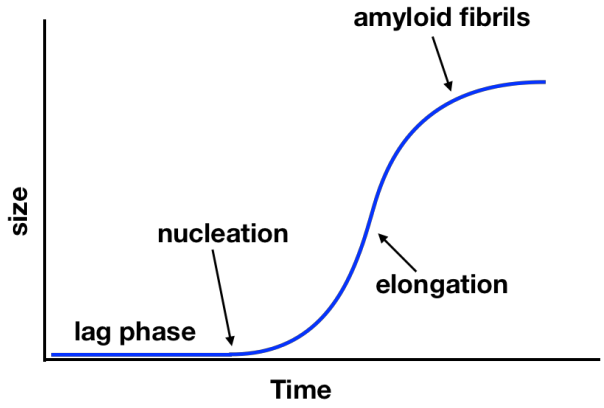


Figure 1.7. A typical aggregation pathway for amyloidogenic proteins.

Before critical size nucleation occurs, amyloidogenic proteins have a long lag phase. The elongation phase involving the addition of monomers is rapid. Once fibrils are formed, the extremely stable structure prevents any spontaneous dissociation.

In the fibril forming pathway of many amyloidogenic IDPs, a lag phase could last months before oligomers exceed a critical size.[45] The accumulation of seeding nuclei reaching critical size in the lag phase promotes a steep increase in the aggregate sizes in the elongation phase, where monomers are added to the oligomer nucleus. Eventually, extremely stable amyloid fibrils are formed (Figure 1.7).[46] Due to the seeding effect at the beginning of the elongation phase, amyloidogenic proteins can act as prions to exaggerate other protein oligomerizations. When IDP oligomers above a critical size get transported to other cells, they may accelerate the aggregation of other proteins at the edge of the aggregation funnel.[47]

1.4 Research Objective

Protein dynamics and structure are critical to carrying out specific functions in the cell. Layering a detailed picture of conformational dynamics over known protein structures can therefore provide unprecedented insights into the molecular origins of protein function. In Chapter 2, our aim is to better understand the unusual interaction between Icn2 and its siderophore targets, in the context of bacterial infection and sterile inflammation. In chapter 3, we attempt to study the change in aggregation properties from normal to mutant transition in huntingtin exon 1 (Httex1).

Chapter Two: Change in the conformational dynamics of lipocalin 2 (Lcn2) upon siderophore binding

This chapter is currently being prepared as a manuscript.

2.1 Summary

The protein-ligand interaction is modelled by lipocalin 2 (Lcn2) and siderophores. The small molecules, siderophores, are produced mainly by bacteria in need of scavenging finite amounts of iron from the environment. Bacteria in the gastrointestinal tract use enterobactin (Ent) for iron acquisition, while Lcn2 sequesters Ent as a host regulation strategy on gut microbiota. Other Lcn2 functions have been actively explored in inflammatory conditions, including its manipulation of reactive oxygen species (ROS) accumulation and cytotoxicity. The diverse functions of Lcn2 involve the interaction or the absence of siderophores. Therefore, studying Lcn2-siderophore complexes contributes to the understanding of the releasing mechanism and siderophore selectivity. Implementing isothermal titration calorimetry (ITC), hydrogen-deuterium exchange (HDX), and native electrospray ionization mass spectrometry (ESI MS), we confirm the conformational shift to a more unfolded structure upon siderophore binding. Ferric 2,3-DHBA-Lcn2 is the most dynamic in its structure and readily releases siderophore intracellularly. Ferric Ent-Lcn2 with a less flexible structure, is less susceptible to gas-phase unfolding and dissociation. An intermediate affinity interaction between Ent and Lcn2 unfolds the protein as well.

2.2 Introduction

2.2.1 Gut microbiota symbiosis and bacterial siderophores

In recent decades, the presence of microbiota in the gastrointestinal tract has been recognized as a large contributing factor toward human health and overall homeostasis.[48] Together with resident and transient bacteria, the total number of microbes are estimated to be in the same order of magnitude as human cells, with a biomass of around 0.2kg.[49] The microbes are actively engaged in symbiosis with the host so that both organisms gain advantages in this relationship. Research shows that mice grown in a sterile environment (germ-free mice) have metabolic and immune defects.[50,51] More in-depth studies reveal this bi-directional interaction cross a single layer of gut epithelial cells. In this symbiotic process, microbes utilize plant polysaccharides that are otherwise useless to humans to synthesize essential amino acids (arginine and glutamine) and vitamin K for the host.[52,53] Other benefits of gut microbiota include the prevention of colonization of pathogens through secretion of various bacitracin, a type of antibiotic produced by bacteria, to inhibit growth or kill pathogenic bacteria.[54,55]

Despite flourishing of a large gut microbe population, the gastrointestinal tract can be considered as a rather hostile environment, in which microbes need to constantly scavenge for nutrients and evade the host immune system. Generally, the host immune system is designed to eliminate all non-self, but the resident microbes have trained the immune system to tolerate their colonization at an early stage of the development.[56] The remaining challenge becomes how to survive with limited local metal availability, especially iron. To acquire enough iron for DNA replication and protein cofactors, bacteria implement various siderophores, specialized in chelating iron at high affinities.[57] Upon secretion, siderophores scavenge iron and transport it back to the bacteria.[58,59]

2.2.2 Enterobactin and its bidentate intermediate

Grouped based on their functional groups, the three major types of siderophores are carboxylates, hydroxamates, and catecholates (Figure 2.1). At physiological pH, the deprotonated carboxyl or hydroxyl groups coordinate and capture iron. Multiple functional groups can coexist in a single siderophore to enhance iron affinity. Siderophores that consist of more than one type of functional group are categorized as mixed type.[60] Enterobactin (Ent), with the highest known affinity towards iron ($K_d = 10^{-52}$ M), is utilized by both common enteric and pathogenic bacteria.[61] The strong interaction with iron comes from three identical catecholate units.[62] Starting with chorismic acid converting into 2,3-dihydroxybenzoic acid (DHBA), this first step of the enterobactin biosynthesis pathway is catalyzed by enzymes EntB, EntC and EntA. Followed by EntE and EntF, three 2,3-DHBA are then attached to three L-serine and linked together through serine internal cyclization (Figure 2.2).[63,64] The hydroxyl oxygen atoms are aligned with the amide proton facing outwards at physiological pH. This orientation favors the binding of iron at the hydroxyl groups. Upon binding, subsequent deprotonation and peptide bond rotation, iron is fully encapsulated by a hexadentate metal cage formed at the centre (Figure 2.3a).[65] Although the binding affinity is substantially weaker, the intermediate 2,3-DHBA is identified to be a siderophore as well.[66] According to previous studies, the complex 3[2,3-DHBA]Fe is predicted to form a similar iron cage (Figure 2.3b).[67]

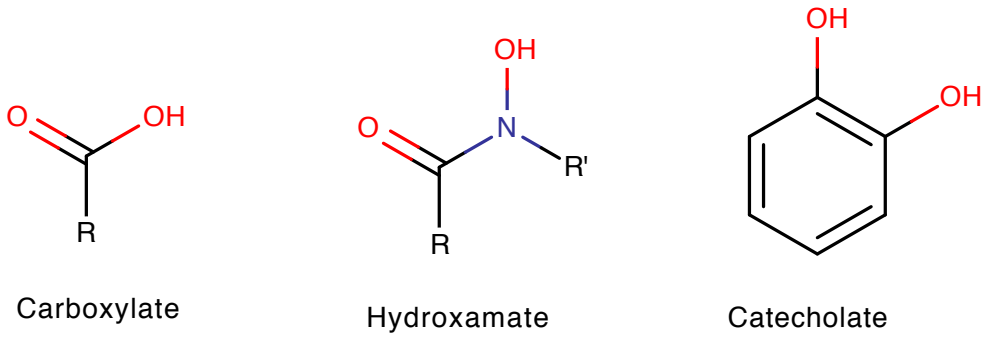


Figure 2.1. Functional groups of the three major types of siderophores.

Iron interacts with siderophores at the deprotonated oxygens in different functional groups. Multiple functional groups or multiple types can cooperate in a single siderophore molecule.

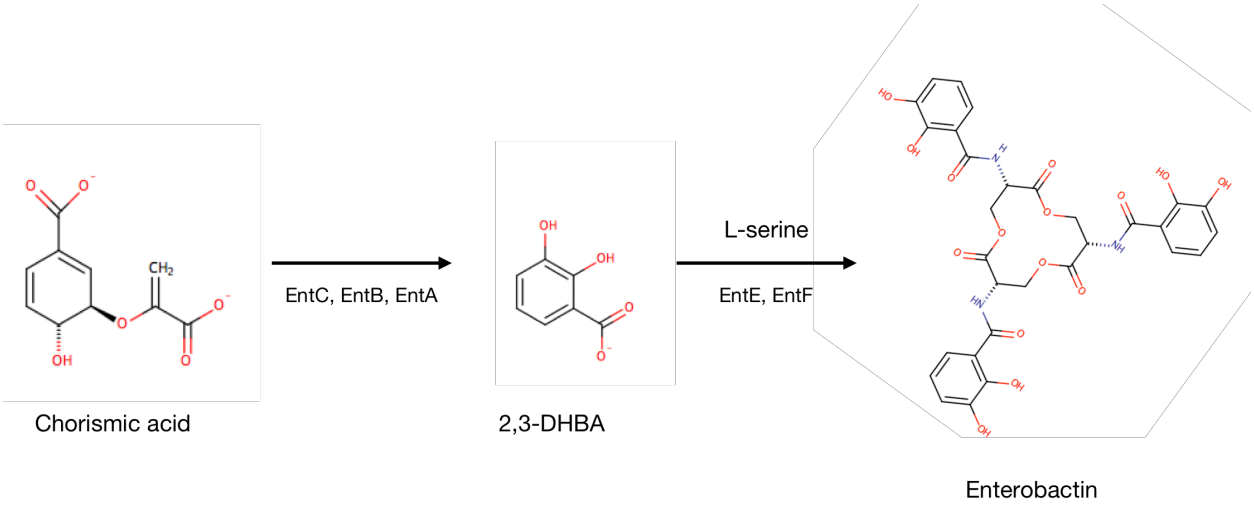


Figure 2.2. The biosynthesis pathway of enterobactin.

In *E. coli*, enterobactin is synthesized by a group of five enzymes. For simplification, only one intermediate is shown in the pathway. The L-serine is attached to 2,3-DHBA through a peptide bond at the carboxyl group.

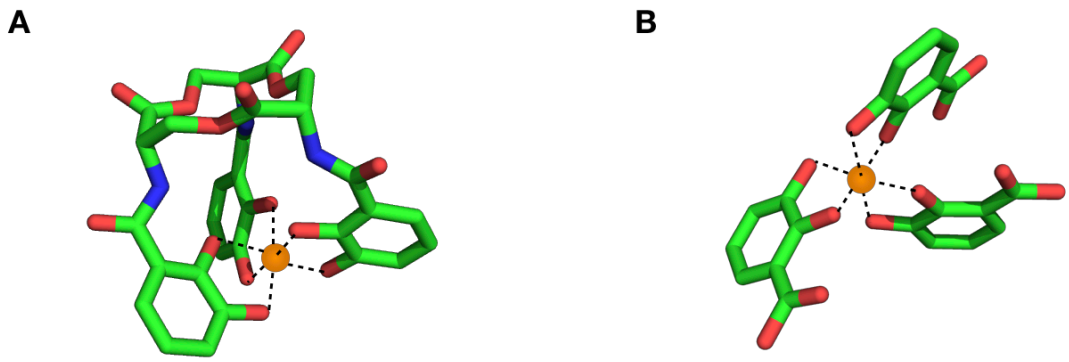


Figure 2.3. The iron cage of catecholate siderophores.

The iron cage of (A) enterobactin consists of three catecholate units and six deprotonated hydroxyl groups. (B) Three 2,3-DHBA are proposed to form a similar hexadentate iron cage based on previously reported results.[67] The iron molecules are shown as a brown sphere at the centre.

2.2.3 Lcn2 battles against iron-stealing bacteria

Naturally, when bacteria evolve a mechanism to acquire iron, the host develops a counteractive strategy against iron stealing. A classic tug-of-war model describes this constant battle. While Ent chelates and brings iron back into bacteria through the FepA receptor, the host utilizes lipocalin 2 (Lcn2, or siderocalin) as a defence mechanism to sequester ferric Ent and prevent bacterial iron acquisition (Figure 2.4).[68,69] Not only does this stop iron transportation into bacteria, but Lcn2 also helps to internalize ferric siderophores into the cells and contributes to host iron homeostasis.[70] A recent study found that Ent-Fe associates with the ATP synthase subunit α and promotes iron influx into the mitochondria (Figure 2.4).[71] The interaction between Ent and ATP synthase α provides a possible destination after siderophore internalization. However, the detailed

releasing mechanism from Lcn2 and whether other siderophores behave in a similar manner remains unclear.

Although Lcn2 is an effective defence protein, it only successfully chelates certain siderophores. Lcn2 has a β-barrel with a shallow ligand-binding cavity or calyx.[72,73] To evade the shallow calyx of Lcn2, bacteria glycosylate Ent into a bulkier salmochelin.[74] Many other structurally distinct siderophores can also deliver iron into bacteria without interference from Lcn2.[75]

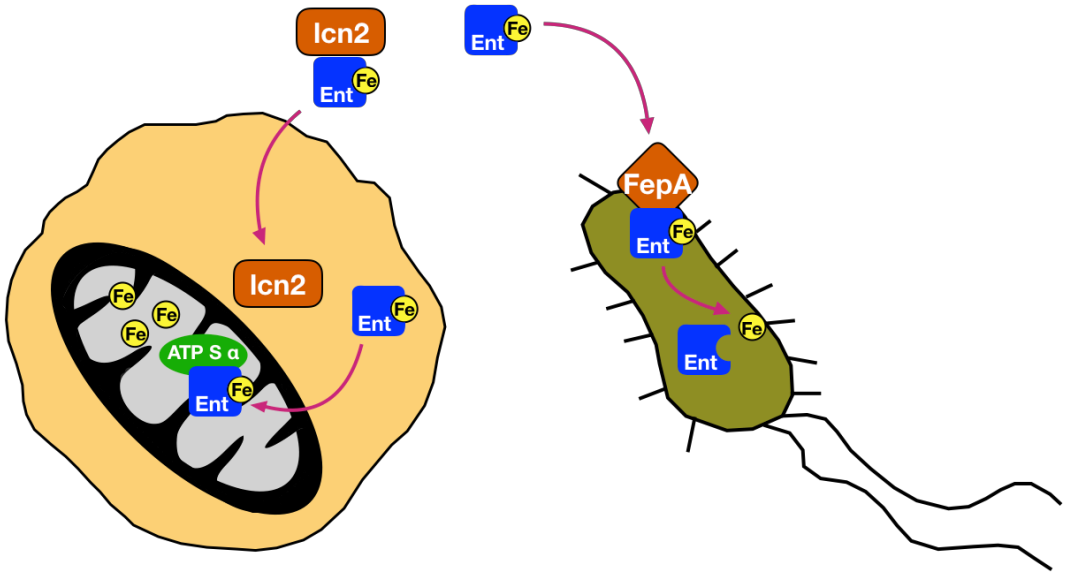


Figure 2.4. Iron tug-of-war between host cells and bacteria.

The better-understood siderophore, Ent, is used to illustrate the iron tug-of-war. The bacteria gain iron via FepA receptor, while host cells rescue iron via Lcn2. The Ent pathway is based on *E. coli* iron uptake. A plausible pathway of increasing intracellular iron concentration by Lcn2-siderophore complexes is shown without a detailed internalization mechanism.

2.2.4 Beyond its antimicrobial role: Lcn2 functions in the absence of infection

Other than its antimicrobial function, prolonged elevation of circulating Lcn2 is also associated with various diseases and inflammation.[76–79] However, the exact pathway of how Lcn2 is linked to inflammation is still somewhat vague. Lcn2 appears to prevent inflammation by promoting bacterial clearance in inflammatory bowel disease.[80] On the other hand, seemingly contradictory evidence implies a pro-inflammatory effect of Lcn2 in low-grade inflammation.[81] Under chronic neuroinflammatory conditions, Lcn2 promotes inflammation through the TNFα pathway.[82] Other evidence in obese and diabetic models affirms the pro-inflammatory hypothesis.[78,83,84] Promoting inflammation or not, Lcn2 is a master regulator in many processes even without infections.[77] Because Lcn2 has no iron affinity by itself, the mammalian siderophore hypothesis arises. A hydroxylmate siderophore with structural similarity to bacterial 2,3-DHBA, 2,5-dihydroxy benzoic acid (2,5-DHBA) is the best candidate for a mammalian siderophore.[85] Other research has argued that the host utilizes siderophores produced by resident gut bacteria, instead of synthesizing a specific mammalian siderophore.[86]

2.2.5 Iron overload and reactive oxygen species (ROS) accumulation

Combining both anti- and pro-inflammatory functions of Lcn2, a dual role of Lcn2 is suggested based on its free (apo-Lcn2) and iron-loaded (holo-Lcn2) forms.[70] A recent study hypothesizes that apo-Lcn2 prolongs cell survival upon intercellular accumulation, but holo-Lcn2 promotes cell death through accumulating ROS.[67] Although ROS is produced under physiological conditions as a pivotal signalling molecule, excess ROS initiates oxidative stress, cytotoxicity, and even apoptosis.[87,88] The formation of hydroxide radicals, one type of intracellular ROS, is directly catalyzed by iron(II) via Fenton-like

chemistry.[89] Cells regulate free iron by incorporating it into protein, such as ferritin or other heme-containing proteins.[90] However, cells overloaded with iron inevitably accumulate excess ROS.[91] Because Lcn2 traffics iron across the cell membrane, the nature of the complex and iron-releasing efficiency are the key features that we aim to investigate.

2.3 Experimental methods

2.3.1 Lcn2 expression and purification

2.3.1.1 Apo-Lcn2 expression

Plasmid pPROEX HTb mLipocalin-2 was received as a dry sample on a Whatman #1 filter paper from the Wang lab (University of Hong Kong). Filter paper containing DNA sample was cut into small pieces and submerged in 100 µl nuclease-free water. The solution was vortexed and incubated at room temperature for 5 min before removing the filter paper by centrifugation. Plasmid concentration was measured by Nanodrop 2000 (Thermo Scientific).

To transform chemically competent BL21(DE3) (NEB# C2527I), 80 ng of reconstituted plasmid was added into 50µl of competent cells. The mixture was chilled on ice for 30 min before heat shocked at 42˚C for 10 sec. After another 5-minute incubation on ice, 940 µl of SOC outgrowth medium (2% vegetable peptone, 0.5% yeast extract, 10 mM NaCl, 2.5mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added for rapid growth at 37˚C for 1 hr with vigorous shaking at 250 rpm. The culture was plated on ampicillin (100µg/ml) LB-agar plates and incubated at 37˚C O/N. A single colony was used to inoculate a 20 ml culture (LB broth with 100 µg/ml ampicillin) growing at 37˚C O/N. Stock bacteria were made by mixing overnight culture and 50% autoclaved glycerol at a 1-to-1 ratio, and stored at -80˚C.

For large scale expression, 5 ml O/N cultures were used to inoculate 500 ml LB broth. The bacteria were incubated at 37˚C for 4 hrs with 200 rpm shaking. Afterwards, the induction took place under the same conditions for 2 hrs with 1 mM IPTG (1M stock concentration). Bacteria were pelleted by centrifugation at 6,000xg for 25 min. The pellet was stored at -20˚C for downstream purification.

