

Electronic Thesis and Dissertation Repository

---

10-28-2022 2:15 PM

# Conformational Dynamics and Aggregation of Thermally Stressed Proteins Studied by Hydrogen/Deuterium Exchange Mass Spectrometry

Nastaran Nosrat Tajoddin, *The University of Western Ontario*

Supervisor: Konermann, Lars, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Chemistry

© Nastaran Nosrat Tajoddin 2022

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Analytical Chemistry Commons](#), [Biochemistry Commons](#), and the [Biophysics Commons](#)

---

## Recommended Citation

Nosrat Tajoddin, Nastaran, "Conformational Dynamics and Aggregation of Thermally Stressed Proteins Studied by Hydrogen/Deuterium Exchange Mass Spectrometry" (2022). *Electronic Thesis and Dissertation Repository*. 8949.  
<https://ir.lib.uwo.ca/etd/8949>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

Proteins perform various biological functions, e.g., as enzymes or transporters. In addition to naturally occurring proteins, the use of protein therapeutic drugs for treating cancer and other diseases is a rapidly growing area. A thorough biophysical characterization of proteins and protein therapeutics opens the door to a more comprehensive understanding of their role in health and disease. This dissertation aims to expand the capabilities of an existing technique (Hydrogen Deuterium Exchange Mass Spectrometry, HDX-MS), which is widely used for probing protein structure and dynamics. Conventionally, HDX-MS experiments are performed as a function of labelling time. Here we aim to establish temperature as a complementary variable. Our goal was to unravel the interplay between thermally induced protein dynamic motions, unfolding, and aggregation.

Chapter 2 examined the effects of protein heating, using myoglobin (Mb) as model system. MS was used to track deuterium uptake in response to increasing temperature at various labelling time points. The resulting data were captured using a comprehensive temperature- and time-dependent HDX data analysis framework. The HDX trends were dissected into contributions from “chemical” labelling, as well as local and global protein dynamics. Experimental profiles started with shallow slopes and showed a sharp increase close to the melting temperature. Our analysis revealed that local dynamics dominate at low temperatures, while global events become prevalent closer to the melting point.

Chapter 3 studied the mechanism of thermally induced Mb aggregation. Upon heating, Mb produced amorphous aggregates. The extent of aggregation was measured by centrifugation and UV-Vis spectroscopy as a function of protein concentration, temperature, and time. From these data, we conclude that aggregation likely proceeds from globally unfolded proteins rather than from semi-unfolded species. The data obtained this way paved the way toward extensive molecular dynamics simulations of protein aggregation.

In Chapter 4, we tested the applicability of the thermodynamic framework developed in Chapter 2 to a monoclonal antibody (NISTmAb), representing a model system of a typical protein therapeutic. Differential scanning calorimetry revealed the presence of three successive melting points, reflecting the different stability of the CH2, CH3, and Fab regions. HDX-MS was performed to comprehensively characterize the conformational dynamics of NISTmAb as a function of time and temperature. Global analysis of the entire data set yielded insights into the enthalpic and entropic behavior of different segments. The unfolding of the Fab domain (which has the highest melting temperature) was found to be closely coupled to aggregation. In summary, we developed a method that provides in-depth information on the thermodynamic behavior of thermally stressed proteins based on HDX-MS experiments, and we demonstrated the applicability of this method to proteins of vastly different sizes and complexity.

## Keywords

protein dynamics | hydrogen-deuterium exchange | mass spectrometry | myoglobin | protein aggregation | protein thermal stability | monoclonal antibody | differential scanning calorimetry | circular dichroism | thermodynamics

## Summary for Lay Audience

Proteins are essential biological molecules in every living organism. Proteins also play an important role as therapeutics. For example, insulin has been used for decades in diabetic patients, and antibody drug conjugates are increasingly being used for treating various types of cancer. Unfavorable solvent conditions (such as heat and extremes of pH) can cause protein unfolding and aggregation. Protein aggregation is also a common characteristic of many neurodegenerative disorders such as Alzheimer's and Parkinson's diseases. Pharmaceutical tests of new protein drugs routinely use thermal unfolding and aggregation assays to assess the stability and shelf life of therapeutic proteins. However, there remains an urgent need to develop new and improved methods for monitoring protein thermal stability. With this in mind, the aim of the current thesis was to develop a better understanding of thermally stressed proteins.

In Chapter 2, myoglobin, a small globular protein was used as model system. Myoglobin was incubated in heavy water, and mass spectrometry was used to track deuterium uptake in response to increasing temperature. We developed a novel model to interpret the temperature- and time-dependent deuterium uptake data. Our analysis revealed that local structural fluctuations dominate the protein behavior at low temperatures, while global unfolding/refolding became prevalent close to the melting point.

Chapter 3 studied the mechanism of thermally induced myoglobin aggregation. The extent of aggregation was measured by centrifugation and UV-Vis spectroscopy as a function of protein concentration, temperature, and time. From these data, we concluded aggregation proceeds from totally unfolded proteins rather than from semi-folded intermediates.

In Chapter 4, we tested the applicability of the thermodynamic framework developed in Chapter 2 to a monoclonal antibody, representing a model system of a therapeutic protein. Differential scanning calorimetry revealed that this multi-domain protein has three melting points. Incubation of the antibody in heavy water at different temperatures and time points, followed by mass spectrometry analysis, was conducted to comprehensively characterize the conformational dynamics of the protein. Our data yielded insights into the

thermodynamic behavior of different protein segments. In summary, we developed a novel method that provides in-depth information on the thermodynamic behavior of thermally stressed proteins.

## Co-Authorship Statement

The work in Chapter 2 was published in the following article:

Tajoddin, N. N.; Konermann, L. Analysis of Temperature-Dependent H/D Exchange Mass Spectrometry Experiments. *Anal. Chem.* 2020, 92 (14), 10058-10067. DOI: 10.1021/acs.analchem.0c01828.

The work in Chapter 3 has been incorporated into the following article:

Ng, Y. K.; Tajoddin, N. N.; Scrosati, P. M.; Konermann, L. Mechanism of Thermal Protein Aggregation: Experiments and Molecular Dynamics Simulations on the High-Temperature Behavior of Myoglobin. *J. Phys. Chem. B* 2021, 125 (48), 13099-13110. DOI: 10.1021/acs.jpcc.1c07210.

The work in Chapter 4 was published in the following article:

Tajoddin, N. N.; Konermann, L. Structural Dynamics of a Thermally Stressed Monoclonal Antibody Characterized by Temperature-Dependent H/D Exchange Mass Spectrometry. *Anal. Chem.* 2022, DOI: 10.1021/acs.analchem.2c03931

The original draft for each of the above articles was prepared by the author. Subsequent revisions were performed by the author and Dr. Lars Konermann. The molecular dynamics simulation in Chapter 3 were performed by Yuen Ki Ng, with the help of Pablo M. Scrosati. Those simulations are *not* an integral part of this thesis. All other experimental work and data analysis were performed by the author under the supervision of Dr. Konermann.

## Acknowledgements

First of all, I would like to express my highest gratitude to my supervisor, Professor Lars Konermann, who has not only prompted my passion for science but also is an inspiration for good manners and kindness. He is the true definition of a leader and the ultimate role model. Being patient and organized, he helped me through all the stages of my project and taught me how to tackle different challenges.

I would also like to give my special thanks to my family for their continuous support and encouragement even from a long distance. The constant love and sincerity from my parents have always motivated me in all the steps to make me in the present stage. Deepest thanks also to my siblings, Yasaman and Ali, who have always been by my side in every hardship. I am also grateful to my friend Hedyeh, who was always like a big sister and life coach to me, and my little friend Hannah, whose smile and energy have always inspired me directly and indirectly. I am appreciative to have my friend Megan by my side on vicissitudes of life.

I am grateful to have worked with incredible lab mates. Very special thanks to my friend, Elnaz, who helped me with her camaraderie and support. I would also express my gratitude to my former labmates, Haidy, Victor, and Yiming who helped and mentored me when I start my project in the laboratory. I would also thank Pablo who was always ready to troubleshoot problems and fix them. A special thanks to Kirsty for doing simulations and cooperation with me.

Last but not least, I like to thank my committee members Dr. Len Luyt, Dr. Ken Yeung, Dr. Vera Tai and Dr. Chris Thibodeaux for their valuable guidance and support.

## Table of Contents

Abstract.....	i
Summary for Lay Audience.....	iii
Co-Authorship Statement.....	v
Acknowledgements.....	vi
Table of Contents.....	i
List of Symbols and Abbreviations.....	iv
Chapter 1: Introduction.....	1
1.1 Proteins.....	1
1.1.1 Protein Structure.....	1
1.1.2 Protein Folding and Dynamics.....	3
1.1.3 Antibodies.....	5
1.2 Native vs. Unfolded Proteins.....	10
1.2.1 Chemical Denaturants.....	12
1.2.2 Thermal Unfolding.....	15
1.2.3 Protein Aggregation.....	16
1.3 Methods for Studying Protein Thermal Stability.....	18
1.3.1 Circular Dichroism (CD) Spectroscopy.....	18
1.3.2 Differential Scanning Calorimetry (DSC).....	20
1.4 Mass Spectrometry.....	22
1.4.1 Fundamentals of Mass Spectrometry.....	22
1.4.2 Ion Source.....	23
1.4.3 Mass Analyzer.....	24
1.4.4 High Performance Liquid Chromatography Mass Spectrometry (HPLC-MS).....	26

1.5	Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) .....	27
1.5.1	HDX Fundamentals .....	28
1.5.2	Peptide Mapping .....	31
1.5.3	Data Analysis in HDX-MS .....	32
1.6	Scope of Thesis .....	33
1.7	References .....	35
2	Chapter 2. Analysis of Temperature-Dependent H/D Exchange Mass Spectrometry Experiments.....	39
2.1	Introduction.....	39
2.2	Methods.....	42
2.2.1	Materials .....	42
2.2.2	Optical Experiments.....	42
2.2.3	H/D Exchange Mass Spectrometry .....	43
2.3	Results and Discussion .....	46
2.3.1	Spectroscopic Characterization of Global Unfolding.....	46
2.3.2	Temperature- and Time-Dependent HDX Experiments .....	48
2.3.3	Developing a Temperature-Dependent HDX-MS Model.....	51
2.4	Conclusions.....	76
2.5	References.....	78
3	Chapter 3. Mechanism of Thermal Protein Aggregation: Experiments and Molecular Dynamics Simulations on the High Temperature Behavior of Myoglobin.....	82
3.1	Introduction.....	82
3.2	Methods.....	85
3.3	Results and Discussion .....	86
3.3.1	Thermal Aggregation Experiments.....	86

3.3.2	Relationship between Global Unfolding and Aggregation.....	87
3.3.3	Overview of Aggregation MD Data .....	90
3.3.4	Conclusions.....	92
3.4	References.....	94
4	Chapter 4. Structural Dynamics of a Thermally Stressed Monoclonal Antibody Characterized by Temperature-Dependent H/D Exchange Mass Spectrometry.....	99
4.1	Introduction.....	99
4.2	Methods.....	103
4.2.1	Experimental Procedure.....	103
4.2.2	Data Analysis .....	105
4.3	Results and Discussion .....	109
4.3.1	DSC Unfolding Experiments .....	109
4.3.2	T-Dependent HDX Analyses for Selected Peptides. ....	112
4.4	Conclusions.....	133
4.5	References.....	136
5	Chapter 5. Conclusions .....	141
5.1	Summary.....	141
5.2	Future Directions .....	143
5.2.1	Develop Instrumentation and Software to Analyze T-dependent HDX- MS Data .....	143
5.2.2	Computational Simulations of Aggregation for Large Multi-Domain Proteins .....	143
	APPENDIX I-PERMISSIONS .....	145
	APPENDIX II- Curriculum Vitae.....	148

## List of Symbols and Abbreviations

HDX-MS	hydrogen deuterium exchange mass spectrometry
Mb	myoglobin
UV-Vis	ultraviolet-visible
Ig	immunoglobulin
RBD	receptor binding domain
mAbs	monoclonal antibodies
ADCs	antibody-drug conjugates
FDA	food and drug administration
CD	circular dichroism
PEM	photo-elastic modulator
DSC	differential scanning calorimetry
MS	mass spectrometry
LC	liquid chromatography
GC	gas chromatography
CE	capillary electrophoresis
ESI	electrospray ionization

HPLC	high performance liquid chromatography
UPLC	ultra-performance liquid chromatographic
EI	electron ionization
CI	chemical ionization
APCI	atmospheric pressure chemical ionization
MALDI	matrix-assisted laser desorption ionization
DESI	desorption electrospray ionization
PS	paper spray ionization
TOF	time-of-flight
FT-ICR	fourier transform ion cyclotron resonance
DC	direct current
RF	radio frequency
P	protection factor
DDA	data-dependent acquisition
DIA	data independent acquisition
HDX	hydrogen/deuterium exchange
MRE	mean residue ellipticity
MD	molecular dynamics

## Chapter 1: Introduction

### 1.1 Proteins

Proteins are biological macromolecules that are involved in all physiological processes. Genetic information is encoded in DNA, but it is put into action in the form of proteins. Proteins are involved in the catalysis of biochemical reactions; they act as receptors for cellular signals, facilitate the transport of small molecules, provide cellular structure, and they protect the body against pathogens. Proteins have long been used medicinally, for example, insulin<sup>1</sup> (for treating diabetes), gamma-globulin<sup>2</sup> (for boosting short-term immunity), and protein-containing vaccines<sup>3</sup> (for providing long-term immunity against infectious agents). In recent years, numerous protein therapeutics have emerged that are now being used to treat a number of diseases, including cancers and genetic disorders, in a highly specific fashion.<sup>4,5</sup>

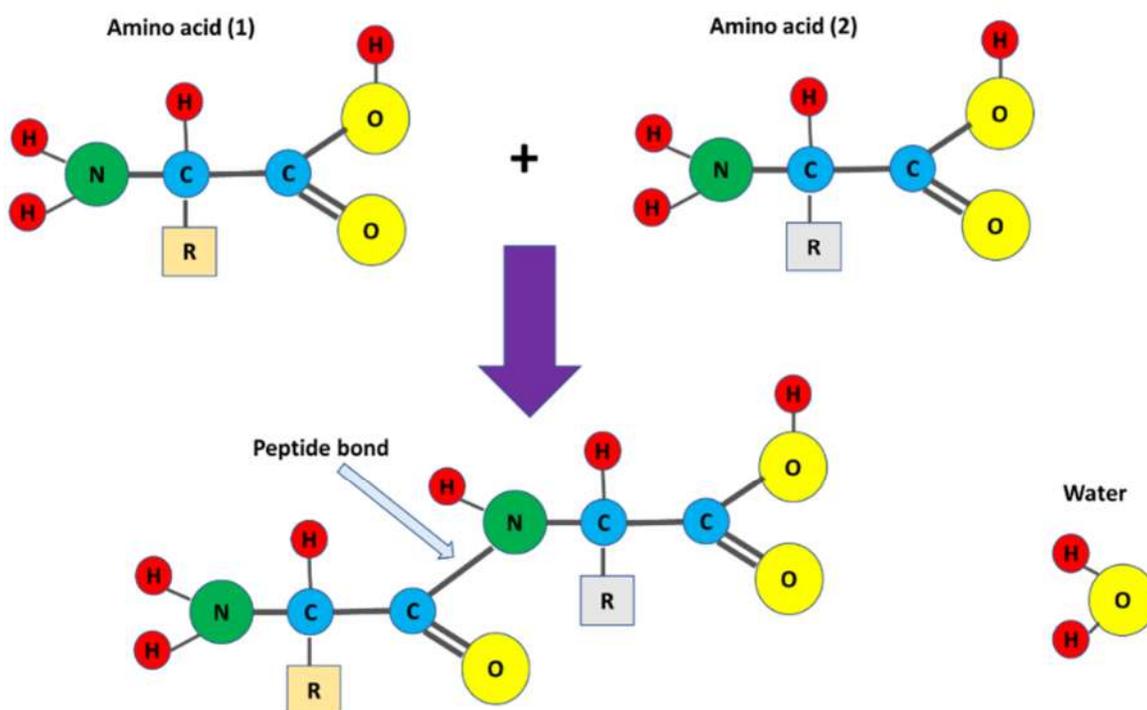
#### 1.1.1 Protein Structure

A protein typically consists of hundreds (sometimes thousands) of amino acids. Twenty different types of canonical amino acids can be distinguished by their unique side chains (“R” groups), which can be hydrophilic or hydrophobic. The amino acids with hydrophobic side chains include alanine (Ala/A), valine (Val/V), proline, leucine (Leu/L), isoleucine (Ile/I), methionine (Met/M), tryptophan (Trp) and phenylalanine (Phe/F).

On the other hand, the side chains of polar amino acids favor residing in an aqueous environment. The amino acids with side chain that are polar but uncharged include serine (Ser/S), threonine (Thr/T), tyrosine (Tyr/Y), asparagine (Asn/N), and glutamine (Gln/Q). Residues that carry a net charge at pH 7 are most hydrophilic. Two of them are negatively charged; glutamic acid (Glu/E) and aspartic acid (Asp/D), while two others are positively charged; lysine (Lys/K) and arginine (Arg/R). Depending on pH, histidine (His/H) can carry a charge or not, and for pH below 6 it is positively charged.

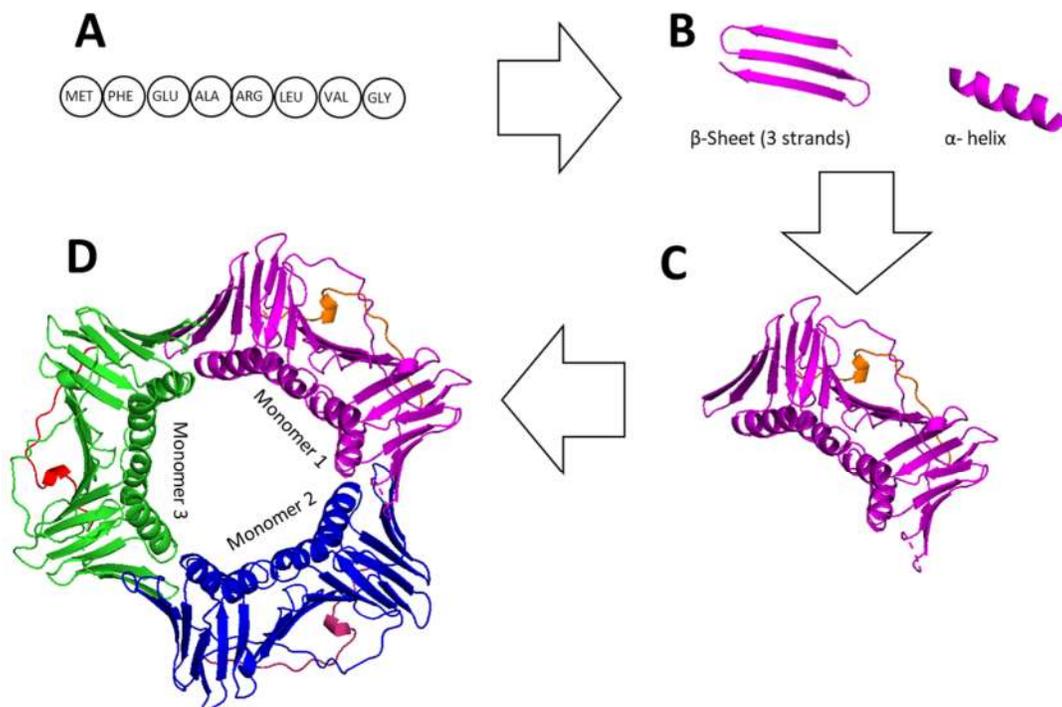
Peptide bonds allow amino acids to create polypeptide chains. A peptide bond is generated by a condensation reaction between the carboxyl group of one amino acid and the amino

group of the next residue (Figure 1.1). In the cell, these reactions are catalyzed by ribosomes. The sequence of amino acids is called primary structure. This sequence also determines the final three-dimensional structure of each folded protein. Intramolecular hydrogen bonds between backbone amide and carbonyl groups give rise to secondary structure elements such as  $\alpha$ -helices or  $\beta$ -sheets.



**Figure 1.1** *Peptide bond formation between two amino acids.*

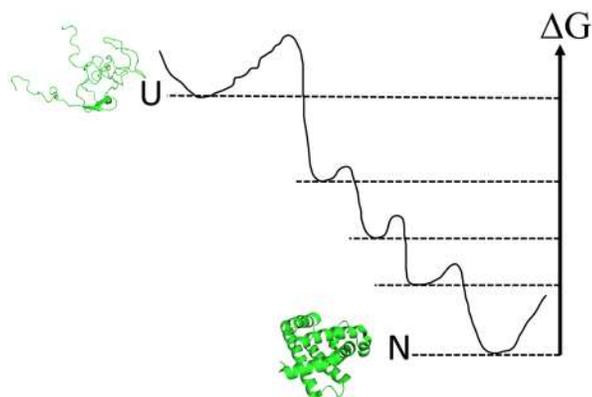
Biologically active proteins have highly ordered tertiary structures that mediate specific functions. These biologically active “native” conformations are created by consolidating secondary structure elements. The hydrophobic effect is one of the driving forces for the formation of tertiary structures. In addition, there are contributions from other interactions such as salt bridges, van der Waals contacts, hydrogen bonds, and disulfide bonds. Some proteins assemble further into quaternary structures consisting of several subunits. Each subunit has its distinct primary, secondary and tertiary structure. (Figure 1.2)



**Figure 1.2.** Protein structure (PDB ID 1AXC). (A) primary structure. (B) Secondary structure. (C) tertiary structure (D) quaternary structure.

### 1.1.2 Protein Folding and Dynamics

In principle, each protein can adopt many different structures. According to the Boltzmann distribution, the equilibrium population of any possible conformer is given by its free energy. Under physiological solvent conditions, the native state (N) has the lowest free energy, so the formation of this conformation from the unfolded state (U) is a spontaneous process. This fundamental principle was uncovered by Christian Anfinsen, who received the 1972 Chemistry Nobel Prize.<sup>6</sup> On its folding trajectory from U to N, the protein may pass through several intermediates. Intermediates along the folding pathway have a successively lower free energy, thereby guiding the protein toward N on the free energy landscape.<sup>7</sup> (Figure 1.3)



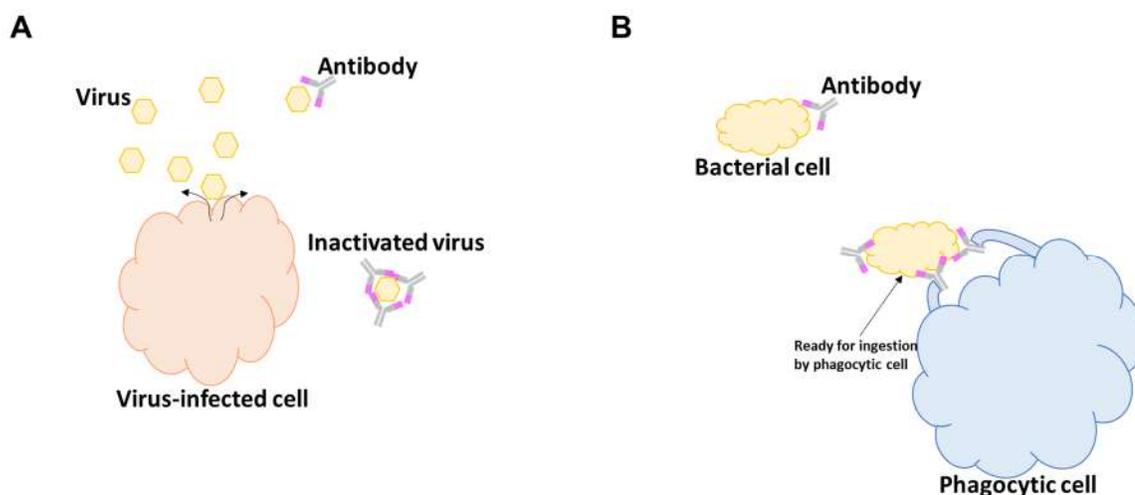
**Figure 1.3** *Proteins fold from the unfolded state to the native state via several intermediates. Each conformation is characterized by its free energy relative to N.*

Although it is undeniable that proteins require a properly folded structure to carry out their biological tasks, native proteins in solution are not static but fluctuate between different conformational states.<sup>8</sup> The interconversion rates between these conformers are governed by the barrier heights separating them, and their populations are determined by Boltzmann weights, as noted above. Protein function is closely correlated with these dynamic motions. Proteins are the molecular machines of biological systems, and proper dynamics have been an integral aspect of their evolutionary selection.<sup>9</sup> These conformational dynamics range from fast vibrations of interatomic bonds to complete folding/unfolding fluctuations. These events can occur on time scales ranging from ps to minutes.<sup>10</sup> It has taken the effort of theorists and experimentalists from the fields of physics, chemistry, and biology to gradually uncover the connections between structure, dynamics, and function.<sup>11</sup>

As an example of the essential role of conformational dynamics, myoglobin, a small protein that stores oxygen in muscle cells, was the first protein to have its three-dimensional structure deciphered.<sup>12</sup> However, it was instantly realized that dynamic motions are required for its biological function since the crystallized protein lacks an open pathway for oxygen to enter and exit the binding site on the heme.<sup>13</sup> Fluctuations of the distal histidine, which blocks the path in the crystal structure, appeared to mediate a quick route for oxygen to reach its binding site.<sup>14</sup> Myoglobin was used as a model system for several multidisciplinary studies in protein science, and it was even referred to as the "hydrogen atom of biology".<sup>15</sup> The experiments in Chapters 2 and 3 will focus on myoglobin as well.

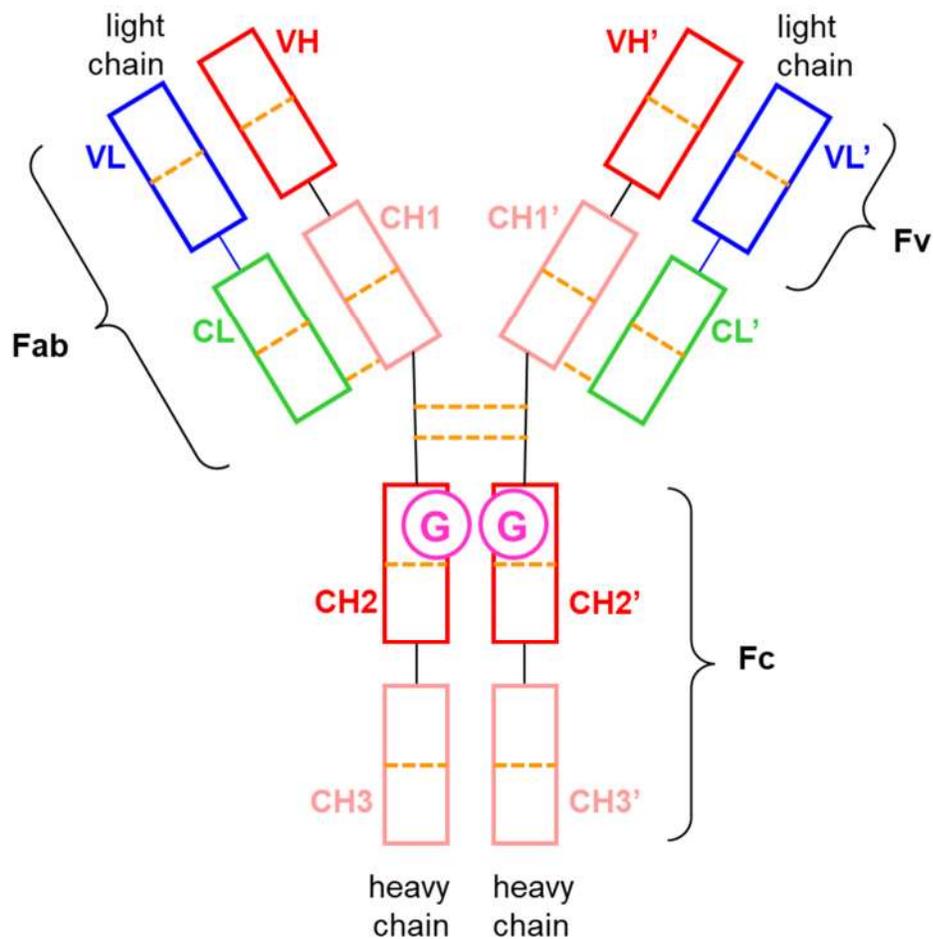
### 1.1.3 Antibodies

Antibodies are glycoproteins that belong to the Immunoglobulin (Ig) family and are part of the immune system. They can recognize and destroy pathogens such as viruses and bacteria. Contact of these pathogens with B-cells causes the production of antibodies. The pathogens that trigger an immune response are called antigen. Antibodies can detect and bind with specific surface features of the antigen, leading to antigen neutralization and/or phagocytosis<sup>16</sup> (Figure 1.4).



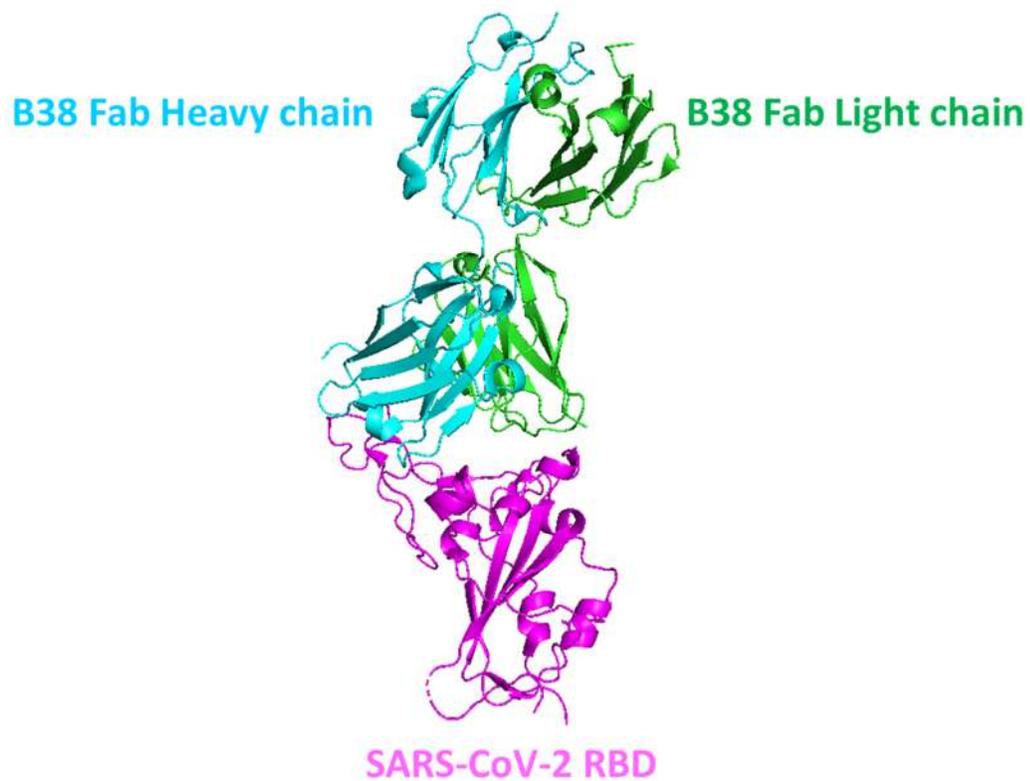
**Figure 1.4** *Antibody functions. (A) Neutralization; antibodies can bind and inactivate the antigen by neutralizing viral infectivity. (B) Opsonization; antibody bind to the antigen and a phagocytic cell seizes the antibody-antigen complex and ingests it.*

Antibodies are Y-shaped structures that consist of two identical heavy chains and two identical light chains that are linked by disulfide bonds. Each heavy chain has one variable domain (VH) and three constant domains (CH1, CH2, CH3). Both light chains have one variable (VL) and one constant domain (CL). Each of the two Fab arms comprises CH1 and CL, along with VH and VL, which form the Fab moiety that is responsible for antigen binding. The Fc substructure consists of CH2/CH2' and CH3/CH3' domains, the former being decorated with N-linked glycans.<sup>17</sup> (Figure 1.5)



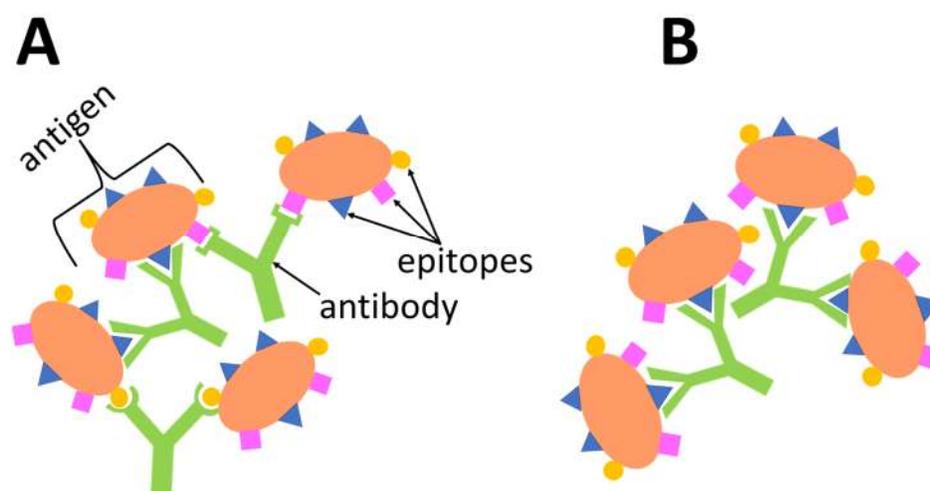
**Figure 1.5** *Cartoon representation of an IgG1 mAb. Disulfides are represented as dashed orange lines, “G” indicates glycans.*

The SARS-CoV-2 virus caused a worldwide COVID-19 pandemic. Specific antibodies generated by the immune system can neutralize the virus and prevent it from being infectious. Such “neutralizing” antibodies block the receptor binding domain (RBD) on the SARS-CoV-2 spike protein, preventing it from binding with the ACE2 receptor on the surface of the lung and other tissues.<sup>18</sup> Figure 1.6 illustrates how the RBD interacts with the Fab domain of an antibody.



**Figure 1.6** Structure of COVID-19 virus spike receptor-binding domain complexed with a B38 neutralizing antibody (pdb code 7BZ5).

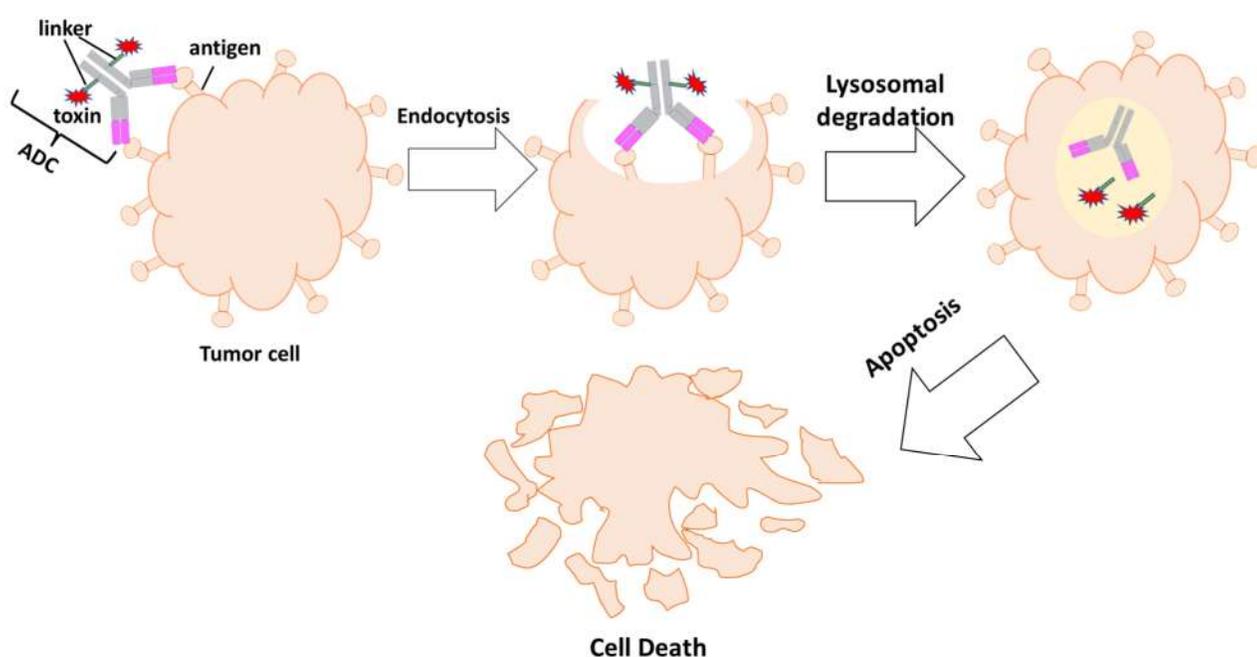
Monoclonal antibodies (mAbs) are laboratory-produced molecules with the same homogeneous sequence and structure that bind to only one unique epitope on the antigen. In contrast, polyclonal antibodies represent a mix of various proteins that bind to multiple epitopes (Figure 1.7).



**Figure 1.7** Antibody (green) binding to antigens (orange). (A) polyclonal antibody can bind to multiple epitopes. (B) monoclonal antibodies bind to the same epitope.

Therapeutic mAbs were introduced as drug candidates in the 1980s by Jerne, Köhler and Milstein, who jointly awarded the 1984 Nobel Prize in Physiology or Medicine. mAbs are becoming increasingly popular as novel types of medicines.<sup>19</sup> Antibody-drug conjugates (ADCs) are mAbs that have been coupled to cytotoxic chemicals.<sup>20</sup> They originated as a targeted therapy for treating diseases and are extensively used for cancer treatment. ADCs can be utilized for drug delivery by binding receptors on the surface of cancerous cells. Paul Ehrlich, a physicist who won the 1913 Nobel Prize, was the first to propose this strategy, which he called the ‘magic bullet’ that relies on the directed transport of chemotoxic chemicals to treat microbial infections or tumors. However, it took over 40 years to achieve successful clinical trials on ADCs.<sup>21</sup> One therapeutic option for the treatment of cancer is traditional chemotherapy. However, this approach is frequently linked to poor therapeutic response and significant toxicity against normal healthy tissues due to its low selectivity towards tumor cells. ADCs, as opposed to traditional chemotherapy, target and destroy tumor cells while minimizing the damage to healthy tissues.<sup>22</sup>

ADCs have three components, (1) the mAb that selectively binds to the antigens on the surface of tumor cells, (2) a cytotoxic substance that can kill target cells once it has been absorbed and released from the mAb, (3) the linker that binds the cytotoxic agent to the mAb. This linker has to have high stability in the circulatory system, but in the tumor cell it must be able to release the cytotoxic molecule. The mechanism of ADCs for treating cancer is as follows (Figure 1.8): 1) the mAb selectively attaches to the tumor cell; 2) it is internalized in the cell by endocytosis ; 3) lysosomes degrade it; and 4) the cytotoxic molecule is released leading to cell death (apoptosis).<sup>23, 24</sup>



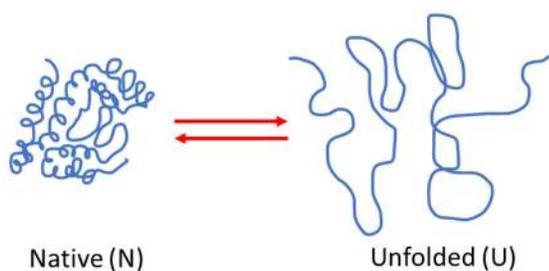
**Figure 1.8.** *Antibody drug conjugate (ADC) mechanism as a targeted therapy for cancer.*

In 2000, the first ADC was approved by the United States Food and Drug Administration (FDA) to treat acute myeloid leukemia, which is a cancer of the blood and bone marrow.<sup>25</sup> From 2011 to 2018, FDA approved four other ADCs, whereas from 2019 to 2020 five more ADCs approved by FDA and entered the market for treating cancer.<sup>27</sup> Currently, over 100 ADCs are being investigated in the clinical stages.<sup>23</sup>

## 1.2 Native vs. Unfolded Proteins

Under physiological solvent conditions (ambient temperature, no denaturants, near-neutral pH), the native state “N” is the preferred conformation. Hydrophobic amino acids are predominantly buried in the protein core, while polar residues are mostly found on the surface.<sup>28</sup> According to the iceberg model, the poor solubility of nonpolar solutes in water is mainly due to entropic reasons. Bulk water is highly dynamic, corresponding to a high entropy. The formation of a partially ordered layer of “iceberg water” around hydrophobic surfaces induce entropically unfavorable ordering, thereby promoting the formation of a hydrophobic core. In contrast, hydrophilic/charged residues on the exterior of natively folded proteins interact favorably with water, mainly via hydrogen bonds.<sup>29</sup>

By exposing a protein to denaturing agents (discussed in more detail below), the native state can be disrupted, resulting in protein unfolding. Denaturants play a crucial role for *in vitro* experiments for probing protein stability. The unfolded state “U” is biologically inactive and has traditionally been described as a random coil. However, it is now well established that U can retain considerable residual structure.<sup>30</sup> For cooperative unfolding transitions, there is an apparent two-state equilibrium involving N and U, without any intermediates (Figure 1.9).<sup>31</sup>



**Figure 1.9.** *Reversible two-state unfolding equilibrium. The native state is highly ordered, whereas the unfolded state has a largely disordered structure.*

The equilibrium constant of a two-state unfolding reaction is

$$K_{eq} = \frac{[U]}{[N]}$$

### Equation 1.1

Where [U] and [N] are the equilibrium concentrations of the unfolded and native states, respectively. The free energy difference  $\Delta G^\circ = G_U^\circ - G_N^\circ$  determines how strongly U and N are populated. The relationship between the free energy of unfolding and the equilibrium constant can be expressed as follows

$$K_{eq} = \exp\left(\frac{-\Delta G^\circ}{RT}\right)$$

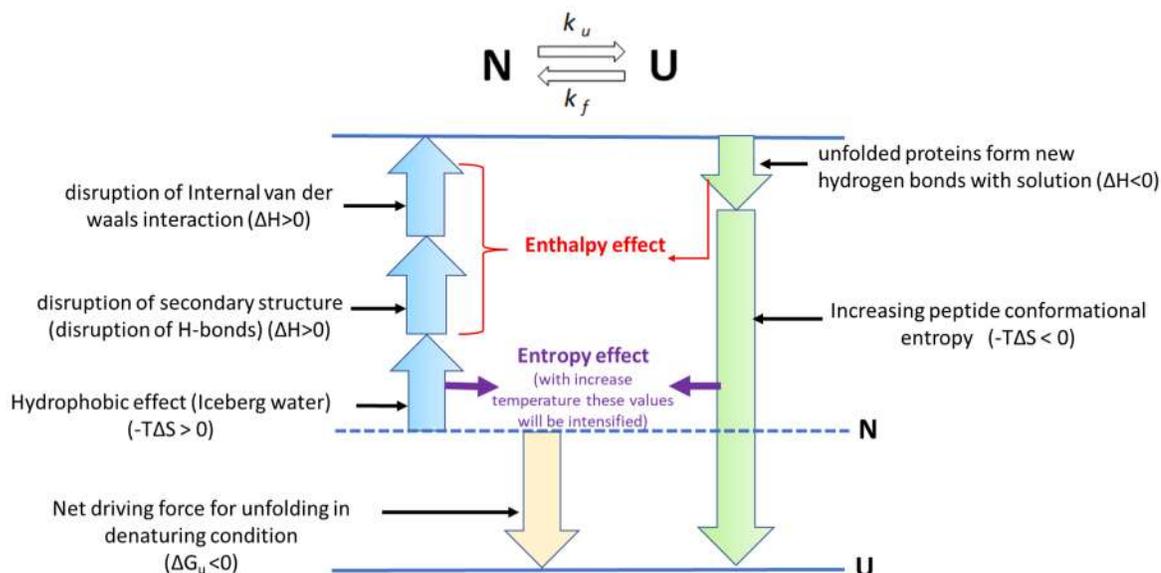
### Equation 1.2

Where  $R$  is the gas constant and  $T$  is the absolute temperature.  $\Delta G$  depends on enthalpy and entropy according to

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

### Equation 1.3

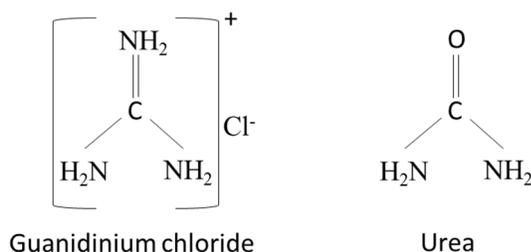
$N$  is stable if the  $N \rightleftharpoons U$  equilibrium has a positive  $\Delta G$ ; conversely, unfolding proceeds spontaneously when  $\Delta G < 0$ . Various factors contribute to the enthalpic ( $\Delta H$ ) and entropic ( $\Delta S$ ) terms in equation 1.3. For  $\Delta H$ , by exposing a protein to denaturants, H-bonds, van der Waals and hydrophobic interactions are destabilized. On the other hand, the unfolded protein forms new H-bonds with the surrounding water that stabilize U. With respect to  $\Delta S$ , the unfolded state has more conformational freedom and, therefore a higher conformational entropy. On the other hand, surrounding formerly buried hydrophobic groups with iceberg water after unfolding is entropically unfavorable (Figure 1.10).<sup>32</sup>



**Figure 1.10.** *The thermodynamics of protein unfolding depend on various factors. For the scenario described here, the native state will spontaneously unfold.*

### 1.2.1 Chemical Denaturants

By exposing natively folded proteins to chemical denaturants such as guanidinium chloride (GdmCl) and urea, the stability of the native state N is reduced, and the protein will transition to the unfolded state U (Figure 1.11). Concentrations required for unfolding are typically  $\sim 6$  M for GdmCl, and  $\sim 8$  M for urea.



