9-20-2018 10:00 AM

Hydrogen-Deuterium Exchange Mass Spectrometry and Molecular Dynamics Simulations for Studying Protein Structure and Dynamics

Yiming Xiao
The University of Western Ontario

Supervisor
Lars Konermann
The University of Western Ontario

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

Abstract

Deciphering properties of proteins is essential for improving human health and aiding in the development of new pharmaceuticals. While many investigations have focused on protein structures, it is equally important to probe protein dynamics, because conformational fluctuations are often intricately connected with protein function. This dissertation uses hydrogen-deuterium exchange (HDX) mass spectrometry (MS) and molecular dynamics (MD) simulations to study protein dynamics, for improving the understanding of protein folding/unfolding mechanisms, and to uncover aspects related to ligand binding and allosteric regulation. The common thread throughout this thesis is that previous studies have greatly underestimated the role of conformational fluctuations for the function of the three proteins studied in this work (and likely for many other proteins as well).

Proteins possess surfactant-like properties, resulting in a high affinity for gas/water interfaces. Chapter 2 uses HDX-MS for probing the conformational dynamics of myoglobin (Mb) in the presence of N_2 bubbles. HDX/MS relies on the principle that unfolded and/or highly dynamic regions undergo faster deuteration than tightly folded segments. In bubble-free solution Mb displays regular dynamics as under the native environment. However, in the presence of N_2 bubbles, some mixed dynamics take place which shows protein are in native/semi-denaturing environments. To explain the observed deuteration kinetics, we propose a dynamic model that quantitatively reproduces the experimentally observed data: “semi-unfolded” ↔ “native” ↔ “globally unfolded” → “aggregated”.

Chapter 3 focuses on osteoprotegerin (OPG), which acts as a receptor activator of nuclear factor κB ligand (RANKL) decoy receptor and hinders bone resorption by inhibiting the interaction between RANKL and its binding partner, receptor activator of nuclear factor κB (RANK). OPG exists as monomer and dimer, both of which are biologically active. The transformation between monomer and dimer can be regulated by heparan sulfate (HS), but the details of this regulatory processes remains unclear. OPG undergoes fast EX2 exchange

which reveals the high dynamics of the protein. Site-specific changes in HDX rates accurately identify the HS and RANKL binding regions. Our HDX data also discovered that the RANKL binding patterns are somewhat different for OPG dimers and monomers. A mechanism is proposed for formation of the RANKL/OPG/HS ternary complex, according to which HS-mediated C-terminal contacts on OPG lower the entropic penalty for RANKL binding.

Chapter 4 represents the center piece of this thesis. It explores the allosteric regulation of S100A11, a dimeric EF-hand protein with two hydrophobic target binding sites. An annexin peptide (Ax) served as the target. The allosteric mechanism was probed by HDX/MS, complemented by microsecond MD simulations. Consistent with experimental data, MD runs in the absence of Ca^{2+} and Ax culminated in target binding site closure. In simulations on [Ca_4 S100] the target binding sites remained open. These results capture the essence of allosteric control, revealing how Ca^{2+} prevents binding site closure. Both HDX/MS and MD data showed that the metalation sites become more dynamic after Ca^{2+} loss. However, these enhanced dynamics do not represent the primary trigger of the allosteric cascade. Instead, a labile salt bridge acts as an incessantly active “agitator” that destabilizes the packing of adjacent residues, causing a domino chain of events that culminates in target binding site closure. Overall, this thesis highlights how the combination of HDX/MS and computational techniques can provide detailed insights into the nature of protein conformational fluctuations, and their implications for protein function,

Keywords: protein dynamics | hydrogen-deuterium exchange | mass spectrometry | myoglobin | osteoprotegerin | S100A11 | molecular dynamics simulation

Statement of Co-Authorship

The works in Chapters 2 and 4 were published in the following articles, respectively:

> Xiao, Y.; Konermann, L., Protein structural dynamics at the gas/water interface examined by hydrogen exchange mass spectrometry. *Protein Sci*. 2015, 24 (8), 1247-1256. Reproduced with permission © 2015, John Wiley and Sons.
>
> Xiao, Y.; Shaw, G. S.; Konermann, L., Calcium-Mediated Control of S100 Proteins: Allosteric Communication via an Agitator/Signal Blocking Mechanism. *J. Am. Chem. Soc.* 2017, 139 (33), 11460-11470. Reproduced with permission © 2017, American Chemical Society.
>
> The work in Chapter 3 has been incorporated into the following article:
>
> Xiao, Y.; Li, M.; Larocque, R.; Zhang, F.; Malhotra, A.; Linhardt, R.J.; Konermann, L.; Xu, D., Dimerization interface of osteoprotegerin revealed by hydrogen-deuterium exchange mass spectrometry. Manuscript accepted on *JBC*. 2018, in press.

The original draft for each of the above articles was prepared by the author. Subsequent revisions were by the author and Dr. Lars Konermann. All experimental work and data analysis were performed by the author under the supervision of Dr. Konermann.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor Dr. Lars Konermann, who is a true gentleman with a great passion for science. Working with him is a wonderful experience, and I enjoy my every second of being in the lab. He motivates and inspires me to explore an alien field that is entirely new to me. Whenever I was facing difficulties, I can always get help from his knowledge and experiences. He never minds whether my question is reasonable or ignorant and replies to me with extreme patience. I can never finish my Doctoral study without the guidance from him. In my future path, no matter where I am, I will always respect Lars as my teacher, my mentor, not only in science but also the spirit to both work and life.

I also want to thank all my wonderful lab mates that overlap with me in the last five years. Their personalities and expertise improve me in many aspects. Thanks to Victor, you sacrifice so much time on my mediocre writing; thanks to Haidy, you are the mum to the lab; thanks to Maryam, you keep bringing joy to the lab; thanks to Vincent, you raise the fashion bar of the lab; thanks to Angela, you are a great listener; thanks to all other lab mates: Siavash, Dupe, Aisha, Robert, Justin, Sherry, you all means a lot to me. Thank you all for showing up in my life.

My collaborators have all been an excellent source of support. Many thanks to Dr. Gary Shaw for providing guidance and valuable input in our collaborative project. I would like to acknowledge Dr. Ding Xu; without your endless supply of protein sample and scientific data, there will be no Chapter 3 in my thesis. Thank you all for your scientific knowledge and professional help.

At last, I want to thank my Family. You mean the world to me. Dan, thank you for tolerating my childishness. You have always been the sunshine in the darkest days. My baby girl, Dariel, I can never get tired of being with you. Thank you for choosing me to be your father. Thank you, mum, your endless love accompanies me through my life. I consider me to be the luckiest person in the world to have such a great and supportive family.

Contents

List of Appendices

List of Symbols and Abbreviations

UV-Vis	ultraviolet-visible
I_0	initial light intensity
I	transmitted intensity
A	absorbance
ε	extinction coefficient
l	length of the light path in the cuvette
c	sample concentration
CD	circular dichroism
A_L, A_R	absorbance for L and R polarized light
NMR	nuclear magnetic resonance
M	net magnetization vector
EM	electron microscopy
CCD	charge coupled device
MS	mass spectrometry
LC	liquid chromatography
HPLC	high performance liquid chromatography
UPLC	ultra performance liquid chromatography
GC	gas chromatography
CE	capillary electrophoresis

ESI	electrospray ionization
m/z	mass to charge ratio
z	charge state
z_R	number of charges
ε_0	vacuum permittivity
γ	surface tension
r	radius of the droplet
IEM	ion evaporation model
CRM	charged residue model
CEM	chain ejection model
EI	electron ionization
CI	chemical ionization
APCI	atmospheric pressure chemical ionization
MALDI	matrix-assisted laser desorption ionization
DESI	desorption electrospray ionization
PS	paper spray
RF	radio frequency
DC	direct current
TOF	time of flight
U	pusher voltage

HO•	hydroxyl radical
HDX	hydrogen deuterium exchange
E_a	activation energy
T	temperature
R	gas constant
DDA	data-dependent acquisition
DIA	data independent acquisition
P	protein
L	ligand
PL	protein-ligand complex
GPCR	G protein-coupled receptor
MD	molecular dynamics
PME	Particle-Mesh-Ewald
PBC	periodic boundary conditions
NVT	constant number of particles, volume, and temperature
NPT	constant number of particles, pressure, and temperature
RMSD	root mean square deviations
RMSF	root mean square fluctuation
Mb	myoglobin
OPG	osteoprotegerin

RANK	receptor activator of nuclear factor-κB
RANKL	RANK ligand
CRD	cysteine-rich domain
HS	heparan sulfate
Ax	annexin peptide

Chapter 1: Introduction

1.1 Proteins

Proteins are among the most important macromolecules in living organisms. Together with polysaccharides, nucleic acids, and membrane lipids, proteins participate in almost every biological process. For example, many proteins act as enzymes that catalyze chemical reactions associated with metabolism. Also, proteins can mediate structure and movement, e.g., actin and myosin in muscle cells and tubulin in microtubules. Some proteins are involved in cell signaling, immune response, and cell cycle control. Protein research is a highly interdisciplinary field that spans across chemistry, biology, pharmacology, physics, and engineering.[1-3]

1.1.1 Protein Structure

Proteins are polymers consisting of amino acids. The amino acid sequence constitutes the primary structure, and it is determined by the corresponding gene (Figure 1a). Protein chains spontaneously fold and form higher order structures. The most common two secondary structures are α-helix (Figure 1b) and β-sheet. These secondary structure elements can further assemble into tertiary structure. One main driving force for forming tertiary structures is the hydrophobic effect, but other contacts such as salt bridges, disulfide bonds, and hydrogen bonds also help stabilize tertiary structure (Figure 1c), thereby defining the ways in which the protein folds and functions. In some cases, several folded chains assemble and form a quaternary structure (Figure 1d) that is capable of performing complex biological functions. For example, ATP synthase, the enzyme creates ATP in cells, has over 20 subunits.[4] Also, some proteins accommodate cofactors such as heme groups in the case of hemoglobin and myoglobin.

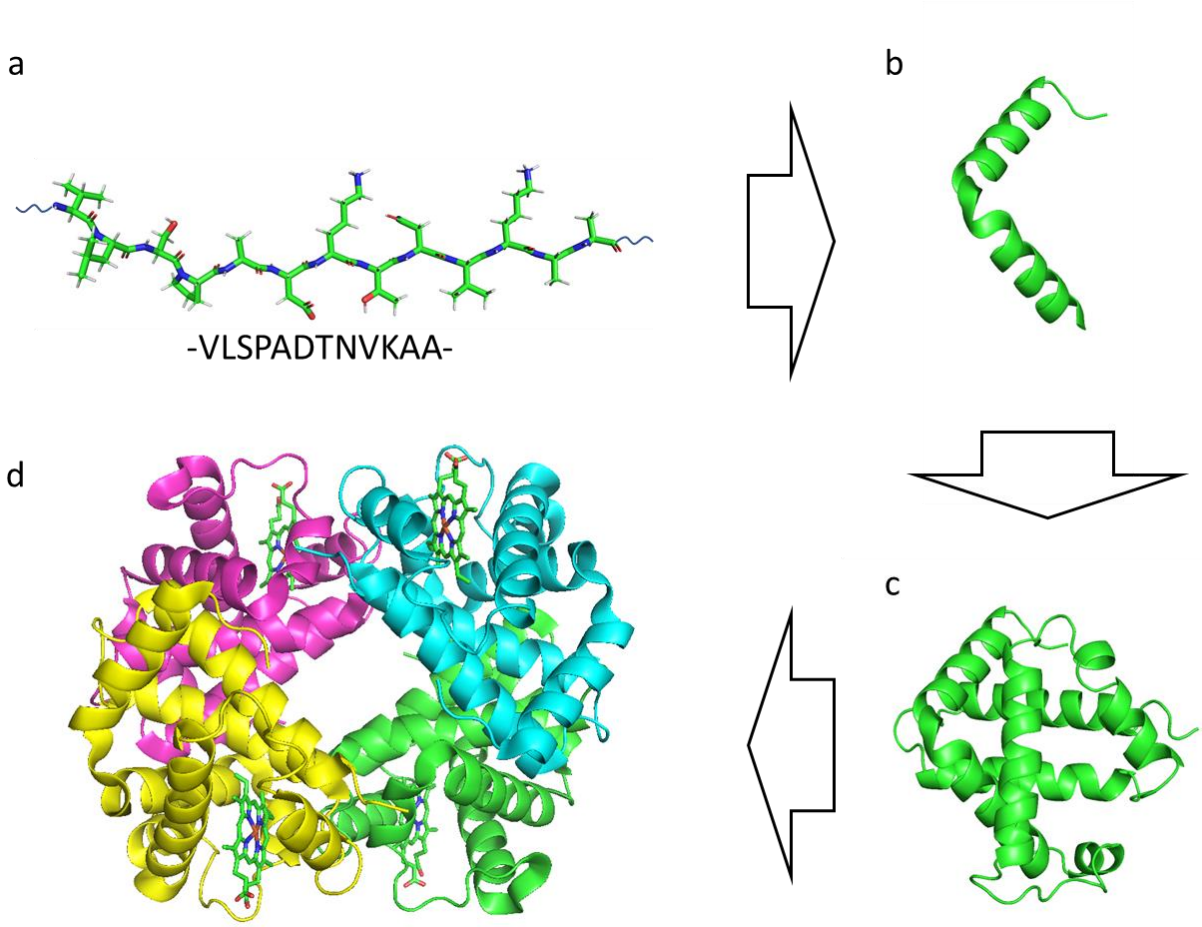


Figure 1. Levels of *protein structures. (a) primary structure: single letter residues codes are used to represent the amino acid sequence. (b) secondary structure: α-helix with a kink is used as an example. (c) tertiary structure; deoxyhemoglobin α-subunit. (d) quaternary structure: full tetramer structure of deoxyhemoglobin. (PDB:2hhb)*[5]

For correctly performing biological functions, proteins need to reach their "proper" (native) conformation. As an example, in trypsin a properly structured S1 catalytic pocket is essential for peptide/protein cleavage.[6] (Figure 2) Trypsin can only perform its function when catalytic residues are located correctly. How can proteins precisely fold into these delicate structures? How are proteins able to adopt these complicated conformations from a simple linear sequence of amino acids? These questions have intrigued scientists for many decades. In 1969, Cyrus Levinthal pointed out his famous "Levinthal's paradox".[7] Hypothetically, if a protein is composed of 100 amino acids and each residue can adopt

3×3 conformations (via rotation of N-C_α bonds and C-C_O bonds), the protein would have roughly $3^{200} \approx 10^{95}$ possible conformation. Levinthal initially envisioned that protein folding is a random search process. If the protein were able to survey 10^{10} conformations per second finding its native state would require around 10^{77} years which is longer than the existence of the universe (15 billion years). It is concluded that the folding of proteins cannot be achieved through random conformational searches. A more efficient mechanism must exist.

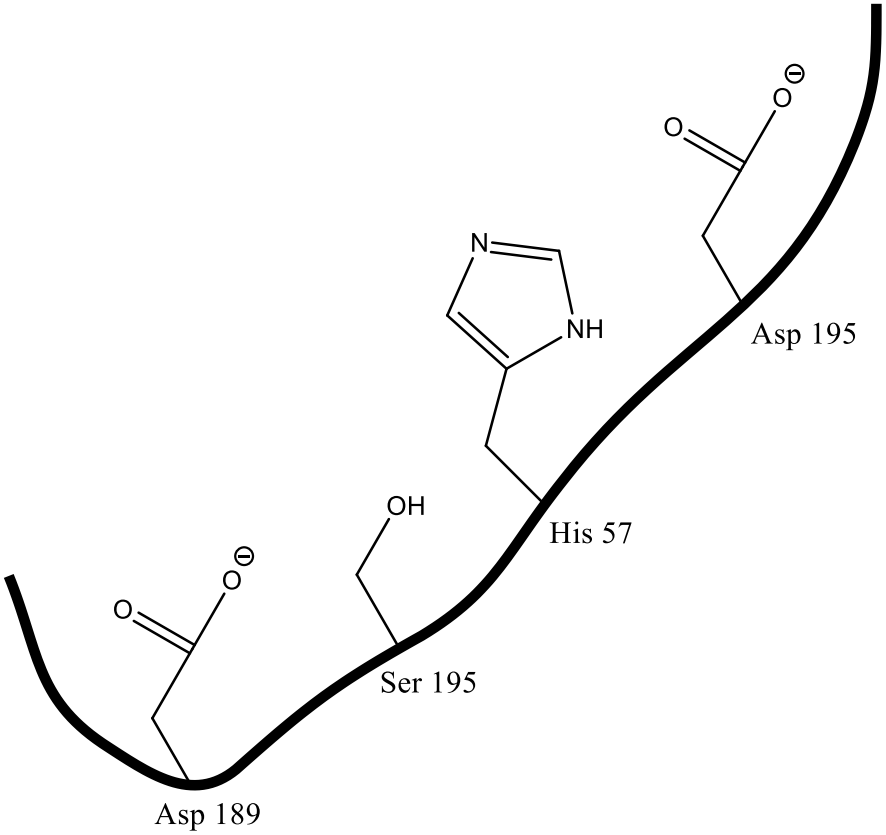


Figure 2. *Trypsin catalytic amino acids at S1 pocket.*

Folding funnel/energy landscape theory is the most successful hypothesis so far for solving Levinthal's paradox. (Figure 3) Imagine a ball sitting on a sloped surface, naturally, driven by the gravitational potential energy the ball would roll downhill to the lowest point on the surface. A similar thing also happens in the case of proteins: while a protein is fully unfolded and exists as a random coil, it has a high free energy due to unfavorable ϕ and ψ

angles (dihedral angles of the roation about Cα-N and C-Cα bonds), water-exposed hydrophobic residues, etc. The protein molecule has a tendency to transit to a conformation with lower free energy, akin to skiing down a slope from higher (energy) ground. Under this circumstance, aimless random conformational searching is suppressed. It makes protein folding possible in a short period (10^{-6} to 10^{3} seconds)[8]. Also similar to skiing, there could be multiple pathways in protein folding. The folding path might pass through local minima where folding-intermediates or molten globules are generated. Eventually, the protein reaches the free energy minimum and finishes the folding process. Alternatively, if the protein pursues a wrong folding path, it may misfolded and/or aggregate.[9]

The main driving force in protein folding is the hydrophobic effect which can be explained by using the "iceberg" model. Many residues in proteins are hydrophobic (valine, leucine, isoleucine, etc.). The existence of "iceberg"-like water molecules that is packed around these exposed hydrophobic sites greatly decreases the solvent entropy which represents an unfavorable scenanrio according to the second law of thermodynamics. The amount of "icebergs" can be minimized as hydrophobic residues in the protein get packed within the protein core during folding. Water molecules are then released back to the bulk solution and increase the entropy of the system. Folding is also driven by enthalpic factors such as the formation of new H-bonds, disulfide bonds, salt bridges and van der Waals contacts. In short, protein folding is driven by the tendency of the system to decrease its free energy.

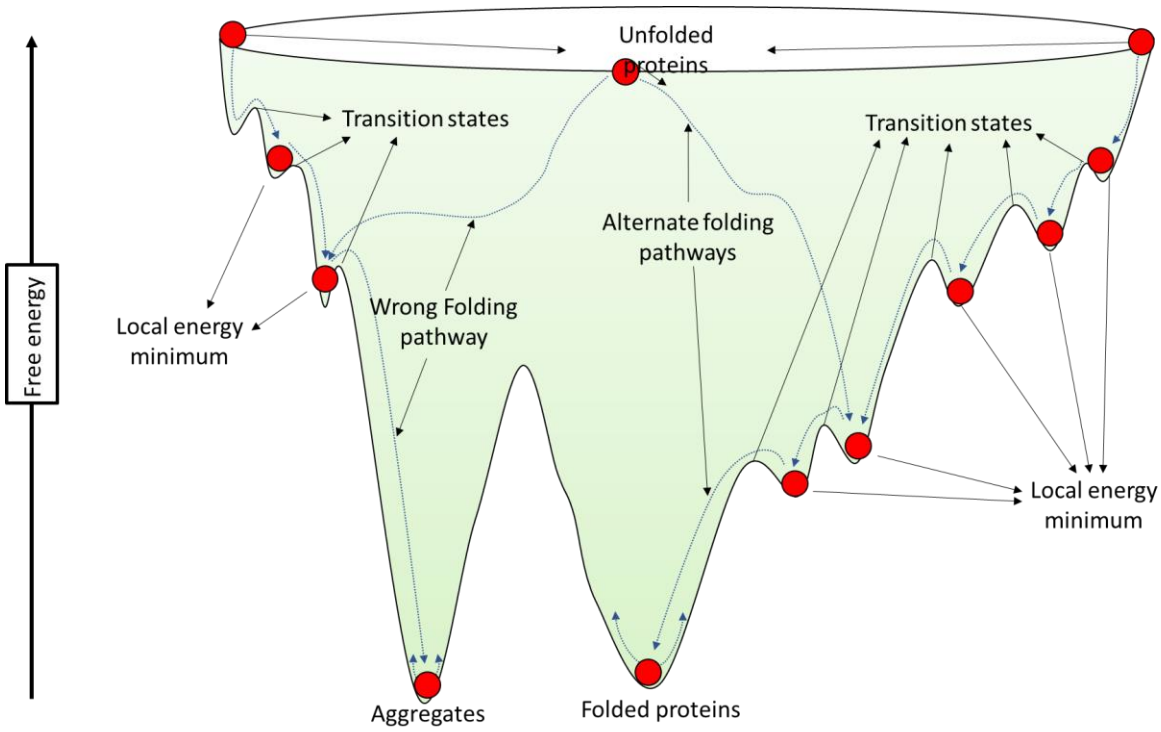


Figure 3. *Schematic energy landscape. Protein conformations are represented as red balls. It is assumed that thermal motion allows the protein to freely explore the energy landscape to reach a (global or local) minimum. Folding and aggregation can have many alternative pathways with different local minima and transition states.*

1.1.2 Protein Dynamics

There is no doubt that proteins need to fold to correct structures in order to perform their biological functions. However, these structures is not static. Proteins in solution are an ensemble of many similar conformations that undergo rapid interconversion (protein dynamics). For example, multiple possible NMR-derived structures of S100A11 were superimposed in Figure 4; each of them contributes to the conformational ensemble that exists in solution.

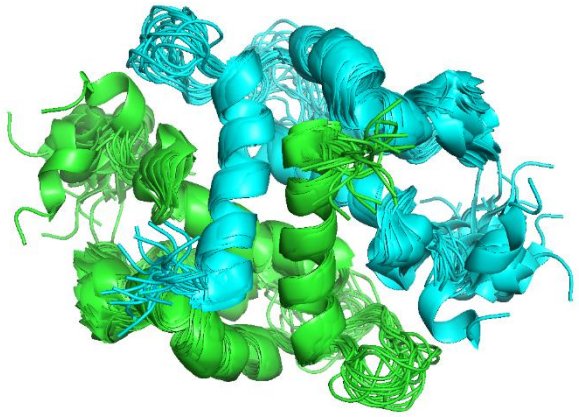


Figure 4. *19 Superimposed NMR structures of S100A11. Green and cyan represent the two monomers in the protein complex (pdb code 1NSH).*[10]

Protein dynamics comprise events taking place on time scales from ps to minutes; and the amplitude can be small or large: from side chain rotation to global transitions between protein conformations.[11,12] Potein dynamics are highly related to protein function. In enzymes, protein dynamics are believed to be the key for catalytic turnover.[13,14] For example, in ATP synthase, the γ subunit (the central shaft) gets destabilized during catalysis[15]; and the dynamics of chymotrypsin are intensified during substrate turnover.[16] Protein-ligand interactions also affect protein dynamics, generally resulting in a stabilization in the vicinity of the binding site due to the formation of new intermolecular contacts. By monitoring such dynamic changes, it is often possible to locate ligand binding sites, e.g., in the context of epitope mapping and drug-protein interaction studies.[17,18] In contrast, by studying the change of protein dynamic properties from ligand binding, researchers can obtain deeper insight on how ligands impact protein.[19] For instance, the dynamics of ligand binding pocket can be studied by computational methods to direct the design of inhibitors.[20,21]

Studying protein dynamics is not straightforward because it is nearly impossible to "see" how proteins move. However, sophisticated and advanced instrumental and experimental designs are now capable of providing quite detailed insights. Nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) are both capable of providing such information.[22,23] More details are discussed in section 1.2.

1.2 Traditional Methods for Studying Protein Structures

1.2.1 UV/Vis Spectroscopy

UV/Vis spectroscopy is a technique that measures the ability of a sample to absorb light. Its principle, Beer's law, can be written as

$$log_{10}\frac{I_0}{I} = A = \varepsilon \cdot l \cdot c$$

Equation 1.1

where I_0 is the initial light intensity; I is the light intensity after the light passes through the sample; A is the absorbance; ε is the wavelength-dependent extinction coefficient that is unique to the sample; l is the length of the light path in the cuvette, and c is the sample concentration. The most common use of UV/vis spectroscopy in protein research is to determine protein concentration. The protein's extinction coefficient depends on the number of tryptophan, tyrosine and phenylalanine.[24] Thus, protein concentrations can be obtained by measuring the absorbance at 280 nm.

UV/Vis spectroscopy can also be used in study protein folding, this approach is most useful for proteins that contain UV-Vis active cofactors. In bacteriorhodopsin (BR), a retinal incorporates in the center of the protein. For native BR in the purple membrane an

absorbance maximum shows at 568 nm, indicating the retinal is covalently bound to Lys216.[25] In a denaturing SDS environment, the absorbance spectrum is blue-shifted to 392 nm indicating the disruption of the native retinal-protein linkage.[26] The straightforward nature of UV-Vis spectroscopy causes it to be widely used for protein studies. However, only limited structural information is obtainable in this way.

1.2.2 Circular Dichroism (CD) Spectroscopy

Similar to UV-Vis spectroscopy, CD also measures the difference of light before and after passing through the sample. However, CD monitors another light property, polarization. Light can be circularly polarized clockwise (right, R) or counter-clockwise (left, L). While polarized light passes through the sample, any chiral center will have different absorbance for L and R polarized light (A_L, A_R). The difference of A_L-A_R is what constitutes a CD spectrum. The far-UV region, between 180 nm and 250 nm in the CD spectrum reports on secondary structure: Two negative peaks (222 and 208 nm) and one positive peak (192 nm) represent α helices ; parallelly, one negative peak (218 nm) and one positive peak (195 nm) serve signs as β sheet; one negative peak at 195 nm reflect the presence of random coil.[27,28] Examples of CD far-UV spectra are shown in Figure 5. However, despite its widespread use, CD spectroscopy still cannot yield residue/atom level structural information.

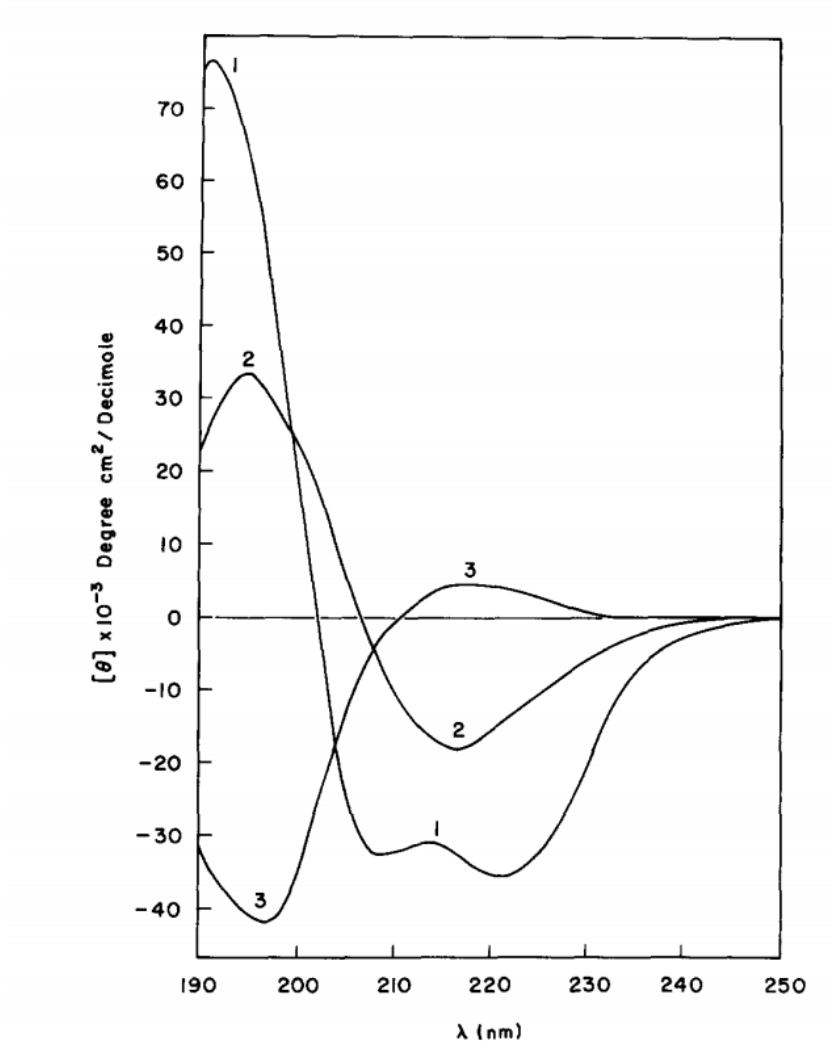


Figure 5. *CD spectra of proteins with characteristic secondary structure.*[29] *Respectively, (1) a-helix; (2) β-sheet; (3) random coil.*[30]

1.2.3 X-ray Crystallography

X-ray crystallography has been the gold standard in protein biophysics for many years. It can provide high-resolution structural information (down to 0.9 Å resolution). Under optimal conditions it can even reveal the position of hydrogen atoms.[31]

X-ray crystallography is based on diffraction. The X-ray wavelengths used for this purpose

are on the order of 0.1 nm, which is comparable to chemical bond lengths. This is the reason why X-rays can be diffracted while passing through a protein molecule. However, a single molecule usually cannot generate a diffraction signal strong enough to be recorded. In a protein crystal, a 3D array of the same molecule amplifies the diffraction signals and protein structures can be uncovered with the help of Fourier transform analyses. Roughly 90% of all structures stored in the protein data bank (PDB) have been generated using X-ray crystallography.[32]

X-ray crystallography is good at revealing the precise 3D-structure of proteins all the way to very large (MDa) complexes. However, dynamic aspects are difficult to uncover.[33,34] Some protein intermediates have been trapped and crystallized, but that requires sophisticated experimental designs, complicated data analysis tools, and often sheer luck. In general, crystallization of proteins can be challenging and time-consuming.[35] Growing high-quality crystals of a new protein can take years even with the help of automated high throughput systems.[36]

1.2.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR might be the most widely used principal analytical technique. Its research objects include almost everything in chemistry: from small molecules to polymers, from solution to solid, from inorganic to biological molecules. The spin of atomic nuclei creates a magnetic moment. When the atom is placed in a strong magnetic field, the magnetic moment of the nucleus can be aligned with the external magnetic field in a parallel or antiparallel fashion and therefore create an energy split (Zeeman effect). The spinning nuclei together create a net magnetization vector (M). M can be “tilted” away from its equilibrated position by well-tuned radio-frequency (RF) electromagnetic radiation and precess in its new orientation, thereby inducing a current in the detector (receiver coil). The electron cloud around the nuclei can shield the magnetic field effect on the nuclei and therefore alter the frequency and the amplitude of the current (chemical shift). Thus, NMR

can be used to study the chemical properties of the sample and reaction mechanisms.

For protein studies, NMR takes on a distinctly important role. Benefiting from the instrumental and the experimental developments (i.e., 2D-NOESY, 3D-^{15}N/^{13}C NOSEY-HSQC, etc.), NMR can also provide 3D protein structures. Over 10,000 protein structures in the PDB have been obtained using NMR.[32] NMR can also be used for probing protein dynamics[37], especially on intrinsically disordered proteins that are a considerable challenge for X-ray crystallography.[38] The upper size limit in protein NMR is often around 30 kDa. However, with the help of sophisticated experimental design it is now possible to examine larger systems (over 200k Da).[39,40] Unfortunately, such experiments often focus more on functional/dynamical instead of generating atomic level structures.[41] The relatively high sample concentrations (often in the 100 μM to 1 mM range) also limit NMR application to some degree, because such samples tend to aggregate.