2.3.1.2 Nickle immobilized metal ion affinity chromatography (Ni-IMAC)

Induced bacteria pellet was resuspended in 20 ml equilibrium buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) and lysed by sonication (20% amplitude, 15 sec pulse, 15 sec intervals, 5 min). Crude protein was collected in the supernatant by centrifugation at 4˚C, 18,000 rpm for 40 min. Total protein supernatant was filtered with Acrodisc syringe filters (0.45 µm, Pall Corporation). A column with Nickel Sepharose 6 Fast Flow resin (GE Healthcare Life Sciences) was prepared by washing with 2 column volumes (CV) of ddH_2O and 5 CV of IMAC equilibrium buffer. Binding of nickel beads and crude protein was initiated by incubation at 4˚C for 30 min. The column was further washed with 20 CV of IMAC wash buffer 1 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and 5 CV of IMAC wash buffer 2 (50 mM NaH_2PO_4, 300 mM NaCl, 50 mM imidazole, pH 8.0). Purified protein was eluted with 6 CV of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole). All flow-through fractions were collected for 12.5% SDS-PAGE analysis.

2.3.1.3 Lcn2 concentration and storage

Purified Lcn2 fractions were pooled together and dialyzed against a total of 8 L phosphate saline buffer (1X PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM

KH_2PO_4, pH 7.4) using cellulose dialysis tubing (12 kDa to 14 kDa cut-off, Thermo Fisher Scientific). The extinction coefficient was computed to be 31985 M^{-1} cm^{-1} by ExPASy ProtParam tool. Lcn2 concentration was estimated by OD_{280} before concentrating with Amicon Ultra-15 centrifugal filters (10kDa MW cut-off). Final protein concentration was measured by Bradford assay. Lcn2 was stored at a concentration higher than 1 mg/ml in 1X PBS at -80˚C.

2.3.2 Lcn2-siderophore complex formation

Stock siderophores 2,3-DHBA and Ent were dissolved in DMSO at concentrations of 300mM and 15mM respectively. Ratios of Lcn2:2,3-DHBA (1:6) or Lcn2:Ent (1:2) were used to form complexes with Lcn2 in the absence of iron. Excess siderophore were used to favor the formation of the complex and limit free Lcn2 in the solution. When forming holo-Lcn2 complexes, $FeCl_3$ (10 mM in ddH_2O) and siderophores were mixed at room temperature before introducing Lcn2 into the mixture. Based on preliminary results of stoichiometry in the complex, the ratio of 2,3-DHBA:Fe(III) was fixed to 3:1.[67] Iron-laden $[Fe(III)Ent]^{3-}$ was formed at a 1:1 ratio. All complexes were formed in 1X PBS (pH 7.4) at 20 μM Lcn2 concentration. The mixture was incubated at room temperature for 30 min. Any unbound iron or siderophores were washed away during Zeba spin column desalting (Thermo Fisher) prior to MS analysis.

2.3.3 Isothermal Titration Calorimetry (ITC)

Degassed samples were analyzed by a MicroCal Thermo Vac unit for ITC experiments. The titrant, siderophores, in the syringe were injected into a protein solution in the thermally controlled cell. The association of Lcn2 and ligand was performed at 20˚C with

4 μM protein and 0.08 mM Ent. Although the Ent stock was dissolved in DMSO, the final concentration in the cell is only 0.56% (v/v) after dilution with 1X PBS. Higher concentrations of protein (20 μM) and 2,3-DHBA (0.3 mM) were used due to the lower affinity.[92] A final DMSO concentration was calculated to be 0.0267%.

Lcn2 was freshly dialyzed in 20 ml of 1X PBS prior to ITC, and the dialyzed buffer was used to dilute the siderophore stocks in order to prevent any possible buffer component mismatch. The same concentration of DMSO was added to both reference and analyte cells for buffer consistency. The titration started by 2 ml injection and 1 min pause, followed by 34 subsequences with 300 sec intervals. The first data point was excluded from the analysis. The baseline was corrected with the heat of dilution, measured by titrating ligands into 1X PBS with the same percentage of DMSO. When applicable, a sequential binding model was used to fit and describe the extracted curve.[93]

2.3.4 Intrinsic fluorescence emission scan

Apo-lcn2 or siderophore-siderocalin complexes were prepared at 5 μM in 1X PBS as described above. The samples were excited at a wavelength of 280 nm (bandwidth 9 nm). The emission fluorescence was measured on a Synergy H4 (BioTek). The emission spectra were collected at 2 nm intervals over a 300 to 400 nm range at 22˚C. Results were baseline corrected with buffer and siderophore emission spectra.

2.3.5 ESI MS of native and partially-unfolded Lcn2 complexes

Stock protein was dialyzed into 100mM NH_4Ac (pH 6.8) for downstream mass spectrometry applications. Both apo-Lcn2 and Lcn2-siderophore complexes were identified by ESI MS (Waters Synapt G2-S) with direct infusion at 2 μM protein

concentration and 20ul/min flow rate. To unfold the protein, Lcn2 was diluted 10X in 10% acetic acid with or without 5% ACN. The instrument was operated in positive ion mode with 2.5 kV source voltage and 30V sampling cone. Trap and transfer collision energy were set to 15 V and 10 V respectively.

2.3.6 Ion Mobility (IM) MS

2.3.6.1 Arrival time distribution (ATD) and analysis

The tune profile of native ESI-MS was kept the same in the IM MS experiments with additional ion mobility conditions. The travelling wave was set to 600 m/s velocity and 30V height. IMS cell was filled with nitrogen gas flowing at 90 ml/min, giving a pressure of 0.319 mbar. The enhanced duty cycle (EDC) delay coefficient was set to 1.41. The ATD was the extracted ion mobility chromatogram at a single charge state.

2.3.6.2 Collisional induced unfolding (CIU) and Collisional induced dissociation (CID)

Identical IM MS settings from section 2.3.6.1 were used for both CIU and CID. The trap cell collisional voltage was ramped from 10 V to 80 V. In CIU experiments, the drift time is plotted against the trap collisional energy (CE). Intensities are represented by a heat map. In CID experiments, changes in mass were monitored with increasing trap CE. The intensity damping of the 10+ charge state was graphed against CE. The Boltzmann sigmoidal function was used to describe the intensity changes.

2.3.6.3 Collision cross section (CCS) calculation

In the travelling wave ion mobility system, CCS can only be calculated experimentally due to the constantly changing electric field. Three denatured proteins, equine

cytochrome C (CytC), horse heart myoglobin and bovine ubiquitin in MeOH:H_2O:acetic acid (49:49:2) were used as calibrants. The same instrumental conditions were applied (section 2.3.6.2) with minimal trap CE at 10 eV. CCS (Ω_c) was corrected with literature CCS (Ω) from Prof. David Clemmer's Cross-Section Database (6), where z is the charge state, e_c is the electron charge in Coulomb, m is the molecular weight of protein calibrant, and M_{N2} is the molecular weight of nitrogen.

$$\Omega_c = \frac{\Omega}{e_c \times z\sqrt{\frac{1}{m}+\frac{1}{M_{N_2}}}} \tag{6}$$

The drift time was corrected for EDC based on (7). The $\ln(\Omega_c)$ is plotted against ln(td') and fitted to linear regression (8) to find X and A.

$$t_d' = t_d \times \frac{EDC \times \sqrt{m/z}}{1000} \tag{7}$$

$$\ln(\Omega_c) = X \ln(t_d') + A \tag{8}$$

The td' is corrected again with experimentally determined exponential factor (X) with equation (9). The CCS($\Omega_{unknown}$) of the target protein is calculated from fit-determined constant (A) and td'' by equation (10). When experimental t_d is a series of values, this sequence of equations can be used to calculate a series of CCS as Y-axis in CIU fingerprints.

$$t_d'' = z \times t_d'^{X}\sqrt{\frac{1}{m}+\frac{1}{M_{N_2}}} \tag{9}$$

$$\Omega_{unknown} = A t_d'' \tag{10}$$

2.3.7 TRESI HDX MS

2.3.7.1 Protease-agarose bead cross-linking

To crosslink pepsin from porcine gastric mucosa (Sigma-Aldrich, #P6887) or protease XIII from *Aspergillus saitoi* (Sigma-Aldrich, #P2143) to N-hydroxysuccinimide(NHS) activated agarose beads, 20 mg of protease and 50 mg of agarose beads were mixed in 1 ml ligation buffer (0.1M $NaPO_4$, 0.15M NaCl, pH 5) overnight at 4˚C on an orbital rotator. The beads were then washed twice with 1 ml blocking buffer (1M Tris, pH 5). At the end of each wash, beads were collected by centrifugation at 2,000xg for 2 min. Subsequently, protease beads were incubated in blocking buffer for another 2 hrs at 4˚C with rotation. The cross-linked agarose beads were then washed and stored in 10% acetic acid (pH 2.5).

2.3.7.2 Global HDX data acquisition and analysis

The microfluidic set-up for HDX experiments was assembled as previously described.[19,94] Briefly, HDX took place in a capillary-based kinetic mixer, allowing for various reaction times. The inner glass capillary (OD 150.8 µm, ID 75.3µm) and outer metal capillary (120 mm in length, 28RW, ID 177.8 µM) created a narrow inter-capillary space. The quenching acid was delivered with the same sized glass capillary. At the end, analyte solution was sent into MS via a metal capillary (33RW, OD 203.2 µM, ID 101.6 µM). All deuterium, protein, and acid were delivered by syringes pumps (Pump 11 Pico Plus Elite, Harvard Apparatus).

In the global HDX set up, the protein was sprayed directly into the mass spectrometer after HDX labelling. Based on flow rates of 5 µl/min protein and 10 µl/min deuterium, seven-time points were collected from 0.25 sec to 5.12 sec. Native ESI spectra were

taken before labelling and at each time point. The mass increases were determined by the deconvoluted mass of each spectrum.

2.3.7.3 Microfluidic device fabrication

After HDX labelling and acid quenching, the analyte solution was passed through a glass capillary (ID 75.3μm) into a poly (methyl methacrylate) (PMMA) microfluidic device with immobilized proteases. The microfluidic device was composed of two PMMA substrates at dimensions 0.83'' x 2'' x 0.25'', one blank and another with an etched protease chamber made using a Versa Laser™ engraver. The two PMMA substrates were fused together using acrylic weld.[94]

2.3.7.4 Data acquisition and analysis for digested workflow of TRESI HDX

In the bottom-up HDX experiments, the deuterated protein was quenched with 10% acetic acid and 5% acetonitrile before online digestion in a PMMA device. A mixture of pepsin and protease XIII linked agarose beads at a ratio 1:1 was used in the protease chamber. With protein and deuterium flow rates each at 2 μl/min, the HDX reaction times were calculated to be 0.68 sec and 3.4 sec at 1 mm and 10 mm inner capillary pull-back positions, respectively. Excess acid at a flow rate of 16 μl/min was used to dilute deuterium and ensure sufficient quenching.

Peptide sequences were first searched based on unique mass using the ExPASy FindPept tool and confirmed by MS/MS fragmentation. The percentage deuterium uptake for each peptide was analyzed in Mass Spec Studio.[95] Differences in deuterium uptake between the complex and apo-lcn2 were reported. The differences for all time points were summed up in the final figure.

2.4 Results and Discussion

2.4.1 Purification of apo-lcn2

The pellet after IPTG induction and supernatant following sonication were monitored for Lcn2 over-expression levels. The calculated molecular weight from the amino acid sequence, 24.2 kDa, positioned apo-Lcn2 around the 25 kDa marker band. A significantly thickened band around 25 kDa represented a successful induction. Sonication supernatant was hard to interpret due to a large amount of protein present. IMAC wash fractions were collected one column volume after each wash buffer was loaded. IMAC wash buffer 1 cleaned most of the bacterial proteins. IMAC wash buffer 2 resulted in some loss of Lcn2 product, however one column purification was possible with the lower yield. A single band around 25 kDa in the 250mM imidazole elution fraction signalled a clean Lcn2 product (Figure 2.5).

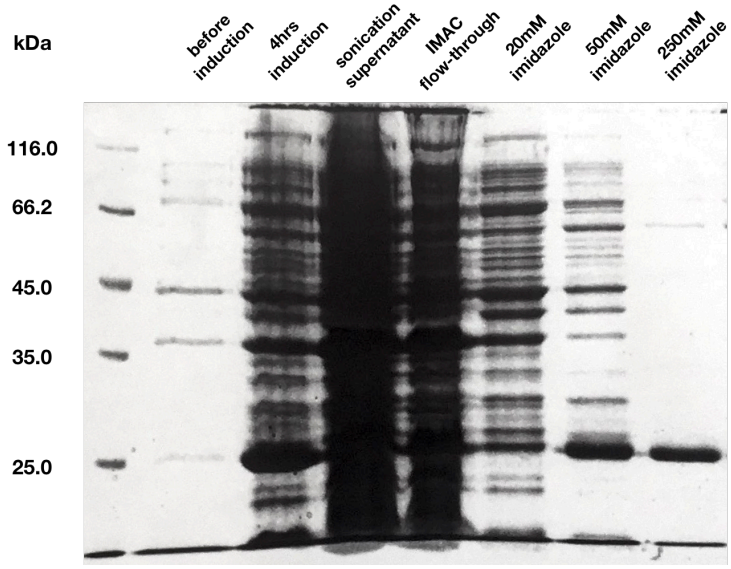


Figure 2.5. The Ni-IMAC single column purification of Lcn2.

From left to right, the second and third lanes were pellets before and after induction. The fourth lane was sonication supernatant after cell debris removal. Immediately following were the flow through of Ni-IMAC, two washes with different imidazole concentrations, and elution of the final product.

2.4.2 The change in thermodynamic parameters from siderophore interaction

To study the shifts in system thermodynamics of Lcn2-siderophore complex formation, ITC experiments were performed by titrating Ent and 2,3-DHBA into the protein solution. Because of the high K_d of ferric Ent, a very low protein concentration was required to obtain a reasonable dimensionless sigmoidicity constant (c) according to equation (11), where M is the protein concentration and n is the stoichiometry. The association constant (K_a) is another expression of affinity, related to K_d by equation (12). Considering the small heat released with low concentration, the experiment with ferric Ent was not feasible in the current set-up.

$$\mathbf{c = K_a \times M \times n} \tag{11}$$

$$\mathbf{K_a = K_d^{-1}} \tag{12}$$

2,3-DHBA gave an uninterpretable result, with noisy heat release close to the baseline (Figure 2.6). The interaction between 2,3-DHBA and Lcn2 was deemed not present at a protein concentration of 20 μM. On the other hand, the binding event was clear in the Ent Lcn2 pair despite the large error in curve fitting with a sequential binding model. The large uncertainty gave a less meaningful value but in the same magnitude K_{d1} (1.1 ± 9.7 nM) as reported in literature. The biphasic curve was clear evidence for two-event binding, where a second event in addition to the initial ligand association changed the system enthalpy (Figure 2.7).

The second event indicated by the ITC curve has several possibilities, including a second binding site,[96,97] ligand-induced dimerization,[98] or perturbation of protein folding/unfolding equilibrium.[99,100] As the crystal structure of the human homolog NGAL revealed, Lcn2 only has one cup-shaped cavity suitable for Ent.[73] Both dimerization and unfolding are plausible based on ITC data alone. Since Lcn2 is reported to be both

monomeric and dimeric, the possibility of dimerization is examined by native MS ESI MS.[101]

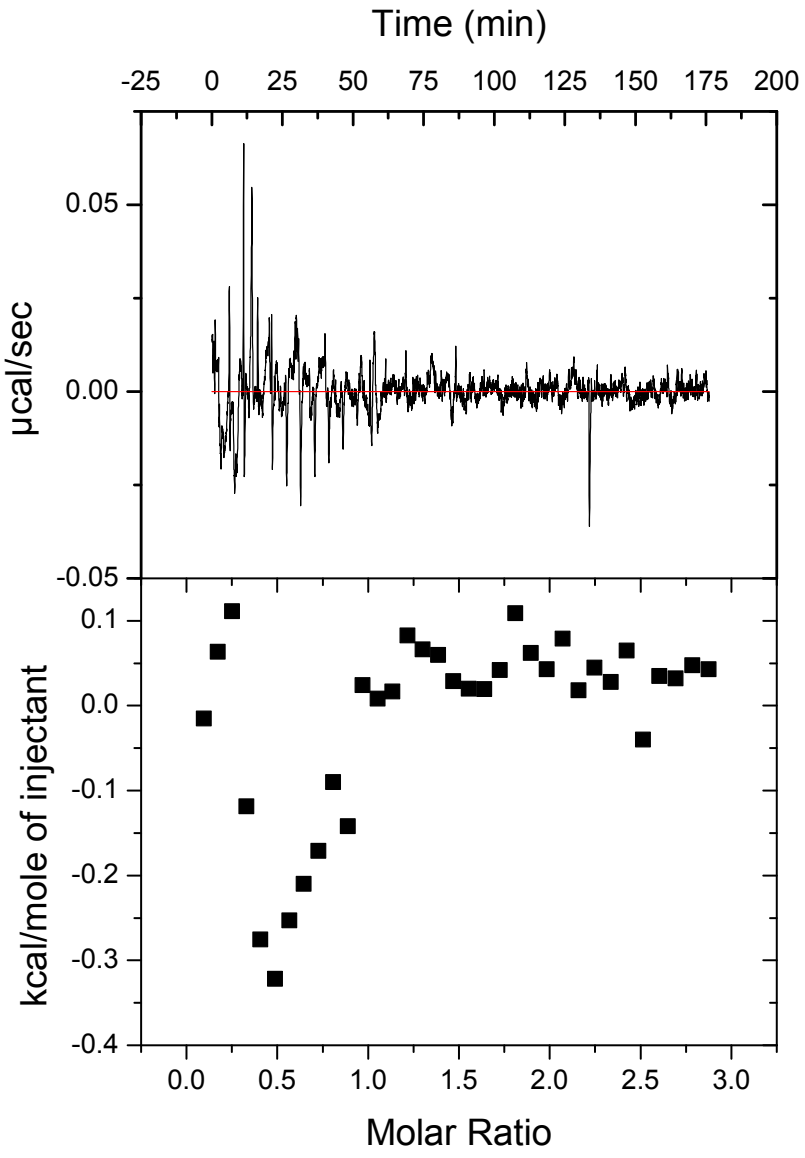


Figure 2.6. ITC of 2,3-DHBA and Lcn2 in the absence of iron.

The heat released from each injection is shown in the top panel. The integrated heat on the bottom panel cannot be fit with any known models. Both graphs are corrected with the heat of dilution.

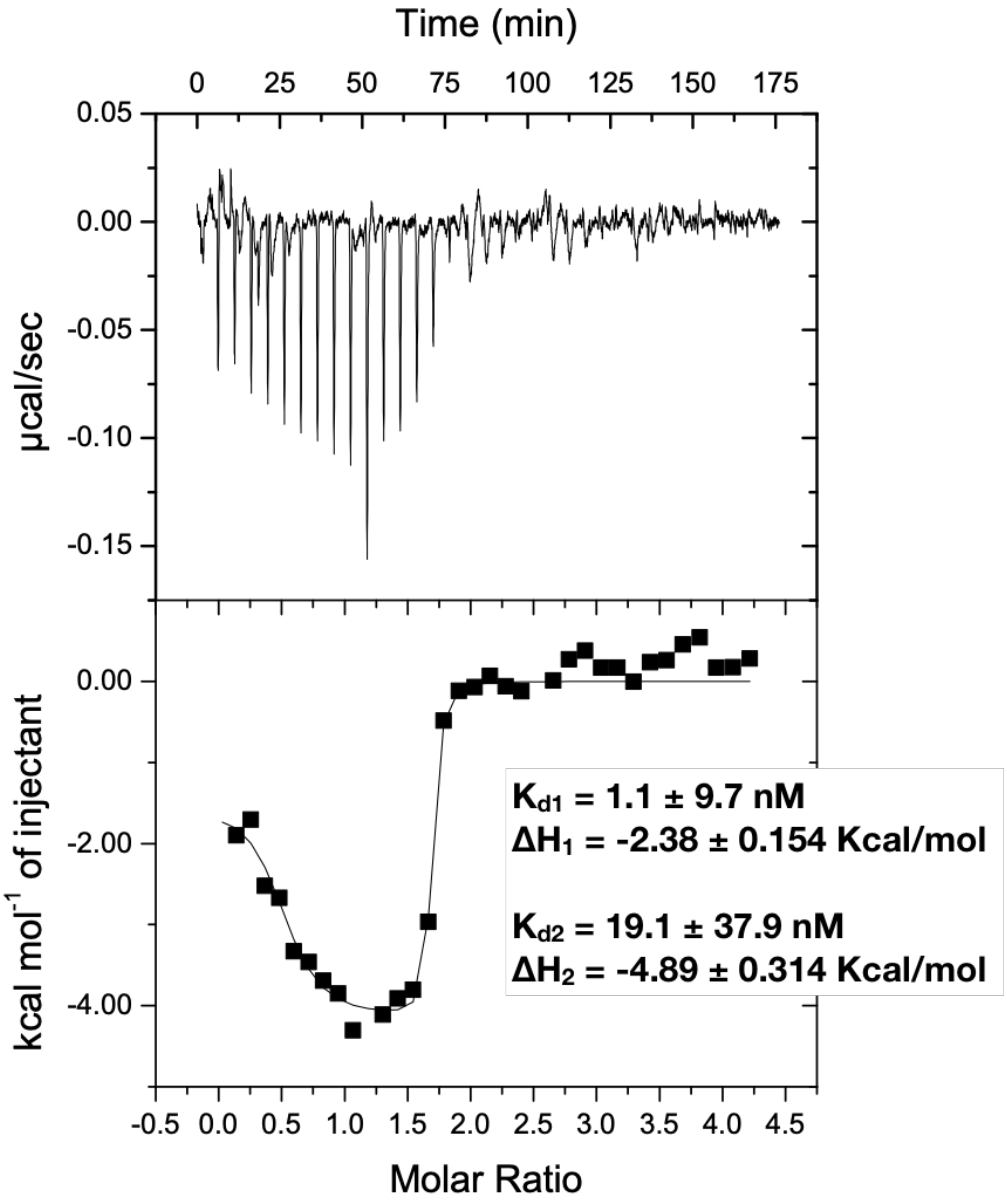


Figure 2.7. ITC of enterobactin and Lcn2 in the absence of iron.