**Figure 1.11.** *Guanidinium Chloride and urea are common chemical denaturant.*

Both polar and nonpolar side chains have a higher solubility in denaturant solutions than in pure water. It has been suggested that these denaturants interact favorably with all parts of the protein and reduce the hydrophobic effect. The extent of these interactions is maximized when the protein is unfolded. Computer simulations suggest that both denaturants reduce the hydrophobic effect by modifying the internal structure of water.<sup>33</sup>

From equation 1.2, the stability of a protein can be expressed as

$$\Delta G^{\circ}_{N \rightarrow U} = -RT \ln K_{eq} = -RT \ln \frac{[U]_{eq}}{[N]_{eq}}$$

**Equation 1.4**

$\Delta G^{\circ}_{N \rightarrow U}$  is found to change linearly with denaturant concentration  $[D]$  according to

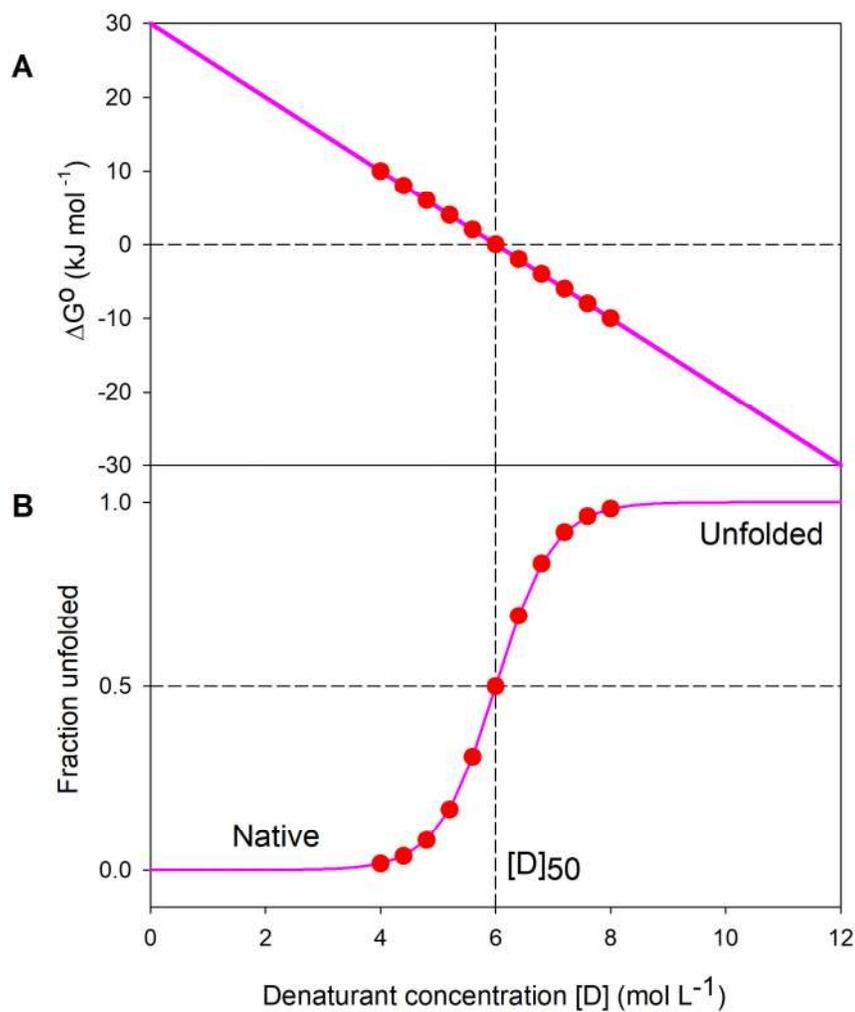
$$\Delta G^{\circ} = \Delta G^{\circ}_{water} - m[D]$$

**Equation 1.5**

The slope of this equation ( $m$ ) relates to the change in the protein's solvent-accessible surface area upon unfolding.<sup>34</sup> Typical  $m$  values for urea and GdHCl are 4 and 12 kJ/mol/M, respectively. Figure 1.12 A shows a typical curve of  $\Delta G^{\circ}$  versus  $[D]$ . In Figure 1.12 B, the fraction of unfolded protein  $f_U = [U]/([N]+[U])$  is plotted as a function of denaturant concentration, using the Boltzmann expression  $f_U = \exp(-\Delta G/RT) / [1 + \exp(-\Delta G/RT)]$ .  $[D]_{50}$  is the denaturant concentration for which half of the proteins are unfolded, and the other half is still native. Under these condition  $\Delta G^{\circ}=0$  and  $[U]=[N]$ . Therefore:

$$\Delta G^{\circ}_{water} = m [D]_{50}$$

**Equation 1.6**



**Figure 1.12.** (A) Free energy ( $\Delta G^\circ$ ) profile of chemical unfolding as a function of denaturant concentration  $[D]$ , based on equation 1.5. While the parameters were:  $\Delta G^\circ_{\text{water}} = 30 \text{ kJ mol}^{-1}$ ;  $m = 5 \text{ kJ mol}^{-1} \text{ M}^{-1}$ ;  $T = 298 \text{ K}$ . (B) Fraction of unfolded protein  $f = [U]/([U] + [N])$  as a function of  $[D]$ , calculated from  $\Delta G^\circ$  values shown in (a). When  $[D]_{50} = 6 \text{ M}$ ,  $\Delta G^\circ$  becomes zero and  $[N] = [U]$ .

## 1.2.2 Thermal Unfolding

Thermal unfolding assays are widely used for assessing protein stability, particularly for mAbs and other therapeutic proteins.<sup>35</sup> In these experiments, a protein is exposed to gradually increasing temperatures, and  $f_U(T)$  is measured until the protein is completely unfolded.<sup>36</sup>

Equation 1.3 shows that  $\Delta G$  is dependent on temperature. From the simple facts that  $\Delta H > 0$  (unfolding is endothermic) and  $\Delta S > 0$  (unfolding increases entropy), it follows that an increase in  $T$  will result in  $\Delta G < 0$  once the temperature is raised beyond a specific value which is referred to as melting temperature  $T_m$ . In other words, for  $T > T_m$  the protein will be unfolded. The situation can be more complicated because  $\Delta H$  and  $\Delta S$  change as a function of temperature (Equations 1.7 and 1.8).<sup>37</sup>

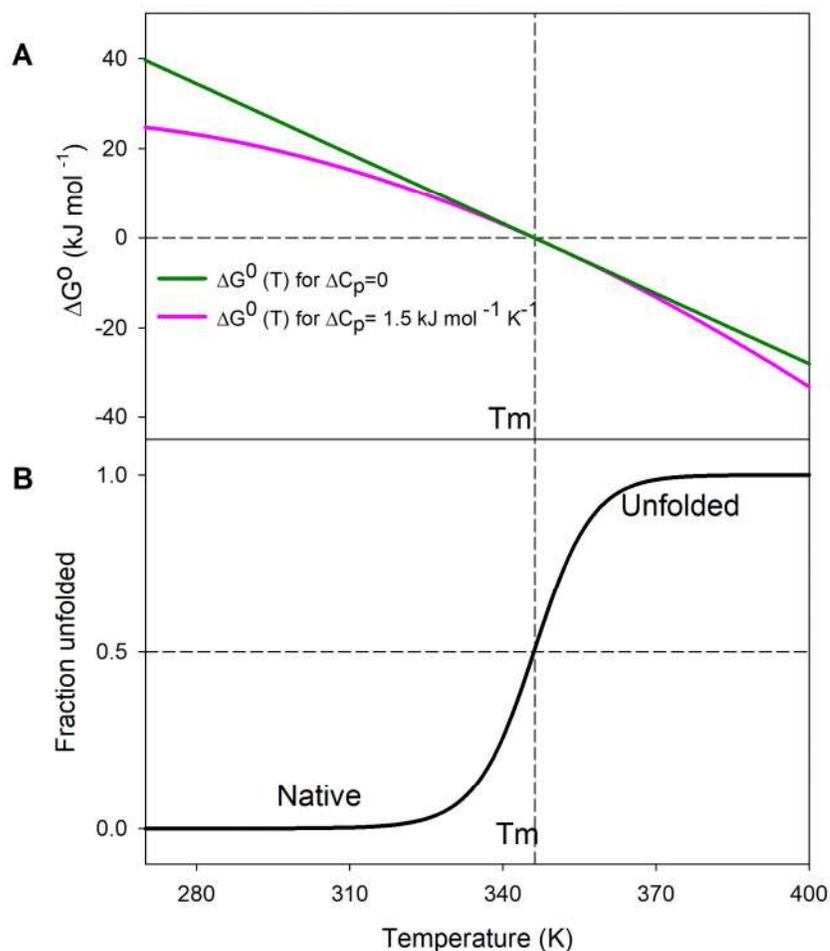
$$\Delta H^\circ = \Delta H^\circ_R + \Delta C_p(T - T_R)$$

**Equation 1.7**

$$\Delta S^\circ = \Delta S^\circ_R + \Delta C_p \ln \frac{T}{T_R}$$

**Equation 1.8**

The temperature dependence expressed in equations 1.7 and 1.8 implies that  $\Delta G^\circ(T)$  profiles are curved instead of being a straight line (Figure 1.13). When we assume  $\Delta C_p = 0$ ,  $\Delta G$  depends linearly on temperature, while for  $\Delta C_p > 0$ , the curvature of  $\Delta G$  is readily apparent (Figure 1.13A). As for chemical unfolding,  $f_U$  can be calculated from the  $\Delta G(T)$  profiles.



**Figure 1.13.** (A) plot of the free energy of thermal unfolding ( $\Delta G^\circ$ ) as a function of temperature. (B) Fraction of unfolded protein, calculated from the free energy profile. Parameters used:  $\Delta C_p = 1.5 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ,  $\Delta H^\circ_R = 180 \text{ kJ mol}^{-1}$  and  $\Delta S^\circ_R = 0.52 \text{ kJ mol}^{-1} \text{ K}^{-1}$ . Here, the melting temperature of  $T_m = 350 \text{ K}$  is identical to the reference temperature  $T_R$ .

### 1.2.3 Protein Aggregation

Protein-protein interaction plays an essential role in biology, for example, when several protein chains assemble into highly ordered quaternary structures. However, some types of protein-protein contact are detrimental to human health.<sup>38</sup> The interior of biological cells is highly crowded, giving rise to possible nonspecific interactions between proteins and other molecules that can lead to aggregate formation. Protein aggregation depends on

solution conditions, such as temperature, pH, salt and buffer condition, agitation, surfactant effects, and protein concentration.<sup>39</sup> Aggregated proteins are often insoluble, resulting in precipitates that give rise to “cloudy” or “milky” solutions. After synthesizing a polypeptide chain on a ribosome, the chain may fold to its native structure via one or more partially folded intermediates. In the cell, the situation is even more complex due to the involvement of chaperones, i.e., proteins that ensure efficient folding and prevent aggregation.<sup>40</sup> Unfolded or partially folded proteins can undergo aggregation:

Native → Aggregated

Native ⇌ Partially Folded → Aggregated

Native ⇌ Unfolded → Aggregated

Aggregates can be disordered and amorphous, or they can be highly ordered and fibrillar. The latter are called amyloid fibrils; they are rich in  $\beta$ -sheet structure and can form from almost any amino acid sequence.<sup>40</sup> These  $\beta$ -sheet structures are not related to native conformations; for instance, although native myoglobin has an all  $\alpha$ -helical globular structure, it can transition into a fundamentally different conformation that shows typical characteristics of amyloid fibrils.<sup>41</sup>

Misfolding and aggregation are associated with a host of human diseases,<sup>42</sup> known as amyloidosis. Many of these are neurodegenerative disorders such as Alzheimer's, where aggregates form in the brain, or type II diabetes, where deposits form in the pancreas.<sup>43</sup> Some amyloidogenic proteins called “prions” can transmit their misfolded shape onto normal protein via a templating mechanism, and they are responsible for Creutzfeldt-Jakob disease in humans or “mad cow disease” (BSE) in cattle.<sup>44</sup>

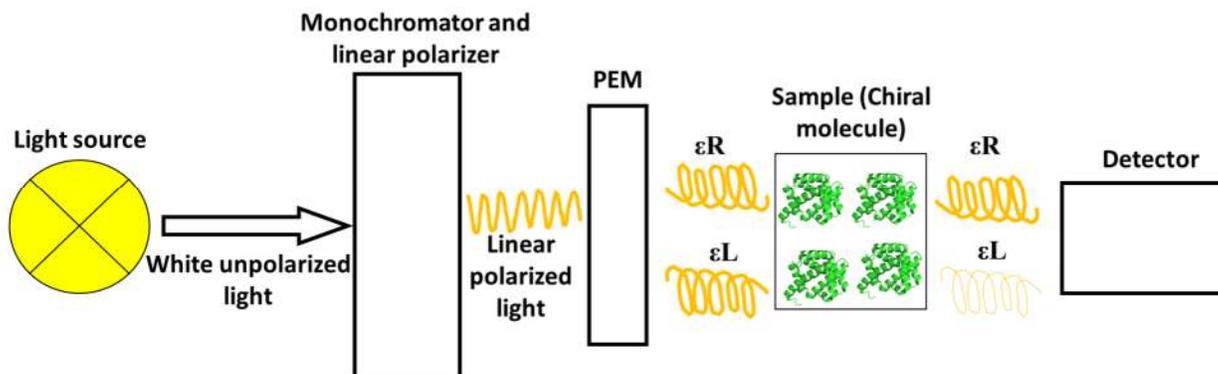
Aggregation is also a major concern during the production and storage of biopharmaceuticals. For commercializing protein therapeutics, physical and chemical instability are key challenges.<sup>45</sup> Protein aggregation is the most common source of instability throughout the product development pipeline.<sup>46</sup>

Aggregation is caused by non-native intermolecular interactions. The native state only has the lowest free energy when considering an isolated protein in solution (or a protein at a low concentration). The native state is typically only about 20-80 kJ mol<sup>-1</sup> lower in free energy than the unfolded state. This thermodynamic stability is marginal, considering that single covalent bonds have energies of ~300 kJ mol<sup>-1</sup>.<sup>47</sup> Aggregates represent the lowest free energy state in concentrated protein solutions, making aggregation thermodynamically favorable under such conditions. The mechanisms underlying protein aggregation remain poorly understood, and the relationship between unfolding and aggregation is a subject of great interest. The route from the native state to various aggregates may proceed through different monomeric precursor species. The exact nature of these precursor species remains to be established.<sup>48</sup>

## 1.3 Methods for Studying Protein Thermal Stability

### 1.3.1 Circular Dichroism (CD) Spectroscopy

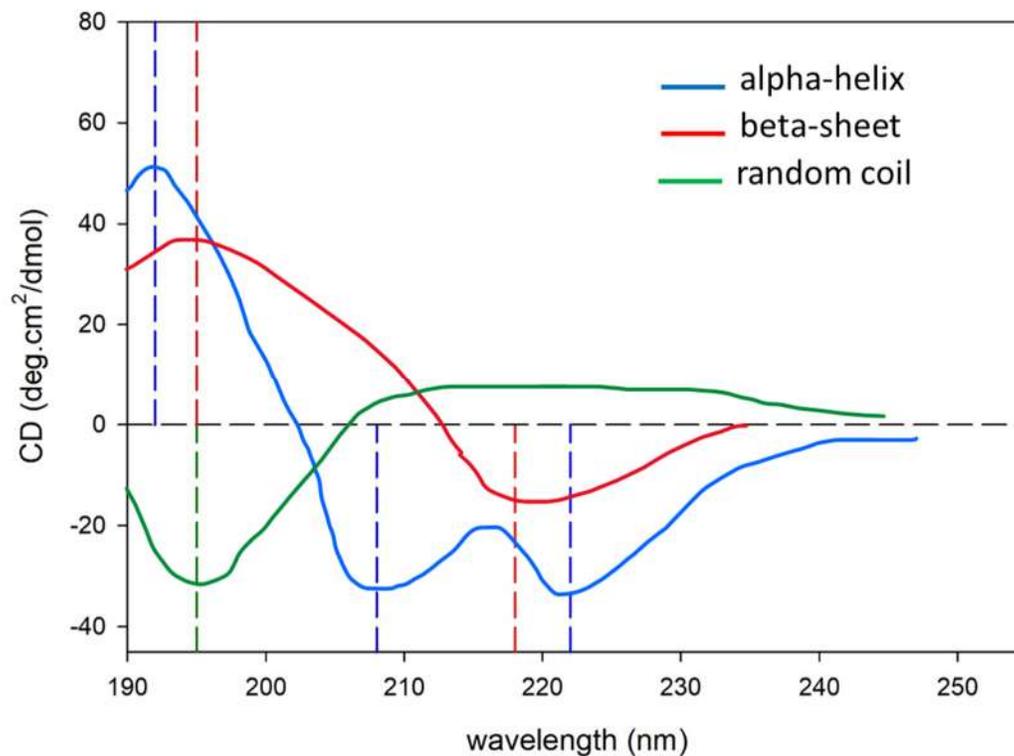
Circular dichroism spectroscopy is an absorption-based technique widely used for monitoring protein secondary structure and tertiary structure.<sup>49</sup> There are two ways that light can be circularly polarized: left and right. In CD spectroscopy, a monochromator is used select linearly polarized light of a specific wavelength. Circularly polarized light is generated by passing this beam through a modulating device, often a photo-elastic modulator (PEM). Chiral molecules absorb left ( $\epsilon_L$ ) and right ( $\epsilon_R$ ) circular polarized light differently. A CD spectrum measures the difference in light before and after passing through the protein solution (Figure 1.14). A CD spectrum is a plot of ( $\epsilon_L - \epsilon_R$ ) versus wavelength ( $\lambda$ ).<sup>50</sup> This wavelength dependence is uncovered by slowly scanning the monochromator.



*Figure 1.14. Schematic layout of a CD experiment.*

Far-UV CD signals (180-250 nm) of the polypeptide backbone report on protein secondary structure. A negative peak at 195 nm indicates the prevalence of random coil structure, while two negative peaks (222 and 208 nm) and one positive peak (192 nm) indicate a high percentage of  $\alpha$ -helical segments. In addition, one negative peak (218 nm) and one positive peak (195 nm) serve as indicators for  $\beta$  sheets (Figure 1.15).<sup>50</sup>

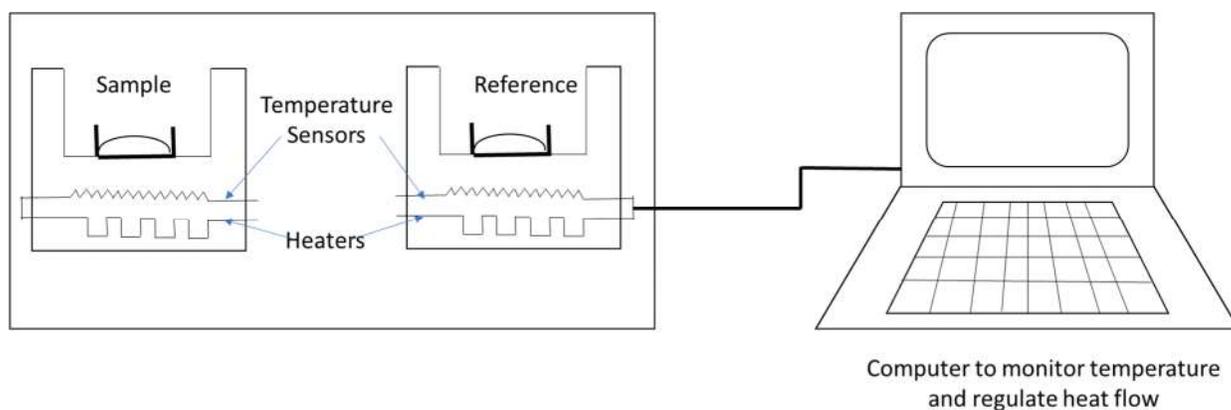
One of CD spectroscopy's most typical applications is monitoring conformational changes, i.e., unfolding  $f_U(T)$ . 222 nm represents the most common detection wavelength. Many CD instruments have cuvettes that allow the temperature of the sample to be controlled for monitoring thermal unfolding. Although widely used, CD spectroscopy cannot provide structural information at the residue level.<sup>51</sup>



**Figure 1.15.** Cartoon diagram of CD spectra that represent principal polypeptide secondary structures ( $\alpha$ -helix,  $\beta$ -sheet and random coil).

### 1.3.2 Differential Scanning Calorimetry (DSC)

The heat capacity  $C_p(T)$  of a protein can be determined using differential scanning calorimetry (DSC). Throughout the experiment, the sample and reference are kept at the same temperature. The sample and reference holder temperature rise linearly as a function of time, and heat flow into or out of a sample is monitored (Figure 1.16).<sup>52</sup>



**Figure 1.16.** *Schematic layout of a DSC instrument.*

The ratio of heat flow to heating rate reflects the heat capacity according to

$$C_p = \frac{\Delta H}{\Delta T} = \frac{dH}{dT}$$

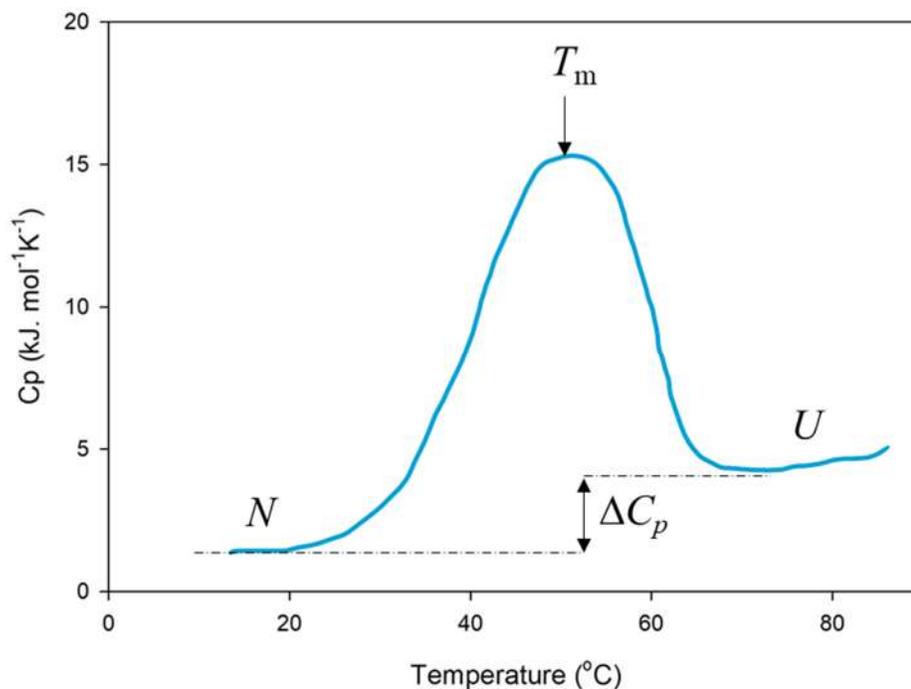
### Equation 1.9

The enthalpy of unfolding ( $\Delta H$ ) is obtained by integrating the area under the  $C_p(T)$  peak, and the baseline offset between U and N represents the  $\Delta C_p$  term that was introduced in equations. 1.7 and 1.8.

$$\Delta C_p = C_p(U) - C_p(N) > 0$$

### Equation 1.10

Similar to CD-based unfolding experiments, DSC only reports on global unfolding transitions. However, for multi-domain proteins such as mAbs, DSC often shows several maxima that reflect the sequential unfolding of different regions at different melting temperatures.



**Figure 1.17.** Cartoon depiction of a DSC thermogram that shows the thermal denaturation of a protein.

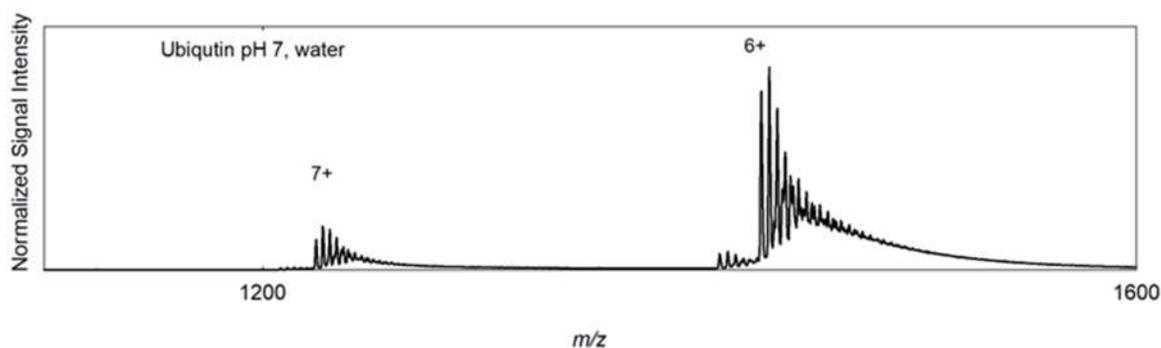
## 1.4 Mass Spectrometry

### 1.4.1 Fundamentals of Mass Spectrometry

Mass spectrometry (MS) is a widely used analytical tool for numerous (bio)analytical applications. In recent years this technique has become particularly important for protein-related studies, i.e., proteomics and “biophysical” MS. The origins of MS go back to the early 20<sup>th</sup> century.<sup>53</sup> MS can be coupled with analytical separation techniques such as liquid chromatography (LC)<sup>54</sup>, gas chromatography (GC)<sup>55</sup> and capillary electrophoresis (CE)<sup>56</sup>. MS measures the mass to charge ratio ( $m/z$ ) of ions in the gas phase. Therefore, an ion source and a mass analyzer are essential components of any MS instrument. Additional information can be provided by incorporating other features such as ion mobility devices, collision cells, etc.

## 1.4.2 Ion Source

As the first step of any MS experiments, the analyte(s) of interest must be ionized and subsequently transported into the vacuum of the mass analyzer. A variety of ionization techniques have been developed over the years. Every type has its unique advantages and drawbacks. The most effective and extensively used ionization method for biological samples is electrospray ionization (ESI).<sup>57</sup> This technique generates intact gaseous ions directly from analytes in the solution. Consequently, ESI can be easily coupled with high performance liquid chromatography (HPLC) and ultra-performance liquid chromatographic (UPLC). This capability significantly broadens the scope of MS for various analytical applications.<sup>58</sup> Additionally, ESI is a "soft" ionization method, that does not rupture covalent bonds. Native MS refers to a unique "flavor" of ESI, where the aim is to preserve noncovalent connections within the protein as well as contacts to weakly bound binding partners.<sup>59</sup> Figure 1.18 displays an example of a native ESI mass spectrum.



**Figure 1.18.** A native mass spectrum of ubiquitin at pH 7

Multiply charged  $[M + zH]^{z+}$  ions are produced via ESI. The mass to charge ratio  $m/z$  of a specific ion can be calculated as

$$\frac{m}{z} = \frac{M + z \times 1.008}{z}$$

**Equation 1.11**

where  $M$  is the mass of the neutral analyte, 1.008 is the proton mass, and  $z$  is the charge state. This equation assumes that protonation is solely responsible for the analyte charge. Under experimental conditions, some cationization (involving  $\text{Na}^+$  or  $\text{K}^+$ ) may take place as well.

Although ESI is the most common method for biological samples, it is not the only ionization method. Numerous other ionization sources can be used, such as electron ionization (EI), chemical ionization (CI), atmospheric pressure chemical ionization (APCI), matrix-assisted laser desorption ionization (MALDI), desorption electrospray ionization (DESI), paper spray ionization (PS), and several others.<sup>60</sup>

### 1.4.3 Mass Analyzer

The mass analyzer's purpose is to measure the  $m/z$  as well as the relative abundance of gaseous ions produced by the ion source. Mass analyzers come in a variety of technologies, including quadrupole, time-of-flight (TOF), ion trap, Orbitrap, and Fourier transform ion cyclotron resonance (FT-ICR) instruments.<sup>61</sup> Quadrupoles and time-of-flight (TOF) mass spectrometers will be explained briefly here.

#### 1.4.3.1 Quadrupole Mass Analyzer

In a quadrupole mass analyzer, four cylindrical rods are arranged in parallel. Opposing rod pairs receive a direct current (DC) and a radio frequency (RF) voltage. For any combination of RF and DC voltages, only ions with one specific  $m/z$  have stable trajectories, allowing them to pass through the quadrupole and arrive at the detector. All other ions have unstable trajectories, collide with the rods, and are neutralized. The Mathieu equations can be used to predict the ion trajectories for every given combination of RF and DC voltages.<sup>62</sup> By scanning the RF and DC voltage amplitudes, one can change the  $m/z$  range that is being transmitted, thereby generating a mass spectrum.

Quadrupoles can also serve as ion guides that transmit any  $m/z$  when operating them in the RF-only mode (without a DC voltage). Consequently, quadrupoles are frequently employed in MS/MS applications as adjustable ion transmission and selection devices.<sup>63</sup> Quadrupoles are exceptionally resilient tools that have remained a standard in environmental and pharmaceutical labs despite their poor resolution and long scanning durations.<sup>64</sup>

### 1.4.3.2 Time of Flight (TOF) Mass Analyzers

Ions are separated by a TOF mass analyzer based on how quickly they pass through a flight tube. By applying a voltage,  $U$ , an ion pusher accelerates ions in a flight tube.<sup>65</sup> The ion velocity  $v$  can be calculated by using the equation

$$zU = \frac{1}{2}mv^2$$

**Equation 1.12**

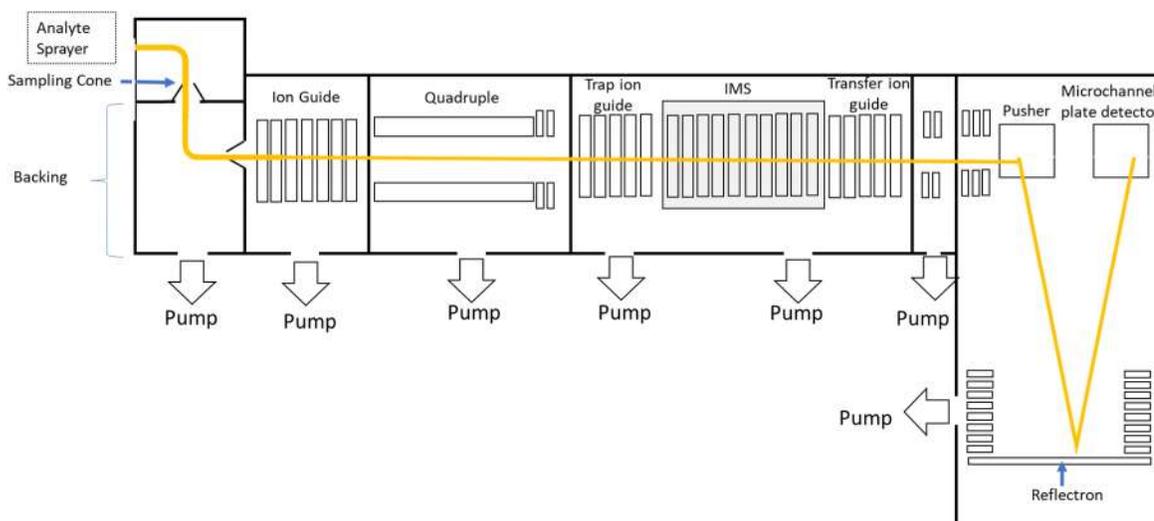
The equation for the time of flight ( $t$ ) that an ion needs to travel through the flight tube with length  $l$  is given by:

$$t = \frac{l}{v} = l \left( \frac{m}{2zU} \right)^{1/2} = l(2U)^{-1/2} \left( \frac{m}{z} \right)^{1/2}$$

**Equation 1.13**

This equation shows that the time of flight depends on the  $m/z$  of the ion. Thus, ions with different  $m/z$  values have different flight times, which causes them to be separated in the flight tube. Lower  $m/z$  ions reach the detector first, followed by their heavier counterparts. Modern TOF mass spectrometers also employ a reflectron to enhance the spectral resolution. Ultimately, a time-to-digital converter that assigns flight times to specific  $m/z$  values,<sup>66</sup>

One of the most widely used instrument designs for biomolecular analysis, known as Q-TOF, results from combining quadrupole and TOF technology.<sup>67</sup> (Figure 1.19).



**Figure 1.19.** Schematic layout of the Q-TOF instrument. The yellow line indicates the ion path.

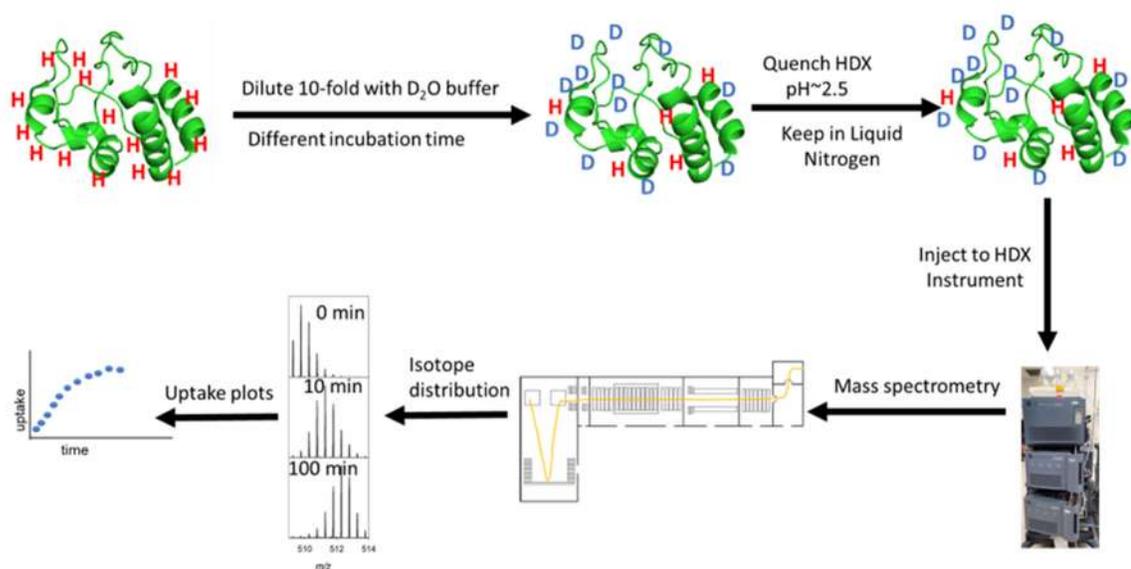
#### 1.4.4 High Performance Liquid Chromatography Mass Spectrometry (HPLC-MS)

High performance Liquid Chromatography (HPLC) is a technique for separating a sample into its components. In this technique, the mobile phase is a pressurized liquid that mediates convective analyte transport through a column that is packed with solid adsorbent material (stationary phase). This stationary phase has different affinity for different components in the sample, causing different retention behavior for different components, and thereby separating the mixtures as they elute from the column. In HPLC-MS, physical separation of liquid chromatography is combined with the capability of measuring the  $m/z$  of analytes to enhance chemical analysis. In summary, HPLC separates different mixtures of components (proteins, peptides, etc.), while MS provides information for detecting each separated component.<sup>68</sup> UPLC is a term used by one specific manufacturer (Waters INC); it refers to the same principle as HPLC, albeit with improved performance due to smaller particle size and higher pressures.

## 1.5 Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS)

Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS) has become a widely used method for studying protein structure and dynamics. This technique relies on the fact that backbone amide hydrogens can be exchanged with deuterium upon protein exposure to a  $D_2O$  solvent environment. The deuteration rates are highly sensitive to protein structure and dynamics. In the native state, most backbone NH sites are involved in intramolecular H-bonding with backbone CO groups. These H-bonded sites are protected and undergo relatively slow HDX.<sup>69</sup> In contrast, disordered and unfolded regions lack stable H-bonds, resulting in much more rapid HDX.<sup>70</sup>

During typical “continuous labeling” HDX-MS experiments, a protein in  $H_2O$  is diluted into  $D_2O$  solution under the desired conditions (usually buffer at pD 7). Aliquots are removed at specified time points, and HDX is quenched by rapidly decreasing the pH to 2.5 and immersing the sample in liquid nitrogen. The deuterium uptake into the protein is then measured as a function of labeling time by MS (Figure 1.20).



**Figure 1.20** Workflow of a typical HDX-MS experiment.

Alternatively, the experiments can be performed in a “pulsed HDX” fashion, where deuteration only takes place for a relatively brief period of time, while some other variables are altered. This second approach is helpful for studying transient folding intermediates.<sup>71</sup>

HDX experiments can be performed on intact proteins or in a spatially resolved fashion at the peptide level. In the latter case, the sample is proteolytically digested after quenching by an acidic protease such as pepsin. For modern workflows, pepsin is immobilized on a digestion column designed for HPLC or UPLC. After digestion, the sample is loaded onto a trapping column for desalting. Subsequently, the peptides are separated on an analytical C18 column. An unavoidable drawback of this workflow is the occurrence of some back exchange during which deuterated peptides lose a certain percentage of their deuterium as they pass through the HPLC. Complete back exchange occurs in amino acid side chains, while a relatively large fraction (often around 70% or more) of amide backbone deuterium can be retained. To minimize amide backbone back exchange, the digestion column is cooled to 15 °C, while the remaining flow path is kept just above 0 °C. The HPLC gradient and flow rates are also optimized to minimize the retention time. Finally, the peptides are analyzed online by ESI-MS, such that deuterium-induced mass shifts can be measured for each peptide. Each amide backbone HDX event increases the protein (or peptide) mass by 1 Da. HDX-MS can be applied for identifying, mapping, and pinpointing the appearance of dynamically fluctuating flexible regions.<sup>72</sup>

### 1.5.1 HDX Fundamentals

NH, OH, and SH groups in proteins can undergo deuteration upon exposure to D<sub>2</sub>O. Because side chain sites undergo complete back exchange during HPLC, we only focus on backbone NH sites. Most backbone NH sites in native proteins are involved in H-bonds (in  $\alpha$ -helices and  $\beta$ -sheets). This H-bonding implies that the corresponding sites exchange very slowly. Their deuteration is mediated by conformational fluctuations (opening/closing transitions) of the protein.<sup>32</sup> These fluctuations may correspond to local, sub-global, or global events.

First, let us consider a scenario where HDX happens at completely open (random coil) backbone amide sites that are fully solvent accessible and not involved in any H-bonding. Deuteration under these conditions is referred to as “chemical exchange”, and it takes place with the pseudo first-order rate constant  $k_{ch}$ . Chemical exchange can be catalyzed by either acid or base according to

$$k_{ch} = k_H [H^+] + k_{OH} [OH^-]$$

**Equation 1.14**

where  $k_H$  and  $k_{OH}$  are the second-order rate constants for acid and base catalysis. These values depend on the adjacent amino acid side chains. Hence, the primary structure impacts HDX rates. For a polyalanine (which is considered to be a typical model),  $k_{ch}$  has its minimum at pH 2.5. Thus, at pH > 2.5 the reaction is catalyzed by base. In other words, for experiments performed under physiological conditions (pH 7)  $k_{OH}$  plays the dominant role, while  $k_H$  effects are negligible.

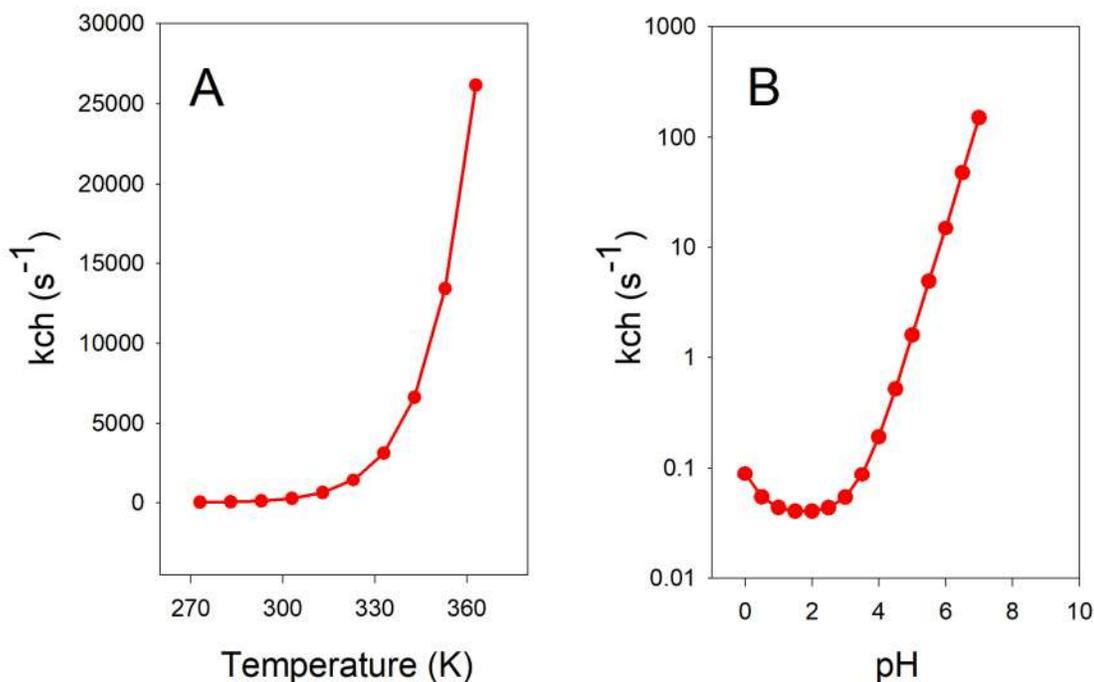
$k_H$  and  $k_{OH}$  depend on temperature  $T$  as governed by the Arrhenius equation

$$k_{OH} = A \exp^{-\frac{E_a}{RT}}$$

**Equation 1.15**

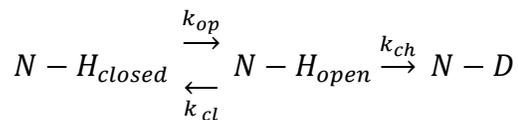
where  $E_a$  represents the activation energy and  $R$  is the gas constant.

The dependence of  $k_{ch}$  on pH and temperature for different protein sequences was uncovered by Englander et al.<sup>73-75</sup> In Figure 1.21,  $k_{ch}$  for the third residue in polyalanine is calculated based on their work. These figures illustrate why HDX can be quenched under UPLC conditions by using pH 2.5 and ~0° C.



**Figure 1.21.** Dependence of the chemical rate constant  $k_{ch}$  on (A) Temperature (B) pH for the third residue in a polyalanine chain.

For backbone NH sites that are involved in H-bonds, HDX proceeds much more slowly compared to the  $k_{ch}$  expressions discussed above. For HDX to occur, there have to be transient H-bond opening events during which H-donor and acceptor are separated. Such opening/closing events are mediated by thermal fluctuations of the protein, i.e., by its conformational dynamics (equation 1.16).



**Equation 1.16**

In this equation,  $k_{op}$  and  $k_{cl}$  are opening and closing rate constants that describe the protein conformational fluctuations. Equation 1.16 gives rise to two limiting cases. For  $k_{cl} \gg k_{ch}$

the situation is referred to as EX2, where numerous opening/closing events are required before deuteration takes place. The overall HDX rate constant ( $k_{HDX}$ ) under these conditions is given by

$$k_{HDX} = \frac{k_{op}}{k_{cl}} k_{ch}$$

**Equation 1.17**

In contrast to EX2, when  $k_{ch} \gg k_{cl}$ , amides will exchange during the first opening event, such that  $k_{HDX} = k_{op}$ . This so-called EX1 mechanism is characterized by bimodal isotopic distribution in the mass spectra because nondeuterated and fully deuterated species coexist in the sample.<sup>76</sup>

protection factor  $P$  is commonly used to report how much slower the measured  $k_{HDX}$  value is relative to  $k_{ch}$  (Equation 1.18).

$$P = \frac{k_{ch}}{k_{HDX}}$$

**Equation 1.18**

Each amide hydrogen in a folded protein has a specific protection factor. Protection factors for loops, termini and unfolded regions are lower (1-100) than those for tightly folded regions (up to  $10^6$  and higher).

## 1.5.2 Peptide Mapping

Usually, the first step to processing bottom-up HDX data analysis is peptide mapping. Data-dependent acquisition (DDA) is the traditional peptide mapping method in which the most abundant precursor ions are selected for fragmentation, and MS/MS is applied to identify the peptides via their  $b$  and  $y$  fragment ions. This process may have to be repeated several times to identify all ions, and a survey scan and precursor selection is needed. Besides, the collision energy may have to be optimized, which makes the processing time

consuming. Also, during MS survey scans, only high-intensity ions will be picked up, and important low-abundance species may be overlooked.<sup>77</sup> Data independent acquisition (DIA) is an alternative method for peptide mapping in which all ions are fragmented without precursor selection. This second approach is also known as MS<sup>E</sup>. The collision cell in the mass spectrometer is alternating between low and high collision energy, such that precursor ions and their fragments can be recorded quasi-simultaneously as peptides elute off the column. The retention time in LC-MS and signal intensity for each precursor are correlated using computer software, such that precursor and fragment ions can be assigned to one another.<sup>78</sup>

### 1.5.3 Data Analysis in HDX-MS

Back exchange is unavoidable in HDX-MS, especially during enzymatic digestion and UPLC separation. To correct for back exchange, two types of reference samples have to be prepared: a protein sample representing the maximum possible HDX uptake ( $m_{100}$ ), and a sample under the quenched condition that was exposed to deuterated at with minimum time ( $m_0$ ). The deuteration percentage for each time point  $\%D(t)$  can be corrected by the following equation:

$$\%D(t) = \frac{m_t - m_0}{m_{100} - m_0} \times 100\%$$

**Equation 1.19**

Where  $m_t$  is the mass of peptide at a specific incubation time and temperature.

The deuteration behavior of a peptide with  $N$  non-proline residues covering amino acids  $k$  to  $(k+N-1)$  can be described as

$$\%D(t) = \frac{1}{(N-2)} \sum_{i=(k+2)}^N [1 - \exp(-k_{HDX,i} \times t)]$$

**Equation 1.20**

where  $k_{HDX}$  is the HDX rate constant of NH  $i$ . Summation starts at  $k+2$  because the first two residues undergo complete back exchange during LC. From typical peptide-resolved data it is not possible to extract  $k_{HDX,i}$  for each individual amide. Instead, fits usually employ one or two exponentials that yield average  $k_{HDX}$  values.

## 1.6 Scope of Thesis

Understanding the properties of mAbs and other proteins requires a comprehensive characterization of their structures and dynamics, their thermodynamic properties, unfolding behavior and aggregation propensity. The aim of the current thesis is to expand the analytical capabilities of HDX-MS to address all these topics.

Chapter 2 develops an HDX-MS framework that considers how elevated temperatures affect protein structure and dynamics. In typical HDX-MS experiments, a native protein is incubated in D<sub>2</sub>O at a constant temperature, and aliquots are removed at selected time points. In addition to time-dependent HDX-MS, there is a rapidly growing interest in conducting these measurements in a temperature-dependent fashion. Temperature-dependent data can yield a much more comprehensive view of protein dynamics. We were able to quantitatively describe the temperature- and time-dependent HDX behaviour of the model protein Mb.

Many challenges remain when it comes to understanding the mechanisms whereby proteins undergo thermal aggregation. Chapter 3 focuses on the aggregation behavior of Mb. We examined the response of Mb to heating to gain insights into the interplay of unfolding and aggregation. Specifically, we attempted to determine what exactly constitutes the aggregation-prone species in solution; possible candidates include the native state, partially folded conformers, and the unfolded state.

In Chapter 4, our thermodynamic model from Chapter 2 was applied to the NIST reference mAb, a complicated multi-domain system where dynamics, unfolding, and aggregation are

closely intertwined. Our experiments aimed to demonstrate that complete HDX profiles can be generated in as little as 30 s. With future automation of this workflow, it should be possible to implement temperature-dependent HDX-MS workflows for high-throughput applications in the pharmaceutical industry. Overall, it is hoped that the current work will encourage practitioners to explore the use of temperature-dependent HDX-MS, instead of being confined to traditional time-domain measurements.