1.2.5 Cryo-Electron Microscopy (cryo-EM)

Another technique that can provide protein structure is cryo-EM. The 2017 Chemistry Nobel Prize was awarded for the development of cryo-EM. In the words of the Nobel committee, cryo-EM has brought biochemistry "into a new area." EM is not a new technique, and its application on biological systems can be traced back to the 1930/40's.[42,43] However, biological samples cannot be directly viewed by EM in solution; sample dehydration or fixation is needed for EM. These sample preparation steps may introduce artifacts. Cryo-EM solves this problem and makes the observation of "aqueous samples" possible. The principle of cryo-EM is to cool the sample to below liquid nitrogen temperature quickly. The cooling process is so fast that vitreous ice is formed instead of regular crystalline ice, and the structures of biological molecules can be well preserved. Molecules in vitreous ice are randomly oriented and leave a "trace" image by interacting with the electron beam. 3D molecular structures can be obtained by combining thousands of images through computer programs. Cyro-EM does not require crystallized samples,

and this important feature makes getting some challenging protein structures possible, e.g., membrane protein complexes. (Figure 6) Recent improvements of the charge coupled device (CCD) detectors and computing algorithms used for cryo-EM have dramatically enhanced the resolution of this technique to near X-ray crystallography level (from ~20 Å to below 4 Å).[44] All these developments make cryo-EM a very attractive tool for protein structural biology. However, the ~US$ 10M price tag on cryo-EM instruments limits its application.[45]

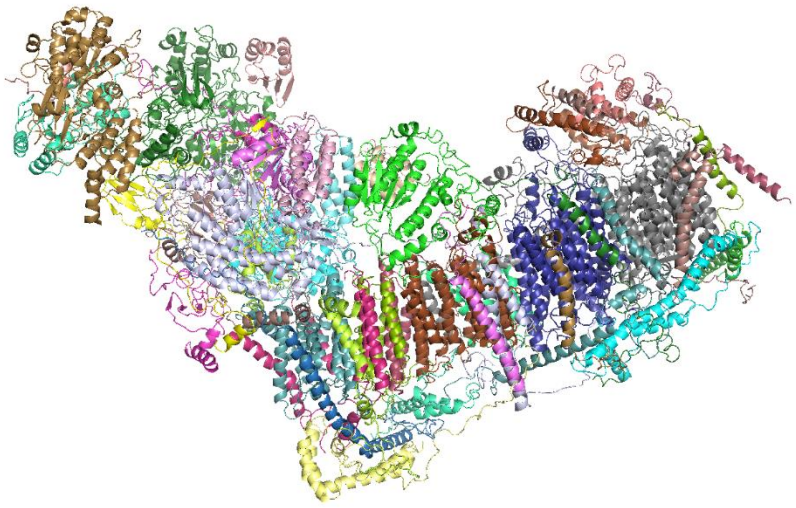


Figure 6. *Structure of mammalian respiratory complex I (NADH: ubiquinone oxidoreductase). Complex I is one of the largest complexes in the mammalian cell and contains over 45 subunits. This model was generated by cryo-EM (Resolution 4.16 Å, pdb file 5LDW).*[46]

1.3 Mass Spectrometry

Mass spectrometry (MS) has been a vital analytical tool in chemistry, biology, and other areas for many years. In the early 20th century, MS first demonstrated the existence of isotopes.[47] After nearly a century of development on instrument and data analysis, MS can now reach a resolution of ~1 million. This high resolution dramatically expands the usage of MS from identifying small molecules[48] to characterizing complicated protein complexes.[49] MS can be coupled with chromatographic separation techniques such as

liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE), etc. Also, multiple components (collision cells, ion mobility modules, etc.) can be combined with MS, and these tools provide additional dimensions of information.

Generally, MS can be divided into several key parts: an ion source, a mass analyzer, and an ion detector (Figure 7). The following sections will briefly explain the basis of MS.

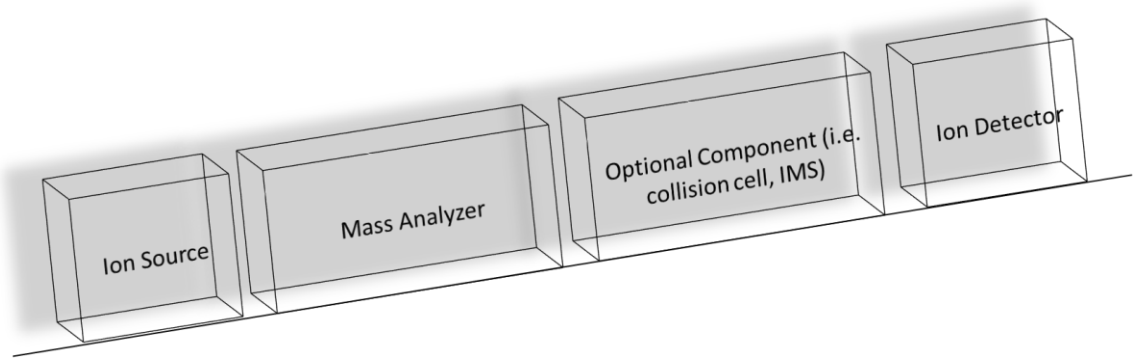


Figure 7. *Schematic of a mass spectrometer*

1.3.1 Ion Source

1.3.1.1 Electrospray Ionization (ESI)

MS can be imaged as a precise "balance" for measuring the mass (more accurately, the mass-to-charge-ratio *m/z*) of molecules and atoms. But no single molecule or atom can be really "put" on a balance to measure: analytes need to be ionized first and transferred into the vacuum of the mass analyzer. Various types of ion sources are commercially available. Each type has its advantages and disadvantages. For biological samples, electrospray ionization (ESI) represents the most successful and widely adopted ionization technique. ESI can transform solution phase molecules into gas phase ions. Therefore, ESI can interface with high performance liquid chromatography (HPLC) and ultra performance liquid chromatography (UPLC). This feature greatly extends MS application in analytical chemistry. Also, ESI is a "soft" ionization technique which means it is capable of analyte

ionization without rupturing covalent bonds. In many cases, even the preservation of noncovalent contacts is possible, which is the basis of native MS experiments.[50]

The basic layout of an ESI source is shown in Figure 8. The sample solution is pumped through a metal capillary which is attached to a high voltage supply (±2~6 kV).[51] The strong electrical field at the tip of the capillary creates a Taylor cone. At the tip of the cone, caused by Coulumbic repulsion which overcomes the surface tension, a mist of droplets that are enriched in charged analytes is ejected. With the help of source heating and a flow of nebulizing gas, droplets quickly shrink and vaporize to transfer charged ions into the gas phase for downstream analysis.

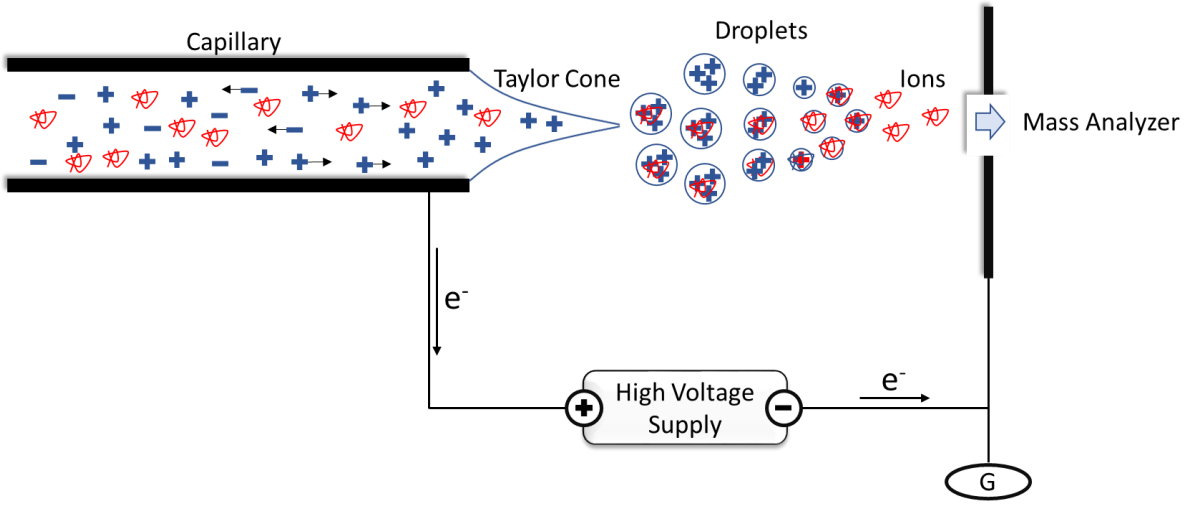


Figure 8. *Schematic layout of an ESI source (positive mode). Charge and analytes (i.e., protein) are represented in blue and red, respectively. The flow of electron is indicated by arrow. G denotes the grounded wire.*

An example of an ESI mass spectrum is shown in Figure 9. A typical attribute of the spectrum is the presence of multiple charge states. The mass to charge ratio m/z of a specific ion is given by

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

Equation 1.2

where z is the (variable) charge state; M is the mass of the neutral analyte, and 1.008 refers to the proton mass (in Da). This equation assumes that the analyte charge is entirely due to protonation.

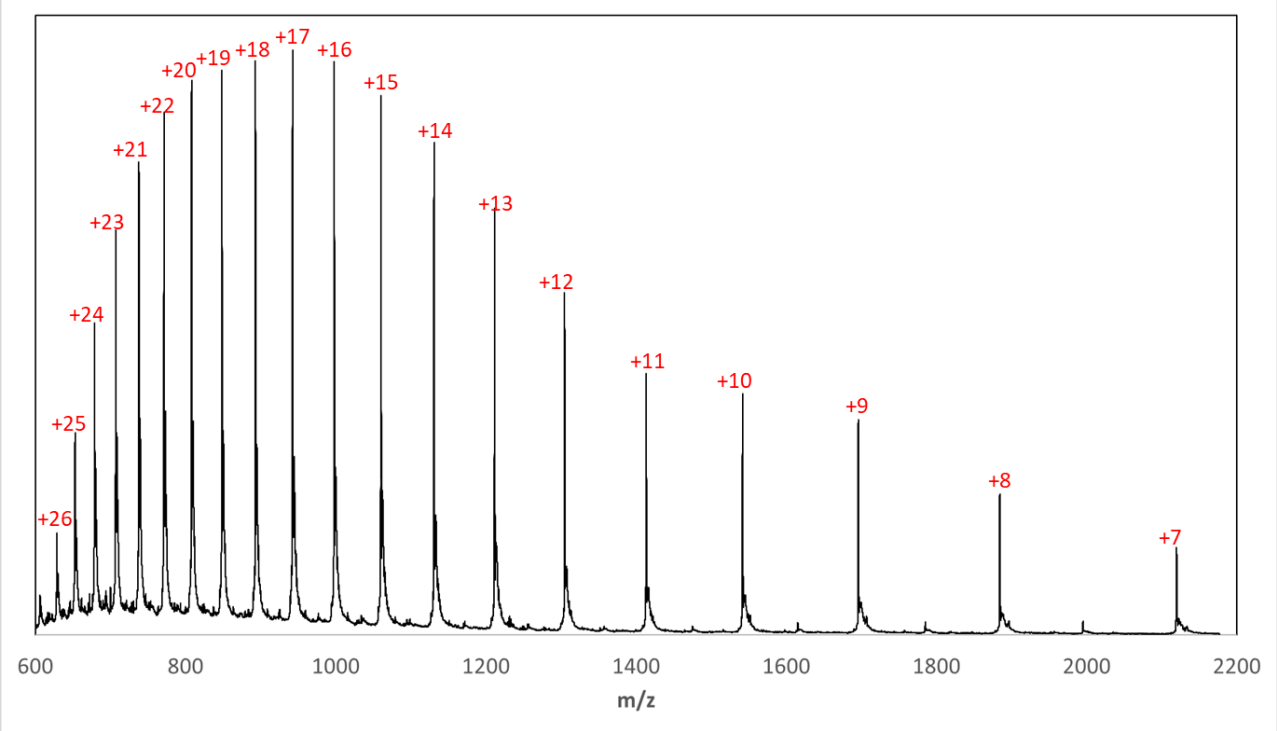


Figure 9. *ESI mass spectrum of apo-myoglobin. The charge states of each peaks are denoted in red.*

The presence of high charge states can be advantageous for large analytes: a high mass is converted to a much lower m/z value. This is important since many mass analyzers have a limited m/z range. However, high charge states can also turn into a disadvantage as the spectrum can become overly complicated, thereby creating the need for higher resolution mass analyzers.

How ions are transferred from the aqueous phase to the gas phase during the ESI process is a particularly interesting question. Several models have attempted to explain the ESI process for various analytes.[52] The first of these is the ion evaporation model (IEM, Figure 10a).[53] The IEM is generally assumed to explain the behavior of small molecule analytes. After the charged droplets leaves the Taylor cone, the droplet size keeps decreasing due to solvent evaporation, and the charge density of the droplet increases until the Coulombic

repulsion overcomes surface tension. This situation is referred as Rayleigh limit, which can be described as using the equation[54]

$$z_R = \frac{8\pi}{e}\sqrt{\varepsilon_0 \gamma r^3}$$

Equation 1.3

where z_R is the number of charges; e = 1.60*10^{-19} C; ε_0 is the vacuum permittivity, γ is the surface tension, and r is the radius of the droplet. The IEM envisions that once the droplet reaches the Rayleigh limit, charged analytes with low MW are ejected from the droplet surface. The other ESI model is charged residue model (CRM, Figure 10b).[55,56] This model applies mainly to compact macromolecules (e.g., native proteins). While a (large) analyte is in a droplet near the Rayleigh limit, charges of the droplet are transferred to the analyte as the droplet evaporates to dryness.[57]

Recently, Konermann et al. proposed an additional model[58] that describes how unfolded proteins behave during ESI. Typically, a globular protein has a hydrophobic core and a hydrophilic surface. Once a protein gets denatured (e.g., in formic acid from the mobile phase in UPLC/LC), its hydrophobic residues in the core become solvent accessible as the chain unravels. In this case, and the protein migrates to the surface of the droplet for minimizing the contact with water molecules.[52] Then the unfolded protein gets ejected from the droplet due to Coulombic repulsion.[59,60] This model is referred to as chain ejection model (CEM, Figure 10c).

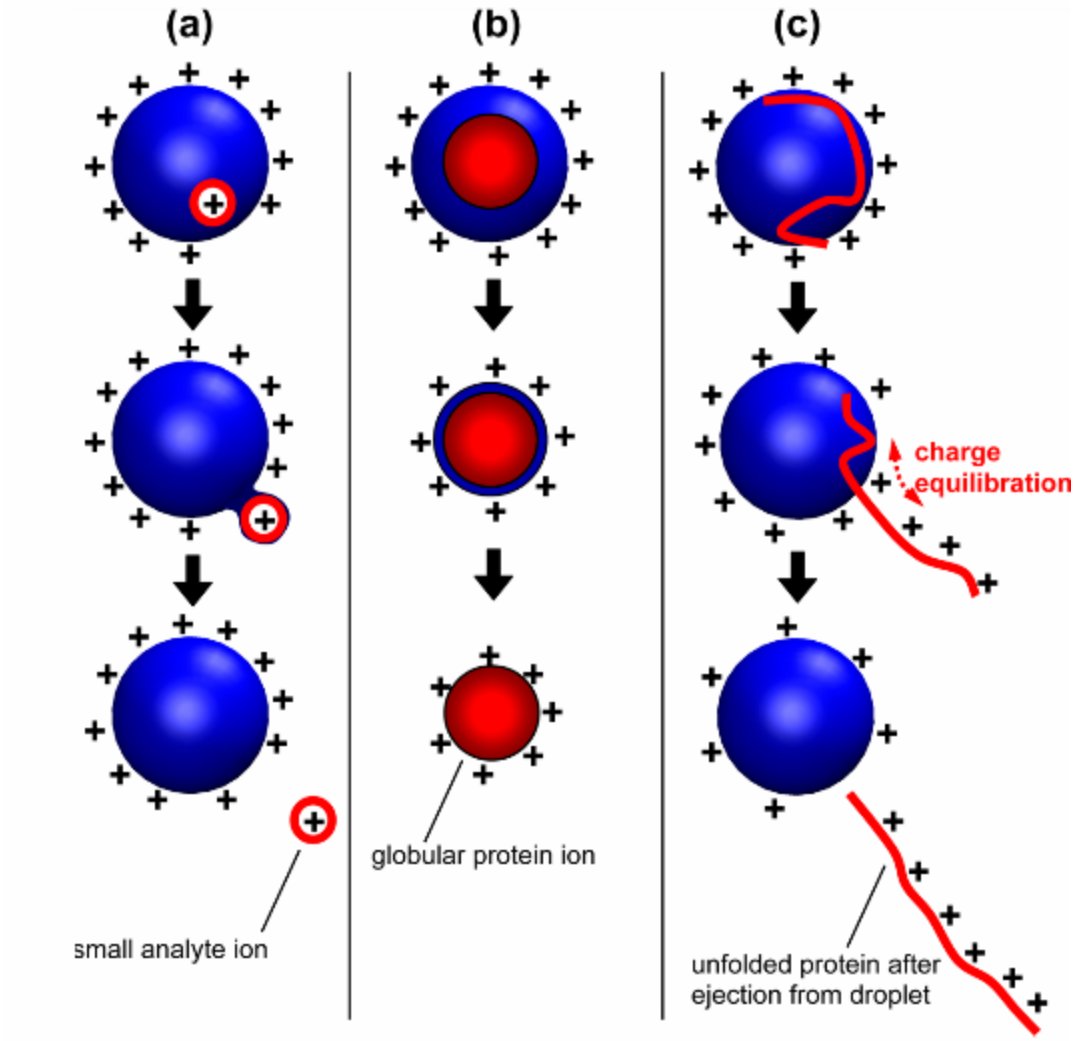


Figure 10. *Summary of three ESI mechanisms (a) IEM: small molecular weight analyte gets ejected from ESI droplet, (b) CRM: a globular folded protein gets released into the gas phase, (c) CEM: an unfolded protein molecule is ejected into the gas phase from the droplet.*[59]

1.3.1.2 Other Ionization Techniques

ESI is vital for biological samples, but it is not the only ionization technique. Besides ESI, many other ionization sources are available: these include electron ionization (EI), chemical ionization (CI), atmospheric pressure chemical ionization (APCI), matrix-

assisted laser desorption ionization (MALDI), desorption electrospray ionization (DESI), paper spray (PS) ionization and several others. All these techniques have their unique strengths and features. EI and CI are considered to be "traditional" ionization techniques that are used mostly for small molecules (a few hundred Da) analytes. Extensive disruption of covalent bonds is the norm with these techniques, which are considered to be quite harsh.[61,62] APCI is similar to CI, but the ionization process happens at atmospheric pressure, which makes it suitable for LC-MS.[63] MALDI is an ionization technique that uses laser pulses to create ions from analyte molecules within an energy-absorbing matrix.[64] MALDI is mostly used for analyzing biomolecules, and it is also important for imaging applications.[65] In DESI, a stream of ESI droplets hits a surface, picks up analytes, and brings them into the gas phase.[66] PS is another variant of ESI. The sample in liquid form (blood, urine, saliva, etc.) is applied to a piece of paper that is kept at high voltage.[67] The spray is formed at the tip of the paper sheet, and ions enter the MS just as in regular ESI. PS probably represents the cheapest high throughput ionization techniques.[68]

1.3.2 Mass Analyzer

After analytes are ionized, a mass analyzer is needed to for detection, ion manipulation, and ion counting such that a mass spectrum can be generated. Several different mass analyzers have been commercialized. Here will briefly discuss quadrupoles and time-of-flight (TOF) instruments.

1.3.2.1 Quadrupoles

Quadrupole mass analyzers are small, simple, and low-cost. They have a reasonable resolution, mass range, and excellent sensitivity. Quadrupoles can also be used for other

purposes: precursor selection, collision cells, and ion guides. In a triple quadrupole mass spectrometer, three quadrupoles are aligned in series. The first (Q1) and third (Q3) ones are regular mass analyzers; the second one (Q2) is employed as a collision cell/ion guide. By setting the Q1 and Q3 into scan mode or setting them to specific m/z values, triple quadrupoles can perform different types of experiments.[69]

Figure 11 shows the schematic principle of a quadrupole. As implied by the name, a quadruple is composed of four rods that are connected to power supplies. All four rods are linked to a combination of direct current (DC) and radio frequency (RF) voltage. By keeping the RF/DC ratio constant and scanning the voltage amplitudes, the whole m/z range can be explored. Alternatively, RF and DC can be set to specific fixed values for transmitting certain m/z values only. A quadrupole can also be set up to RF only mode (DC = 0). In this case, the quadrupole behaves as an ion guide, and all almost all ions pass through.[70]

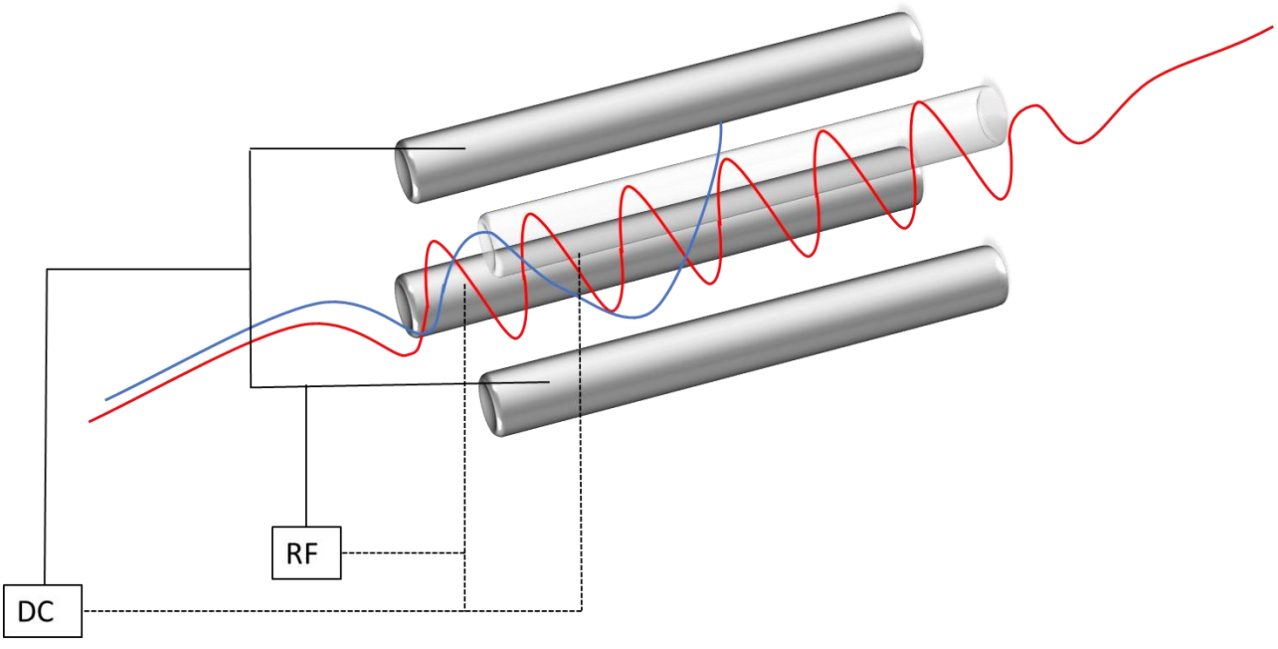


Figure 11. *Schematic of quadrupole operation. The ion path is demonstrated in blue and red. Red line: detected ions with the stable trajectory. Blue line: undetected ion with an unstable trajectory.*

1.3.2.2 Time of Flight Mass Analyzer (TOF-MS)

TOF-MS is another widely used mass analyzer. It is robust and can have high resolution (over 20,000) at a relatively low cost.[71] Figure 12 shows the schematic of an orthogonal acceleration TOF-MS. After the ion beam enters the pusher chamber, ions are accelerated by a pulsed electric field. This pulse gives all ions potential energy as governed by the ionic charge. The potential energy is then transformed into kinetic energy. Thus, ions with different m/z have different velocities. By measuring how long ions travel from pusher to detector, the m/z of the ion can be determined.

The following equation directs the TOF operation:

$$\sqrt{\frac{m}{z}} = \frac{t\sqrt{2eU}}{l}$$

Equation 1.4

where U is the pusher voltage, l is the path length, and t is the time of flight in the instrument. For traditional TOF instruments, the path length is the length of the apparatus (as knowns as "linear TOF").[72] With the application of a reflectron (Figure 12), the resolution is dramatically improved. The reflection provides an electric field that reverses the direction of the ions. The reflectron not only increases the flight path length, but also corrects artifacts such as ions with the same m/z that have inadvertent kinetic energy differences created during acceleration in the pusher. The ions with higher kinetic energy penetrate deeper into the reflectron, such that their flight time becomes slightly longer. In this way the kinetic energy variation is compensated, and the resolution of the mass analyzer is greatly enhanced.

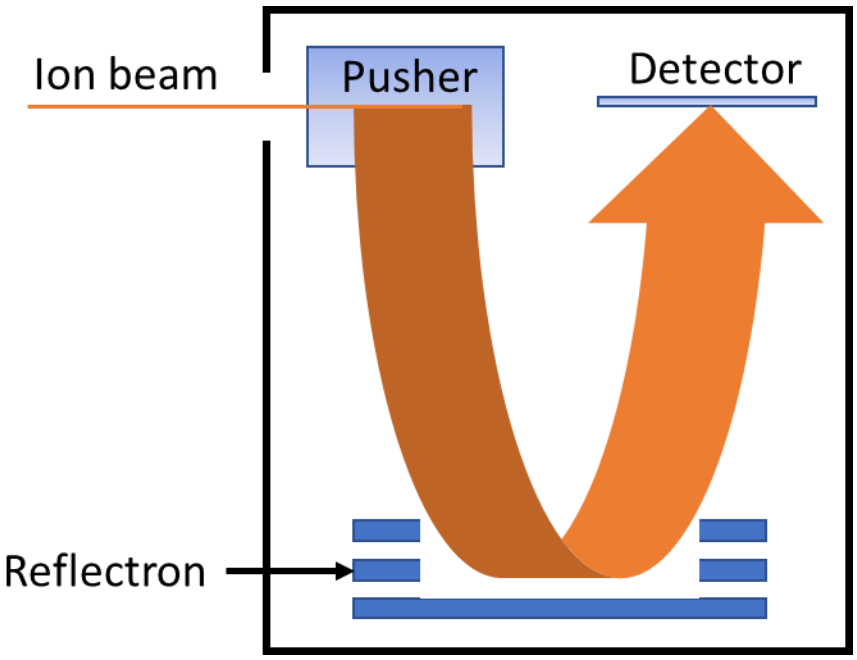


Figure 12. *Schematic layout of a TOF mass analyzer (see text for details).*

1.3.3 MS Approaches to Study Biological Molecules

Various MS techniques are available that can cover multiple areas including protein structure, kinetics, and dynamics. The conceptually simplest approach is known as "native" ESI-MS. It takes advantage of ESI as a soft ion source which can preserve native protein structures and interactions in the gas phase.[50] By analyzing the mass spectrum of protein complexes under native condition, protein-protein/ligand binding stoichiometry can be uncovered.[73] Another technique for characterizing the structure of multi-subunit proteins is chemical cross-linking.[74] In this technique, bifunctional coupling reagents that react with amino acid side chains preform linking reactions. By digesting crosslinked protein and analyzing the tryptic peptides, protein/complex structural models can be built.[75]

Chemical labeling is another technique that gains more and more traction in biological MS. This approach can be divided into selective chemical labeling and nonselective chemical labeling. For the former, chemical tags that react with specific amino acid side chains

(cysteine, lysine, arginine, etc.) are used to label proteins.[76] The labeling pattern is dependent on solvent accessibility of the side chains, and their intrinsic reactivities.[77] Nonselective chemical labeling can universally tag all (or most) of the solvent-exposed amino acids on a protein. The most common nonselective labeling reagent is hydroxyl radical (HO•).[78] With its universality, the mapping of protein solvent accessibility can be achieved in a single experiment.[79]

1.4 Hydrogen-Deuterium Exchange (HDX)

1.4.1 History

The use of hydrogen-deuterium exchange (HDX) can be traced back to Linderstrøm-Lang's work in the 1950s.[80] In their work, many fundamental concepts such as EX1 and EX2 exchange, fast exchange in the amino acid side chain, acid and base catalysis, etc. have been proposed. These concepts have been used to direct the HDX field until today. The detection methods of HDX kept getting developed. In the 1980s, it has already been possible to obtain site-resolved exchanging kinetics through NMR.[81] Over the past 20 years, MS has entered the HDX field. Johnson and Walsh pioneered to use ESI-MS to measure HDX rates.[82] They were the first to use LC-MS to measure the mass of deuterated peptides. This work laid the foundation of bottom-up HDX-MS. The easy operation and high efficiency of HDX-MS quickly made it the method of choice for most practitioners. Software was introduced to automate the data analysis.[83] For improving throughput, the automated sample preparation[84] and integration with UPLC[85] were also applied around ten years ago. All these improvements built HDX-MS to become a powerful tool for protein dynamics study and was widely applied to elucidate various aspects: ligand binding, protein structure exploration, antigen justification, etc. HDX-MS is now well accepted in both academic and industrial laboratories.

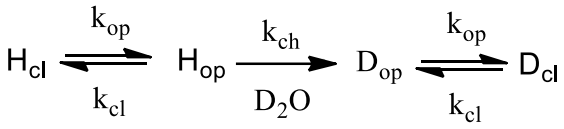
1.4.2 Fundamentals of HDX

In proteins, only hydrogens in O-H, N-H, and S-H are labile and can exchange with water hydrogens. This H-H exchange is not detectable by MS since there is no mass change. In contrast, when D_2O is used as solvent, H-D exchange became measurable by MS since D is 1 Da heavier, and every H→D exchange event will increase the protein mass. This exchange is bi-directional, which means D in the protein can also be replaced by solvent H. There is no fundamental difference between the two exchanges. Like most other studies, we will focus on exchange-in, where D replaces H. Back exchange (or exchange-out) refers to condition where the reverse scenario applies.

In the amide backbone, HDX rates are slower by about two orders of magnitudes compared to those hydrogens in sidechains and protein termini.[86,87] While hydrogens in amino acid side chains reaches exchange equilibrium in seconds, backbone amide hydrogen would take from minutes to hours, in some extreme cases even days, to achieve full equilibrium. Therefore, the exchange of side chain hydrogens is more challenging to measure and observe in MS since in LC-MS deuterium will be rapidly exchanged out when a gradient contained H_2O is used. That makes backbone amide hydrogen the primary information carrier in HDX experiments.

Why can HDX be used to study protein conformations and dynamics? This answer to this question can be ascribed into the activities of H-bonds of proteins in solution. For simplifying the situation, only H-bonds involved in backbone amide H-bonding will be discussed. In well-structured region (e.g., α-helices, β-sheets), backbone amide hydrogen are protected by intra-molecular H-bonds. This protection dramatically reduces the HDX rate constant, k_{HDX}, by as much as 10^{-6}.[88] Exchange of these backbone amide hydrogens requires temporary H-bond breaking ("opening") events that are a manifestation of protein dynamics. Therefore, by measuring HDX rate it is possible to obtain valuable information regarding protein dynamics, which is unavailable through other techniques, such as X-ray crystallography. Simply speaking, highly dynamic regions exchange fast, whereas tightly folded regions show slow HDX.

HDX of backbone amides is usually discussed within the framework of a simple exchange-in reaction diagram:



Equation 1.5

Initially, the amide NH is in a "closed" environment with an intact NH···OC contact. Due to protein thermal fluctuations, such H-bonds will transiently break (or "open") and then subsequently close again. The rate constants of these steps are denoted as k_{op} (opening transition) and k_{cl} (closing transition). While a site is "open", the backbone amide hydrogen is labile (H_{op}) and can be replaced with deuterium from the bulk solution (k_{ch}). After the exchange, the backbone amide deuterium (D_{cl}) can be reformed to new H-bond and be protected again (D_{cl}). In a deuterium rich environment, the whole reaction is pushed to the right. Similarly, back exchange can take place, e.g. during LC separation of deuterated peptides, in a H_2O environment encountered during the downstream analysis process.