Both the heat released per injection (top panel) and the integrated heat (bottom panel) are baseline corrected. A sequential binding model is the best in fitting the U-shaped curve. The large error in the dissociation constants comes from the low heat and model fitting. As a result, enthalpy changes are less accurate despite the seemingly less uncertainty.

2.4.3 Verification of siderocalin-siderophore complex formation

In the absence of iron, siderocalin and siderophore were allowed to interact at twice the stoichiometry ratio to form the complex. Two steps were involved in creating a holo-Lcn2 complex. Iron-chelated ferric enterobactin or ferric 2,3-DHBA always formed first. Apo-Lcn2 was then added in the solution to form a ferric complex so that the competition between iron-loaded and iron-free siderophore was minimized. Complex formation was verified by both intrinsic fluorescence quenching and native ESI MS.

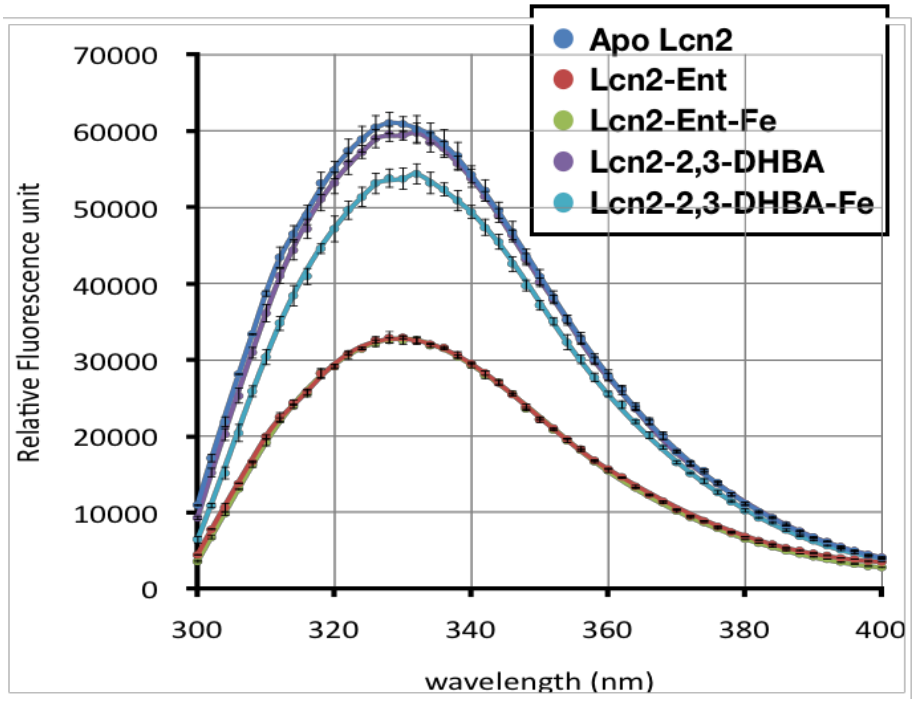


Figure 2.8. Intrinsic fluorescence quenching in Lcn2-siderophore complexes. The emission spectrum (n=3) of Lcn2 at 280 nm excitation was collected in the presence of siderophores. The spectrum was baseline corrected with buffer and siderophore emission. The emissions of Lcn2-Ent and Lcn2-Ent-Fe were similar and overlapping.

According to the amino acid sequence, eleven tyrosines and two tryptophans are located in the β barrel of Lcn2. Therefore, ligand binding at the barrel was predicted to change the intrinsic fluorescence intensity. Indeed, when Lcn2 was excited at wavelength

280 nm, fluorescence quenching was observed for ferric enterobactin, enterobactin and ferric 2,3-DHBA (Figure 2.8). In agreement with the ITC experiment, 2,3-DHBA failed to produce meaningful intrinsic fluorescence changes above the standard deviation level. Enterobactin in both iron-associated and iron-free forms showed similar levels of quenching (Figure 2.8). This observation was consistent with literature reported affinities at 0.4nM and 4nM for ferric Ent and Ent respectively.

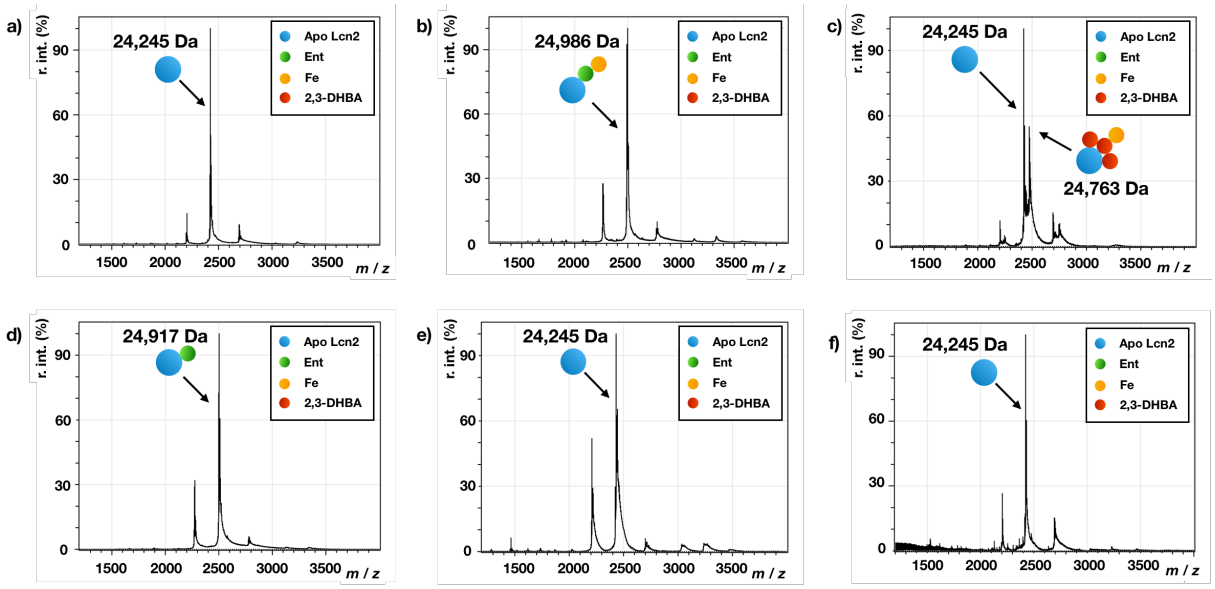


Figure 2.9. Native ESI MS of Lcn2 and various siderophores.

(a) Lcn2 formed complexes with (b) ferric Ent, (c) three 2,3-DHBA with iron, and (d) Ent. The absence of a complex peak at higher *m/z* was shown in Lcn2 mixed with (e) 2,3-DHBA or (f) 2,5-DHBA. All mixtures were kept at the same protein concentration and acquired using identical instrument conditions. The presence of complex peaks was consistent in (b)(c)(d) in multiple replicates (n>3), although the ratio between apo-Lcn2 and bound Lcn2 fluctuated.

The formation of Lcn2-Ent and Lcn2-Ent-Fe were clearly indicated by a mass shift in ESI MS (Figure 2.9 a,b,d). An increase in mass was also observed with 2,3-DHBA but only in the presence of iron (Figure 2.9 c). Repeatedly, 2,3-DHBA alone was unable to bind to Lcn2 (Figure 2.9e). The addition of the most popular mammalian siderophore candidate, 2,5-DHBA, also did not trigger any complex formation (Figure 2.9f). Relaying on the stoichiometry information from ESI MS, both Ent complexes are identified to be at a 1:1 association. However, three 2,3-DHBA chelating one iron is bound to Lcn2. Likely, six hydroxyl groups from three 2,3-DHBA orient themselves into a hexadentate cage for iron, similar to the way Ent chelates iron. The lack of interaction between 2,5-DHBA and Lcn2 favours the hypothesis that a specific mammalian siderophore might not exist. Instead, the circulating bacterial siderophores are utilized by the host in Lcn2-mediated iron homeostasis.[86]

In addition to confirming the binding partners for Lcn2, ESI native MS can help verify if any dimerization is present. Despite the naturally occurring Lcn2 dimer in vivo, there is no evidence of induced dimerization in ESI MS. The low intensity peaks above 3,000 *m/z* were likely artificial (and non-covalent) dimers produced from the ESI source.[102] Therefore, the second event indicated by the ITC measurements is not oligomerization.

2.4.4 Siderophore-siderocalin complex stability

The stability of the siderocalin-siderophore complexes were investigated using three techniques: acid-induced unfolding, collisional induced dissociation (CID), and collisional induced unfolding (CIU). During acid-induced unfolding, the protein solution was diluted 10X into 10% acetic acid to a final pH of 2.5. The same ESI MS parameters were applied under acidic conditions. Charge states at lower *m/z* regions were more intense for

siderocalin-siderophore complexes compared to apo-Lcn2. Interestingly, emerging charges states preserved the same deconvoluted mass as the original folded species (Figure 2.10).

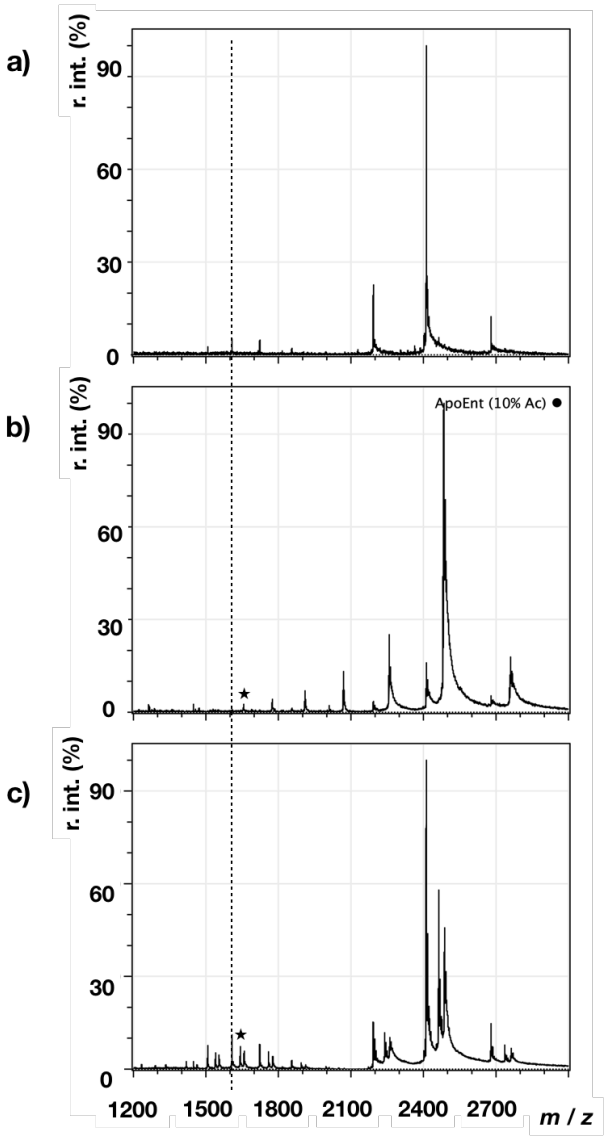


Figure 2.10. The acid induced unfolding of Lcn2 and Lcn2-siderophore complexes. The dashed line corresponds to the 15+ charge state of (a) apo-Lcn2 in the unfolded charge envelope. The stars mark the same charge state in (b) Lcn2-Ent and (c) Lcn2-2,3-DHBA-Fe. All unfolded complexes retained their corresponding siderophores, causing a shift to a higher *m/z* than the apo-Lcn2 peak.

Unfolded protein is more efficiently protonated and is shifted towards higher charge states at a lower *m/z* range. Therefore, the charge states occurring at the leftmost positions represent the most unfolded species that can be detected. The ligand-bound peak of this charge state indicates an acid-stable siderocalin-siderophore complex. Most likely, acidic conditions alone are not sufficient for siderophore release intracellularly without another protein to transfer the iron or siderophore to.

CID measures mass changes as the complex collides with gas molecules with increasing collisional energy (CE). A lower energy required to dissociate the complex is translated into a less stable interaction. The Lcn2-Ent complex fully dissociated at a CE of 60 eV, marked by a flat intensity change at CE range 60 to 80 eV (Figure 2.11a). For ferric Ent, with a higher affinity towards Lcn2, the corresponding complex fully dissociated at a CE of 80 eV (Figure 2.11b). All complexes had some unsaturated apo-Lcn2 to start with, but only the interaction between 2,3-DHBA and Lcn2 produced intermediates, Lcn2-[2]2,3-DHBA and Lcn2-[1]2,3-DHBA. The change of intensities for the intermediates occurred at CEs of 25 eV and 30 eV, respectively (Figure 2.11c).

Data points of apo-Lcn2 after a CE of 60 eV are eliminated from Boltzmann sigmoidal curve fitting since the decreases in intensities possibly come from the reduced transmission efficiency at high CE. Because Lcn2-Ent-Fe dissociation does not plateau at a CE of 80 eV, this complex is concluded to have a stronger protein-ligand interaction compared to Lcn2-Ent. The stepwise dissociation of Lcn2-2,3-DHBA-Fe reveals that iron is only properly chelated with three 2,3-DHBA. Once the first 2,3-DHBA is released, iron dissociates as well.

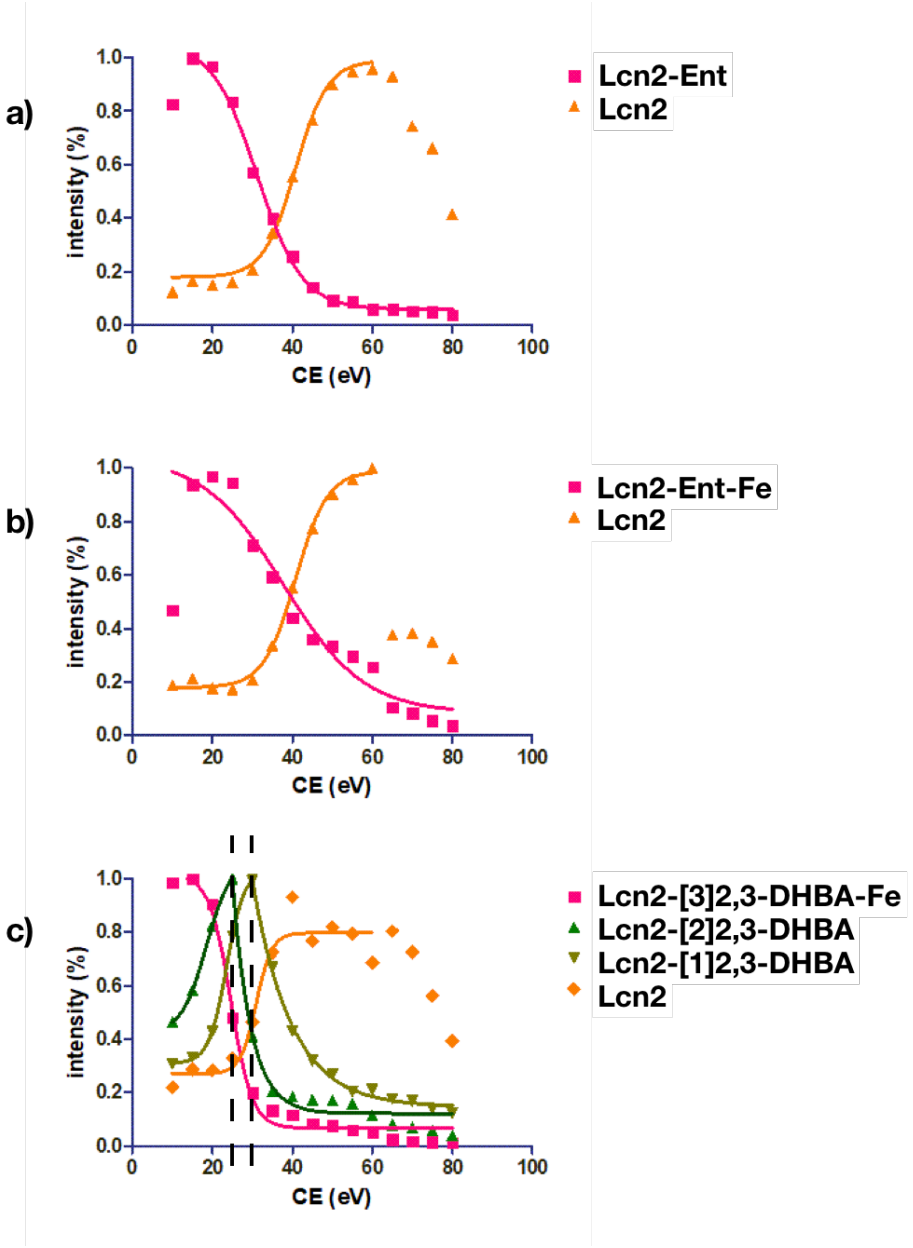


Figure 2.11. The collisional induced dissociation of the Lcn2-siderophore complexes.

The intensity shift of the 10+ charge state was monitored. The increased intensities of apo-Lcn2 in (a) Lcn2-Ent and (b) Lcn2-Ent-Fe were observed. (c) The step-wise dissociation with multiple species was captured in Lcn2-2,3-DHBA-Fe. The two dashed lines mark the splitting of forming and dissociation events of Lcn2 together with two 2,3-DHBA (Lcn2-[2]2,3-DHBA) or one 2,3-DHBA (Lcn2-[1]2,3-DHBA).

While CID measures the tightness of binding between Lcn2 and siderophore, CIU measures the readiness of complex unfolding. All drift times were converted into CCS for CIU fingerprint plots. A calibration curve with unfolded cytochrome c, myoglobin, and ubiquitin is shown in Figure 2.12 with R^2 0.99. The conversion was based on the CCS database of Prof. David Clemmer. The CIU plot of Lcn2-2,3-DHBA-Fe was not shown since the complex dissociated at very low CE. The three major unfolding states of apo-Lcn2 was compared to both Ent complexes (Figure 2.13a). Lcn2-Ent was more stable at both transitions (Figure 2.13b,d). Lcn2-Ent-Fe had a more stable first transition but a similar second transition (Figure 2.13c,e).