## 1.7 References

- 1 Bliss, M., *Diabetes Care* **1993**, *16*, 4.
- 2 Hooper, J. A., *LymphoSign Journal* **2015**, *2*, 181.
- 3 Gilbert, S. C., M. Plebanski, S. J. Harris, C. E. M. Allsopp, R. Thomas, G. T. Layton, A. V. S. Hill, *Nature Biotechnology* **1997**, *15*, 1280.
- 4 Zheng, Z., X. Chen, X. Cai, H. Lin, J. Xu, X. Cheng, *Cell Death Discovery* **2022**, *8*, 250.
- 5 Sun, W., W. Zheng, A. Simeonov, *Am J Med Genet A* **2017**, *173*, 2307.
- 6 Anfinsen, C. B., *Science* **1973**, *181*, 223.
- 7 Englander, S. W., L. Mayne, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 8253.
- 8 Uzawa, T., S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, T. Fujisawa, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1171.
- 9 Feng, R. P., M. Gruebele, C. M. Davis, *Nat. Commun.* **2019**, *10*,
- 10 Otten, R., J. Villali, D. Kern, F. A. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 17004.
- 11 Miller, M. D., G. N. Phillips, Jr., *J Biol Chem* **2021**, *296*, 100749.
- 12 Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662.
- 13 Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, V. C. Shore, *Nature* **1960**, *185*, 422.
- 14 Scott, E. E., Q. H. Gibson, J. S. Olson, *J Biol Chem* **2001**, *276*, 5177.
- 15 Frauenfelder, H., B. H. McMahon, P. W. Fenimore, *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 8615.
- 16 Hutchinson, I. V., H. Zola, *Cellular Immunology* **1978**, *36*, 161.
- 17 Ionescu, R. M., J. Vlasak, C. Price, M. Kirchmeier, *J. Pharm. Sci.* **2008**, *97*, 1414.
- 18 Wu, Y., F. Wang, C. Shen, W. Peng, D. Li, C. Zhao, Z. Li, S. Li, Y. Bi, Y. Yang, Y. Gong, H. Xiao, Z. Fan, S. Tan, G. Wu, W. Tan, X. Lu, C. Fan, Q. Wang, Y. Liu, C. Zhang, J. Qi, G. F. Gao, F. Gao, L. Liu, *Science* **2020**, *368*, 1274.
- 19 Ecker, D. M., S. D. Jones, H. L. Levine, *Mabs* **2015**, *7*, 9.
- 20 Lambert, J. M., A. Berkenblit, *Annual review of medicine* **2018**, *69*, 191.

- 21 Ziemianowicz, D. S., D. Saltzberg, T. Pells, D. A. Crowder, C. Schrader, M. Hepburn, A. Sali, D. C. Schriemer, *Mol. Cell. Proteomics* **2021**, *20*, 12.
- 22 Ponziani, S., G. Di Vittorio, G. Pitari, A. M. Cimini, M. Ardini, R. Gentile, S. Iacobelli, G. Sala, E. Capone, D. J. Flavell, R. Ippoliti, F. Giansanti, *International Journal of Molecular Sciences* **2020**, *21*,
- 23 Fu, Z., S. Li, S. Han, C. Shi, Y. Zhang, *Signal Transduct Target Ther* **2022**, *7*, 93.
- 24 Walsh, S. J., J. D. Bargh, F. M. Dannheim, A. R. Hanby, H. Seki, A. J. Counsell, X. Ou, E. Fowler, N. Ashman, Y. Takada, A. Isidro-Llobet, J. S. Parker, J. S. Carroll, D. R. Spring, *Chem Soc Rev* **2021**, *50*, 1305.
- 25 Bross, P. F., J. Beitz, G. Chen, X. H. Chen, E. Duffy, L. Kieffer, S. Roy, R. Sridhara, A. Rahman, G. Williams, R. Pazdur, *Clinical Cancer Research* **2001**, *7*, 1490.
- 26 Norsworthy, K. J., C. W. Ko, J. E. Lee, J. Liu, C. S. John, D. Przepiorka, A. T. Farrell, R. Pazdur, *The Oncologist* **2018**, *23*, 1103.
- 27 Jin, Y., M. A. Schladetsch, X. Huang, M. J. Balunas, A. J. Wiemer, *Pharmacology & Therapeutics* **2022**, *229*, 107917.
- 28 Dill, K. A., J. L. MacCallum, *Science* **2012**, *338*, 1042.
- 29 Bakker, H. J., *Nature* **2012**, *491*, 533.
- 30 Jha, A. K., A. Colubri, K. F. Freed, T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099.
- 31 Kaya, H., H. S. Chan, *Proteins* **2003**, *52*, 524.
- 32 Southall, N. T., K. A. Dill, A. D. J. Haymett, *J. Phys. Chem. B* **2002**, *106*, 521.
- 33 England, J. L., V. S. Pande, G. Haran, *J. Am. Chem. Soc.* **2008**, *130*, 11854.
- 34 Myers, J. K., C. N. Pace, J. M. Schotz, *Protein Sci.* **1995**, *4*, 2138.
- 35 Segawa, S.-I., M. Nakayama, M. Sakane, *Biopolymers* **1981**, *20*, 1691.
- 36 Breuker, K., F. W. McLafferty, *Angew. Chem.* **2005**, *44*, 4911.
- 37 Privalov, P. L., N. N. Khechinashvili, *J. Mol. Biol.* **1974**, *86*, 665.
- 38 Pastore, A., P. A. Temussi, *Curr. Op. Struct. Biol.* **2012**, *22*, 30.
- 39 Ricchiuto, P., A. V. Brukhno, S. Auer, *J. Phys. Chem. B* **2012**, *116*, 5384.
- 40 Dobson, C. M., *Nature* **2003**, *426*, 884.

- 41 Fändrich, M., M. A. Fletcher, C. M. Dobson, *Nature* **2001**, *410*, 165.
- 42 Knowles, T. P. J., M. Vendruscolo, C. M. Dobson, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384.
- 43 Chiti, F., C. M. Dobson, *Annu. Rev. Biochem.* **2017**, *86*, 27.
- 44 Soto, C., *Trends Biochem. Sci.* **2011**, *36*, 151.
- 45 Wang, W., *International Journal of Pharmaceutics* **1999**, *185*, 129.
- 46 Wang, W., *Int J Pharm* **2005**, *289*, 1.
- 47 Chi, E. Y., S. Krishnan, T. W. Randolph, J. F. Carpenter, *Pharmaceutical Research* **2003**, *20*, 1325.
- 48 Hartl, F. U., M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* **2009**, *16*, 574.
- 49 Sreerama, N., S. Y. Venyaminov, R. W. Woody, *Analytical Biochemistry* **2000**, *287*, 243.
- 50 Kelly, S. M., T. J. Jess, N. C. Price, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005**, *1751*, 119.
- 51 Benjwal, S., S. Verma, K. H. Rohm, O. Gursky, *Protein Sci* **2006**, *15*, 635.
- 52 Reading, M., A. Luget, R. Wilson, *Thermochimica Acta* **1994**, *238*, 295.
- 53 Zhu, J., K. R. Vinothkumar, J. Hirst, *Nature* **2016**, *536*, 354.
- 54 Perez de Souza, L., S. Alseekh, F. Scossa, A. R. Fernie, *Nature Methods* **2021**, *18*, 733.
- 55 Richer, J., J. Spencer, M. Baird, *J. Chem. Educ.* **2006**, *83*, 1196.
- 56 Cai, J., J. Henion, *Journal of Chromatography A* **1995**, *703*, 667.
- 57 Konermann, L., *Science Progress* **1998**, *81*, 123.
- 58 Ma, Y.-L., H.-L. Qin, W.-J. Liu, J.-Y. Peng, L. Huang, X.-P. Zhao, Y.-Y. Cheng, *Digestive Diseases and Sciences* **2009**, *54*, 2655.
- 59 Gavriilidou, A. F. M., F. P. Holding, D. Mayer, J. E. Coyle, D. B. Veprintse, R. Zenobi, *Biochemistry* **2018**, *57*, 1685.
- 60 Awad, H., M. M. Khamis, A. El-Aneed, *Applied Spectroscopy Reviews* **2015**, *50*, 158.
- 61 El-Aneed, A., A. Cohen, J. Banoub, *Appl. Spectrosc. Rev.* **2009**, *44*, 210.

- 62 March, R., J. F. J. Todd, *Quadrupole Ion Trap Mass Spectrometry*. Second edition ed., Editor, Wiley, New York ; Toronto, **2005**.
- 63 Girard, E., L. P. Morin, S. Robert, J. Boivin, V. Vincent, F. Garofolo, *J. Chromatogr. Sci.* **2010**, *48*, 778.
- 64 Geib, T., L. Sleno, R. A. Hall, C. S. Stokes, D. A. Volmer, *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1404.
- 65 Voinov, V. G., M. L. Deinzer, J. S. Beckman, D. F. Barofsky, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 607.
- 66 Nettels, D., I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2655.
- 67 van den Heuvel, R. H. H., E. van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. van der Oost, J. Hoyes, A. J. R. Heck, *Analytical Chemistry* **2006**, *78*, 7473.
- 68 Matuszewski, B. K., M. L. Constanzer, C. M. Chavez-Eng, *Analytical Chemistry* **2003**, *75*, 3019.
- 69 Skinner, S. P., G. Radou, R. Tuma, J. J. Houwing-Duistermaat, E. Paci, *Biophys. J.* **2019**, *116*, 1194.
- 70 Englander, S. W., L. M. Bai, T. R. Sosnick, *Protein Sci.* **1997**, *6*, 1101.
- 71 Konermann, L., J. Pan, Y. Liu, *Chem. Soc. Rev.* **2011**, *40*, 1224.
- 72 Vahidi, S., B. B. Stocks, Y. Liaghati-Mobarhan, L. Konermann, *Anal. Chem.* **2013**, *85*, 8618.
- 73 Bai, Y., J. S. Milne, L. Mayne, S. W. Englander, *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.
- 74 Connelly, L., H. Jang, F. T. Arce, S. Ramachandran, B. L. Kagan, R. Nussinov, R. Lal, *Biochemistry* **2012**, *51*, 3031.
- 75 Nguyen, D., L. Mayne, M. C. Phillips, S. W. Englander, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1936.
- 76 Swint-Kruse, L., A. D. Robertson, *Biochemistry* **1996**, *35*, 171.
- 77 Distler, U., J. Kuharev, P. Navarro, Y. Levin, H. Schild, S. Tenzer, *Nat. Methods* **2014**, *11*, 167.
- 78 Egertson, J. D., A. Kuehn, G. E. Merrihew, N. W. Bateman, B. X. MacLean, Y. S. Ting, J. D. Canterbury, D. M. Marsh, M. Kellmann, V. Zabrouskov, C. C. Wu, M. J. MacCoss, *Nature Methods* **2013**, *10*, 744.

## 2 Chapter 2. Analysis of Temperature-Dependent H/D Exchange Mass Spectrometry Experiments

### 2.1 Introduction

Native proteins undergo incessant thermal motions, from local fluctuations to global unfolding/refolding.<sup>1,2</sup> These dynamics reflect the fact proteins continuously explore their conformational space. The population of each conformer depends on its free energy,<sup>2</sup> while interconversion rates are governed by activation barriers.<sup>3</sup> Protein dynamics are linked to biological function such as catalysis,<sup>4</sup> energy conversion,<sup>5</sup> and signaling.<sup>6</sup> However, protein fluctuations can also generate structures that act as gateway to cytotoxic aggregates.<sup>7,8</sup>

Hydrogen/deuterium exchange (HDX) mass spectrometry (MS) is one of the most widely used tools for interrogating protein dynamics.<sup>5, 8-15</sup> HDX-MS complements other techniques such as Förster resonance energy transfer, time-resolved X-ray diffraction, and NMR spin relaxation studies.<sup>16</sup> HDX-MS monitors the deuteration of backbone NH groups upon exposure to D<sub>2</sub>O. In native proteins most of these sites are engaged in NH···OC hydrogen bonds. According to the widely accepted Linderstrøm-Lang model (equation 2.1)<sup>2,17</sup> native state HDX is mediated by opening/closing (unfolding/refolding) fluctuations that transiently disrupt H-bonds.



**Equation 2.1**

Here,  $k_{op}$  and  $k_{cl}$  are the opening and closing rate constants, and  $k_{ch}$  is the “chemical” rate constant.<sup>18</sup> HDX usually proceeds in the EX2 limit ( $k_{cl} \gg k_{ch}$ ),<sup>2</sup> where each NH undergoes many opening/closing transition before it is deuterated. The overall HDX rate constant in this case is

$$k_{HDX} = K_{op} \times k_{ch}$$

**Equation 2.2**

where  $K_{op} = k_{op}/k_{cl}$ . Hence, HDX kinetics are governed by  $NH_{open}$  sites, even though the population of these sites tends to be very low (typically  $\ll 1\%$ ).<sup>2, 19, 20</sup> Dynamic regions (larger  $K_{op}$ ) exhibit faster HDX, while labeling of rigid segments (smaller  $K_{op}$ ) proceeds more slowly.

In conventional HDX-MS a protein is incubated in D<sub>2</sub>O at constant temperature. Aliquots sampled at various time points are subjected to proteolysis, followed by LC/MS to uncover the deuteration percentage (%D) of individual peptides. In addition to these conventional measurements, there is growing interest in temperature-dependent HDX-MS. The latter can provide a more comprehensive view of protein dynamics.<sup>21-29</sup> Of particular importance is the characterization of thermally stressed protein drugs (such as therapeutic antibodies) to assess their stability and aggregation propensity.<sup>15, 30, 31</sup>

It is well known that elevated temperatures tend to enhance HDX,<sup>21-29</sup> but the origins of this effect are non-trivial. Temperature controls HDX via two avenues. (1) The labeling chemistry ( $k_{ch}$  in equation 2.1) accelerates quasi-exponentially with temperature<sup>18</sup> due to a combination of Arrhenius behavior and changes in the concentration of OD<sup>-</sup> catalyst.<sup>32, 33</sup> (2) Temperature governs the Boltzmann populations of  $NH_{open}$  states,<sup>2</sup> and it alters the protein energy landscape.<sup>34</sup> HDX-MS aims to uncover protein behavior (contribution 2), but unfortunately this aspect tends to be masked by temperature-dependent changes of  $k_{ch}$  (contribution 1). Preliminary steps have been taken to unravel this problem,<sup>20, 25, 35-37</sup> but a comprehensive strategy for analyzing HDX-MS data as a function of temperature ( $T$ ) and time ( $t$ ) is still lacking. The current work fills this void by deconvoluting experimental HDX-MS data into the two aforementioned contributions.

Protein stability studies usually rely on a two-state approximation involving the native state N and the unfolded state U.<sup>34</sup> The free energy of the global N  $\leftrightarrow$  U equilibrium is

$$\Delta G_{glob} = \Delta H_{glob} - T \Delta S_{glob}$$

**Equation 2.3**

N is stable as long as  $\Delta G_{glob} > 0$ . Simple analyses often assume that enthalpy ( $\Delta H_{glob} > 0$ ) and entropy ( $\Delta S_{glob} > 0$ ) are constant, such that  $\Delta G_{glob}$  depends linearly on  $T$ .<sup>38</sup> Heating causes  $\Delta G_{glob}$  to turn negative at the melting temperature  $T_m$ , thereby triggering unfolding. A more thorough treatment of equation 2.3 has to consider  $\Delta C_p$ , the heat capacity difference between U and N which causes  $\Delta H_{glob}$  and  $\Delta S_{glob}$  to become  $T$ -dependent ( $\Delta C_p > 0$ )<sup>34</sup>

$$\Delta H_{glob}(T) = \Delta H_{glob}(T_m) + \Delta C_p \times (T - T_m)$$

**Equation 2.4a**

$$\Delta S_{glob}(T) = \Delta S_{glob}(T_m) + \Delta C_p \times \ln(T/T_m)$$

**Equation 2.4b**

where  $\Delta S_{glob}(T_m) = \Delta H_{glob}(T_m)/T_m$ .<sup>38</sup> Equation 2.4 implies that  $\Delta G_{glob}(T)$  is curved. As a result, proteins are most stable at an intermediate temperature. Raising the temperature beyond  $T_m$  triggers heat-induced unfolding. Cooling causes destabilization as well; depending on the magnitude of  $\Delta C_p$  this can cause cold-unfolding.<sup>39-41</sup>

Equations 2.3 and 2.4 are key pillars of protein thermodynamics.<sup>34, 38</sup> However, it is undisputed that two-state  $N \leftrightarrow U$  models do not fully capture the protein behavior, especially at ambient  $T$  where local fluctuations dominate over global unfolding/refolding.<sup>1, 2</sup> For addressing this deficiency and for interpreting  $T$ -dependent HDX-MS data it is necessary to combine the Linderstrøm-Lang model (equations 2.1, 2.2)<sup>2, 17</sup> with the thermodynamic principles expressed in equations 2.3 and 2.4.

Building on the aforementioned ideas, the current work devises a strategy for analyzing  $T$ -dependent HDX-MS data. Like many previous studies on protein thermodynamics,<sup>39</sup> folding,<sup>42</sup> fluctuations,<sup>1</sup> and aggregation,<sup>43</sup> we chose the heme protein myoglobin (Mb) as model system. Mb has a globular native structure, where a hydrophobic core is surrounded by solvent-exposed polar and charged residues.<sup>44</sup> We tracked the HDX response of Mb to

changes in  $T$ , and we captured the resulting data using a comprehensive  $T$ - and  $t$ -dependent HDX data analysis framework.

## 2.2 Methods

### 2.2.1 Materials

Horse-heart ferri-Mb (Sigma, St. Louis, MO) samples were centrifuged to remove small amounts of insoluble debris. All solutions contained 50 mM sodium phosphate buffer and 100 mM NaCl at a pH meter reading adjusted to 7.2 (corresponding to pD 7.6 for D<sub>2</sub>O-based solutions).<sup>18, 45</sup> This value was  $T$ -independent with deviations of less than  $\pm 0.1$ , consistent with previous reports on the temperature stability of phosphate buffer.<sup>46</sup>

### 2.2.2 Optical Experiments

Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter (Easton, MD) with a 1 mm cuvette using 5  $\mu$ M Mb between 20 °C and 96 °C. Unfolding profiles were generated by monitoring the CD signal at 222 nm which is characteristic of  $\alpha$ -helical secondary structure.<sup>47</sup> These CD<sub>222</sub> profiles were analyzed by using<sup>38, 48</sup>

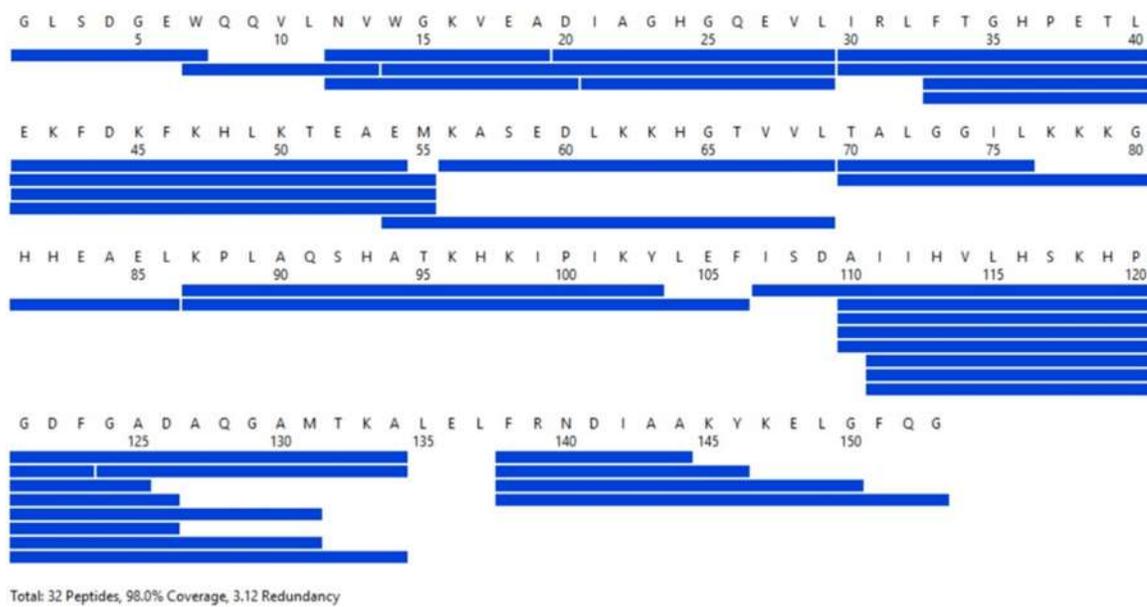
$$CD_{222} = \frac{(y_N + m_N T) + (y_U + m_U T) \exp(-\Delta G_{glob}/RT)}{1 + \exp(-\Delta G_{glob}/RT)}$$

**Equation 2.5**

where  $(y_N + m_N T)$  and  $(y_U + m_U T)$  are the sloped pre- and post-transition baselines, respectively, with  $\Delta G_{glob} = \Delta H_{glob}(1 - T/T_m)$ .

### 2.2.3 H/D Exchange Mass Spectrometry

HDX was performed by adding 50  $\mu\text{M}$  Mb solution in  $\text{H}_2\text{O}$  buffer to  $\text{D}_2\text{O}$  labeling buffer in a 1:9 ratio, for a final protein concentration of 5  $\mu\text{M}$  in 90%  $\text{D}_2\text{O}$ . All measurements were performed with careful temperature control, by placing the samples in Eppendorf tubes that were immersed in a water bath during labeling. An ice/water mix was used for measurements at 0  $^\circ\text{C}$ , while HDX between 23  $^\circ\text{C}$  and 80  $^\circ\text{C}$  was performed by employing a heated circulating bath. Prior to initiating HDX,  $\text{D}_2\text{O}$  labeling buffer was pre-equilibrated at the desired temperature, while Mb was kept at 23  $^\circ\text{C}$  to avoid aggregation in the stock solution. Two types of HDX experiments were performed. (i) [variable  $T$ , constant  $t$ ] was conducted by using a deuteration time of  $t = 30$  s. (ii) For [constant  $T$ , variable  $t$ ] aliquots were removed at  $t = 30$  s, 10 min, and 100 min. HDX was quenched by mixing with HCl, resulting in a pH meter reading of 2.4. This was followed by flash freezing and storage in  $\text{N}_2(l)$ . For analysis, the samples were rapidly thawed and injected into an Acquity HDX-UPLC (Waters, Milford, MA). Digestion was performed on an immobilized pepsin column (Thermo Fisher) at 15  $^\circ\text{C}$ . Peptides were separated on a 1.7  $\mu\text{m}$  BEH130 C18 2.1 mm  $\times$  100 mm column using a 20 min water/acetonitrile gradient in the presence of 0.1% formic acid at  $\sim 0$   $^\circ\text{C}$ . The sequence coverage was 98% (Figure 2.1).



**Figure 2.1.** Peptic digestion map, showing the HDX sequence coverage (image generated by Waters DynamX).

Blanks were injected to prevent carryover. In addition, the pepsin column was washed with 1.5 M guanidine hydrochloride in water/acetonitrile/formic acid (95.2/4/0.8) after each digestion step. The UPLC was coupled to a Waters Synapt G2 mass spectrometer with a lock spray dual electrospray source. The identity of each peptide was confirmed by MS<sup>E</sup> on non-deuterated samples with data analysis by Waters PLGS 2.5.3, based on the known Mb sequence (pdb 1wla<sup>44</sup>). The capillary voltage and desolvation temperature were +2.8 kV and 250 °C respectively. The centroid mass of each peptide isotope distribution was calculated using DynamX 3.0 (Waters) and converted to percent deuteration (%D) according to  $\%D = [(m - m_0)/(m_{100} - m_0)] \times 100\%$ , where  $m$  is the centroid  $m/z$  for the peptide of interest.  $m_0$  and  $m_{100}$  correspond to minimally and fully deuterated controls, respectively. The former were prepared by adding Mb to pre-quenched D<sub>2</sub>O buffer, followed by flash freezing.  $m_{100}$  samples were prepared similar to  $m_0$  samples, except that they were incubated for 3 days at 37 °C. Back-exchange levels determined from  $m_{100}$  samples were (38 ± 14)%, similar to previous reports.<sup>35, 49</sup> All %D values are averages of three independent replicates; error bars represent standard deviations.

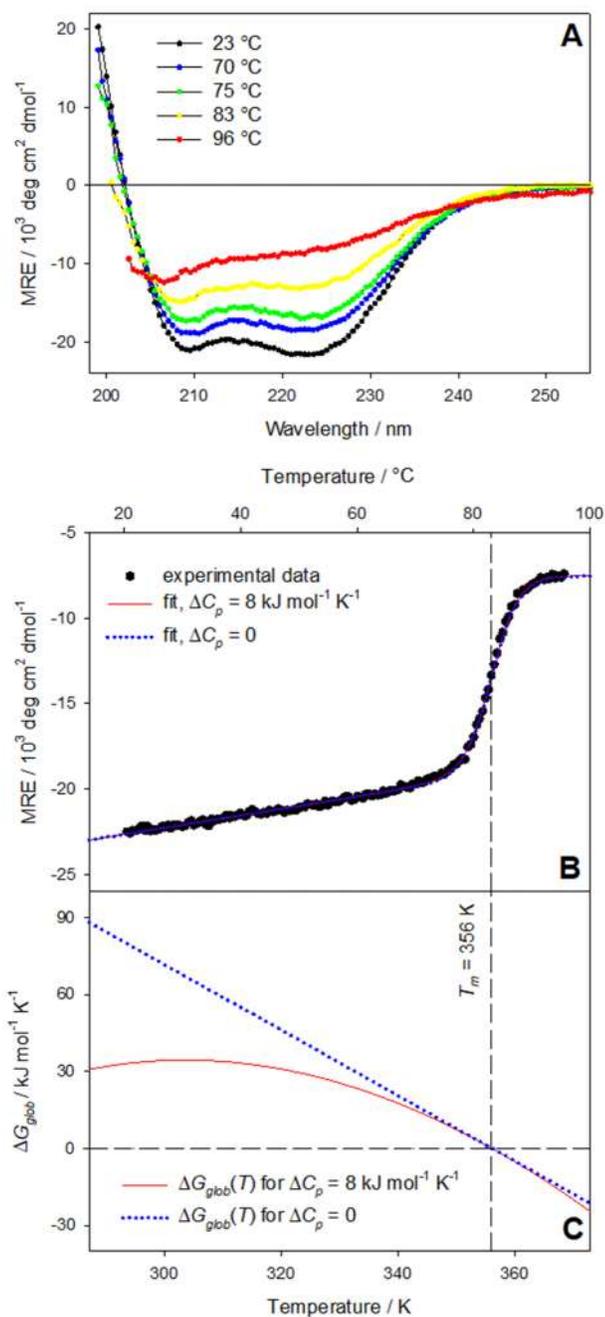
Elevated temperatures may cause protein aggregation<sup>50</sup> which would complicate the interpretation of HDX and CD experiments. Aggregation can be suppressed by working with dilute solutions.<sup>42, 51</sup> For the experiments of this work we therefore used a relatively low Mb concentration (5 μM), where aggregation was negligible for at least 20 min even when heated to 358 K (85 °C). This assertion is based on aggregation assays on Mb that had been heated for different  $t$  and  $T$ , followed by centrifugation and UV-Vis analyses of the supernatant.

## 2.3 Results and Discussion

### 2.3.1 Spectroscopic Characterization of Global Unfolding

Heat-induced unfolding of Mb was probed by CD spectroscopy. Spectra acquired at different  $T$  intersected at  $\sim 204$  nm (Figure 2.2A). This isodichroic point is consistent with a global N  $\leftrightarrow$  U transition,<sup>39,52</sup> justifying the application of equations 2.3 and 2.4. Thermal unfolding profiles were recorded by monitoring CD<sub>222</sub> at different heating rates (1 °C min<sup>-1</sup> and 4 °C min<sup>-1</sup>) in both H<sub>2</sub>O and in D<sub>2</sub>O. All of the resulting CD<sub>222</sub> profiles were very similar, with  $T_m = 356.2 \pm 0.6$  K and  $\Delta H_{glob}(T_m) = 453 \pm 20$  kJ mol<sup>-1</sup> (Figure 2.2B).

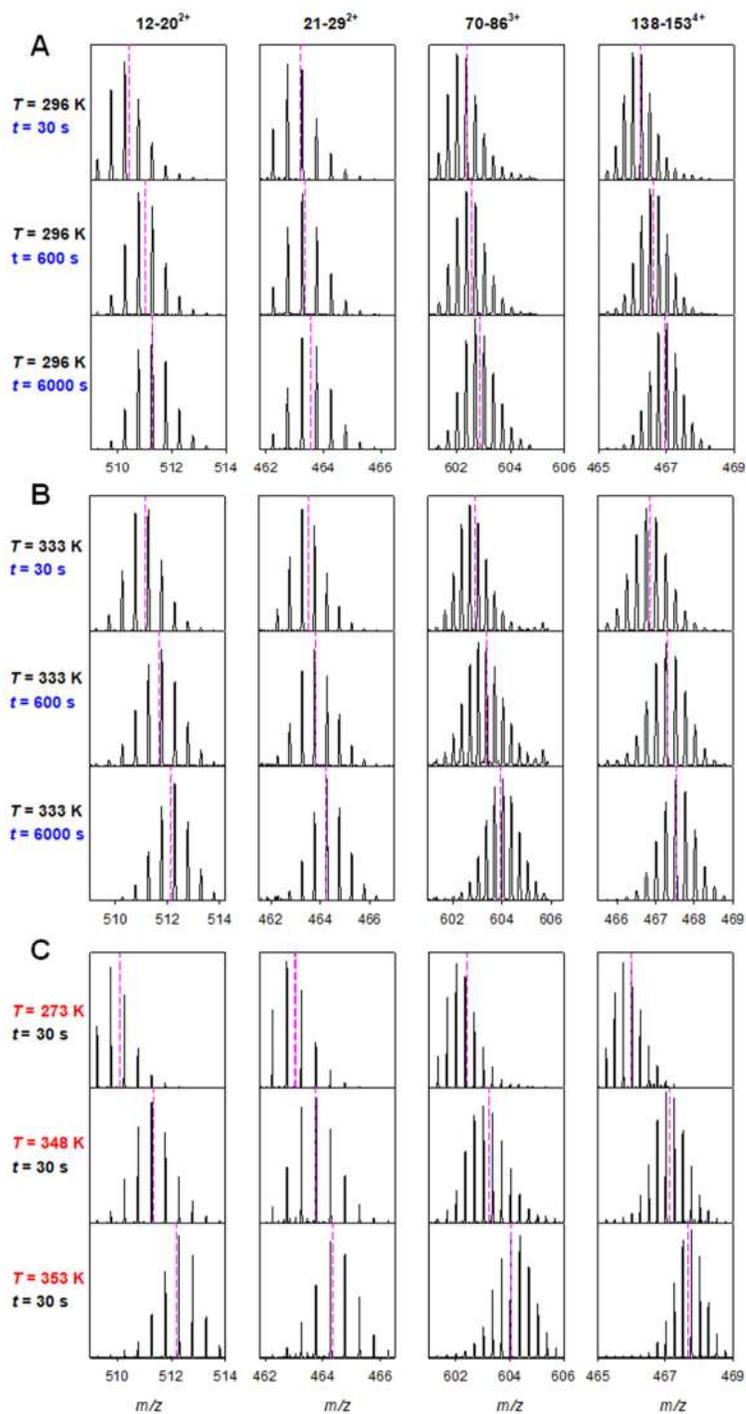
As is common practice in optical melting experiments,<sup>38</sup> the aforementioned analysis assumed that  $\Delta C_p = 0$ , such that  $\Delta G_{glob}$  depends linearly on temperature (Figure 2.2C, dotted line, equation 2.3). Alternatively, one can perform an analysis with  $\Delta C_p = 8$  kJ mol<sup>-1</sup> K<sup>-1</sup>, a value determined by differential scanning calorimetry (DSC).<sup>39,53</sup> This  $\Delta C_p$  value with the  $T_m$  and  $\Delta H_{glob}(T_m)$  noted above corresponds to a curved  $\Delta G_{glob}$  profile (Figure 2.2C, solid line, equation 2.4). Both  $\Delta G_{glob}$  profiles match the experimental data equally well (Figure 2.2B), but the curved profile in Figure 2.2C is more realistic because it takes into account the DSC-derived  $\Delta C_p$  value.<sup>39,53</sup> This curved  $\Delta G_{glob}$  profile will serve as starting point for the analysis of  $T$ -dependent HDX-MS data.



**Figure 2.2** Thermal unfolding of 5  $\mu\text{M}$  Mb studied by CD spectroscopy. (A) CD spectra acquired at various temperatures, displaying mean residue ellipticity (MRE) vs. wavelength. (B) Unfolding profile monitored at 222 nm, heating rate 1  $^\circ\text{C min}^{-1}$ . Black symbols are experimental data, colored lines represent fits for two different values of  $\Delta C_p$ . (C) Free energy  $\Delta G_{\text{glob}}(T)$  of the  $N \leftrightarrow U$  equilibrium according to equation 2.3 and 2.4; parameters were extracted from the fits of panel B.

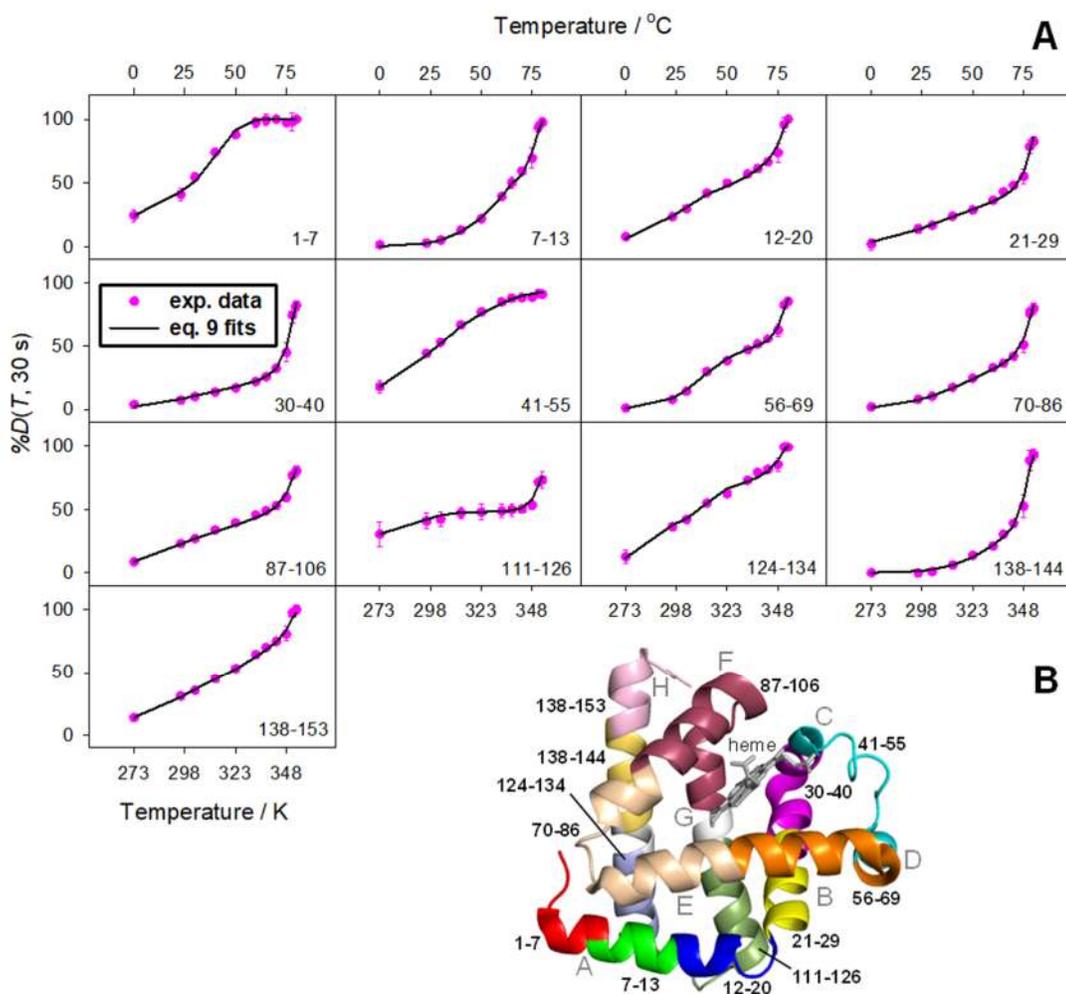
### 2.3.2 Temperature- and Time-Dependent HDX Experiments

Figures 2.3A and 2.3B exemplify peptide isotope distributions acquired after exposing Mb to D<sub>2</sub>O for different times at  $T = 296$  K (23 °C). With increasing  $t$  the spectra shifted to higher  $m/z$  because of backbone deuteration. A complementary perspective was obtained by conducting HDX in a  $T$ -dependent fashion between 273 K and 353 K, while keeping the labeling time  $t$  constant. We chose  $t = 30$  s, which resulted in a wide dynamic range of %D value. Mass spectra measured in this way demonstrate that increased  $T$  dramatically enhances the extent of deuteration (Figure 2.3C). In fact, the  $T$ -induced spectral changes were much larger than those seen in the  $t$ -dependent data of Figure 2.3A and 2.3B.



**Figure 2.3** HDX-MS isotope distributions of selected peptic peptides. (A) Data acquired after different labeling time intervals  $t$  at a constant temperature of  $T = 296\text{ K}$  ( $23\text{ }^{\circ}\text{C}$ ). (B) Same as in panel A, but for  $T = 333\text{ K}$  ( $60\text{ }^{\circ}\text{C}$ ). (C) Data acquired at different temperatures  $T$  for a constant labeling time of  $t = 30\text{ s}$ . Vertical dashed lines indicate centroid  $m/z$  values.

$T$ -dependent  $\%D$  profiles revealed interesting differences for the various Mb regions (Figure 2.4). Many of the profiles showed very shallow slopes between 273 K and  $\sim 340$  K, e.g., 30-40 and 111-126. These shallow slopes are surprising, considering that  $k_{ch}$  increases quasi-exponentially with  $T$ .<sup>18</sup> Several profiles had an upwards kink at 348 K. This feature is most pronounced for 111-126, but it also affects 12-20, 21-29, 30-40, 56-69, 87-106, and 124-134. Figure 2.13 summarizes all  $T$ - and  $t$ -dependent data, including peptides not shown in Figure 2.4.



**Figure 2.4** (A) Deuteration percentage  $\%D$  as a function of temperature  $T$  for a labeling time of  $t = 30\text{ s}$ . The sequence range of each peptide is indicated. Pink symbols represent experimental data, black lines are fits based on equation 2.10. (B) Locations of panel A peptides in the Mb crystal structure. The cartoon also highlights the eight  $\alpha$ -helices A-H.

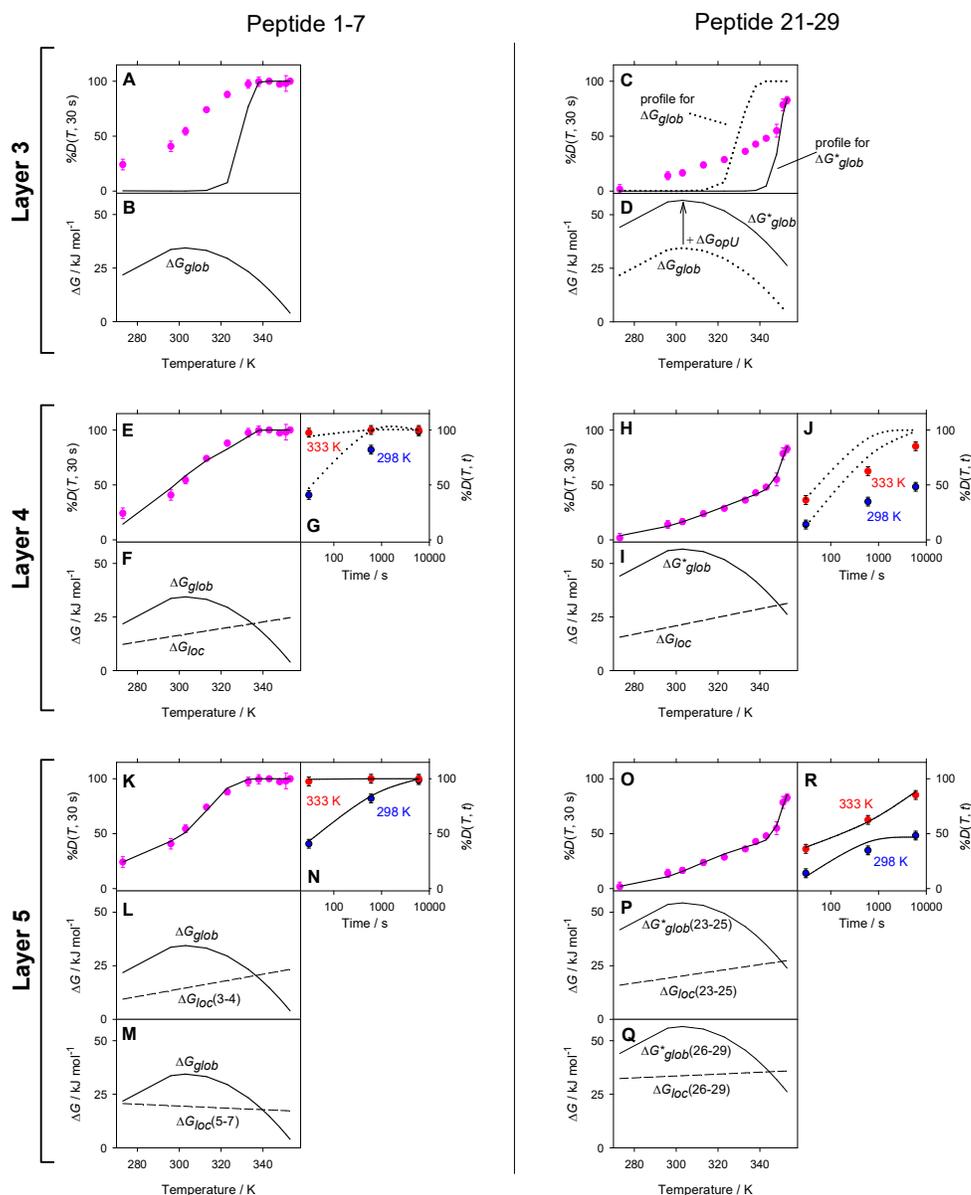
### 2.3.3 Developing a Temperature-Dependent HDX-MS Model

We now examine how  $T$ -dependent changes in labeling chemistry and protein dynamics manifest themselves in the HDX behavior, ultimately yielding a model that accounts for all the experimental data. The deuteration behavior of a peptide with  $N$  non-proline residues covering amino acids  $k$  to  $(k+N-1)$  can be described as<sup>54-56</sup>

$$\%D(T, t) = \frac{1}{(N - 2)} \sum_{i=(k+2)}^N [1 - \exp(-k_{HDX,i}(T) \times t)]$$

**Equation 2.6**

where  $k_{HDX,i}(T)$  is the HDX rate constant of NH  $i$ . Summation starts at  $k+2$  because the first two residues undergo complete back exchange during LC.<sup>18, 35, 57</sup> Layer by layer, the following sections describe a strategy to capture the behavior of  $k_{HDX,i}(T)$  in equation 2.6 (Figure 2.5). We were guided by the tenets that a good model will (i) be able to quantitatively match the experimental data, (ii) be as simple as possible, and (iii) have a minimum number of adjustable parameters.



**Figure 2.5** Layer-by-layer development of a model for interpreting HDX-MS data recorded as a function of temperature  $T$  and labeling time  $t$ . Circles represent experimental  $\%D(T, t)$  data for peptides 1-7 and 21-29. Note how the quality of the fits improves for  $A \rightarrow K/N$  and  $C \rightarrow O/R$ . For Layers 1 and 2, see text. Layer 3: Global dynamics with and without  $\Delta G_{opU}$ . Layer 4: Inclusion of simple local dynamics. Layer 5: Inclusion of local dynamics where each peptide was dissected into two segments that undergo independent opening/closing. Layer 5 parameters for Peptide 1-7:  $\Delta G_{opU(3-4)} = 0$ ,  $\Delta H_{loc(3-4)} = -38$ ,  $\Delta S_{loc(3-4)} = -174$ ,  $\Delta G_{opU(5-7)} = 0$ ,  $\Delta H_{loc(5-7)} = 33$ ,  $\Delta S_{loc(5-7)} = 43$ . For Peptide 21-29:  $\Delta G_{opU(23-25)} = 20$ ,  $\Delta H_{loc(23-25)} = -23$ ,  $\Delta S_{loc(23-25)} = -143$ ,  $\Delta G_{opU(26-29)} = 22$ ,  $\Delta H_{loc(26-29)} = 20$ ,  $\Delta S_{loc(26-29)} = -44$  ( $\Delta G_{opU}$  and  $\Delta H_{loc}$  in  $\text{kJ mol}^{-1}$ ;  $\Delta S_{loc}$  in  $\text{J mol}^{-1} \text{K}^{-1}$ ).