From equation 1.5, both H_{cl} and H_{op} can be grouped into $[H_{unexchanged}]$ which represents all the unexchanged backbone amide hydrogens in the system.

$$H_{unexchanged} = H_{cl} + H_{op}$$

Equation 1.6

The population changing of $[H_{unexchanged}]$ can be described as:

$$\frac{d[H_{unexchanged}]}{dt} = -k_{ch}[H_{op}]$$

Equation 1.7

The population change of H_{op} can be described as

$$\frac{d[H_{op}]}{dt} = -(k_{cl} + k_{ch})\left[H_{op}\right] + k_{op}[H_{cl}]$$

Equation 1.8

During the HDX continuous labeling process, it will always be a period that the generation and depletion of H_{op} are the same and this leaves $d[H_{op}]/dt$ (equation 1.8) equal to near 0. Under these steady-state conditions equation 1.8 can be rearranged to

$$(k_{cl} + k_{ch})\left[H_{op}\right] = k_{op}[H_{cl}]$$

Equation 1.9

By substituting equation 1.6 into equation 1.9 we find

$$(k_{cl} + k_{ch})\left[H_{op}\right] = k_{op}(H_{unexchanged} - H_{op})$$

Equation 1.10

$$\left[H_{op}\right] = \frac{k_{op}\left[H_{unexchanged}\right]}{k_{cl} + k_{ch} + k_{op}}$$

Equation 1.11

By combining equation 1.7 and equation 1.11, it follows that

$$\frac{d[H_{unexchanged}]}{dt} = -\frac{k_{ch}k_{op}}{k_{cl} + k_{ch} + k_{op}}\left[H_{unexchanged}\right]$$

Equation 1.12

Here we have a new equation that is very close to first order reaction ($d[A]/dt = -k[A]$). Then, it can be further simplified to

$$\frac{d[H_{unexchanged}]}{dt} = -k_{HDX}[H_{unexchanged}]$$

Equation 1.13

where

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

Equation 1.14

k_{HDX} is the rate constant that for backbone amide hydrogen replaced by solvent deuterium. In most experimental condition, the protein is more likely to stay folded (closed conformation) and k_{cl} is much larger than k_{op}. Therefore, equation 1.14 can be simplified to:

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

Equation 1.15

The relative magnitude of k_{cl} and k_{ch} establishes two limiting HDX scenarios. If $k_{ch} >> k_{cl}$, the site will undergo complete HDX after the first opening event. These conditions are commonly seen under mildly denaturing conditions. If these slow fluctuations take place cooperatively (i.e., if they affect all or many sitest in parallel) two populations can be observed: The corresponding mass spectra will show a low mass and a high mass contribution (Figure 13).[89] As time proceeds, the low mass population (unlabeled protein) keeps decreasing while the high mass population increases. This scenario is referred to as EX1, and the k_{HDX} equation simplifies to

$$k_{HDX} = k_{op}$$

Equation 1.16

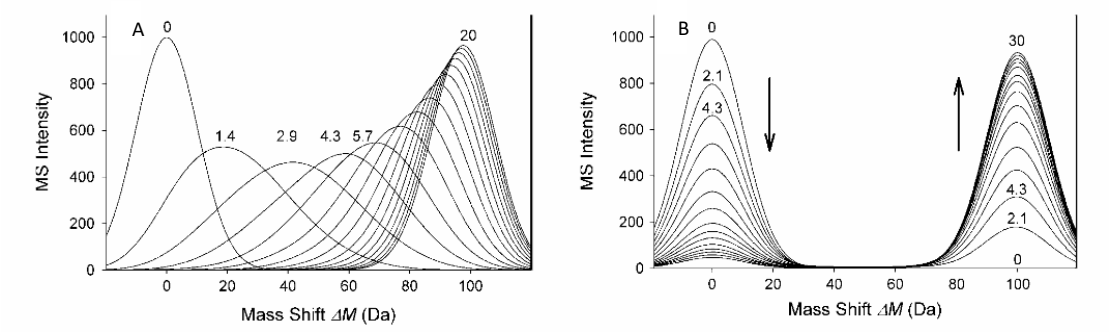


Figure 13. *Simulated HDX kinetics under EX2 (A) and EX1 (B) conditions. The numbers on the top of each peak denote time points in second. It is rare to see these pure scenarios. In reality, EX1 and EX2 are often superimposed, i.e. the protein undergoes fluctuations on many different time scales that affect different regions.*[89]

In contrast to EX1, EX2 scenario happens when $k_{ch} << k_{cl}$ and this is often seen in native environments, which are typical in protein/protein-ligand studies. Under these conditions many opening events are required before HDX takes place at any given sites. In mass spectrum, EX2 conditions give rise to a single mass distribution that gradually shifts to higher m/z as time increases. (Figure 13) Similar to equation 1.11 in EX1, the k_{HDX} equation can be rewritten to

$$k_{HDX} = (k_{op}/k_{cl})k_{ch}$$

Equation 1.17

Equation 1.12 can be further simplified to

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

Equation 1.18

The hydrogen exchange on peptide backbone amide hydrogen can be catalyzed by H^+, OH^- and water. Each catalyst has its unique intrinsic rate as in

$$k_{ch} = k_{acid}[H^+] + k_{base}[OH^-] + k_{H_2O}[H_2O]$$

Equation 1.19

Thus, k_{ch} is highly pH dependent. A plot of log(k_{ch}) vs. pH is V-shaped (Figure 14), and the lowest point is around pH 2.5 – 3. This pH range is used as for HDX quenching. At pH 3, the exchange halftime has been increased to ~20 minutes. This provides a chance to analyze HDX by LC-MS (a typical gradient is around 10~20 minutes), but is still too fast for many other experimental methods.

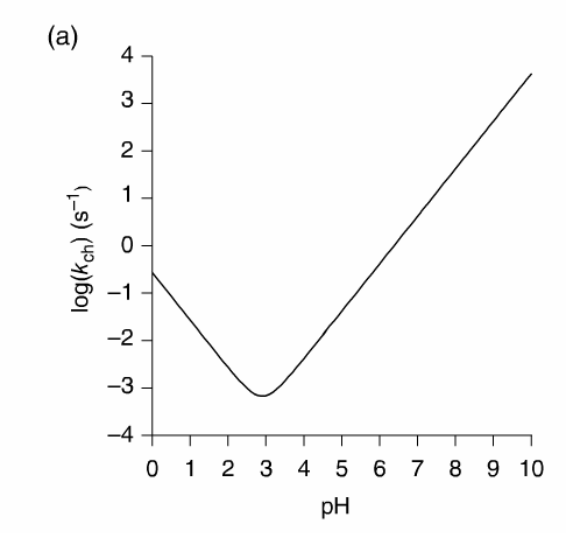


Figure 14. *k_{ch} of a random coiled poly-alanine chain at 20 °C under various pH.*[90]

Luckily, HDX is also affected by temperature as governed by Arrhenius equation

$$k_{ch} = Ae^{-\frac{E_a}{RT}}$$

Equation 1.19

where E_a is the activation energy, T is the temperature, A is an empirical constant, and R is the gas constant (8.314 J mol^{-1} K^{-1}). As temperature drops, k_{ch} decreases exponentially. k_{ch} can be lowered ~ 10 times by lowering the temperature from 25 to 0°C. Therefore, HDX quenching takes place at ~ 0°C *and* at pH 2.5 - 3.

1.4.3 HDX Experiment Design

For HDX experiments protein samples are diluted into D_2O buffer; then aliquots are removed and quenched at selected labeling times by lowing the pH to ~3 and cooling to 0°C (or flash frozen in liquid nitrogen). Afterward, the deuterated protein sample can be directly injected into the mass spectrometer, and the mass changes can be observed at the intact protein level.[91] More detailed information can be derived when incorporating enzymatic digestion is incorporated into the workflow. This step requires a specific enzyme that is active under quenching condition. Although other types of enzymes are available[92], porcine pepsin remains to be the gold standard due to its robust nature, low price and well-established protocols.[93] The activity of pepsin is highest at pH 2-4, which matches well with the HDX quenching condition. The simplest approach is to mix target protein and pepsin in a 1:1 mass ratio for a few minutes. More sophisticated workflows involve pepsin immobilization on a column for on-line UPLC experiments that provide highly consistent data.[94] (Figure 15). The analysis processes are separated into three successive steps. 1. Sample loading: the deuterated sample is injected into the sample loop. 2. Trapping: By valve switching, auxiliary solvent manager (ASM) pumps acidified water (pH 2.5, 0.1% formic acid in water) through the sample loop, then carries sample into pepsin column where the protein is digested. The resulting peptides are trapped in a short (2 cm) C18 reverse phase trapping column. 3. Injection: binary solvent manager (BSM) pumps an acetonitrile/water gradient through the trap column. The trapped peptides are eluted and further separated by a downstream analytical C18 column. Finally, HDX data are collected by MS.

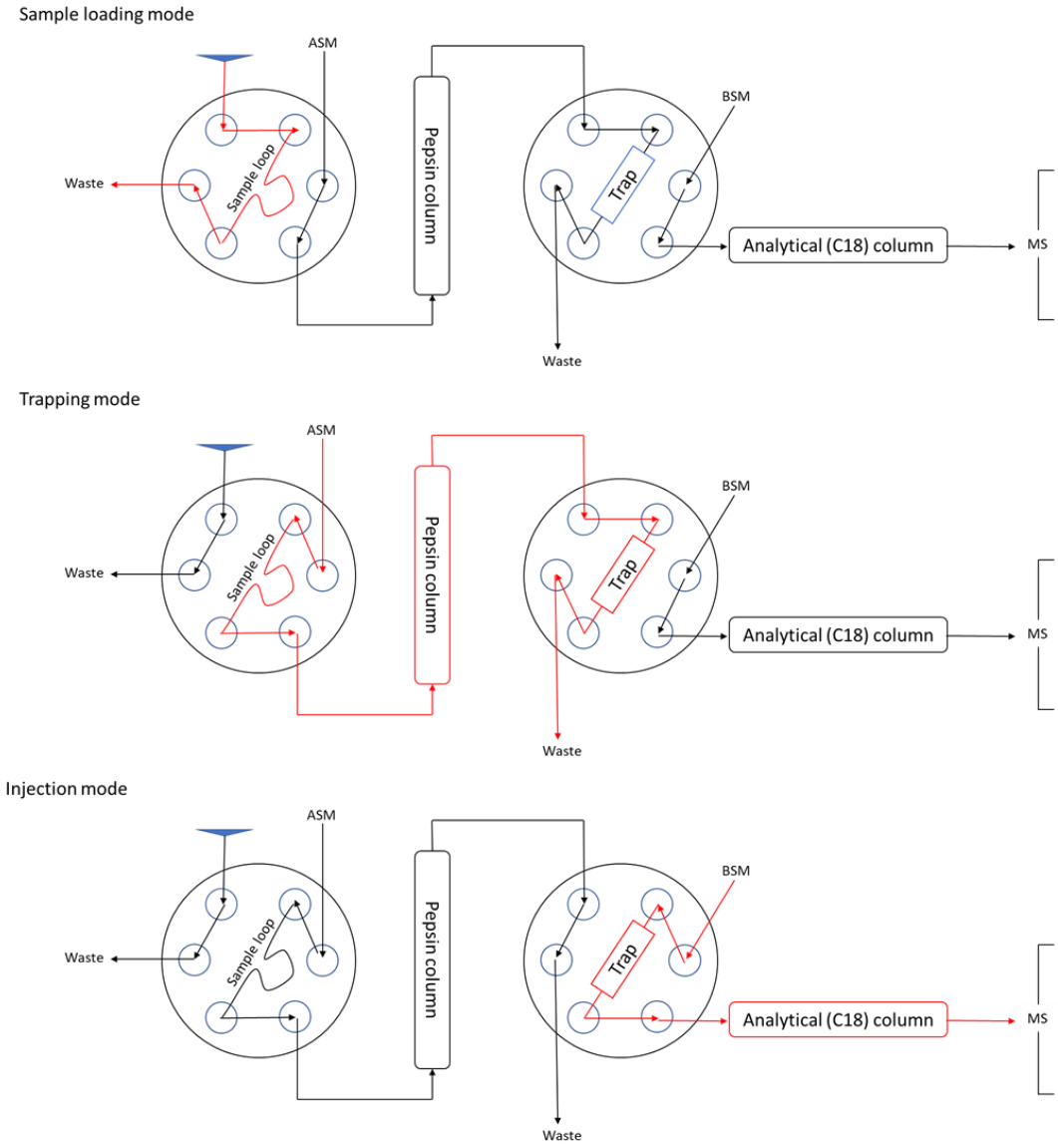


Figure 15. *Schematic representation of HDX online digestion module. Lines (red and black) shows the solvent flow path. Red lines indicate where deuterated samples are located during specific steps.*

The modification of HDX instruments does not end here. For increasing digestion efficiencies, high pressure on the pepsin column has been proven to be effective.[95] The integration of an electrochemical module into an HDX online digestion system was reported which is capable of reducing disulfide bonds for better digestion coverage.[96] The full automation of HDX (from sample production to sample injection) makes high throughput studies possible and can enhance the efficiency of epitope mapping, drug

screening, etc.[84]

Several designs of HDX-MS experiments have been proposed for various purposes. These methods can be categorized into continuous-labeling and pulsed labeling. As implied by the name, continuous labeling means that a protein is incubated in labeling buffer and deuterium "continuously" exchanges into the protein. The mass shift is monitored as a function of time. Time points cover the range from seconds to hours, sometimes even to days. One of the primary usages of continuous labeling is to compare the difference between protein states. By interpreting the differences between HDX rates in spatially-resolved experiments, it may be possible to identify ligand binding sites, allosteric effect induced by ligand binding, dynamic changes upon binding, etc. Continuous-labeling is the most commonly used HDX-MS strategy in all of them.[97]

Another approach is pulsed labeling HDX. In pulsed labeling the protein is often studied under non-equilibrium conditions, e.g., during protein folding. Hence one key application of pulsed HDX is the detection and characterization of folding intermediates.[98] One typical experiment design is to transfer denatured proteins into a folding environment quickly. Subsequently, while the protein folds, it is exposed to an HDX pulse (in millisecond scale) and quenched subsequently with acidification. By controlling the refolding time, transient intermediates can be identified on the basis of their unique deuterium uptake properties. It is advantageous to design the HDX pulse such that it results in an all-or-nothing scenario: e.g., at pD 9, unprotected amides are nearly fully exchanged while marginally protected amides (partially folded) only have less than 5% deuteration levels.[88] Although finely tuning is still required, pulsed labeling HDX remains as a powerful tool in the protein-folding study arsenal.

1.4.4 Peptide Mapping

Peptide mapping is usually the first step in bottom-up HDX. The traditional peptide mapping method is data-dependent acquisition (DDA). During the LC-MS run, MS1

survey scans identify the most intense precursor ions and apply MS/MS to identify the peptide. This process may have to be repeated several times to identify all ions. Plus, optimization of collision energies is needed,[99] which can make the process quite time-consuming. Also, some low abundance peptide may be missed since MS1 survey scans only pick up ions that stronger than a certain threshold.

The development of data independent acquisition (DIA) has tremendously facilitated the tedious workflow. The central idea in DIA is *all* ions are fragmented without precursor selection during the LC-MS process. The collision cell in the MS is switched back and forth between low and high energy. Therefore, precursors and their fragments are measured at the same time. The signal intensity and retention time between each precursor and corresponding fragments must be well correlated. Thus, the precursors (peptide) can be identified from its paired fragments with the help of computer software/algorithms.[100] Therefore, the peptide mapping processing is greatly simplified and can sometimes be performed in a single run.

1.4.5 Data Analysis in HDX-MS

Once HDX data are obtained, the question becomes how to extract the protein dynamic behavior from the measured isotope distributions. Here we are going to take bottom-up HDX-MS as an example to explain the general method of data analysis.

For a typical HDX experiment, it is impossible to completely avoid back exchange, especially when on-line digestion and HPLC/UPLC separation are applied. For correcting the back exchange, two types of control samples need to be prepared: 1st, m_0: a protein sample that is exposed to deuterated buffer with minimum time under quenched condition. 2nd, m_{100}: a protein sample that has attained maximum deuteration. The corrected deuteration percentage (D_t) at each specific time point *t* can then be calculated from the following equation:[101]

$$D_t = \frac{m_t - m_0}{m_{100} - m_0} \times 100\%$$

Equation 1.20

where m_t is the centroid mass of deuterated peptide at the particular time point. An HDX uptake curve then can be plotted as D_t vs. time.

In principle, k_{HDX} (discussed in section 1.4.2), the rate constant of HDX, can be acquired by fitting the HDX uptake curve. However, each amide has its own rate constant which generally results in multi-exponential kinetics. For simplifying the analysis, the amide hydrogens are often grouped into three categories: fast, intermediate and slow. A suitable fitting equation under such conditions is shown as the following:[101]

$$D_t = N_1\left(1 - e^{-k_1 t}\right) + N_2\left(1 - e^{-k_2 t}\right) + N_3(1 - e^{-k_3 t})$$

Equation 1.21

where N_1 and N_2 and N_3 are the numbers of fast, intermediate and slow exchanging amide hydrogens, respectively; and k_1, k_2, and k_3 are the corresponding HDX rate constants. Since k_3 usually is very close to 0, that brings the 3rd term in the equation 1.24 also near 0. Therefore, the equation is often further simplified to

$$D_t = N_1\left(1 - e^{-k_1 t}\right) + N_2\left(1 - e^{-k_2 t}\right)$$

Equation 1.22

This two components equation usually works reasonable well with most EX2 HDX-MS data. For EX1 data, different population need to be separated first before fitting with equation 1.24 or 1.25.

1.4.6 Applications of HDX

The application of HDX covers various fields. One of the major aspects is to investigate protein-ligand interactions. The interaction between protein and ligand is the trigger of plenty of biological activities which are important for enzymatic catalysis, the function of the immune system, transmembrane ion transport, etc. The protein-ligand interactions are usually non-covalent and reversible, as expressed in

$$P + L \rightleftharpoons PL$$

Equation 1.23

where P is the protein; L is the ligand; PL is the protein-ligand complex. The dissociation constant (K_d) equals to [P][L]/[PL]. In most case, binding of a ligand (e.g., a drug molecule) decreases the deuteration rate around the ligand binding region, which implies the ligand-protein complex are more folded/protected.[19,102] However, in some scenarios, ligand binding can increase the hydrogen-deuterium rate. The oxygenation of hemoglobin is such a case: the HDX rate is increased at the heterodimers interface while oxygen binds.[103]

The discovery of small molecule therapeutics can also benefit from using HDX to monitor protein-ligand interaction. On G protein-coupled receptors (GPCRs), HDX-MS has emerged to be one important tool to explore its conformational and allosteric changes upon ligand/drug binding. In one study, HDX was used to identify ligand-binding loops on the β2-adrenergic receptor, that where not resolved in crystal structures.[104] HDX can also be used to direct the development of therapeutic antibodies by mapping epitopes upon protein complexes.[105,106] The application of HDX-MS can be further used to test protein therapeutics. One example involves insulin analogs that were probed by HDX for their pharmacokinetics and stability.[107]

1.5 Molecular Dynamics (MD) Simulations

As previously discussed, proteins are in constant movement, i.e., they undergo thermal motions that are coupled to the dynamics of the surrounding solvent.[108] HDX methods represent an experimental tool to probe these protein dynamics. Computer simulations (in particular, MD methods) are a valuable tool to directly visualize protein movements and conformational changes from an atomistic perspective. One significant advantage of MD simulation is that it can reveal movements of molecules between the start and the end of a process/transformation. This is a vital complement to many other experimental methods since most time only the start and the end of an experiment are observed and analyzed.

1.5.1 Background of MD Simulations

In a perfect world with unlimited resource and time, the behavior and trajectories of all molecules could be predicted *ab initio* by the time-dependent Schrödinger equation. However, realistically, one has to make approximations to be able to obtain simulation results in reasonable time. MD simulations represent a computational technique that has a relatively low computational cost. MD uses Newton's law to calculate the motion of atoms:

$$F = ma$$

Equation 1.24

The forces applied to the atoms are provided by the potential energy calculated from force fields. Force fields are an assembly of empirical parameters that define interactions between atoms.[109,110] These parameters can be classed into two groups: bonded and non-bonded interactions. Bonded interactions describe the behavior of covalent bonds, including stretching and bending, torsion potentials of bond rotation, etc. Nonbonded interactions are described as Lennard-Jones and Coulombic electrostatic potentials. Nonbonded interactions are calculated between any pair of atoms. If two atoms are

separated by a large distance, these nonbonded interactions become rather weak. For reducing computational cost, a cut-off value is commonly used for Lennard-Jones potential since it decays very fast. The case is more complicated for Coulombic interactions especially when periodic boundary conditions (see below) are applied. This is because the Coulomb potential decays much slower than the Lennard-Jone potential at long range. A simple Coulomb cut-off would induce severe atifacts.[111] Therefore, the Partical-Mesh-Ewald (PME) algorithm is used to calculate Coulombic interactions beyond the cut-off distance at relatively low computational cost[112], allowing better efficiency and accuracy in large systems, such as proteins in solution.

Periodic boundary conditions (PBC) allow simulating quasi-infinite systems, thereby eliminating surface artifact that would otherwise be encountered in bulk solution studies.[112] The idea of PBCs is that the actual simulation box is surrounded with identical copies of itself in all directions. If an object leaves the box on one side, it will re-enter from the other side with the same speed and direction (Figure 16). In this way, an extensive, homogeneous system can be simulated with no boundaries in MD.

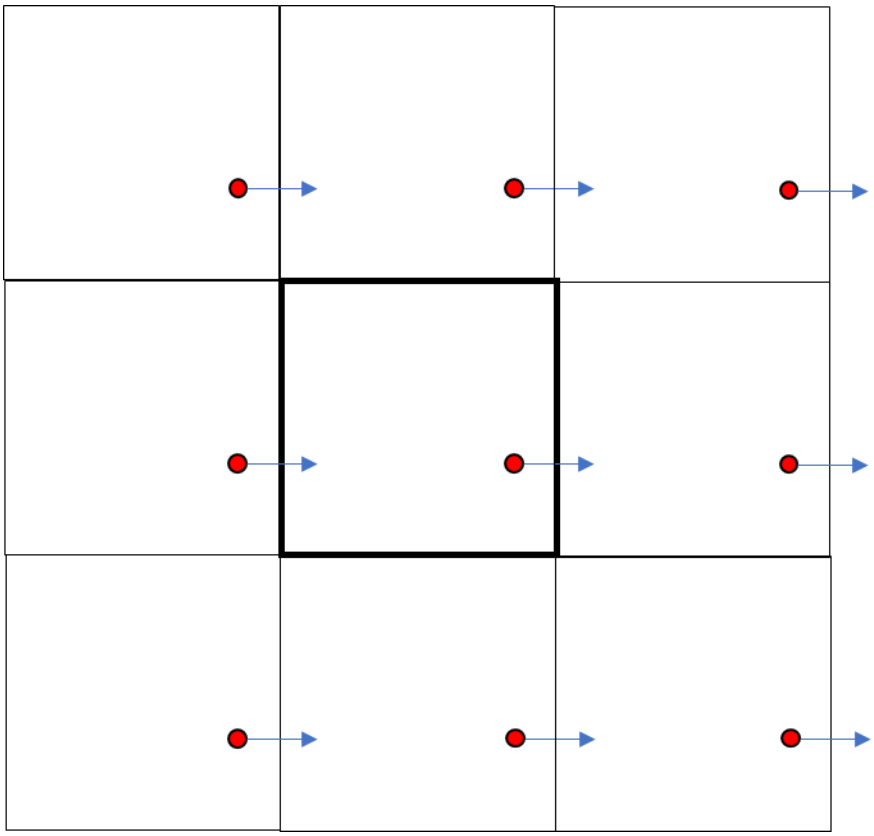


Figure 16 *Schematic of periodic boundary conditions (PBC)*

MD is an excellent tool for many applications, but it cannot simulate the formation and breaking of covalent bonds. That precludes the application of MD techniques to simulating chemical reactions. Despite this limitation, MD techniques allow simulations on many interesting bio-macromolecules on nanosecond to millisecond time scales, which is nearly impossible for quantum or DFT methods.

A number of MD packages are available. Some have been commercialized, such as Abalone II; but there are also quite a few free/open source MD platforms. Gromacs is one of them. It is versatile, easy to use, and can be modified to suit the user's needs. In the rest of section 1.5, we will take Gromacs as an example to discuss the steps of a typical MD simulation.

1.5.2 Steps to Perform a MD Simulation

The first step of a MD simulation is to introduce the starting structure into the system. A static "picture" of the molecule which describes all atoms and their interactions is needed for this step; the "picture" is named as *topology*. In Gromacs, the *topology* file is generated by the program *pdb2gmx*. *pdb2gmx* reads data from protein's pdb file and adds all the missing hydrogen atoms. *pdb2gmx* allows the user to select the forcefield, water model, protonation sites, etc. Once finished, *pdb2gmx* puts the molecule in a simulation box and the box with user-defined interactions.

After protein coordinates are introduced into the MD simulation, water and ions can be added into the box. This allows the protein to be simulated in a solvated environment and the input of ions makes the environment closer to the physiological state. These newly added molecules may be placed into unrealistic positions and could generate large forces that might crash the system. For this reason, it is necessary to perform an energy minimization procedure to "relax" the system. Energy minimization allows all atoms to move by a small distance in the direction that lowers the potential energy.

In the energy minimization step, solvent (water and ions) is optimized with itself, but not necessarily with the protein. Therefore, further action is needed to equilibrate the solvent around the protein. The first phase of the equilibration step is conducted under constant temperature, volume and number of particles, also known as NVT. The initial velocities of all atoms can be assigned randomly from a Maxwell-Boltzmann distribution at the desired temperature. NVT equilibration stabilizes the system around the required temperature. After that, the pressure/density of the system can be equilibrated through an NPT step, which has the constant temperature, pressure, and the number of particles. After these equilibration steps are finished, the system is ready for the actual simulation run.

The MD simulation runs in time steps: the system will move with the assigned velocity without disturbance for one time step (typically 2 fs). The new coordinates of particles are recorded in the trajectory files. Potential energy and velocities of particles are then

recalculated. With the new velocities and coordinates, the system runs for another time step and repeats until the whole simulation finishes.

1.5.3 Trajectory Analysis

Once the MD simulation finishes, the trajectory files are ready to be analyzed. The most straightforward tool is to compare the protein structure before and after the MD simulation. The standard way to do this is to calculate the root mean square deviations (RMSD) between protein structures. RMSD reports by how much a structure deviate from a reference structure. The equation is

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} D_i^2}$$

Equation 1.25

where D_i is the distance between atom i and its reference position; N is the number of atoms. The higher the RMSD, the larger the difference between structures.

It is also interesting to track the protein's dynamics during the simulation. In MD simulation the fluctuation of atoms or amino acid residues can be measured by root mean square fluctuation (RMSF). RMSF is very similar to RMSD except the position deviation is averaged over time. Also, RMSD is generally referred to entire or segment of the protein; RMSF is more specific for single residue or atom. The Equation of RMSF is shown as the following

$$RMSF = \sqrt{\frac{1}{t}\sum_{i=1}^{t} d_i^2}$$

Equation 1.26

where t is the time, d is the distance between atom i and its reference at the different time point. For protein, normally only RMSF of backbones are calculated since side chains are way more fluctuating even in tightly packed secondary structure.

Since all the atoms/molecules coordinates of all the time points are stored in the trajectory file, a movie of molecular movement can be made to present the protein conformational change or coordinating with its ligand.[113] This makes data presented more attractive and straightforward. The analysis of MD data is unrestricted. From MD data, one can analyze the center of mass, radius of gyration, distance/bond/dihedral angles between two particles, H-bond formation to explore the changes in the system.

Recently, with the help of state of the art supercomputers, the time scale of MD has been pushed into the ms scale, which brings the MD simulation closer to experimental conditions.[114] MD has contributed tremendously to various fields including protein-folding[115], surface chemistry[116], instrument development[117], and will continue to expand into new application areas.

1.6 Scope of Thesis

HDX-MS is an ideal tool to study dynamic features of proteins. HDX rates reflects whether the local region is protected and folded. MD simulation is another technique that can explore protein dynamics. HDX-MS and MD simulations complement each other in time scales and resolution: HDX visualizes opening/closing reactions that take place within microseconds to minutes, and information is obtained at the level of peptides; MD simulations probe the behavior of proteins on a ns to ms time scale, and information is obtained for each atom in the system.

The objective of this work is to investigate protein dynamics of three different systems. Different aspects of proteins are explored: protein folding, protein-ligand interactions, and

allosteric effects. While these efforts primarily provide very fundamental information, they could ultimately help the biotechnological industry, and assist in the development of new pharmaceuticals.

Chapter 2 was inspired by previous work done by Hedges[118] to explore how proteins interact with gas/liquid interfaces (in the form of bubbles). While the work of Hedges focused on proteins in ESI droplets, our aim is to explore the behavior in bulk solution, e.g., in a bioreactor. This Chapter used HDX-MS to probe the protein dynamics of a model protein (myoglobin) in the presence of N2 bubbles. The data suggest that proteins on the interface adopt a globally unfolded, long-lived state that can proceed in two directions: refolding to the native state or aggregation. A simple computational model was devised to quantitatively reproduce the experimental kinetics.

Chapter 3 continues to use HDX-MS to explore protein dynamic properties upon ligand binding. The OPG/RANKL/HS system is crucial for regulating bone formation. HDX-MS revealed HS-binding site on OPG; also, RANKL induced stabilization of OPG was observed. A schematic mechanism is proposed based on the HDX data to explain how the HS-induced stabilization effect could promote the binding affinity between OPG and RANKL.

Finally, Chapter 4 combines MD simulations and HDX-MS to investigate allosteric effects in a protein-ligand system. The S100A11-Annexin A1 complex is studied in this project. Both HDX-MS and MD data shows the loss of Ca^{2+} made the binding site more dynamics. However, our data suggest the enhanced dynamics is not the actual trigger of the allosteric effects induced by Ca^{2+} binding. Instead, Ca^{2+} appears to block the signals that emanate from a remote agitator. This blocking mechanism challenges existing paradigms related to the mechanism of allosteric regulation in proteins.

1.7 References

(1) Brannigan, J. A.; Wilkinson, A. J., *Nat. Rev. Mol. Cell Biol.* **2002,** *3* (12), 964-70.

(2) Shea, J. E.; Brooks, C. L., 3rd, *Annu. Rev. Phys. Chem.* **2001,** *52*, 499-535.

(3) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K., *Nat. Biotechnol.* **2007,** *25* (2), 197-206.

(4) Okuno, D.; Iino, R.; Noji, H., *J. Biochem* **2011,** *149* (6), 655-664.

(5) Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R., *J. Mol. Biol.* **1984,** *175* (2), 159-174.

(6) Ma, W.; Tang, C.; Lai, L., *Biophys. J.* **2005,** *89* (2), 1183-93.

(7) Levinthal, C., *J. Chim. Phys.* **1968,** *65*, 44-45.

(8) Naganathan, A. N.; Munoz, V., *J. Am. Chem. Soc.* **2005,** *127* (2), 480-1.

(9) Dill, K. A.; Chan, H. S., *Nat. Struct. Biol.* **1997,** *4* (1), 10-9.

(10) Dempsey, A. C.; Walsh, M. P.; Shaw, G. S., *Structure* **2003,** *11* (7), 887-97.

(11) Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D., *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82* (15), 5000-4.

(12) Henzler-Wildman, K.; Kern, D., *Nature* **2007,** *450* (7172), 964-72.

(13) Kerns, S. J.; Agafonov, R. V.; Cho, Y. J.; Pontiggia, F.; Otten, R.; Pachov, D. V.; Kutter, S.; Phung, L. A.; Murphy, P. N.; Thai, V.; Alber, T.; Hagan, M. F.; Kern, D., *Nat. Struct. Mol. Biol.* **2015,** *22* (2), 124-31.

(14) Kovermann, M.; Aden, J.; Grundstrom, C.; Sauer-Eriksson, A. E.; Sauer, U. H.; Wolf-Watz, M., *Nat. Commun.* **2015,** *6* (1), 7644.

(15) Vahidi, S.; Bi, Y.; Dunn, S. D.; Konermann, L., *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (9), 2412-7.

(16) Liuni, P.; Jeganathan, A.; Wilson, D. J., *Angew. Chem. Int. Ed. Engl.* **2012,** *51* (38), 9666-9.

(17) Deng, B.; Lento, C.; Wilson, D. J., *Anal. Chim. Acta* **2016,** *940*, 8-20.

(18) Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R., *Drug Discov. Today* **2014,** *19* (1), 95-102.

(19) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R., *Expert Rev. Proteomics* **2011,** *8* (1), 43-59.