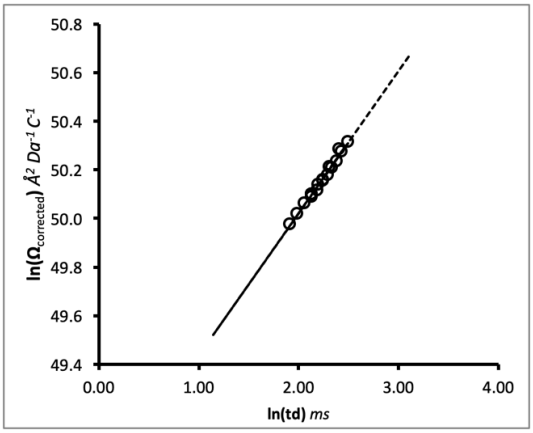


Figure 2.12. The calibration curve of TWIMS with cytochrome C, myoglobin, and ubiquitin.

The detailed calculation of CCS correction was described in section 2.3.6. Charge states with multiple conformations were eliminated. $\ln(\Omega_{corrected}) = x*\ln(t_d'') + A$ ($A = 1.64x10^{21} \pm 0.08\%$, $x = 0.587 \pm 2.93\%$).

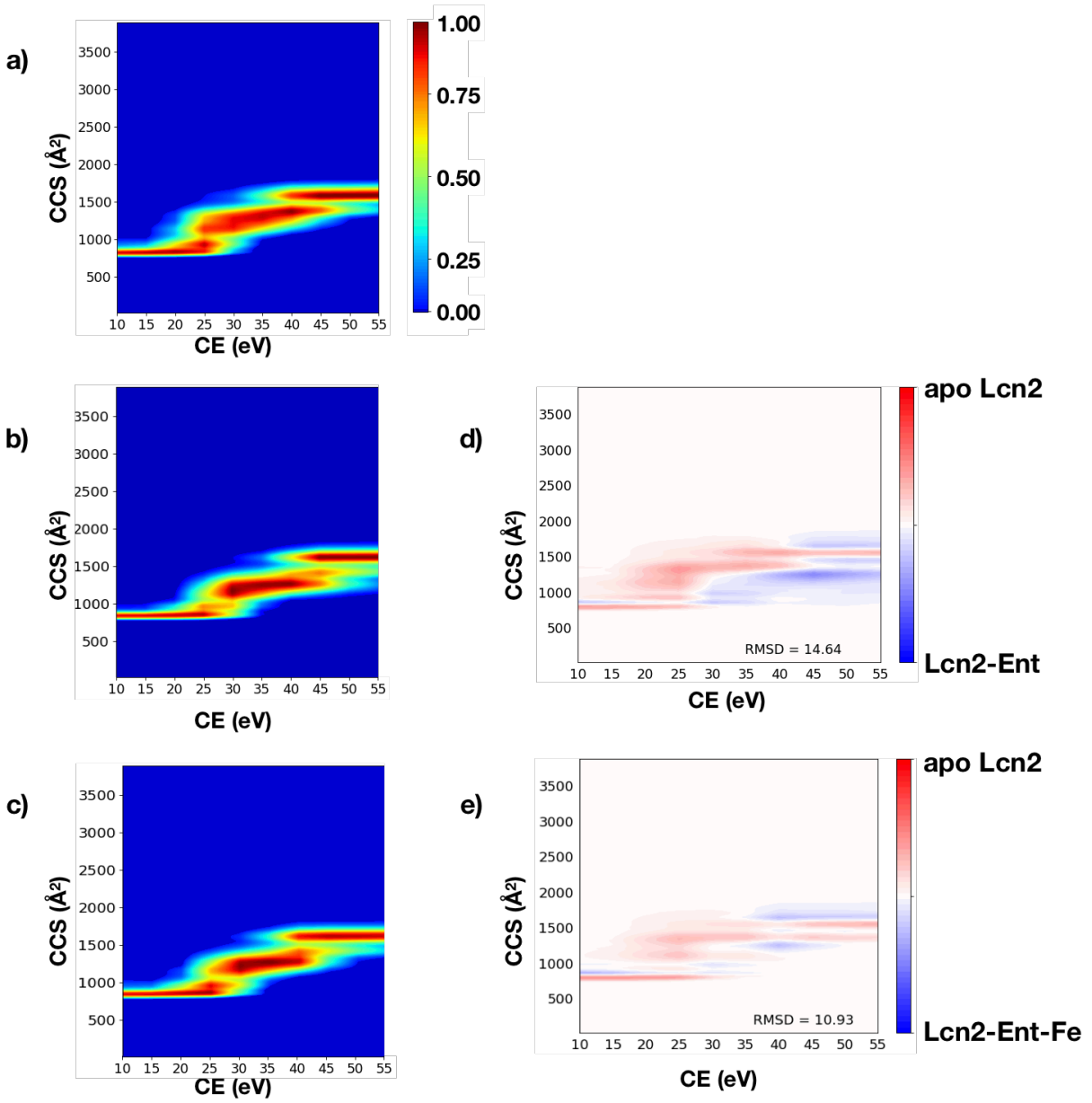


Figure 2.13. The collisional induced unfolding of free Lcn2 complexed with different siderophores.

The gas-phase unfolding (n=2) of (a) apo-Lcn2 had two obvious transitions. The same transitions were observed in (b) Lcn2-Ent and (c) Lcn2-Ent-Fe at different CEs. Due to the early dissociation of the Lcn2-2,3-DHBA-Fe complex, a CIU fingerprint was not generated. The changes in gas-phase unfolding of (d) Lcn2-Ent and (e) Lcn2-Ent-Fe was compared to apo-Lcn2.

Taken together, Lcn2-Ent-Fe has a stronger protein-ligand interaction than its iron-free counterpart, corresponding to a complex less susceptible to gas-phase unfolding. In the gas-phase, the first transition of unfolding at CE 25-30 eV is at the same energy required for complex dissociation. The overlapping CE range implies that the dissociation might be initiated by unfolding in the gas phase. However, Lcn2 at low pH retains the siderophores in its unfolded form in the solution phase.

2.4.5 Global dynamics: ligand-induced Lcn2 unfolding

To examine the possibility of a shifting folding/unfolding equilibrium by ligand association as the second event in ITC, the compactness of the Lcn2 structure was investigated by IM-MS arrival time distribution (ATD). The ATD revealed a single conformation for the 10+ charge state. A ligand-bound complex showed an increase in collisional cross-section with a longer drift time, indicating a reduction in protein compactness (Figure 2.14). Due to the direct increase in mass, a positive shift in the CCS alone cannot definitively prove unfolding of the protein.

In terms of the overall hydrogen bond network intactness, all complexes exhibited a disruption marked by higher deuterium uptake. The increase of the deuterium level possibly came from a ligand-induced conformational switch (Figure 2.15). Although the enthalpy usually drops at the binding site via protein-ligand interaction, the averaged dynamics cross the whole protein could rise. This increase in structural flexibility has previously been shown as a plausible theory with a modified heme ligand.[103]

Possibly because of the higher rotational and translational freedom ($S_{r/t}$) presented in the catecholate units, Ent without iron has the greatest level of distortion in the protein structure. Siderophores with lower degrees of freedom, 2,3-DHBA-Fe and Ent are locked

in their shapes by the iron cage. Due to the semi-rigid ferric siderophore conformation, the introduction of such siderophores disrupts the Lcn2 structure less.

Collectively, the ATD in IM-MS and global HDX rate changes suggest a shift in Lcn2 folding/unfolding equilibrium towards the less structured state. Therefore, unfolding is the perfect candidate for the second event in a biphasic ITC. However, the overall changes in protein structure provide limited information on specific changes occurring at the calyx. TR HDX experiments are employed to study these local changes in protein structure.

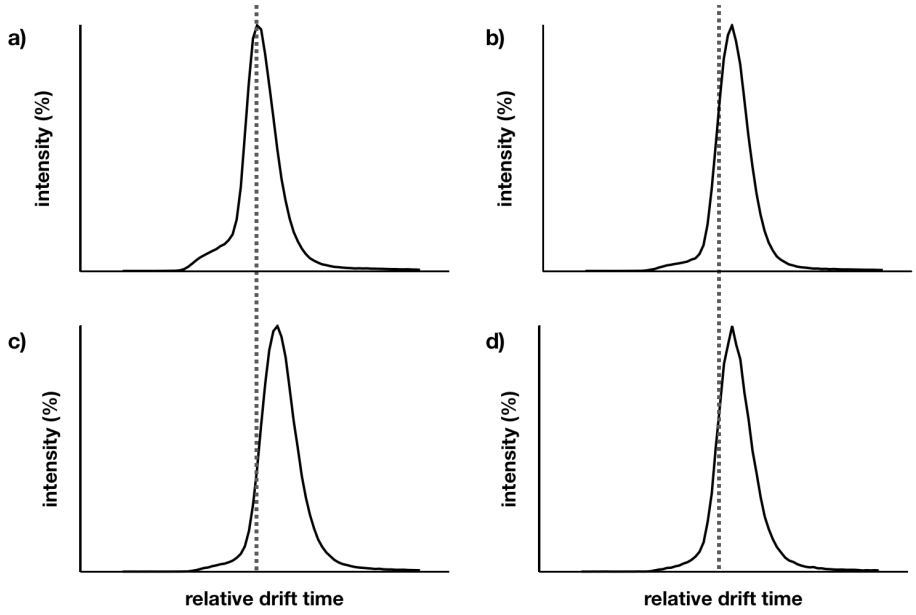


Figure 2.14. ATD of Lcn2 with different siderophores.

(a) Apo-Lcn2 has a single conformation for the 10+ charge state. Extended drift times were observed in (b) Lcn2-Ent, (c) Lcn2-Ent-Fe, and (d) Lcn2-2,3-DHBA-Fe. A reference dashed line represents the centroid of apo-Lcn2.

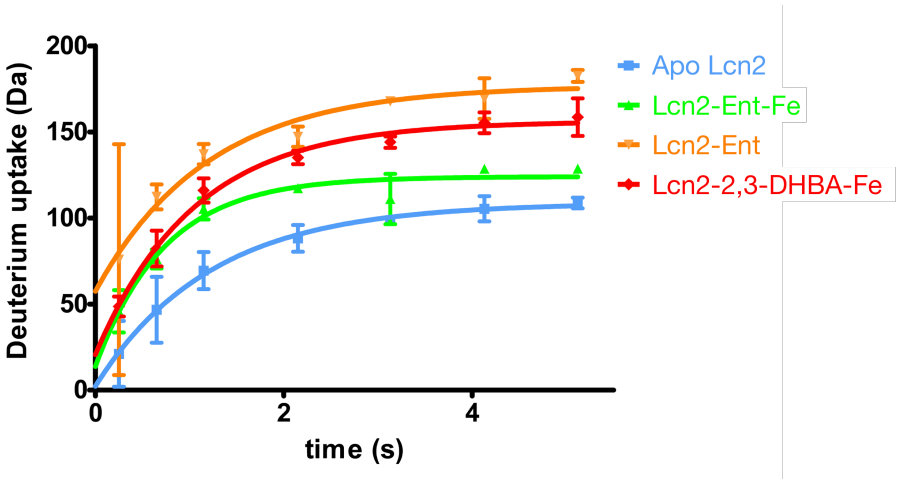


Figure 2.15. Global HDX showing variability in conformational dynamics between complexes.

The global changes in deuterium uptake levels were measured at seven different reaction times from 0.25 sec to 5.12 sec. The standard deviation was computed from technical (n>3) and biological (n≥3) replicates.

2.4.6 Proteolytic digestion optimization

Due to the acid stability of Lcn2, proteolytic digestion optimization is critical for adequate peptide coverage. Four different protease combinations, 100% pepsin, 100% protease XIII, 80:20 pepsin:protease XIII, and 50:50 pepsin:protease XIII, were tested. The digestion profiles are shown in Figure 2.16. Because the measured deuterium exchange level was averaged cross the selected peptide, the shorter peptides were preferred to improve the spatial resolution. The lengthy peptide from 100% protease XIII digestion was less ideal (Figure 2.16b). The mixture of both proteases at 50:50 gave optimal digestion with the highest peptide coverage at 64% (Figure 2.16d).

Other than the protease combination, additives in the quenching acid can also help with the digestion efficiency. For example, an organic solvent can help unfold the protein in addition to acidic conditions alone. Addition of 2% MeOH, or 5% ACN is tested for this

purpose. Regarding both unfolding and digestion efficiency, 5% ACN in 10% acetic acid was preferred (Figure 2.17). Eventually, the sequence coverage was increased to 73%.

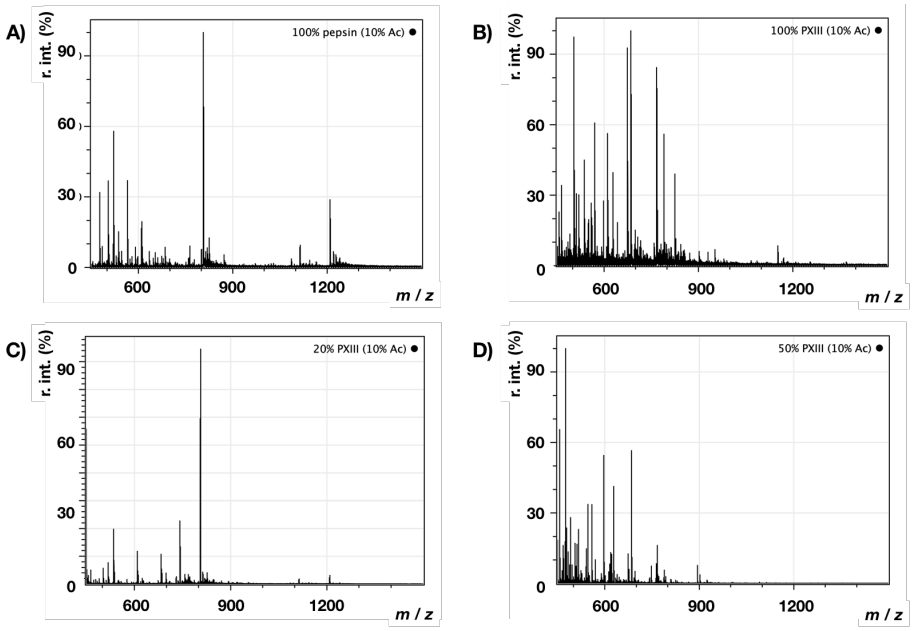


Figure 2.16. Optimization of digestion efficiency with four different ratios of proteases.

The ratio of pepsin and protease XIII immobilized agarose beads was altered in the protease chamber. (a) 100% pepsin preferably cleaved off the His-tag, leaving the rest of the protein poorly digested. (b) 100% protease XIII improved the digestion efficiency, but many peptides generated were more than 15 amino acids. (c) Pepsin:protease XIII at a 8:2 ratio provided no advantage on digestion compared to 100% pepsin. (d) Pepsin: protease XIII at a 5:5 ratio was the optimal choice with a suitable peptide amount and length.

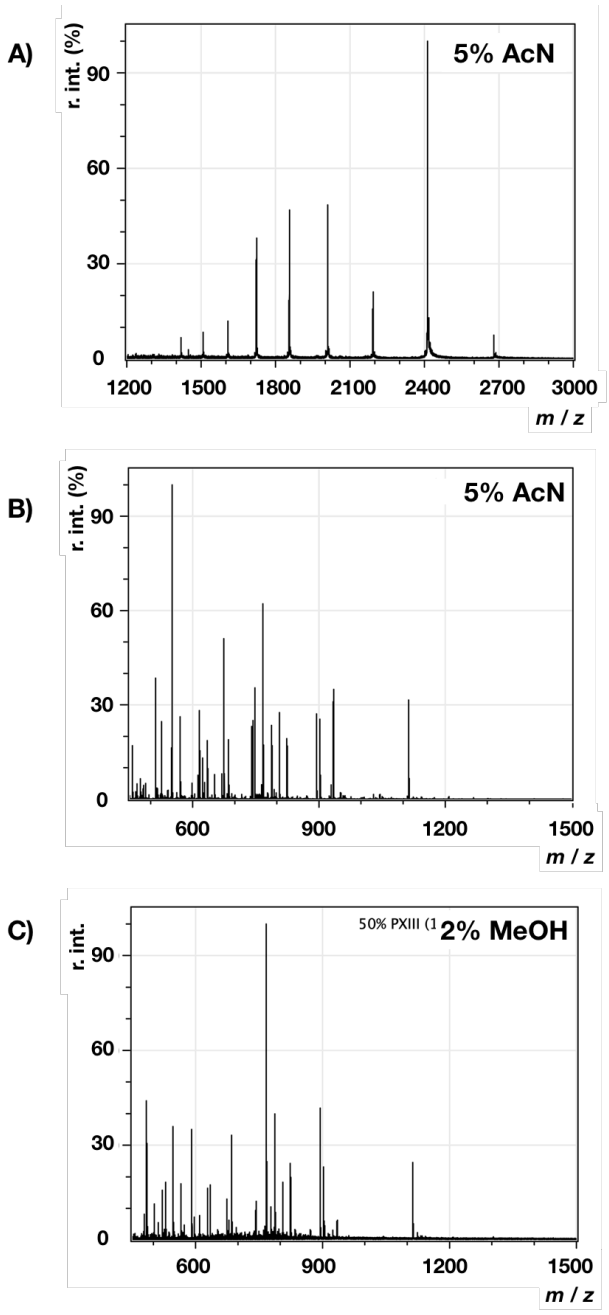


Figure 2.17. Optimization of digestion efficiency via additives in quenching buffer. With 10% acetic acid as the quenching solution, (a) 5% ACN partially unfolded Lcn2 in a continuous-flow set-up. The digested profile by adding (b) 5% ACN had a more suitable peptide diversity and length than (c) 2% MeOH as an additive in the quenching acid.

2.4.7 Regions of ligand-induced unfolding revealed by TRESI HDX MS

The disrupted region on Lcn2 was mapped by TR HDX, a structural technique that is sensitive in probing subtle differences between apo-Lcn2 and Lcn2-siderophore complexes. The protein was quenched and digested prior to MS analysis, so that the deuterium exchanged information was locked to a specific peptide. Difference plots, subtracting the complexes' deuterium uptake with apo-Lcn2, showed overall increases of percentage deuterium (and thus increase in dynamics) in red. Lcn2-Ent revealed a wide destabilization cross the whole protein, but the largest distortion was localized around the binding sites predicted from the human homolog of Lcn2 (Figure 2.18a). A similar region of the β barrel was disturbed in Lcn2-Ent-Fe as well (Figure 2.18b). The association with ferric 2,3-DHBA displayed a similar but weaker calyx destabilization. The protective effect at the 3_{10} helix was undoubtedly more pronounced in Lcn2-2,3-DHBA-Fe (Figure 2.18c).

Lower levels of destabilization for Lcn2-2,3-DHBA-Fe conflicts with the global HDX results, which suggests that this complex causes the second most pronounced disruption in the Lcn2 structure. However, the true deuterium uptake might be higher than what is shown in the TR HDX difference plot. Based on the native ESI spectra from section 2.4.3, the formation of Lcn2-2,3-DHBA-Fe complex is unsaturated, leaving a substantial amount of apo-Lcn2 in the system. The differences are therefore reduced when TR HDX monitors the averaged percentage deuterium level of peptides from both species. Ent and ferric Ent form more stable complexes with Lcn2 and are skewed towards saturation, so the apparent percentage of deuterium uptake is inevitably higher. In this case, global HDX provides a more accurate comparison of destabilization between complexes, while TR HDX can localize the regions of change in protein structure.