### 2.3.3.1 Layer 1: Temperature Dependence of $k_{ch}$ .

Before focusing on protein dynamics, we examine the chemical rate constant  $k_{ch}$  that governs the deuteration of  $NH_{open}$  (equation 2.1).  $k_{ch}$  depends on  $T$ , pD, and neighboring side chains.<sup>18, 58</sup> Its  $T$ -dependence is often described using the Arrhenius expression<sup>18</sup>

$$k_{ch}(T) = k_{ch_{298}} \times \exp \left( -\frac{E_{a_{eff}}}{R} (T^{-1} - [298 \text{ K}]^{-1}) \right)$$

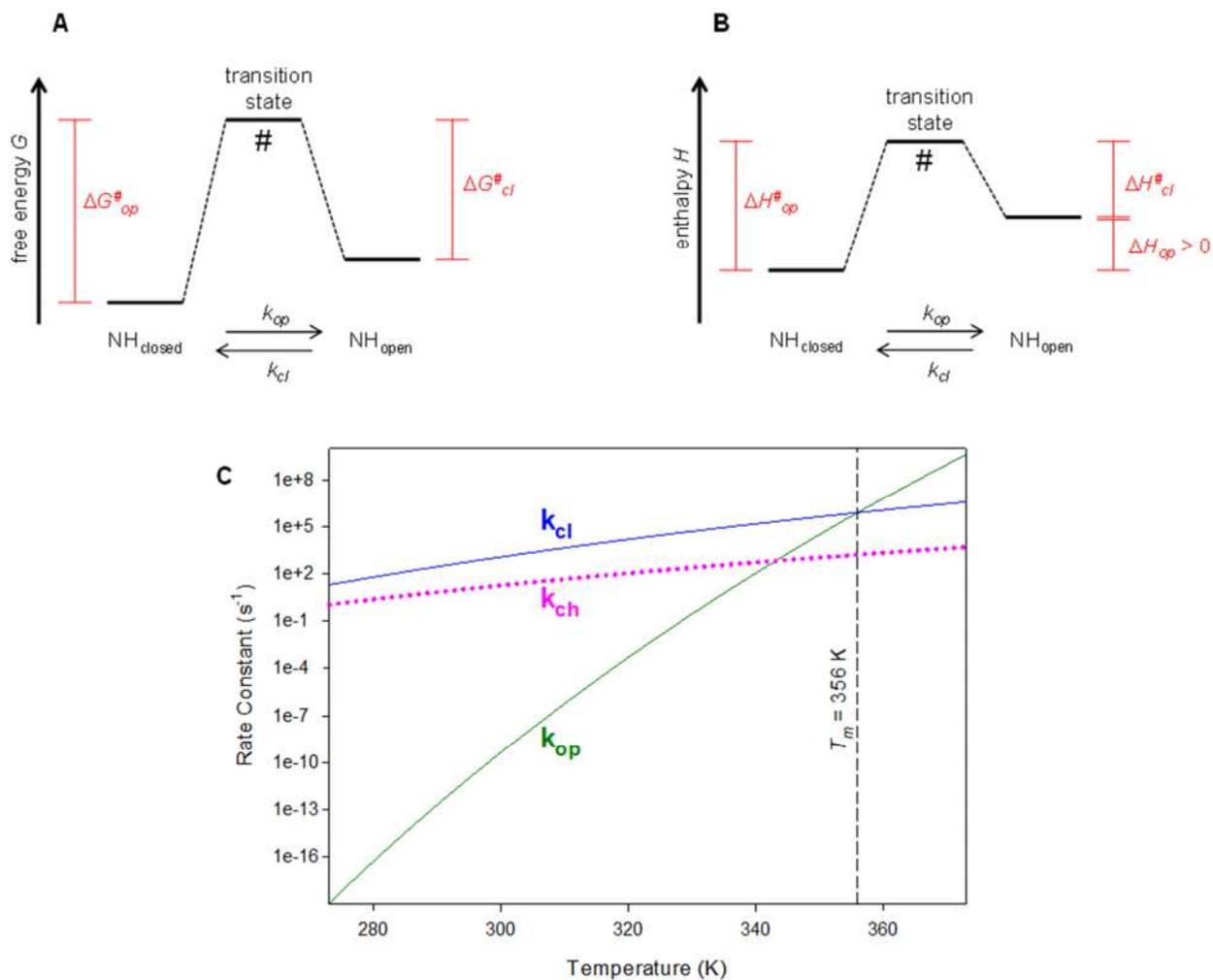
**Equation 2.7a**

where  $k_{ch_{298}}$  is a reference value for 298 K, and the effective activation energy  $E_{a_{eff}}$  is 71.1 kJ mol<sup>-1</sup> (Figures 2.6, 2.7).<sup>58</sup> However, equation 2.7a is only an approximation.<sup>18</sup> The actual  $E_a$  for chemical exchange is 12.6 kJ mol<sup>-1</sup>.<sup>18</sup> The  $T$ -dependence of  $k_{ch}$  arises largely from changes in  $[OD^-]$  as governed by the  $T$ -dependent ionization constant  $K_{D2O} = [D^+][OD^-]$ ,<sup>32</sup> keeping in mind that HDX in near-neutral solution is  $OD^-$  catalyzed.<sup>33</sup> The analyses below use the more accurate expression

$$k_{ch}(T) = k_{B_{298}} \times \exp \left( -\frac{E_a}{R} (T^{-1} - [298 \text{ K}]^{-1}) \right) \times [OD^-](T)$$

**Equation 2.7b**

with the reference value  $k_{B_{298}}$  for base-catalyzed exchange (Figure 2.7).<sup>18</sup> Between 0 °C and ~60 °C the  $k_{ch}(T)$  profiles predicted by equations 2.7a and 2.7b are similar. At 80 °C the equation 2.7a value is 34% larger than that of equation 2.7b; at higher  $T$  the differences become more pronounced (Figure 2.7).



**Figure 2.6.** Temperature dependence of backbone amide H-bond opening/closing, and its relationship to the temperature dependence of  $k_{ch}$ . (A) Transition state theory model of the  $NH_{closed} \leftrightarrow NH_{open}$  equilibrium.<sup>48, 59</sup>  $\Delta G_{op}^{\#}$  and  $\Delta G_{cl}^{\#}$  are the activation barrier heights for opening and closing, respectively. The corresponding rate constants are

$$k_{op} = \kappa \frac{kT}{h} \exp\left(-\frac{\Delta G_{op}^{\#}}{RT}\right) \quad k_{cl} = \kappa \frac{kT}{h} \exp\left(-\frac{\Delta G_{cl}^{\#}}{RT}\right)$$

( $k$  = Boltzmann constant,  $h$  = Planck constant,  $R$  = gas constant,  $T$  = temperature, and  $\kappa$  = transmission coefficient). The activation free energies can be dissected into enthalpic and entropic contributions according to  $\Delta G^{\#} = \Delta H^{\#} - T\Delta S^{\#}$ , such that

**Figure 2.6 Caption (continued):**

$$k_{op} = C_{op} T \exp\left(-\frac{\Delta H_{op}^{\#}}{RT}\right) \quad k_{cl} = C_{cl} T \exp\left(-\frac{\Delta H_{cl}^{\#}}{RT}\right)$$

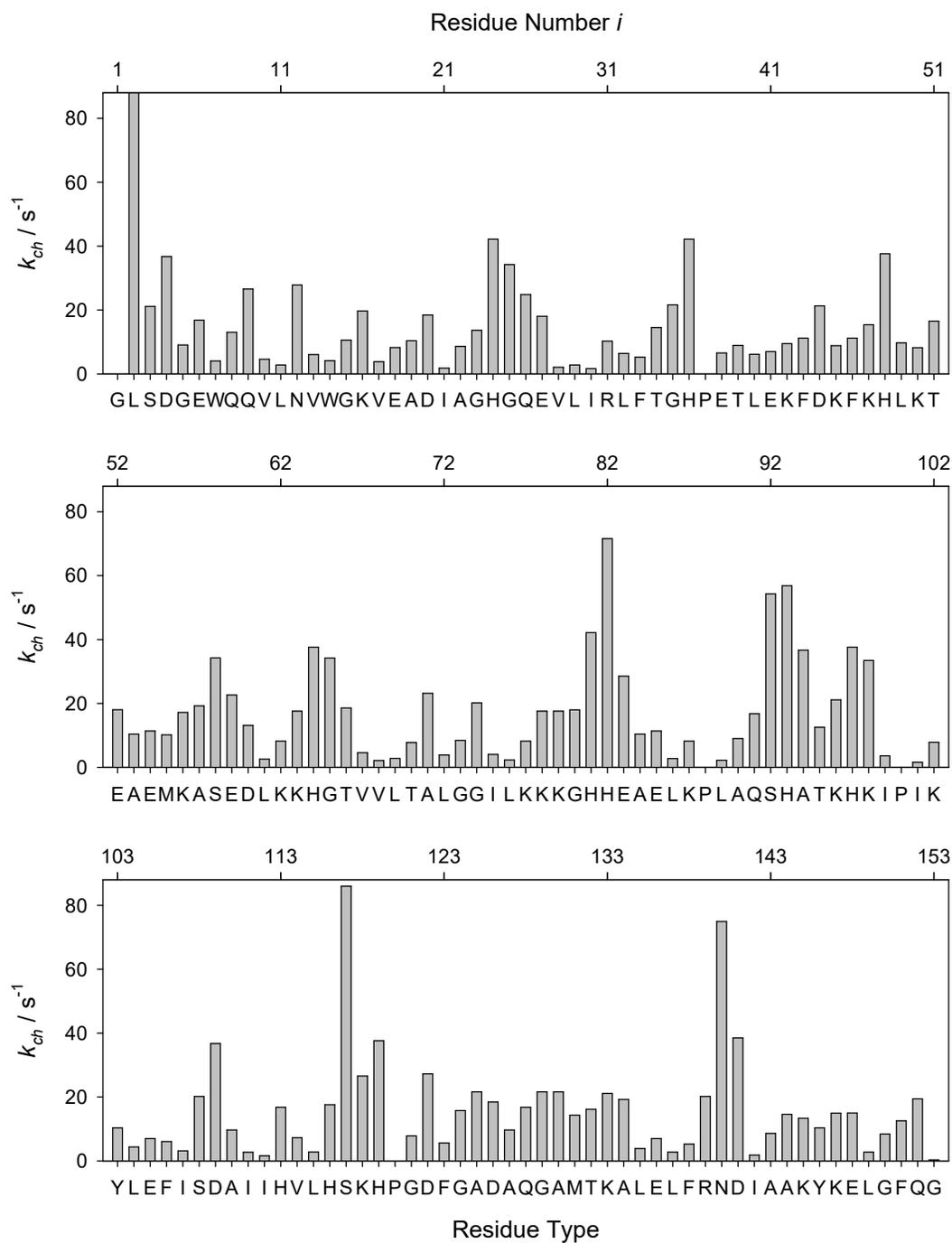
$$\text{with } C_{op} = \kappa \frac{k_B}{h} \exp\left(\frac{\Delta S_{op}^{\#}}{R}\right) \quad C_{cl} = \kappa \frac{k_B}{h} \exp\left(\frac{\Delta S_{cl}^{\#}}{R}\right)$$

When expressed in this way, it becomes clear that the temperature dependence of  $k_{op}$  and  $k_{cl}$  is governed by the activation enthalpies  $\Delta H_{op}^{\#}$  and  $\Delta H_{cl}^{\#}$ , while the entropy terms can be incorporated into the  $T$ -independent prefactor. Heat is required to dissociate H-bonds ( $\Delta H_{op} > 0$ ). This implies  $\Delta H_{op}^{\#} > \Delta H_{cl}^{\#}$  as illustrated in (B), causing  $k_{op}$  to depend more strongly on temperature than  $k_{cl}$ .

Panel (C) illustrates how  $k_{op}$  and  $k_{cl}$  change with temperature. The numerical parameters were chosen to resemble the global Mb unfolding data of Figure 2.2, i.e.,  $\Delta H_{op} = 453 \text{ kJ mol}^{-1}$  and  $\Delta H_{cl}^{\#} = 100 \text{ kJ mol}^{-1}$  (estimated from literature data<sup>60, 61</sup>) such that  $\Delta H_{op}^{\#} = 553 \text{ kJ mol}^{-1}$ .  $C_{cl}$  was arbitrarily chosen as  $10^{18} \text{ s}^{-1} \text{ K}^{-1}$  to ensure EX2 conditions with  $k_{cl} = 20 \text{ s}^{-1}$  at 273 K. This determines the value of  $C_{op} = 3 \times 10^{84} \text{ s}^{-1} \text{ K}^{-1}$  to ensure that  $k_{op} = k_{cl}$  at  $T_m = 356 \text{ K}$ .

Also included in (C) is a temperature-dependent  $k_{ch}$  profile, calculated using the Arrhenius parameters of ref.<sup>18</sup> for poly-alanine at pD = 7.6 (equation 2.8a).

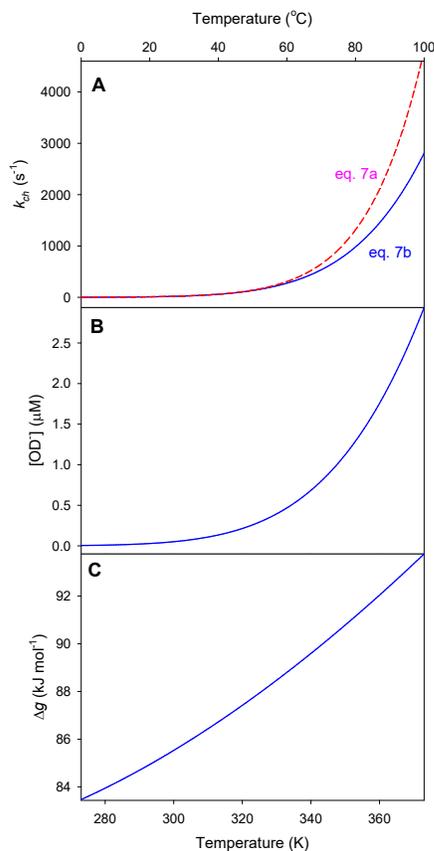
Key conclusion from the data presented in this Figure: A protein that exhibits EX2 behavior ( $k_{cl} \gg k_{ch}$ ) at low temperature is likely to remain in the EX2 regime when the temperature is raised. This is illustrated by in panel (C), where  $k_{cl}$  remains at least two orders of magnitude above  $k_{ch}$  throughout the entire range from 273 K to 373 K.



**Figure 2.7.**  $k_{ch,i}$  values for backbone NH sites along the Mb sequence, for pD 7.6 and 298 K. These values were calculated using Excel files from the Englander Laboratory (<http://hx2.med.upenn.edu/download.html>).<sup>58</sup>

### 2.3.3.2 Layer 2: Verifying the EX2 Regime

The interpretation of HDX-MS data requires different approaches, depending on whether deuteration proceeds in the EX1 or in the EX2 regime.<sup>5, 8-15</sup> EX2, defined as  $k_{cl} \gg k_{ch}$  (equation 2.1), represents the most common scenario; it is associated with unimodal isotope distribution that gradually shift to higher mass.<sup>2, 12, 19, 20</sup> EX1 ( $k_{cl} \ll k_{ch}$ ) is less common. Cooperative EX1 dynamics cause bimodal isotope distributions.<sup>12, 19, 20</sup> The unimodal nature of the spectra in this work confirms that HDX proceeds in the EX2 regime (Figure 2.3). Skeptics might bring up an interesting issue in this context. Because  $k_{ch}$  increases with  $T$ ,<sup>18</sup> heating might cause a transition from EX2 ( $k_{cl} \gg k_{ch}$ ) at low  $T$  to EX1 ( $k_{cl} \ll k_{ch}$ ) at high  $T$ . However, such an EX2  $\rightarrow$  EX1 transition is unlikely, because  $k_{cl}$  and  $k_{ch}$  both increase with  $T$  (Figure 2.8). We conclude that equation 2.2, which represents the central paradigm of the EX2 regime, represents a reasonable foundation for analyzing the HDX behavior of Mb.



**Figure 2.8.** (A) Temperature dependence of  $k_{ch}$ , calculated using the simple Arrhenius expression of equation 2.8a (main text), and by explicitly taking into account how  $[OD^-]$  changes with  $T$  (equation 2.8b). All calculations are based on  $pD = \text{const.} = 7.6$ , keeping in mind that the phosphate-buffered solutions used of the current work are stable against  $T$ -induced changes.<sup>46</sup> The data shown here are for poly-alanine with  $k_{ch}(298 \text{ K}) = 15 \text{ s}^{-1}$ .<sup>18</sup>

(B) Concentration of  $OD^-$  vs. temperature, calculated as follows:

The ionization constant of  $D_2O$  is  $K_{D_2O}(T) = [D^+] \times [OD^-](T)$

such that  $pOD(T) = -\log(K_{D_2O}(T)) - pD$

or  $pOD(T) = -\log(\Delta g(T)/RT) - pD$

The  $OD^-$  concentration ( $M$ ) is thus given by  $[OD^-](T) = 10^{-pOD(T)}$

(C)  $\Delta g(T)$  is the free energy change associated with the  $D_2O \leftrightarrow D^+ + OD^-$  equilibrium.<sup>32</sup>  $\Delta g(T)$  is required for calculating  $[OD^-](T)$ .

$$\Delta g(T) = \Delta h(T) - T\Delta s(T)$$

with the enthalpy

$$\Delta h(T) = \Delta h(298 \text{ K}) + \Delta c_p (T - 298 \text{ K})$$

and the entropy

$$\Delta s(T) = \Delta s(298 \text{ K}) + \Delta c_p \ln(T/298 \text{ K})$$

where  $\Delta h(298 \text{ K}) = 59.8 \text{ kJ mol}^{-1}$ ,  $\Delta s(298 \text{ K}) = -85.5 \text{ J mol}^{-1} \text{ K}^{-1}$ , and  $\Delta c_p = -229.3 \text{ J mol}^{-1} \text{ K}^{-1}$ . These parameters imply that  $k_{B,298} = 3.45 \times 10^8$  in equation 2.8b, to ensure that  $k_{ch}(298 \text{ K}) = 15 \text{ s}^{-1}$ .

Here we use lower case symbols for solvent-related thermodynamic parameters, whereas upper case symbols refer to the protein.

Individual NH sites exhibit single-exponential EX2 kinetics only if the  $NH_{open}$  population in equation 2.1 is small ( $k_{cl} \gg k_{op}$ ).<sup>2, 17</sup> To ensure that this criterion is being met, our analysis will not consider data beyond 353 K (80 °C) where the population of U reaches ~22% (Figure 2.5). In the next few paragraphs will use two peptides, 1-7 and 21-29, to illustrate the remaining layers of the model (Figure 2.4).

### 2.3.3.3 Layer 3: Protein Dynamics at High Temperature (Close to 353 K).

There is some uncertainty regarding the exact nature of the opening/closing fluctuations in equation 2.1.<sup>62, 63</sup> As a starting point we try to equate these fluctuations with global N ↔ U transitions such that equation 2.2 turns into

$$k_{HDX,i}(T) = \exp[-\Delta G_{glob}(T)/RT] \times k_{ch,i}(T)$$

#### Equation 2.8a

where  $\Delta G_{glob}(T)$  is the free energy derived from CD melting data (Figure 2.2C). Figure 2.5A-D compares the predictions of equation 2.8a with experimental %D( $T$ , 30 s) data for peptides 1-7 and 21-29. We ignore, for now, the mismatch at low  $T$  and focus on higher temperatures. For peptide 1-7 the equation 2.8a prediction agrees well with experiments for  $T \geq 338$  K, where %D approaches 100% (Figure 2.5A). In contrast, peptide 21-29 shows major discrepancies at high  $T$ , with experimental %D values that are much lower than predicted by equation 2.8a (dotted line in Figure 2.5C).

The high-temperature mismatch in Figure 2.5C implies that the unfolded state in the CD-detected N ↔ U equilibrium retains some HDX protection for peptide 21-29. This interpretation is consistent with reports of residual structure in many other unfolded proteins,<sup>64, 65</sup> which causes NH sites to exchange slower than in dipeptides.<sup>57, 58, 66</sup> In other words, HDX for peptide 21-29 must involve a two-step opening process (N ↔ U<sub>cl</sub> ↔ U<sub>op</sub>), where the first step corresponds to the CD-detected  $\Delta G_{glob}(T)$ . In the context of equation 2.1, U<sub>cl</sub> represents a “closed” (exchange-incompetent) state. The subsequent U<sub>cl</sub> ↔ U<sub>op</sub> transition generates the HDX-competent “open” state, and this opening event is associated with an additional free energy  $\Delta G_{opU}$ . Overall, the N ↔ U<sub>op</sub> equilibrium thus has a free energy of

$$\Delta G^*_{glob}(T) = \Delta G_{glob}(T) + \Delta G_{opU}$$

#### Equation 2.8b

(Figure 2.5D). For peptide 21-29,  $\Delta G_{opU} = 22 \text{ kJ mol}^{-1}$  generates a HDX profile that agrees quite well with the experimental data above 348 K (solid in Figure 2.5C). In summary, we account for residual protection of heat-unfolded Mb by modifying equation 2.7a according to

$$k_{HDX,i}(T) = \exp[ -\Delta G^*_{glob}(T)/RT ] \times k_{ch,i}(T)$$

**Equation 2.8c**

where  $\Delta G^*_{glob}(T)$  is defined in equation 2.8b and  $\Delta G_{opU} \geq 0$  represents a fitting parameter. U segments without residual protection have  $\Delta G_{opU} = 0$  (such as 1-7, Figure 2.5B). In principle,  $\Delta G_{opU}$  will depend on  $T$  as expressed in equation 2.4. However,  $\Delta G_{opU}$  makes its presence felt only in a narrow range close to  $T_m$ , such that the  $\Delta H$  and  $\Delta S$  components of  $\Delta G_{opU}$  cannot be determined. To avoid the use of such ill-defined parameters, we will therefore assume that  $\Delta G_{opU}$  is  $T$ -independent.

#### 2.3.3.4 Layer 4: Local Fluctuations.

The Layer 3 considerations refer to relatively high temperatures (around 353 K) where global unfolding/refolding starts to make its presence felt.<sup>34</sup> At lower  $T$  global dynamics are less prevalent, and HDX occurs mainly via local fluctuations.<sup>2, 63</sup> Englander et al.<sup>2</sup> suggested that the interplay of local and global dynamics can be captured by replacing  $K_{op}$  in equation 2.2 with the sum  $K_{op}(\text{local}) + K_{op}(\text{global})$ . We build on this idea by expanding equation 2.2 according to

$$k_{HDX,i}(T) = [ K_{op\_loc}(T) + K^*_{op\_glob}(T) ] \times k_{ch,i}(T)$$

**Equation 2.9**

The equilibrium constant  $K_{op\_loc}$  in equation 2.9 describes local opening/closing fluctuations which are associated with the free energy  $\Delta G_{loc}(T)$ , such that  $K_{op\_loc} = \exp[ -$

$\Delta G_{loc}(T)/RT$  ] with  $\Delta G_{loc}(T) = \Delta H_{loc} - T\Delta S_{loc}$ . To avoid “overfitting”, we make the approximation that these local events have  $\Delta C_p = 0$ . As discussed in Layer 3,  $K^*_{op\_glob} = \exp[-\Delta G^*_{glob}(T)/RT]$  describes the  $N \leftrightarrow U_{op}$  equilibrium. Equation 2.9 expresses the idea<sup>2</sup> that any NH site can reach an open state via two types of dynamics, such that  $k_{HDX}$  is the sum of both kinetic channels. When putting it all together, equation 2.9 turns into

$$k_{HDX,i}(T) = ( \exp[-(\Delta H_{loc} - T\Delta S_{loc})/RT] + \exp[-(\Delta G_{glob}(T) + \Delta G_{opU})/RT] ) \times k_{ch,i}(T)$$

**Equation 2.10**

Fitting of the adjustable parameters ( $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$ ) results in greatly improved agreement with the experimental  $\%D(T, 30 \text{ s})$  profiles for both test peptides (Figure 2.5E and H). Here we assumed that all NH sites in any given peptide share the same thermodynamic parameters, a limitation that will be improved upon in the following section.

### 2.3.3.5 Layer 5: Inclusion of time-Dependent Data.

For extending our model to labeling times beyond  $t = 30 \text{ s}$  we included data at  $t = 600 \text{ s}$  and  $6000 \text{ s}$  at 296 K and 333 K. These temperatures were chosen because they provided  $\%D(T, t)$  values that covered a wide dynamic range. The inclusion of these additional time points provides much more stringent constraints for the model parameters.

Layer 4 provides a poor description of the HDX data for  $t > 30 \text{ s}$  (Figure 2.5G, J). Luckily, a minor modification is sufficient to remedy this mismatch. So far, we assumed that all NH sites in a peptide share the same  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$ . We now eliminate this unrealistic restriction. In principle each NH site should have its own  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$ . However, such an approach would imply an unrealistically large number of fitting parameters ( $\sim 148 \times 3 = 444$ ), generating a mathematically underdetermined situation. To avoid this problem, we chose a compromise where each peptide was divided into segments. NH sites in each segment share the same  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$ . This segmentation concept is borrowed from

the widely accepted foldon model,<sup>2</sup> where groups of adjacent NH sites open/close with the same thermodynamic parameters. In Figure 2.5 we dissected peptides into two segments, e.g., peptide 1-7 was divided into segments 3-4 and 5-7. Residues 1-2 were not considered because of back exchange.<sup>18, 35, 57</sup> After this modification, equation 2.10 provided excellent agreement with the experimental data for all  $T$  and  $t$  (Figures 2.5K/N and O/R).

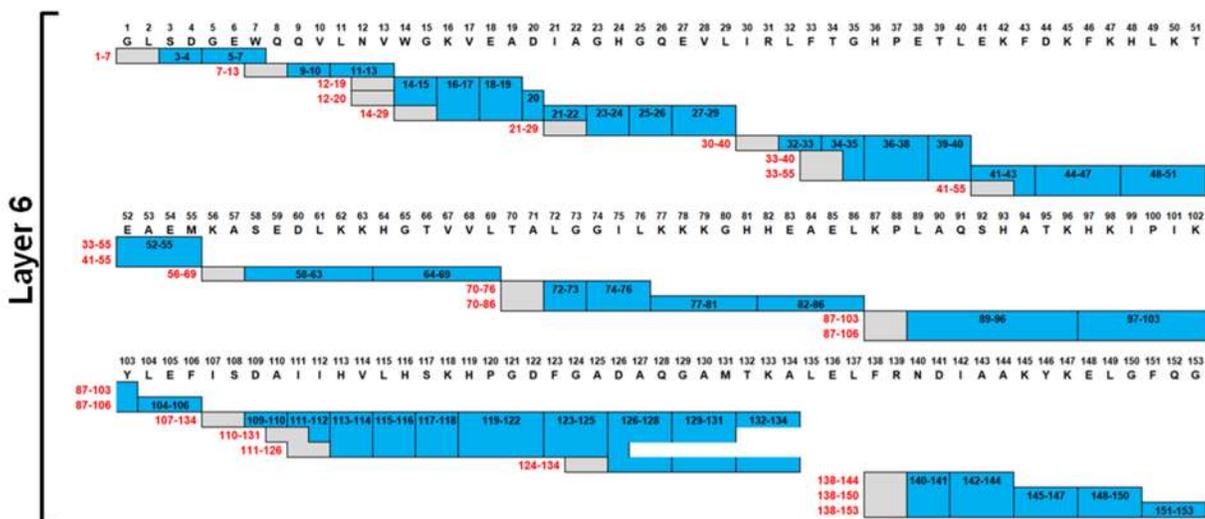
### 2.3.3.6 Layer 6: Global Fitting

Many Mb regions were covered by overlapping peptides. Rather than fit each peptide individually (as in Layer 5), this overlap allows the implementation of a global fitting strategy (Note: this term refers to a data analysis method;<sup>55, 56, 67-69</sup> it is unrelated to “global” unfolding). Segments that are shared across multiple peptides were modeled using the same  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$ . Global fitting improves the robustness and accuracy of results compared to single-curve analyses, and it reduces the number of parameters.<sup>55, 56, 67-69</sup> The 22 peptides with the highest S/N were subjected to global fitting (Figure 2.9) by minimizing

$$X^2 = \sum_{peptides} \sum_T \sum_t (\%D_{exp} - \%D_{calc})^2$$

**Equation 2.11**

where the summation includes multiple peptides,  $T$ , and  $t$  values.  $\%D_{exp}$  refers to experimental data.  $\%D_{calc}$  values were calculated using equation 2.10. Preliminary segment boundaries were first determined by analyzing one peptide at a time. If treating a peptide as a single segment did not yield an acceptable fit, it was divided into two, then three segments, etc. For global fitting these preliminary boundaries had to be slightly adjusted to ensure consistency across overlapping peptides. In the end, Mb was dissected into 44 segments, each of which had its own  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$  ( $44 \times 3 = 132$  parameters, segment boundaries are denoted as vertical lines in Figure 2.9).



**Figure 2.9.** Overlapping peptides (red) used for global fitting, illustrating Layer 6 of the modeling strategy developed here. Blue vertical segments share the same  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$  across different peptides; the residue range for each segment is indicated. The first two residues of each peptide (gray) were not considered due to back exchange. Preliminary segment boundaries were first determined by analyzing one peptide at a time. If treating the peptides as a single segment did not yield an acceptable fit, it was divided into two, then three segments, etc. For global fitting these preliminary boundaries had to be slightly adjusted to ensure consistency across overlapping peptides (vertical lines in the figure above).

### 2.3.3.7 Discussion of Global Fitting Strategy

Global fitting generally improves the robustness and accuracy of parameters compared to single-curve analyses. At the same time, the number of parameters required for describing the whole data set is reduced.<sup>55, 56, 67-69</sup> The procedure used in our work involved 22 peptides that were dissected into 44 segments, for a total of  $44 \times 3 = 132$  fitting parameters. One can contrast this to traditional HDX-MS strategies that use expressions such as

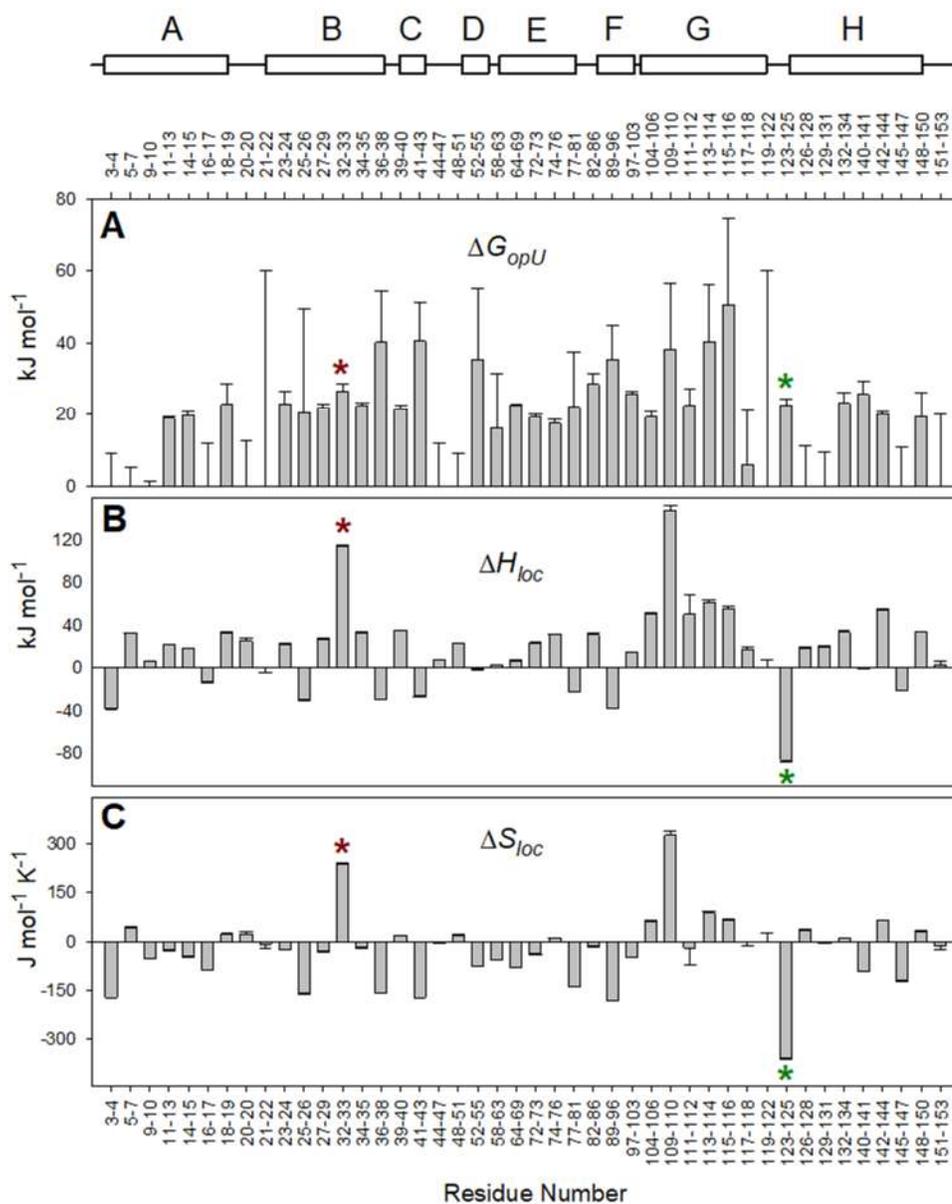
$$\%D = a_0 + a_1(1 - \exp[-k_{app,1} \times t]) + a_2(1 - \exp[-k_{app,2} \times t])$$

**Equation 2.12**

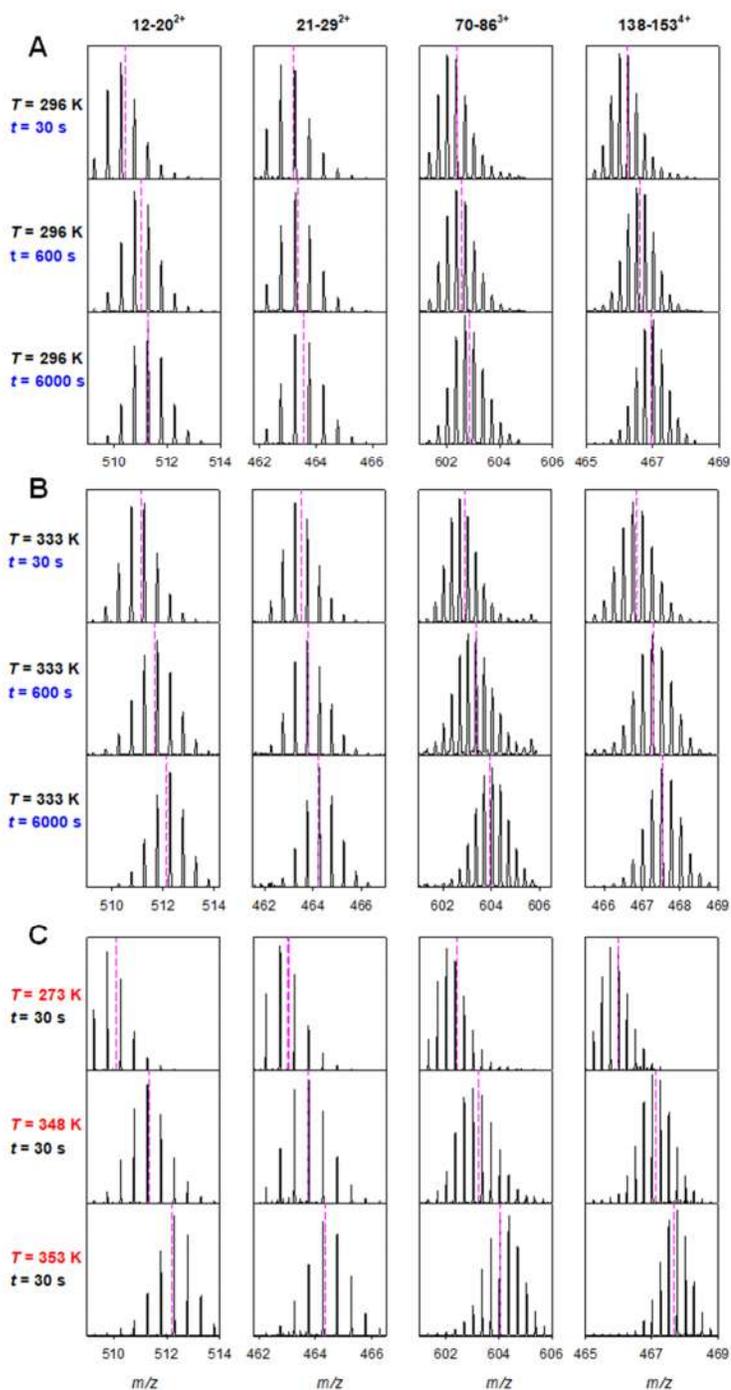
with five parameters per peptide (or more, when using additional exponentials).<sup>25,54</sup> For 22 peptides and two temperatures that traditional method would require a minimum of  $22 \times 2 \times 5 = 220$  parameters. Here we probed the HDX properties of Mb at *eleven* (not two) temperatures. In other words, the number of fitting parameters in our global analysis is low, compared to traditionally used approaches. More importantly, the parameters obtained here ( $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$ ) directly report on first-principle protein properties. This is in contrast to the  $a_i$  and  $k_{app\_i}$  values of the equation above, which are difficult to interpret in a structural/thermodynamic context.

### 2.3.3.8 Applying the Model to Experimental HDX Data.

HDX-MS data for Mb acquired as a function of  $T$  and  $t$  were analyzed using equation 2.10. Gratifyingly, our model produced excellent fits for all 22 peptides, illustrated in Figure 2.4A for selected  $\%D(T, 30 \text{ s})$  profiles. Fitted  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$  parameters are compiled in Figure 2.10.



**Figure 2.10.**  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$  determined by global fitting of temperature- and time-dependent HDX-MS data for 44 segments along the Mb sequence. Shown at the top is the Mb secondary structure with helices A-H. Plots of  $\Delta G_{glob}^*(T)$  and  $\Delta G_{loc}(T)$  derived from the parameters shown here are summarized in Figure 2.11. Errors bars indicate by how much each value could be altered to cause a 10% increase of  $X^2$ . Asterisks highlight two segments that are further examined in Figure 2.12.



**Figure 2.11.** HDX-MS isotope distributions of selected peptic peptides. (A) Data acquired after different labeling time intervals  $t$  at a constant temperature of  $T = 296$  K (23 °C). (B) Same as in panel A, but for  $T = 333$  K (60 °C). (C) Data acquired at different temperatures  $T$  for a constant labeling time of  $t = 30$  s. Vertical dashed lines indicate centroid  $m/z$  values.

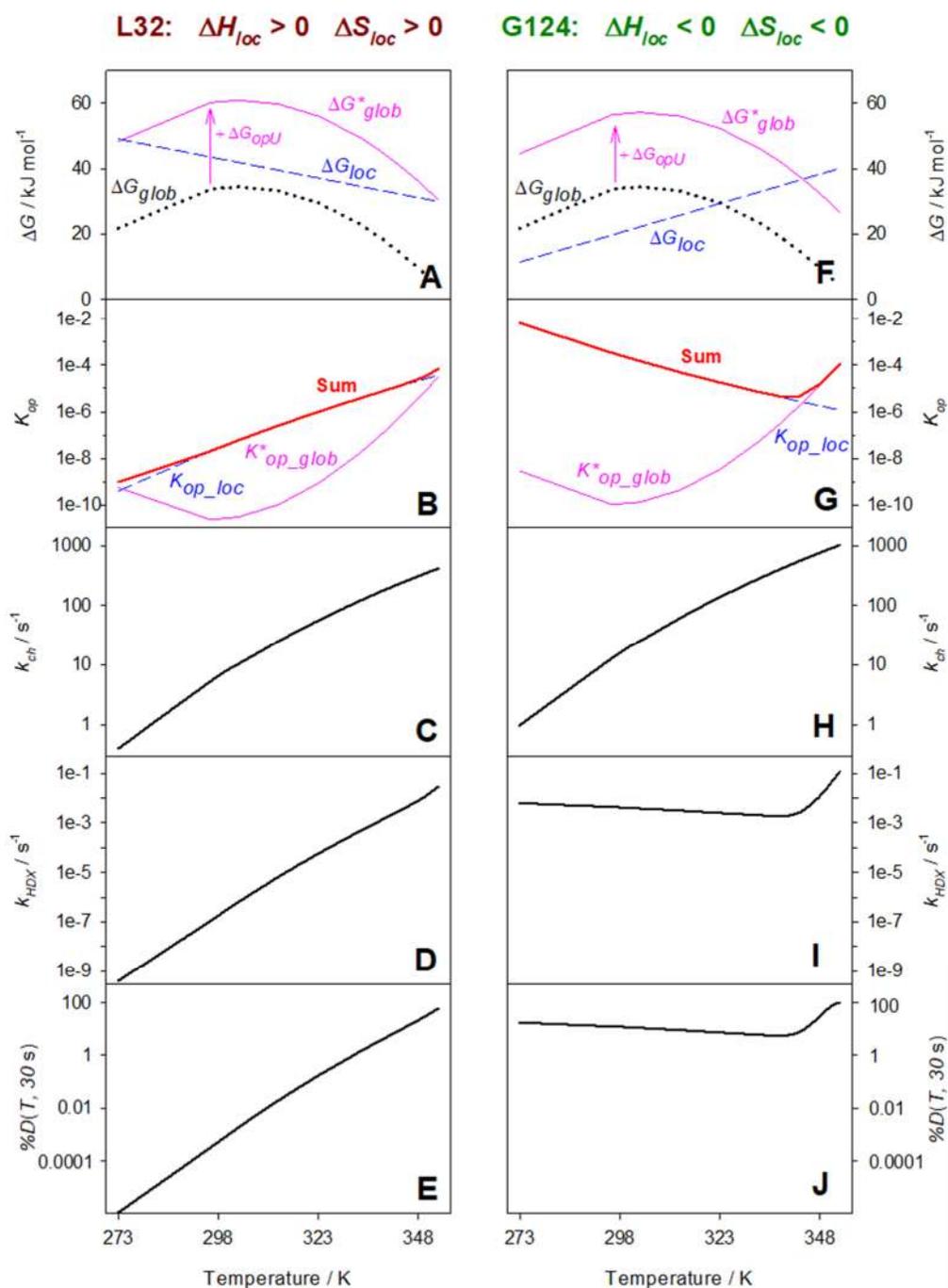
**$\Delta G_{opU}$  Values.** Consistent with previous observations,<sup>57, 58, 64-66</sup> our data show that the CD-detected globally unfolded state retains residual protection. Deuteration of  $U_{cl}$  is mediated by opening/closing transitions with a free energy  $\Delta G_{opU}$ . Most of the fitted  $\Delta G_{opU}$  values fell in the range of 20 to 40 kJ mol<sup>-1</sup> (Figure 2.10A). The similarity of  $\Delta G_{opU}$  across much of the protein could suggest the occurrence of cooperative  $U_{cl} \leftrightarrow U_{op}$  fluctuations.<sup>2</sup> However, the presence of independently fluctuating segments cannot be ruled out.

**$\Delta H_{loc}$  and  $\Delta S_{loc}$  Values.** Before examining the remaining fitting parameters, we note that  $N \leftrightarrow U$  equilibria can follow different thermodynamic scenarios. (i) Unfolding is often endothermic ( $\Delta H > 0$ ) because energy is required to dissociate contacts within the protein that stabilize  $N$  (e.g., H-bonds). Under these conditions the enhanced conformational freedom of  $U$  usually results in  $\Delta S > 0$ .<sup>38, 39</sup> In the context of global Mb unfolding this scenario applies for  $T > 303$  K, where the slope of  $\Delta G_{glob}(T)$  is negative (Figure 2.5B). (ii)  $N \leftrightarrow U$  equilibria can also be exothermic ( $\Delta H < 0$ ) with  $\Delta S < 0$ ,<sup>39, 70</sup> as seen in Figure 2.5B for  $T < 303$  K where  $\Delta G_{glob}(T)$  has a positive slope. The origin of this second scenario remains poorly understood,<sup>41</sup> but the assembly of tightly H-bonded shell water around unfolded regions likely plays a role.<sup>71</sup>

The aforementioned two scenarios are well established for *global*  $N \leftrightarrow U$  equilibria.<sup>39, 70, 71</sup> It is reasonable to assume that these concepts also extend to *local* closed  $\leftrightarrow$  open fluctuations. Consistent with this expectation, our HDX fits yielded positive as well as negative  $\Delta H_{loc}$  and  $\Delta S_{loc}$  values (Figure 2.10B, C). Segments with  $\Delta H_{loc} > 0$  and  $\Delta S_{loc} > 0$  were located mostly within long helices (such as B, G, and H), corresponding to scenario (i) described above. Conversely, most segments with  $\Delta H_{loc} < 0$  and  $\Delta S_{loc} < 0$  (scenario ii) were found at helix/loop boundaries (Figure 2.10B, C). This spatial distribution is intriguing, but from the data presented here it cannot be decided why these specific regions follow one scenario versus the other.

**Interplay of Local and Global Dynamics.** It is gratifying that equation 2.10 can capture the entire HDX-MS data set for different  $T$  and  $t$ . We will illustrate the underlying  $T$ -dependent local and global dynamics for two residues, L32 and G124 (Figure 2.12); these

were selected because of their particularly large  $|\Delta H_{loc}|$  and  $|\Delta S_{loc}|$  (asterisks in Figure 2.10B, C).