(20) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A., *J. Med. Chem.* **2004,** *47* (8), 1879-81.

(21) Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M., *J. Med. Chem.* **2008,** *51* (18), 5843-55.

(22) Kern, D.; Eisenmesser, E. Z.; Wolf-Watz, M., *Methods Enzymol.* **2005,** *394*, 507-24.

(23) Bai, Y., *Chem. Rev.* **2006,** *106* (5), 1757-68.

(24) Gill, S. C.; von Hippel, P. H., *Anal. Biochem.* **1989,** *182* (2), 319-26.

(25) Wang, J.; Heyes, C. D.; El-Sayed, M. A., *J. Phys. Chem. B.* **2002,** *106* (3), 723-729.

(26) Pan, Y.; Brown, L.; Konermann, L., *J. Mol. Biol.* **2009,** *394* (5), 968-81.

(27) Barrow, C. J.; Yasuda, A.; Kenny, P. T. M.; Zagorski, M. G., *J. Mol. Biol.* **1992,** *225* (4), 1075-1093.

(28) Greenfield, N.; Fasman, G. D., *Biochemistry* **1969,** *8* (10), 4108-16.

(29) Curtis, J. W., *Proteins: Struct., Funct., Bioinf.* **1990,** *7* (3), 205-214.

(30) Wagner, G.; Wüthrich, K., *J. Mol. Biol.* **1982,** *160* (2), 343-361.

(31) Adler, A. J.; Greenfield, N. J.; Fasman, G. D., Academic Press: 1973; Vol. 27, pp 675-735.

(32) Podjarny, A.; Howard, E.; Mitschler, A.; Chevrier, B.; Lecomte, C.; Guillot, B.; Pichon-Pesme, V.; Jelsch, C., *Europhysics News* **2002,** *33* (4), 113-117.

(33) Burley, S. K.; Berman, H. M.; Christie, C.; Duarte, J. M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C., *Protein Sci.* **2018,** *27* (1), 316-330.

(34) Mozzarelli, A.; Rossi, G. L., *Annu. Rev. Biophys. Biomol. Struct.* **1996,** *25* (1), 343-65.

(35) Read, R. J., *Structure* **1996,** *4* (1), 11-14.

(36) McPherson, A., *Methods* **2004,** *34* (3), 254-65.

(37) Bard, J.; Ercolani, K.; Svenson, K.; Olland, A.; Somers, W., *Methods* **2004,** *34* (3), 329-47.

(38) Kilby, P. M.; Van Eldik, L. J.; Roberts, G. C., *Protein Sci.* **1997,** *6* (12), 2494-503.

(39) Dyson, H. J.; Wright, P. E., *Nat. Rev. Mol. Cell Biol.* **2005,** *6*, 197.

(40) Sprangers, R.; Kay, L. E., *Nature* **2007,** *445* (7128), 618-22.

(41) Gelis, I.; Bonvin, A. M.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G., *Cell* **2007,** *131* (4), 756-69.

(42) Frueh, D. P.; Goodrich, A. C.; Mishra, S. H.; Nichols, S. R., *Curr. Opin. Struct. Biol.* **2013,** *23* (5), 734-9.

(43) Ruska, H., *Naturwissenschaften* **1941,** *29* (24), 367-368.

(44) Ruska, H.; Borries, B.; Ruska, E., *Arch. Gesamte Virusforsch.* **1939,** *1* (1), 155-169.

(45) Elmlund, D.; Le, S. N.; Elmlund, H., *Curr. Opin. Struct. Biol.* **2017,** *46*, 1-6.

(46) Bonnington, L.; Lindner, I.; Gilles, U.; Kailich, T.; Reusch, D.; Bulau, P., *Anal. Chem.* **2017,** *89* (16), 8233-8237.

(47) Zhu, J.; Vinothkumar, K. R.; Hirst, J., *Nature* **2016,** *536* (7616), 354-358.

(48) Donnan, F. G., *J. Soc. Chem. Ind., London* **1923,** *42* (36), 861-861.

(49) Herniman, J. M.; Langley, G. J.; Bristow, T. W. T.; O'Connor, G., *J. Am. Soc. Mass Spectrom.* **2005,** *16* (7), 1100-1108.

(50) Gault, J.; Donlan, J. A.; Liko, I.; Hopper, J. T.; Gupta, K.; Housden, N. G.; Struwe, W. B.; Marty, M. T.; Mize, T.; Bechara, C.; Zhu, Y.; Wu, B.; Kleanthous, C.; Belov, M.; Damoc, E.; Makarov, A.; Robinson, C. V., *Nat. Methods* **2016,** *13* (4), 333-6.

(51) Marcoux, J.; Robinson, C. V., *Structure* **2013,** *21* (9), 1541-50.

(52) Berkel, G. J.; Kertesz, V., *Anal. Chem.* **2007,** *79* (15), 5510–5520.

(53) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., *Anal. Chem.* **2013,** *85* (1), 2-9.

(54) Walter, R.; Ballardin, A.; Schwartz, I. L.; Gibbons, W. A.; Wyssbrod, H. R., *Proc. Natl. Acad. Sci. U.S.A.* **1974,** *71* (11), 4528-4532.

(55) Rayleigh, L., *Lond. Edinb. Dubl. Phil. Mag.* **1882,** *14* (87), 184-186.

(56) Dole, M.; Hines, R. L.; Mack, L. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., *Macromolecules* **1968,** *1* (1), 96-97.

(57) De, J. F., *Anal. Chim. Acta* **2000,** *406* (1), 93-104.

(58) Hogan, C. J., Jr.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L., *Anal. Chem.* **2009,** *81* (1), 369-77.

(59) Konermann, L.; Rodriguez, A. D.; Liu, J., *Anal. Chem.* **2012,** *84* (15), 6798-804.

(60) Ahadi, E.; Konermann, L., *J. Phys. Chem. B* **2012,** *116* (1), 104-12.

(61) Becker, K. H.; Tarnovsky, V., *Plasma Sources Sci. Technol.* **1995,** *4* (2), 307.

(62) Field, F. H., *Acc. Chem. Res.* **1968,** *1* (2), 42-49.

(63) Thomson, B. A., *J. Am. Soc. Mass Spectrom.* **1998,** *9* (3), 187-193.

(64) El-Aneed, A.; Cohen, A.; Banoub, J., *Appl. Spectrosc. Rev.* **2009,** *44* (3), 210-230.

(65) McDonnell, L. A.; Heeren, R. M., *Mass Spectrom. Rev.* **2007,** *26* (4), 606-43.

(66) Ifa, D. R.; Wu, C.; Ouyang, Z.; Cooks, R. G., *Analyst* **2010,** *135* (4), 669-81.

(67) Zhang, Y.; Ju, Y.; Huang, C.; Wysocki, V. H., *Anal. Chem.* **2014,** *86* (3), 1342-6.

(68) Zhang, Z. P.; Xiao-Ning, L.; Zheng, Y. J., *Chin. J. Anal. Chem.* **2014,** *42* (1), 145-152.

(69) Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C., *Anal. Chem.* **1990,** *62* (20), 2162-2172.

(70) Hang, W.; Lewis, C.; Majidi, V., *Analyst.* **2003,** *128* (3), 273-280.

(71) Guilhaus, M.; Selby, D.; Mlynski, V., *Mass Spectrom. Rev.* **2000,** *19* (2), 65-107.

(72) Opsal, R. B.; Owens, K. G.; Reilly, J. P., *Anal. Chem.* **1985,** *57* (9), 1884-1889.

(73) Leney, A. C.; Heck, A. J. R., *J. Am. Soc. Mass Spectrom.* **2017,** *28* (1), 5-13.

(74) Arora, B.; Tandon, R.; Attri, P.; Bhatia, R., *Curr. Protein Pept. Sci.* **2016,** *17* (999), 1-1.

(75) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R., *Mol. Cell. Proteomics* **2010,** *9* (8), 1634-49.

(76) Spicer, C. D.; Davis, B. G., *Nat. Commun.* **2014,** *5*, 4740.

(77) Hager-Braun, C.; Tomer, K. B., *Biochemistry* **2002,** *41* (6), 1759-66.

(78) Vahidi, S.; Konermann, L., *J. Am. Soc. Mass Spectrom.* **2016,** *27* (7), 1156-64.

(79) Xu, G.; Chance, M. R., *Chem. Rev.* **2007,** *107* (8), 3514-43.

(80) Hvidt, A.; Linderstrom-Lang, K., *Biochim. Biophys. Acta* **1954,** *14* (4), 574-5.

(81) Johnson, R. S.; Walsh, K. A., *Protein Sci.* **1994,** *3* (12), 2411-8.

(82) Woods, V. L., Jr.; Hamuro, Y., *J. Cell. Biochem. Suppl.* **2001,** *Suppl 37* (37), 89-98.

(83) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; He, Y.; Hendrickson, C. L.; Marshall, A. G.; Griffin, P. R., *Anal. Chem.* **2006,** *78* (4), 1005-14.

(84) Wu, Y.; Engen, J. R.; Hobbins, W. B., *J. Am. Soc. Mass Spectrom.* **2006,** *17* (2), 163-7.

(85) Liepinsh, E.; Otting, G., *Magn. Reson. Med.* **1996,** *35* (1), 30-42.

(86) Jansson, E. T.; Lai, Y. H.; Santiago, J. G.; Zare, R. N., *J. Am. Chem. Soc.* **2017,** *139* (20), 6851-6854.

(87) Konermann, L.; Pan, J.; Liu, Y. H., *Chem. Soc. Rev.* **2011,** *40* (3), 1224-34.

(88) Konermann, L.; Tong, X.; Pan, Y., *J. Mass Spectrom.* **2008,** *43* (8), 1021-36.

(89) Smith, D. L.; Deng, Y.; Zhang, Z., *J. Mass Spectrom.* **1998,** *32* (2), 135-146.

(90) Katta, V.; Chait, B. T., *Rapid Commun. Mass Spectrom.* **1991,** *5* (4), 214-7.

(91) Cravello, L.; Lascoux, D.; Forest, E., *Rapid Commun. Mass Spectrom.* **2003,** *17* (21), 2387-93.

(92) Yoshitomo, H.; J., C. S.; S., M. K.; J., T. S.; A., M. J., *Rapid Commun. Mass Spectrom.* **2008,** *22* (7), 1041-1046.

(93) Wang, L.; Pan, H.; Smith, D. L., *Mol. Cell. Proteomics* **2002,** *1* (2), 132-8.

(94) Ahn, J.; Jung, M. C.; Wyndham, K.; Yu, Y. Q.; Engen, J. R., *Anal. Chem.* **2012,** *84* (16), 7256-62.

(95) Zhao, D. S.; Gregorich, Z. R.; Ge, Y., *Proteomics* **2013,** *13* (22), 3256-3260.

(96) Wales, T. E.; Engen, J. R., *Mass Spectrom. Rev.* **2006,** *25* (1), 158-70.

(97) Pan, J.; Rintala-Dempsey, A. C.; Li, Y.; Shaw, G. S.; Konermann, L., *Biochemistry* **2006,** *45* (9), 3005-13.

(98) Maclean, B.; Tomazela, D. M.; Abbatiello, S. E.; Zhang, S.; Whiteaker, J. R.; Paulovich, A. G.; Carr, S. A.; Maccoss, M. J., *Anal. Chem.* **2010,** *82* (24), 10116-24.

(99) Guo-Zhong, L.; C., V. J. P.; C., S. J.; Dan, G.; V., G. M.; J., G. S., *Proteomics* **2009,** *9* (6), 1696-1719.

(100) Zhang, Z.; Smith, D. L., *Protein Sci.* **1993,** *2* (4), 522-31.

(101) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., *Anal. Chim. Acta* **2012,** *721*, 7-21.

(102) Sowole, M. A.; Konermann, L., *Anal. Chem.* **2014,** *86* (13), 6715-22.

(103) West, Graham M.; Chien, Ellen Y. T.; Katritch, V.; Gatchalian, J.; Chalmers, Michael J.; Stevens, Raymond C.; Griffin, Patrick R., *Structure* **2011,** *19* (10), 1424-1432.

(104) Zhang, Q.; Willison, L. N.; Tripathi, P.; Sathe, S. K.; Roux, K. H.; Emmett, M. R.; Blakney, G. T.; Zhang, H. M.; Marshall, A. G., *Anal. Chem.* **2011,** *83* (18), 7129-36.

(105) Sevy, A. M.; Healey, J. F.; Deng, W.; Spiegel, P. C.; Meeks, S. L.; Li, R., *J. Thromb. Haemost.* **2013,** *11* (12), 2128-2136.

(106) Nakazawa, S.; Hashii, N.; Harazono, A.; Kawasaki, N., *Anal. Biochem.* **2012,** *420* (1), 61-67.

(107) Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroe, I. R.; Swenson, J.; Young, R. D., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (13), 5129-34.

(108) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., *J. Phys. Chem. B.* **2001,** *105* (28), 6474-6487.

(109) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., *J. Phys. Chem. B* **1998,** *102* (18), 3586-616.

(110) Saito, M., *J. Chem. Phys.* **1994,** *101* (5), 4055-4061.

(111) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., *J. Chem. Phys.* **1995,** *103* (19), 8577-8593.

(112) Essmann, U.; Perera, L.; L., B. M., *J. Chem. Phys.* **1995**.

(113) Humphrey, W.; Dalke, A.; Schulten, K., *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.

(114) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B., ACM: Portland, Oregon, **2009**, 1-11.

(115) Duan, Y.; Kollman, P. A., *Science* **1998,** *282* (5389), 740-744.

(116) Hu, H.; Sun, Y., *Appl. Phys. Lett.* **2013,** *103* (26), 263110.

(117) Welch, D. A.; Mehdi, B. L.; Hatchell, H. J.; Faller, R.; Evans, J. E.; Browning, N. D., *Adv Struct Chem Imaging.* **2015,** *1* (1), 1.

(118) Hedges, J. B.; Vahidi, S.; Yue, X.; Konermann, L., *Anal. Chem.* **2013,** *85* (13), 6469-76.

Chapter 2: Protein Structural Dynamics at the Gas/Water Interface Examined by Hydrogen Exchange Mass Spectrometry

2.1 Introduction

Under physiological conditions most globular proteins adopt a highly ordered conformation, comprising a hydrophobic core and a hydrophilic exterior.[1] This native state can be destabilized by exposure to extremes of pH, high or low temperature, and denaturants such as urea.[2-4] These physical and chemical agents can result in various degrees of unfolding, from semi-structured molten globules all the way to the random coil state.[5-7] Non-native conformers formed in this way are often prone to aggregation.[8,9]

Proteins have become a valuable commodity – not only as enzymes for industrial processes but also as vaccines and therapeutic agents. Protein drugs such as monoclonal antibodies and antibody-drug conjugates have a rapidly growing market share.[10,11] The formulation, processing, and packaging of these proteins may induce various degradation processes.[12,13] In other words, structural changes and/or aggregation may take place even if the proteins are never exposed to any of the "classical" denaturing agents mentioned above.[14-16] Especially for protein drugs these issues are of concern because degradation may affect their efficacy and immunogenicity.[16] Early investigations considered shear stress experienced during pumping and stirring to be a major cause of degradation.[17] Recent studies found shear effects on proteins to be negligible,[13,18-20] although shear can certainly damage protein-producing cells in stirred-tank reactors.[11,21,22]

It has now been established that a key factor associated with the degradation of proteins during processing is the presence of gas/water interfaces in the form of bubbles or foam.[14,15,23-27] Proteins exhibit a high affinity for these interfaces. Protein-induced foaming

in bioreactors can lead to overflow and other problems during large-scale industrial operations.[28] On the other hand, foam stabilization via protein adsorption is beneficial for many food and beverage formulations, exemplified by the frothy appearance of whisked egg-white and the foaming of beer.[25,29] The texture and shelf life of aerated food products greatly depends on the capability of proteins to prevent the coalescence of gas bubbles.[30] Similar considerations apply to the interaction of proteins with oil/water interfaces, a factor that is important for the stabilization of emulsions such as milk. [31]

The affinity of proteins for gas/water interfaces is related to the amphiphilic nature of the polypeptide chains. In a manner that is reminiscent of low molecular-weight surfactants, hydrophobic residues tend to orient themselves towards the gas phase, whereas hydrophilic moieties will remain solvated by water. The tendency of previously buried hydrophobic sites to maximize interactions with the gaseous side of the interface often forces adsorbed proteins to adopt non-native conformations[14,25,28,29] that readily aggregate.[14,15,23-27]

Despite the considerable interest in surface denaturation phenomena, the exact nature of protein structure and dynamics at gas/water interfaces remains poorly understood. Previous work in this area employed probes such as CD, fluorescence, or FTIR spectroscopy, light and small-angle X-ray scattering, as well as surface tension measurements.[14,15,25,31] Protein losses due to aggregation have been quantified by UV-Vis absorbance measurements[14,27] and radioactivity assays.[26]

Hydrogen/deuterium exchange (HDX) mass spectrometry (MS) has emerged as a versatile tool for examining protein structural features and conformational fluctuations under a wide range of conditions. [32-40] This technique relies on the fact that exposure of a protein to D_2O will induce the replacement of amide backbone hydrogens with deuterium.[41-43] Limited proteolysis of the labeled protein with subsequent LC separation and mass analysis of the resulting peptides allows HDX kinetics to be measured in a spatially-resolved manner.

In near-neutral solution *unstructured* protein regions undergo rapid HDX with a "chemical" rate constant k_{ch} that is on the order of 650 min^{-1}.[43] Deuteration in *folded* segments proceeds much more slowly, mediated by structural fluctuations that transiently disrupt N-H···O=C hydrogen bonds while providing solvent access to amide backbone

sites.[44] These thermally activated dynamics cause each protein to constantly cycle through all of its Boltzmann-accessible conformations.[44] Although the steady-state concentration of most excited conformers is low, the measured HDX behavior is largely governed by these conformational dynamics. The opening and closing rate constants associated with HDX-relevant structural fluctuations are referred to as k_{op} and k_{cl}, respectively. Two different regimes can be distinguished. Under most conditions the open states have a relatively short lifetime ($k_{cl} >> k_{ch}$), such that numerous opening/closing events are required before all NH groups associated with a given structural transition are deuterated. This so-called EX2 limit is recognizable in HDX/MS as a continuous shift of the peak envelopes to higher mass, with an overall deuteration rate constant [45-48]

$$k_{HDX} = k_{op} / k_{cl} \times k_{ch}$$

Equation 2.1

Alternatively, proteins may undergo cooperative dynamics where the open state is quite long-lived ($k_{cl} << k_{ch}$) such that complete deuteration of the corresponding NH sites will take place during the first opening event. These "EX1" dynamics give rise to bimodal distributions where the relative intensity of the high mass component increases with time, and where[45-48]

$$k_{HDX} = k_{op}$$

Equation 2.2

Using met-myoglobin (Mb) as model system, the current work employs HDX/MS to examine protein structure and dynamics at the gas/water interface. Mb is an archetypical globular protein[49] with a native state that comprises eight helices. Most nonpolar side chains are sequestered in a well-developed hydrophobic core. The protein binds heme in a nonpolar pocket, with Fe^{3+} ligation by the proximal His93. The choice of Mb for the experiments of this study is based on previous work that revealed Mb to be highly susceptible to the presence of gas/water interfaces.[27] By conducting comparative experiments on Mb solutions with and without N_2 sparging we obtain detailed insights into

the mechanism by which gas/water interfaces promote unfolding and aggregation. We develop a kinetic model that is capable of quantitatively describing the experimentally observed interplay between aggregation and EX1/EX2 conformational fluctuations. It appears that this study marks the first use of HDX/MS as a tool for exploring surface denaturation effects.

2.2 Methods

2.2.1 Materials

Equine skeletal muscle Mb was purchased from Sigma (St. Louis, MO). HCl, Na_2HPO_4 and NaH_2PO_4 were obtained from Caledon (Georgetown, ON, Canada). D_2O was procured from Aldrich (St. Louis, MO). LC-MS grade H_2O was purchased from Optima (Fair Lawn, NJ). All protein solutions were prepared in 50 mM phosphate buffer. pH values were measured using a Fisher (Waltham, MA) AB15 pH meter. All experiments were carried out at a pH meter reading of 7.1, without correction for isotope effects.

2.1.1 Gas/Water Interface Exposure and HDX

200 μL aliquots of 5 μM Mb in 90% D_2O were placed in Eppendorf tubes. N_2 was bubbled through these solutions at 5 L h^{-1} using a syringe needle with an inner diameter of 0.8 mm. This procedure was carried out at room temperature (23 ± 1 ˚C). After various deuteration time periods ranging from 1 to 20 minutes the gas flow was stopped. Insoluble aggregates formed during this treatment were removed by centrifugation at 10,000 rpm for 30 s. HDX was quenched by adding HCl to the supernatant, resulting in a pH meter reading of 2.4.

Subsequently, samples were flash frozen in liquid nitrogen until further analysis. Times reported in the text above refer to the bubble-exposure period, not including of the ~60 s required for centrifugation, centrifuge deceleration, and sample handling prior to freezing. Measurements for quantifying protein loss due to aggregation were conducted using non-deuterated buffer solutions. Control samples for all measurements were prepared exactly as described above, but without N_2 bubbling.

2.1.2 Mass Spectrometry

Frozen deuterated samples were rapidly thawed to ~0 ˚C. Intact protein HDX/MS employed a reversed-phase column for desalting (BEH C4, 1.7 µm, 2.1 × 50 mm, Waters, Milford, MA) with a 10 min water/acetonitrile gradient at 200 µL min^{-1} in the presence of 0.1% formic acid. The sample loop had a 20 µL volume and the amount of Mb per injection was 100 pmol. The column, injector and solvent lines were kept in ice for maintaining a temperature close to 0 ˚C to minimize H/D back exchange. The column outlet was connected to a Waters Synapt HDMS quadrupole time-of-flight mass spectrometer. Intact protein analyses focused on the 18+ charge state which exhibited the highest S/N ratio.

Peptide-level deuteration measurements were conducted using a nanoACQUITY UPLC with HDX technology (Waters)[50] Online digestion was performed using a 2.1 mm × 30 mm POROS pepsin column (Life Technologies/Applied Biosystems, Carlsbad, CA) at 15 ˚C. A 20 min water/acetonitrile (0.1% formic acid) gradient was used for desalting and peptide separation at 0 ˚C. Peptide mass analysis was performed on a Waters Synapt G2 mass spectrometer. The identity of each peptide was confirmed by tandem MS on non-deuterated samples based on the known Mb sequence.[49] Peptide mass spectra were recorded with an electrospray capillary voltage of +2.8 kV, 30 V cone voltage and a desolvation temperature of 250 ˚C. Mass calibration was performed using 2 µg μL^{-1} NaI in 50:50 water: isopropanol at a capillary voltage of 1.2 kV. Deuteration percentages were calculated as $(m - m_0)/(m_{100} - m_0)$. Zero-time-point controls (m_0) were prepared by exposing the protein to 90% D_2O buffer under quench conditions. Fully exchanged control

samples (m_{100}) were prepared by incubating Mb in 90% D_2O at pD 2.4 / 37 ˚C for 12 h. All experiments were carried out in triplicate. Error bars represent standard deviations.

2.2 Results and Discussion

2.2.1 Protein Aggregation in the Presence of Gas Bubbles

Sparging of Mb solutions with N_2 bubbles results in the formation of insoluble aggregates. The assay used for examining this process involves precipitate removal by centrifugation, followed by LC/MS analysis of the supernatant. The loss of soluble protein in the presence of N_2 bubbles can be approximated as an exponential decay with an apparent rate constant of k_{agg} = 0.054 min^{-1}. In contrast, the protein concentration in bubble-free control samples remains constant (Figure 1).

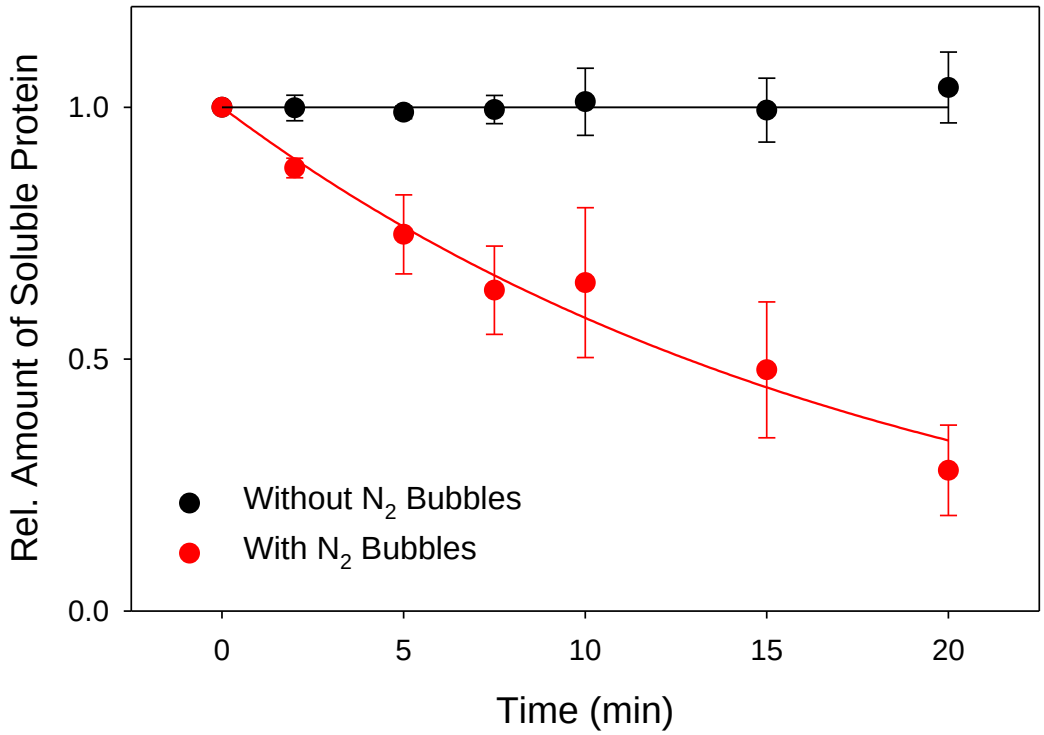


Figure 1. *Relative amount of soluble protein in the presence and absence of N_2 bubbles, as monitored by ESI-MS signal intensity measurements. Protein loss in the presence of bubbles is attributed to aggregation. The red curve is an exponential fit with an apparent rate constant of 0.054 min^{-1}.*

The occurrence of Mb aggregation under the conditions of our experiments is consistent with earlier observations on other proteins.[15,23-27] In particular, it has been noted[14] that sparging is characterized by a nascent interface area that continuously renews itself. These conditions are much more effective in causing aggregation than the "static" surface of regular protein samples.[14] A likely explanation for this difference is that the surface of regular samples is decorated with a metastable thin film of adsorbed protein. This film prevents bulk solution proteins from coming into direct contact with the air/water interface. Sparging, on the other hand, continuously produces bare air/water interface that will readily adsorb proteins. Bubbles rise until they burst, leaving previously adsorbed proteins in non-native conformations that are prone to aggregation. Alternatively, aggregation may begin already while proteins are still in contact with the bubble surface.

2.2.2 Intact Protein HDX/MS

Mb conformational dynamics in the absence and in the presence of N_2 bubbles were probed using HDX/MS. As a first step the Mb behavior was monitored at the intact protein level. In agreement with previous reports,[40,51] bubble-free control samples display EX2 kinetics with mass distributions that gradually shift to higher m/z, (Figure 2A). A drastically different behavior was observed in the presence of N_2 bubbles, with spectra that are bimodal for $t > 2$ min (Figure 2B). The decreasing relative intensity of the low mass component and the increasing intensity of the high mass component represent the hallmark of EX1 exchange.[45,46] In addition, a gradual shift of the low mass component is apparent in Figure 2B which reveals that the protein also undergoes EX2 deuteration. The spectra in Figure 2B are normalized such that each time point has the same integrated area. This representation is helpful for recognizing the occurrence of combined EX1/EX2 dynamics [45] in the sparged samples, but it masks the ongoing loss of soluble protein due to aggregation (as discussed above, Figure 1). A more realistic view of the protein behavior is provided in Figure 2C, where the HDX mass spectra of the sparged samples are displayed using actual signal intensities such that aggregation effects become easily recognizable.

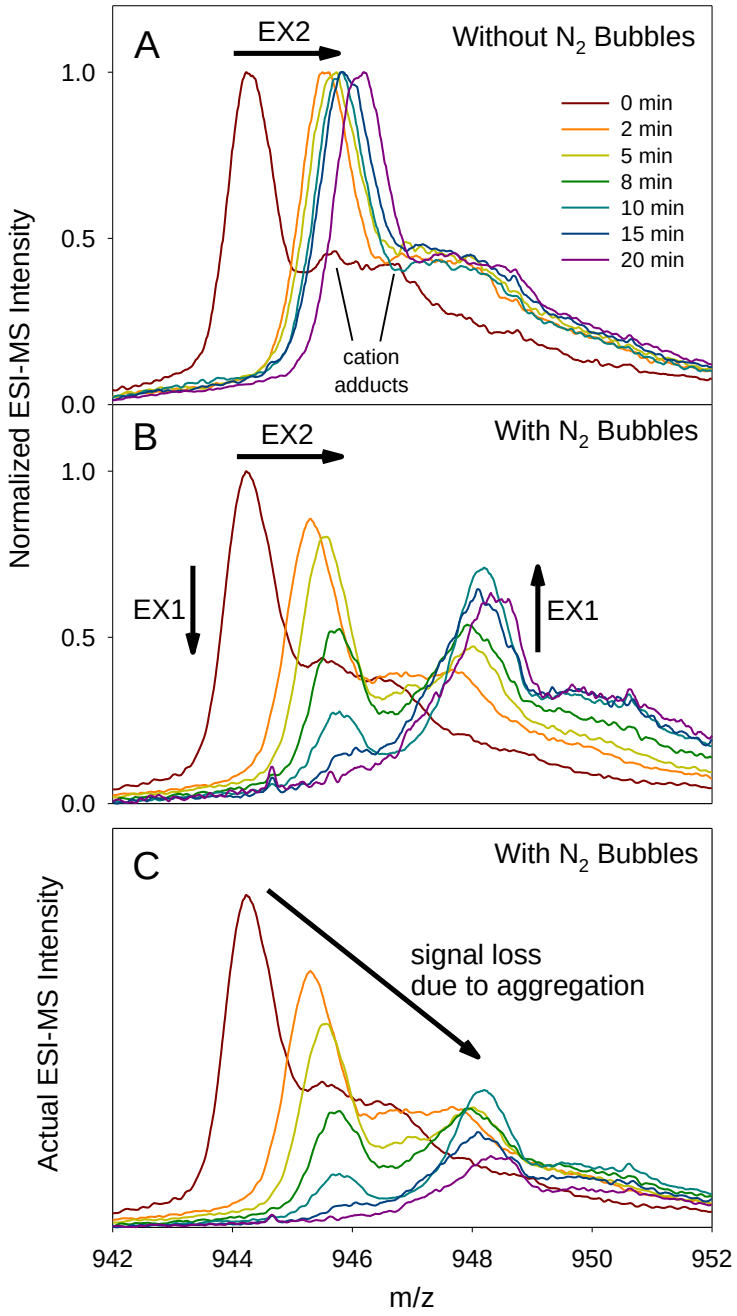


Figure 2. *Mb mass spectra recorded at different HDX time points in the absence (A) and in the presence (B), (C) of N_2 bubbles. The data shown here correspond to the 18+ charge state. Time-dependent shifts in peak maxima (EX2 behavior) are highlighted in panels A, B using arrows. In addition, for panel B time-dependent changes in the relative intensity of high mass vs. low mass component (EX1 behavior) are emphasized. The spectra in (B) were normalized such that each time point represents the same integrated intensity. (C) The same data as in panel B, but with actual intensities that reflect the time-dependent decrease in overall signal due to aggregation. Unresolved cation adducts are highlighted for the t = 0 spectrum in panel A.*

By tracking the low mass peak maxima in Figure 2 it is found that the EX2 kinetics without and with bubble sparging are quite similar. Both EX2 deuteration profiles approach a maximum level around 55%, with k_{HDX}^{EX2} values of 0.79 min^{-1} and 0.56 min^{-1}, respectively (Figure 3A). For EX1 analyses one has to determine the signal intensities associated with the high mass and the low mass components, i.e., I_{high} and I_{low}. The kinetics can then be determined by using the relationship[52]

$$EX1\ Deuteration = \frac{I_{high}}{I_{low} + I_{high}}$$

Equation 2.3

Figure 2 reveals the occurrence of peak tailing due to cation adduction.[53] This type of artifact is common in intact protein HDX/MS.[54] Cation adduction is observed regardless of experimental conditions (see for example Figure 2A). Overlap of adduct signals with the high mass EX1 component renders the data analysis somewhat difficult. The EX1 data of Figure 2B were therefore analyzed in a semi-quantitative fashion by assuming that both components can be described using the peak shape of the t = 0 profiles, subject to the appropriate mass shifts. On the basis of equation 2.3 these data were then converted to an EX1 profile (Figure 3B). These data can be described by an exponential curve with k_{HDX}^{EX1} = 0.084 min^{-1}. The scatter in the experimental data points is attributed to the challenges outlined above.