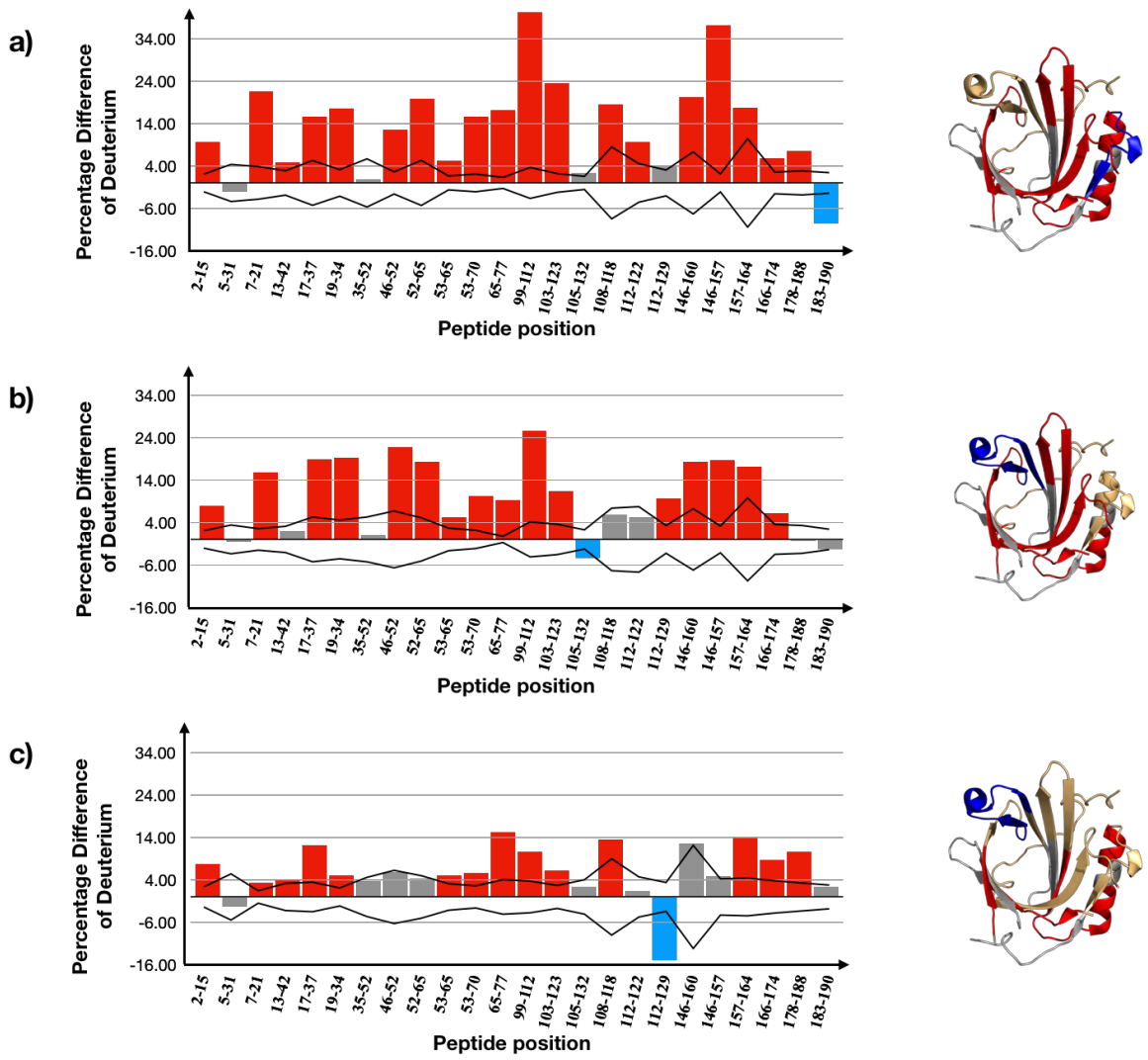


Figure 2.18. The difference plots of TR HDX results of multiple Lcn2-siderophore complexes.

The differences between (a) Lcn2-Ent, (b) Lcn2-Ent-Fe, and (c) Lcn2-2,3-DHBA-Fe and apo-Lcn2 were averaged from technical replicates (n>3) and summed for biological replicates (n=3). The propagated errors of standard deviation were represented by solid back lines. Only differences above the errors were identified as significant increases (red) or decreases (blue). For easier visualization, the corresponding changes were mapped onto a crystal structure (PDB ID: 3S26) on the right. The non-significant regions were coloured in gold while grey areas had no sequence coverage.

2.5 Conclusion

The siderocalin-siderophore interaction displays two-event binding, involving a second thermodynamic change other than the initial ligand association. The presence of various complexes is verified by native ESI MS and intrinsic fluorescence quenching. Ent-Fe, Ent and 2,3-DHBA-Fe are identified as in vitro binding partners. After eliminating the possibility of a secondary binding site and dimerization, a ligand-induced protein unfolding event is most likely the second thermodynamic change. The unfolding event is further confirmed with IM-MS ATD and global HDX results.

In-depth analysis via TR HDX reveals that the distortion of the protein structure spreads across the whole β-barrel for all complexes at different levels. Lcn2 unfolds to several conformers depending on the siderophore, showing greater levels of deuterium uptake for Lcn2-Ent to lower levels for Lcn2-Ent-Fe and Lcn2-2,3-DHBA-Fe. Between both Ent complexes, Lcn2-Ent is easier to unfold or dissociate in the gas phase. Between both ferric ligands, Lcn2-2,3-DHBA-Fe is easier to unfold or dissociate in the gas phase. A cell-based assay also supports that the easier-dissociated and more unfolded 2,3-DHBA-Fe-Lcn2 releases its siderophore faster for higher cytotoxicity (Appendix 1, unpublished results from collaborator Gary Sweeney).

Chapter Three. Studying the aggregation potential of Huntingtin N terminal fragment mutants

3.1 Summary

Huntington's disease (HD) is an inherited genetic disorder caused by mutations in the Huntingtin (Htt) protein. The length of the polyglutamine (polyQ) stretch dictates the age of onset and the extent to which the individual experiences symptoms. Although the exact routine function of Htt is unclear, normal Htt interacts with many other proteins and plays a pivotal role in early development.[104,105] The limited knowledge regarding its pathology suggests that mutant Htt is cleaved into many polyQ-containing N terminal fragments, which aggregate into amyloid fibrils and inclusion bodies in the brain.[106]

This study focuses on Htt exon 1 (Httex1), the shortest fragment found in vivo, as a model to study the structural transition of the N terminal fragments passing their pathological threshold.[107] Because of the sharp change in cytotoxicity when exceeding 39 polyQ, the conformational shift with longer polyQs is thought to contribute to this pathological switch. In many pre-existing expression systems, Httex1 is expressed together with a solubilisation tag. The first step would be obtaining a more native-like conformation via tag removal. Limited analysis has been performed on Httex1 with a short polyQ length, but the fast aggregation rate of mutated Httex1 hinders further analysis.

3.2 Introduction

3.2.1 Huntington's disease and Huntingtin (Htt)

While as equally devastating as other neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease, Huntington's disease (HD) is much less recognizable due to its low prevalence worldwide. Based on 2016 statistics, 5.96 to 13.7 per 100,000 people are affected by HD in different regions.[108] The low but clustered

prevalence can be explained by its inherited pathology. The autosomal dominant mutation of Huntingtin (Htt) was identified to be the cause of Huntington's disease in 1993.[109,110] Based on Mendel's autosomal dominant pattern of inheritance, the descendant of an affected parent has 50% probability of displaying symptoms, but the likelihood of developing HD without any family history is low.[111,112]

The Htt gene contains variable lengths of CAG repeats (Table 1). As the most major prognosis factor, CAG repeats higher than 39 are predicted to show symptoms. The general population has CAG repeats lower than 35, with a mean length of 26.[109,113,114] A link has been established between polyQ length and age of onset as well; the longer the repeats, the earlier the onset (Table 1).[115] Most patients first reveal symptoms in middle age after having children, but early onset HD before 20 years old is observed as well (juvenile HD). In the most extreme case known, the patient started presenting clinical motor manifestation at 18 months.[116] No effective treatment is currently available, but genetic testing can provide information regarding the likelihood of HD onset based on CAG repeat length.

Table 1: A link between CAG repeat length and HD pathology.

The CAG repeat length determines the poly-glutamine expansion in the Huntingtin protein. Based on large surveys and databases, a relationship between phenotype and polyQ length has been established.

PolyQ length	Diagnosis	Phenotype
< 35	normal	no clinical symptoms
35-39	carrier	may or may not have symptoms
39-75	HD	middle age onset (mean age at 41)
> 75	juvenile HD	early onset under age of 20

HD patients suffer from both cognitive and motor dysfunctions mainly due to the selective death of brain cells. Neuronal cell death at the basal ganglia causes brain mass loss and the characteristic erratic movement of HD. As HD progresses, loss of motor control inevitably results in the inability to speak and swallow.[117,118] Unfortunately, with this incurable condition, HD patients resemble healthy individuals pre-symptomatically but usually die 15-20 years afterward. The common causes of death are pneumonia and heart disease since patients have difficulty synchronizing movements.[119,120]

3.2.2 Poly glutamine (polyQ) expansion of Htt

Unusual for a protein of this size, full-length huntingtin protein is 300 kDa but completely soluble intracellularly. Htt consists of a polyglutamine (polyQ) stretch with variable lengths starting from the eighteenth amino acid from the N terminus, followed are the proline-rich region (PRR) and multiple HEAT repeats. The first seventy amino acids (N17), polyQ and PRR belong to the Htt exon 1 (Httex1).[121]

Directly coded by CAG repeats, the polyQ expansion has always been the focus of research. However, the polyQ motif is not unique to Htt but is present in many other disease-causing and essential human proteins.[122,123,124] Htt, in its normal or mutant form, interacts with over 100 proteins, many via the polyQ region.[125] For example, mutant Htt has been shown to interact with CREB binding protein (CBP) and decrease transcription levels.[126,127] Normal functions of Htt are suggested to be in the early development stages based on evidence that both homogenous (-/-) and heterogeneous (-/+) Htt knockout mice die in the embryonic stage.[104,105] However, neither the pathological mechanism nor normal pathway is clear for Htt.

3.2.3 Htt N terminal fragmentation and Htt exon 1 aggregation

Despite the little known about Htt, its cleavage-prone property is commonly accepted.[128,129] Upon protease cleavage, Htt generates many N terminal fragments containing a polyQ region. Eventually, those fragments form amyloid-like fibrils and inclusion bodies.[106] Studies show that the aggregation of N terminal fragments is polyQ length dependent, but the length of the fragment is irrelevant.[87,130] Implied, only N17 and polyQ are involved in the aggregation potential of various mutants. Indeed, N17 changes the subcellular localization of N terminal fragments and modulates aggregation through post-translational modification and its transient helical property.[131,132,133] The structural differences between pathological and mutated Htt from simple CAG repeats are believed to hold the answer to the sharp change in cytotoxicity.

Figure 3.1. A schematic representation of the Httex1 constructs.

A solubilisation tag is attached at the N terminus with a linker, containing a EK specific cleavage site. The exon 1 gene has three parts, N terminal 17 amino acid (Nt17), polyQ, and a proline-rich region (PRR). The His tag at the C terminus is used for IMAC purification.

PolyQ length has been observed to change the helical tendency of the N17 region and might regulate the aggregation rate through N17.[134] The PRR region following polyQ has no significant function in the normal Htt protein, so the aggregation is less likely to be controlled by this region.[135] Exon 1 (N17, polyQ, and PRR) is the smallest fragment observed in vivo, and is readily available as a bacterial construct.[107] Httex1 is convenient

as a model of the N terminal fragment, however, almost all purification systems for Httex1 produces a tagged protein for intercellular solubilization purposes. The first step of this study will be generating tag-free proteins for more native-like structures. The specific constructs used in this work are listed in Figure 3.1.

3.3 Experimental Methods

3.3.1 Thioredoxin (TrxA) tagged Htt exon 1 induction and expression

Plasmids with various polyglutamine lengths (pET32a-HD16Q, pET32a-HD25Q, pET32a-HD39Q, pET32a-HD46Q) were purchased from Addgene. All plasmids were transformed into BL21 cells following the same protocol described in section 2.3.1.1. Expression was performed on small volume (100 ml to 200 ml) cultures and pooled together into a single pellet for downstream purification. Cultures were grown at 37˚C for 2 hrs to reach OD_{600} 0.6. Bacteria were induced with 0.7mM IPTG at 18˚C O/N. Induced bacteria were pelleted at 6,000 xg for 25min. The pellet was lysed by either osmotic shock (OS) or sonication. The OS procedure was adopted from the protocol developed in the Bjorkman lab. Briefly, ice cold equilibrium buffer (15 mM Tris, 20% sucrose, pH 8.0) with a ¼ volume of the culture was used to resuspend the pellet by gentle swirling and rocking at 4˚C for 30 min. The pellet was collected again at 15,000 xg for 30 sec at 4˚C. The second resuspension used OS buffer (15 mM Tris, pH 8.0). A 60 min incubation at 4˚C with rocking was followed by centrifugation at 15,000 xg for 10min at 4˚C to extract peripheral proteins. The supernatant was collected. Sonication was performed with the parameters listed in section 2.3.1.2, except 30% amplitude was used.

3.3.2 Ni-IMAC purification

The first nickel IMAC purification of TrxA-tagged Httex1 used a similar protocol as described in section 2.3.1.2 with some minor modifications. The same phosphate based buffer system was used, but the washing buffers had concentrations of 20mM, 50mM, and 100mM imidazole respectively. Phosphate buffer with 500mM imidazole was used to elute a relatively concentrated product, to avoid any further concentration steps. Excess imidazole was eliminated by dialyzing into enterokinase (EK) cleavage buffer (50mM NaH_2PO_4, 50mM NaCl, 1mM EDTA pH 6.0). 1 mM EDTA was added to the protein prior to dialysis to help prevent aggregation.

3.3.3 TrxA tag cleavage

For HD16Q, the lowest aggregation potential variant, the TrxA tag cleavage was performed at room temperature (RT) without agitation O/N. Protein concentration was estimated by OD_{280} after dialysis. One IU of light-chain EK (stock concentration at 33 IU/ml) cleaved 50 µg of protein under these conditions. For HD46Q, the fastest aggregating variant, TrxA was cleaved by full-length EK for one hour at RT without agitation with the same protein:enzyme ratio. Tag-free Httex1 was immediately diluted to at least twice the volume with CIX loading buffer (50mM NaH_2PO_4, pH 6.0). The dilution slowed down the aggregation and lowered the NaCl concentration for downstream CIX.

3.3.4 Cation ion exchange (CIX) chromatography

EK-cleaved Httex1 was loaded onto a CIX gravity column pre-equilibrated with CIX loading buffer. The column was further washed with CIX wash buffer 1 (50mM NaH_2PO_4, 50mM NaCl, pH 6.0) and CIX wash buffer 2 (50mM NaH_2PO_4, 100mM NaCl, pH 6.0).

Purified tag-free Httex1 was eluted with 500mM NaCl (in 50mM NaH2PO4, pH 6.0). All purified tag-free Httex1 were pooled together from 1 ml elution fractions.

3.3.5 Htt exon 1 concentration and storage

CIX elution buffer was used as the storage buffer for Httex1. All TrxA-Httex1 protein after IMAC purification was visualized on 12.5% SDS-PAGE, while EK-cleaved Httex1 was identified on 15% SDS-PAGE. In both cases, 200 V was applied for 43 min. The concentration was estimated using the Bradford assay. Since Httex1 interacts with Bradford reagent poorly due to a lack of aromatic residues, results can be less accurate. Concentration was not performed because the tag-free sample would easily aggregate in the process.

3.3.6 ESI native MS

The final tag-free products were buffer exchanged using a HiTrap desalting column (5 ml, GE Healthcare). The column was equilibrated with 5 CV 500mM NH_4Ac (pH 6.8) before sample loading. The common loading and elution volumes are listed in Table 2. Protein at around 3μM flowed at a rate of 20 μl/min into the ESI source. The instrument (Waters Synapt G2-S) was operated in positive ion mode with a source voltage of 2.5 kV, temperature of 100˚C, and 25 V sampling cone. Mild collision was applied at 10 V and 6 V for Trap and Transfer cells respectively. The IMS cell operated at 650 m/s wave velocity and 40 V wave height.

Table 2: The standard loading and elution of a HiTrap desalting column.

A 5ml HiTrap column can only desalt a maximum of 1.5 ml sample at a time. With enough sample available, larger loading volume was always preferable to minimize the dilution.

Loading volume (ml)	Elution volume (ml)	Dilution factor
0.25	1.0	4
0.5	1.5	3
1.0	2.0	2
1.5	2.0	1.3

3.3.7 Proteolytic digestion

In preparation for TR HDX, different protease combinations (100% protease XIII, 100% pepsin, and 50:50 protease XIII:pepsin) were tested for best digestion results. All digestions were carried out in the protease chamber, with deuterium substituted with 500 mM NH_4Ac. The quenching acid was 2% formic acid with 1% MeOH. The flow rates were 3 µl/min for both protein and buffer lines. The acid flow rate was set to be 15µl/min. The digestion profile was optimized using HD16Q.

3.4 Results and Discussion

3.4.1 HD16Q purification

After O/N induction, the bacterial culture was pelleted for cell lysis. Suggested by the purification protocol from the Bjorkman group (the same lab that produced the pET32a-Httex1 constructs), the pellet was lysed with osmotic shock (OS).[136] Because TrxA is secreted to the periplasm of *E. coli* cell, the fusion protein can be released by osmotic shock. The yield of osmotic shock was low since most of the TrxA-HD16Q remained in the pellet (Figure 3.2). The cell lysis was improved by sonication with 30% amplitude.[137] The lysed protein mixture was loaded onto the IMAC column. The IMAC elution contained

many impurities that can be cleaned up further with a second column after the TrxA tag is cleaved (Figure 3.3a).

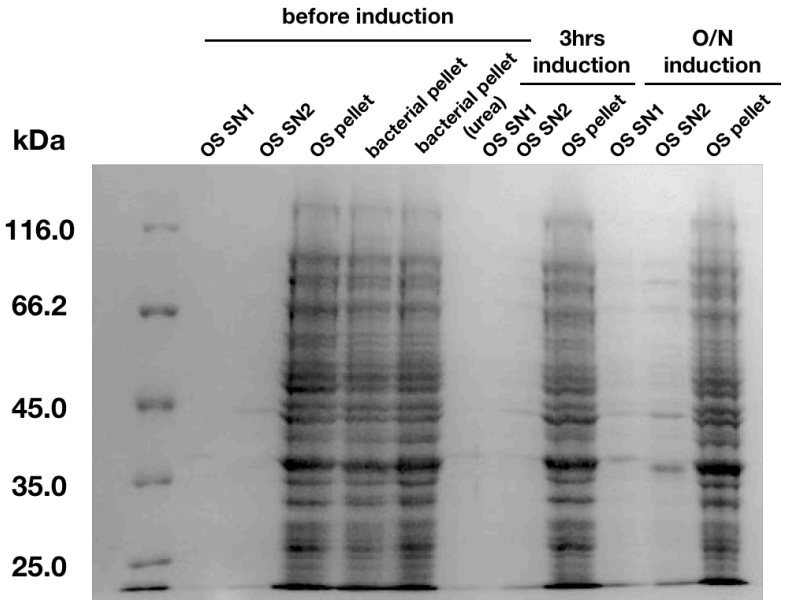


Figure 3.2. HD16Q pellet lysis via osmotic shock.

The supernatants (S/N) after incubating in equilibrium buffer or osmotic shock (OS) buffer were collected. The OS pellet was used to check the lysis efficiency. Bacterial pellet refers to the cultured pellet before lysis, dissolved in SDS-PAGE running buffer or 8M urea. The same OS protocol was used to extract peripheral proteins after 3 hrs 37˚C induction or O/N 18˚C induction. The apparent band at a molecular weight of around 35 kDa was the TrxA-HD16Q fusion, overlapping with *E. coli* proteins.

Enterokinase (EK), the protease cleaving at a specific site (Asp-Asp-Asp-Asp-Lys) just before Httex1, is a heterodimer with light and heavy chains. The light-chain is the catalytic subunit.[138] TrxA-HD16Q was cleaved with light-chain EK overnight at RT without agitation, resulting in around 70% HD16Q product (Figure 3.3b). Httex1 (pI 7.07) can be well separated from the other proteins in the mixture, TrxA-Httex1 (pI 5.25), TrxA (pI 4.31) and EK (pI 5.20), with cation exchange chromatography (CIX). Both TrxA and EK did not bind

to the column, but a small amount of uncleaved TrxA-HD16Q remained in the elution (Figure 3.3b).