**Figure 2.12.** Temperature-dependent HDX events, illustrated for the backbone NH sites of L32 (left) and G124 (right). (A, F) Free energy profiles for local and global fluctuations,

**Figure 2.12 Caption (continued):**

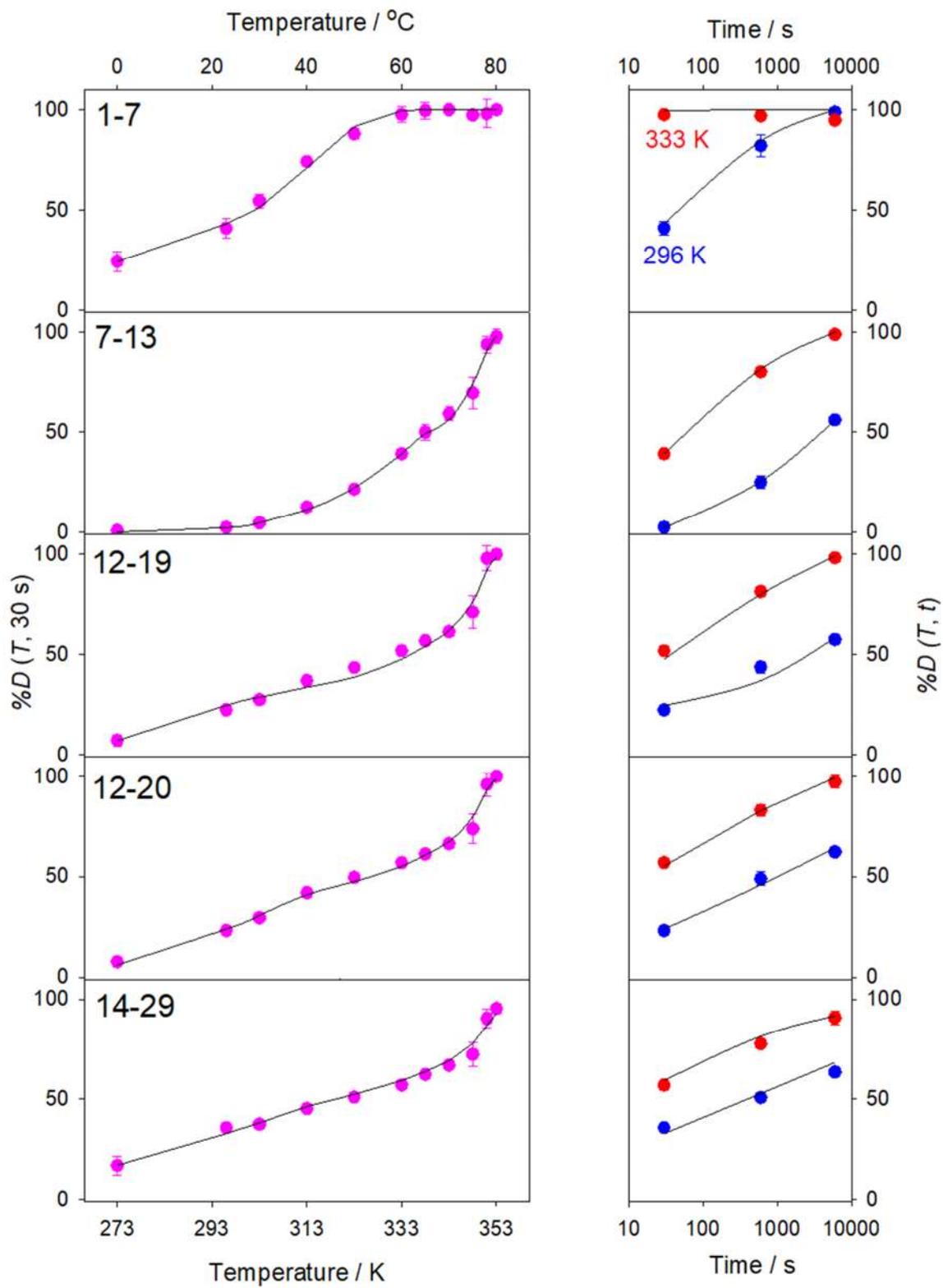
*calculated from  $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$  parameters of Figure 2.10. (B, G) Conversion of free energy to equilibrium constants, along with their sum (red lines, eq. 8). (C, H)  $k_{ch,32}$  and  $k_{ch,124}$ , calculated using eq. 6b. (D, I)  $k_{HDX,32}$  and  $k_{HDX,124}$ , calculated as the product of “Sum” in panels B/G, and  $k_{ch}$  in panels C/H (eq. 8). (E, J) %D for a labeling time of 30 s. Note that except for A/F, all panels have logarithmic y-axes.*

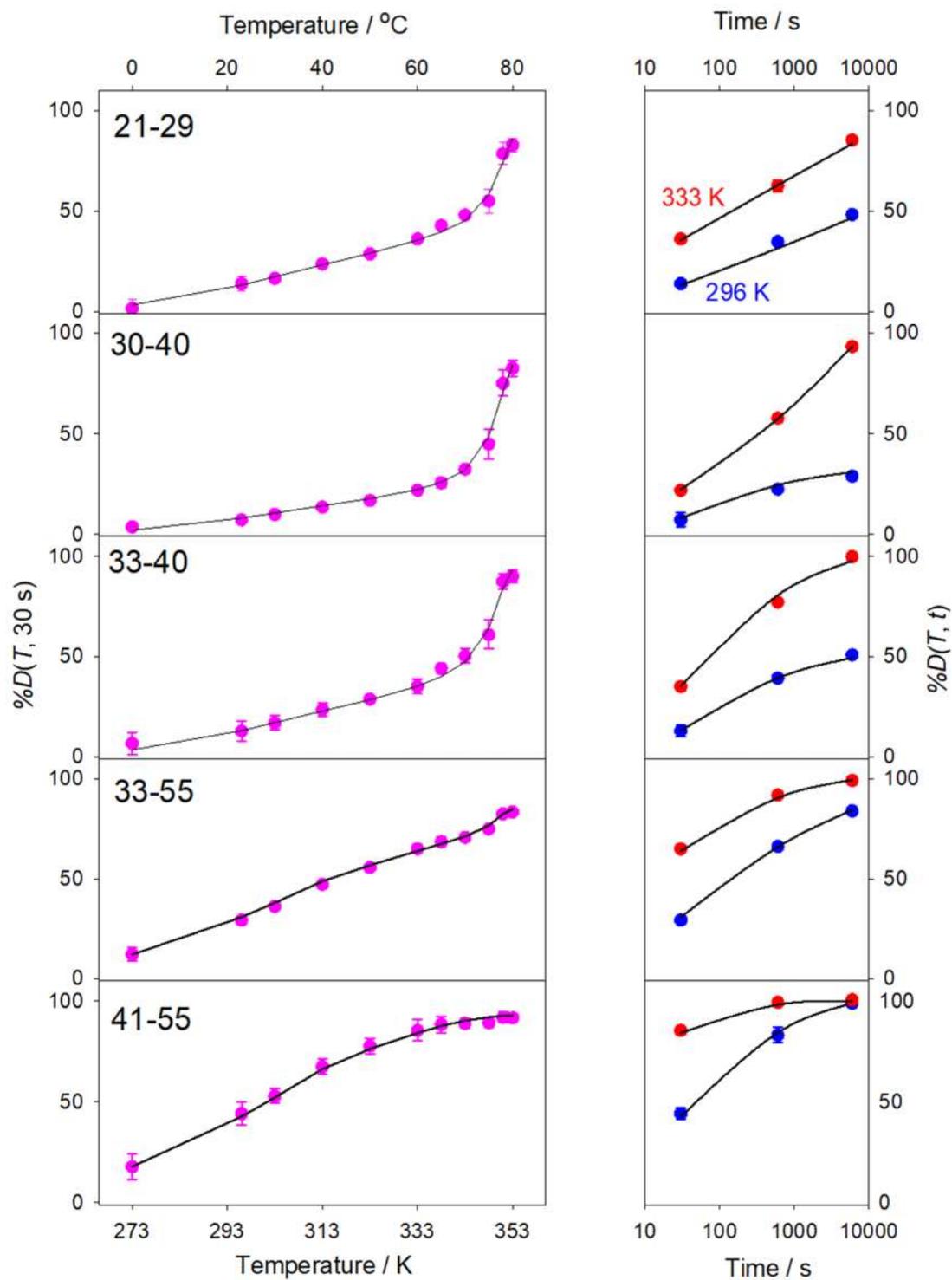
The HDX behavior of any NH site is governed by the opening/closing fluctuation with the lowest  $\Delta G$  (the largest  $K_{op}$ ).<sup>2</sup> Accordingly, HDX of L32 is dominated by local fluctuations throughout the entire  $T$  range, while global events are negligible (Figure 2.12A, B). As expressed in equation 2.9, multiplication of the  $K_{op}$  “Sum” (red line in Figure 2.12B) with  $k_{ch}$  (Figure 2.12C) yields  $k_{HDX}$  (Figure 2.12D). The resulting %D( $T$ ) profile increases toward 100% in a near-exponential fashion (Figure 2.12E). This steep %D( $T$ ) rise results from three effects: 1.  $k_{HDX}$  increases with  $T$  (Figure 2.12D). 2. The  $\exp(-T^{-1})$  dependence in equation 2.10 favors open conformations at high  $T$ . 3.  $\Delta G_{loc}(T)$  has a negative slope ( $\Delta S_{loc} > 0$ ) which further promotes the open state at high  $T$  (Figure 2.12A).

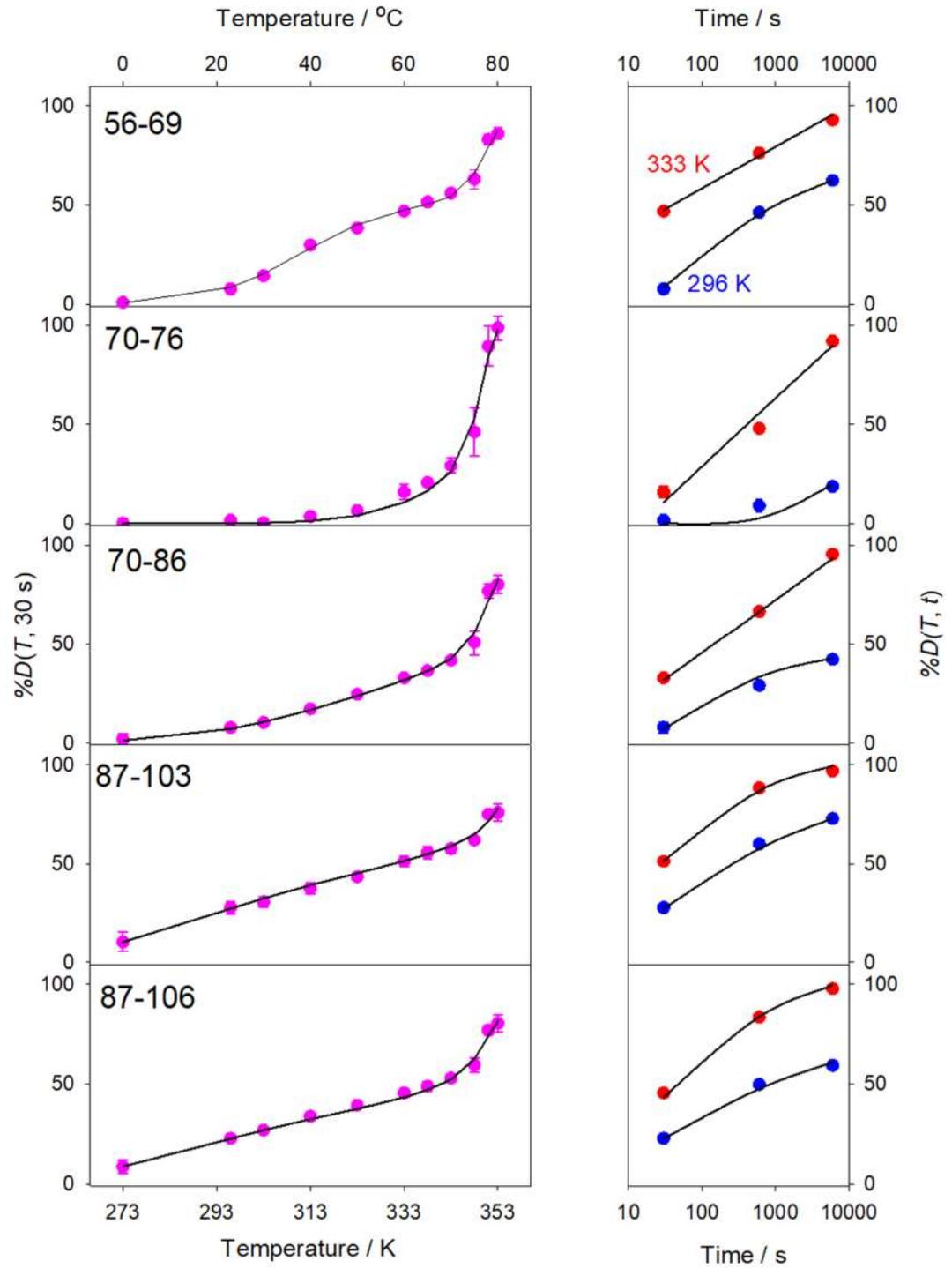
The experimental %D( $T$ , 30 s) profiles of many Mb peptides had very shallow slopes at low  $T$ , followed by a sudden increase close to  $T_m$  (e.g. peptide 111-126 in Figure 2.4A). It would be impossible to model these data if all NH sites had rapidly increasing %D profiles similar to that of Figure 2.12E. A look at G124 reveals how our model solves this problem. G124 has  $\Delta S_{loc} < 0$ , resulting in a positive slope for  $\Delta G_{loc}$  that suppresses local opening with increasing  $T$ . At 343 K the  $\Delta G_{loc}$  and  $\Delta G^*_{glob}$  curves intersect; beyond this temperature HDX takes place mainly via global dynamics (Figure 2.12F). As a result,  $k_{HDX}(T)$  decreases slightly between 273 K and 343 K, followed by a sudden upward kink (Figure 2.12I). This behavior is echoed in the %D( $T$ , 30 s) profile of Figure 2.12J as well as the experimental data of peptide 111-126 (Figure 2.4A).

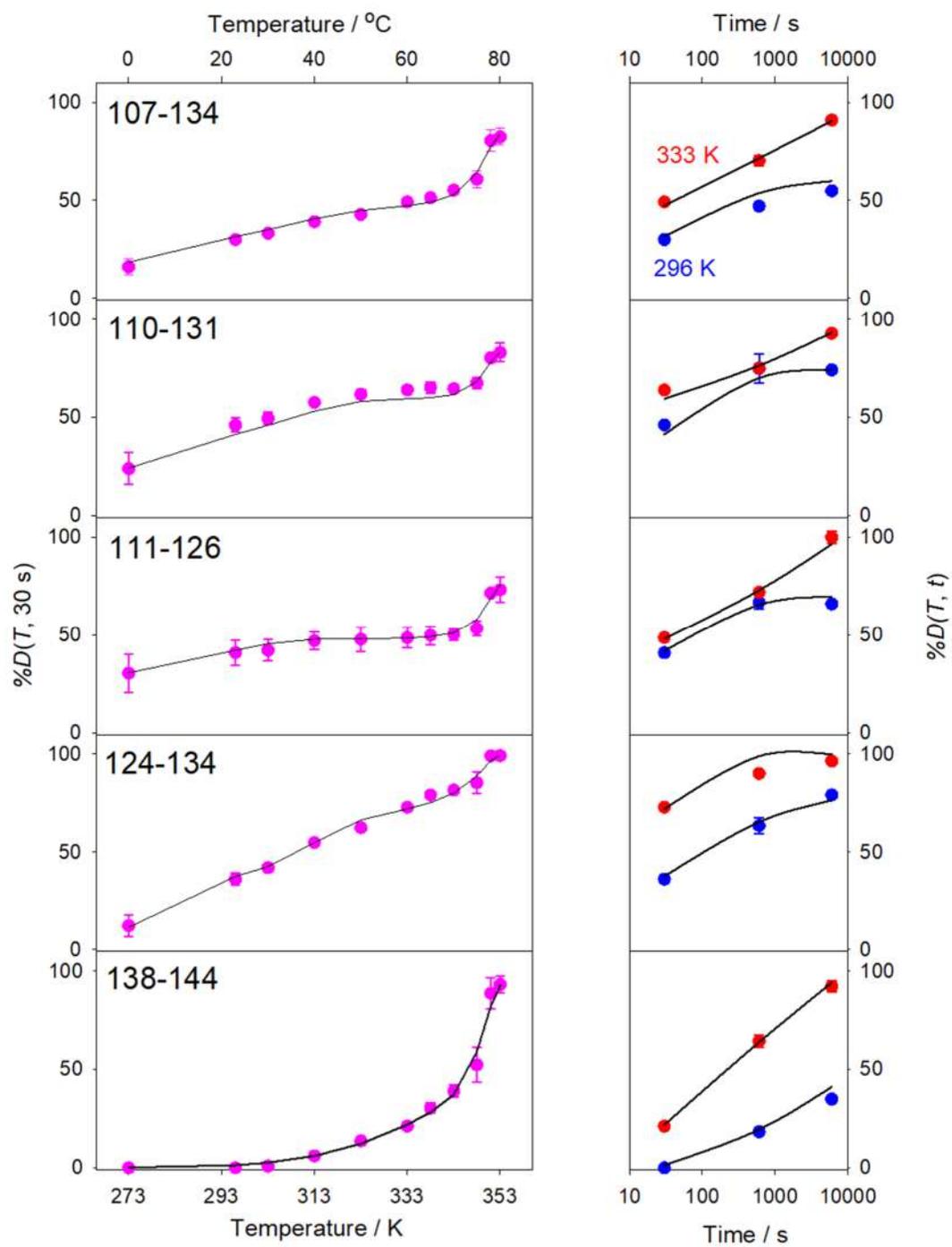
In summary, Mb dynamics at low  $T$  are dominated by local fluctuations. For some of these local fluctuations  $\Delta G_{loc}(T)$  decreases, while for others  $\Delta G_{loc}(T)$  increases with  $T$ . The latter account for the very shallow slopes that were experimentally observed for many  $\%D(T)$  profiles at low  $T$ . The conspicuous kink of the experimental  $\%D(T)$  data at around 343 K results from a crossover of  $\Delta G_{loc}(T)$  and  $\Delta G_{glob}(T)$ , marking the point at which global fluctuations start to dominate.

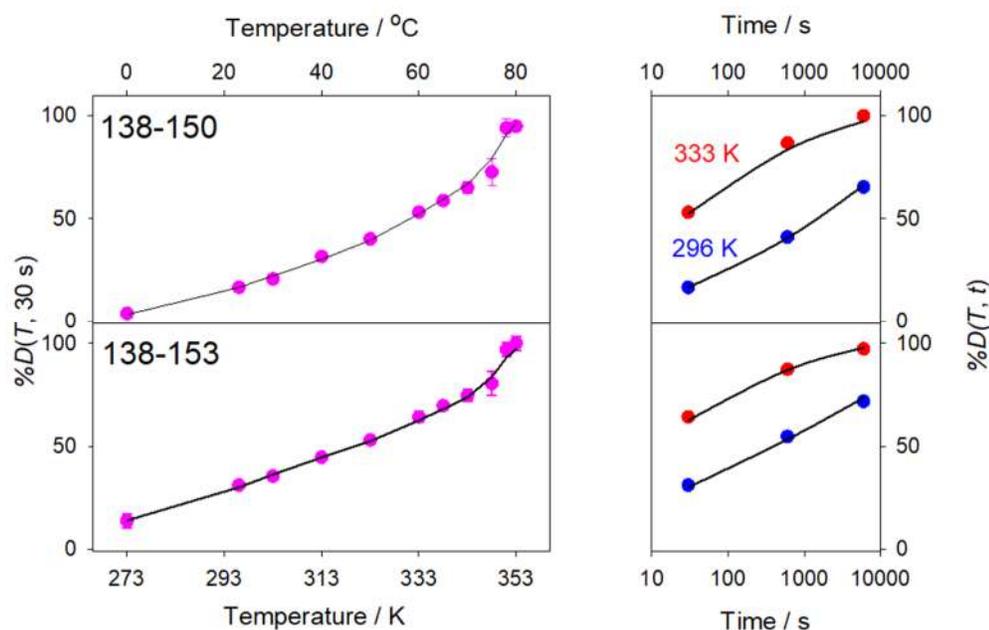
Figure 2.13 summarizes all  $T$ - and  $t$ -dependent data, including peptides not shown in Figure 2.4.











**Figure 2.13.** Complete experimental HDX-MS data set (colored dots), with fits (black lines) based on equation 2.11 obtained by global analysis of overlapping peptides.

## 2.4 Conclusions

For the first time, the current work provides a thermodynamic model that can quantitatively describe temperature-dependent protein HDX-MS data. For conventional time-dependent HDX-MS experiments it is common to use multi-exponential fits with apparent rate constants.<sup>54</sup> Unfortunately, that conventional approach only yields a phenomenological description of the data, and the fitting parameters obtained are difficult to interpret in a structural/thermodynamic context. This is in contrast to the temperature-dependent model developed here, where the fitting parameters directly report on first-principle protein properties ( $\Delta G_{opU}$ ,  $\Delta H_{loc}$ , and  $\Delta S_{loc}$ ).

The analysis strategy presented here captures the interplay of local fluctuations (which dominate at low  $T$ ) and global unfolding/refolding (which becomes prevalent closer to  $T_m$ ). Some of the local fluctuations are associated with  $\Delta H_{loc} > 0$  and  $\Delta S_{loc} > 0$ , representing the canonical scenario<sup>38, 39</sup> where thermal energy is required to disrupt local noncovalent

contacts, and where locally unfolded segments are more disordered than in the native state. On the other hand, there are also local events with  $\Delta H_{loc} < 0$  and  $\Delta S_{loc} < 0$ . The structural origin of this second scenario remains to be fully elucidated, but it may be caused by tightly H-bonded water around the locally unfolded regions.<sup>39, 70, 71</sup> The second scenario is indispensable for modeling the shallow slopes seen for many of the experimental HDX profiles at low temperature. Another essential ingredient of our model is residual HDX protection of the globally unfolded state, implying that U undergoes transient opening transitions that are associated with  $\Delta G_{opU}$ . This phenomenon is consistent with studies on many other proteins.<sup>57, 58, 64-66</sup>

This work provides practitioners with a tool for analyzing HDX-MS data across a wide range of temperatures, e.g., for assessing the thermal stability of protein therapeutics.<sup>15, 30, 31</sup> The key steps required for applying this method are as follows: (i) Ensure that the protein does not aggregate in the  $T$  range of interest. (ii) Determine global unfolding parameters ( $T_m$ ,  $\Delta H_{glob}(T_m)$ ,  $\Delta C_p$ ) by CD spectroscopy, DSC, or from the literature. (iii) Measure %D for overlapping peptides over a wide temperature range and at different time points and verify EX2 behavior. (iv) Perform global fitting on the basis of equation 2.10.

An interesting aspect of temperature-dependent HDX-MS is the dramatically shortened time scale. Traditional room temperature assays routinely employ labeling times up to many hours,<sup>5, 8-15</sup> whereas the current work demonstrates that complete HDX profiles can be generated in as little as 30 s (Figure 2.4A). With robotic technology it should be possible to develop temperature-dependent HDX-MS workflows for high-throughput applications such as excipient screening<sup>30</sup> or drug candidate binding tests.<sup>11</sup> In any case, it is hoped that the current work will encourage practitioners to explore the use of temperature-dependent HDX-MS, instead of being confined to traditional time-domain measurements.

## 2.5 References

- 1 Frauenfelder, H., G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, R. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129.
- 2 Bai, Y., T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **1995**, *269*, 192.
- 3 Neupane, K., D. A. N. Foster, D. R. Dee, H. Yu, F. Wang, M. T. Woodside, *Science* **2016**, *352*, 239.
- 4 Hanoian, P., C. T. Liu, S. Hammes-Schiffer, S. Benkovic, *Accounts Chem. Res.* **2015**, *48*, 482.
- 5 Murcia Rios, A., S. Vahidi, S. D. Dunn, L. Konermann, *J. Am. Chem. Soc.* **2018**, *140*, 14860.
- 6 Boehr, D. D., R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, *5*, 789.
- 7 Balchin, D., M. Hayer-Hartl, F. U. Hartl, *Science* **2016**, *353*, 42.
- 8 Rob, T., P. Liuni, P. K. Gill, S. L. Zhu, N. Balachandran, P. J. Berti, D. J. Wilson, *Anal. Chem.* **2012**, *84*, 3771.
- 9 Pirrone, G. F., R. E. Iacob, J. R. Engen, *Anal. Chem.* **2015**, *87*, 99.
- 10 Rand, K. D., M. Zehl, T. J. D. Jorgensen, *Acc. Chem. Res.* **2014**, *47*, 3018.
- 11 Marciano, D. P., V. Dharmarajan, P. R. Griffin, *Curr. Op. Struct. Biol.* **2014**, *28*, 105.
- 12 Percy, A. J., M. Rey, K. M. Burns, D. C. Schriemer, *Anal. Chim. Acta* **2012**, *721*, 7.
- 13 Martens, C., M. Shekhar, A. M. Lau, E. Tajkhorshid, A. Politis, *Nat. Protoc.* **2019**, *14*, 3183.
- 14 Habibi, Y., K. A. Uggowitzer, H. Issak, C. J. Thibodeaux, *J. Am. Chem. Soc.* **2019**, *141*, 14661.
- 15 Masson, G. R., M. L. Jenkins, J. E. Burke, *Expert. Opin. Drug Discov.* **2017**, *12*, 981.
- 16 van den Bedem, H., J. S. Fraser, *Nat. Methods* **2015**, *12*, 307.
- 17 Hvidt, A., S. O. Nielsen, *Adv. Protein Chem.* **1966**, *21*, 287.

- 18 Bai, Y., J. S. Milne, L. Mayne, S. W. Englander, *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.
- 19 Konermann, L., J. Pan, Y. Liu, *Chem. Soc. Rev.* **2011**, *40*, 1224.
- 20 Xiao, H., J. K. Hoerner, S. J. Eyles, A. Dobo, E. Voigtman, A. I. Mel'Cuk, I. A. Kaltashov, *Protein Sci.* **2005**, *14*, 543.
- 21 Xiao, P., D. Bolton, R. A. Munro, L. S. Brown, V. Ladizhansky, *Nat. Commun.* **2019**, *10*,
- 22 van de Waterbeemd, M., A. Llauro, J. Snijder, A. Valbuena, A. Rodriguez-Huete, M. A. Fuertes, P. J. de Pablo, M. G. Mateu, A. J. R. Heck, *Biophys. J.* **2017**, *112*, 1157.
- 23 Lim, X. X., A. Chandramohan, X. Y. E. Lim, J. E. Crowe, S. M. Lok, G. S. Anand, *Structure* **2017**, *25*, 1391.
- 24 Offenbacher, A. R., S. S. Hu, E. M. Poss, C. A. M. Carr, A. D. Scouras, D. M. Prigozhin, A. T. Iavarone, A. Palla, T. Alber, J. S. Fraser, J. P. Klinman, *ACS Central Sci.* **2017**, *3*, 570.
- 25 Liang, Z.-X., T. Lee, K. A. Resing, N. G. Ahn, J. P. Klinman, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9556.
- 26 Banks, D. D., J. Zhang, C. C. Siska, *Mol. Pharm.* **2014**, *11*, 3431.
- 27 Man, P., M. Fabry, I. Sieglöva, D. Kavan, P. Novak, A. Hnizda, *BBA-Proteins Proteomics* **2019**, *1867*, 376.
- 28 Rist, W., T. J. D. Jorgensen, P. Roepstorff, B. Bukau, M. P. Mayer, *J. Biol. Chem.* **2003**, *278*, 51415.
- 29 Cirri, E., S. Brier, R. Assal, J. C. Canul-Tec, J. Chamot-Rooke, N. Reyes, *Elife* **2018**, *7*,
- 30 Toth, R. T., S. E. Pace, B. J. Mills, S. B. Joshi, R. Esfandiary, C. R. Middaugh, D. D. Weis, D. B. Volkin, *J. Pharm. Sci.* **2018**, *107*, 1009.
- 31 Wang, G. B., P. V. Bondarenko, I. A. Kaltashov, *Analyst* **2018**, *143*, 670.
- 32 Covington, A. K., R. A. Robinson, R. G. Bates, *J. Phys. Chem.* **1966**, *70*, 3820.
- 33 Perrin, C. L., *Accounts Chem. Res.* **1989**, *22*, 268.
- 34 Privalov, P. L., N. N. Khechinashvili, *J. Mol. Biol.* **1974**, *86*, 665.
- 35 Coales, S. J., S. Y. E. J. E. Lee, A. Ma, J. A. Morrow, Y. Hamuro, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3585.

- 36 Goswami, D., S. Devarakonda, M. J. Chalmers, B. D. Pascal, B. M. Spiegelman, P. R. Griffin, *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1584.
- 37 Klinman, J. P., *Accounts Chem. Res.* **2015**, *48*, 449.
- 38 Swint, L., A. D. Robertson, *Protein Sci.* **1993**, *2*, 2037.
- 39 Privalov, P. L., Y. V. Griko, S. Y. Venyaminov, V. P. Kutysenko, *J. Mol. Biol.* **1986**, *190*, 487.
- 40 Graziano, G., *Phys. Chem. Chem. Phys.* **2010**, *12*, 14245.
- 41 Yang, C., S. Jang, Y. Pak, *Nat. Commun.* **2014**, *5*,
- 42 Hughson, F. M., P. E. Wright, R. L. Baldwin, *Science* **1990**, *249*, 1544.
- 43 Fändrich, M., M. A. Fletcher, C. M. Dobson, *Nature* **2001**, *410*, 165.
- 44 Maurus, R., C. M. Overall, R. Bogumil, Y. Luo, A. G. Mauk, M. Smith, G. D. Brayer, *Biochim. Biophys. Acta* **1997**, *1341*, 1.
- 45 Glasoe, P. K., F. A. Long, *J. Phys. Chem.* **1960**, *64*, 188.
- 46 Reineke, K., A. Mathys, D. Knorr, *Int. J. Food Prop.* **2011**, *14*, 870.
- 47 Greenfield, N. J., *Anal. Biochem.* **1996**, *235*, 1
- 48 Fersht, A. R., *Structure and Mechanism in Protein Science*. Editor, W. H. Freeman & Co., New York, **1999**.
- 49 Kaltashov, I. A., C. E. Bobst, R. R. Abzalimov, *Anal. Chem.* **2009**, *81*, 7892.
- 50 Dong, A. C., T. W. Randolph, J. F. Carpenter, *J. Biol. Chem.* **2000**, *275*, 27689.
- 51 Bianco, V., M. Alonso-Navarro, D. Di Silvio, S. Moya, A. L. Cortajarena, I. Coluzza, *J. Phys. Chem. Lett.* **2019**, *10*, 4800.
- 52 Holtzer, M. E., A. Holtzer, *Biopolymers* **1992**, *32*, 1675.
- 53 Hermans, J., G. Acampora, *J. Am. Chem. Soc.* **1967**, *89*, 1547.
- 54 Smith, D. L., Y. Deng, Z. Zhang, *J. Mass Spectrom.* **1997**, *32*, 135.
- 55 Skinner, S. P., G. Radou, R. Tuma, J. J. Houwing-Duistermaat, E. Paci, *Biophys. J.* **2019**, *116*, 1194.
- 56 Fajer, P. G., G. M. Bou-Assaf, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1202.

- 57 Keppel, T. R., B. A. Howard, D. D. Weis, *Biochemistry* **2011**, *50*, 8722.
- 58 Nguyen, D., L. Mayne, M. C. Phillips, S. W. Englander, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1936.
- 59 Yang, W. Y., M. Gruebele, *Nature* **2003**, *423*, 193.
- 60 Scalley, M. L., D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10636.
- 61 Mallamace, F., C. Corsaro, D. Mallamace, S. Vasi, C. Vasi, P. Baglioni, S. V. Buldyrev, S. H. Chen, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 3159.
- 62 Persson, F., B. Halle, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10383.
- 63 Best, R. B., M. Vendruscolo, *Structure* **2006**, *14*, 97.
- 64 Ratcliff, K., S. Marqusee, *Biochemistry* **2010**, *49*, 5167.
- 65 Bowler, B. E., *Curr. Op. Struct. Biol.* **2012**, *22*, 4.
- 66 Zhu, S. L., A. Shala, A. Bezginov, A. Sljoka, G. Audette, D. J. Wilson, *PloS One* **2015**, *10*, 1.
- 67 Hamuro, Y., *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 486.
- 68 Kan, Z. Y., B. T. Walters, L. Mayne, S. W. Englander, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16438.
- 69 Gessner, C., W. Steinchen, S. Bedard, J. J. Skinner, V. L. Woods, T. J. Walsh, G. Bange, D. P. Pantazatos, *Sci. Rep.* **2017**, *7*,
- 70 Ascolese, E., G. Graziano, *Chem. Phys. Lett.* **2008**, *467*, 150.
- 71 Dias, C. L., T. Ala-Nissila, M. Karttunen, I. Vattulainen, M. Grant, *Phys. Rev. Lett.* **2008**, *100*, 118101.

### 3 Chapter 3. Mechanism of Thermal Protein Aggregation: Experiments and Molecular Dynamics Simulations on the High Temperature Behavior of Myoglobin

#### 3.1 Introduction

The native conformations of typical globular proteins are tightly folded. These proteins possess a hydrophobic core, while the exterior is dominated by charged and polar side chains that interact favorably with water.<sup>1</sup> In addition to the hydrophobic effect, native proteins are stabilized by H-bonds, van der Waals contacts, salt bridges, and (sometimes) disulfide bonds.<sup>2, 3</sup> Unfolding can be triggered by exposure to non-physiological temperatures, extremes of pH, or chemical denaturants.<sup>4-6</sup> Many unfolding transitions show two-state behavior ( $N \rightarrow U$ ),<sup>7-10</sup> although partially folded intermediates can become populated for some proteins.<sup>11-13</sup>

Protein aggregation is an enigmatic phenomenon that is closely intertwined with the question how proteins fold and unfold.<sup>14, 15</sup> Aggregation refers to the assembly of proteins into non-native higher-order structures. The widespread interest in protein aggregates is based on their involvement in various diseases, including Alzheimer's and Parkinson's, ALS, and many others.<sup>16-21</sup> Also, aggregation can limit the shelf life and efficacy of protein therapeutics such as monoclonal antibodies.<sup>22</sup> Aggregates of such protein therapeutics have been linked to adverse immune responses.<sup>23</sup> Understanding the mechanisms of protein aggregation *in vitro* and *in vivo*, therefore, is of great interest for a wide range of applications.

Protein aggregates come in many shapes and sizes.<sup>14, 16</sup> Their heterogeneous and disordered nature usually precludes the application of high-resolution structure determination

methods. Even the use of standard spectroscopic tools (such as CD spectroscopy) is challenging due to light scattering and solubility issues.<sup>24</sup> As a result, the structures of aggregates and their formation mechanisms remain poorly understood.<sup>14, 16, 25</sup> Many practitioners will have witnessed aggregation in samples that were unstable. Such aggregates tend to be amorphous, and micrometer sized, causing Rayleigh scattering that gives the degraded samples a turbid (“cloudy”) appearance.<sup>24, 26-28</sup>

Amorphous aggregates are insoluble, and they can be collected as a pellet after spinning the samples in a standard centrifuge.<sup>26, 27, 29, 30</sup> Various numerical approaches have been applied for modeling the formation kinetics of amorphous aggregates.<sup>28, 31</sup> Aggregation of most proteins is based solely on noncovalent contacts.<sup>16</sup> The intermolecular clustering of hydrophobic side chains appears to be highly important in this context,<sup>14, 32, 33</sup> sometimes in combination with H-bonding across intermolecular  $\beta$ -sheets.<sup>24, 34, 35</sup> Other types of interactions may participate as well,<sup>14</sup> including intermolecular disulfide bridges for Cys-containing proteins.<sup>36, 37</sup> Amyloid fibrils are a special type of aggregate that is relatively ordered and has a cross- $\beta$  structure.<sup>16, 38, 39</sup> Although amyloid is associated with numerous diseases,<sup>16</sup> the actual cytotoxic species are likely not full-length fibrils but smaller oligomers.<sup>20, 40-42</sup>

Aggregation can be promoted by exposing proteins to destabilizing conditions.<sup>43</sup> Heating is particularly effective in this regard,<sup>44-47</sup> especially for solutions that are highly concentrated.<sup>14, 29, 33</sup> In contrast, heating at low concentration favors reversible unfolding such that thermodynamic parameters can be measured, e.g., by differential scanning calorimetry (DSC) or in optical experiments.<sup>10, 48</sup> Chemical denaturants such as urea and

guanidinium hydrochloride tend to solubilize non-native chains, such that aggregation is less prevalent than with heating.<sup>24, 49</sup>

Two main aggregation mechanisms have been proposed in the literature. Many studies emphasize the role of partially folded intermediates as aggregate precursors (N → Partially Folded → Aggregated).<sup>16, 17, 45-47, 50, 51</sup> Others envision that aggregation commences from globally unfolded conformers (N → U → Aggregated).<sup>35, 52-55</sup> In either case, aggregates start out as small soluble complexes that grow into larger insoluble assemblies as more and more chains associate with the initial nuclei.<sup>21, 30, 56</sup> In addition to aggregation in bulk solution (which is the topic of the current study), there is also the possibility of aggregation at liquid/vapor interfaces, e.g., in solutions that contain air bubbles.<sup>57, 58</sup>

Molecular dynamics (MD) simulations have become a key tool for exploring protein folding and dynamics.<sup>59-63</sup> Surprisingly, this technique remains under-utilized when it comes to protein aggregation, as there are only relatively few MD investigations in this area. Most of those studies have focused on amyloid formation<sup>20, 64-69</sup> using simple coarse-grained force fields with or without explicit solvent.<sup>68</sup> Some others explored specific aspects of aggregation, such as cytotoxic SOD1 oligomers,<sup>40</sup> urea effects on short peptides,<sup>70</sup> and cataract formation from crystallin.<sup>33</sup> However, there have been very few attempts to model the formation of amorphous aggregates from common globular proteins using atomistic MD simulations with explicit solvent.

To enhance the general understanding of protein aggregation, the current work examines the behavior of myoglobin (Mb) at elevated temperature. Mb is a well suited model for this purpose, because of its paradigmatic role in earlier studies related to protein structure determination,<sup>71</sup> folding/unfolding,<sup>5, 11</sup> thermodynamics,<sup>48</sup> conformational fluctuations,<sup>72</sup>

and amyloid formation.<sup>73</sup> Native Mb has a globular structure that comprises eight helices (A-H), with a hydrophobic core and a hydrophilic exterior.<sup>74</sup> Among its 153 amino acids there are no Cys residues, eliminating complications related to disulfide bonding. Heating of Mb triggers the formation of amorphous (i.e., non-amyloid)<sup>14</sup> aggregates.<sup>24, 35</sup> IR spectroscopy suggests that these aggregates have partial  $\beta$ -sheet structure, and that some of the  $\beta$ -sheets are involved in intermolecular H-bonding.<sup>24,35</sup> Just like for other aggregates, hydrophobic contacts and other types of interactions likely play a role as well.<sup>14, 32, 33</sup> Molecular details of Mb aggregation remain unknown. Here we perform experiments and atomistic MD simulations in explicit solvent to uncover mechanistic aspects of heat-induced Mb aggregation. We find that aggregation proceeds through the interaction of globally unfolded chains, and we provide detailed insights of the highly dynamic events that culminate in the formation of higher order assemblies.

## 3.2 Methods

H Horse-heart ferri-Mb was from Sigma (St. Louis, MO). Initial stock solutions prepared at room temperature were centrifuged to remove small amounts of insoluble debris. All samples contained 50 mM sodium phosphate buffer and 100 mM NaCl in water at neutral pH. Circular dichroism (CD) data were acquired between 20 °C and 96 °C on a Jasco J-810 instrument (Easton, MD) with a 1 mm cuvette using 5  $\mu$ M Mb. Unfolding profiles were generated by monitoring the CD signal at 222 nm. These profiles were analyzed using the expression<sup>2, 10</sup>

$$CD_{222} = \frac{(y_N + m_N T) + (y_U + m_U T) \exp(-\Delta G(T)/RT)}{1 + \exp(-\Delta G(T)/RT)}$$

**Equation 3.1**

where the free energy of unfolding is  $\Delta G(T) = \Delta H(1 - T/T_m)$ , and where  $T_m$  denotes the melting temperature. Following established protocols,<sup>10</sup> the enthalpy of unfolding ( $\Delta H$ ) was assumed to be constant for the temperature range considered here. The  $(y_N + m_N T)$  and  $(y_U + m_U T)$  terms in equation 3.1 represent the pre- and post-transition baselines, respectively. From the fitted parameters, the fraction of globally unfolded protein  $f_U$  can be calculated as

$$f_U = \frac{\exp(-\Delta G(T)/RT)}{1 + \exp(-\Delta G(T)/RT)}$$

**Equation 3.2**

Aggregation assays were conducted by immersing Mb samples at concentrations between 5  $\mu\text{M}$  and 100  $\mu\text{M}$  in a  $T$ -controlled water bath for various time intervals (6 s to 100 min). The samples were then centrifuged (10 min, 13000 g) for precipitate removal, and the supernatant was analyzed using a Cary 100 spectrophotometer (Varian, Mississauga, ON) to quantify the leftover soluble Mb at 409 nm. Some of the samples had to be diluted to ensure absorbance readings  $< 1$ .

### 3.3 Results and Discussion

#### 3.3.1 Thermal Aggregation Experiments.

As an initial step, we characterized the heat-induced aggregation of Mb experimentally, with the goal of establishing realistic conditions for subsequent MD simulations. Thermal

aggregation was probed by heating of aqueous Mb solutions, followed by centrifugation for removal of precipitated aggregates. UV-Vis spectroscopy was then used to quantify the residual (non-aggregated) protein. Aggregation was negligible for samples that had been heated for 20 min regardless of protein concentration, as long as the solution temperature did not exceed 348 K (Figure 3.1A). At higher temperatures aggregation became prevalent, especially for high concentrations. For example, exposure of 100  $\mu\text{M}$  Mb to 358 K for 20 min caused almost complete aggregation (Figure 3.1A). For temperatures above 348 K, aggregation became more prevalent when the heat exposure time was increased (Figure 3.1B). Overall, the assays of Figure 3.1A,B confirm the expected trends,<sup>11, 29</sup> i.e., an increase of aggregation with increasing temperature, protein concentration, and time.

### 3.3.2 Relationship between Global Unfolding and Aggregation.

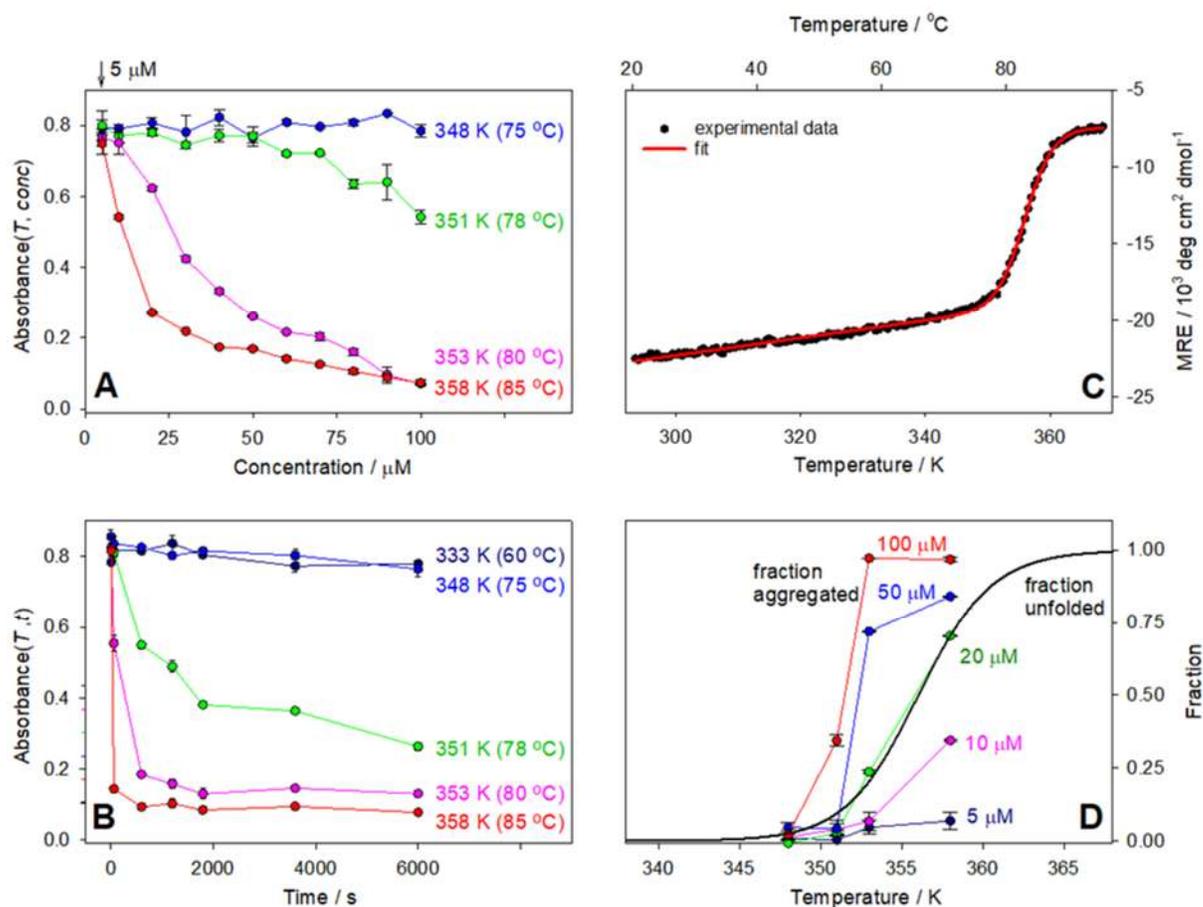
The aforementioned experiments revealed that Mb aggregation can be prevented by using low protein concentrations (5  $\mu\text{M}$ ). These conditions were used for unfolding experiments that monitored changes in  $\alpha$ -helicity by CD spectroscopy at 222 nm.<sup>75</sup> Thermal unfolding data generated in this way are exemplified in Figure 3.1C. CD melting experiments were repeated five times at scan rates of 1 K  $\text{min}^{-1}$  and 4 K  $\text{min}^{-1}$ . The profiles obtained in this way were all very similar. Fits on the basis of equation 3.1 resulted in  $T_m = 356.2 \pm 0.6$  K and  $\Delta H = 450 \pm 20$  kJ  $\text{mol}^{-1}$ .

Evidently, equation 3.1 describes the experimental melting profiles very well (Figure 3.1C). This equation is based on a two-state model.<sup>10</sup> In other words, our data indicate that thermally-induced global breakdown of the Mb helical structure can be approximated as a N  $\rightarrow$  U two-state process. This finding is supported by DSC data,<sup>48</sup> and by the results from

chapter 2 that showed Mb has a melting at  $\sim 204$  nm isodichroic point that represents a hallmark of two-state helix  $\rightarrow$  coil transitions. Despite the two-state character of this transition, IR data reveal that global unfolding is preceded by subtle conformational changes around 330 K,<sup>76</sup> i.e.,  $\sim 26$  K below the global unfolding transition of Figure 3.1C. Mass spectrometry experiments have attributed this pre-transition to heme loss from the protein.<sup>77</sup>

Figure 3.1D compares Mb aggregation with the fraction of globally unfolded protein ( $f_U$ , from equation 3.2). Even at the highest concentration, aggregation is negligible up to  $T = 348$  K where the fraction of globally unfolded protein remains close to zero ( $f_U \approx 3\%$  at 348 K). For higher temperatures  $f_U$  rises sharply, concomitant with dramatically increased aggregation. Hence, the aggregation propensity is closely correlated with the fraction of globally unfolded Mb. In contrast, the aforementioned heme loss at  $\sim 330$  K is *not* sufficient for triggering aggregation. Taken together, these data strongly suggest that thermal aggregation of Mb proceeds from the globally unfolded state, via a mechanism that can be expressed (in simplified form) as  $N \rightarrow U \rightarrow \text{Aggregated}$ . IR data on Mb support this mechanism<sup>35</sup> which also seems to be operative for some other proteins.<sup>52-54</sup> We do not rule out that there are proteins that aggregate via partially folded intermediates,<sup>16, 17, 45-47, 50, 51</sup> but aggregation of heated Mb appears to result from the interaction of globally unfolded chains.

Overall, the experiments of Figure 3.1 identify suitable conditions for the subsequent MD runs. Accordingly, we performed aggregation simulations on heated proteins that were completely unfolded, rather than using semi-folded conformations.