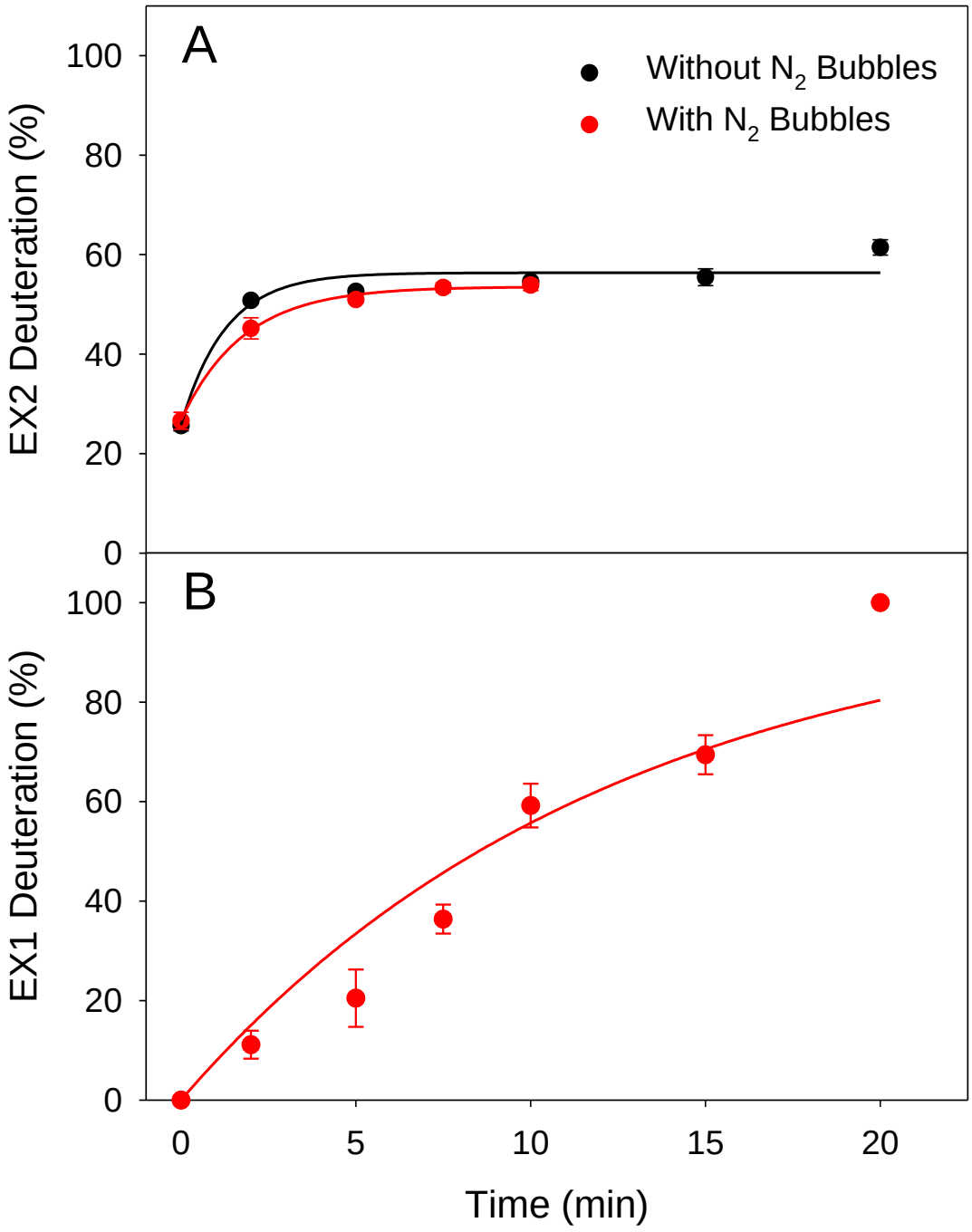


Figure 3. *Intact Mb HDX profiles extracted from the data of Figure 2. (A) EX2 kinetics in the absence and in the presence of N_2 bubbles. Curves are exponential fits with kHDXEX2 = 0.79 min^{-1} and 0.56 min^{-1}, respectively. The "red" profile shows fewer data points because the EX2 component becomes indiscernible for t > 10 min in the presence of N2 bubbles. (B) EX1 kinetics observed in the presence of bubbles. The exponential fit has kHDXEX1 = 0.084 min^{-1}.*

Summarizing the intact protein data of Figures 1-3 it can be stated that the exposure of Mb to N_2 bubbles results in a combination of EX1 and EX2 conformational dynamics, in conjunction with aggregation. Without sparging only EX2 behavior is observed.

2.2.3 Proteolytic Digestion HDX/MS

For gaining additional insights into the Mb behavior in the presence of N_2 bubbles we analyzed the HDX/MS behavior of the protein at the peptide level. For short peptides it is difficult to differentiate between EX1 and EX2 behavior. Our attention was therefore focused on relatively large segments that comprised 15 to 35 residues. In the presence of N_2 bubbles all of these peptides display bimodal mass distributions, as exemplified in Figure 4. This finding implies the EX1 structural dynamics correspond to *global* opening/closing transitions that affect the entire Mb structure, not just individual parts of the protein. Gaussian decomposition allows a determination of EX1 kinetic profiles on the basis of equation 2.3. The resulting kinetic plots confirm that the EX1 dynamics for all peptides are very similar, as would be expected for global conformational fluctuations (Figure 5A). Global analysis of this data set results in $k_{HDX}^{EX1} = 0.089\ min^{-1}$, close to the value obtained from the intact protein data of Figure 3B.

By tracking the maxima of the low mass Gaussian component for each peptide it is possible to monitor the EX2 kinetics of the protein. Data for three representative segments are depicted in Figure 5B-D (residues 1-29, 70-103, and 138-151). The EX2 kinetics for all three peptides are well described by k_{HDX}^{EX2} values on the order of 0.5 min^{-1}, both in the presence and in the absence of N_2 sparging (see caption of Figure 5 for details).

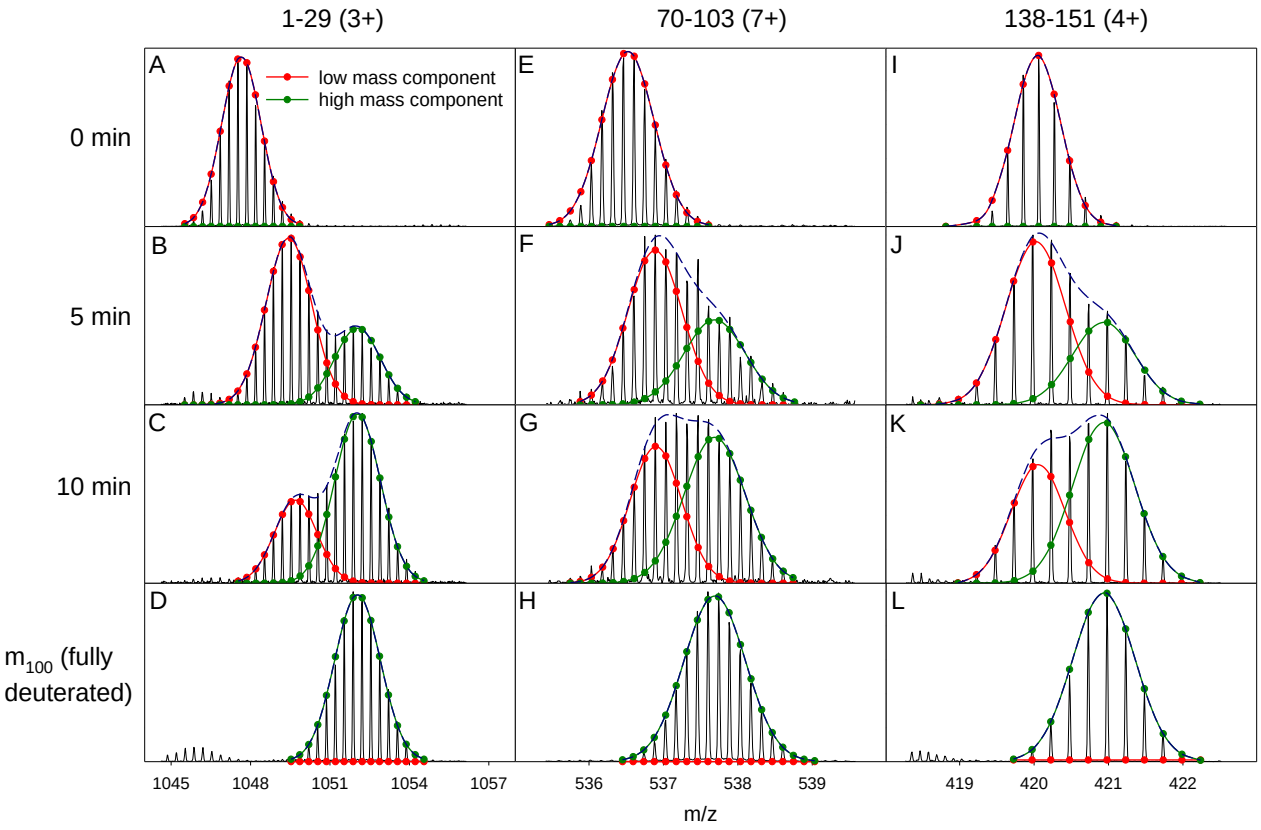


Figure 4. *HDX mass spectra of three representative Mb peptides recorded at different time points in the presence of N_2 bubbles. EX1/EX2 kinetics were analyzed using two-component Gaussian decompositions. The analysis was performed using Microsoft Excel Solver for global fitting that required the peak position and fwhm to be the same for the high mass component of each peptide for all time points.*

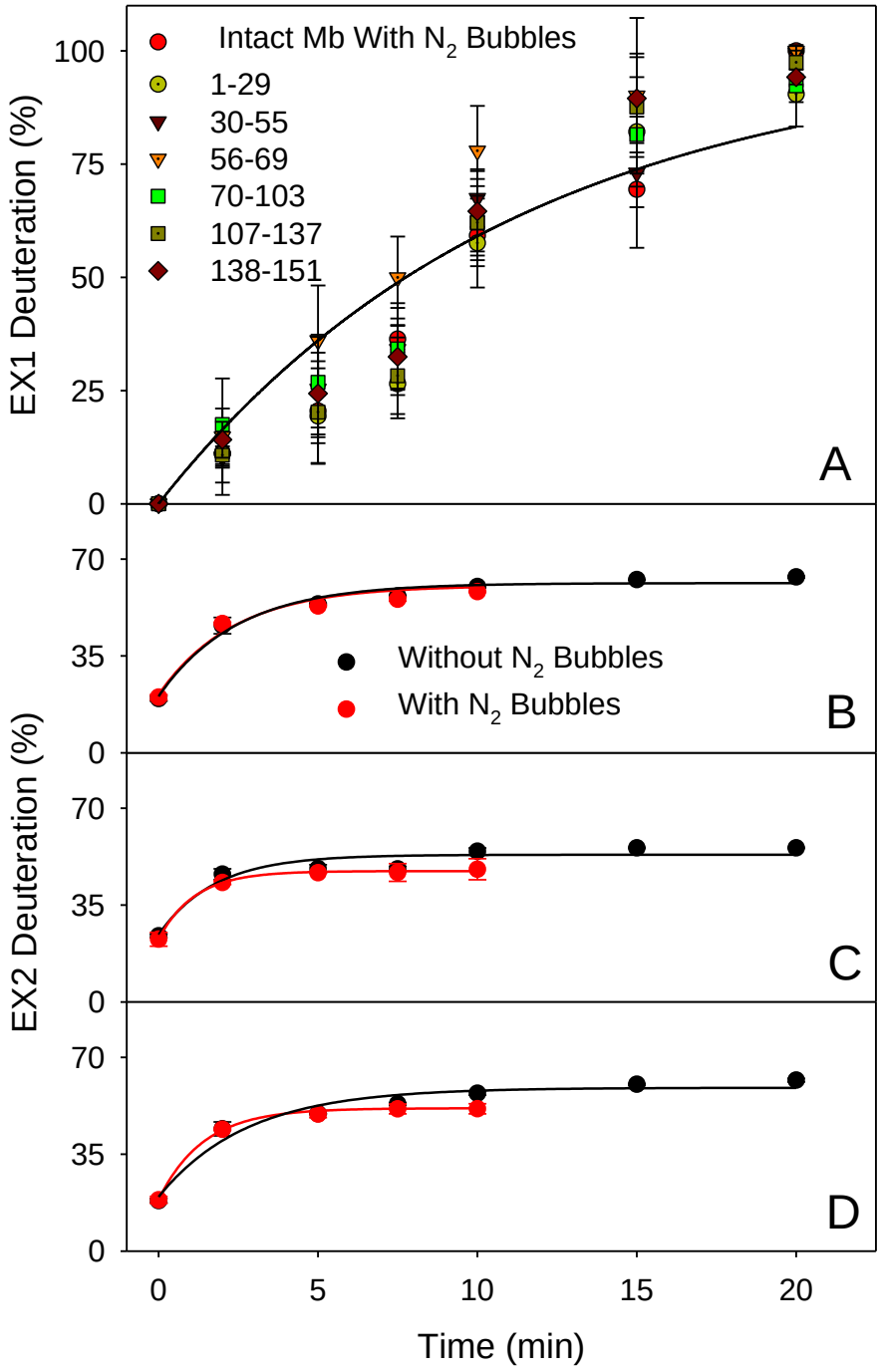


Figure 5. *(A) EX1 kinetics of Mb peptic fragments in the presence of N_2 bubbles. The solid line represents a global fit to the combined data, with k_{HDX}^{EX1} = 0.089 min^{-1}. Intact protein data are included as well. Panels B-D show a comparison of EX2 kinetic profiles recorded in the absence and in the presence of N_2 bubbles. (B) Residues 1-29, k_{HDX}^{EX2} = 0.41 and 0.42 min^{-1}; (C) residues 70-103, k_{HDX}^{EX2} = 0.57 and 0.88 min^{-1}; (D) residues 138-151, k_{HDX}^{EX2} = 0.36 and 0.71 min^{-1}.*

2.2.4 Mb at the Gas/Water Interface: A Simple Kinetic Model

The data presented in the preceding sections demonstrate that Mb in the presence of N_2 bubbles undergoes both EX2 and EX1 fluctuations, as well as aggregation. We will now devise a model that is capable of describing these processes in a quantitative fashion. It is not our aim to provide an atomistic framework, rather we strive to come up with the simplest possible description of the Mb behavior.

In bubble-free solution ~55% of backbone amides are involved in rapid sub-global fluctuations that give rise to EX2 kinetics (Figure 3A). Earlier work has revealed that the corresponding sites are located in the N-terminal region of helix A, as well as helices B, C, D, and the C-terminal half of helix H.[40] We make the simplifying assumption that all of these sites participate in a single type of opening/closing transition, corresponding to "foldon" fluctuations[44] between a semi-unfolded species and the native state. These dynamics may be described using the notation $OC \leftrightarrow CC$, where the first letter designates the status ("*O*" open, or "*C*" closed) of the 55% of amides that participates in the EX2 dynamics. The second letter signifies the residual 45% that do not participate in the EX2 dynamics.

Upon exposure to gas bubbles the EX2 kinetics do not change very much (Figure 3A). However, N_2 sparging gives rise to additional EX1 conformational transitions that expose *all* amides to the solvent (Figure 5A), implying the involvement of surface-adsorbed conformers that are globally unfolded ("*OO*"). Inclusion of this species extends the model to $OC \leftrightarrow CC \leftrightarrow OO$. As a final step, aggregation has to be considered. The lack of protein loss in bubble-free solution suggests that neither *OC* nor *CC* are particularly susceptible to aggregation. This leaves *OO* as the most likely aggregation-prone species. *OO* can either refold to *CC*, or it can form aggregates. This kinetic competition can be incorporated by modifying the model to

OC ↔ *CC* ↔ *OO* → *aggregated* (see Figure 6A for a complete description that includes rate constants). A cartoon representation of the four kinetic species is provided in Figure 6B, emphasizing their different degrees of "openness", as well as the interaction of nonpolar residues (orange) with the gas/water interface[14,15,23-29] in the *OO* state.

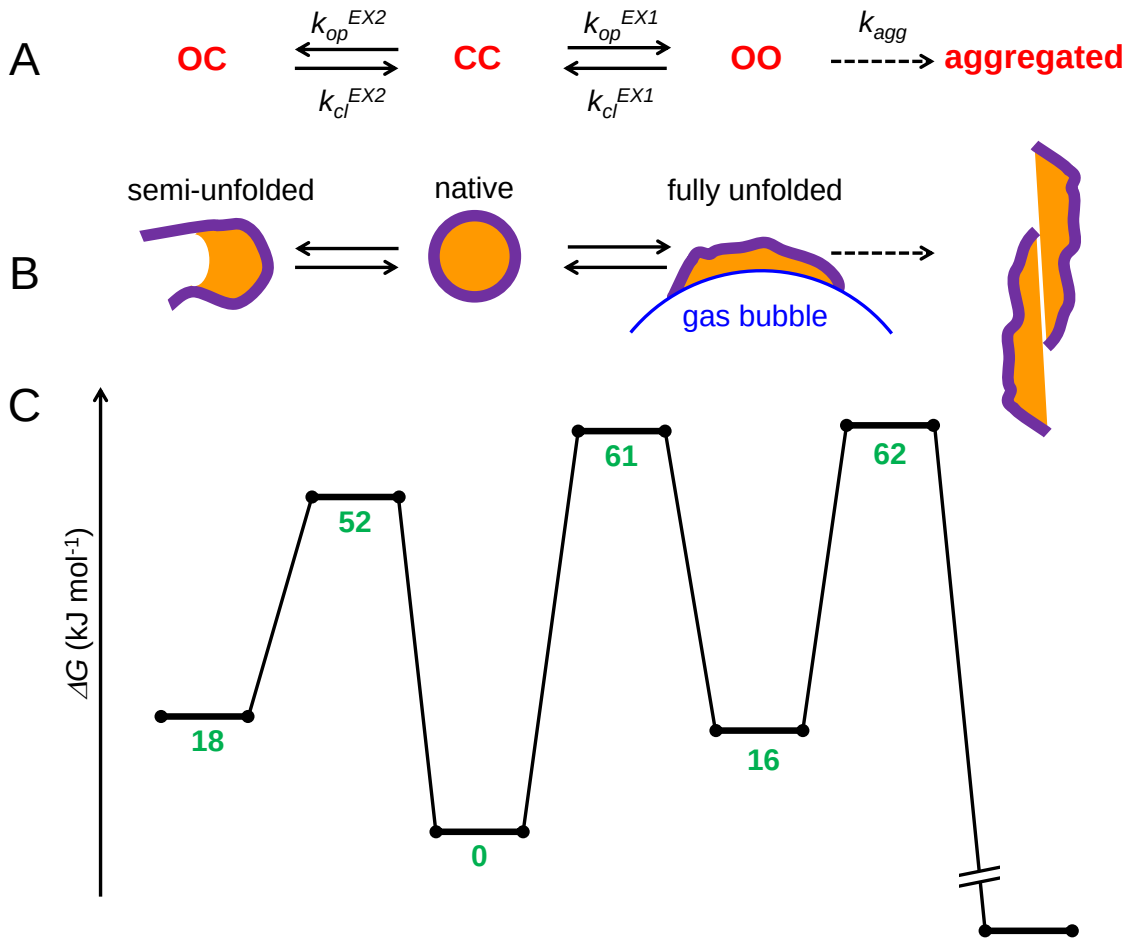


Figure 6. *Kinetic model of the protein dynamics in the presence of N_2 bubbles. Backbone amides are conceptually divided into two groups, each of which can either be open ("O") or closed ("C"). (A) The CC native state undergoes EX2 fluctuations to the semi-unfolded conformer OC. Alternatively, CC can undergo EX1 fluctuations to the fully unfolded species OO which is adsorbed to the gas/water interface. The latter can either refold or aggregate. (B) The same model as in panel A, with cartoon representations for each of the four species. Orange color represents hydrophobic residues. (C) Free energy profile of the protein. Numbers indicate ΔG (in kJ mol^{-1}) relative to the CC native state.*

The appropriateness of the model can be scrutinized by testing whether it is capable of reproducing the experimental HDX and aggregation kinetics. We previously developed a procedure for simulating deuteration processes that are associated with the interconversion of different conformers.[45] That method iteratively tracks the behavior of a population of bit strings (proteins), where each position (backbone amide site) is occupied by either 0 (hydrogen) or 1 (deuterium). Opening/closing dynamics of user-defined protein regions take place with probabilities governed by the corresponding k_{op} and k_{cl} values. Open sites undergo $0 \rightarrow 1$ conversion (deuteration) with a probability that depends on k_{ch}. The mass shift distribution of the population is analyzed at selected time points by adding the number of 1s in each of the proteins. It was already demonstrated that a $OC \leftrightarrow CC \leftrightarrow OO$ model is capable of describing the simultaneous occurrence of EX1 and EX2 kinetics (see Figure 5 in ref. [45]). For the current work we extend this framework by incorporating an aggregation step, as envisioned in Figure 6. This is achieved by implementing a reaction path that removes *OO* with a probability *W* that is given by

$$W = 1 - \exp(-k_{agg} \times \Delta\tau)$$

Equation 2.4

where k_{agg} is the aggregation rate constant and $\Delta\tau$ is the iteration time step.[45]

Given the minimalist nature of the model, a number of simplifications are required. We assume that chemical exchange proceeds with $k_{ch} = 650$ min^{-1} at all sites, representing the amino acid-averaged value at neutral pH.[43] For EX2 deuteration ($k_{cl} >> k_{ch}$) it is assumed that $k_{cl}^{EX2} = 6500$ min^{-1}. Conversely, for EX1 kinetics ($k_{cl} << k_{ch}$) we use $k_{cl}^{EX1} = 65$ min^{-1}. The experimentally measured $k_{HDX}^{EX2} = 0.56$ min^{-1} (Figure 3A) then corresponds to $k_{op}^{EX2} = 5.6$ min^{-1}, as dictated by equation 2.1. For a "pure" EX1 scenario the fitted rate constant of Figure 3B would suggest that $k_{op}^{EX1} = k_{HDX}^{EX1} = 0.084$ min^{-1} (equation 2.2). However, the kinetic competition between k_{cl}^{EX1} and k_{agg} renders equation 2.2 inadequate for the reaction scheme of Figure 6A. To address this point, we empirically settled on a slightly higher value of $k_{op}^{EX1} = 0.12$ min^{-1}. Aggregation of *OO* was approximated as a first-order process with $k_{agg} = 44$ min^{-1}. Figure 7 demonstrates that this choice of parameters results

in excellent agreement between the experimental data (red) and the simulated kinetics (blue).

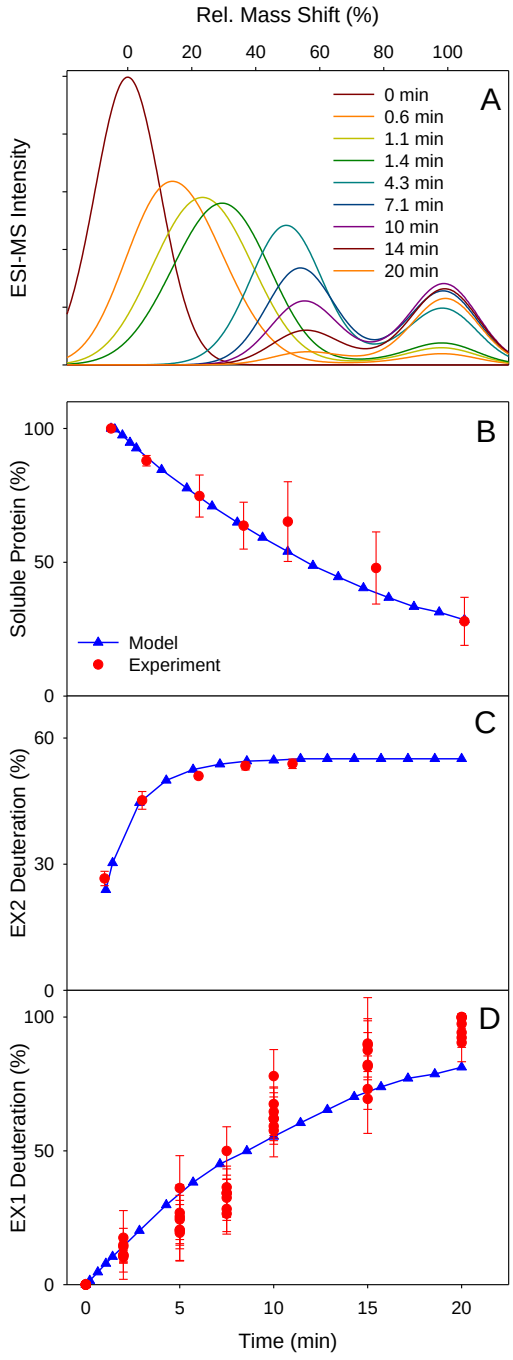


Figure 7. *Simulation results, using the model of Figure 6 for describing the HDX and aggregation kinetics. (A) Intact protein mass shift distributions, (B) aggregation kinetics, (C) EX2 kinetics, (D) EX1 kinetics. Simulation results in panels B-D are displayed in blue, whereas experimental data (from Figures 1, 3A, 5A) are shown in red. Simulations were conducted using the method of ref. 45 for 5000 proteins with rate constants as defined in the text.*

The agreement between modeled and experimental kinetics (Figure 7) suggests that the framework of Figure 6 provides a reasonable approximation of the protein dynamics in the presence of N_2 bubbles. We do not claim that this model is unique; alternative scenarios such as $CC \leftrightarrow OC \leftrightarrow OO \rightarrow aggregated$ might fit the data equally well. However, from our experiments there is no evidence that *OC* is an obligatory intermediate for the formation of *OO*. It was therefore decided to put forward the model of Figure 6, which does not involve a direct $OC \leftrightarrow OO$ transition.

2.2.5 Free Energy Landscape of Mb at the Gas/Water Interface

Having determined the rate constants associated with the model, one can proceed to outline the energy landscape of the protein in the presence of N_2 bubbles (summarized in Figure 6C). The free energy of *OC* and *OO* relative to the *CC* ground state is given by[44,45]

$$\Delta G^{\circ} = -RT \ln \frac{k_{op}}{k_{cl}}$$

Equation 2.5

yielding values of 18 and 16 kJ mol^{-1}, respectively. At first sight, it might seem surprising that *OC* and *OO* possess almost the same free energy, considering that they represent very different degrees of unfolding. This apparent contradiction is resolved when considering that the globally unfolded conformer *OO* is stabilized by adsorption to the gas/water interface.

For estimating the height of the activation barriers ($\Delta G^{\#}$) that separate the protein conformers we use the Kramers equation[55-59]

$$k_{op} = C \exp\left(-\frac{\Delta G^{\#}}{RT} \right)$$

Equation 2.6

with a pre-exponential factor of $C \approx 6 \times 10^9$ min^{-1}.[60] The resulting activation barrier heights for $OC \leftarrow CC$ and $CC \rightarrow OO$ are 52 kJ mol^{-1} and 61 kJ mol^{-1}, respectively. These numbers suggest that the transition state for global unfolding is much more unfavorable than that leading to the semi-unfolded state. Hence, while *OO* is strongly stabilized by adsorption to the gas/water interface, this stabilization is not fully developed in the $CC \rightarrow OO$ transition state.

By switching k_{op} to k_{agg} one can use equation 2.6 to estimate the kinetic barrier for aggregation of *OO*. This approach yields an activation barrier height of 46 kJ mol^{-1}, corresponding to a transition state energy that is (16 + 46) kJ mol^{-1} = 62 kJ mol^{-1} above *CC*. Aggregation can be considered to be quasi-irreversible,[61] implying that the free energy of aggregated Mb in our model is far below that of the metastable *CC* state (Figure 6C).

2.3 Conclusions

The experiments and kinetic simulations of this work provide insights into the conformational dynamics of a model protein (Mb) under conditions where interactions with gas/water interfaces are promoted by N_2 sparging. Our results highlight the destabilizing effects of gas bubbles, as well as the propensity of the protein to aggregate under these conditions. Analysis of the HDX/MS data strongly suggests the involvement of globally unfolded conformers that are adsorbed to the surface of gas bubbles via interactions with previously buried nonpolar residues. This *OO* species represents a kinetic branching point from which the protein can either aggregate, or refold to the native state.

The kinetic framework of Figure 6 describes the experimental data well (Figure 7), but

readers are reminded of the minimalist nature of our model. Many aspects are not considered in detail. For example, it seems possible that bubble-adsorbed *OO* is quite stable, and that the kinetic competition between refolding and aggregation only commences after this conformer is released into solution upon bubble bursting. Also, modeling aggregation as a first-order process with a fixed rate constant glosses over many of the complexities associated with this process. Similarly, describing EX2 dynamics via a single semi-unfolded state certainly represents an oversimplification.[40,44] Despite these limitations, we feel that the model of Figure 6 provides a useful approximation of the protein dynamics in the presence of gas bubbles.

HDX measurements clearly represent an interesting approach for exploring the properties of proteins at interfaces.[62] This technology can provide insights that go beyond those obtainable from commonly used bulk spectroscopic methods.[14,15,25-27,31] It is hoped that the approach introduced in this study will be useful for exploring protein dynamics and aggregation in industrial processes, where bubble-related degradation phenomena are commonly encountered.[14,15,23-27]

2.4 References

(1) Farber, P. J.; Mittermaier, A. *Protein Sci.* **2008**, *17*, 644.

(2) Lim, W. K.; Rosgen, J.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2595.

(3) England, J. L.; Pande, V. S.; Haran, G. *J. Am. Chem. Soc.* **2008**, *130*, 11854.

(4) Zangi, R.; Zhou, R. H.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131*, 1535.

(5) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787.

(6) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099.

(7) Curnow, P.; Booth, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18970.

(8) Hartl, F. U.; Hayer-Hartl, M. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574.

(9) Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. *J. Mol. Biol.* **2012**, *421*, 160.

(10) Kimchi-Sarfaty, C.; Schiller, T.; Hamasaki-Katagiri, N.; Khan, M. A.; Yanover, C.; Sauna, Z. E. *Trends Pharmacol. Sci.* **2013**, *34*, 534.

(11) Chu, L.; Robinson, D. K. *Curr. Opin. Biotechnol.* **2001**, *12*, 180.

(12) Bye, J. W.; Platts, L.; Falconer, R. J. *Biotechnol. Lett.* **2014**, *36*, 869.

(13) Thomas, C. R.; Geer, D. *Biotechnol. Lett.* **2011**, *33*, 443.

(14) Wiesbauer, J.; Prassl, R.; Nidetzky, B. *Langmuir* **2013**, *29*, 15240.

(15) Rudiuk, S.; Cohen-Tannoudji, L.; Huille, S.; Tribet, C. *Soft Matter* **2012**, *8*, 2651.

(16) Philo, J. S.; Arakawa, T. *Curr. Pharm. Biotechnol.* **2009**, *10*, 348.

(17) Tirrell, M.; Middleman, S. *Biotechnol. Bioeng.* **1975**, *17*, 299.

(18) Thomas, C. R.; Nienow, A. W.; Dunnill, P. *Biotechnol. Bioeng.* **1979**, *21*, 2263.

(19) Jaspe, J.; Hagen, S. J. *Biophys. J.* **2006**, *91*, 3415.

(20) Bee, J. S.; Stevenson, J. L.; Mehta, B.; Svitel, J.; Pollastrini, J.; Platz, R.; Freund, E.; Carpenter, J. F.; Randolph, T. W. *Biotech. Bioeng.* **2009**, *103*, 936.

(21) Taticek, R. A.; Lee, C. W. T.; Shuler, M. L. *Curr. Op. Biotechnol.* **1994**, *5*, 165.