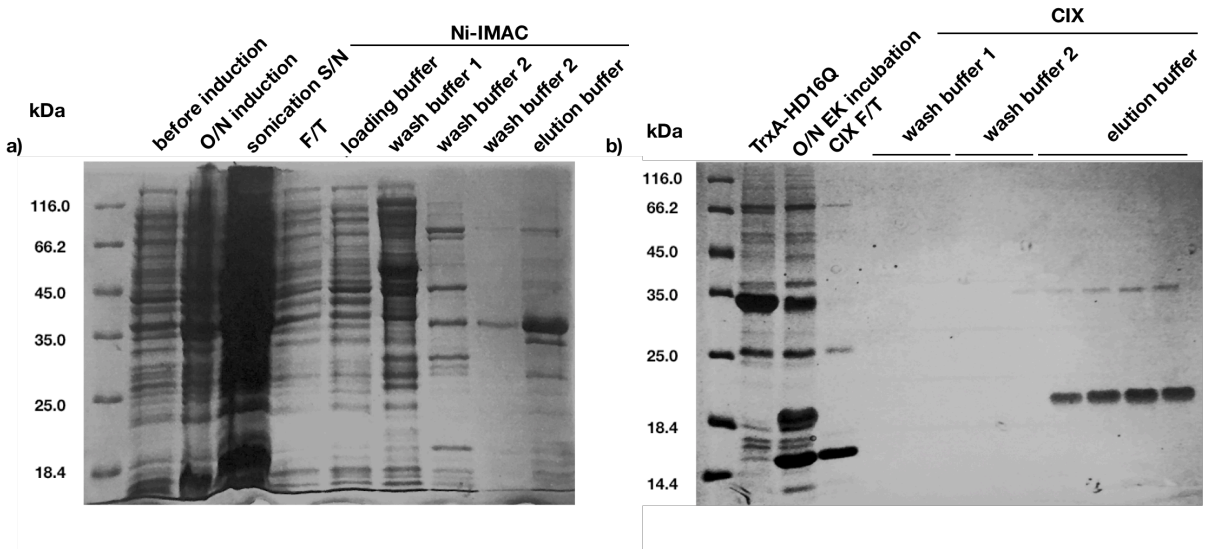


Figure 3.3. Ni-IMAC purification of HD16Q followed by tag removal.

HD16Q was first purified with (a) Ni-IMAC and the cleaved product was further cleaned up with (b) CIX. Ni-IMAC fractions were run on a 12.5% SDS-PAGE while CIX fractions were run on a 15% gel. Multiple elution fractions were collected for CIX, but only the first four were shown on this gel. The apparent molecular weight was around 20kDa for HD16Q (calculated mass of 10.5 kDa).

3.4.2 HD46Q optimization

The purification process is different for the constructs with longer polyglutamine expansions above the pathological threshold. With the solubilizing tag, TrxA-HD46Q was successfully purified by IMAC (Figure D3.4a). However, cleaved HD46Q has a much higher aggregation potential, resulting in a large volume of aggregates. The small amount of soluble protein leftover was not enough for CIX purification (Figure 3.4b). Further optimization of the tag removal time course was performed using only the light-chain EK. Light-chain EK had a much lower efficiency because it only cleaved around half of the

protein after 6 hrs of RT incubation (Figure 3.5). Longer incubation at RT gave rise to aggregates that prevented further cleavage. Incubation at 4˚C had even lower efficiency but slowed down the aggregation. This condition only provided around 30% product after O/N incubation, despite the absence of visible aggregates (Figure 3.5).

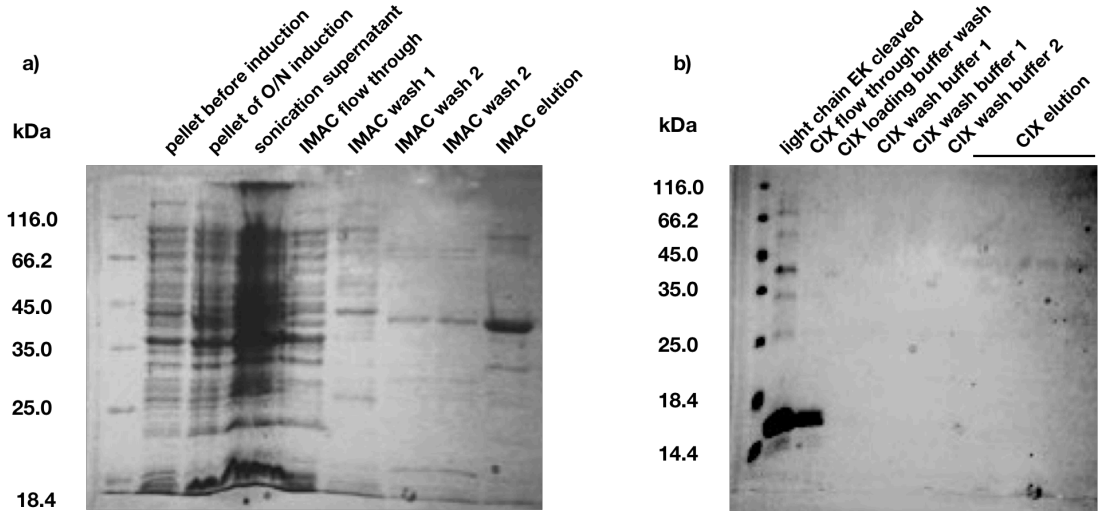


Figure 3.4. Same purification protocol of HD16Q on HD46Q.

(a) The results of Ni-IMAC were visualized on a 12.5% SDS-PAGE. TrxA-HD46Q was successfully purified at an apparent MW of 40 kDa. (b) All fractions of CIX were loaded on a 15% SDS-PAGE. Multiple elutions revealed no signs of HD46Q. The dominant band between MW 18.4 and 14.4 likely corresponds to the TrxA tag.

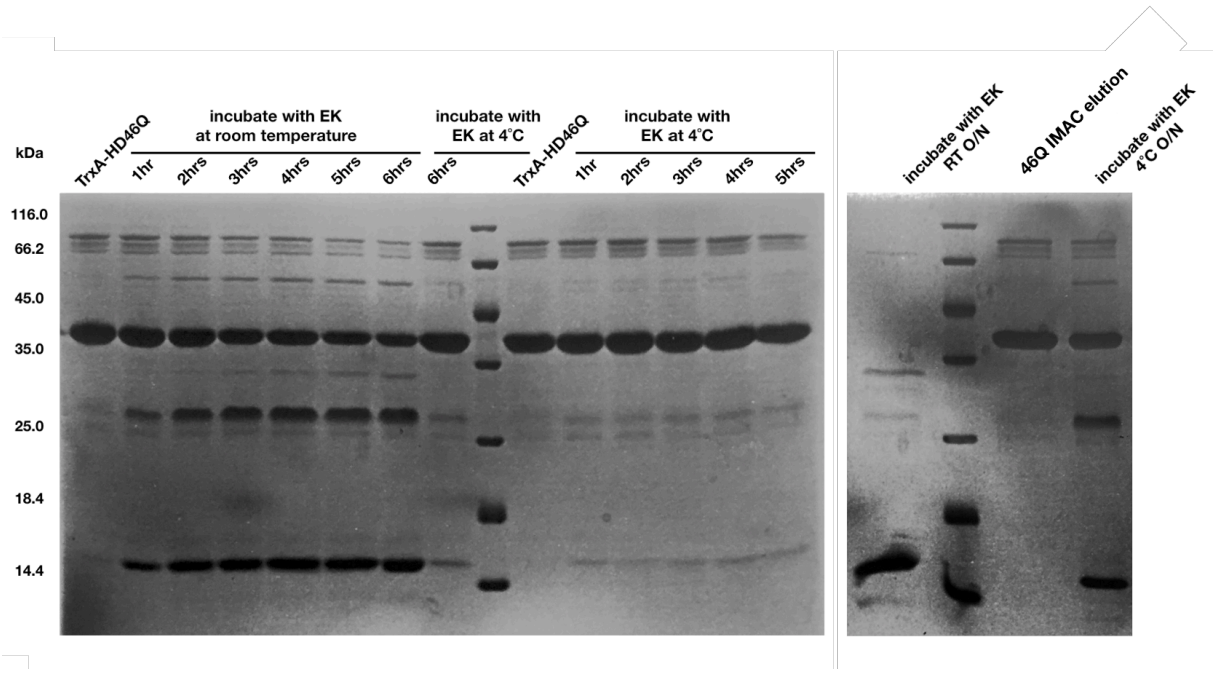


Figure 3.5. Time-course of light-chain enterokinase (EK) tag removal.

The incubation with EK was conducted at RT or 4˚C. Samples were collected every hour for 6 hrs. The enzymatic reaction was stopped by adding SDS-PAGE loading buffer and boiling for 5 min. The agitation of the O/N digestions were kept at a minimum.

Full-length EK was used instead to solve the problem of low efficiency. With the same protein to enzyme ratio, full-length EK cleaved TrxA-HD46Q to near completion in one hour at RT (Figure 3.6a). The tag-free product was loaded onto CIX, but it associated poorly with the column and was lost in the flow through. The final elution only contained very diluted protein together with some possibly uncut TrxA-HD46Q (Figure 3.6b). Another buffer system (50mM MES, pH 6.0) was tested for CIX as well. However, it possessed no advantages regarding EK efficiency or column affinity.

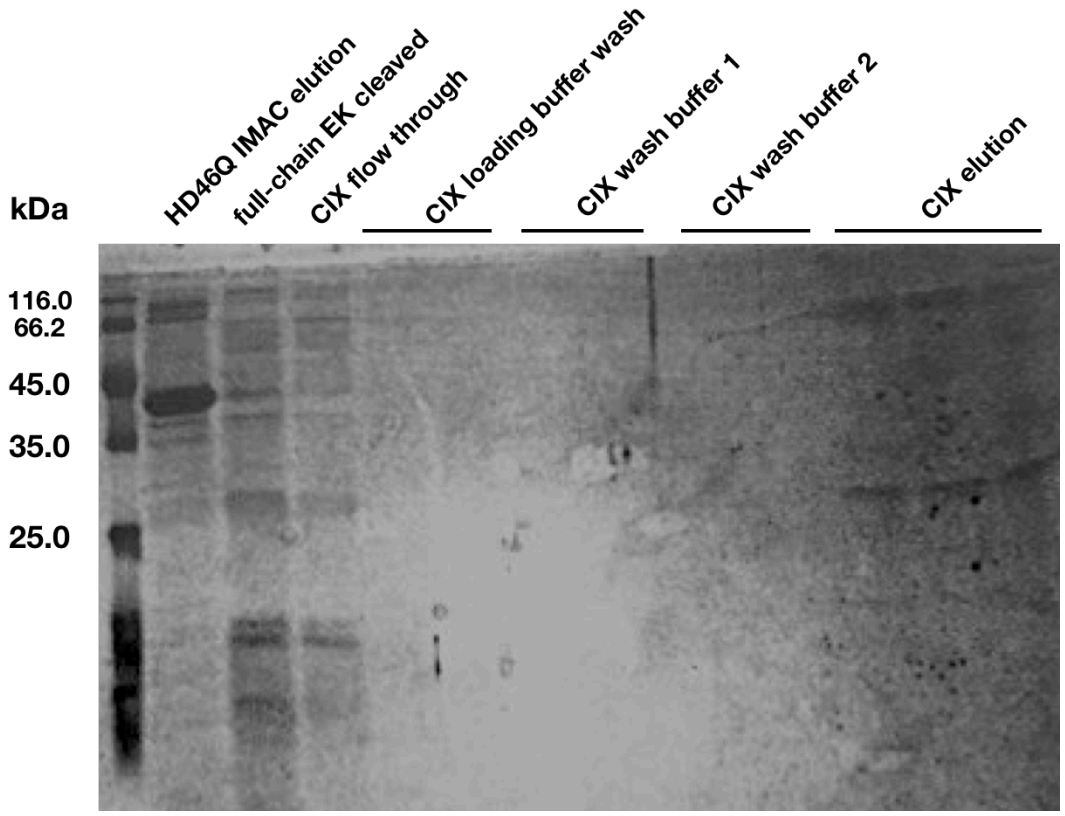


Figure 3.6. TrxA tag of HD46Q cleaved by full chain EK.

The same IMAC protocol was followed for TrxA-Htt46Q purification. After incubation with EK at RT, the aggregates were removed by centrifugation and the supernatant was loaded onto a CIX column. Fractions of CIX were collected and visualized using 15% SDS-PAGE. The apparent masses at 40 kDa and 25 kDa correspond to TrxA-HD46Q and HD46Q respectively.

3.4.3 ESI native MS

Only HD16Q and HD25Q were successfully purified and observed on native ESI MS. HD16Q showed two charge state envelopes with charge states ranging from 7+ to 14+ (Figure 3.7a). The deconvoluted mass constructed from all charge states was 10,305 Da, agreeing with the calculated mass of 10.5 kDa (Figure 3.7b). HD25Q had a similar but more distinct two charge state envelope pattern, with a deconvoluted mass of 11,346 Da

(11.7 kDa calculated) (Figure 3.7c-d). The wide range of charge states in both variants agreed with the unstructured nature of Httex1.

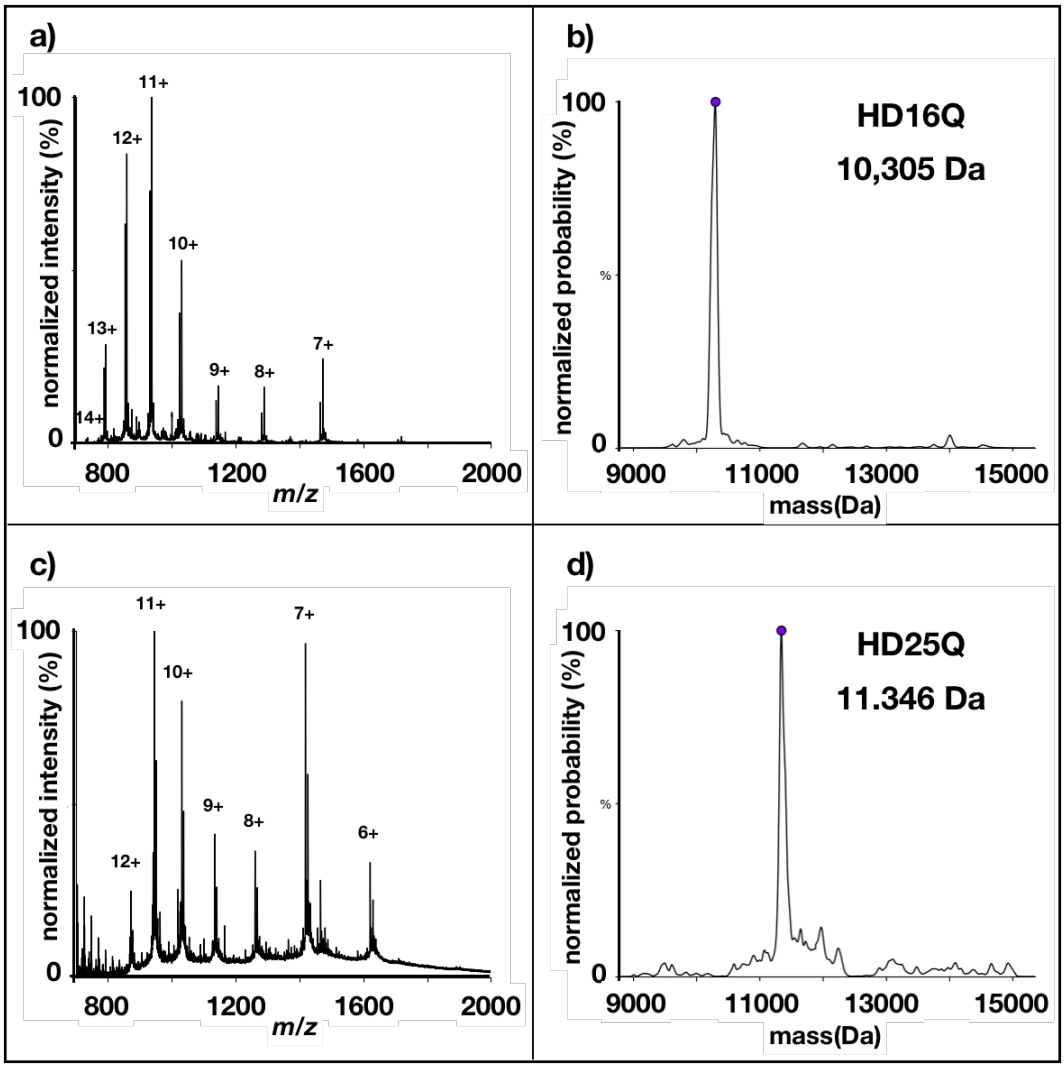


Figure 3.7. Native ESI MS spectra of Httex1 below the pathological length of polyQ. The (a) HD16Q and (c) HD25Q spectra were deconvoluted by UniDec. The peaks of deconvoluted (b) HD16Q and (d) HD25Q were labelled with purple dots.

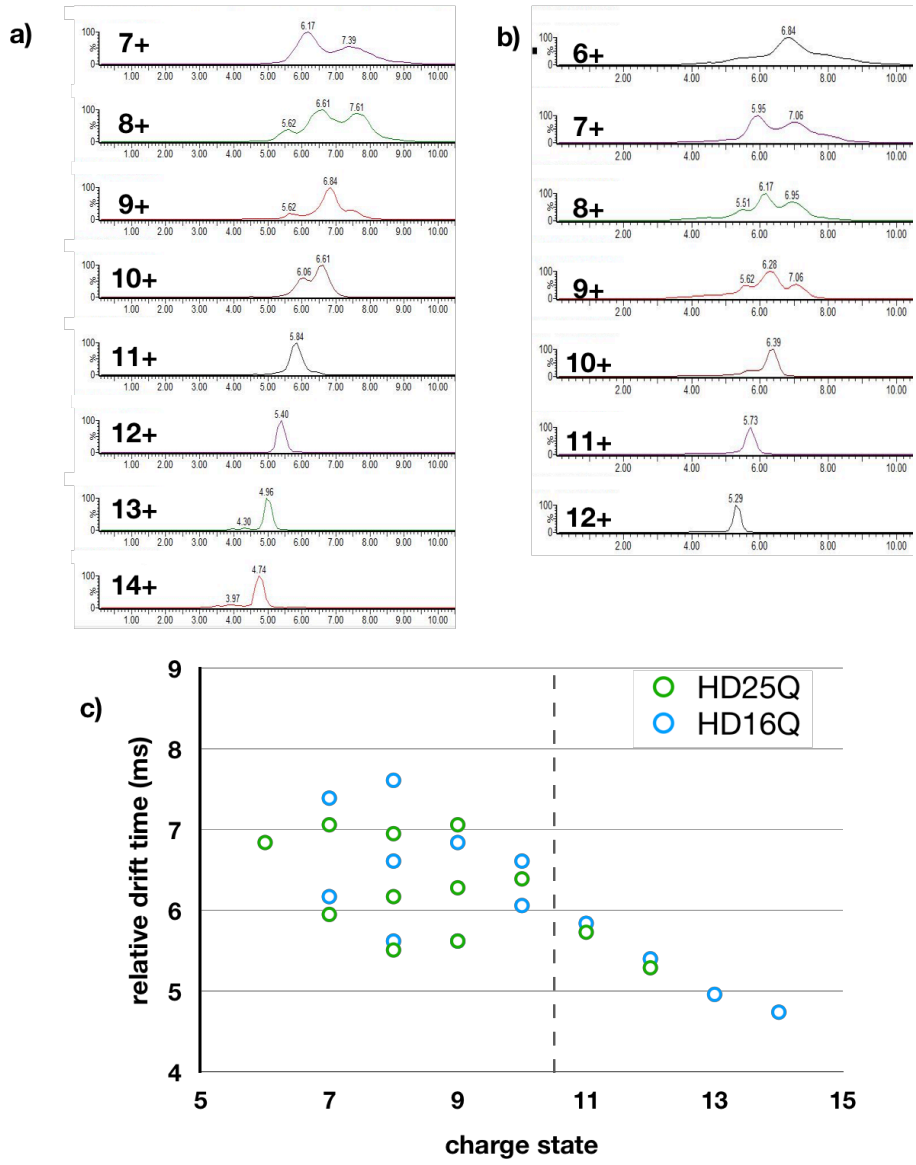


Figure 3.8. The relative drift time distribution of different charge states in HD16Q and HD25Q.