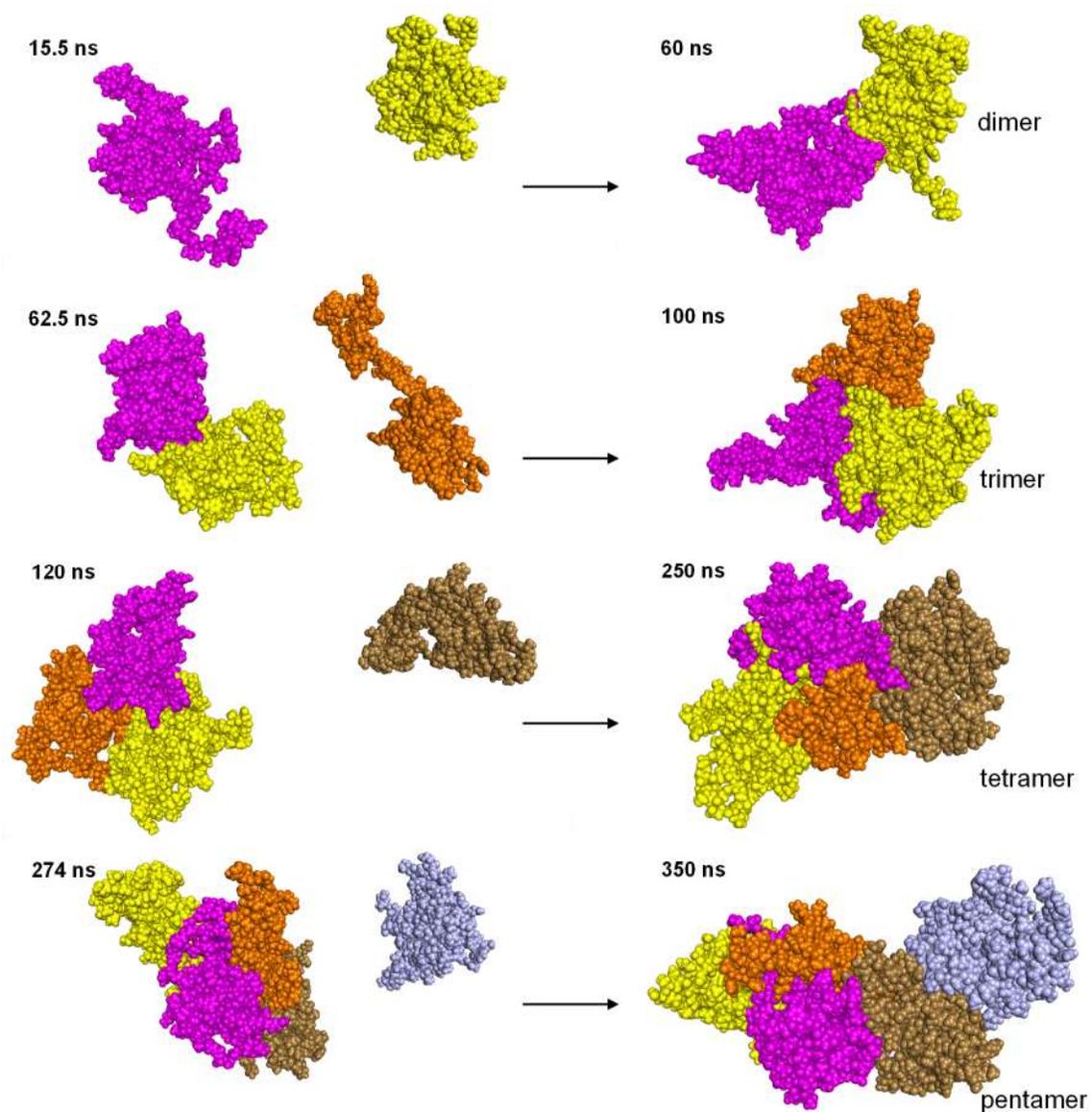


**Figure 3.1.** *Mb* aggregation monitored by UV-Vis spectroscopy as a function of protein concentration, temperature, and time. (A) Aggregation after 20 min of heat exposure at different *Mb* concentrations. (B) Aggregation of 100  $\mu\text{M}$  *Mb* at different temperatures. (C) Thermal unfolding of 5  $\mu\text{M}$  *Mb* monitored by CD spectroscopy at 222 nm, with a fit based on equation 3.1. (D) Colored data represent the fraction of aggregated *Mb* after 20 min of heat exposure at different protein concentrations. Also shown in panel D is the fraction of unfolded protein ( $f_U$ ), deduced from the CD data via equation 3.2

### 3.3.3 Overview of Aggregation MD Data

All simulations were performed at 370 K, in accordance with the experiments of Figure 3.1 which demonstrated that aggregation was most pronounced at the highest temperatures.

Snapshots for one of these runs are shown in Figure 3.2, illustrating the association of two thermally unfolded monomers into a dimer, with subsequent trimer, tetramer, and pentamer formation. The association time required for newly added monomers was variable, ranging from tens to hundreds of nanoseconds. The protein backbone in the pentameric aggregates was mostly coiled, although there were a few short  $\alpha$ -helical and  $\beta$ -sheet segments.



**Figure 3.2.** MD simulation snapshots, illustrating the stepwise assembly of a pentameric Mb aggregate from thermally unfolded monomers. Each monomer is shown in a different color. Panels on the right show the structures of aggregated complexes, just prior to addition of a new monomer into the simulation box. The cumulative simulation time is indicated in each frame. These MD data were generated by Yuen Ki Ng, with the help of Pablo M. Scrosati.<sup>78</sup>

### 3.3.4 Conclusions

The *in vitro* aggregation of heat-unfolded proteins such as Mb ultimately produces micrometer sized (or even larger) amorphous assemblies.<sup>24, 26, 27</sup> Unfortunately, such large systems are beyond the size range that is accessible to atomistic MD simulations. The current work for the first time provides detailed *in silico* insights into the initial steps of these assembly processes, from heat-unfolded monomers to pentameric aggregates. The assembly mode pursued in our MD runs, where aggregates grow via attachment of monomers, will dominate the early stages of aggregation because monomers are the most abundant solution species early during the reaction. Growth via monomer attachment may continue even for larger assemblies, as proposed for A $\beta$  fibrils<sup>68, 79</sup> and other types of aggregates.<sup>21</sup> An alternative scenario for the Mb system studied here is that large aggregates arise from binding of smaller oligomers to one another.

It is tempting to speculate on the implications of our findings for the aggregation of other small, single-domain proteins. The experiments of Figure 1 imply that Mb and some other proteins<sup>35, 52-55</sup> aggregate via globally unfolded chains (N  $\rightarrow$  U  $\rightarrow$  Aggregated). This view contrasts reports that emphasize the role of partially folded species (N  $\rightarrow$  Partially Folded  $\rightarrow$  Aggregated).<sup>16, 17, 45-47, 50, 51</sup> It is possible that both models are adequate, and that the mechanisms are protein specific. Alternatively, global unfolding may play a more central role than previously thought, i.e., formation of U could be an obligatory step for aggregation of most proteins. In other words, we speculate that even proteins with purported “aggregation-prone” semi-folded species might follow a mechanism such as N  $\rightarrow$  Partially Folded  $\rightarrow$  U  $\rightarrow$  Aggregated. Support for this idea comes from the fact that conditions favoring partially folded structures will also give the globally unfolded state a relatively high Boltzmann weight.<sup>80</sup> We reiterate that these considerations apply to small, single-domain proteins. The situation is likely different for larger systems such as IgGs, where individual domains unfold sequentially, and where unfolding of one domain may be sufficient for triggering aggregation.<sup>81, 82</sup> It is hoped that future studies will shed additional light on these and other questions related to protein dynamics and aggregation. The

computational strategies devised in the current work seem well suited for tackling many of these issues.

## 3.4 References

- 1 Dill, K. A., J. L. MacCallum, *Science* **2012**, 338, 1042.
- 2 Fersht, A. R., *Structure and Mechanism in Protein Science*. Editor, W. H. Freeman & Co., New York, **1999**.
- 3 Creighton, T. E., *Proteins*. Editor, W. H. Freeman & Co, New York, **1993**.
- 4 Jha, A. K., A. Colubri, K. F. Freed, T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 13099.
- 5 Eliezer, D., J. Yao, H. J. Dyson, P. E. Wright, *Nat. Struct. Biol.* **1998**, 5, 148.
- 6 Goto, Y., L. J. Calciano, A. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, 87, 573.
- 7 Rhoades, E., M. Cohen, B. Schuler, G. Haran, *J. Am. Chem. Soc.* **2004**, 126, 14686.
- 8 Maxwell, K. L., D. Wildes, A. Zarrine-Afsar, M. A. De Los Rios, A. G. Brown, C. T. Friel, L. Hedberg, J.-C. Horng, D. Bona, E. J. Miller, A. Vallée-Bélisle, E. R. G. Main, F. Bemporad, L. Qiu, K. Teilum, N.-D. Vu, A. M. Edwards, I. Ruczinski, F. M. Poulsen, B. B. Kragelund, S. W. Michnick, F. Chiti, Y. Bai, S. J. Hagen, L. Serrano, M. Oliveberg, D. P. Raleigh, P. Wittung-Stafshede, S. E. Radford, S. E. Jackson, T. R. Sosnick, S. Marqusee, A. R. Davidson, K. W. Plaxco, *Protein Sci.* **2005**, 14, 602.
- 9 Zhou, Y., C. K. Hall, M. Karplus, *Protein Sci.* **1999**, 8, 1064.
- 10 Swint, L., A. D. Robertson, *Protein Sci.* **1993**, 2, 2037.
- 11 Hughson, F. M., P. E. Wright, R. L. Baldwin, *Science* **1990**, 249, 1544.
- 12 El-Baba, T. J., D. W. Woodall, S. A. Raab, D. R. Fuller, A. Laganowsky, D. H. Russell, D. E. Clemmer, *J. Am. Chem. Soc.* **2017**, 139, 6306.
- 13 Gu, Z. Y., J. A. Zitzewitz, C. R. Matthews, *J. Mol. Biol.* **2007**, 368, 582.
- 14 Hartl, F. U., M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* **2009**, 16, 574.
- 15 Vendruscolo, M., *Curr. Op. Struct. Biol.* **2012**, 22, 138.
- 16 Chiti, F., C. M. Dobson, *Annu. Rev. Biochem.* **2017**, 86, 27.
- 17 Sekhar, A., J. A. O. Rumfeldt, H. R. Broom, C. M. Doyle, G. Bouvignies, E. M. Meiering, L. E. Kay, *Elife* **2015**, 4, 33.

- 18 Bernstein, S. L., N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condrón, G. Bitan, D. B. Teplow, J.-E. Shea, B. T. Ruotolo, C. V. Robinson, M. T. Bowers, *Nat. Chem.* **2009**, *1*, 326.
- 19 Zhu, S. L., A. Shala, A. Bezginov, A. Sljoka, G. Audette, D. J. Wilson, *PloS One* **2015**, *10*, 1.
- 20 Nguyen, P. H., A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J. E. Shea, N. V. Dokholyan, A. De Simone, B. Y. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. M. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. X. Chen, B. Hnath, B. Strodel, R. Kayed, S. Lesne, G. H. Wei, F. Sterpone, A. J. Doig, P. Derreumaux, *Chem. Rev.* **2021**, *121*, 2545.
- 21 Ahmed, R., G. Melacini, *Biophys. Chem.* **2021**, *269*, 10.
- 22 Kaltashov, I. A., C. E. Bobst, J. Pawlowski, G. B. Wang, *J. Pharm. Biomed. Anal.* **2020**, *184*,
- 23 Moussa, E. M., J. P. Panchal, B. S. Moorthy, J. S. Blum, M. K. Joubert, L. O. Narhi, E. M. Topp, *J. Pharm. Sci.* **2016**, *105*, 417.
- 24 Dong, A. C., T. W. Randolph, J. F. Carpenter, *J. Biol. Chem.* **2000**, *275*, 27689.
- 25 Kazlauskas, R., *Chem. Soc. Rev.* **2018**, *47*, 9026.
- 26 Penzkofer, A., J. Shirdel, P. Zirak, H. Breitkreuz, E. Wolf, *Chem. Phys.* **2007**, *342*, 55.
- 27 Greene, D. G., S. Modla, N. J. Wagner, S. I. Sandler, A. M. Lenhoff, *Biophys. J.* **2015**, *109*, 1716.
- 28 Borgia, M. B., A. A. Nickson, J. Clarke, M. J. Hounslow, *J. Am. Chem. Soc.* **2013**, *135*, 6456.
- 29 Bianco, V., M. Alonso-Navarro, D. Di Silvio, S. Moya, A. L. Cortajarena, I. Coluzza, *J. Phys. Chem. Lett.* **2019**, *10*, 4800.
- 30 Fink, A. L., *Fold. Des.* **1998**, *3*, R9.
- 31 Stranks, S. D., H. Ecroyd, S. Van Sluyter, E. J. Waters, J. A. Carver, L. von Smekal, *Phys. Rev. E* **2009**, *80*, 13.
- 32 Das, U., G. Hariprasad, A. S. Ethayathulla, P. Manral, T. K. Das, S. Pasha, A. Mann, M. Ganguli, A. K. Verma, R. Bhat, S. K. Chandrayan, S. Ahmed, S. Sharma, P. Kaur, T. P. Singh, A. Srinivasan, *PloS One* **2007**, *2*, e1176.

- 33 Wong, E. K., V. Prytkova, J. A. Freites, C. T. Butts, D. J. Tobias, *Biochemistry* **2019**, *58*, 3691.
- 34 Mendoza, V. L., K. Antwi, M. A. Baron-Rodriguez, C. Blanco, R. W. Vachet, *Biochemistry* **2011**, *50*, 6711.
- 35 Meersman, F., L. Smeller, K. Heremans, *Biophys. J.* **2002**, *82*, 2635.
- 36 Yang, M., C. Dutta, A. Tiwari, *J. Phys. Chem. B* **2015**, *119*, 3969.
- 37 Futami, J., A. Miyamoto, A. Hagimoto, S. Suzuki, M. Futami, H. Tada, *Sci. Rep.* **2017**, *7*, 10.
- 38 Ow, S. Y., D. E. Dunstan, *Protein Sci.* **2014**, *23*, 1315.
- 39 Radamaker, L., Y. H. Lin, K. Annamalai, S. Huhn, U. Hegenbart, S. O. Schonland, G. Fritz, M. Schmidt, M. Fandrich, *Nat. Commun.* **2019**, *10*, 1103.
- 40 Sangwan, S., A. Zhao, K. L. Adams, C. K. Jayson, M. R. Sawaya, E. L. Guenther, A. C. Pan, J. Ngo, D. M. Moore, A. B. Soriaga, T. D. Do, L. Goldschmidt, R. Nelson, M. T. Bowers, C. M. Koehler, D. E. Shaw, B. G. Novitch, D. S. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8770.
- 41 Zhang, Y., D. L. Rempel, J. Zhang, A. K. Sharma, L. M. Mirica, M. L. Gross, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14604.
- 42 Laganowsky, A., C. Liu, M. R. Sawaya, J. P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. B. Soriaga, M. Landau, P. K. Teng, D. Cascio, C. Glabe, D. Eisenberg, *Science* **2012**, *335*, 1228.
- 43 Stathopoulos, P. B., G. A. Scholz, Y. M. Hwang, J. A. O. Rumfeldt, J. R. Lepock, E. M. Meiering, *Protein Sci.* **2004**, *13*, 3017.
- 44 Yan, Y. B., Q. Wang, H. W. He, H. M. Zhou, *Biophys. J.* **2004**, *86*, 1682.
- 45 Longo, L. M., Y. Gao, C. A. Tenorio, G. Wang, A. K. Paravastu, M. Blaber, *Protein Sci.* **2018**, *27*, 431.
- 46 Shivu, B., S. Seshadri, J. Li, K. A. Oberg, V. N. Uversky, A. L. Fink, *Biochemistry* **2013**, *52*, 5176.
- 47 Horvath, G., L. Biczok, Z. Majer, M. Kovacs, A. Micsonai, J. Kardos, O. Toke, *Febs J.* **2017**, *284*, 3637.
- 48 Privalov, P. L., Y. V. Griko, S. Y. Venyaminov, V. P. Kutysenko, *J. Mol. Biol.* **1986**, *190*, 487.
- 49 Clark, E. D., E. Schwarz, R. Rudolph, *Meth. Enzymol.* **1999**, *309*, 217.

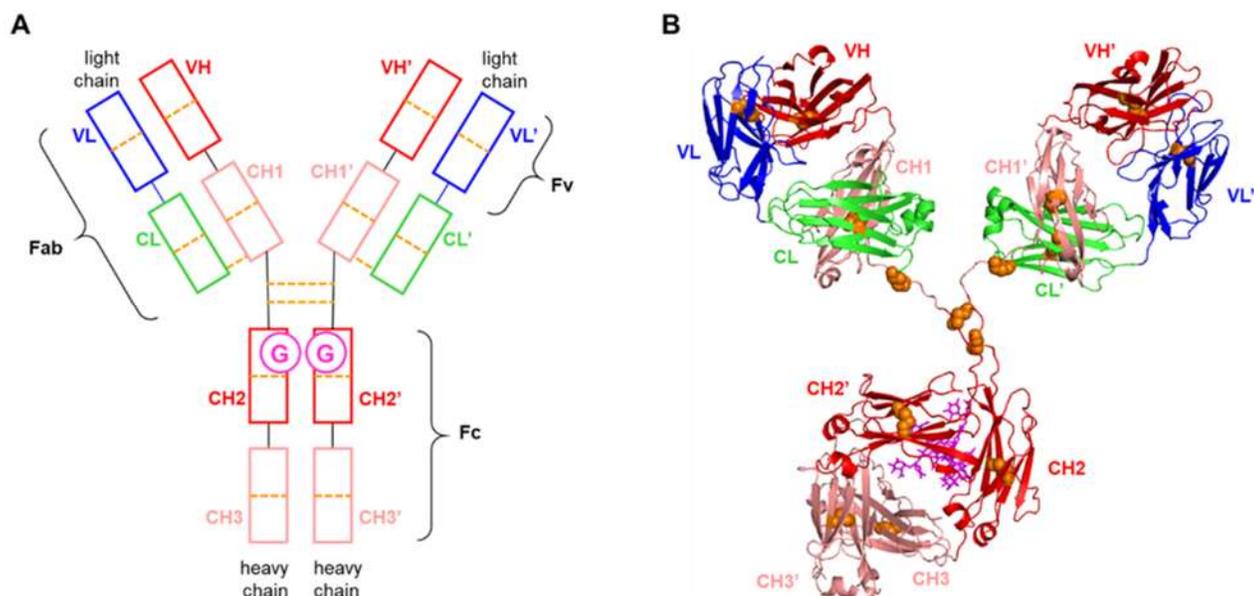
- 50 Clark, P. L., *T. Biochem. Sci.* **2004**, *29*, 527.
- 51 Wetzel, R., *Cell* **1996**, *86*, 699.
- 52 Mashaghi, A., S. Bezrukavnikov, D. P. Minde, A. S. Wentink, R. Kityk, B. Zachmann-Brand, M. P. Mayer, G. Kramer, B. Bukau, S. J. Tans, *Nature* **2016**, *539*, 448.
- 53 Hingorani, K. S., M. C. Metcalf, D. T. Deming, S. C. Garman, E. T. Powers, L. M. Gierasch, *Nat. Chem. Biol.* **2017**, *13*, 369.
- 54 Benjwal, S., S. Verma, K.-H. Röhm, O. Gursky, *Protein Sci.* **2006**, *15*, 635.
- 55 De Young, L. R., K. A. Dill, A. L. Fink, *Biochemistry* **1993**, *32*, 3877.
- 56 Wang, G., R. R. Abzalimov, I. A. Kaltashov, *Anal. Chem.* **2011**, *83*, 2870.
- 57 Xiao, Y., L. Konermann, *Protein Sci.* **2015**, *24*, 1247.
- 58 Wiesbauer, J., R. Prassl, B. Nidetzky, *Langmuir* **2013**, *29*, 15240.
- 59 Piana, S., K. Lindorff-Larsen, D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915.
- 60 Prigozhin, M. B., M. Gruebele, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3372.
- 61 Sultan, M. M., R. A. Denny, R. Unwalla, F. Lovering, V. S. Pande, *Scientific Reports* **2017**, *7*,
- 62 van der Spoel, D., *Curr. Opin. Struct. Biol.* **2021**, *67*, 18.
- 63 Lindahl, E., in *Molecular Modeling of Proteins. Methods in Molecular Biology (Methods and Protocols)*, ed. by A. Kukol, Humana Press, New York, NY, **2015**, Vol. 1215, pp. 3.
- 64 Wei, G. H., N. Mousseau, P. Derreumaux, *Prion* **2007**, *1*, 3.
- 65 Ma, B. Y., R. Nussinov, *Curr. Opin. Chem. Biol.* **2006**, *10*, 445.
- 66 Davidson, D. S., A. M. Brown, J. A. Lemkul, *J. Mol. Biol.* **2018**, *430*, 3819.
- 67 Carballo-Pacheco, M., B. Strodel, *J. Phys. Chem. B* **2016**, *120*, 2991.
- 68 Wu, C., J.-E. Shea, *Curr. Op. Struct. Biol.* **2011**, *21*, 209.
- 69 Ilie, I. M., A. Caflisch, *Chem. Rev.* **2019**, *119*, 6956.
- 70 Cai, Z., J. Li, C. Yin, Z. Yang, J. Wu, R. Zhou, *J. Phys. Chem. B* **2014**, *118*, 48.

- 71 Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662.
- 72 Frauenfelder, H., G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, R. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129.
- 73 Fändrich, M., M. A. Fletcher, C. M. Dobson, *Nature* **2001**, *410*, 165.
- 74 Maurus, R., C. M. Overall, R. Bogumil, Y. Luo, A. G. Mauk, M. Smith, G. D. Brayer, *Biochim. Biophys. Acta* **1997**, *1341*, 1.
- 75 Greenfield, N. J., *Anal. Biochem.* **1996**, *235*, 1
- 76 Yan, Y. B., Q. Wang, H. W. He, X. Y. Hu, R. Q. Zhang, H. M. Zhou, *Biophys. J.* **2003**, *85*, 1959.
- 77 Woodall, D. W., L. W. Henderson, S. A. Raab, K. Honma, D. E. Clemmer, *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 64.
- 78 Ng, Y. K., N. N. Tajoddin, P. M. Scrosati, L. Konermann, *J. Phys. Chem. B* **2021**, *125*, 13099.
- 79 Sanchez, L., S. Madurga, T. Pukala, M. Vilaseca, C. Lopez-Iglesias, C. V. Robinson, E. Giralt, N. Carulla, *J. Am. Chem. Soc.* **2011**, *133*, 6505.
- 80 Englander, S. W., L. Mayne, Z. Y. Kan, W. B. Hu, *Annu. Rev. Biophys.* **2016**, *45*, 135.
- 81 Vermeer, A. W. P., W. Norde, *Biophys. J.* **2000**, *78*, 394.
- 82 Wang, G. B., P. V. Bondarenko, I. A. Kaltashov, *Analyst* **2018**, *143*, 670.

## 4 Chapter 4. Structural Dynamics of a Thermally Stressed Monoclonal Antibody Characterized by Temperature-Dependent H/D Exchange Mass Spectrometry

### 4.1 Introduction

Protein therapeutics play an ever-increasing role in medicine.<sup>1-3</sup> Above all, monoclonal antibodies (mAbs) and mAb-derived antibody-drug conjugates (ADCs) of the IgG1 subclass have had a major impact for treatment of cancer, as well as infectious and autoimmune diseases.<sup>4-7</sup> IgG1 proteins consist of two identical heavy chains and two identical light chains that are linked by disulfide bonds. Each heavy chain has one variable domain (VH) and three constant domains (CH1, CH2, CH3). Both light chains have one variable (VL) and one constant domain (CL). Each of the two Fab arms comprises CH1 and CL, along with VH and VL which form the Fv moiety that is responsible for antigen binding. The Fc substructure consists of CH2/CH2' and CH3/CH3' domains, the former being decorated with N-linked glycans (Figure 4.1A).<sup>7-9</sup>



**Figure 4.1.** (A) Cartoon representation of an IgG1 mAb. Disulfides are represented as dashed orange lines, “G” indicates glycans. (B) Structure of NSTmAb, using the same coloring as in panel A.<sup>10</sup> Disulfide-linked Cys residues are shown in orange, glycans are shown as pink sticks. The Fab moieties have  $2 \times 431$  residues, CH2 and CH3 have  $2 \times 118$  and  $2 \times 109$  residues, respectively.

Protein therapeutic formulations are required to survive for at least two years in solution at 4 °C without aggregation, chemical modification, or surface adsorption.<sup>10, 11</sup> Excipients can help suppress such degradation events. Unfortunately, laborious screening is required to identify suitable conditions for each new product.<sup>11, 12</sup> Monitoring the stability of proteins in real time (years) is incompatible with an efficient product development pipeline. Accelerated screening assays address this issue by exposing proteins to unfavorable thermal, mechanical, or pH environments.<sup>11-14</sup> Thermal stability assays are particularly important, based on the premise that proteins with a high melting temperature ( $T_m$ ) usually degrade more slowly during long-term storage.<sup>11-15</sup>

Differential scanning calorimetry (DSC) is widely used for assessing the thermal stability of mAbs and other protein therapeutics.<sup>7, 12, 15-17</sup> For two-state unfolding of the native state

to the unfolded state, DSC thermograms show a single peak at  $T_m$ . The peak area represents the enthalpy of global unfolding ( $\Delta H_{glob} > 0$ ).<sup>18, 19</sup> Thermodynamic analyses require that unfolding is reversible.<sup>18, 19</sup> Unfortunately, this is not the case for mAbs, which undergo irreversible aggregation after thermal unfolding.<sup>7, 17, 20, 21</sup> DSC profiles of mAbs are further complicated by the presence of several unfolding transitions that arise from the multi-domain architecture of the protein (Figure 4.1A).<sup>7, 14, 17, 20-23</sup> Such data can only be analyzed in a semi-quantitative fashion,<sup>18, 19</sup> relying on  $T_m$  values as indicators of protein robustness.<sup>7, 14, 17, 20-25</sup>

DSC and related techniques report on global unfolding.<sup>18, 19, 26, 27</sup> However, proteins also undergo a wide range of local fluctuations<sup>28, 29</sup> that are crucial for biological function<sup>30-32</sup> and that may play a role during aggregation.<sup>33, 34</sup> A comprehensive characterization of these dynamics is highly desirable. Hydrogen/deuterium exchange (HDX) mass spectrometry (MS) has become a key tool in this context,<sup>35, 36</sup> specifically for protein therapeutics.<sup>3, 12, 25, 37-44</sup> HDX-MS monitors the deuteration of backbone NH sites in D<sub>2</sub>O. HDX is mediated by  $\text{NH}_{\text{closed}} \leftrightarrow \text{NH}_{\text{open}}$  fluctuations, i.e., transient unfolding events that are associated with H-bond disruption.<sup>45</sup> In the EX2 regime<sup>29</sup> HDX proceeds with an overall rate constant

$$k_{HDX} = K_{op} \times k_{ch}$$

#### Equation 4.1

where  $K_{op}$  is the  $\text{NH}_{\text{closed}} \leftrightarrow \text{NH}_{\text{open}}$  equilibrium constant, and  $k_{ch}$  is the “chemical” rate constant.<sup>46</sup>

Traditional HDX-MS experiments monitor deuteration as a function of time ( $t$ ) at constant temperature ( $T$ ). Keeping in mind the need for investigating thermally stressed proteins,<sup>11-15</sup> it is an obvious question whether  $T$  can serve as additional HDX-MS variable. The key problem with  $T$ -dependent HDX-MS is that both terms in equation 4.1 change with  $T$ . It is therefore unclear to what extent HDX alterations can be ascribed to changes of the protein ( $K_{op}$ ) vs. changes of the labeling chemistry ( $k_{ch}$ ). Only the former is of interest for

characterizing protein behavior. With some exceptions,<sup>47</sup> earlier  $T$ -dependent HDX work did not clearly separate the two contributions.<sup>48-54</sup> Using myoglobin as a model system, we recently proposed a thermodynamics model to dissect  $T$ -dependent HDX-MS data into  $K_{op}(T)$  and  $k_{ch}(T)$  contributions (Chapter 2). The key ideas of this strategy are as follows:

Protein dynamics at ambient  $T$  are dominated by local fluctuations of the native state,<sup>29, 55</sup> whereas global unfolding/refolding becomes prevalent in the vicinity of  $T_m$ .<sup>19</sup> To capture the participation of both local and global dynamics, equation 4.1 can be modified to

$$k_{HDX}(T) = [ K_{loc}(T) + K^*_{glob}(T) ] \times k_{ch}(T)$$

**Equation 4.2**

$K_{loc}(T)$  refers to local opening,  $K^*_{glob}(T)$  represents global unfolding and,  $k_{ch}(T)$  can be calculated from tabulated data.<sup>46, 56, 57</sup> When expressing  $K$  values via free energy ( $\Delta G$ ), equation 4.2 becomes

$$k_{HDX}(T) = [ \exp(-\Delta G_{loc}/RT) + \exp(-\Delta G^*_{glob}/RT) ] \times k_{ch}(T)$$

**Equation 4.3**

where local fluctuations are governed by enthalpic ( $\Delta H_{loc}$ ) and entropic ( $\Delta S_{loc}$ ) contributions

$$\Delta G_{loc} = \Delta H_{loc} - T \times \Delta S_{loc}$$

**Equation 4.4**

and global dynamics are determined by the free energy

$$\Delta G^*_{glob} = \Delta G_{glob}(T) + \Delta G_{opU}$$

### Equation 4.5

In equation 4.5,  $\Delta G_{glob}(T)$  is the free energy of global unfolding that can be measured by DSC.<sup>19</sup> The  $\Delta G_{opU} > 0$  term in equation 4.5 accounts for residual protection of the unfolded state.<sup>56, 58-62</sup>

The framework outlined above successfully captured the  $T$ -dependent HDX behavior of myoglobin, which represents a relatively simple protein (Chapter 2). In the current work we explored whether the same strategy can be extended to a much more complicated system, i.e., a thermally stressed IgG1 mAb. We focused on the NIST monoclonal antibody (“NISTmAb”, 150 kDa,  $2 \times 663$  residues, Figure 4.1B), a reference system that has been widely studied using crystallography,<sup>63</sup> computational modeling,<sup>64</sup> DSC,<sup>23, 43</sup> and constant-temperature HDX-MS.<sup>39, 43, 65, 66</sup> We found that different regions of NISTmAb displayed distinct types of  $T$ -dependent HDX-MS profiles, reflecting the interplay of local and global dynamics as well as thermal aggregation. We implemented a global fitting strategy that captured this  $T$ -dependent HDX behavior on the basis of equation 4.1-4.5, yielding detailed insights into the thermodynamic properties ( $\Delta H_{loc}$ ,  $\Delta S_{loc}$ ,  $\Delta G_{opU}$ ) of individual protein segments.

## 4.2 Methods

### 4.2.1 Experimental Procedure

NISTmAb Reference Material 8671 was purchased from the National Institute of Standards and Technology.<sup>64</sup> All solutions contained 50 mM sodium phosphate buffer and 100 mM NaCl. DSC data were recorded on a MicroCal VP-DSC instrument at a scan rate of 1 °C min<sup>-1</sup> using a protein concentration 3.3 μM (0.5 mg/ml) at pH 7.0. Analysis of the DSC data was performed using Microsoft Excel.

*T*-dependent HDX experiments were conducted by mixing 66  $\mu\text{M}$  NISTmAb in  $\text{H}_2\text{O}$  with  $\text{D}_2\text{O}$  in a 1:9 ratio, producing 10  $\mu\text{L}$  of 6.6  $\mu\text{M}$  protein in 90%  $\text{D}_2\text{O}$ . Initial experiments were performed at a pH meter reading of 7.0, but those conditions caused premature HDX saturation at elevated temperature (data not shown). To enhance the dynamic range of our experiments, the data discussed below were recorded using slightly more acidic solutions,<sup>46, 48</sup> i.e., pH meter reading of 6.3 (corresponding to pD 6.7).<sup>67</sup> HDX was performed at 0, 23, 30, 40, 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95  $^\circ\text{C}$  in Eppendorf tubes that were immersed in a *T*-controlled water bath. Prior to HDX,  $\text{D}_2\text{O}$  labeling buffer was pre-equilibrated at the desired temperature, while NISTmAb was kept at room temperature to avoid aggregation of the stock solution. After different time points (15 s, 30 s, 2 min, and 20 min) HDX was quenched by 1:1 mixing with 0  $^\circ\text{C}$  aqueous solution containing 8 M urea and 1 M tris(2-carboxyethyl) phosphine (TCEP·HCl) at pH 2.3. This was followed by flash freezing and storage in liquid  $\text{N}_2$ . For analysis, samples were thawed to 0  $^\circ\text{C}$ , and kept on ice for 2 min to allow for TCEP-mediated disulfide reduction.<sup>68, 69</sup> The samples were then diluted by adding 2 volumes of aqueous formic acid (pH 2.3) to prevent pepsin degradation by TCEP. 60  $\mu\text{L}$  aliquots were injected into an Acquity HDX-UPLC (Waters, Milford, MA). Digestion was performed on an immobilized pepsin column (Waters) at 15  $^\circ\text{C}$ . Peptides were separated on a 1.7  $\mu\text{m}$  BEH130 C18 2.1  $\times$  100  $\text{mm}^2$  column using a 20 min water/acetonitrile gradient with 0.2% formic acid at  $\sim 0$   $^\circ\text{C}$ . To prevent carryover, blanks were injected, and the pepsin column was washed with 1.5 M guanidine hydrochloride in water/acetonitrile/formic acid (95.2/4/0.8) after each sample. The UPLC was coupled to a Waters Synapt G2 electrospray mass spectrometer operated in IMS mode.<sup>70</sup> The identity of each peptide was confirmed by  $\text{MS}^E$  on nondeuterated samples with data analysis by Waters PLGS 2.5.3, based on the sequence reported in the NIST Reference Material 8671 Report of Investigation.<sup>64</sup> The sequence coverage was 72% and 92% for the light and heavy chains, respectively (Figure 4.7). For quantitative analyses we only considered peptides with the highest S/N, lowering the coverage to 72% and 76% (Figure 4.8). Centroid  $m/z$  values were determined using Waters DynamX 3.0 and these were converted to percent deuteration using  $\%D(t, T) = (m - m_0)/(m_{100} - m_0) \times 100\%$ , where  $m$  is the centroid, while  $m_0$  and  $m_{100}$  are minimally and fully deuterated controls, respectively.<sup>71</sup> The former was prepared by adding NISTmAb to prequenched  $\text{D}_2\text{O}$  buffer,

followed by flash freezing.  $m_{100}$  samples were prepared in the same way, except that they were incubated for 8 h at 55 °C prior to flash freezing. All %D values are averages of two independent experiments; error bars represent the deviation between these measurements. Protein aggregation precluded data acquisition for the 20 min time point at the two highest temperatures (90 and 95 °C).

## 4.2.2 Data Analysis

The HDX kinetics of peptides with  $N$  non-proline residues covering amino acids  $k$  to  $(k+N-1)$  was modeled as<sup>72-74</sup>

$$\%D_{calc}(t, T) = \frac{1}{(N-2)} \sum_{i=(k+2)}^N [1 - \exp(-k_{HDX,i} \times t)]$$

**Equation 4.6**

This equation excludes the first two residues which undergo complete back exchange during UPLC separation.<sup>56</sup> Each  $k_{HDX}$  value in equation 4.6 depends on  $T$ ,  $\Delta G_{glob}$ ,  $k_{ch}$ ,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$ , and  $\Delta G_{opU}$  (equation 4.3). Measuring  $T$  is trivial.  $\Delta G_{glob}(T)$  was determined by DSC, as discussed in the following section.  $k_{ch}(T)$  values were calculated for each NH site by initially determining  $k_{ch}$  reference values at 298 K.<sup>46, 56</sup> Most of these values fell into a narrow window around an average of 3.1 s<sup>-1</sup> (Figures 4.2, 4.3A). Then,  $k_{ch}(T)$  was obtained via (chapter 2)

$$k_{ch}(T) = k_{B,298} \times \exp\left(-\frac{E_a}{R}(T^{-1} - [298 K]^{-1})\right) \times [OD^-](T)$$

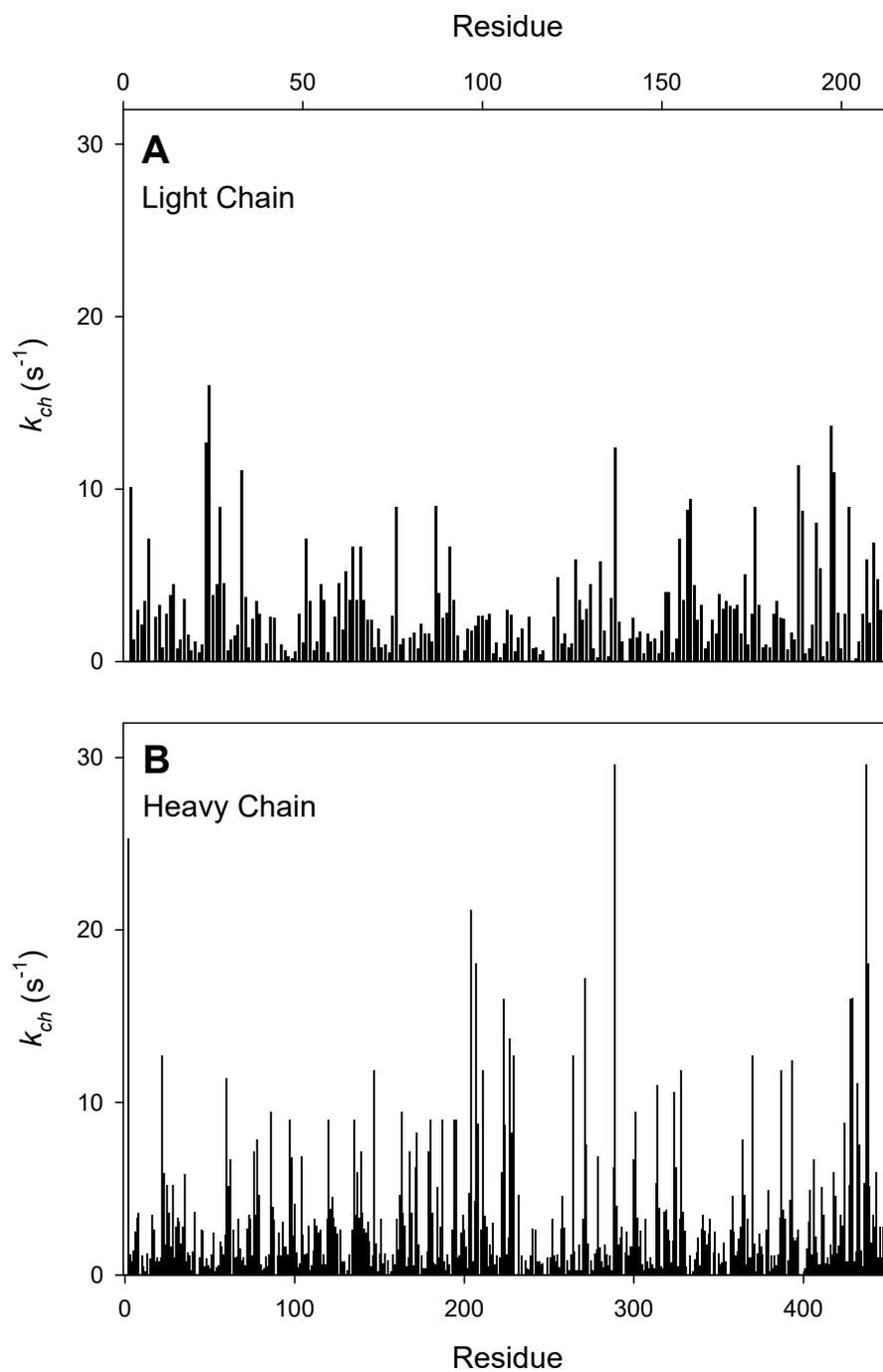
**Equation 4.7**

with<sup>46</sup>  $E_a = 12.6 \text{ kJ mol}^{-1}$ , and  $k_{B\_298}$  chosen such that  $k_{ch}(298 \text{ K})$  from equation 4.7 matched the reference values of Figure 4.2.  $k_{ch}(T)$  profiles generated in this way are exemplified in Figure 4.3A. Finally,  $\Delta H_{loc}$ ,  $\Delta S_{loc}$ , and  $\Delta G_{opU}$  were determined by global least-square fitting that used in Chapter 2 of the experimental  $\%D(t, T)$  data by minimizing the expression

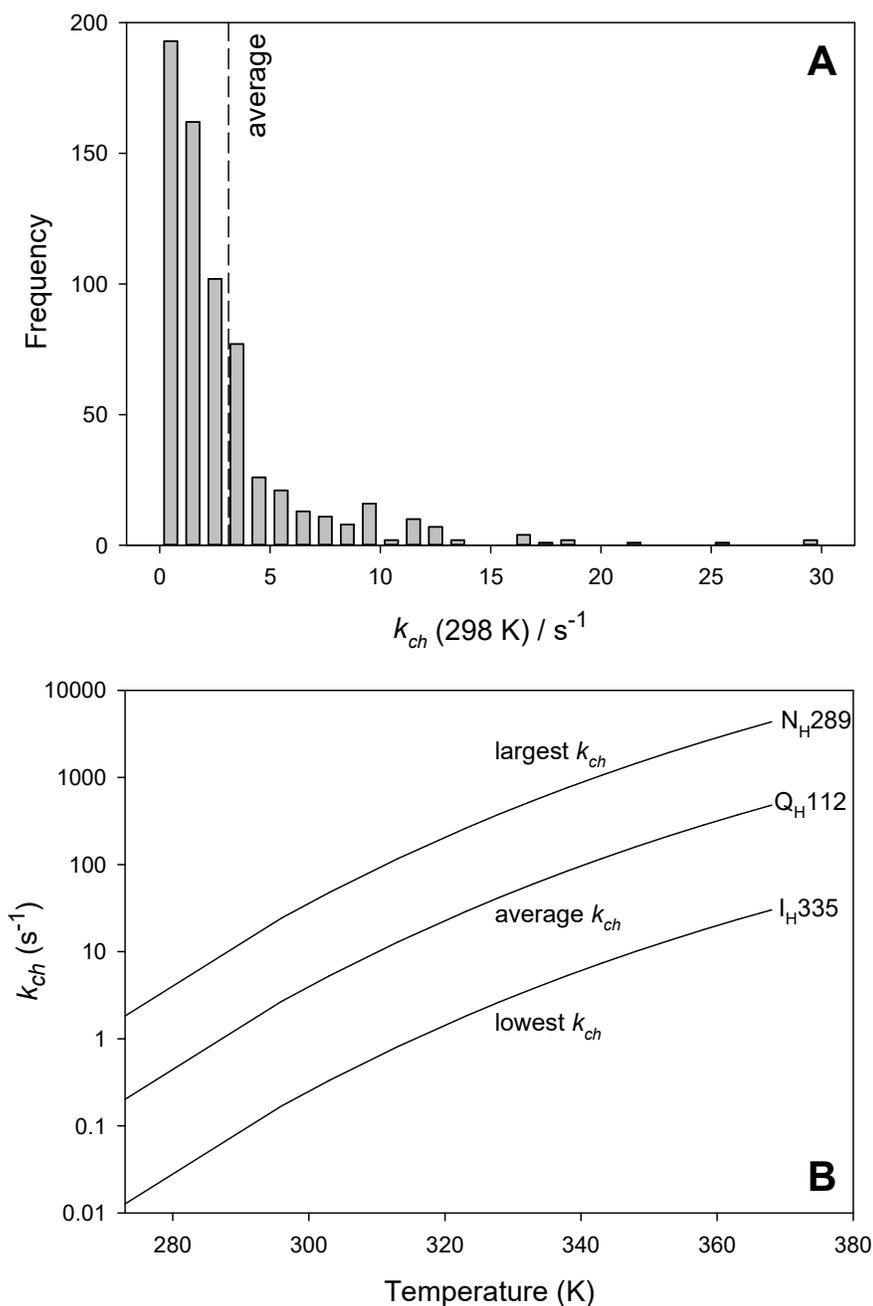
$$X^2 = \sum_{\text{peptides}} \sum_T \sum_t (\%D(t, T) - \%D_{calc}(t, T))^2$$

**Equation 4.8**

where the summation includes all peptides, temperatures, and time points.  $\%D(t, T)$  refers to experimental data, and  $\%D_{calc}(t, T)$  values were calculated using equation 4.6. Additional details are outlined in *Results and Discussion*. All peptide isotope distributions used for our analyses were monomodal, thereby indicating EX2 conditions in accordance with previous data on NISTmAb.<sup>43</sup>



**Figure 4.2.**  $k_{ch}$  values of all NH sites in the (A) light chain and (B) heavy chain of NISTmAb for  $T = 298$  K and  $pD = 6.7$ . The data were generated using *HXrates2018\_HD.xlsx* from the Englander laboratory website.<sup>46, 56</sup>



**Figure 4.3.** (A) Histogram of  $k_{ch}$  values at 298 K for all NH sites in the heavy and light chains, from Figure 4.2. The average  $k_{ch}$  of  $3.1 \text{ s}^{-1}$  is indicated by the vertical dashed line. (B) Temperature dependence of  $k_{ch}$  for three selected backbone amides, representing the fastest and the slowest NH sites, as well as one that is close to the average. The data in panel B were calculated using equation 4.7 (note the logarithmic y-axis).

## 4.3 Results and Discussion

### 4.3.1 DSC Unfolding Experiments

The analysis of  $T$ -dependent HDX-MS data requires knowledge of  $\Delta G_{glob}(T)$  in equation 4.5, that was applied in Chapter 2. For this reason, we characterized the global unfolding of NISTmAb by DSC. Uncorrected DSC data exhibited three endothermic maxima that signify unfolding transitions. These transitions were followed by a drop into the exothermic range at high  $T$ , representing the hallmark of aggregation (Figure 4.4A, black profile). The presence of amorphous aggregates was also apparent from visual inspection of heated samples (Figure 4.12). Consistent with earlier work,<sup>7, 17, 20, 21, 43</sup> our observations show that NISTmAb remains soluble up to roughly 85 °C, while for  $T > 93$  °C the protein behavior is dominated by aggregation and precipitation. We selected data points in these two temperature regions and interpolated them with a polynomial, yielding a background profile that approximates the aggregation contributions to the thermogram (Figure 4.4A, pink symbols and line). Subtraction of this background from the raw data yielded a corrected thermogram that reports on the thermal unfolding of NISTmAb (Figure 4.4B, black solid line).