(22) Joshi, J. B.; Elias, C. B.; Patole, M. S. *Chem. Eng. J.* **1996**, *62*, 121.

(23) Graham, D. E.; Phillips, M. C. *J. Colloid Interface Sci.* **1979**, *70*, 403.

(24) Donaldson, T. L.; Boonstra, E. F.; Hammond, J. M. *J. Colloid Interface Sci.* **1980**, *74*, 441.

(25) Clarkson, J. R.; Cui, Z. F.; Darton, R. C.; Clarkson, J. R. *J. Colloid Interface Sci.* **1999**, *215*, 323.

(26) Cornec, M.; Cho, D.; Narsimhan, G. *J. Colloid Interface Sci.* **1999**, *214*, 129.

(27) Hedges, J. B.; Vahidi, S.; Yue, X.; Konermann, L. *Anal. Chem.* **2013**, *85*, 6469.

(28) Prins, A.; van't Riet, K. *Trends Biotechnol.* **1987**, *5*, 296.

(29) Wilde, P. J. *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 176.

(30) Narsimhan, G. *J. Colloid Interface Sci.* **2010**, *345*, 566.

(31) Zhai, J. L.; Day, L.; Aguilar, M. I.; Wooster, T. J. *Curr. Opin. Colloid Interface Sci.* **2013**, *18*, 257.

(32) Marciano, D. P.; Dharmarajan, V.; Griffin, P. R. *Curr. Op. Struct. Biol.* **2014**, 105.

(33) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Protein Sci.* **2013**, *22*, 530.

(34) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta* **2012**, *721*, 7.

(35) Iacob, R. E.; Engen, J. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003.

(36) Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1481.

(37) Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. D. *Anal. Chem.* **2009**, *81*, 5577.

(38) Rob, T.; Liuni, P.; Gill, P. K.; Zhu, S. L.; Balachandran, N.; Berti, P. J.; Wilson, D. J. *Anal. Chem.* **2012**, *84*, 3771.

(39) Sperry, J. B.; Smith, C. L.; Caparon, M. G.; Ellenberger, T.; Gross, M. L. *Biochemistry* **2011**, *50*, 4038.

(40) Sowole, M. A.; Konermann, L. *Anal. Chem.* **2014**, *86*, 6715.

(41) Baldwin, R. L. *Proteins* **2011**, *79*, 2021.

(42) Jaswal, S. S.; Miranker, A. D. *Protein Sci.* **2007**, *16*, 2378.

(43) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.

(44) Englander, S. W.; Mayne, L.; Krishna, M. M. G. *Quart. Rev. Biophys.* **2007**, *40*, 287.

(45) Konermann, L.; Tong, X.; Pan, Y. *J. Mass Spectrom.* **2008**, *43*, 1021.

(46) Miranker, A.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. *FASEB J.* **1996**, *10*, 93.

(47) Weis, D. D.; Wales, T. E.; Engen, J. R.; Hotchko, M.; Ten Eyck, L. F. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1498.

(48) Hui, X.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'Cuk, A. I.; Kaltashov, I. A. *Protein Sci.* **2005**, *14*, 543.

(49) Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; Mauk, A. G.; Smith, M.; Brayer, G. D. *Biochim. Biophys. Acta* **1997**, *1341*, 1.

(50) Wales, T. E.; Fadgen, K. E.; Gerhardt, G. C.; Engen, J. R. *Anal. Chem.* **2008**, *80*, 6815.

(51) Johnson, R. S.; Walsh, K. A. *Protein Sci.* **1994**, *3*, 2411.

(52) Maier, C. S.; Schimerlik, M. I.; Deinzer, M. L. *Biochemistry* **1999**, *38*, 1136.

(53) Verkerk, U. H.; Kebarle, P. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1325.

(54) Tsui, V.; Garcia, C.; Cavagnero, S.; Siuzdak, G.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **1999**, *8*, 45.

(55) Kramers, H. A. *Physica* **1940**, *7*, 284.

(56) Hanggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Pys.* **1990**, *62*, 251.

(57) Chung, H. S.; Eaton, W. A. *Nature* **2013**, *502*, 685.

(58) Klimov, D. K.; Thirumalai, D. *Phys. Rev. Lett.* **1997**, *79*, 317.

(59) Jacob, M.; Geeves, M.; Holtermann, G.; Schmid, F. X. *Nat. Struct. Biol.* **1999**, *6*, 923.

(60) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597.

(61) Vendruscolo, M. *Curr. Op. Struct. Biol.* **2012**, *22*, 138.

(62) Wang, A. L.; Vo, T.; Le, V.; Fitzkee, N. C. *J. Phys. Chem. B* **2014**, *118*, 14148.

Chapter 3. Binding Interactions of Osteoprotegerin (OPG) with RANKL and Heparan Sulfate Studied by Hydrogen Exchange Mass Spectrometry

3.1 Introduction

Bone is one of the most vital connective tissues in vertebrates. It consists of a collagen matrix that is mineralized with hydroxyapatite ($[Ca_3(PO_4)_2]Ca(OH)_2$).[1] Bone takes on many roles; in conjunction with skeletal muscles it is essential for locomotion, it provides support for soft tissues, stores minerals, etc.[2] Despite its rigid appearance, bone undergoes constant turnover which allows it to respond to mechanical and physiological stimuli.[1] All these processes are part of a delicate balance between bone resorption mediated by osteoclasts, and bone formation mediated by osteoblasts.[3] Bone remodeling is essential for growth, fracture healing, and for maintaining calcium homeostasis.[4,5] An imbalance between bone resorption and formation gives rise to bone disease. Excessive osteoclast activity causes osteoporosis, while unchecked osteoblast activity results in osteopetrosis.[6,7]

Osteoprotegerin (OPG), a soluble ~50 kDa glycoprotein secreted by osteoblasts, plays a central role during bone remodeling.[8-11] Two other key proteins are RANK (receptor activator of nuclear factor-κB) and its binding partner RANKL (RANK ligand).[12-14] OPG, RANKL, and RANK are part of a signaling network that regulates bone resorption by modulating osteoclast development and activation.[14,15] RANK is a trimeric membrane protein on the surface of osteoclast precursors. Its ligand RANKL is a membrane protein expressed by osteoblasts. RANKL binding to RANK triggers the transformation of osteoclast precursors to active osteoclasts, which allows bone resorption to commence. OPG acts as a RANKL decoy receptor. By forming a high affinity RANKL/OPG complex, OPG suppresses RANKL/RANK interactions. In other words, OPG protects bone from

being resorbed by sequestering RANKL.[14,15] An elevated RANKL:OPG ratio serves as a clinical marker for pathological osteoclastogenesis.[16]

OPG can be dissected into several domains (Figure 1a).[10,17] Four cysteine-rich domains (CRD1-CRD4) contain numerous intramolecular disulfide bonds. Asn178 in CRD4 represents a glycosylation site.[18] The second half of the sequence includes two death domains (DD1 and DD2). The C-terminal tail is basic due to a relatively high percentage of Arg and Lys residues. Cys400, the penultimate residue, allows the formation of a disulfide bond with Cys400' of a second OPG chain.[8,10,18,19] We will refer to this SS-linked homodimer as OPG_2^{SS}. We distinguish this covalent dimer from OPG^{SH} which is not covalently linked and instead possesses a free sulfhydryl group at Cys400. Both OPG_2^{SS} and OPG^{SH} bind RANKL with K_d values in the nM range.[19] OPG can also dimerize in the absence of the Cys400 linkage, forming noncovalent $[OPG^{SH}]_2$ complex.[19] The two death domains, as well as the tail domain, are thought to be involved in OPG dimerization.[19]

X-ray crystallography has provided some insights into the interaction of OPG with RANKL.[17,18] Unfortunately, this information only extends to truncated constructs because the full-length proteins have not been amenable to crystallographic investigations. Figures 1b and 1c depict the structure of a 3:3 complex between the RANKL receptor-binding domain (grey) and the CRD1-CRD4 regions of OPG (colored). The central RANKL trimer consists of three β-sandwiches that are linked by hydrophobic contacts. Each of the RANKL subunits interacts with one OPG chain. These interactions are mediated by segments ^{71}TDSWHTSDECVYCSPVCKEL^{90} and ^{111}RYLEIEFC^{118} in CRD2 and CRD3, respectively.[17] Figure 1c shows the remaining OPG regions (DD1, DD2, Tail) in cartoon form due to the lack of crystallographic information for these segments.[17] Given the propensity of OPG to dimerize,[8,10,18,19] the main physiological binding scenario likely involves two subunits of the RANKL trimer that interact with one OPG dimer.[18] The third RANKL subunit can bind a second OPG dimer, but likely with lower affinity since that interaction would only involve a single RANKL chain and a single OPG chain.[19]

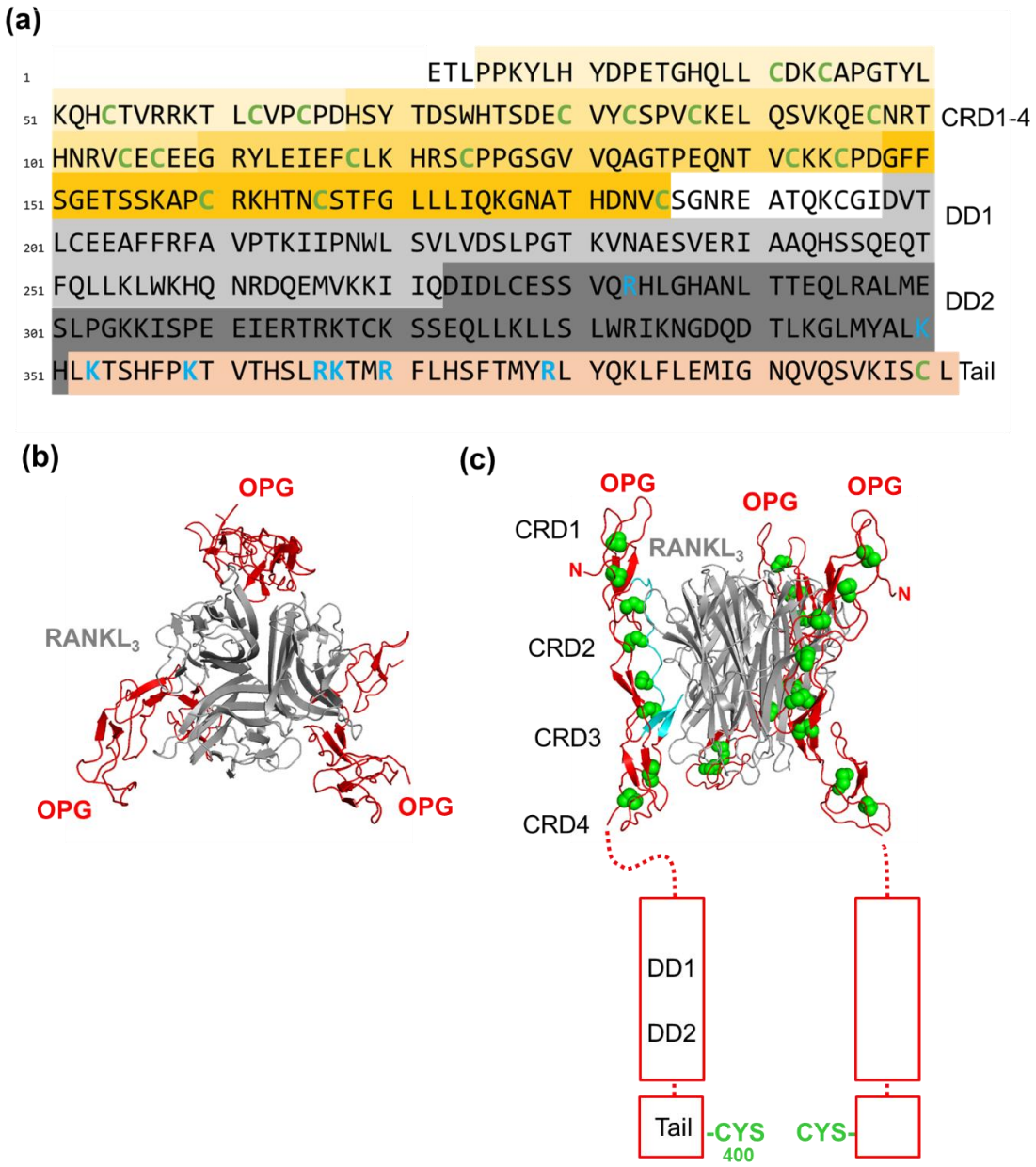


Figure 1. *(a) Mouse OPG sequence and domain architecture. Cys residues involved in disulfide bonds are shown in green. Basic residues in the DD2 and Tail domains are highlighted in blue, with those of particular importance for HS binding underlined.*[20] *Panels (b) and (c) show the X-ray structure of a complex consisting of three truncated OPG chains and trimeric truncated RANKL (pdb code 4E4D).*[17] *RANKL is depicted in gray, OPG is colored. Disulfides are highlighted in green. The crystallographic data available for OPG only cover the N-terminal cysteine-rich domains CRD1-CRD4. (b) View from the OPG N-terminal side. (c) Side view, with key contact regions highlighted in magenta for one of the OPG chains. Also shown in (c) are cartoons of the two OPG death domains (DD1 and DD2), and the tail domain with Cys400 that can form a disulfide with Cys400 of another OPG chain.*

OPG also interacts with heparan sulfate (HS),[21,22] a linear glycosaminoglycan that is present on the surface of virtually all cell types. HS is involved in a wide range of regulatory functions.[23-25] Of note, HS on the osteoblast surface participates in the regulation of bone remodeling.[21] The presence of glucuronate and sulfate groups renders HS negatively charged. HS-protein contacts are generally mediated by electrostatic interactions with cationic (Lys/Arg-rich) regions,[23,24,26] often resulting in conformational changes of the receptor protein.[27] The basic tail domain of OPG has traditionally been considered to be the HS binding site, with affinities in the nM range.[10,11,18,19,22,28]

Recent work (by our collaborator D.X.)[20] demonstrated that HS binding to OPG involves not only the tail domain but also parts of DD2, consistent with the fact that DD2 also contains a high percentage of basic residues (Figure 1a). HS binding was demonstrated for both OPG_2^{SS} and OPG^{SH}. The interaction of HS with OPG^{SH} resulted in dimerization, i.e., the formation of noncovalently bound $[OPG^{SH}]_2$. HS binding to OPG_2^{SS} triggered structural changes, evident from a ~10% reduction in R_g measured by small angle X-ray scattering (SAXS). The SAXS profiles of HS-bound OPG_2^{SS} and $[OPG^{SH}]_2$ were almost superimposable, suggesting similar structures. The presence of RANKL resulted in complexes containing HS, two OPG dimers, and one RANKL trimer. The data obtained were consistent with the view that HS at the osteoblast surface promotes OPG binding to RANKL via OPG sequestration and/or conformational changes.[20]

As noted, thus far crystallographic information is only available for the CRD domains in the N-terminal region of OPG. Those data provide some insights into the RANKL binding sites on OPG.[17,29] However, major knowledge gaps persist with respect to the OPG C-terminal domains. For example, the exact location of HS binding sites in the C-terminal region, and the spatial relationship between CRD domains and death/tail domains remain largely unknown. Also, differences related to the dimerization of OPG^{SH} and OPG_2^{SS}, as well as details of the dimerization interface remain obscure.[30] In addition, crystal structures may not always provide accurate structural information for solution phase protein-protein complexes due to crystal packing artifacts.[31] Therefore, new experimental methods need to be introduced to resolve these conundrums. Hydrogen-deuterium exchange (HDX) mass spectrometry (MS) has emerged as a powerful tool to examine the structure and dynamics

of proteins in solution.[32-37] HDX experiments monitor the exchange of backbone amide hydrogens with deuterium. Changes of exchange rates reflect the conformational alteration of protein influenced by ligand binding and/or protein-protein contacts.[38-44] Although HDX-MS generally benefits from X-ray crystallographic information, the presence of high resolution structural data is not a prerequisite for the successful application of HDX-MS. In this work, we apply HDX-MS to explore and analyze the OPG system in the presence of HS and/or RANKL. The data reveal that HS and RANKL have different effects on OPG^{SH} and OPG_2^{SS}. Based on the observed dynamics changes of OPG under different condition, we were able to pinpoint the HS-binding site on OPG. We also propose a model that describes how HS regulates OPG dimerization and RANKL binding: formation of $[OPG^{SH}]_2$, induced by HS binding, reduces the entropy penalty of OPG-RANKL complexation.

3.2 Experimental

3.2.1 Materials

Full-length mouse OPG was expressed in 293-freestyle cells (ThermoFisher Scientific). OPG_2^{SS} and OPG^{SH} were separated on a heparin-Sepharose column.[20] Stable trimers of $RANKL_{158-316}$ were expressed in *E. coli*. The proteins were purified as described,[20] yielding stock solutions of 8 mg mL^{-1} in 25 mM HEPES buffer with 150 mM NaCl at pH 7.1, and 10 mg mL^{-1} in 20 mM Tris buffer and 150 mM NaCl at pH 7.5, respectively. HS was supplied as heparin-derived dodecasaccharide (Sigma, St. Louis, MO). Tris-(2-carboxyethyl)-phosphine·HCl (TCEP) and other reagents were from Sigma as well.

3.2.2 HDX Samples

HDX samples were prepared in 25 mM HEPES, 50 mM NaCl and 85 % D_2O (HDX buffer) at pH_{read} 7.1 at room temperature (22 ± 1)° C. All samples contained 8 μM OPG_2^{SS} or 16 μM OPG^{SH}. RANKL trimer was added at a concentration of 7.5 μM, and HS at 11.4 μM. One the basis of nM literature K_d values[11,19] for interactions with HS and RANKL, OPG was ~99% saturated with its binding partners under HDX conditions. A stoichiometric excess of RANKL was chosen to suppress the presence of free OPG. This scenario precluded meaningful HDX measurements on RANKL because a significant fraction of RANKL chains remained unbound. The OPG response to ligand binding was probed by conducting HDX-MS on eight types of samples: (1) OPG^{SH} without binding partners, (2) OPG^{SH} + HS, OPG^{SH} + HS + RANKL, (4) OPG^{SH} + RANKL, (5) OPG_2^{SS} without binding partners, (6) OPG_2^{SS} + HS, (7) OPG_2^{SS} + HS + RANKL, (8) OPG_2^{SS} + RANKL. 10 μL aliquots were removed from the HDX solutions at 10 s, 1 min, 10 min, 50 min, 100 min, 200 min. In initial digestion experiments, we found that the high number of disulfides in the OPG CRD1-CRD4 region caused poor sequence coverage, a problem that is commonly encountered with SS-containing proteins.[45-48] After extensive testing and optimization we settled on a disulfide reduction strategy involving TCEP.[47,49] 10 μL aliquots were mixed in a 1:1 volume ratio with ice-cold quenching buffer (8 M urea and 1 M TCEP·HCl) at a measured pH of 2.3. The samples were then flash frozen in liquid nitrogen and stored at -80° C. Prior to analysis the samples were thawed to 0° C, and the liquid samples were kept on ice for 5 min for TCEP-mediated disulfide reduction. The samples were then diluted with three volumes of aqueous formic acid (pH 2.3) to lower the TCEP concentration for protecting the downstream pepsin column. The resulting 60 μL samples were analyzed as outlined below. The reduction/digestion workflow resulted in 60+ peptides for a sequence coverage of ~80%. The digestion patterns of OPG_2^{SS} and OPG^{SH} were slightly different. Although the use of TCEP significantly improved peptic digestion of the CRD regions, the number of peptides originating from these regions was significantly lower than in the DD1/DD2/Tail regions (Figures 2, 3). CRD regions contain 9 disulfide bonds which results in much lower peptide numbers compared to other regions. Zero-time point samples (m_0)

were produced by exposing pre-quenched OPG to D_2O labeling solution. Maximally exchanged controls (m_{100}) were generated by incubation of OPG in HDX solution for 24 h at pH 2.4 and 37 ˚C. In all other aspects, the m_0 and m_{100} samples were treated like the regular time points.

Figure 2. Pep*tic digestion map of OPG^{SH}. Number of peptides = 62, coverage = 75%, redundancy = 3.4. The sequence numbering shown here includes the N-terminal signal peptide (residues 1-21). The protein used for the experiments of the current study started at residue 22.*

Figure 3. *Peptic digestion map of OPG_2^{SS}. Number of peptides = 64, coverage = 86%, redundancy = 3.4. See comments regarding sequence numbering in the caption of Figure 2.*

3.2.3 HDX-MS Analysis

60 μL quenched aliquots were injected into a nanoACQUITY UPLC with HDX technology (Waters, Milford, MA) for digestion, desalting and peptide separation. Online digestions were performed on a POROS pepsin column (2.1 mm × 30 mm, Applied Biosystems, Carlsbad, CA) at 15˚C. The resulting peptides were trapped on a guard column (BEH C18 1.7 um, 2.1 × 5 mm) and separated on a reversed phase column (BEH C18 1.7 um, 1 × 100 mm) using a water/acetonitrile gradient in the presence of 0.1% formic acid at 40 μL min^{-1}. Peptide mass spectra was recorded on a Waters Synapt G2 instrument with source and desolvation temperatures of 80 and 250° C, respectively. The cone voltage was set at 20 V, and the electrospray voltage was 3 kV. The identity of each peptide was confirmed by MS^E and PLGS 2.4.3 (Waters) in initial control experiments conducted without D_2O. HDX kinetic profiles were analyzed by DynamX 3.0 (Waters) and plotted as centroid mass versus D_2O exposure time. Deuteration levels are reported as *HDX Percentage* = $(m_t - m_0)/(m_{100} - m_0)$. All experiments were performed in triplicate. Error bars represent standard deviations.

3.3 Results and Discussion

3.3.1 HDX Isotope Distributions

HDX-MS was employed to interrogate changes in the structure and dynamics of full-length OPG^{SH} and OPG_2^{SS} in response to HS and/or RANKL addition. As an example of unprocessed raw data, Figure 4 shows peptide 347-374, which covers parts of the DD2 and Tail regions. This peptide exhibited dramatically lowered HDX levels in the presence of HS and HS + RANKL for both OPG^{SH} and OPG_2^{SS}. The addition of RANKL alone did not

cause significant effects in this region. These data reaffirm that interactions with HS trigger major conformational changes in OPG,[20] as suggested in previous studies.[21,22] Similar to the behavior seen for all other peptides, Figure 4 reveals that OPG deuteration proceeds with unimodal H/D isotope distributions which reflect EX2 behavior. In other words, HDX is mediated by rapid H-bond opening/closing transition, with closing rate constants >> 1 s^{-1},[50] as commonly seen for many other proteins under native solvent conditions. [34-36]

The complete set of HDX kinetics recorded under all eight conditions is summarized in Figures 5 and 6. One striking feature in this data set are the relatively high exchange levels seen for the majority of the peptides. Typical globular proteins tend to possess rigid segments that remain incompletely deuterated even after hours of D_2O incubation.[32-37] OPG showed a very different behavior, as many peptides were ~50% labeled already after 10 s. After 2 h, deuteration had gone to completion for all peptides, even in the presence of ligands. This behavior reflects the highly dynamic nature of OPG which has thus far precluded crystallization of the full-length protein.[17,18] Difficulties of growing crystals from proteins that show high HDX levels have been noted previously.[51]

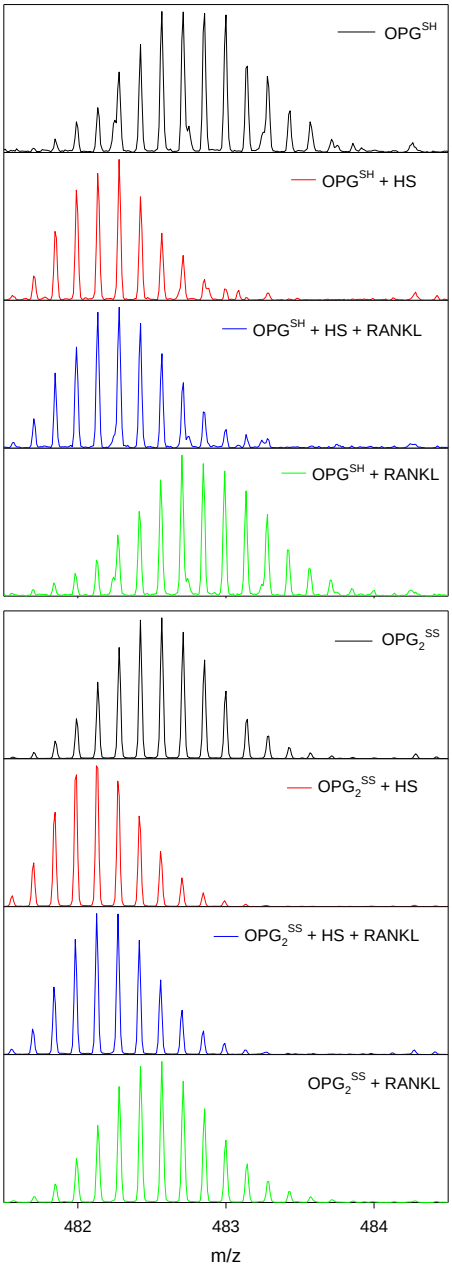


Figure 4. *HDX mass spectra of OPG^{SH} (top panels) and OPG_2^{SS} (bottom panels), representing the eight experimental conditions tested in this work. These data are for peptide 347YALKHLKTSHFPKTVTHSLRKTMRFLHS374 (covering the DD2/Tail region) for an HDX labeling time of 10 s.*

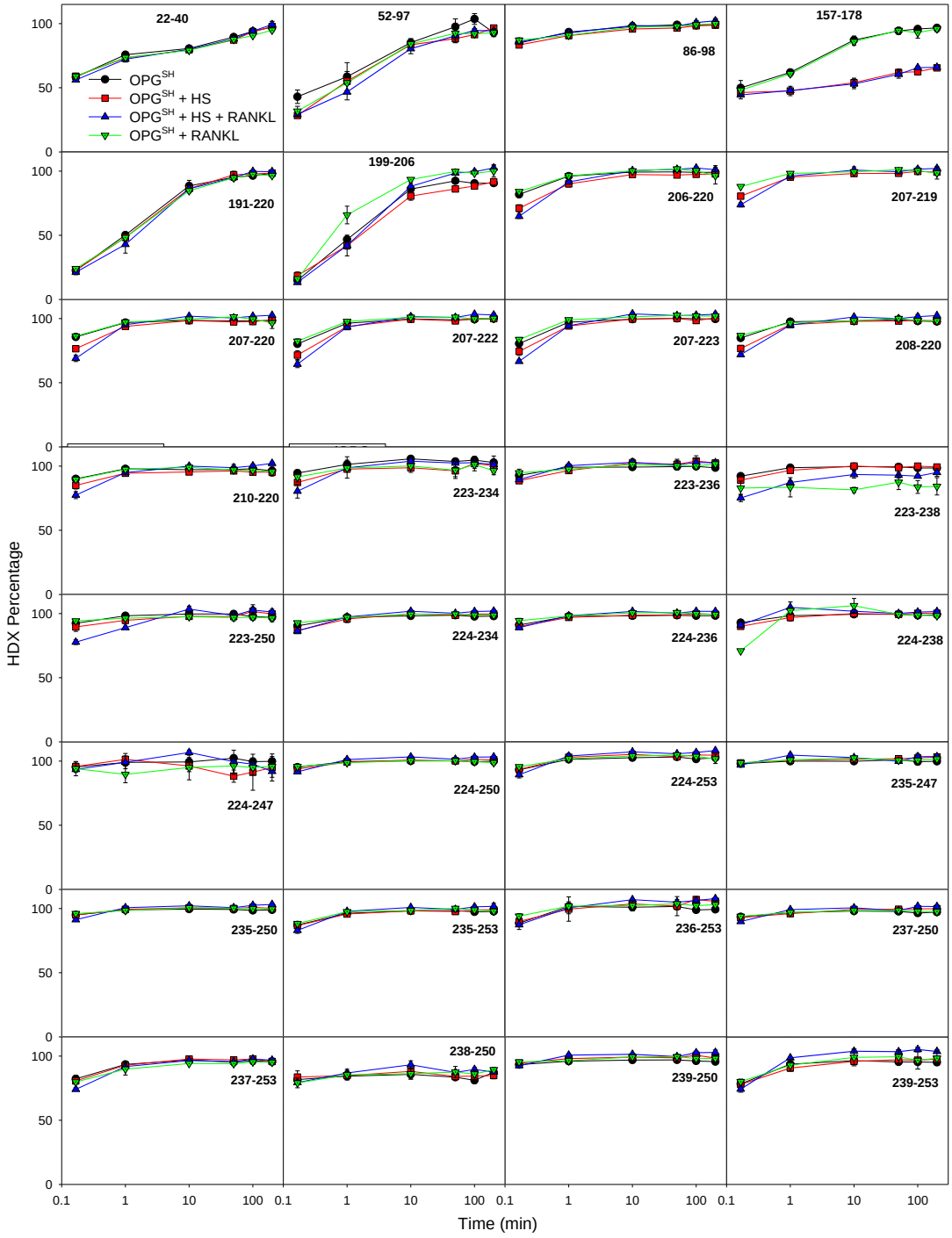


Figure 5. *Deuteration kinetics of OPG^{SH} under four different conditions (without ligands, with HS, with HS and RANKL, with RANKL), as noted in the legend. The figure is continued on the next page.*