The extracted relative drift time for (a) HD16Q and (b) HD25Q were (c) plotted against the charge states. The dashed line separated the partially structured and unfolded charge states.

In both HD16Q and HD25Q, the ATD for each charge state was well resolved (Figure 3.8 a-b). For the unfolded envelope, spanning charges 11+ to 14+, the charge and drift time had a linear relationship (Figure 3.8c). In the structured envelope, each charge state

had multiple drift times (Figure 3.8c). The two charge envelopes and multiple conformers imply that even intrinsically disordered proteins have some partially folded structures. When comparing the partially folded species, HD25Q has a shorter drift time despite the larger MW. Therefore, the reduced drift time may result from a more compact structure of HD25Q.

3.4.4 Proteolytic digestion

To investigate the structure of the N17 region by HDX, efficient proteolytic digestion is essential. Three protease combinations were tested. Although pepsin digested Httex1 to a level where no intact protein was observed in the digestion profile, it only gave a collection of highly charged and lengthy peptides (Figure 3.9a). Both 50% and 100% protease XIII provided smaller peptides with a sacrifice of incomplete digestion (Figure 3.9b-c). The digestion profile from 100% protease XIII had a higher sequence coverage and more peptides in the N17 region (Figure 3.9d). Peptides were identified using the unique sequence of Httex1.

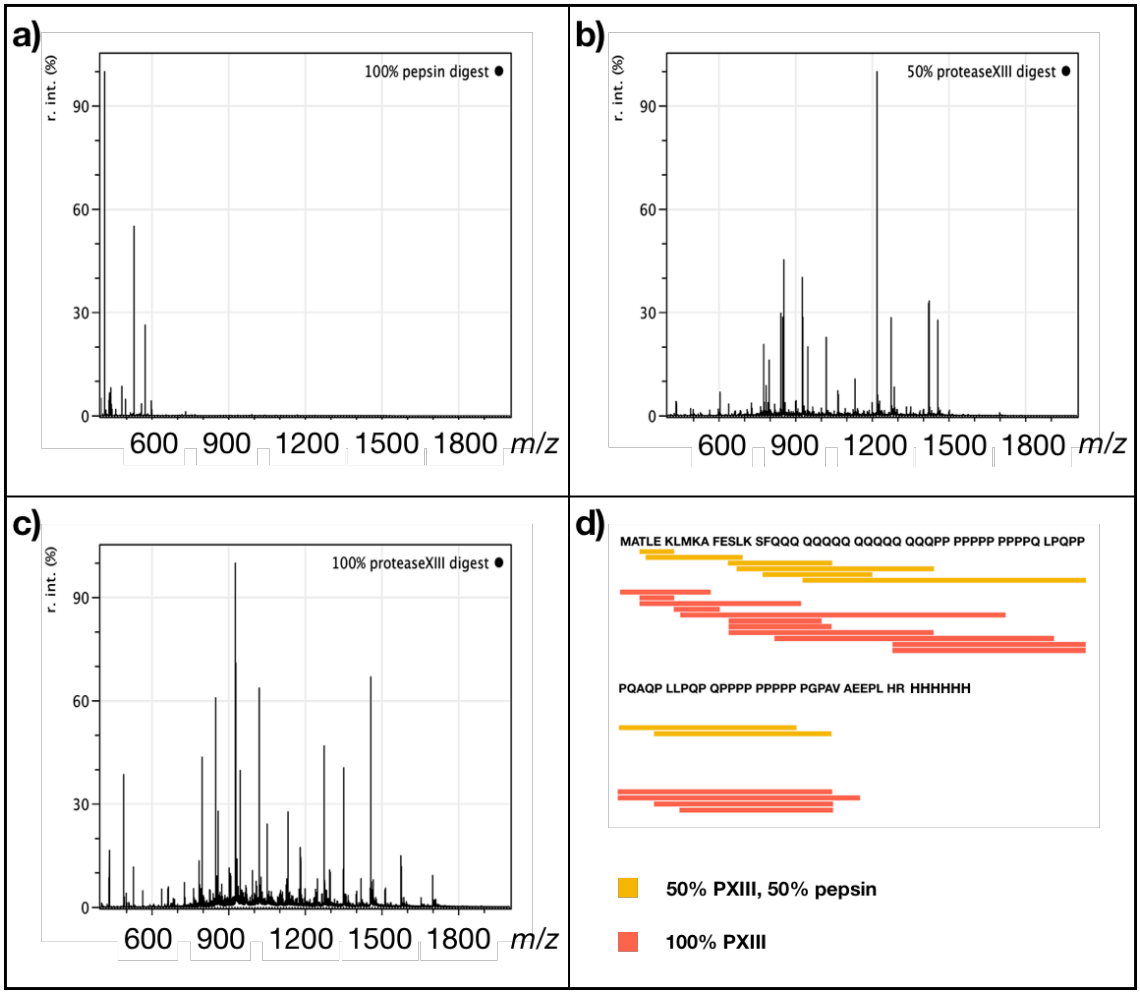


Figure 3.9. Digestion efficiency of various proteases on HD16Q.

The digested profiles of HD16Q were generated by (a) 100% pepsin, (b) 50:50 pepsin:protease XIII mixture, or (c) 100% protease XIII. For easier comparison, (d) peptides from both protease XIII digestions were colour mapped together. Each bar represents a peptide with a corresponding sequence aligned on top. The 50:50 pepsin:protease XIII mixture produced 83% sequence coverage, while 100% protease XIII gave a slightly higher sequence coverage of 89%.

3.5 Conclusion

The purification and tag removal of the Httex1 constructs with a short polyQ expansion was successful. The disordered structure of Httex1 is well-represented by a wide charge

distribution in native ESI MS. A more compact structure is associated with a longer polyQ (HD25Q) below the pathological length. However, the purification of tag-free Httex1 with longer polyQ lengths requires more optimization. Recently, Reif *et al*. have purified tag-free HD43Q. In their protocol, Httex1 is purified with reverse phase HPLC and lyophilized immediately after tag removal.[139] However, Httex1 has transient secondary structure despite being an intrinsically disordered protein. Httex1 loses its structural information after exposure to organic solvent and low pH in the reverse phase column. “Softer” methods are still required in order to study native Httex1 properties after tag removal and before visible aggregation.

The switch between asymptomatic and diseased phenotypes at a critical polyQ length remains a mystery. Although an acute change of structure after 39 glutamine residues has long been the hypothesis, Warner *et al.* showed that normal and mutant Httex1 has similar tadpole structures via fluorescence resonance energy transfer and simulations.[140] Assuming correct modelling, the physiological to pathological switch of Htt may come from the higher fragmentation tendency of longer polyQs. A higher N terminal fragment concentration from a more cleavage-prone mutant Htt unavoidably facilitates aggregation. Both theories have failed to explain the link between age onset and polyQ length. More studies on the molecular pathology of Htt function or dysfunction are needed before an effective therapy can be explored.

Chapter Four: Conclusion and Future Work

4.1 Conclusion

This work applied various MS methods to study protein-ligand and protein-protein interactions. In chapter 2, the complexes of Lcn2 with three of its ligands, ferric 2,3-DHBA, ferric Ent and Ent, were investigated. The results showed that Lcn2 is able to discriminate between ligands and form complexes with different stabilities, although all ligands ultimately unfold the protein. The variations between complexes may distinguish the intracellular releasing mechanism of Lcn2. The MS-based analysis provides structural evidence for different intracellular effects and releasing rates indirectly. In chapter 3, the self-interaction of polyQ proteins leads to aggregation. Httex1 constructs with a polyQ above pathological length were shown to aggregate faster at similar or even lower concentrations compared to its physiological variants. Conceivably, the longer polyQ changes the protein structure or the way Httex1 interacts with each other.

4.2 Future Work

Chapter 2 studies the binding of siderophores to Lcn2, but the releasing mechanism of these ligands is puzzling. It is commonly believed that the low pH (6.5-4.5) conditions of the endosome and lysosome protonates siderophores and promotes their release.[141] However, we observe a majority of folded protein and intact complexes with Lcn2 and siderophores at pH 2.5. This raises questions regarding the intracellular release of siderophores. Since Lcn2 is proposed to recycle, Lcn2 cannot be destroyed by the lysosome if this theory held true.[70] In turn, the transfer of iron to another molecule is more likely. Heme proteins, iron-binding proteins, and vitamin C can be used to probe this transfer.[142] With the incubation of natural intracellular iron chelators with iron-loaded Lcn2

at cytosol or endosomal pH, the transfer of iron can be monitored by native ESI MS if present. Understanding this iron-releasing mechanism can be important in resolving the transportation of excess iron into mitochondria for ROS accumulation.

The most pronounced challenge encountered in Chapter 3 is that Httex1 with longer polyQs form aggregates before any meaningful analysis could be performed. To address that, reverse phase chromatography and lyophilisation are utilized to prevent aggregation. However, the aggregation will continue once Httex1 is in solution again. A better method would be to perform online cleavage of the tag and monitor the aggregation immediately. The enzymatic activity of EK is much weaker but still functional in NH4Ac (Appendix 3). Native mass spectra can be acquired at different time points of the incubation. Due to the distinct masses of various species (EK, TrxA, TrxA-Httex1, and Httex1), the tag-free Httex1 should be easily identified. Utilizing the power of IMS, the monomer and oligomers can be separated and monitored as well. As a diploid species, Httex1 almost always exists in heterogeneity in humans. A more interesting question would be whether a normal Httex1 could slow down the aggregation of a mutant. A similar online monitoring system with a mixture of species might be able to provide this information.

Bibliography

1. Woodward, C., Simon, I. & Tüchsen, E. Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem.* **48**, 135–160 (1982).
2. Brown, K. A. & Wilson, D. J. Bottom-up hydrogen deuterium exchange mass spectrometry: data analysis and interpretation. *Analyst* **142**, 2874–2886 (2017).
3. Resing, K. A., Hoofnagle, A. N. & Ahn, N. G. Modeling deuterium exchange behavior of ERK2 using pepsin mapping to probe secondary structure. *J. Am. Soc. Mass Spectrom.* **10**, 685–702 (1999).
4. Winters, M. S., Spellman, D. S. & Lambris, J. D. Solvent Accessibility of Native and Hydrolyzed Human Complement Protein 3 Analyzed by Hydrogen/Deuterium Exchange and Mass Spectrometry. *J. Immunol.* **174**, 3469–3474 (2005).
5. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
6. Eigen, M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES - Eigen - 1964 - Angewandte Chemie International Edition in English - Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.196400011#. (Accessed: 3rd March 2019)
7. Matthew, J. B. & Richards, F. M. The pH dependence of hydrogen exchange in proteins. *J. Biol. Chem.* **258**, 3039–3044 (1983).
8. Eigen, M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angew. Chem. Int. Ed. Engl.* **3**, 1–19 (1964).

9. Morgan, C. R. & Engen, J. R. Investigating solution-phase protein structure and dynamics by hydrogen exchange mass spectrometry. *Curr. Protoc. Protein Sci.* **Chapter 17**, Unit 17.6.1-17 (2009).
10. Oganesyan, I., Lento, C. & Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* **144**, 27–42 (2018).
11. Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M. & Ten Eyck, L. F. Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *J. Am. Soc. Mass Spectrom.* **17**, 1498–1509 (2006).
12. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1814**, 942–968 (2011).
13. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
14. Xiao, H., Kaltashov, I. A. & Eyles, S. J. Indirect assessment of small hydrophobic ligand binding to a model protein using a combination of ESI MS and HDX/ESI MS. *J. Am. Soc. Mass Spectrom.* **14**, 506–515 (2003).
15. West, G. M. *et al.* Ligand-Dependent Perturbation of the Conformational Ensemble for the GPCR β2 Adrenergic Receptor Revealed by HDX. *Structure* **19**, 1424–1432 (2011).
16. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M. & Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions. *Expert Rev. Proteomics* **8**, 43–59 (2011).

17. Dharmasiri, K. & Smith, D. L. Mass Spectrometric Determination of Isotopic Exchange Rates of Amide Hydrogens Located on the Surfaces of Proteins. *Anal. Chem.* **68**, 2340–2344 (1996).
18. Simmons, D. A., Dunn, S. D. & Konermann, L. Conformational Dynamics of Partially Denatured Myoglobin Studied by Time-Resolved Electrospray Mass Spectrometry with Online Hydrogen−Deuterium Exchange,. *Biochemistry* **42**, 5896–5905 (2003).
19. Rob, T. & Wilson, D. J. A Versatile Microfluidic Chip for Millisecond Time-Scale Kinetic Studies by Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **20**, 124–130 (2009).
20. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill Jr., H. H. Ion mobility–mass spectrometry. *J. Mass Spectrom.* **43**, 1–22 (2008).
21. Okhonin, V., Petrov, A. P., Berezovski, M. & Krylov, S. N. Plug−Plug Kinetic Capillary Electrophoresis: Method for Direct Determination of Rate Constants of Complex Formation and Dissociation. *Anal. Chem.* **78**, 4803–4810 (2006).
22. Du, X. *et al.* Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**, (2016).
23. Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J. Recept. Signal Transduct. Res.* **24**, 1–52 (2004).
24. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Dtsch. Chem. Ges.* **27**, 2985–2993 (1894).
25. Jorgensen, W. L. Rusting of the lock and key model for protein-ligand binding. *Science* **254**, 954–955 (1991).

26. Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chem. Int. Ed. Engl.* **33**, 2375–2378 (1995).
27. Liu, S.-Q. *et al.* Protein Folding, Binding and Energy Landscape: A Synthesis. in *Protein Engineering* (ed. Kaumaya, P.) (InTech, 2012). doi:10.5772/30440
28. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35**, 539–546 (2010).
29. Jäger, M. *et al.* Structure–function–folding relationship in a WW domain. *Proc. Natl. Acad. Sci.* **103**, 10648–10653 (2006).
30. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci.* **92**, 452–456 (1995).
31. Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501–1509 (1985).
32. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680–1685 (1993).
33. Wilkinson, B. & Gilbert, H. F. Protein disulfide isomerase. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1699**, 35–44 (2004).
34. Göthel, S. F. & Marahiel, M. A. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell. Mol. Life Sci. CMLS* **55**, 423–436 (1999).
35. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571 (1996).
36. Wang, X. *et al.* Hsp90 Cochaperone Aha1 Downregulation Rescues Misfolding of CFTR in Cystic Fibrosis. *Cell* **127**, 803–815 (2006).

37. Doyle, S. M., Genest, O. & Wickner, S. Protein rescue from aggregates by powerful molecular chaperone machines. *Nat. Rev. Mol. Cell Biol.* **14**, 617–629 (2013).
38. Eisele, F. & Wolf, D. H. Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143–4146 (2008).
39. Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**, 3–16 (2004).
40. Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
41. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533 (2002).
42. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **37**, 509–516 (2012).
43. Groot, N. S. de *et al.* Evolutionary selection for protein aggregation. *Biochem. Soc. Trans.* **40**, 1032–1037 (2012).
44. Patel, L., Abate, C. & Curran, T. Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**, 572–575 (1990).
45. Frieden, C. Protein aggregation processes: In search of the mechanism. *Protein Sci.* **16**, 2334–2344 (2007).
46. Miti, T., Mulaj, M., Schmit, J. D. & Muschol, M. Stable, Metastable, and Kinetically Trapped Amyloid Aggregate Phases. *Biomacromolecules* **16**, 326–335 (2015).
47. Malinovska, L., Kroschwald, S. & Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1834**, 918–931 (2013).

48. Quigley, E. M. M. Gut Bacteria in Health and Disease. *Gastroenterol. Hepatol.* **9**, 560–569 (2013).
49. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, (2016).
50. Shanahan, F. The host-microbe interface within the gut. *Best Pract. Res. Clin. Gastroenterol.* **16**, 915–931 (2002).
51. Cebra, J. J. Influences of microbiota on intestinal immune system development. *Am. J. Clin. Nutr.* **69**, 1046s–1051s (1999).
52. Zhang, T., Yang, Y., Liang, Y., Jiao, X. & Zhao, C. Beneficial Effect of Intestinal Fermentation of Natural Polysaccharides. *Nutrients* **10**, (2018).
53. Conly, J. M. & Stein, K. The production of menaquinones (vitamin K2) by intestinal bacteria and their role in maintaining coagulation homeostasis. *Prog. Food Nutr. Sci.* **16**, 307–343 (1992).
54. Corr, S. C. *et al.* Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7617–7621 (2007).
55. Rea, M. C. *et al.* Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9352–9357 (2010).
56. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions Between the Microbiota and the Immune System. *Science* **336**, 1268–1273 (2012).
57. Neilands, J. B. Siderophores: Structure and Function of Microbial Iron Transport Compounds. *J. Biol. Chem.* **270**, 26723–26726 (1995).

58. Faraldo-Gómez, J. D. & Sansom, M. S. P. Acquisition of siderophores in Gram-negative bacteria. *Nat. Rev. Mol. Cell Biol.* **4**, 105–116 (2003).
59. Thulasiraman, P. *et al.* Selectivity of Ferric Enterobactin Binding and Cooperativity of Transport in Gram-Negative Bacteria. *J. Bacteriol.* **180**, 6689–6696 (1998).
60. Khan, A., Singh, P. & Srivastava, A. Synthesis, nature and utility of universal iron chelator – Siderophore: A review. *Microbiol. Res.* **212**–**213**, 103–111 (2018).
61. Raymond, K. N., Müller, G. & Matzanke, B. F. Complexation of iron by siderophores a review of their solution and structural chemistry and biological function. in *Structural Chemistry* 49–102 (Springer Berlin Heidelberg, 1984).
62. Pollack, J. R. & Neilands, J. B. Enterobactin, an iron transport compound from Salmonella typhimurium. *Biochem. Biophys. Res. Commun.* **38**, 989–992 (1970).
63. Gehring, A. M., Bradley, K. A. & Walsh, C. T. Enterobactin Biosynthesis in Escherichia coli: Isochorismate Lyase (EntB) Is a Bifunctional Enzyme That Is Phosphopantetheinylated by EntD and Then Acylated by EntE Using ATP and 2,3-Dihydroxybenzoate. *Biochemistry* **36**, 8495–8503 (1997).
64. Reichert, J., Sakaitani, M. & Walsh, C. T. Characterization of EntF as a serine-activating enzyme. *Protein Sci. Publ. Protein Soc.* **1**, 549–556 (1992).
65. Johnstone, T. C. & Nolan, E. M. Determination of the Molecular Structures of Ferric Enterobactin and Ferric Enantioenterobactin Using Racemic Crystallography. *J. Am. Chem. Soc.* **139**, 15245–15250 (2017).
66. Berner, I., Greiner, M., Metzger, J., Jung, G. & Winkelmann, G. Identification of enterobactin and linear dihydroxybenzoylserine compounds by HPLC and ion spray mass spectrometry (LC/MS and MS/MS). *Biol. Met.* **4**, 113–118 (1991).