The baseline-corrected thermogram was deconvoluted into Gaussians using least-square fitting (Figure 4.4B). Melting points determined from Gaussian maxima were  $T_{m1} = 344.4$  K,  $T_{m2} = 357.0$  K, and  $T_{m3} = 362.7$  K, in close agreement with previous work.<sup>23, 43</sup> Regardless of IgG1 sequence, CH2 has the lowest stability,<sup>7, 14, 75</sup> while Fab unfolding is associated with the tallest signal in the thermogram.<sup>7, 14, 22</sup> The three melting points can therefore be assigned as  $T_m(\text{CH2}) = 344.4$  K,  $T_m(\text{CH3}) = 357.0$  K, and  $T_m(\text{Fab}) = 362.7$  K.<sup>14, 43</sup>

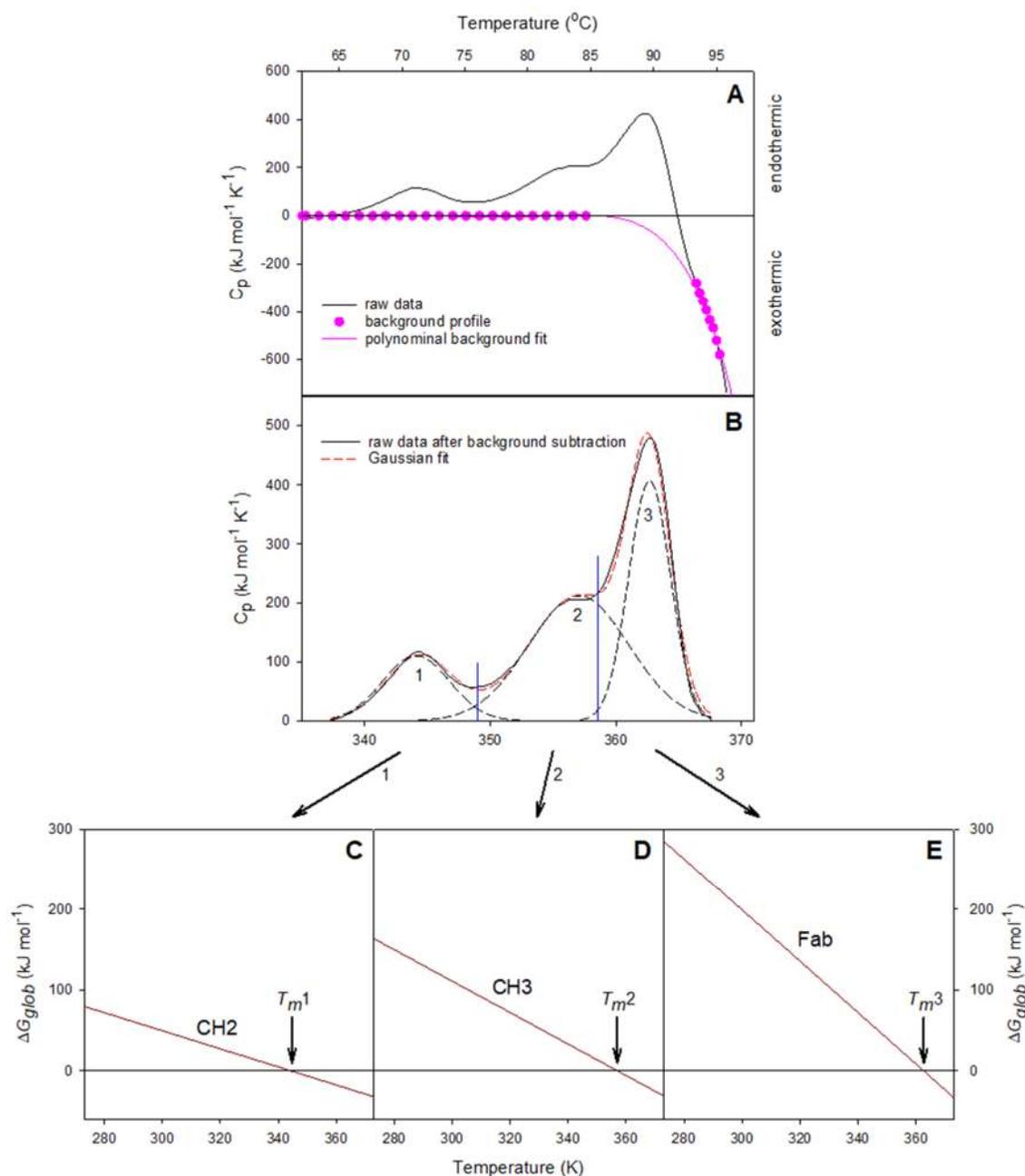
One possibility for estimating  $\Delta H_{glob}$  values associated with the three unfolding transitions would be to use Gaussian areas. However, this is problematic if transitions are asymmetrical and/or show significant overlap (as for  $T > 350$  K in Figure 4.4B).<sup>7, 21</sup> In our case, Gaussian 3 has a smaller area than Gaussian 2, which is inconsistent with the fact that  $\Delta H_{glob}(\text{Fab})$  is known to be larger than  $\Delta H_{glob}(\text{CH3})$ .<sup>7, 14, 22</sup> To sidestep this issue,  $\Delta H_{glob}$  values were determined from the area under the baseline-corrected thermogram, using the

blue vertical lines in Figure 4.4B as integration bounds. The integration results were divided by two because NISTmAb contains two CH2, two CH3, and two Fab (Figure 4.1) and because each of these moieties acts as an independent unfolding unit.<sup>76</sup> The resulting enthalpies are  $\Delta H_{glob}(\text{CH2}) = 385 \text{ kJ mol}^{-1}$ ,  $\Delta H_{glob}(\text{CH3}) = 694 \text{ kJ mol}^{-1}$ , and  $\Delta H_{glob}(\text{Fab}) = 1153 \text{ kJ mol}^{-1}$ . The  $\Delta H_{glob}$  and  $T_m$  values govern the global unfolding free energy of the three moieties according to<sup>27</sup>

$$\Delta G_{glob}(T) = \Delta H_{glob} \times (1 - T/T_m)$$

#### Equation 4.9

which yields the plots of Figure 4.4C-E. In earlier work on myoglobin on Chapter 2, we used a more complex  $\Delta G_{glob}$  expression that considered  $\Delta C_p$  of the unfolding transition, causing  $\Delta G_{glob}(T)$  to be curved. The complexity of IgG1 unfolding dictates that DSC data are generally analyzed using the approximation that  $\Delta C_p = 0$  approximation, corresponding to the linear expression of equation 4.9.<sup>7, 16</sup> Luckily, the  $\Delta C_p$ -related curvature of  $\Delta G_{glob}(T)$  only affects the HDX behavior of proteins that are significantly destabilized at low  $T$ , with possible cold unfolding.<sup>77</sup> For IgG1 proteins such effects can become relevant below  $-20 \text{ }^\circ\text{C}$ ,<sup>78</sup> i.e., not in the  $T$  range considered here. Hence, the linear  $\Delta G_{glob}(T)$  profiles of Figure 4.4C-E are adequate for the purpose of this work.



**Figure 4.4.** DSC analysis of NSTmAb. (A) Uncorrected thermogram, displaying the heat capacity  $C_p$  vs. temperature. The range selected to serve as aggregation background is highlighted by dots. Also shown is the corresponding polynomial background interpolation. (B) Background-corrected thermogram with Gaussian components and overall fit. Blue vertical lines indicate integration bounds for  $\Delta H$  calculations. (C-E)  $\Delta G_{glob}(T)$  profiles extracted from panel B for the CH2, CH3, and Fab moieties.

### 4.3.2 T-Dependent HDX Analyses for Selected Peptides.

We previously applied the thermodynamic framework of equations 4.1-4.7 to myoglobin in Chapter 2. To assess if this framework is also suitable for modeling the  $T$ -dependent HDX behavior of the IgG1 studied here, we initially focused on individual NISTmAb peptides that exhibited different types of  $\%D(t, T)$  profiles (Figures 4.5, 4.6). To reduce the number of fitting parameters, we assumed that residues in close sequence proximity share the same  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  parameters, as suggested by the foldon model.<sup>29, 79</sup> Accordingly, the sequence was dissected into segments of three consecutive residues, each of which was given its own  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  parameters. In a few cases this procedure did not yield satisfactory fits, necessitating the use of two-residue segments and, on rare occasions, one-residue segments. The  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  parameters obtained in this way capture the combined effects of local and global dynamics. Of note, HDX is always mediated by the fluctuation with the lowest  $\Delta G$  (the largest  $K_{op}$ ).<sup>29</sup> Thus, depending on  $T$ , either  $\Delta G_{loc}(T)$  or  $\Delta G^*_{glob}(T)$  may be the prevalent factor for any given NH segment.

Figure 4.5A shows  $\%D(t, T)$  for heavy chain Fab peptide 5-19. Based on the near-exponential rise of  $k_{ch}(T)$  in Figure 4.3B, one might have expected that  $\%D(t, T)$  profiles should exhibit a sudden and very steep increase as  $T$  is raised. Surprisingly, the  $\%D$  data of Figure 4.5A show a very different behavior, i.e., a near-linear increase. Our framework was able to capture this near-linear behavior, evident from the close agreement between fitted and experimental data (Figure 4.5A). Panels B-E illustrate how the overall HDX behavior of peptide 5-19 can be dissected into staggered contributions from individual NH sites, some of which get deuterated at low  $T$  (Figure 4.5D) while others undergo HDX at much higher temperatures (Figure 4.5C). All these deuteration events were attributed to local fluctuations, as evidenced by the  $\Delta G_{loc}$  profiles in Figure 3F-I.

The DSC data of Figure 4.4 revealed that  $T_m(\text{Fab}) = 362.7$  K, implying that global opening/closing fluctuations become prevalent around this temperature.<sup>18, 19</sup> Figure 4.5 illustrates two different scenarios related to the role of these global fluctuations. (1) Amides 7-10 and 15-17 are completely deuterated at  $T_m(\text{Fab})$ , such that global opening/closing has no effect on their HDX behavior (Figure 4.5B/D). Any  $\Delta G_{opU}$  value is compatible with the experimental data under such conditions. The absence of  $\Delta G^*_{glob}$  profiles in Figure 4.5F/H

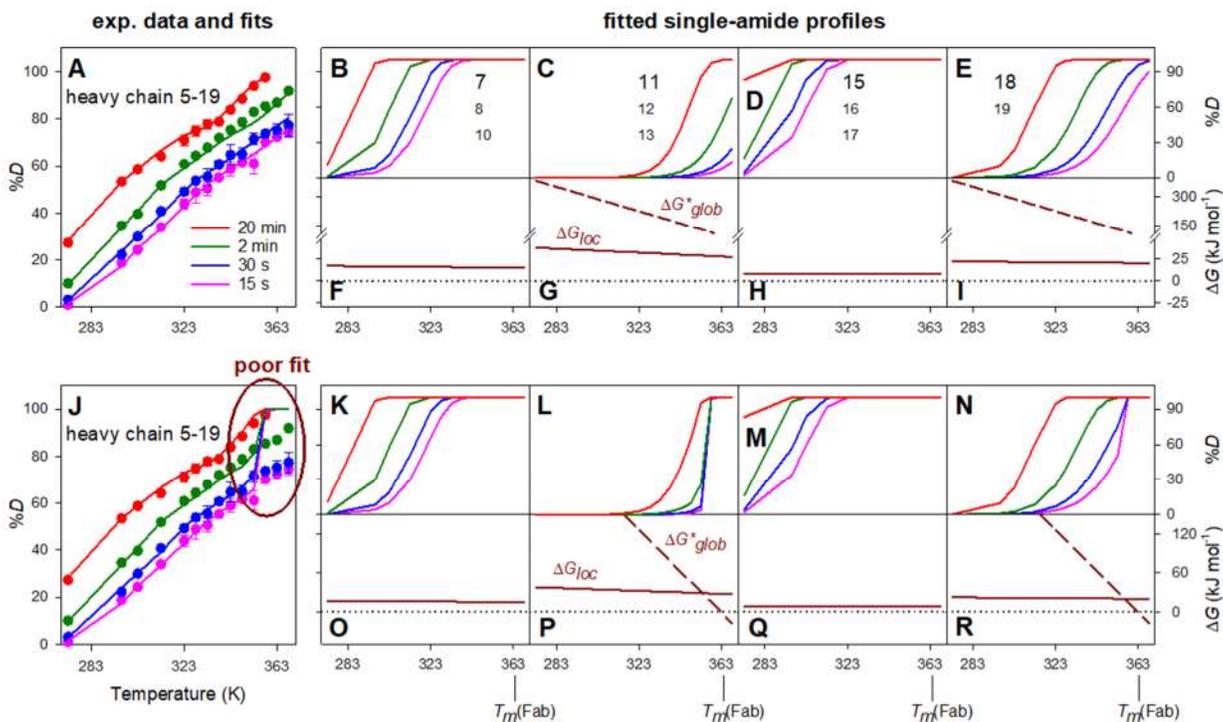
reflects the indeterminate nature of  $\Delta G_{opU}$  in this scenario. (2) Amides 11-13 and 18-19 undergo deuteration around  $T_m(\text{Fab})$ , i.e., range that should be affected by global fluctuations (Figure 4.5C/E). Surprisingly, inclusion of these global fluctuations with  $\Delta G_{opU} = 0$  resulted in fits that were irreconcilable with experiments (Figure 4.5J-R). The situation could be remedied by invoking a  $\Delta G_{opU} \gg 0$  value to ensure that  $\Delta G^*_{glob}$  remained above  $\Delta G_{loc}$  at all temperatures.  $\Delta G_{opU} = 100 \text{ kJ mol}^{-1}$  in Figure 4.5G/I resulted in excellent agreement with the experimental data (Figure 4.5A), but this  $\Delta G_{opU}$  only represents a lower limit because any value greater than  $100 \text{ kJ mol}^{-1}$  would yield an equally good fit. Later, when discussing global fitting results, we will make the case that these very high  $\Delta G_{opU}$  values are related to protein aggregation.

%D profiles of light chain Fab peptide 46-53 displayed a sigmoidal temperature dependence (Figure 4.6A). Profiles with this shape were readily modeled on the basis of local fluctuations (Figure 4.6B-E). Similar to Figure 4.5, global dynamics were not involved in deuteration of peptide 46-53.  $\Delta G_{opU}$  was indeterminate for amides 48-50, while  $\Delta G_{opU} = 100 \text{ kJ mol}^{-1}$  was assigned to 51-53.

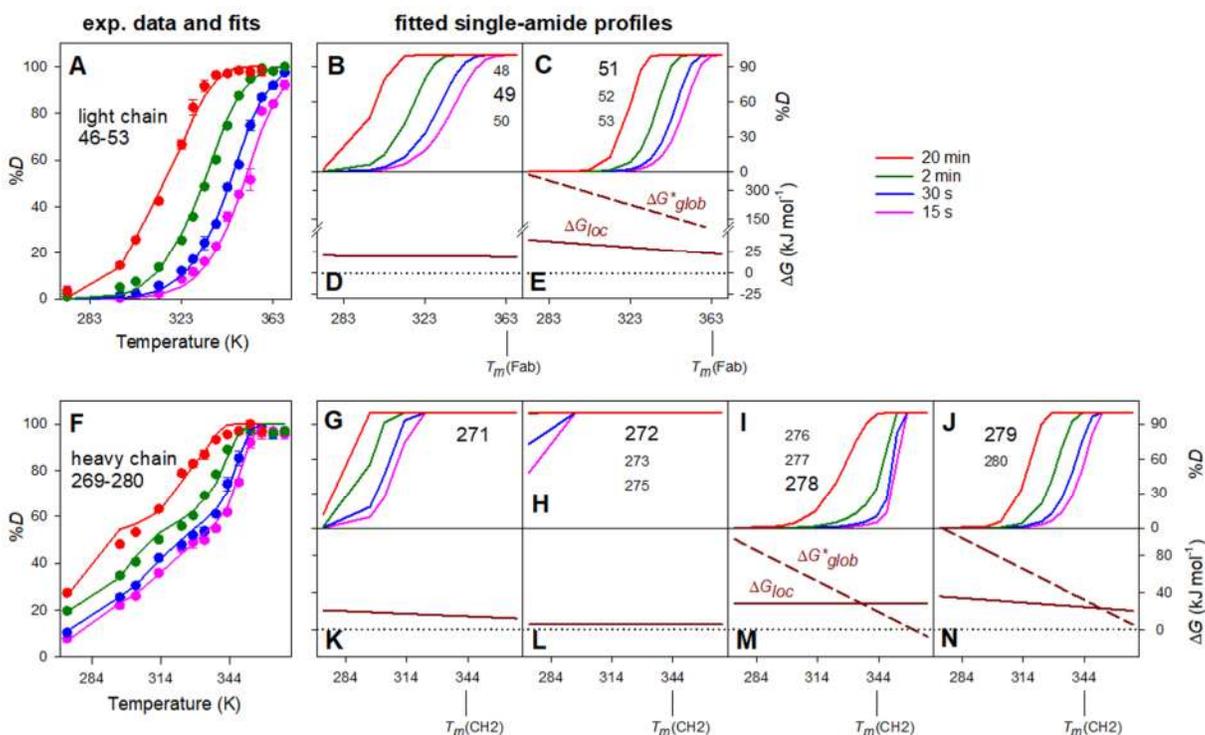
CH2 peptide 269-280 exemplifies a case where global fluctuations did make their presence felt as a steep increase close to  $T_m(\text{CH2})$  in the experimental data (Figure 4.6F). This steep increase is caused by the  $\Delta G^*_{glob}$  profiles that dip below  $\Delta G_{loc}$  close to the melting temperature (Figures 4.6M/N), such that global fluctuations trigger deuteration of amides 276-280. When combined with the local dynamics of sites 271-275 (Figure 4.6K/L), excellent agreement with the experimental data was obtained (Figure 4.6F).

In summary, Figures 4.5 and 4.6 demonstrate that the thermodynamic framework of equations 4.1-4.7 can successfully capture experimental %D profiles that represent vastly different  $T$ -dependent shapes. Most of the deuteration events were attributed to local fluctuations. In addition, one of the peptides illustrated a case of HDX being driven by global dynamics; this scenario is encountered whenever  $\Delta G^*_{glob}$  dips below  $\Delta G_{loc}$  (Figure 4.6M/N). Global dynamics were not involved in the other deuteration processes of Figures 4.5/4. Such a lack of global involvement can arise if amides are already deuterated around  $T_m$  (Figures 4.5B/D, 4.6G/H).  $\Delta G_{opU}$  is indeterminate in such cases, and no numerical

values will be reported. Global dynamics are also irrelevant if they have a large  $\Delta G_{opU}$ , causing  $\Delta G^*_{glob}$  to hover above  $\Delta G_{loc}$  at all temperatures (Figures 4.5G/I, 4E).  $\Delta G_{opU}$  for the latter scenario will be reported as  $100 \text{ kJ mol}^{-1}$ , although this value only represents a lower limit.



**Figure 4.5.** Temperature-dependent HDX kinetics of heavy chain Fab peptide 5-19, analyzed using the thermodynamic model of equation 4.1-4.7. (A) Experimental data (dots) and fits (lines). (B-E) Fitted HDX profiles of backbone sites. Each panel shows data for one NH, identified by a large number. Smaller numbers identify the remaining NH sites in the segment. (F-I) Free energy profiles associated with panels B-E ( $\Delta G_{loc}$  is shown as solid lines,  $\Delta G^*_{glob}$  as dashed lines). (J-R) These bottom panels mirror the top panels, except that  $\Delta G_{opU}$  was changed from  $100 \text{ kJ mol}^{-1}$  to zero in P/R, yielding global dynamics that are irreconcilable with experiments (illustrated in panel J).



**Figure 4.6.** Thermodynamic analysis of temperature-dependent HDX kinetics for (A-E) light chain peptide 46-53 in the Fab region, and (F-N) heavy chain peptide 269-280 in the CH2 region. For additional details, see text and caption of Figure 4.5A-I.

#### 4.3.2.1 Global Fitting of T-Dependent HDX Data

Encouraged by the successful application of our thermodynamic framework to individual peptides (Figures 4.5, 4.6), we aimed to determine  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  values for the entire NISTmAb. Extracting these parameters for each of the 613 non-proline residues would be an insurmountable task. Several safeguards were implemented to narrow down the range of possible solutions. (1) As for the single-peptide analyses discussed previously, we assumed that segments of three (sometimes fewer, see Table 4.1) consecutive residues share the same  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  values. (2) Global fitting is a method that greatly improves the robustness of results<sup>73, 74, 80-82</sup> [note that “global fitting” is unrelated to “global unfolding”]. Here we implemented a global fitting strategy that required  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  parameters to be identical for all segments that were shared across overlapping peptides,

using the least-square optimization strategy of equation 4.8. A total of 80 peptides were included, for a sequence coverage 73%. Figure 4.9 shows these overlapping peptides, as well as segment boundaries.  $\Delta H_{loc}$  was capped at  $100 \text{ kJ mol}^{-1}$ , in line with energetic estimates of local unfolding events.<sup>83</sup> (3) Instead of focusing on  $T$ -dependent HDX data at a single time point, we simultaneously fitted four complete temperature profiles for  $t = 15 \text{ s}$ ,  $30 \text{ s}$ ,  $2 \text{ min}$ , and  $20 \text{ min}$  for all peptides.

Gratifyingly, our global fitting strategy succeeded in finding a set of  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  values that provided excellent agreement between calculated and experimental  $\%D(t,T)$  profiles for all peptides, all temperatures, and all time points (Figure 4.11). The fitted thermodynamic parameters obtained in this way are summarized in Figure 4.10.

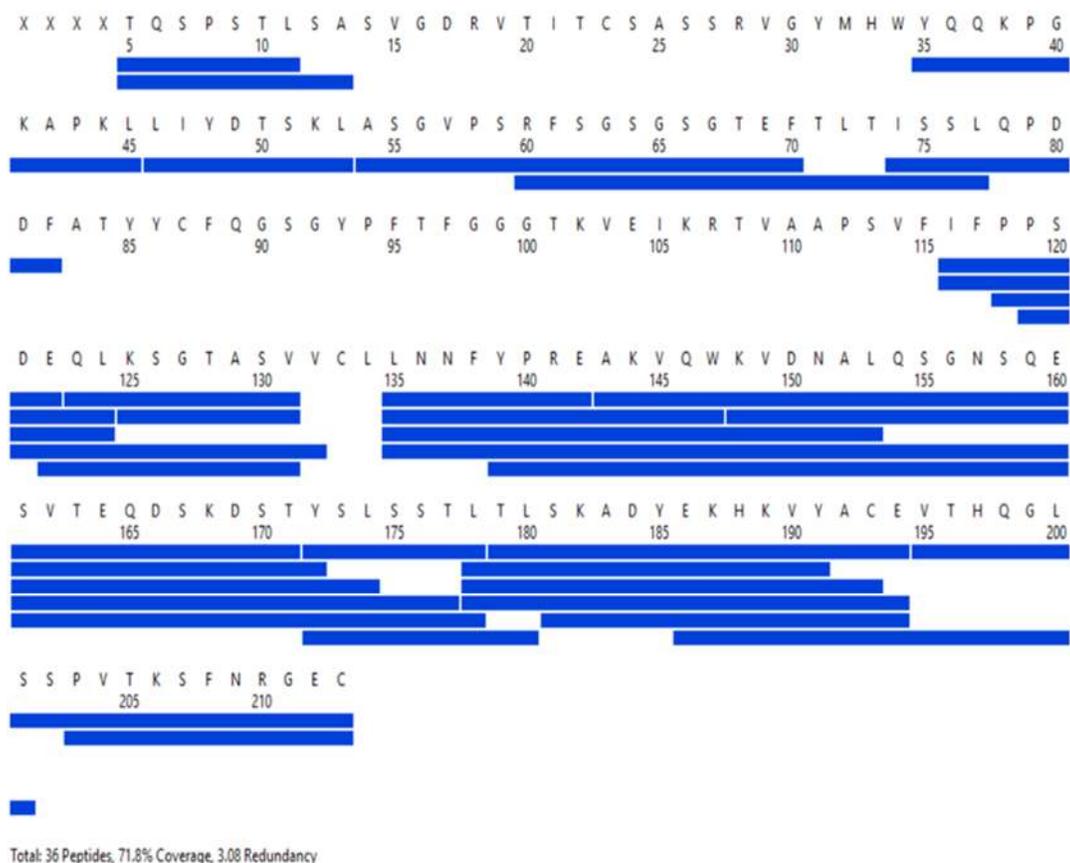
**Table 4.1** (continued following page). Assignment of segments to protein regions and residue range for global fitting results, see Figure 4.3.

<b>LIGHT CHAIN</b>			<b>LIGHT CHAIN (cont'd)</b>		
Segment	Region	Residues	Segment	Region	Residues
1	VL	7-9	36	CL	163-165
2	VL	10-11	37	CL	166-168
3	VL	12,13	38	CL	169-171
4	VL	37-40	39	CL	172-173
5	VL	41-43	40	CL	174-175
6	VL	44-45	41	CL	176-177
7	VL	48-50	42	CL	178
8	VL	51-53	43	CL	180-182
9	VL	56-58	44	CL	183-185
10	VL	59-61	45	CL	186-188
11	VL	62-64	46	CL	189-191
12	VL	65-67	47	CL	192-194
13	VL	68-70	48	CL	195-197
14	VL	71-73	49	CL	198-199
15	VL	74-75	50	CL	200-201
16	VL	76-77	51	CL	202-204
17	VL	78-80	52	CL	205-207
18	VL	81-82	53	CL	208-210
19	CL	120	54	CL	211-213
20	CL	121			
21	CL	122			
22	CL	123-125			
23	CL	126-128			
24	CL	129-131			
25	CL	132			
26	CL	137			
27	CL	138			
28	CL	139			
29	CL	140-142			
30	CL	143-145			
31	CL	146-148			
32	CL	149-151			
33	CL	152-154			
34	CL	155-157			
35	CL	158-160			

<b>HEAVY CHAIN</b>	<b>HEAVY CHAIN (cont'd)</b>	<b>HEAVY CHAIN (cont'd)</b>
--------------------	-----------------------------	-----------------------------

Segment	Region	Residues	Segment	Region	Residues	Segment	Region	Residues
1	VH	7-10	41	CH1	130-134	81	CH3	351-353
2	VH	11_14	42	CH1	135-137	82	CH3	354-356
3	VH	15-17	43	CH1	138-140	83	CH3	357
4	VH	18-19	44	CH1	141-144	84	CH3	358
5	VH	23	45	CH1	145-147	85	CH3	359
6	VH	24-25	46	CH1	148-150	86	CH3	360
7	VH	26-27	47	CH1	151-154	87	CH3	361
8	VH	30	48	CH1	155-158	88	CH3	362-363
9	VH	31	49	CH1	159-160	89	CH3	364-365
10	VH	32	50	CH1	190-191	90	CH3	366-367
11	VH	33-35	51	CH1	192-194	91	CH3	368-369
12	VH	36-38	52	CH1	195-196	92	CH3	370
13	VH	39-41	53	CH1	197-198	93	CH3	371-372
14	VH	42-45	54	CH1	199-200	94	CH3	373-374
15	VH	46-47	55	CH2	247-252	95	CH3	375-376
16	VH	48	56	CH2	253-255	96	CH3	377-379
17	VH	53-55	57	CH2	263-265	97	CH3	380-381
18	VH	56-58	58	CH2	266-268	98	CH3	382-383
19	VH	59-61	59	CH2	269-271	99	CH3	385-386
20	VH	62-64	60	CH2	272-275	100	CH3	387-388
21	VH	65-67	61	CH2	276-278	101	CH3	389-391
22	VH	68-70	62	CH2	279-280	102	CH3	392-393
23	VH	71-73	63	CH2	283-285	103	CH3	394-395
24	VH	74-76	64	CH2	286-288	104	CH3	396-399
25	VH	77-79	65	CH2	289-291	105	CH3	400-401
26	VH	80-82	66	CH2	292-295	106	CH3	416-418
27	VH	85-87	67	CH2	296-296	107	CH3	419-421
28	VH	88-90	68	CH2	312-313	108	CH3	422-424
29	VH	91-92	69	CH2	314-315	109	CH3	425-426
30	VH	93-94	70	CH2	316-317	110	CH3	428-430
31	VH	99-101	71	CH2	318-320	111	CH3	431-433
32	VH	102-104	72	CH2	321-323	112	CH3	434-436
33	VH	105-107	73	CH2	324-326	113	CH3	437-439
34	VH	108-110	74	CH2	327-329	114	CH3	440-442
35	VH	111-113	75	CH2	330-333	115	CH3	443-445
36	VH	114-116	76	CH2	334-337	116	CH3	446-448
37	VH	117-119	77	CH2	338-340	117	CH3	449-450
38	VH	120-122	78	CH2	341-343			
39	CH1	123-125	79	CH2	344-345			
40	CH1	126-129	80	CH3	346-350			

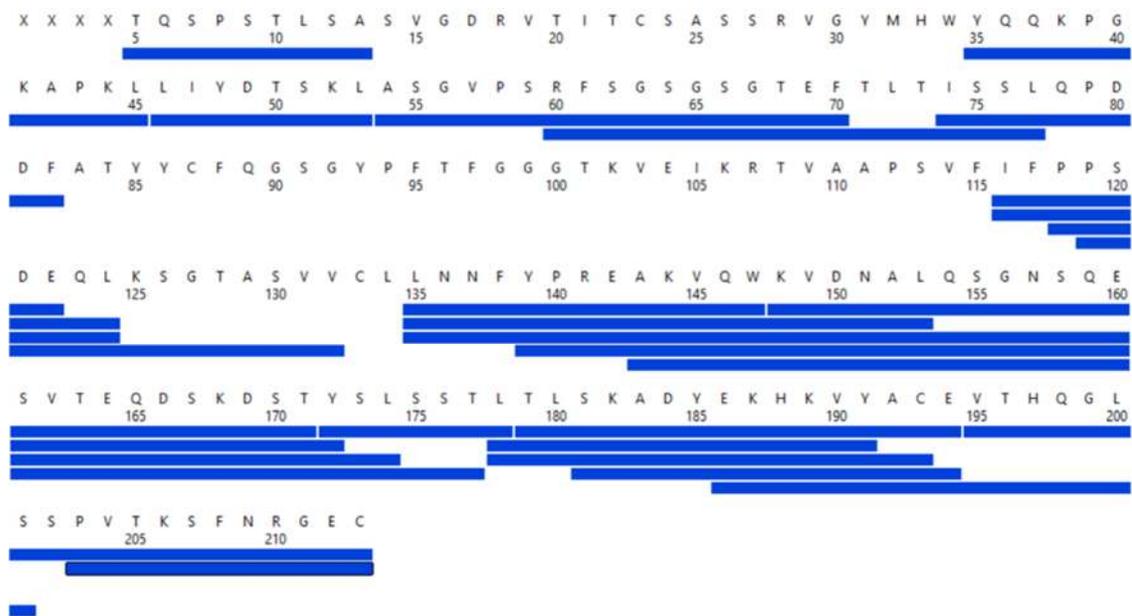


**Figure 4.7 (A).** *NISTmAb light chain peptides detected after peptic digestion.*

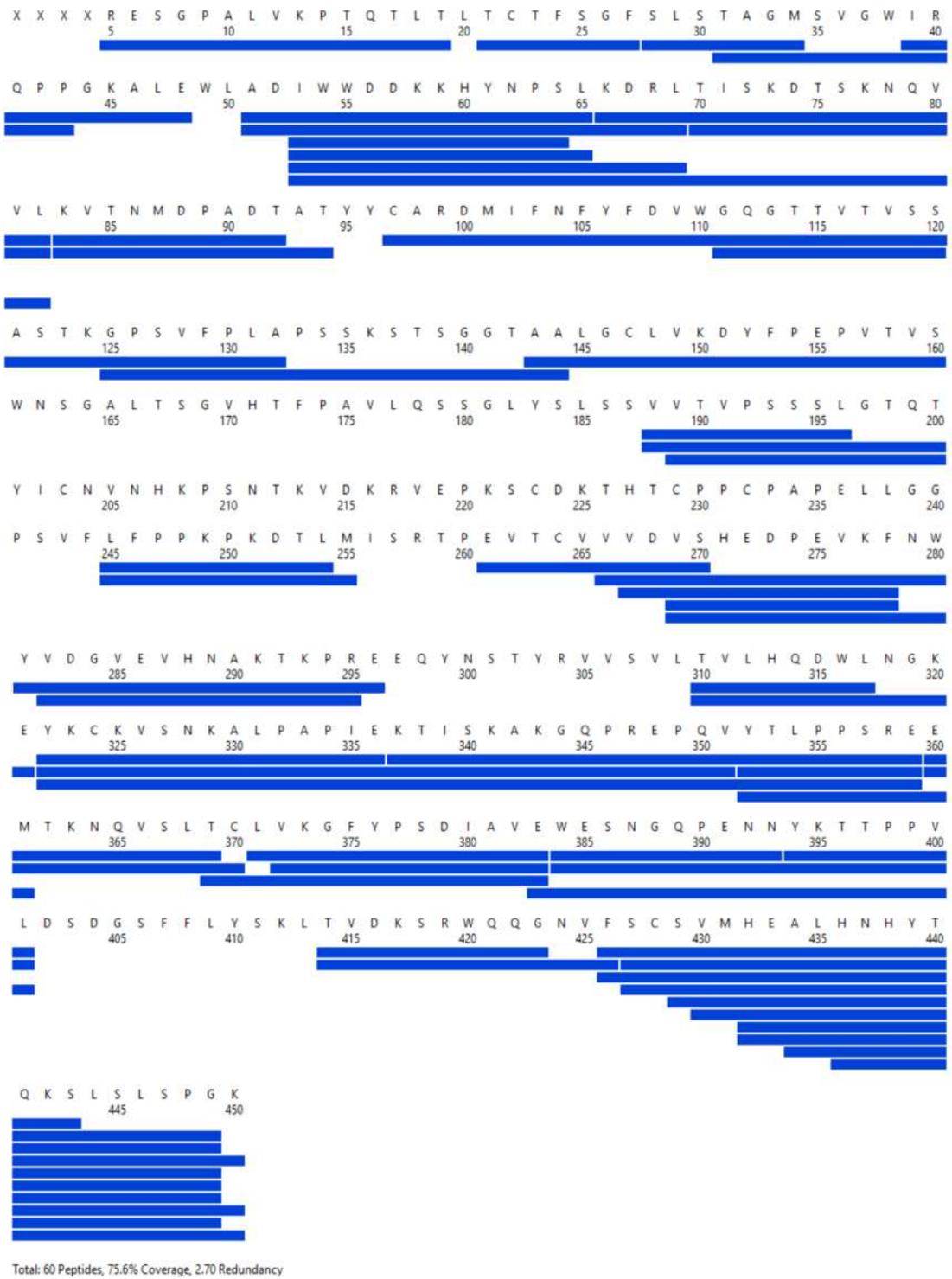


**Figure 4.7. (B)** *NISTmAb heavy chain peptides detected after peptic digestion.*

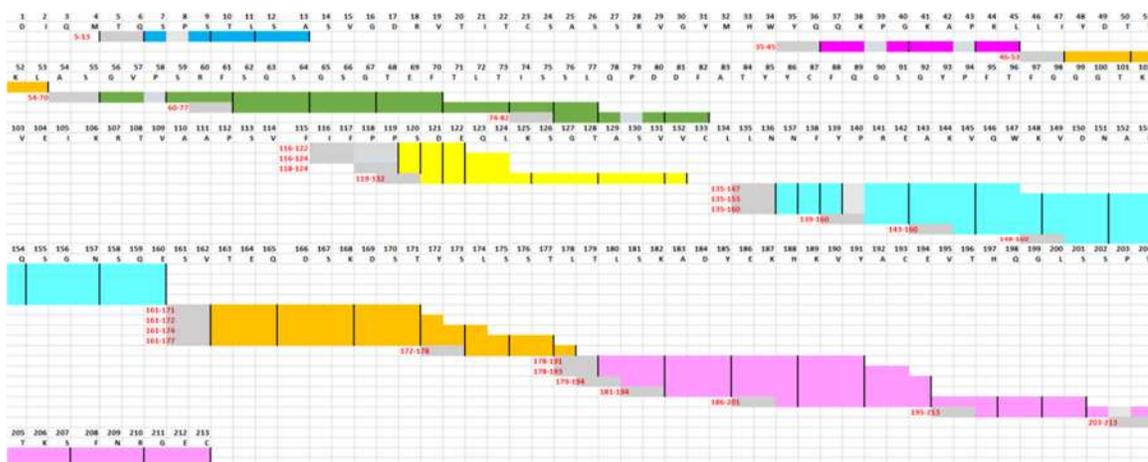
Total: 81 Peptides, 92.2% Coverage, 3.29 Redundancy



**Figure 4.8.** (A) *NISTmAb* light chain peptides used for global analysis.



**Figure 4.8 (B).)** *NISTmAb heavy chain peptides used for global analysis.*



**Figure 4.9.(A)** Light chain peptide map for global fitting. Clusters of overlapping peptides were grouped together as indicated by the different colors. Vertical black lines indicate segment boundaries. All NH sites within a segment and across all overlapping peptides were fitted with the same  $\Delta H_{loc}/\Delta S_{loc}/\Delta G_{opU}$  parameters. The first two residues of each peptide (gray) were not considered due to back exchange. Pro residues are also shown in gray, reflecting their lack of NH sites [continued following page].

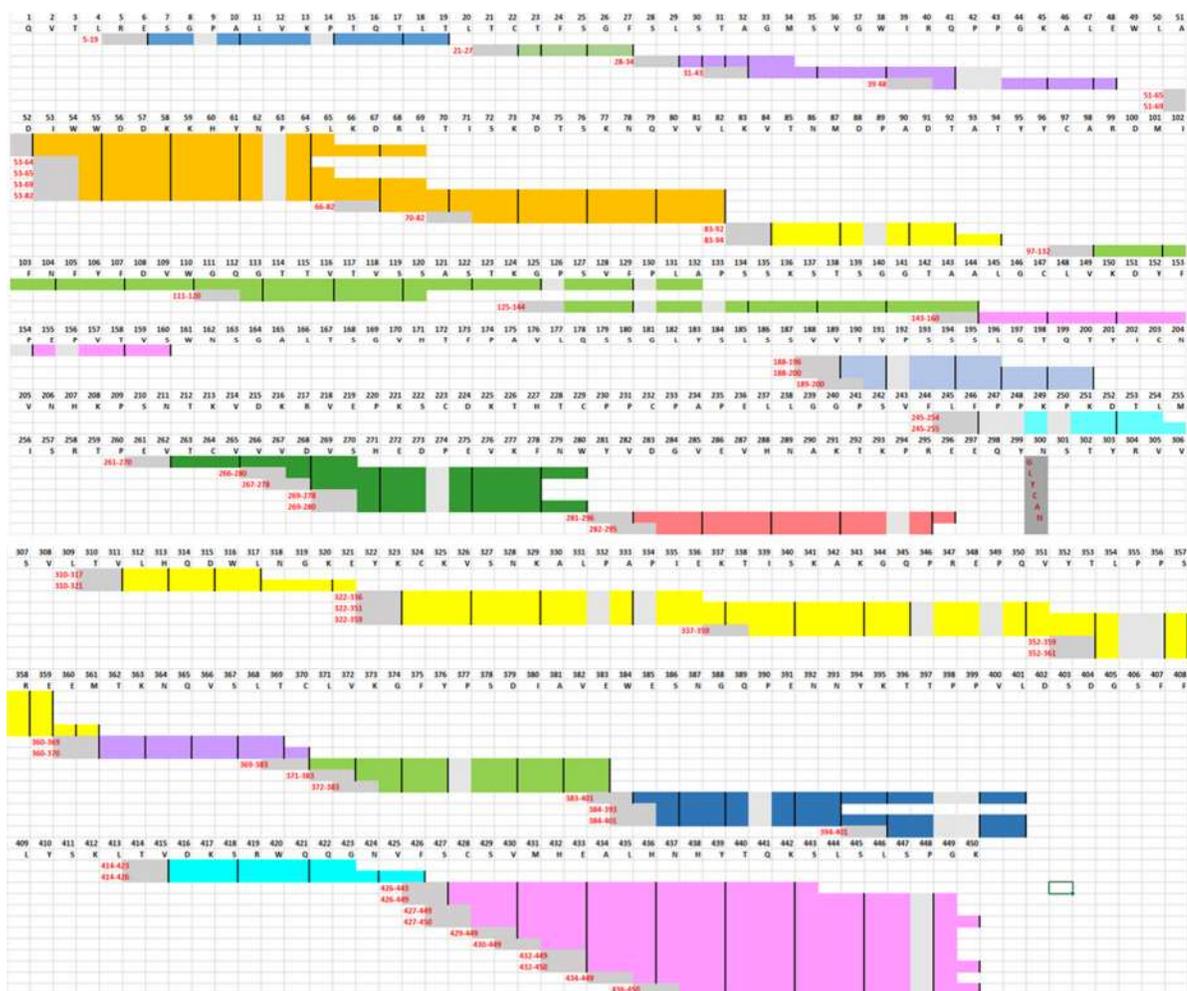


Figure 4.9 (B) Heavy chain peptide map for global fitting.

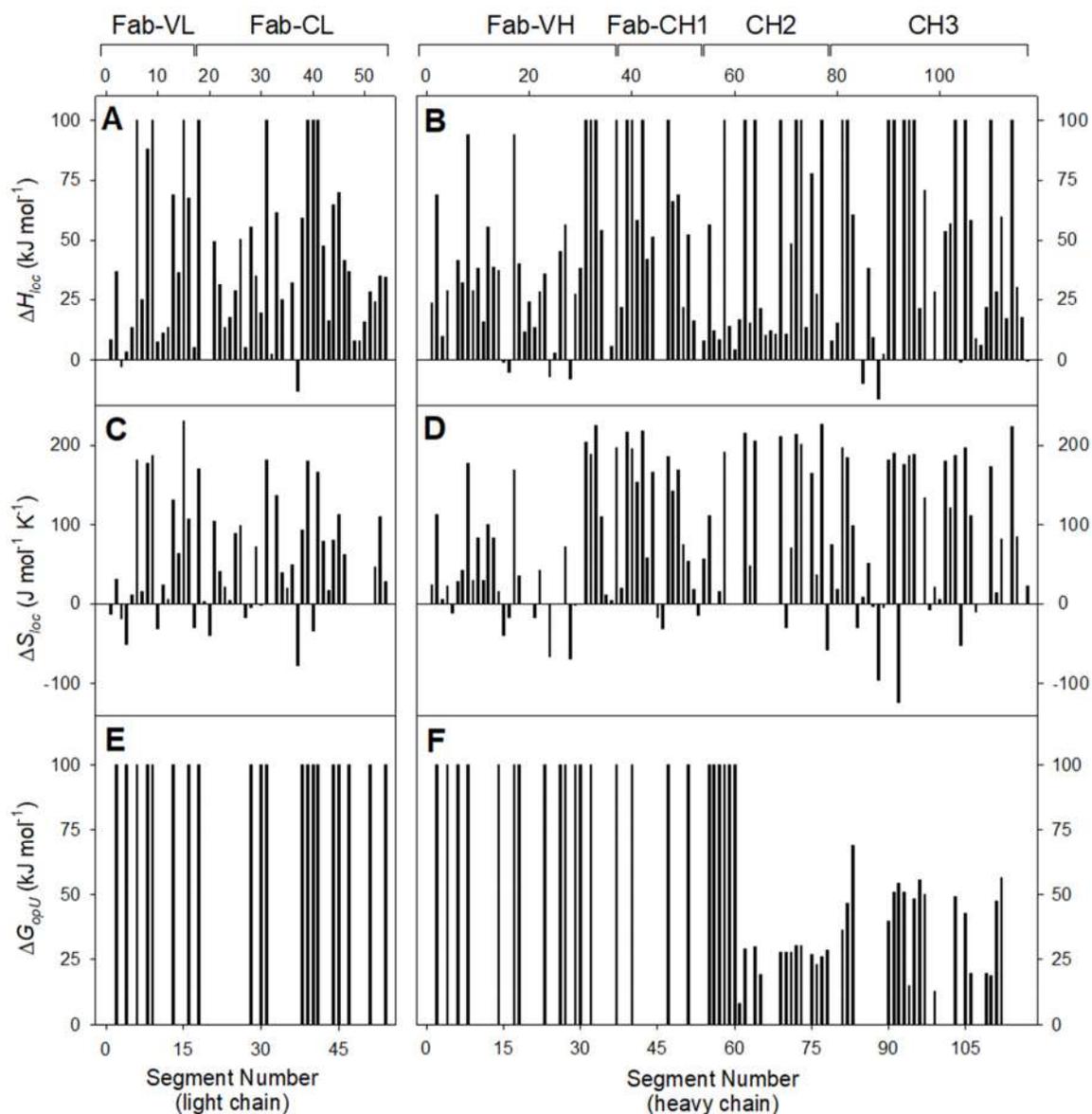
#### 4.3.2.2 Local Dynamics: $\Delta H_{loc}$ and $\Delta S_{loc}$ .