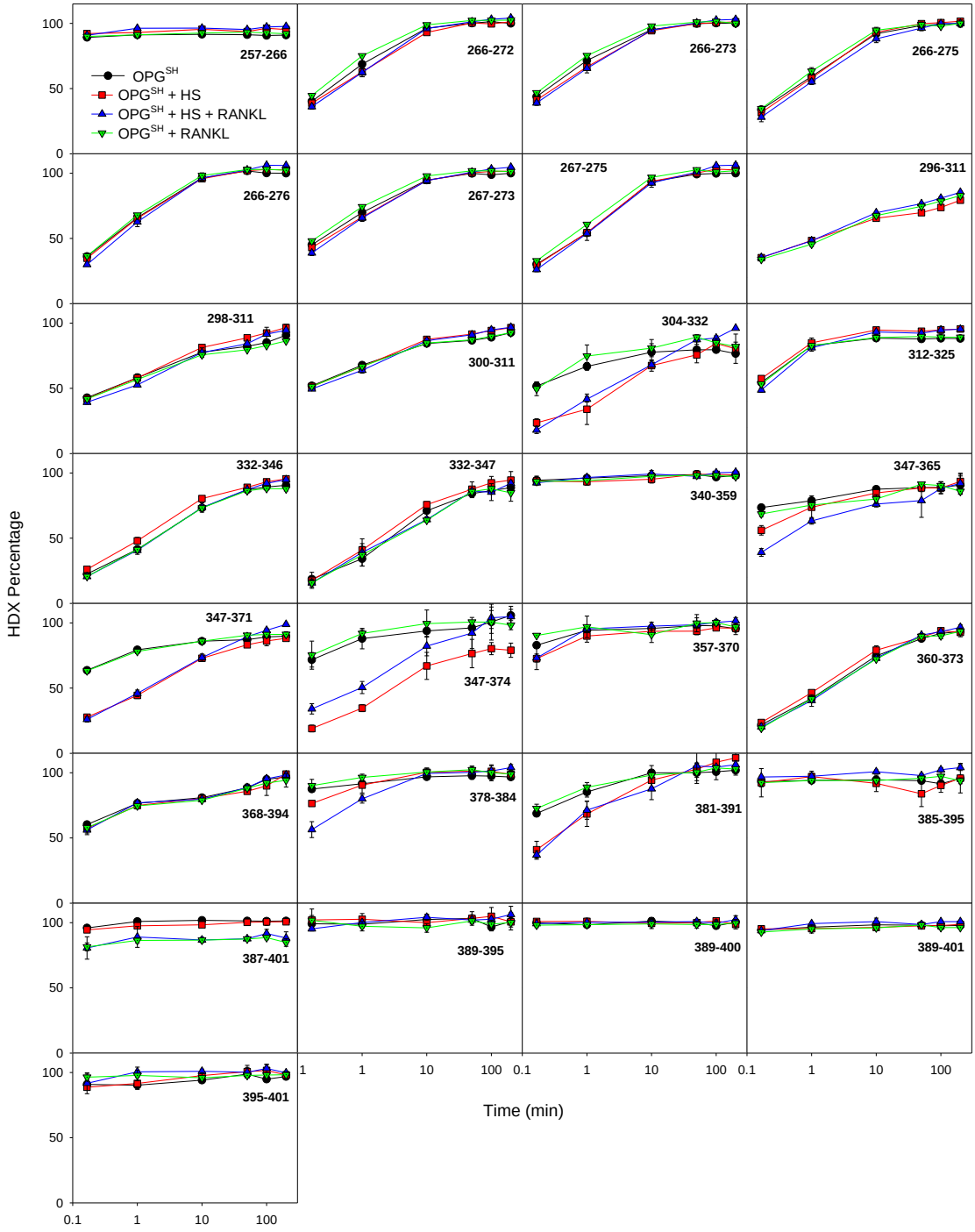


Figure 5. *OPG^{SH} deuteration kinetics – continuation from the previous page.*

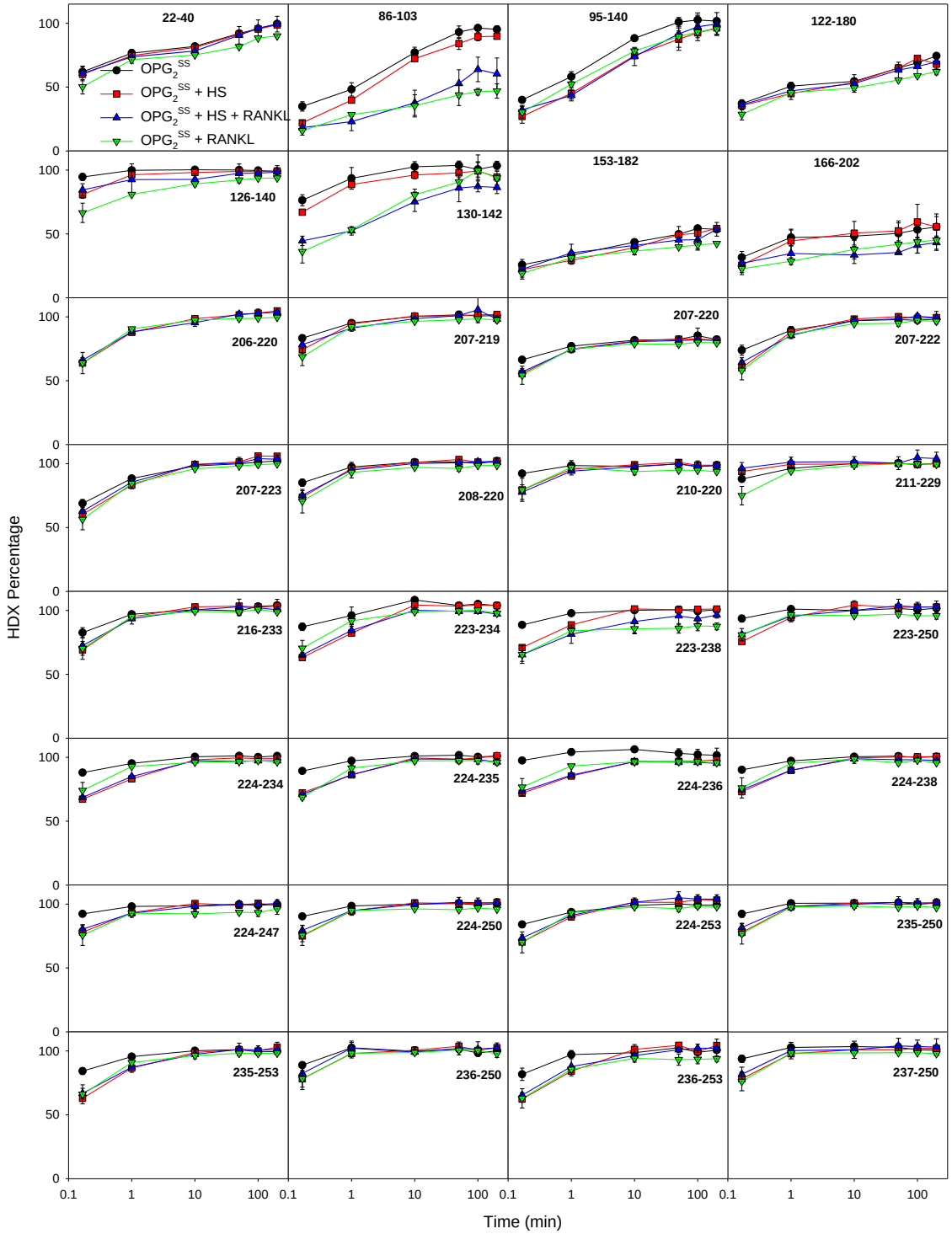


Figure 6. *Deuteration kinetics of OPG_2^{SS} under four different conditions (without ligands, with HS, with HS and RANKL, with RANKL), as noted in the legend.*

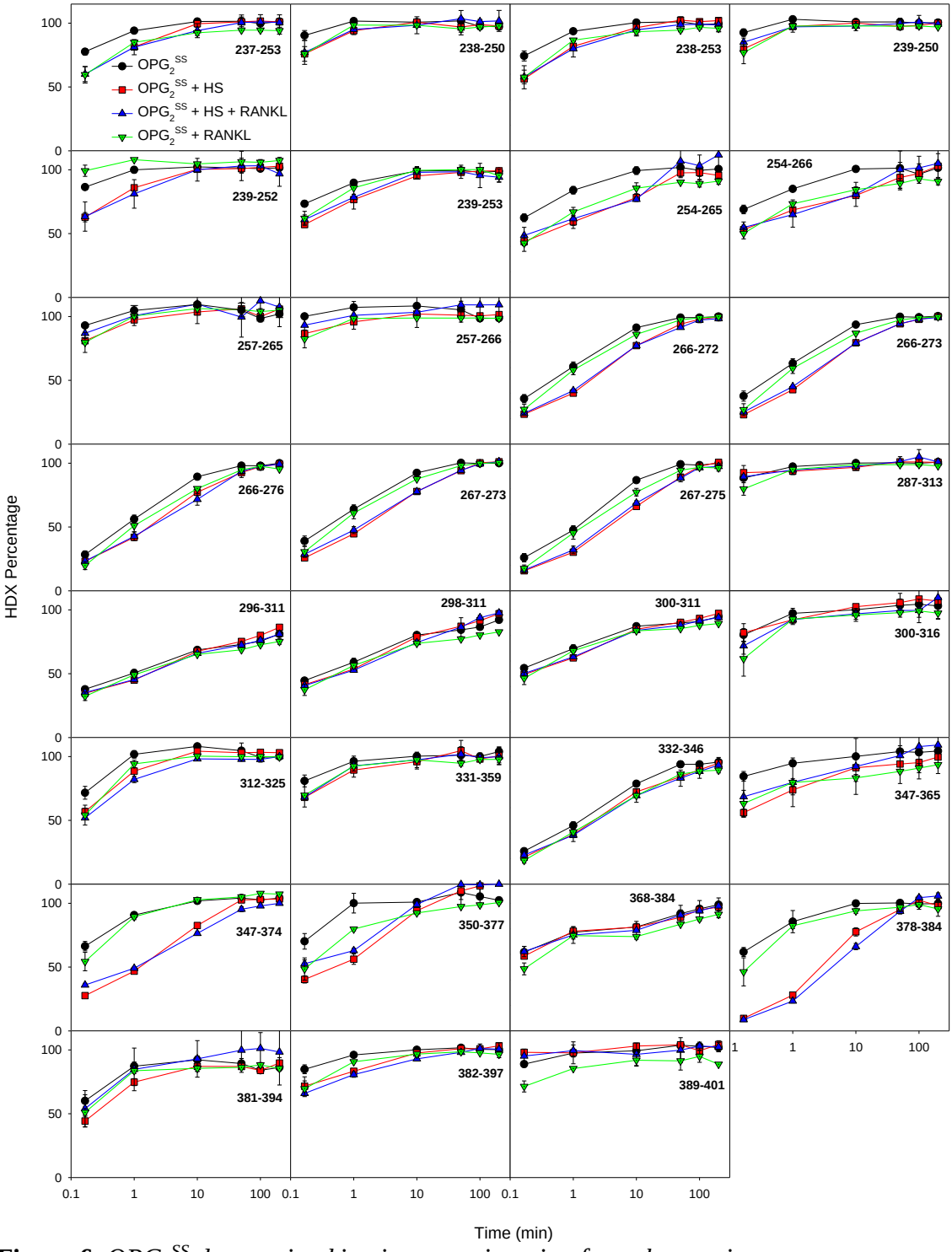


Figure 6. *OPG_2^{SS} deuteration kinetics – continuation from the previous page.*

3.3.2 Ligand-Induced Effects in the CRDs

Representative examples taken from the complete HDX data set (Figure 5, 6) are highlighted in Figure 7, illustrating the behavior of various regions in OPG^{SH} (left hand side) and OPG_2^{SS} (right hand side). As noted above, the peptic cleavage patterns were somewhat different for the two forms. In Figure 7 an effort was made to compare OPG^{SH} and OPG_2^{SS} peptides that cover approximately the same regions. We will discuss these data by moving from the N- to the C-terminus, i.e., top to bottom of Figure 7. The data interpretation is guided by the tenet that ligand-protein or protein-protein interactions often (though not always) cause reduced HDX levels. These HDX changes tend to be most pronounced directly at the interaction site, but allosteric effects may take place as well.[52-55]

CRD1 (Figure 7, peptide 22-40) was found to be quite dynamic, with high deuteration levels that were insensitive to addition of HS and/or RANKL. This behavior is consistent with the view that CRD1 is not involved in HS or RANKL binding, both in the case of OPG^{HS} and OPG_2^{SS}.[17,18,20-22]

CRD2 in OPG^{HS} appeared to be largely unstructured, regardless of the presence of HS or RANKL (Figure 7, 86-98) In OPG_2^{SS} this segment was much more protected even without ligands (Figure 7, peptide 86-103). This behavior strongly suggests the occurrence of noncovalent CRD2-CRD2' binding interactions in OPG_2^{SS}, promoted by the "pre-alignment" of the two OPG chains by the C-terminal disulfide. It is well established that SS-bonds can allosterically trigger distant protein-protein contacts by reducing the entropic penalty that would otherwise be associated with these protein-protein contacts.[56] Significant additional stabilization of CRD2 in OPG_2^{SS} was observed after addition of RANKL; this effect was independent of HS. Hence, OPG_2^{SS} binds RANKL in an HS-independent fashion, confirming earlier size exclusion chromatography data.[20] Very similar effects were seen for CRD3 (Figure 7, 130-142). These findings are consistent with the crystallographic data that identified CRD2 and CRD3 as RANKL binding region.[17,18] Importantly, the structural data of Figure 1 imply that binding of the two OPG chains in

OPG_2^{SS} to RANKL is incompatible with the presence of CRD2-CRD2' contacts.[17,18] Our data thus imply that these contacts in OPG_2^{SS} have to dissociate for RANKL binding to take place.

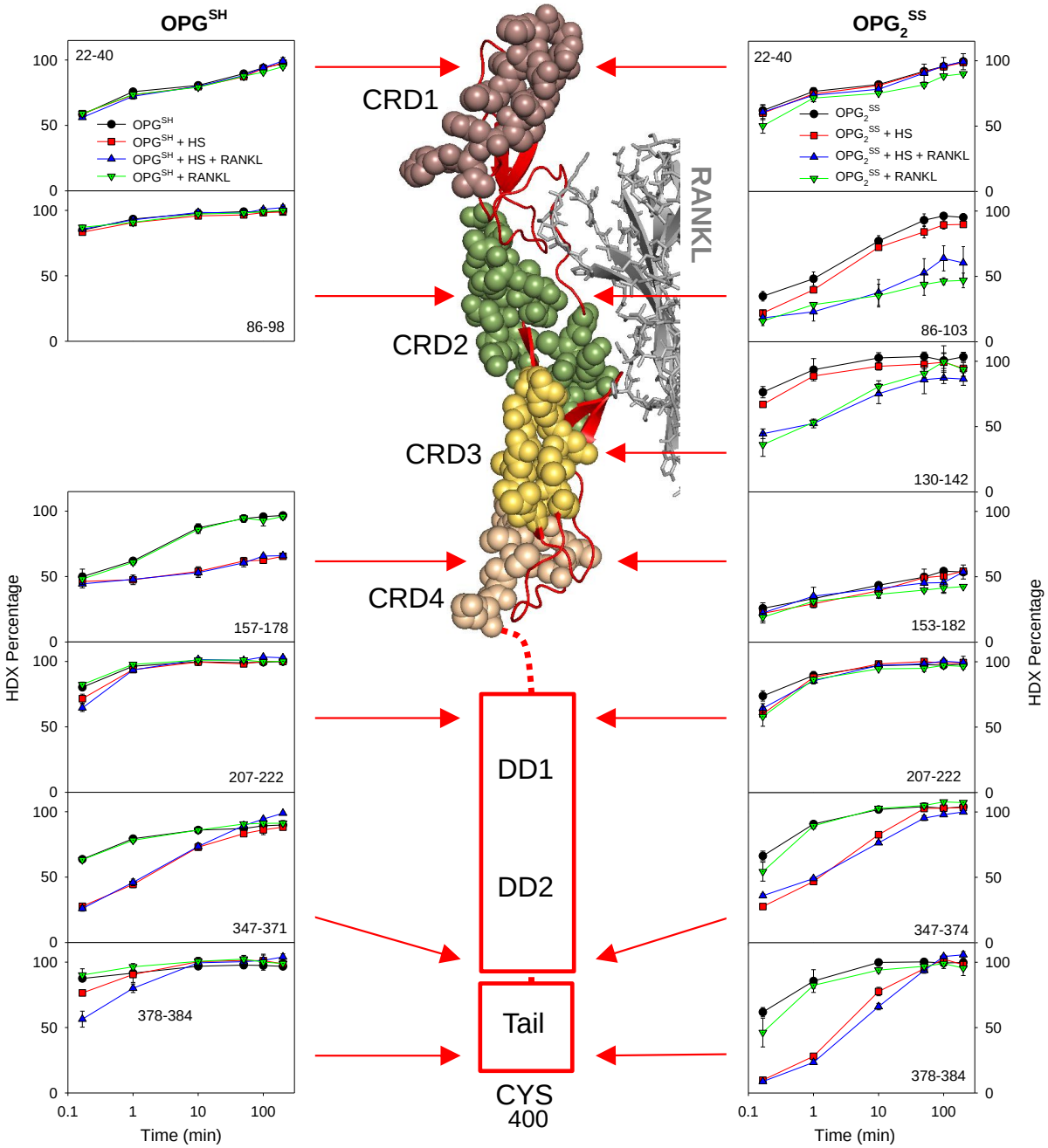


Figure 7. *HDX difference plots for a labeling time of t = 10 s. Panels (a) - (c): HDX changes for OPG^{SH} relative to ligand-free OPG^{SH}. Negative values indicate less deuteration after addition of binding partners, i.e., HS and/or RANKL. (d) Comparison of ligand-free OPG^{SH} and ligand-free OPG_2^{SS}. Panels (e) – (g): HDX changes for OPG_2^{SS} relative to ligand-free OPG_2^{SS}.*

For OPG^{SH} our data do not indicate a strong involvement of CRD2 in RANKL binding, despite minor binding-related effects in CRD2 peptide 52-97 (Figure 7). Hence, CRD3 seems to be the dominant RANKL binding element in OPG^{SH}. Sadly, CRD3 is not covered in the peptide map of OPG^{SH}, such that these contacts cannot be verified directly. Nonetheless, from the deuteration behavior of CRD2 (discussed above), it is clear that the OPG-RANKL binding interactions must be somewhat different for OPG^{SH} and OPG_2^{SS}. In the former case, RANKL contacts are formed mainly with CRD2. For OPG_2^{SS} both CRD2 and CRD3 mediate binding, as seen in the crystal structures.[17,18]

For CRD4 in OPG^{SH}, deuteration was significantly suppressed by HS, while RANKL did not have any effects. Even stronger protection was seen for CRD4 in OPG_2^{SS}, regardless of the solution conditions (Figure 7, 157-178 and 153-182). We attribute this behavior to HS-induced dimerization of OPG^{SH}.[20] The data of Figure 7 strongly suggest that these $[OPG^{SH}]_2$ species interact through CRD4-CRD4' contacts. The relatively strong protection seen in this region for OPG_2^{SS} (regardless of ligands) suggests that similar CRD4-CRD4' contacts are also triggered by the presence of the Cys400 disulfide bond. The lack of such CRD4-CRD4' contacts in the available X-ray structures likely reflects the absence of C-terminal linkages in the truncated crystal constructs.[17,18] Formation of such CRD4-CRD4'contacts represents an allosteric effect, as it is caused by distant factors, i.e., HS binding the case of OPG^{SH}, and the presence of a C-terminal SS bond in the case of OPG_2^{SS}. As noted above, these C-terminal contacts "pre-align" the two OPG chains and thereby lower the entropic penalty associated with the formation of noncovalent intra-chain bonds.[56]

3.3.3 Ligand Effects in the Non-Crystallized Regions

DD1 and DD2 were relatively disordered in both OPG^{SH} and OPG_2^{SS}, regardless of the solution conditions (Figure 7, 207-222). Close inspection of the HDX profiles in these regions nonetheless reveals interesting features. OPG_2^{SS} showed slightly more protection after addition of HS and/or RANKL. Considering the small magnitude of these changes,

they likely do not represent direct ligand-protein interactions, but weak allosteric stabilization from distant binding events[57-60] (i.e., RANKL binding to CRD2 and CRD3, as well as HS binding closer to the C-terminus). DD1 in OPG^{SH} did not respond to isolated RANKL, while HS caused a marked reduction in deuteration for early labeling times (Figure 7, 207-222). Additional stabilization was observed upon RANKL addition to the HS-containing sample. This stabilization likely originates from HS-induced RANKL binding to CRD3 (see above).

HS significantly suppressed deuteration in the range of residues 350-384 for OPG^{SH} and OPG_2^{SS}, consistent with electrostatically-mediated HS binding in this region (Figure 7, 347-371).[20] Farther towards the C-terminus, the Cys400-Cys400' linkage caused a marked stabilization for OPG_2^{SS}, with additional stabilization after HS binding (Figure 7, 378-384). In ligand-free OPG^{SH} this segment was much more dynamic, with slight stabilization by HS and a surprisingly large additional stabilization under HS + RANKL conditions.

3.3.4 HDX Difference Plots.

A complementary perspective of the effects associated with OPG ligand binding and dimerization can be obtained by studying HDX differences for all the available peptides. We will focus on data for t = 10 s (Figure 8), where many of the distinguishing features were more pronounced than at later time points (Figure 5, 6). The most prominent feature in the difference plots are the negative values in the DD2 and Tail regions of both OPG^{SH} and OPG_2^{SS} (Figure 8, a, b, e, f), arising from HS binding to these regions.[20] RANKL binding to OPG^{SH} caused slightly reduced deuteration in CRD2 (Figure 8b, c). More pronounced effects were evident for CRD3 in OPG_2^{SS} (8f, g). A direct comparison between ligand-free OPG^{SH} and OPG_2^{SS} reveals an intricate pattern, consistent with slight stabilization of the Tail and DD1 regions after formation of the Cys400 disulfide bond (Figure 8d). Two possible reasons can cause this effect: First, the Tail domain could be part of the dimerization interface since it is reasonable to deduce the dimerization interface includes the adjacent region of cys400. However, DD1 domain is unlikely also to be a part

of the dimerization interface. The stoichiometry between OPG_2^{SS} and RANKL trimer is more favorable at 1:1 ratio[19], suggesting each monomer in the OPG_2^{SS} would bind to separated binding sites on the same RANKL trimer. The distance between two monomers is over 50 Å. It is implausible that two DD1 domains on different monomers remain in contact while their neighbor domains are separated by such a large distance. That makes the stabilization on DD1 must due to an allosteric effect induced by dimerization.

The sparse appearance of this data set arises from differences in the OPG^{SH} and OPG_2^{SS} digestion patterns, keeping in mind that deuteration differences can only be calculated for peptides that are exactly matched therefore not all peptic HDX data is presented in Figure 8.

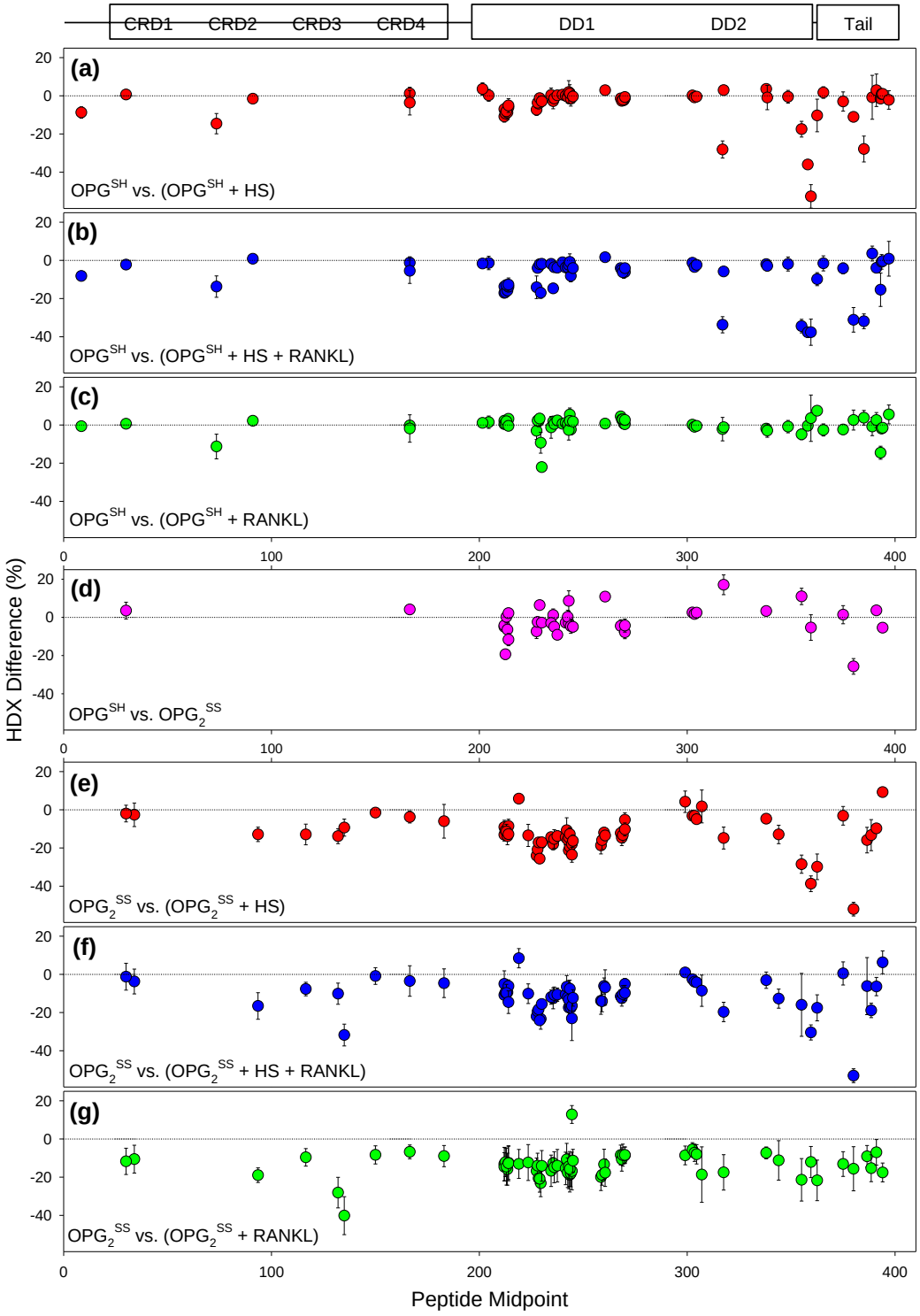


Figure 8. *HDX difference plots for a labeling time of t = 10 s. Panels (a) - (c): HDX changes for OPG^{SH} relative to ligand-free OPG^{SH}. Negative values indicate less deuteration after addition of binding partners, i.e., HS and/or RANKL. (d) Comparison of ligand-free OPG^{SH} and ligand-free OPG_2^{SS}. Panels (e) – (g): HDX changes for OPG_2^{SS} relative to ligand-free OPG_2^{SS}.*

3.4 Conclusions

Figure 9 provides a structural interpretation of our HDX-MS data in cartoon form. With no ligand (HS and/or RANKL) present, OPG^{SH} remains monomeric and highly dynamic, especially in the CRD domains that contain the RANKL binding region. (Figure 9a) Under these conditions, formation of a OPG-RANKL complex would be associated with a large entropic penalty. This might be the reason why OPG^{SH} has a lower affinity compared to OPG_2^{SS}. In contrast, HS-induced $[OPG^{SH}]_2$ formation proceeds with HS-mediated contacts in the tail domain. Under these conditions the two OPG^{SH} chains are pre-aligned for ligand binding. This pre-alignment is triggered by HS binding in the tail domain, but it is synergistically supported by the establishment of CRD4-CRD4' contacts (Figure 7, peptide 157-178), likely via hydrophobic interactions. The formation of $[OPG^{SH}]_2$ complex reduces the entropy penalty of subsequent RANKL binding. In this way, HS can regulate the binding affinity of OPG^{SH} (or $[OPG^{SH}]_2$) for RANKL. Interestingly, once RANKL is bound to the HS-containing $[OPG^{SH}]_2$, it appears to stabilize the HS binding region even further (Figure 7). Hence, the presence of HS favors RANKL binding, while conversely the presence of RANKL favors HS binding. This mutual stabilization of two distinct binding partners will become even more evident in the following Chapter that deals with calcium-mediated peptide interactions in a S100 protein.

Figure 9b illustrates the proposed binding pattern for OPG_2^{SS} to RANKL/HS. In contrast to OPG^{SH}, OPG_2^{SS} has already been "aligned" by dimerization. Therefore, the binding of HS does not strengthen the RANKL binding affinity of OPG_2^{SS}. However, similar to OPG^{SH}, HS also rigidify OPG_2^{SS} N-terminal domains which includes Tail domain, DD1 domain. DD2 domain is also affected but to a lesser extent (Figure 7 and 8). The region with suppressed deuteration correspond to the HS binding region, in agreement previous site-directed mutagenesis experiments.[30] Another interesting feature evident from our data is that non-covalent CRD2-CRD2' interactions in OPG_2^{SS} have to dissociate first in order to allow OPG_2^{SS} interactions with RANKL. In the absence of these dissociation events the

RANKL binding site would be blocked, as evident from the X-ray structural data of Figure 1.

In conclusion, this study reports new structural insights into the RANKL/OPG/HS regulatory system. It can be expected that these new revelations can lead to the discovery of the more detailed regulatory mechanism of OPG in the human body and offer the opportunity to develop therapeutic approaches for tackling osteoporosis and related disorders.

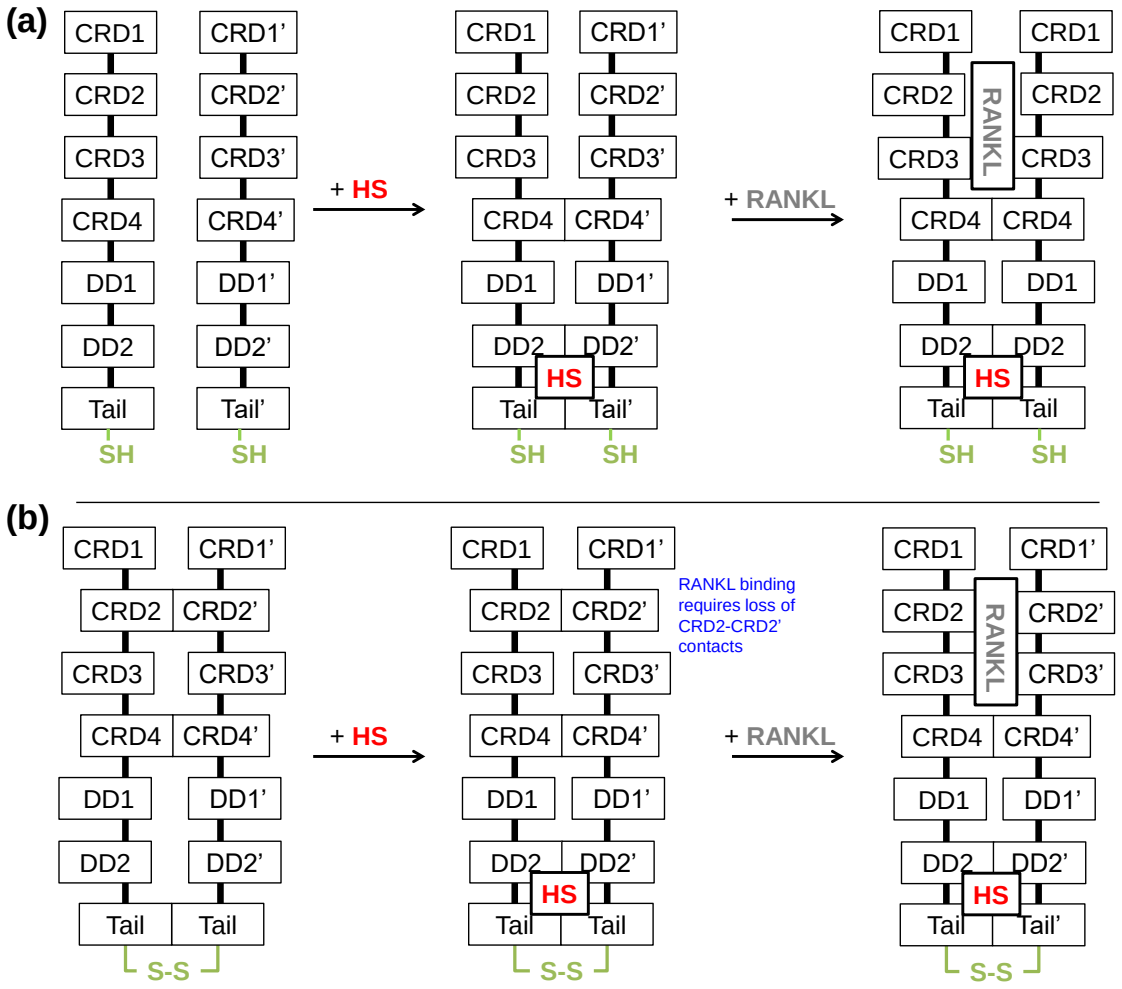


Figure 9. *Summary of proposed changes in structure and dynamics of OPG in response to HS and RANKL binding. (a) HS binding to* OPG^{SH} *triggers formation of a* $[OPG^{SH}]_2$ *complex, this complex subsequently binds RANKL. (b) HS binding to* OPG_2^{SS} *promotes stabilization and inter-chain contacts; subsequent RANKL binding requires some of these contacts (CRD2-CRD2') to dissociate. Each panel shows two OPG chains, domains are represented as boxes. Note how some of these boxes are in direct contact with one another (indicating noncovalent inter-domain contacts), while others are separated by a wide gap (representing weaker or no contacts).*

3.5 References

(1) Datta, H. K.; Ng, W. F.; Walker, J. A.; Tuck, S. P.; Varanasi, S. S. *J. Clin. Pathol.* **2008**, *61*, 577.

(2) Robling, A. G.; Castillo, A. B.; Turner, C. H. In *Annu. Rev. Biomed. Eng*. 2006; Vol. 8, p 455.

(3) Karsenty, G.; Kronenberg, H. M.; Settembre, C. In *Annu. Rev. Cell Devel. Biol.* 2009; Vol. 25, p 629.

(4) Dallas, S. L.; Prideaux, M.; Bonewald, L. F. *Endocrine Rev.* **2013**, *34*, 658.

(5) Teitelbaum, S. L. *Am. J. Pathol.* **2007**, *170*, 427.

(6) Khosla, S.; Oursler, M. J.; Monroe, D. G. *Trends Endocrinol. Metabol.* **2012**, *23*, 576.

(7) Sobacchi, C.; Schulz, A.; Coxon, F. P.; Villa, A.; Helfrich, M. H. *Nat. Rev. Endocrin.* **2013**, *9*, 522.

(8) Simonet, W. S.; Lacey, D. L.; Dunstan, C. R.; Kelley, M.; Chang, M. S.; Luthy, R.; Nguyen, H. Q.; Wooden, S.; Bennett, L.; Boone, T.; Shimamoto, G.; DeRose, M.; Elliott, R.; Colombero, A.; Tan, H. L.; Trail, G.; Sullivan, J.; Davy, E.; Bucay, N.; RenshawGegg, L.; Hughes, T. M.; Hill, D.; Pattison, W.; Campbell, P.; Sander, S.; Van, G.; Tarpley, J.; Derby, P.; Lee, R.; Boyle, W. J. *Cell* **1997**, *89*, 309.