67. Song, E. *et al.* Holo-lipocalin-2-derived siderophores increase mitochondrial ROS and impair oxidative phosphorylation in rat cardiomyocytes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1576–1581 (2018).
68. Annamalai, R., Jin, B., Cao, Z., Newton, S. M. C. & Klebba, P. E. Recognition of Ferric Catecholates by FepA. *J. Bacteriol.* **186**, 3578–3589 (2004).
69. Abergel, R. J. *et al.* The Siderocalin/Enterobactin Interaction: A Link between Mammalian Immunity and Bacterial Iron Transport. *J. Am. Chem. Soc.* **130**, 11524–11534 (2008).
70. Richardson, D. R. 24p3 and Its Receptor: Dawn of a New Iron Age? *Cell* **123**, 1175–1177 (2005).
71. Qi, B. & Han, M. Microbial Siderophore Enterobactin Promotes Mitochondrial Iron Uptake and Development of the Host via Interaction with ATP Synthase. *Cell* **175**, 571-582.e11 (2018).
72. Chu, S. T., Lin, H. J., Huang, H. L. & Chen, Y. H. The hydrophobic pocket of 24p3 protein from mouse uterine luminal fluid: fatty acid and retinol binding activity and predicted structural similarity to lipocalins. *J. Pept. Res. Off. J. Am. Pept. Soc.* **52**, 390–397 (1998).
73. Goetz, D. H. *et al.* Ligand preference inferred from the structure of neutrophil gelatinase associated lipocalin. *Biochemistry* **39**, 1935–1941 (2000).
74. Bister, B. *et al.* The structure of salmochelins: C-glucosylated enterobactins of Salmonella enterica. *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* **17**, 471–481 (2004).

75. Golonka, R., Yeoh, B. S. & Vijay-Kumar, M. The Iron Tug-of-War between Bacterial Siderophores and Innate Immunity. *J. Innate Immun.* 1–14 (2019). doi:10.1159/000494627

76. Abella, V. *et al.* The potential of lipocalin-2/NGAL as biomarker for inflammatory and metabolic diseases. *Biomarkers* **20**, 565–571 (2015).

77. Moschen, A. R., Adolph, T. E., Gerner, R. R., Wieser, V. & Tilg, H. Lipocalin-2: A Master Mediator of Intestinal and Metabolic Inflammation. *Trends Endocrinol. Metab. TEM* **28**, 388–397 (2017).

78. Jang, Y., Lee, J. H., Wang, Y. & Sweeney, G. Emerging clinical and experimental evidence for the role of lipocalin-2 in metabolic syndrome. *Clin. Exp. Pharmacol. Physiol.* **39**, 194–199 (2012).

79. Banjara, M. Lipocalin-2: a New Regulator of Non-Pathogen-Associated Neuroinflammation. *Int. J. Clin. Exp. Neurol.* **2**, 8–15 (2014).

80. Toyonaga, T. *et al.* Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages. *Sci. Rep.* **6**, 35014 (2016).

81. Yan, Q.-W. *et al.* The adipokine lipocalin 2 is regulated by obesity and promotes insulin resistance. *Diabetes* **56**, 2533–2540 (2007).

82. Naudé, P. J. W. *et al.* Lipocalin 2: novel component of proinflammatory signaling in Alzheimer's disease. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **26**, 2811–2823 (2012).

83. Pereira, S. S. & Alvarez-Leite, J. I. Low-Grade Inflammation, Obesity, and Diabetes. *Curr. Obes. Rep.* **3**, 422–431 (2014).

84. Zhao, P., Elks, C. M. & Stephens, J. M. The induction of lipocalin-2 protein expression in vivo and in vitro. *J. Biol. Chem.* **289**, 5960–5969 (2014).
85. Liu, Z. *et al.* Regulation of mammalian siderophore 2,5-DHBA in the innate immune response to infection. *J. Exp. Med.* **211**, 1197–1213 (2014).
86. Kim, D. H. Bacterial Siderophores Promote Animal Host Iron Acquisition and Growth. *Cell* **175**, 311–312 (2018).
87. Latunde-Dada, G. O. Ferroptosis: Role of lipid peroxidation, iron and ferritinophagy. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 1893–1900 (2017).
88. Schieber, M. & Chandel, N. S. ROS Function in Redox Signaling and Oxidative Stress. *Curr. Biol. CB* **24**, R453–R462 (2014).
89. Hiraishi, H. *et al.* Role for iron in reactive oxygen species-mediated cytotoxicity to cultured rat gastric mucosal cells. *Am. J. Physiol.* **260**, G556-563 (1991).
90. Arosio, P., Elia, L. & Poli, M. Ferritin, cellular iron storage and regulation. *IUBMB Life* **69**, 414–422 (2017).
91. Bystrom, L. M., Guzman, M. L. & Rivella, S. Iron and Reactive Oxygen Species: Friends or Foes of Cancer Cells? *Antioxid. Redox Signal.* **20**, 1917–1924 (2014).
92. Correnti, C. *et al.* Siderocalin/Lcn2/NGAL/24p3 does not drive apoptosis through gentisic acid mediated iron withdrawal in hematopoietic cell lines. *PloS One* **7**, e43696 (2012).
93. Freiburger, L. A., Auclair, K. & Mittermaier, A. K. Elucidating Protein Binding Mechanisms by Variable-c ITC. *ChemBioChem* **10**, 2871–2873 (2009).

94. Knox, R., Lento, C. & Wilson, D. J. Mapping Conformational Dynamics to Individual Steps in the TEM-1 β-Lactamase Catalytic Mechanism. *J. Mol. Biol.* **430**, 3311–3322 (2018).

95. Rey, M. *et al.* Mass spec studio for integrative structural biology. *Struct. Lond. Engl. 1993* **22**, 1538–1548 (2014).

96. Brown, A. Analysis of cooperativity by isothermal titration calorimetry. *Int. J. Mol. Sci.* **10**, 3457–3477 (2009).

97. Brautigam, C. A. Fitting two- and three-site binding models to isothermal titration calorimetric data. *Methods San Diego Calif* **76**, 124–136 (2015).

98. Kovriguine, E. Resolving Three-State Ligand-Binding Mechanisms By Isothermal Titration Calorimetry: A Simulation Study. *BioRxiv Beta* (2017).

99. Shrake, A. & Ross, P. D. Ligand-induced biphasic protein denaturation. *J. Biol. Chem.* **265**, 5055–5059 (1990).

100. Seo, M.-H., Park, J., Kim, E., Hohng, S. & Kim, H.-S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **5**, 3724 (2014).

101. Strong, R. K. *et al.* Expression, purification, crystallization and crystallographic characterization of dimeric and monomeric human neutrophil gelatinase associated lipocalin (NGAL). *Acta Crystallogr. D Biol. Crystallogr.* **54**, 93–95 (1998).

102. Cai, L., Rubin, J., Han, W., Venge, P. & Xu, S. The origin of multiple molecular forms in urine of HNL/NGAL. *Clin. J. Am. Soc. Nephrol. CJASN* **5**, 2229–2235 (2010).

103. Konermann, L., Rodriguez, A. D. & Sowole, M. A. Type 1 and Type 2 scenarios in hydrogen exchange mass spectrometry studies on protein–ligand complexes. *Analyst* **139**, 6078–6087 (2014).

104. Nasir, J. *et al.* Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell* **81**, 811–823 (1995).

105. Zeitlin, S., Liu, J. P., Chapman, D. L., Papaioannou, V. E. & Efstratiadis, A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat. Genet.* **11**, 155–163 (1995).

106. Petrakis, S., Schaefer, M. H., Wanker, E. E. & Andrade-Navarro, M. A. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *Bioessays* **35**, 503–507 (2013).

107. Landles, C. *et al.* Proteolysis of Mutant Huntingtin Produces an Exon 1 Fragment That Accumulates as an Aggregated Protein in Neuronal Nuclei in Huntington Disease. *J. Biol. Chem.* **285**, 8808–8823 (2010).

108. Baig, S. S., Strong, M. & Quarrell, O. W. The global prevalence of Huntington's disease: a systematic review and discussion. *Neurodegener. Dis. Manag.* **6**, 331–343 (2016).

109. Nørremølle, A. *et al.* Trinucleotide repeat elongation in the Huntingtin gene in Huntington Disease patients from 71 Danish families. *Hum. Mol. Genet.* **2**, 1475–1476 (1993).

110. Everett, C. M. & Wood, N. W. Trinucleotide repeats and neurodegenerative disease. *Brain J. Neurol.* **127**, 2385–2405 (2004).

111. Walker, F. O. Huntington's disease. *The Lancet* **369**, 218–228 (2007).

112. Zheng, Z. & Diamond, M. I. Chapter 6 - Huntington Disease and the Huntingtin Protein. in *Progress in Molecular Biology and Translational Science* (ed. Teplow, D. B.) **107**, 189–214 (Academic Press, 2012).

113. Brinkman, R. R., Mezei, M. M., Theilmann, J., Almqvist, E. & Hayden, M. R. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am. J. Hum. Genet.* **60**, 1202–1210 (1997).

114. Rubinsztein, D. C. *et al.* Phenotypic Characterization of Individuals with 30–40 CAG Repeats in the Huntington Disease (HD) Gene Reveals HD Cases with 36 Repeats and Apparently Normal Elderly Individuals with 36–39 Repeats. *Am. J. Hum. Genet.* **59**, 16–22 (1996).

115. Andresen, J. M. *et al.* The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. *Ann. Hum. Genet.* **71**, 295–301 (2007).

116. Nicolas, G. *et al.* Juvenile Huntington disease in an 18-month-old boy revealed by global developmental delay and reduced cerebellar volume. *Am. J. Med. Genet. A.* **155A**, 815–818 (2011).

117. Estrada Sánchez, A. M., Mejía-Toiber, J. & Massieu, L. Excitotoxic Neuronal Death and the Pathogenesis of Huntington's Disease. *Arch. Med. Res.* **39**, 265–276 (2008).

118. Jakel, R. J. & Maragos, W. F. Neuronal cell death in Huntington's disease: a potential role for dopamine. *Trends Neurosci.* **23**, 239–245 (2000).

119. Heemskerk, A.-W. & Roos, R. A. C. Aspiration pneumonia and death in Huntington's disease. *PLoS Curr.* **4**, (2012).

120. Haines, J. L. & Conneally, P. M. Causes of death in Huntington disease as reported on death certificates. *Genet. Epidemiol.* **3**, 417–423 (1986).

121. Li, W., Serpell, L. C., Carter, W. J., Rubinsztein, D. C. & Huntington, J. A. Expression and Characterization of Full-length Human Huntingtin, an Elongated HEAT Repeat Protein. *J. Biol. Chem.* **281**, 15916–15922 (2006).

122. Schaefer, M. H., Wanker, E. E. & Andrade-Navarro, M. A. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res.* **40**, 4273–4287 (2012).

123. Fan, H.-C. *et al.* Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell Transplant.* **23**, 441–458 (2014).

124. Totzeck, F., Andrade-Navarro, M. A. & Mier, P. The Protein Structure Context of PolyQ Regions. *PLoS ONE* **12**, (2017).

125. Goehler, H. *et al.* A Protein Interaction Network Links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's Disease. *Mol. Cell* **15**, 853–865 (2004).

126. Steffan, J. S. *et al.* The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6763–6768 (2000).

127. Jiang, H., Nucifora, F. C., Ross, C. A. & DeFranco, D. B. Cell death triggered by polyglutamine-expanded huntingtin in a neuronal cell line is associated with degradation of CREB-binding protein. *Hum. Mol. Genet.* **12**, 1–12 (2003).

128. Lunkes, A. *et al.* Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions. *Mol. Cell* **10**, 259–269 (2002).

129. Wellington, C. L. *et al.* Caspase Cleavage of Mutant Huntingtin Precedes Neurodegeneration in Huntington's Disease. *J. Neurosci.* **22**, 7862–7872 (2002).

130. Fodale, V. *et al.* Polyglutamine- and Temperature-Dependent Conformational Rigidity in Mutant Huntingtin Revealed by Immunoassays and Circular Dichroism Spectroscopy. *PLOS ONE* **9**, e112262 (2014).

131. Deguire, S. M. *et al.* N-terminal Huntingtin (Htt) phosphorylation is a molecular switch regulating Htt aggregation, helical conformation, internalization, and nuclear targeting. *J. Biol. Chem.* jbc.RA118.004621 (2018). doi:10.1074/jbc.RA118.004621

132. Chaibva, M. *et al.* Acetylation within the First 17 Residues of Huntingtin Exon 1 Alters Aggregation and Lipid Binding. *Biophys. J.* **111**, 349–362 (2016).

133. Pandey, N. K. *et al.* The 17-residue-long N terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms. *J. Biol. Chem.* **293**, 2597–2605 (2018).

134. Bravo-Arredondo, J. M. *et al.* The Folding Equilibrium of Huntingtin Exon-1 Monomer Depends on its Polyglutamine Tract. *J. Biol. Chem.* jbc.RA118.004808 (2018). doi:10.1074/jbc.RA118.004808

135. Neveklovska, M., Clabough, E. B. D., Steffan, J. S. & Zeitlin, S. O. Deletion of the huntingtin proline-rich region does not significantly affect normal huntingtin function in mice. *J. Huntingt. Dis.* **1**, 71–87 (2012).

136. Bennett, M. J. *et al.* A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11634–11639 (2002).

137. Owens, G. E., New, D. M., West, A. P. & Bjorkman, P. J. Anti-PolyQ Antibodies Recognize a Short PolyQ Stretch in Both Normal and Mutant Huntingtin Exon 1. *J. Mol. Biol.* **427**, 2507–2519 (2015).

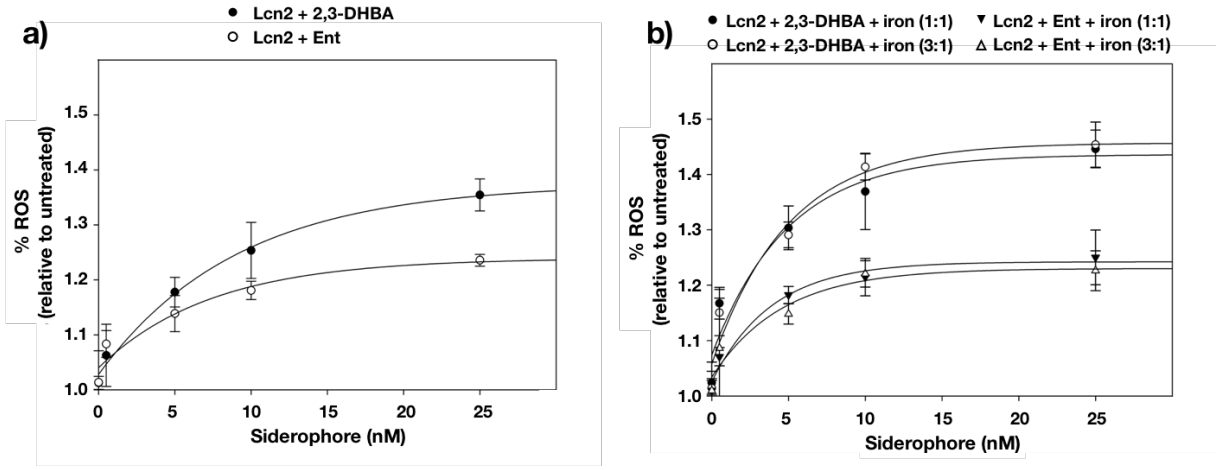
138. Gasparian, M. E., Bychkov, M. L., Dolgikh, D. A. & Kirpichnikov, M. P. Strategy for improvement of enteropeptidase efficiency in tag removal processes. *Protein Expr. Purif.* **79**, 191–196 (2011).

139. Reif, A., Chiki, A., Ricci, J. & Lashuel, H. A. Generation of Native, Untagged Huntingtin Exon1 Monomer and Fibrils Using a SUMO Fusion Strategy. *J. Vis. Exp. JoVE* (2018). doi:10.3791/57506

140. Warner, J. B. *et al.* Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. *J. Am. Chem. Soc.* **139**, 14456–14469 (2017).

141. Maxfield, F. R., Willard, J. M. & Lu, S. *Lysosomes: Biology, Diseases, and Therapeutics*. (John Wiley & Sons, 2016).

142. Hallberg, L., Brune, M. & Rossander, L. The role of vitamin C in iron absorption. *Int. J. Vitam. Nutr. Res. Suppl. Int. Z. Vitam.- Ernahrungsforschung Suppl.* **30**, 103–108 (1989).



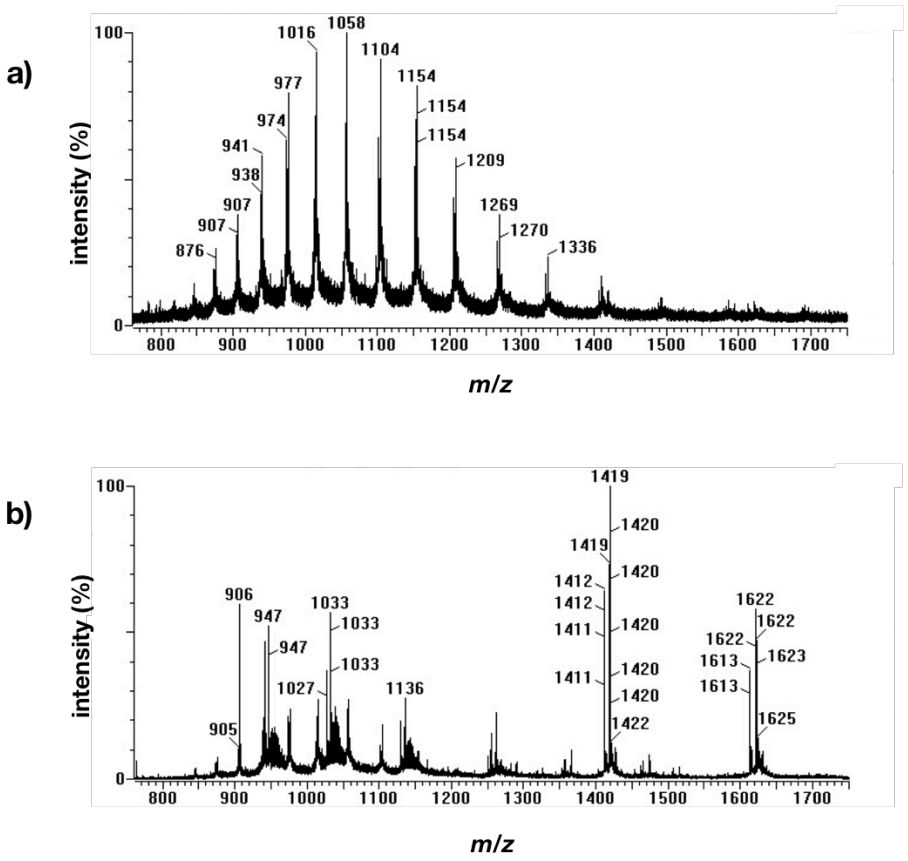
Appendix 1: Cell-based assay for intracellular ROS accumulation.

The accumulation of intracellular ROS was tested in H9C2 cells. (a) The ROS fold changes with limited iron in the complete growth media were monitored in a dose-dependent manner. (b) Supplementary Fe3+ at a 1:1 or 3:1 siderophore:iron ratio increased the ROS responses.

MSDKI IHLTD DSFDT DVLKA DGAIL VDFWA EWCGP CKMIA PILDE IADEY QGKLT VAKLN IDQNP GTAPK YGIRG IPTLL LFKNG EVAAT KVGAL SKGQL KEFLD ANLAG SGSGE RQHMD SPDLG T**DDDD K**
MATLE KLMKA FESLK SFQQQ QQQQQ QQQQQ QQQQQ QQQPP PPPPP PPPPQ LPQPP PQAQP LLPQP QPPPP PPPPP PGPAV AEEPL HR **HHHHHH**

Appendix 2: Sequence of TrxA-HD16Q.

The sequence of the TrxA tag is highlighted in orange. DDDKX is the recognition site for enterokinase, which cleaves immediately after K. The six histidines at the C terminus of the sequence is un-removable. TrxA-HD16Q is shown as an example because the only difference between the four constructs is the length of glutamine.



Appendix 3: Full-chain EK cleaved TrxA-HD16Q in 500mM NH_4Ac.

After 4hrs of incubation at RT with 1.5 IU/mg full-chain EK, (a) TrxA-HD16Q was partially digested into (b) HD16Q. The most intense peak in the in HD16Q was the 7+ charge state. Although more than one species was present in the cleaved product, the tag-free protein was clearly identifiable.