Almost all local fluctuations of NISTmAb were characterized by  $\Delta H_{loc} > 0$  and  $\Delta S_{loc} > 0$  (Figure 4.10A-D), causing  $\Delta G_{loc}$  to decrease with increasing temperature (equation 4.4). In other words, the free energy penalty associated with local unfolding became less severe at high  $T$ . The endothermic ( $\Delta H_{loc} > 0$ ) nature of these local fluctuations reflects the fact that energy input is required to disrupt intramolecular contacts that are associated with local  $NH_{closed} \rightarrow NH_{open}$  events. These contacts include  $NH \cdots OC$  hydrogen bonds,<sup>83</sup> but also

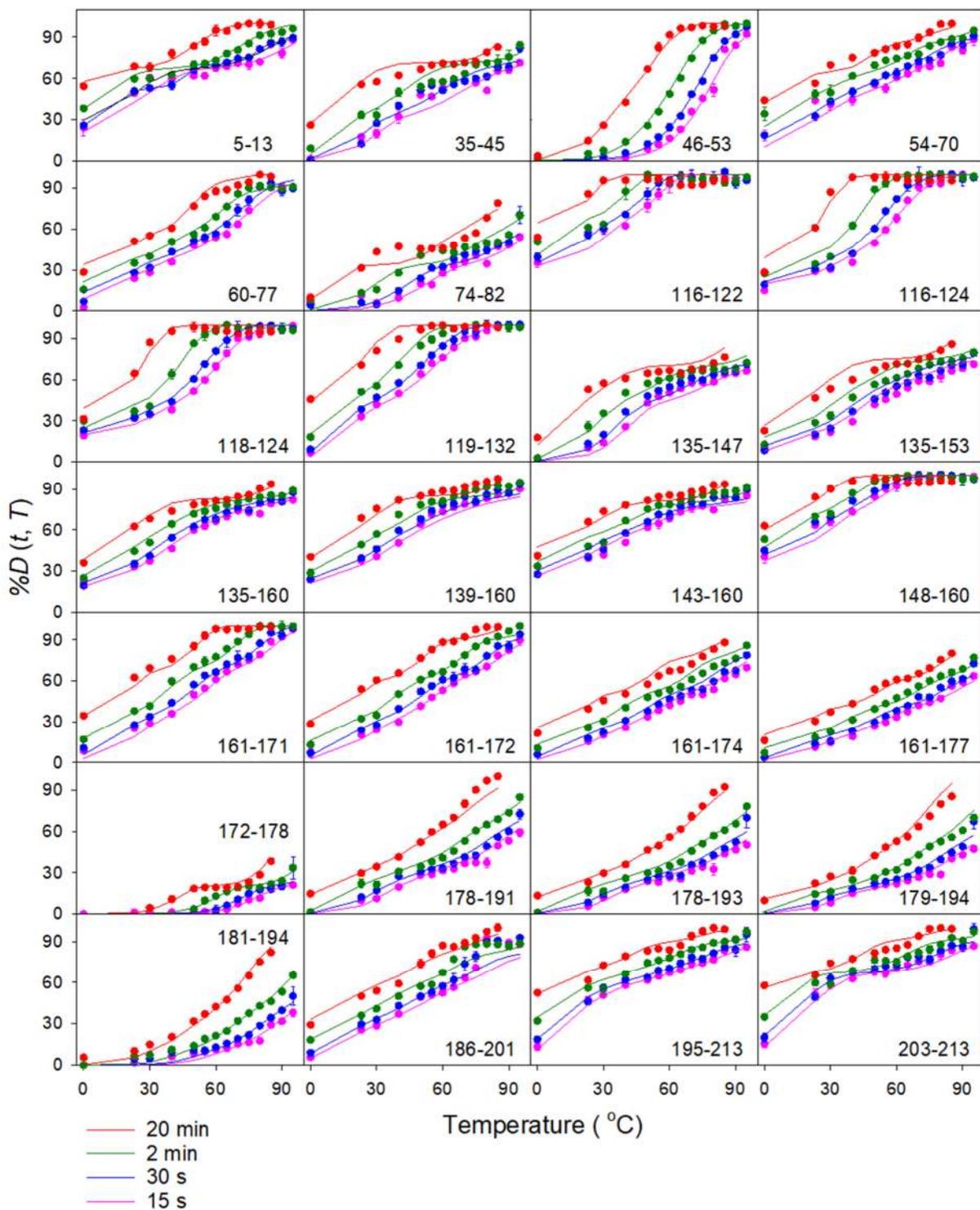
hydrophobic and van der Waals interactions, as well as salt bridges.<sup>84</sup> The entropy increase ( $\Delta S_{loc} > 0$ ) of the  $NH_{closed} \rightarrow NH_{open}$  transitions reflects the increased conformational freedom of the locally unfolded segments.<sup>84</sup>

The prevalence of  $\Delta H_{loc} > 0$  and  $\Delta S_{loc} > 0$  for local fluctuations of NISTmAb is somewhat different from myoglobin, where a multitude of segments showed “noncanonical” behavior with  $\Delta H_{loc} < 0$  and  $\Delta S_{loc} < 0$  (showed already in chapter 2). Local unfolding is enthalpically driven in such noncanonical cases, likely by the tight hydration of the unfolded segments.<sup>85</sup>

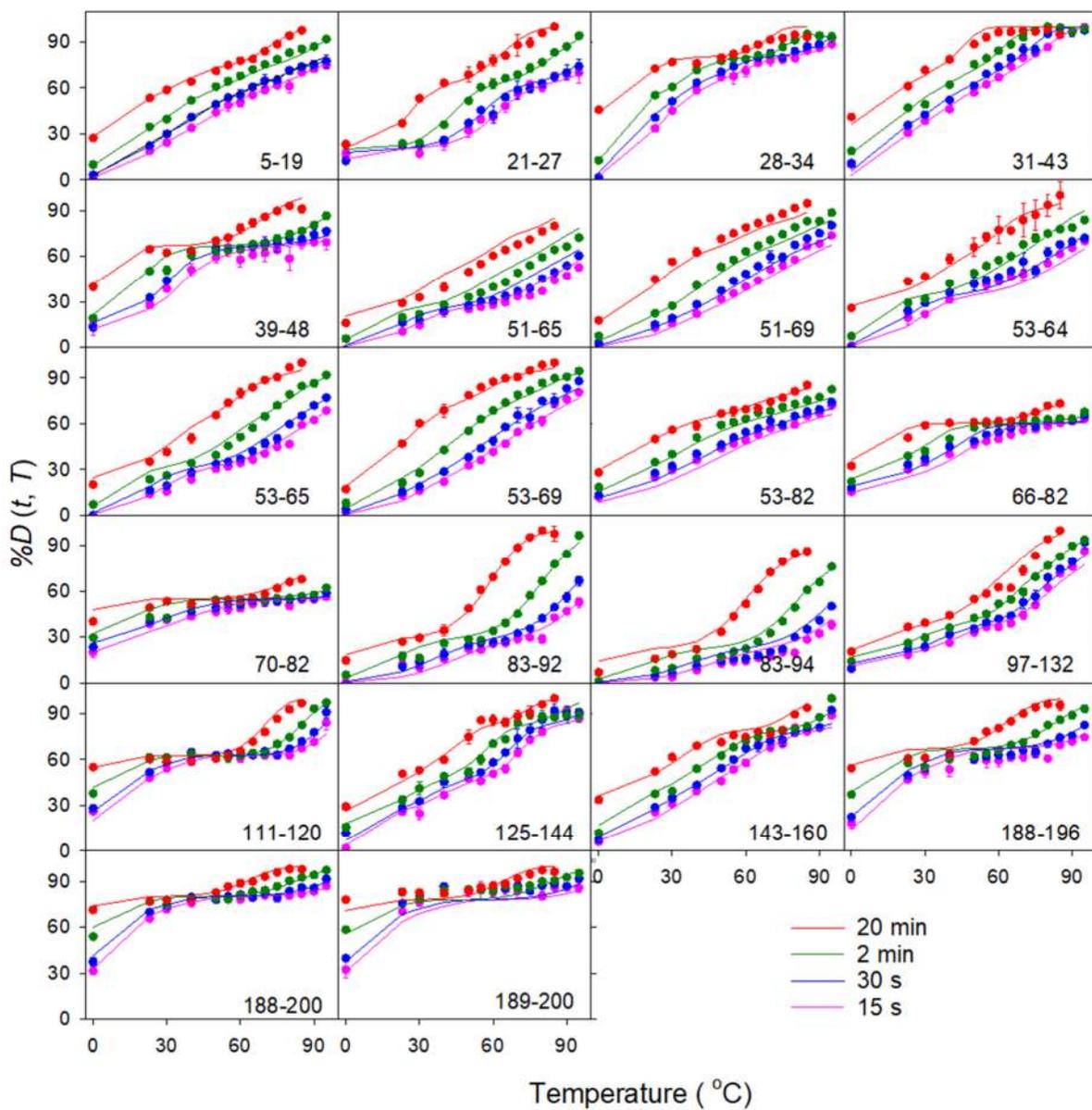
<sup>86</sup> In NISTmAb only very few segments displayed noncanonical behavior, seen from the sparsity of negative data points in Figure 4.10A-D.



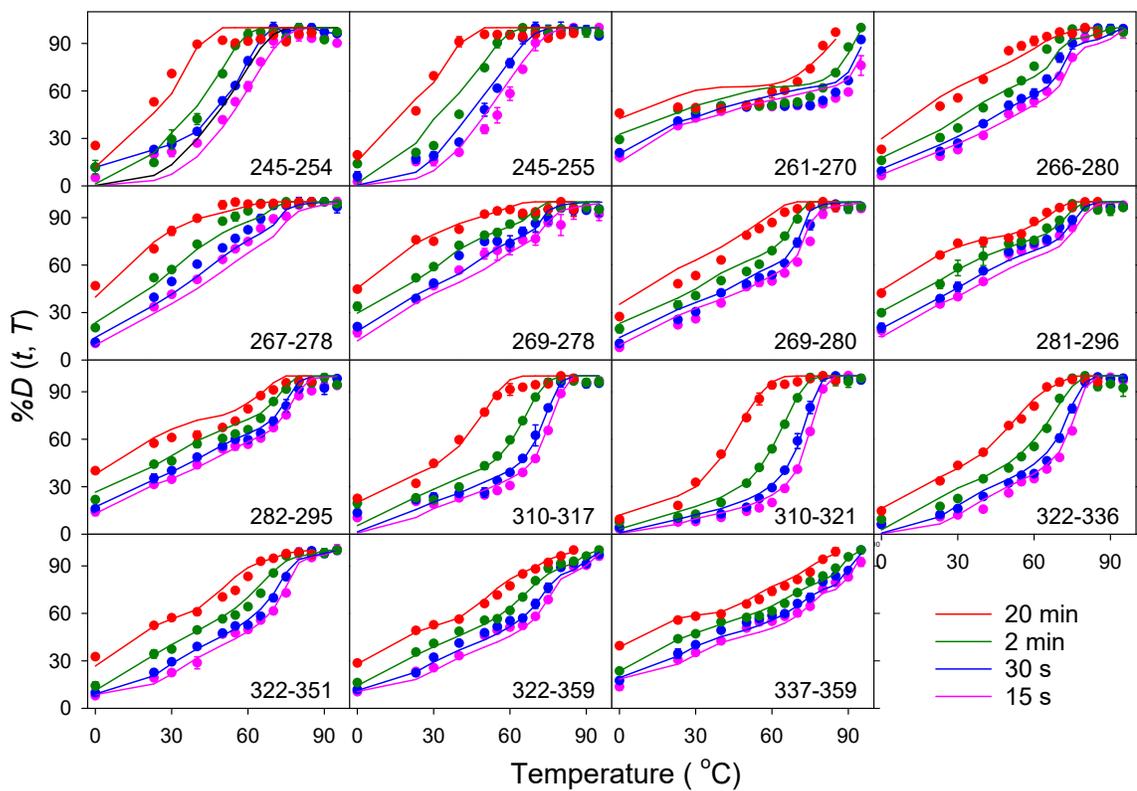
**Figure 4.10.** Thermodynamic parameters obtained by global fitting of all experimental %D profiles at all temperatures and all time points (see Figure 4.11 for fitted curves). (A, B)  $\Delta H_{loc}$ , (C, D)  $\Delta S_{loc}$ , (E, F)  $\Delta G_{globU}$ . Missing data in panels E/F correspond to indeterminate  $\Delta G_{globU}$  values. See Table 4.1 for the assignment of segment numbers to residue numbers.



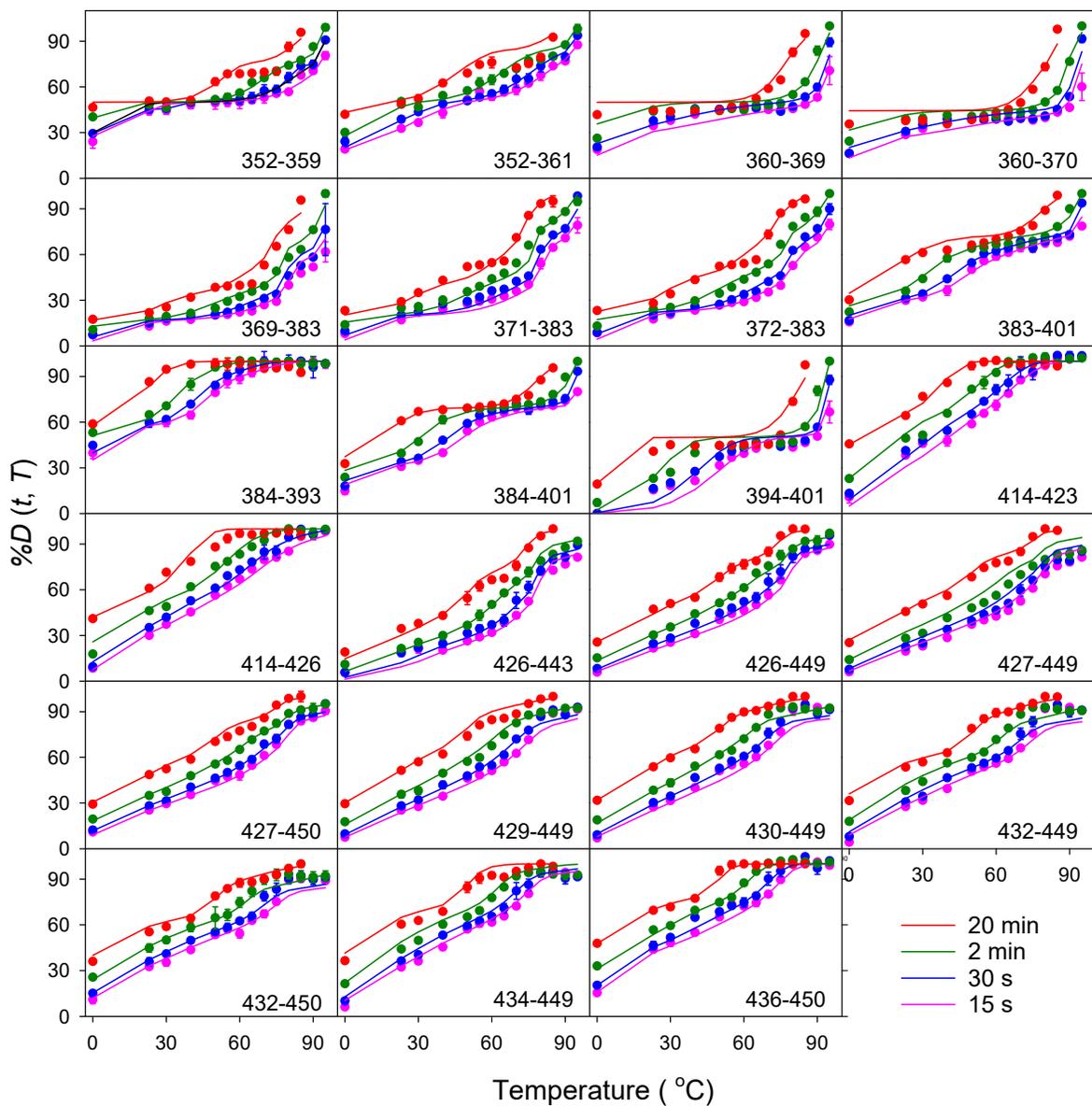
**Figure 4.11. (A)** Experimental  $\%D(t, T)$  data (points) and global fits (lines) for light chain Fab peptides. Error bars are smaller than the data points in most cases.



**Figure 4.11 (B)** Experimental  $\%D(t, T)$  data (points) and global fits (lines) for heavy chain Fab peptides. Error bars are smaller than the data points in most cases.



**Figure 4.11. (C)** Experimental  $\%D(t, T)$  data (points) and global fits (lines) for heavy chain CH2 peptides. Error bars are smaller than the data points in most cases.



**Figure 4.11 (D)** Experimental  $\%D(t, T)$  data (points) and global fits (lines) for heavy chain  $CH_3$  peptides. Error bars are smaller than the data points in most cases.

### 4.3.2.3 Global Dynamics of CH2 and CH3 Moieties

Global N  $\leftrightarrow$  U fluctuations between the native and the unfolded state become prevalent in the vicinity of  $T_m$ .<sup>19</sup> For small single-domain proteins these global fluctuations simultaneously affect all residues. However, for large proteins with distinct domains (such as IgG1s) the situation is different. In this work, “global” does not refer to the entire NISTmAb, but it separately applies to the CH2, CH3, and Fab moieties. Approximating these moieties as individual global units is justified because each of them undergoes unfolding with a characteristic  $T_m$ , quite independently of the other protein regions.<sup>7, 14, 22, 75</sup>

We will first discuss global dynamics of CH2 and CH3. It might be expected that N  $\leftrightarrow$  U fluctuations close to the corresponding  $T_m$  values are equivalent to the NH<sub>closed</sub>  $\leftrightarrow$  NH<sub>open</sub> dynamics that mediate HDX (equation 4.1). In such a case, backbone amides in the thermally unfolded moieties would be completely unprotected, akin to NH sites in dipeptide model compounds.<sup>46</sup> However, it is known that many thermally unfolded proteins possess residual protection, suggesting that globally-mediated HDX requires two steps (N  $\leftrightarrow$  U<sub>closed</sub>  $\leftrightarrow$  U<sub>open</sub>), where the second step produces NH<sub>open</sub> sites that are required for deuteration. Within our thermodynamic framework, such two-step transitions are captured via equation 4.5, where  $\Delta G_{glob}(T)$  refers to the N  $\leftrightarrow$  U<sub>closed</sub> transitions that can be probed by DSC, while  $\Delta G_{opU}$  refers to U<sub>closed</sub>  $\leftrightarrow$  U<sub>open</sub> (showed in Chapter 2). Figure 4.10F shows that many CH2 and CH3 segments had  $\Delta G_{opU}$  values around 20 – 50 kJ mol<sup>-1</sup>, revealing that HDX of the corresponding segments indeed proceeds via N  $\leftrightarrow$  U<sub>closed</sub>  $\leftrightarrow$  U<sub>open</sub>. In free energy plots such as Figure 4.6M/N, these  $\Delta G_{opU}$  contributions shift  $\Delta G^*_{glob}$  upward, rendering the formation of NH<sub>open</sub> sites less favorable than the initial N  $\leftrightarrow$  U<sub>closed</sub> events. To summarize, CH2 and CH3 global fluctuations do not directly involve fully unprotected random coil structures. Instead, the significant  $\Delta G_{opU}$  contributions in Figure 4.10F highlight the presence of non-native intrachain contacts in the unfolded state. These non-native contacts have to transiently open up for HDX to proceed.

#### 4.3.2.4 Fab Global Dynamics and Protein Aggregation

For the single Fab peptide examined in Figure 4.5, the inclusion of global fluctuations produced irreconcilable discrepancies with the experimental data (Figure 4.5J). Agreement between fits and experiments could only be achieved by invoking a very large  $\Delta G_{opU}$  value, which we arbitrarily assigned as  $100 \text{ kJ mol}^{-1}$ . The fitting results of Figure 4.10E/F reveal that this  $\Delta G_{opU} = 100 \text{ kJ mol}^{-1}$  scenario applied to the entire Fab moiety for both heavy and light chains, implying that global dynamics were not involved in Fab deuteration. This Fab behavior is in striking contrast to CH2 and CH3, where global fluctuations were required for fitting the experimental data.

How is it possible that global fluctuations of Fab do not contribute to deuteration? We attribute this peculiar effect to irreversible NISTmAb aggregation, which takes place close to  $T_m(\text{Fab})$  as reported in previous studies,<sup>7, 17, 20, 21, 43</sup> and as seen in the DSC data of Figure 4.4B. This aggregation was also apparent by visual inspection of heated samples (Figure 4.12). The Linderstrom-Lang HDX model relies on  $\text{NH}_{\text{closed}} \leftrightarrow \text{NH}_{\text{open}}$  fluctuations that are fully reversible (equation 4.1).<sup>29</sup> The aforementioned  $\text{N} \leftrightarrow \text{U}_{\text{closed}} \leftrightarrow \text{U}_{\text{open}}$  dynamics of the CH2 and CH3 moieties fall under this umbrella. However, the Fab behavior is different. The fact that irreversible aggregation starts to take place around  $T_m(\text{Fab})$  implies that global Fab dynamics cannot be described as  $\text{N} \leftrightarrow \text{U}_{\text{closed}} \leftrightarrow \text{U}_{\text{open}}$ , but that they produce aggregates via either  $\text{N} \leftrightarrow \text{U}_{\text{closed}} \rightarrow \text{Aggregated}$ , or  $\text{N} \leftrightarrow \text{U}_{\text{closed}} \leftrightarrow \text{U}_{\text{open}} \rightarrow \text{Aggregated}$ . Conventional HDX-MS analyses strategies (equation 4.1) are not applicable under such irreversible scenarios. Our HDX-MS data are consistent with the view that aggregated NISTmAb is resistant to peptic digestion, akin to other types of aggregates.<sup>87</sup> Any protease-resistant species is unobservable in HDX-MS, and it will therefore not contribute to the experimentally observed %D values. Instead, the experimental data only report on the subpopulation of proteins that remain soluble. For this reason, the very large  $\Delta G_{opU}$  of  $100 \text{ kJ mol}^{-1}$  throughout the Fab moiety (Figure 4.10) should not be interpreted in the context of  $\text{U}_{\text{closed}} \leftrightarrow \text{U}_{\text{open}}$  transitions (Chapter 2). Under the conditions encountered here, these very large  $\Delta G_{opU}$  values represent the hallmark of irreversible aggregation.

Finally, we note that gaps in the bar diagrams of Figure 4.10E/F should *not* be interpreted as  $\Delta G_{opU} = 0$ . Instead, they represent segments where  $\Delta G_{opU}$  was indeterminable, i.e., where any value  $\Delta G_{opU} > 0$  is compatible with the experimental data. The origin of this effect has been discussed above (Figures 4.5 and 4.6). This nature of these indeterminate values is very different from the  $\Delta G_{opU} = 100 \text{ kJ mol}^{-1}$  data points. In the former case, any value is consistent with the experiments; in the latter case, only a very large value provides a good fit with the measured %D values.



**Figure 4.12.** *Photograph of  $6.6 \mu\text{M}$  NISTmAb in an Eppendorf vial after 10 min at  $95 \text{ }^\circ\text{C}$ , pH 6.32. The presence of aggregated protein (white amorphous precipitate) is readily apparent.*

## 4.4 Conclusions

Traditional HDX-MS analyses involve fitting of time-dependent data to a sum of exponentials, providing amplitudes and apparent rate constants for individual peptides.<sup>72</sup> Although such analyses can uncover interesting trends, the fits are largely descriptive instead of directly uncovering intrinsic protein features. The analysis strategy pursued in

this work goes one step further, i.e., it uncovers thermodynamic parameters that govern the structural dynamics of different protein regions for different time points, and across temperatures ranging from near freezing to almost boiling. We originally devised the framework used here for the small single-domain protein myoglobin (Chapter 2). It was not obvious if the same strategy would be applicable to NISTmAb, a complicated multi-domain system where dynamics, unfolding, and aggregation are closely intertwined. Fortunately, global analysis of  $%D(t,T)$  produced excellent fits for all NISTmAb peptides, verifying the suitability of our strategy for complex protein therapeutics.

The application of our thermodynamic model to NISTmAb requires a disclaimer regarding the role of  $\Delta G_{opU}$ . Originally, this parameter was designed to capture residual protection of a globally unfolded protein (Chapter 2), as seen in the current work for CH2 and CH3. However, the entire Fab region had to be fitted with an extremely large  $\Delta G_{opU}$  value of 100 kJ mol<sup>-1</sup> not because of residual protection, but to account for aggregation.

Our data reveal that increasing temperature causes enhanced deuteration. The biophysical foundation of this effect is surprisingly complicated, as it reflects the combination of at least four factors. (1) To a large extent, the  $T$ -dependence of  $%D$  arises from the near-exponential increase of  $k_{ch}(T)$ , illustrated in Figure 4.3B. (2)  $K_{loc}$  and  $K^*_{glob}$  in equation 4.5 govern the population of NH<sub>open</sub> sites which are required for HDX. These equilibrium constants can be expressed via their respective  $\Delta G$  values (equation 4.3). Even if these  $\Delta G$  values were constant, increasing temperature would boost HDX because the population of Boltzmann-excited NH<sub>open</sub> sites is proportional to  $\exp(\Delta G/RT)$ . (3) At relatively low  $T$ , where protein dynamics are governed by local fluctuations, the population of NH<sub>open</sub> sites is further enhanced by the fact that  $\Delta G_{loc}$  becomes less positive with increasing  $T$  for most segments, i.e., those with  $\Delta H_{loc} > 0$  and  $\Delta S_{loc} > 0$  (Figure 4.10A-D). (4)  $\Delta G^*_{glob}$  becomes less positive with increasing  $T$ , with a slope that is steeper than that of  $\Delta G_{loc}$ . This steep slope causes  $\Delta G^*_{glob}$  to drop below  $\Delta G_{loc}$  in the vicinity of  $T_m$ , such that global fluctuations become the dominant contributor to deuteration (exemplified in Figure 4.6M/N). The analysis strategy pursued in this work allows to clearly distinguish between the four factors. Factors #1 and #2 are rather trivial, whereas factors #3 and #4 provide a direct window into the inner workings of the protein.

Overall, we hope that this work will encourage practitioners to consider time and temperature as equivalent variables in HDX-MS experiments, specifically for investigations of mAbs and other protein therapeutics. Complete deuteration of highly protected NH sites can take days at ambient temperature,<sup>45</sup> whereas complete deuteration can be achieved within ~20 minutes upon heating the samples. With suitable automation, it should be possible to streamline the temperature-dependent HDX workflow employed here, such that the dynamic features across the entire protein become experimentally accessible in very short time windows. The thermodynamic model used here represents a quantitative analysis tool for deciphering the various factors that govern the temperature- and time-dependence of protein deuteration. Spatially-resolved data obtained by *T*-dependent HDX-MS provide a more detailed picture than DSC or related techniques that only report on global unfolding.

## 4.5 References

- 1 Leader, B., Q. J. Baca, D. E. Golan, *Nature Reviews Drug Discovery* **2008**, *7*, 21.
- 2 Lagassé, H. A. D., A. Alexaki, V. L. Simhadri, N. H. Katagiri, W. Jankowski, *F1000 Faculty Rev.* **2017**, *113*, 1.
- 3 Kaltashov, I. A., C. E. Bobst, J. Pawlowski, G. B. Wang, *J. Pharm. Biomed. Anal.* **2020**, *184*,
- 4 Castelli, M. S., P. McGonigle, P. J. Hornby, *Pharmacol. Res. Perspect.* **2019**, *7*, 11.
- 5 Lambert, J. M., A. Berkenblit, *Annual review of medicine* **2018**, *69*, 191.
- 6 Vidarsson, G., G. Dekkers, T. Rispens, *Frontiers in Immunology* **2014**, *5*,
- 7 Ionescu, R. M., J. Vlasak, C. Price, M. Kirchmeier, *J. Pharm. Sci.* **2008**, *97*, 1414.
- 8 Chen, W., L. Kong, S. Connelly, J. M. Dendle, Y. Liu, I. A. Wilson, E. T. Powers, J. W. Kelly, *ACS Chem. Biol.* **2016**, *11*, 1852.
- 9 Liu, H. C., K. May, *Mabs* **2012**, *4*, 17.
- 10 Bye, J. W., L. Platts, R. J. Falconer, *Biotechnol. Lett.* **2014**, *36*, 869.
- 11 Goldberg, D. S., R. A. Lewus, R. Esfandiary, D. C. Farkas, N. Mody, K. J. Day, P. Mallik, M. B. Tracka, S. K. Sealey, H. S. Samra, *J. Pharm. Sci.* **2017**, *106*, 1971.
- 12 Toth, R. T., S. E. Pace, B. J. Mills, S. B. Joshi, R. Esfandiary, C. R. Middaugh, D. D. Weis, D. B. Volkin, *J. Pharm. Sci.* **2018**, *107*, 1009.
- 13 Capelle, M. A. H., R. Gurny, T. Arvinte, *Eur. J. Pharm. Biopharm.* **2007**, *65*, 131.
- 14 Menzen, T., W. Friess, *J. Pharm. Sci.* **2013**, *102*, 415.
- 15 Arthur, K. K., N. Dinh, J. P. Gabrielson, *J. Pharm. Sci.* **2015**, *104*, 1548.
- 16 Garidel, P., A. Eiperle, M. Blech, J. Seelig, *Biophys. J.* **2020**, *118*, 1067.
- 17 Wang, G. B., P. V. Bondarenko, I. A. Kaltashov, *Analyst* **2018**, *143*, 670.
- 18 Zhou, Y., C. K. Hall, M. Karplus, *Protein Sci.* **1999**, *8*, 1064.
- 19 Privalov, P. L., N. N. Khechinashvili, *J. Mol. Biol.* **1974**, *86*, 665.
- 20 Vermeer, A. W. P., W. Norde, *Biophys. J.* **2000**, *78*, 394.

- 21 Akazawa-Ogawa, Y., H. Nagai, Y. Hagihara, *Biophys. Rev.* **2018**, *10*, 255.
- 22 Garber, E., S. J. Demarest, *Biochem. Biophys. Res. Commun.* **2007**, *355*, 751.
- 23 Lang, B. E., K. D. Cole, *Biotechnol. Prog.* **2017**, *33*, 677.
- 24 Fukuda, J., T. Iwura, S. Yanagihara, K. Kano, *Anal. Sci.* **2015**, *31*, 1233.
- 25 Edgeworth, M. J., J. J. Phillips, D. C. Lowe, A. D. Kippen, D. R. Higazi, J. H. Scrivens, *Angew. Chem. Int. Ed.* **2015**, *54*, 15156.
- 26 Niesen, F. H., H. Berglund, M. Vedadi, *Nat. Protocols* **2007**, *2*, 2212.
- 27 Swint, L., A. D. Robertson, *Protein Sci.* **1993**, *2*, 2037.
- 28 Frauenfelder, H., G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, R. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129.
- 29 Bai, Y., T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **1995**, *269*, 192.
- 30 Hanoian, P., C. T. Liu, S. Hammes-Schiffer, S. Benkovic, *Accounts Chem. Res.* **2015**, *48*, 482.
- 31 Murcia Rios, A., S. Vahidi, S. D. Dunn, L. Konermann, *J. Am. Chem. Soc.* **2018**, *140*, 14860.
- 32 Xiao, Y., G. S. Shaw, L. Konermann, *J. Am. Chem. Soc.* **2017**, *139*, 11460.
- 33 Balchin, D., M. Hayer-Hartl, F. U. Hartl, *Science* **2016**, *353*, 42.
- 34 Rob, T., P. Liuni, P. K. Gill, S. L. Zhu, N. Balachandran, P. J. Berti, D. J. Wilson, *Anal. Chem.* **2012**, *84*, 3771.
- 35 Masson, G. R., J. E. Burke, N. G. Ahn, G. S. Anand, C. H. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. H. Faber, R. A. Garlish, P. R. Griffin, M. L. Gross, M. Guttman, Y. Hamuro, A. J. R. Heck, D. Houde, R. E. Jacob, T. J. D. Jorgensen, I. A. Kaltashov, J. P. Klinman, L. Konermann, P. Man, L. Mayne, B. D. Pascal, D. Reichmann, M. Skehel, J. Snijder, T. S. Strutzenberg, E. S. Underbakke, C. Wagner, T. E. Wales, B. T. Walters, D. D. Weis, D. J. Wilson, P. L. Wintrode, Z. Zhang, J. Zheng, D. C. Schriemer, K. D. Rand, *Nat. Methods* **2019**, *16*, 595.
- 36 James, E. I., T. A. Murphree, C. Vorauer, J. R. Engen, M. Guttman, *Chem. Rev.* **2022**, *122*, 7562.
- 37 Hageman, T. S., M. S. Wrigley, D. D. Weis, *Anal. Chem.* **2021**, *93*, 6980.
- 38 Pirrone, G. F., H. Wang, N. Canfield, A. S. Chin, T. A. Rhodes, A. A. Makarov, *Analytical Chemistry* **2017**, *89*, 8351.

- 39 Hudgens, J. W., E. S. Gallagher, I. Karageorgos, K. W. Anderson, J. J. Filliben, R. Y. C. Huang, G. D. Chen, G. M. Bou-Assaf, A. Espada, M. J. Chalmers, E. Harguindey, H. M. Zhang, B. T. Walters, J. Zhang, J. Venable, C. Steckler, I. Park, A. Brock, X. J. Lu, R. Pandey, A. Chandramohan, G. S. Anand, S. N. Nirudodhi, J. B. Sperry, J. C. Rouse, J. A. Carroll, K. D. Rand, U. Leurs, D. D. Weis, M. A. Al-Naqshabandi, T. S. Hageman, D. Deredge, P. L. Wintrode, M. Papanastasiou, J. D. Lambris, S. Li, S. Urata, *Anal. Chem.* **2019**, *91*, 7336.
- 40 Trabjerg, E., R. U. Jakobsen, S. Mysling, S. Christensen, T. J. D. Jorgensen, K. D. Rand, *Anal. Chem.* **2015**, *87*, 8880.
- 41 Zhu, S. L., P. Liuni, L. Ettore, T. Chen, J. Szeto, B. Carpick, D. A. James, D. J. Wilson, *Biochemistry* **2019**, *58*, 646.
- 42 Bonnington, L., I. Lindner, U. Gilles, T. Kailich, D. Reusch, P. Bulau, *Anal. Chem.* **2017**, *89*, 8233.
- 43 Hamuro, Y., M. G. Derebe, S. Venkataramani, J. F. Nemeth, *Protein Sci.* **2021**, *30*, 1686.
- 44 Houde, D., J. Arndt, W. Domeier, S. Berkowitz, J. R. Engen, *Anal. Chem.* **2009**, *81*, 2644.
- 45 Skinner, J. J., W. K. Lim, S. Bedard, B. E. Black, S. W. Englander, *Protein Sci.* **2012**, *21*, 996.
- 46 Bai, Y., J. S. Milne, L. Mayne, S. W. Englander, *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.
- 47 Calugareanu, D., I. R. Moller, S. G. Schmidt, C. J. Loland, K. D. Rand, *J. Mol. Biol.* **2022**, *434*, 15.
- 48 Coales, S. J., S. Y. E, J. E. Lee, A. Ma, J. A. Morrow, Y. Hamuro, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3585.
- 49 Gao, S., J. P. Klinman, *Curr. Op. Struct. Biol.* **2022**, *75*, 102434.
- 50 Liang, Z.-X., T. Lee, K. A. Resing, N. G. Ahn, J. P. Klinman, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9556.
- 51 Goswami, D., S. Devarakonda, M. J. Chalmers, B. D. Pascal, B. M. Spiegelman, P. R. Griffin, *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1584.
- 52 Klinman, J. P., *Accounts Chem. Res.* **2015**, *48*, 449.
- 53 Xiao, H., J. K. Hoerner, S. J. Eyles, A. Dobo, E. Voigtman, A. I. Mel'Cuk, I. A. Kaltashov, *Protein Sci.* **2005**, *14*, 543.

- 54 Xiao, P., D. Bolton, R. A. Munro, L. S. Brown, V. Ladizhansky, *Nat. Commun.* **2019**, *10*,
- 55 Best, R. B., M. Vendruscolo, *Structure* **2006**, *14*, 97.
- 56 Nguyen, D., L. Mayne, M. C. Phillips, S. W. Englander, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1936.
- 57 Covington, A. K., R. A. Robinson, R. G. Bates, *J. Phys. Chem.* **1966**, *70*, 3820.
- 58 Ratcliff, K., S. Marqusee, *Biochemistry* **2010**, *49*, 5167.
- 59 Bowler, B. E., *Curr. Op. Struct. Biol.* **2012**, *22*, 4.
- 60 Keppel, T. R., B. A. Howard, D. D. Weis, *Biochemistry* **2011**, *50*, 8722.
- 61 Zhu, S. L., A. Shala, A. Bezginov, A. Sljoka, G. Audette, D. J. Wilson, *PloS One* **2015**, *10*, 1.
- 62 Ng, Y. K., N. N. Tajoddin, P. M. Scrosati, L. Konermann, *J. Phys. Chem. B* **2021**, *125*, 13099.
- 63 Gallagher, D. T., I. Karageorgos, J. W. Hudgens, C. V. Galvin, *Data Brief* **2018**, *16*, 29.
- 64 Bergonzo, C., D. T. Gallagher, *J. Res. Natl. Inst. Stand. Technol.* **2021**, *126*, 6.
- 65 Pabon, J. P. R., B. A. Kochert, Y. H. Liu, D. D. Richardson, D. D. Weis, *J. Pharm. Sci.* **2021**, *110*, 2355.
- 66 Groves, K., A. Cryar, S. Cowen, A. E. Ashcroft, M. Quaglia, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 553.
- 67 Glasoe, P. K., F. A. Long, *J. Phys. Chem.* **1960**, *64*, 188.
- 68 Xiao, Y., M. Li, R. Larocque, F. Zhang, A. Malhotra, J. Chen, R. J. Linhardt, L. Konermann, D. Xu, *J. Biol. Chem.* **2018**, *293*, 17523.
- 69 Zhang, H. M., S. M. McLoughlin, S. D. Frausto, H. L. Tang, M. R. Emmett, A. G. Marshall, *Anal. Chem.* **2010**, *82*, 1450.
- 70 Distler, U., J. Kuharev, P. Navarro, Y. Levin, H. Schild, S. Tenzer, *Nat. Methods* **2014**, *11*, 167.
- 71 Peterle, D., T. E. Wales, J. R. Engen, *Anal. Chem.* **2022**, *94*, 10142.
- 72 Smith, D. L., Y. Deng, Z. Zhang, *J. Mass Spectrom.* **1997**, *32*, 135.

- 73 Skinner, S. P., G. Radou, R. Tuma, J. J. Houwing-Duistermaat, E. Paci, *Biophys. J.* **2019**, *116*, 1194.
- 74 Fajer, P. G., G. M. Bou-Assaf, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1202.
- 75 Ito, T., K. Tsumoto, *Protein Sci.* **2013**, *22*, 1542.
- 76 Feige, M. J., S. Walter, J. Buchner, *J. Mol. Biol.* **2004**, *344*, 107.
- 77 Privalov, P. L., Y. V. Griko, S. Y. Venyaminov, V. P. Kutysenko, *J. Mol. Biol.* **1986**, *190*, 487.
- 78 Lazar, K. L., T. W. Patapoff, V. K. Sharma, *mAbs* **2010**, *2*, 42.
- 79 Weinkam, P., J. Zimmermann, F. E. Romesberg, P. G. Wolynes, *Acc. Chem. Res.* **2010**, *43*, 652.
- 80 Hamuro, Y., *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 486.
- 81 Kan, Z. Y., B. T. Walters, L. Mayne, S. W. Englander, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16438.
- 82 Gessner, C., W. Steinchen, S. Bedard, J. J. Skinner, V. L. Woods, T. J. Walsh, G. Bange, D. P. Pantazatos, *Sci. Rep.* **2017**, *7*,
- 83 Pace, C. N., H. L. Fu, K. L. Fryar, J. Landua, S. R. Trevino, D. Schell, R. L. Thurlkill, S. Imura, J. M. Scholtz, K. Gajiwala, J. Sevcik, L. Urbanikova, J. K. Myers, K. Takano, E. J. Hebert, B. A. Shirley, G. R. Grimsley, *Protein Sci.* **2014**, *23*, 652.
- 84 Kazlauskas, R., *Chem. Soc. Rev.* **2018**, *47*, 9026.
- 85 Yang, C., S. Jang, Y. Pak, *Nat. Commun.* **2014**, *5*,
- 86 Dias, C. L., T. Ala-Nissila, M. Karttunen, I. Vattulainen, M. Grant, *Phys. Rev. Lett.* **2008**, *100*, 118101.
- 87 Sole-Domenech, S., A. V. Rojas, G. G. Maisuradze, H. A. Scheraga, P. Lobel, F. R. Maxfield, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 1493.

## 5 Chapter 5. Conclusions

### 5.1 Summary

The work in this thesis expands the analytical capabilities of HDX-MS to study thermal stability and thermodynamics of small globular protein to large, multidomain protein. Detailed information on chemical and physical stability is required to develop and commercialize novel protein drugs. In this regard, it is essential to establish advanced methods for providing comprehensive information for protein stability.

H/D exchange (HDX) mass spectrometry (MS) is a widely used technique for interrogating protein structure and dynamics. Backbone HDX is mediated by opening/closing (unfolding/refolding) fluctuations. In traditional HDX-MS, proteins are incubated in D<sub>2</sub>O as a function of time at constant temperature ( $T$ ). There is an urgent need to complement this traditional approach with experiments that probe proteins in a  $T$ -dependent fashion, e.g., for assessing the stability of therapeutic antibodies. In **Chapter 2**, a fundamental workflow is built for analyzing temperature- dependent HDX-MS. A key problem with such studies is the absence of strategies for interpreting HDX/MS data in the context of  $T$ -dependent protein dynamics. Specifically, it has not been possible thus far to separate  $T$ -induced changes in the chemical labeling step ( $k_{ch}$ ) from thermally enhanced protein fluctuations. Focusing on myoglobin, we solved this problem by dissecting  $T$ -dependent HDX-MS profiles into contributions from  $k_{ch}(T)$ , as well as local and global protein dynamics. Experimental profiles started off with surprisingly shallow slopes that seemed to defy the quasi-exponential  $k_{ch}(T)$  dependence. Just below the melting temperature ( $T_m$ ) the profiles showed a sharp increase. Our analysis revealed that local dynamics dominate at low  $T$ , while global events become prevalent closer to  $T_m$ . About half of the backbone NH sites exhibited a canonical scenario, where local opening/closing was associated with positive  $\Delta H$  and  $\Delta S$ . Many of the remaining sites had negative  $\Delta H$  and  $\Delta S$ , thereby accounting for the shallowness of the experimental HDX-MS profiles at low  $T$ . In summary, this chapter provides practitioners with the tools to analyze proteins over a wide temperature range, paving the way toward  $T$ -dependent high-throughput screening applications by HDX-MS.

Proteins that encounter unfavorable solvent conditions are prone to aggregation, a phenomenon that remains poorly understood. In **Chapter 3** we focused on myoglobin (Mb) as model protein. Upon heating, Mb produces amorphous aggregates. Thermal unfolding experiments at low concentration (where the aggregation is negligible), along with centrifugation assays, imply that Mb aggregation proceeds via globally unfolded conformers. This contrasts studies on other proteins that emphasized the role of partially folded structures as aggregate precursors. Molecular dynamics (MD) simulations were performed to gain insights into the mechanism by which heat-unfolded Mb molecules associate with one another. Binding was mediated by hydrophobic contacts, along with salt bridges that involved hydrophobically embedded Lys residues. Overall, this Chapter provides insights into protein aggregation mechanisms.

In **Chapter 4**, we focused on NISTmAb, a complicated multi-domain system where dynamics, unfolding, and aggregation are closely intertwined and we uncover thermodynamic parameters that govern the structural dynamics of different protein regions for different time points, and across temperatures ranging from near freezing to almost boiling. In this regard the framework from chapter 2 was applied. Global analysis of  $%D(t,T)$  produced excellent fits for all NISTmAb peptides, verifying the suitability of our strategy for complex protein therapeutics. Our data reveal that increasing temperature causes enhanced deuteration. The biophysical foundation of this effect is surprisingly complicated, as it reflects the combination of at least four factors. The analysis strategy pursued in this work allows to clearly distinguish between the four factors. Overall, we hope that this work will encourage practitioners to consider time and temperature as equivalent variables in HDX-MS experiments, specifically for investigations of mAbs and other protein therapeutics. Complete deuteration of highly protected NH sites can take days at ambient temperature, whereas complete deuteration can be achieved within ~20 minutes upon heating the samples. With suitable automation, it should be possible to streamline the temperature dependent HDX workflow employed here, such that the dynamic features across the entire protein become experimentally accessible in very short time windows. The thermodynamic model used here represents a quantitative analysis tool for deciphering the various factors that govern the temperature- and time-dependence of protein

deuteration. Spatially resolved data obtained by  $T$ -dependent HDX-MS provide a more detailed picture than DSC or related techniques that only report on global unfolding.

## 5.2 Future Directions

### 5.2.1 Develop Instrumentation and Software to Analyze $T$ -dependent HDX-MS Data

Temperature-dependent HDX-MS can provide valuable information about protein thermodynamics and fluctuation, and this method would be very beneficial for assessing protein therapeutic stability in biopharmaceutical companies. However, a key obstacle to employing this strategy is the manual experiment procedure and time-consuming data processing. It may come as no surprise that the global fitting of Chapters 2 and 4 required many weeks of painstaking work, involving multiple highly complex Excel files. At the current stage of development, such strategies will likely not be adopted by researchers in pharmaceutical companies. It is hoped that future work will streamline this analysis process via the development of dedicated software. Similarly, temperature-dependent HDX-MS experiments are quite laborious. Automated equipment is available for traditional time-domain measurements. It is hoped that it will be possible to adapt such robotic platforms for temperature-dependent measurements.

### 5.2.2 Computational Simulations of Aggregation for Large Multi-Domain Proteins

Monoclonal antibodies are multi-domain proteins that have three different melting temperatures. Individual domains unfold sequentially, and our observation shows antibody remains soluble up to the temperature all the three domains are unfolded. This is in contrast to the theory that the unfolding of one domain would be sufficient for triggering aggregation. It raises the question of how unfolded domain is soluble and does not trigger aggregate.

The aggregation of protein therapeutics is one of the main challenges in biopharmaceutical companies, and the computational simulation for this big system would provide invaluable information for their thermal stability, which can be beneficial for the drug industry. Moreover, it would shed additional light on other questions related to protein dynamics and thermodynamics.

## APPENDIX I-PERMISSIONS



RightsLink



Home



Help ▾



Email Support



Sign in



Create Account

### Analysis of Temperature-Dependent H/D Exchange Mass Spectrometry Experiments



**Author:** Nastaran N. Tajoddin, Lars Konermann

**Publication:** Analytical Chemistry

**Publisher:** American Chemical Society

**Date:** Jul 1, 2020

*Copyright © 2020, American Chemical Society*

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

### Mechanism of Thermal Protein Aggregation: Experiments and Molecular Dynamics Simulations on the High-Temperature Behavior of Myoglobin



**Author:** Yuen Ki Ng, Nastaran N. Tajoddin, Pablo M. Scrosati, et al

**Publication:** The Journal of Physical Chemistry B

**Publisher:** American Chemical Society

**Date:** Dec 1, 2021

*Copyright © 2021, American Chemical Society*

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

### Structural Dynamics of a Thermally Stressed Monoclonal Antibody Characterized by Temperature-Dependent H/D Exchange Mass Spectrometry



**Author:** Nastaran N. Tajoddin, Lars Konermann

**Publication:** Analytical Chemistry

**Publisher:** American Chemical Society

**Date:** Oct 1, 2022

*Copyright © 2022, American Chemical Society*

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}, Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

## APPENDIX II- Curriculum Vitae

**Name:** Nastaran Nosrat Tajoddin

**Post-secondary Education and Degrees:** University of Birjand  
Birjand, Iran  
2007-2012 B.Sc., Applied chemistry

University of Tabriz  
Tabriz, Iran  
2012-2015 M.Sc., Nanoscience and Nanotechnology

**Honours and Awards:** Hadis Hayatdavoudi Graduate Scholarship in Chemistry  
The University of Western Ontario  
2021

**Related Work Experience** Teaching Assistant  
The University of Western Ontario  
2018-2022

### Publications:

- Y. K. Ng, N. N. Tajoddin, P. M. Scrosati, and L. Konermann, *J. Phys. Chem. B*, vol. 125, no. 48, pp. 13099–13110, 2021.
- N. N. Tajoddin and L. Konermann, *Anal. Chem.*, vol. 92, no. 14, pp. 10058–10067, 2020.
- T. Hossiennejad, M. Daraie, M. M. Heravi, and N. Nosrat Tajoddin. *J. Inorg. Organomet. Polym. Mater.*, vol. 27, pp. 861–870, 2017.
- S. Mohammadi, B. Sarkhosh, and N. Nosrat Tajoddin. *Journal of Mater. Sci. Mater. Electron.*, vol. 28, pp. 9456–9463, 2017.