(9) Udagawa, N.; Takahashi, N.; Yasuda, H.; Mizuno, A.; Itoh, K.; Ueno, Y.; Shinki, T.; Gillespie, M. T.; Martin, T. J.; Higashio, K.; Suda, T. *Endocrinology* **2000**, *141*, 3478.

(10) Yamaguchi, K.; Kinosaki, M.; Goto, M.; Kobayashi, F.; Tsuda, E.; Morinaga, T.; Higashio, K. *J. Biol. Chem.* **1998**, *273*, 5117.

(11) Baud'huin, M.; Duplomb, L.; Teletchea, S.; Lamoureux, F.; Ruiz-Velasco, C.; Maillasson, M.; Redini, F.; Heymann, M. F.; Heymann, D. *Cytokine Growth Factor Rev.* **2013**, *24*, 401.

(12) Lacey, D. L.; Timms, E.; Tan, H. L.; Kelley, M. J.; Dunstan, C. R.; Burgess, T.; Elliott, R.; Colombero, A.; Elliott, G.; Scully, S.; Hsu, H.; Sullivan, J.; Hawkins, N.; Davy, E.; Capparelli, C.; Eli, A.; Qian, Y. X.; Kaufman, S.; Sarosi, I.; Shalhoub, V.; Senaldi, G.; Guo, J.; Delaney, J.; Boyle, W. J. *Cell* **1998**, *93*, 165.

(13) Yasuda, H.; Shima, N.; Nakagawa, N.; Yamaguchi, K.; Kinosaki, M.; Mochizuki, S.; Tomoyasu, A.; Yano, K.; Goto, M.; Murakami, A.; Tsuda, E.; Morinaga, T.; Higashio, K.; Udagawa, N.; Takahashi, N.; Suda, T. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 3597.

(14) Boyce, B. F.; Xing, L. *Arthritis Research & Therapy* **2007**, *9*, S1.

(15) Boyle, W. J.; Simonet, W. S.; Lacey, D. L. *Nature* **2003**, *423*, 337.

(16) Hofbauer, L. C.; Schoppet, M. *J. Am. Med. Assoc.* **2004**, *292*, 490.

(17) Nelson, Christopher A.; Warren, Julia T.; Wang, Michael W. H.; Teitelbaum, Steven L.; Fremont, Daved H. *Structure* **2012**, *20*, 1971.

(18) Luan, X.; Lu, Q.; Jiang, Y.; Zhang, S.; Wang, Q.; Yuan, H.; Zhao, W. J.; Wang, J.; Wang, X. *J. Immunol.* **2012**, *189*, 245.

(19) Schneeweis, L. A.; Willard, D.; Milla, M. E. *J. Biol. Chem.* **2005**, *280*, 41155.

(20) Li, M.; Yang, S.; Xu, D. *J. Biol. Chem.* **2016**, *291*, 24160.

(21) Mansouri, R.; Jouan, Y.; Hay, E.; Blin-Wakkach, C.; Frain, M.; Ostertag, A.; Le Henaff, C.; Marty, C.; Geoffroy, V.; Marie, P. J.; Cohen-Solal, M.; Modrowski, D. *Cell Death Dis.* **2017**, *8*.

(22) Theoleyre, S.; Tat, S. K.; Vusio, P.; Blanchard, F.; Gallagher, J.; Ricard-Blum, S.; Fortun, Y.; Padrines, M.; Redini, F.; Heymann, D. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 460.

(23) Xu, D.; Esko, J. D. *Annu. Rev. Biochem.* **2014**, *83*, 129.

(24) Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030.

(25) Kraemer, P. M. *Biochemistry* **1971**, *10*, 1437.

(26) Minsky, B. B.; Dubin, P. L.; Kaltashov, I. A. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 758.

(27) Jin, L.; Abrahams, J. P.; Skinner, R.; Petitou, M.; Pike, R. N.; Carrell, R. W. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14683.

(28) Willard, D.; Chen, W. J.; Barrett, G.; Blackburn, K.; Bynum, J.; Consler, T.; Hoffman, C.; Horne, E.; Iannone, M. A.; Kadwell, S.; Parham, J.; Ellis, B. *Protein Expr. Purif.* **2000**, *20*, 48.

(29) Luan, X.; Lu, Q.; Jiang, Y.; Zhang, S.; Wang, Q.; Yuan, H.; Zhao, W.; Wang, J.; Wang, X. *The Journal of Immunology* **2012**, *189*, 245.

(30) Li, M.; Yang, S.; Xu, D. *J. Biol. Chem.* **2016**, *291*, 24160.

(31) Alexopoulos, J. A.; Guarnéa, A.; Ortega, J. *J. Struct. Biol.* **2012**, *179*, 202.

(32) Rand, K. D.; Zehl, M.; Jorgensen, T. J. D. *Acc. Chem. Res.* **2014**, *47*, 3018.

(33) Marciano, D. P.; Dharmarajan, V.; Griffin, P. R. *Curr. Op. Struct. Biol.* **2014**, *28*, 105.

(34) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta* **2012**, *721*, 7.

(35) Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'Cuk, A. I.; Kaltashov, I. A. *Protein Sci.* **2005**, *14*, 543.

(36) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158.

(37) Maaty, W. S.; Weis, D. D. *J. Am. Chem. Soc.* **2016**, *138*, 1335.

(38) Konermann, L.; Pan, J.; Liu, Y. H. *Chem. Soc. Rev.* **2011**, *40*, 1224.

(39) Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'cuk, A. I.; Kaltashov, I. A. *Protein Sci.* **2005**, *14*, 543.

(40) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158.

(41) Rand, K. D.; Zehl, M.; Jorgensen, T. J. *Acc. Chem. Res.* **2014**, *47*, 3018.

(42) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta* **2012**, *721*, 7.

(43) Marciano, D. P.; Dharmarajan, V.; Griffin, P. R. *Curr. Opin. Struct. Biol.* **2014**, *28*, 105.

(44) Maaty, W. S.; Weis, D. D. *J. Am. Chem. Soc.* **2016**, *138*, 1335.

(45) Wongkongkathep, P.; Li, H.; Zhang, X.; Loo, R. R. O.; Julian, R. R.; Loo, J. A. *Int. J. Mass Spectrom.* **2015**, *390*, 137.

(46) Trabjerg, E.; Jakobsen, R. U.; Mysling, S.; Christensen, S.; Jorgensen, T. J. D.; Rand, K. D. *Anal. Chem.* **2015**, *87*, 8880.

(47) Zhang, H. M.; McLoughlin, S. M.; Frausto, S. D.; Tang, H. L.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **2010**, *82*, 1450.

(48) Zhang, J.; Loo, R. R. O.; Loo, J. A. *Int. J. Mass Spectrom.* **2015**, *377*, 546.

(49) Zhao, D. S.; Gregorich, Z. R.; Ge, Y. *Proteomics* **2013**, *13*, 3256.

(50) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.

(51) Pantazatos, D.; Kim, J. S.; Klock, H. E.; Stevens, R. C.; Wilson, I. A.; Lesley, S. A.; Woods, V. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 751.

(52) Bobst, C. E.; Zhang, M.; Kaltashov, I. A. *J. Mol. Biol.* **2009**, *388*, 954.

(53) Sowole, M. A.; Alexopoulos, J. A.; Cheng, Y.-Q.; Ortega, J.; Konermann, L. *J. Mol. Biol.* **2013**, *425*, 4508.

(54) Burke, J. E.; Babakhani, A.; Gorfe, A. A.; Kokotos, G.; Li, S.; Woods, V. L.; McCammon, J. A.; Dennis, E. A. *J. Am. Chem. Soc.* **2009**, *131*, 8083.

(55) Shukla, A. K.; Westfield, G. H.; Xiao, K.; Reis, R. I.; Huang, L.-Y.; Tripathi-Shukla, P.; Qian, J.; Li, S.; Blanc, A.; Oleskie, A. N.; Dosey, A. M.; Su, M.; Liang, C.-R.; Gu, L.-L.; Shan, J.-M.; Chen, X.; Hanna, R.; Choi, M.; Yao, X. J.; Klink, B. U.; Kahsai, A. W.; Sidhu, S. S.; Koide, S.; Penczek, P. A.; Kossiakoff, A. A.; Woods Jr, V. L.; Kobilka, B. K.; Skiniotis, G.; Lefkowitz, R. J. *Nature* **2014**, *512*, 218.

(56) Creighton, T. E. *Proteins*; W. H. Freeman & Co: New York, 1993.

(57) Deredge, D.; Li, J. W.; Johnson, K. A.; Wintrode, P. L. *J. Biol. Chem.* **2016**, *291*, 10078.

(58) Donovan, K. A.; Zhu, S. L.; Liuni, P.; Peng, F.; Kessans, S. A.; Wilson, D. J. *J. Biol Chem.* **2016**, *291*, 9244.

(59) Underbakke, E. S.; Iavarone, A. T.; Chalmers, M. J.; Pascal, B. D.; Novick, S.; Griffin, P. R.; Marletta, M. A. *Structure* **2014**, *22*, 602.

(60) Sheff, J. G.; Farshidfar, F.; Bathe, O. F.; Kopciuk, K.; Gentile, F.; Tuszynski, J.; Barakat, K.; Schriemer, D. C. *Mol. Cell. Proteomics* **2017**, *16*, 428.

Chapter 4. Calcium-Mediated Control of S100 Proteins: Allosteric Communication via an Agitator/Signal Blocking Mechanism

4.1 Introduction

Signaling proteins possess specific sites for binding their biological targets, such as nucleic acids or other proteins. Many of these interactions are controlled by allosteric effectors,[1-3] i.e., species that modulate the protein's target affinity by interacting with locations remote from the target binding site.[4-6] The dysregulation of allosteric interactions can lead to cancer and other diseases.[7] Also, allostery is a central element of drug action mechanisms.[8,9]

Classical models of allostery envision that effector binding triggers conformational changes that cause distant target binding sites to open or close.[1,2,4] Recent studies have expanded this view by emphasizing the role of conformational dynamics.[10-12] For example, it has been proposed that proteins fluctuate between co-existing conformers, and that allosteric effectors shift these equilibria towards structures with higher or lower target affinities.[6,13-16] Regardless of the exact mechanism, allostery requires the transmission of signals to target binding sites, often over tens of Ångstroms.[11,15] Such signals travel along dynamically coupled residues.[16,18,19] Allosteric signal propagation pathways have been identified on the basis of mutational[17] and evolutionary data,[18] NMR spectroscopy,[16,19-22] X-ray crystallography,[1,23,24] molecular dynamics (MD) simulations,[6,9,25-30] and hydrogen-deuterium exchange (HDX) mass spectrometry (MS).[31-34] Even after allosteric pathways have been identified, however, the exact mechanisms by which remote sites communicate often remains unclear.

Calcium represents one of the most common allosteric regulators. The intracellular Ca^{2+} concentration under resting conditions is low, but it rises sharply following stimuli that cause the opening of membrane channels. This Ca^{2+} influx can activate signaling proteins, allowing them to bind their targets. Subsequent deactivation relies on membrane pumps that return the calcium concentration to its resting value, such that protein-target complexes dissociate again.[35]

A particularly important group of calcium-dependent signaling proteins is the S100 family which is involved in numerous functions and diseases.[36-38] Here we focus on S100A11, a typical homo-dimeric S100 family member that binds four Ca^{2+} and two target proteins.[39] One of these targets[40,41] is annexin A2, a protein that interacts with the plasma membrane.[42] Calcium-loaded S100A11 can link two annexin A2 chains, thereby forming a complex required for membrane repair, specifically in cancer cells.[43,44]

Each S100A11 subunit comprises four helices. The helix III/IV loop forms a high-affinity EF-hand calcium binding site. A lower affinity pseudo EF-hand is formed by the helix I/II loop (Figure 1a).[35,45] In the Ca^{2+}-bound form helices III/IV adopt a near-perpendicular orientation that results in an open target binding site. Figure 1b shows the structure of S100A11 with four Ca^{2+} and two annexin peptides.[39] We will refer to this complex as [Ca_4 S100 Ax_2]. Square brackets indicate components that are bound in a complex, the annexin peptide is denoted as Ax, and "A11" is omitted to streamline the notation. In apo-S100, helices III/IV are in a more antiparallel orientation, resulting in a closed target binding site that precludes interactions with Ax (Figure 1c).[46] Similar transitions take place for many other Ca^{2+}-dependent proteins.[35,47]

The cooperative nature of allosteric regulation can be illustrated by using a thermodynamic cycle.[48-50] Figure 1d considers two possible pathways for the conversion of apo-S100 to [Ca_4 S100 Ax_2]. Steps 1 and 2 represents the "canonical" scenario where metalation takes place first, followed by Ax binding. In this case calcium enhances the protein's Ax binding affinity by $\Delta\Delta G° = -RT \ln C_{Ca}$, where the factor $C_{Ca} > 1$ reflects how the Ax binding equilibrium responds to the presence of calcium.[48,49] Alternatively, one may consider a pathway where Ax binds first (step 3), followed by metalation (step 4). In this second

scenario, Ax enhances the metal binding affinity by $\Delta\Delta G° = -RT \ln C_{Ax}$.[48,49,51] Thus, Ca^{2+} binding enhances the target affinity, and target binding enhances the Ca^{2+} affinity. This implies that allosteric signals must be able to travel from effector sites to target binding sites and *vice versa*. Both scenarios are associated with the same affinity enhancement, because $C_{Ca} = C_{Ax} = C$.[48-50] For the system considered here $C \approx 10$.[46,51-53]

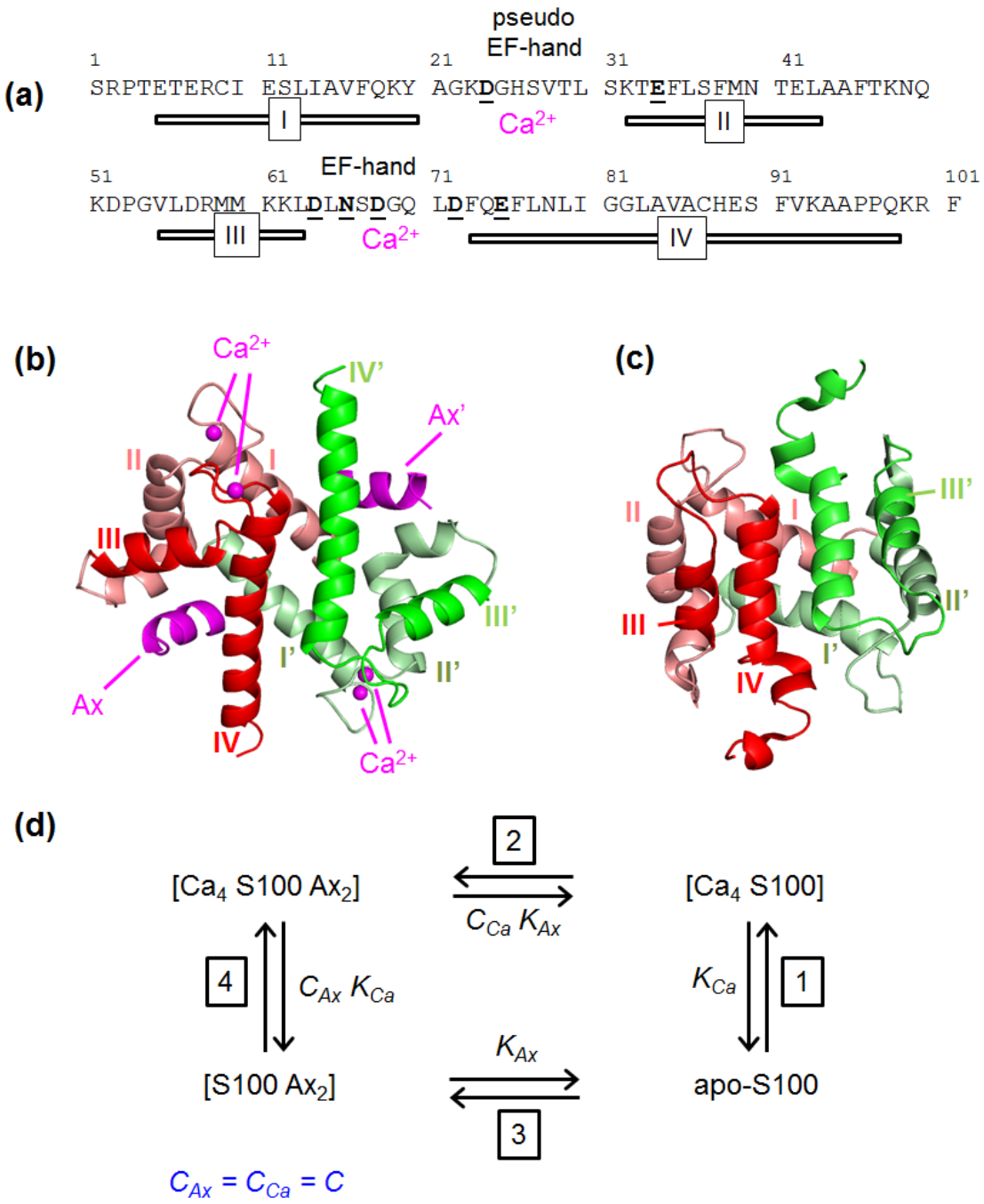


Figure 1. *(a) Sequence of rabbit S100A11; helices and calcium binding sites are indicated; side chains involved in Ca^{2+} binding are underlined. (b) Crystal structure 1QLS[39] of S100A11 bound to four Ca^{2+} and two annexin peptides (Ax). (c) NMR structure of apo-S100A11 (pdb file 1NSH).[46] (d) Thermodynamic cycle, illustrating the transition between apo-S100 and [Ca_4 S100 Ax_2] along two paths. Steps 1 & 2 refer to Ca^{2+} binding followed by Ax binding. Steps 3 & 4 refer to Ax binding followed by Ca^{2+} binding. The overall equilibrium constant for complex formation is the same along both paths, implying that K_{Ca} (C_{Ca} K_{Ax}) = K_{Ax} (C_{Ax} K_{Ca}) such that C_{Ca} = C_{Ax} = C.*

A number of studies have explored Ca^{2+}-dependent allosteric effects in S100 family members and other EF-hand proteins[35,54-57] The mechanism by which metal binding sites control target binding sites over a substantial distance nonetheless remains poorly understood.[46,58] For the system considered here, this distance is roughly 26 Å (measured from the center of an Ax helix to the EF-hand Ca^{2+} in the same subunit)[39] Simply speaking, the field is dominated by the view that calcium loss enhances the conformational freedom of the EF- and pseudo EF-hands, and that this increased flexibility triggers a sequence of events that culminates in target binding site closure.

Here we scrutinized the mechanism of S100 allosteric regulation by employing microsecond molecular dynamics (MD) simulations in explicit water, complemented by hydrogen/deuterium exchange mass spectrometry (HDX/MS). Protein structure and dynamics were probed in the presence/absence of calcium and Ax, corresponding to the complexes [Ca_4 S100 Ax_2], [Ca_4 S100], [S100 Ax_2] and apo-S100. Our data reveal an unexpected regulatory mechanism. Closure of the target binding site represents the culmination of an allosteric cascade that does *not* originate at the metalation sites. Instead, binding site closure is triggered by structural perturbations emanating from a labile salt bridge that acts as incessantly active "agitator". Bound calcium stabilizes the target binding sites in an open conformation by blocking allosteric signals produced by the agitator. To our knowledge, the current work marks the first time that such an allosteric mechanism has been documented.

4.2 Methods

4.2.1 Proteins and Reagents

Rabbit S100A11 was expressed and purified as described.[46,59,60] As in previous work,[61] a C9S variant was used to avoid inappropriate disulfide formation. The expected monomer

mass (11281 Da) was verified by MS (11281.5 ± 1) Da. Ax (acetyl-STVHEILSKLSLEGD) was synthesized by BioBasic (Markham, ON). This peptide was used because of its higher solubility and binding affinity,[61] compared to the annexin A1 peptide in the 1QLS X-ray structure.[39] D_2O (Aldrich, St. Louis, MO), KCl (Caledon, Georgetown, ON), HEPES, EDTA (Sigma), $CaCl_2$, and HCl (Caledon) were used as received.

4.2.2 HDX Mass Spectrometry

Deuteration was performed at 23 ± 1 °C in 90 % D_2O, 50 mM KCl, 20 mM HEPES, at a pH meter reading of 7.1. These solution conditions were chosen to ensure consistency with previous NMR studies, which demonstrated that background electrolyte concentrations >> 50 mM can interfere with Ca^{2+} binding.[46,61] The protein concentration was 2.5 μM (as dimer). Other components were added as needed, i.e., 5 mM $CaCl_2$ for experiments conducted in the presence of Ca^{2+}, 0.2 mM EDTA for Ca^{2+}-free samples, and with or without 70 μM Ax. All concentrations refer to the final values under HDX conditions. From published dissociation constants it can be estimated that ~99.5 % of the metalation sites were occupied in the Ca^{2+} containing solutions,[62] and ~95 % of the protein was bound to Ax in [Ca_4 S100 Ax_2] samples.[61] 100 μL aliquots were removed at various time points between 1 min and 2 h after initiation of labeling. The aliquots were quenched to pH 2.3 by addition of HCl on ice, flash frozen in liquid nitrogen, and stored at - 80 ˚C. The aliquots were thawed to ~ 0˚C and injected into an HDX nanoACQUITY UPLC[63] (Waters, Milford, MA). Digestion was performed on a 2.1 mm × 30 mm POROS pepsin column (Applied Biosystems, Carlsbad, CA) at 15˚C. Peptides were trapped on a Waters BEH C18 (1.7 μm, 2.1 × 5 mm) column, and separated on a BEH C18 (1.7 μm, 1 × 100 mm) column using a water/acetonitrile gradient with 0.1% formic acid at 40 μL min^{-1}. Data analysis focused on 17 peptides that cover the sequence in a contiguous fashion (Figure 2). Peptide masses were measured on a Waters Synapt G2 Q-TOF mass spectrometer. Peptide identities were confirmed by MS/MS. For correction of in-exchange, zero time point controls (m_0) were

performed by exposing the protein to labeling buffer under quench conditions. Fully exchanged controls (m_{100}) were prepared by incubation in labeling buffer at pH 2.4 and 37 ˚C for 24 h. HDX temporal profiles were obtained from centroid mass values at time *t*, m_t, obtained by DynamX 3.0 (Waters). Deuteration levels are reported as *percent deuteration* = $(m_t - m_0)/(m_{100} - m_0)$, where m_t represented the centroid mass of a peptide at time *t*. All samples displayed uncorrelated HDX with gradual shifts of the isotope envelopes to higher mass, as commonly seen in the EX2 regime.[64] All HDX data represent an average of triplicate measurements. Error bars indicate standard deviations.

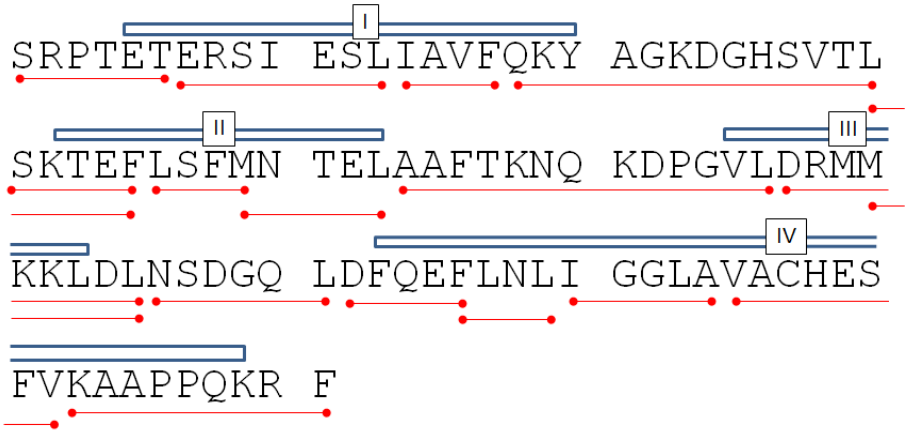


Figure 2. *Peptic digestion map of S100A11, illustrating the peptides that were used for HDX/MS data analysis.*

4.2.3 MD Simulations

MD simulations were conducted using GROMACS 5 with GPU acceleration[65] at 298 K. The coordinates of human S100A11 bound to four Ca^{2+} and two annexin peptides (pdb file 1QLS)[39] served as starting structure for all runs. Human S1000A11 and the rabbit variant used in our HDX/MS experiments have slightly different sequences (84% sequence identity). To ensure that MD results and HDX/MS data were directly comparable we swapped several amino acids in the 1QLS starting structure using the mutagenesis and sculpting routines implemented in PyMol (Schrödinger), as detailed in Figure 3. Similarly,

the annexin A1 peptide of the 1QLS structure was transformed to the homologous Ax sequence. Simulations were conducted using the CHARMM 36 force field[66] with TIP3P water.[67] This combination has previously been shown to yield reliable folding and equilibrium dynamics.[68] The protein was placed in a periodic box with a minimum distance of 7 Å between protein and the box surface. ~8000 water molecules were added, as well as 70 mM K^+ and Cl^-, plus additional ions to ensure charge neutrality. After one round of energy minimization atomic velocities were initiated from a Maxwell-Boltzmann distribution with random seeds. This was followed by 100 ps NVT and NPT equilibration runs. Production runs were conducted with a 2 fs time step, using a Parrinello-Rahman barostat and Berendsen thermostat.[69] Bonds were constrained by the linear constraint solver algorithm.[70] Short-range electrostatic and van der Waals cutoffs were set to 10 Å. Long-range electrostatics were treated using Particle Mesh Ewald summation,[71] with PME order = 4 and Fourier spacing = 0.16 nm. Two independent 1 μs simulations were conducted for each of the conditions [Ca_4 S100 Ax_2], [Ca_4 S100], [S100 Ax_2] and apo-S100, for a total of eight 1 μs runs. Major structural changes were observed only for apo-S100, prompting us to perform an additional 200 ns run for this species to verify reproducibility. All repeat runs employed slightly different starting coordinates and initial velocities.

```
          1           11          21          31          41
pig    MA KRPTETERCI ESLIAIFQKH AGRDGNNTKI SKTEFLIFMN TELAAFTQNQ
rabbit    SRPTETERCI ESLIAVFQKY AGKDGHSVTL SKTEFLSFMN TELAAFTKNQ
MD        SRPTETERCI ESLIAVFQKY AGKDGHSVTL SKTEFLSFMN TELAAFTKNQ

          51          61          71          81          91
pig       KDPGVLDRMM KKLDLDSDGQ LDFQEFLNLI GGLAIACHDS FIKSTQK
rabbit    KDPGVLDRMM KKLDLNSDGQ LDFQEFLNLI GGLAVACHES FVKAAPPQKR F
MD        KDPGVLDRMM KKLDLNSDGQ LDFQEFLNLI GGLAVACHES FVKSTQK
           1           11          21          31          41
```

Figure 3. *Sequence comparison of porcine (1QLS)[39] and rabbit (1NSH)[46] S100A11. Amino acids that differ are highlighted in red. Also shown is the sequence used for the MD simulations of this work. It corresponds to the rabbit variant, except for the C-terminus where the shorter STQK element of the porcine protein was retained. This shorter C-terminus permitted the use of a smaller simulation box with less water, thereby significantly reducing computational cost. This figure also defines the residue numbering used in this study.[46]*

4.3 Results and Discussion

4.3.1 Molecular Dynamics Simulations

We conducted 1 μs MD simulations on [Ca_4 S100 Ax_2], [Ca_4 S100], [S100 Ax_2], and apo-S100. Using the [Ca_4 S100 Ax_2] X-ray coordinates[39] as starting point, structural changes were triggered by removal of calcium and/or Ax. All simulations converged towards relaxed (meta)stable conformers within ~0.5 μs (Figure 4). As a benchmark, we verified that [Ca_4 S100 Ax_2] maintained a conformation close to the initial X-ray structure (Figure 5a). Runs without calcium and Ax produced structures consistent with the apo-S100 NMR coordinates (Figure 5b). These tests confirm that the MD conditions used are adequate.

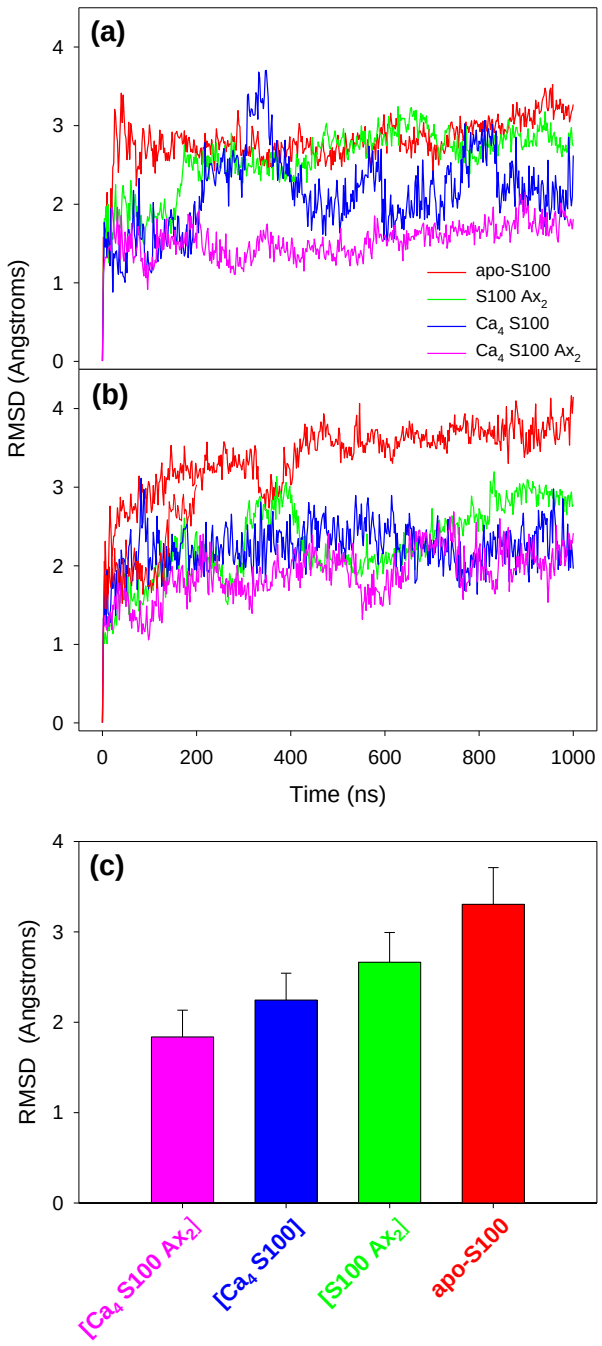


Figure 4: *RMSD plots, illustrating how the protein structures evolved during 1 μs simulations, relative to the common starting conformation. Panels (a), (b) display data for two sets of replicate MD runs. Occasional large fluctuations (e.g., in (a) for Ca_4 S100 around 350 ns) arise from fraying of terminal helices. In (c), data from the 500 ns - 1 μs range from both sets of simulations were averaged. All RMSDs were calculated for backbone heavy atoms from both subunits, focusing on helical regions. The sequence range used was I: 6-22, II: 34-48, III: 56-63, IV: 74-86.*

Target binding site conformations can be characterized by examining the orientation of helices III/IV (Figure 5c). For helix angles of ~115° or less the binding sites are open and capable of accommodating Ax. Larger values (around 155°) represent closed binding sites.[39,46,72] Most S100 structural studies[39,46,72] reported the same angle for both subunits in the homodimer due to symmetry constraints imposed during analysis[73] (Figure 5c, gray bars). However, in solution[74] and in our unconstrained MD simulations the helix III/IV and helix III'/IV' angles will not be identical. Figure 5c therefore displays two angles for each MD structure. Runs for [Ca_4 S100 Ax_2] and [Ca_4 S100] maintained open binding sites with angles close to those of the corresponding experimental structures.[39,72] Experimental data for [S100 Ax_2] are not available; the two MD runs conducted for this species culminated in somewhat different outcomes. In the first instance both binding sites stayed open. The second [S100 Ax_2] run yielded dissimilar angles of ~85° for helices III/IV (open), and ~144° for III'/IV' (closed). Finally, the simulated apo-S100 angles all corresponded to closed binding sites (Figure 5b), as noted above.[46]

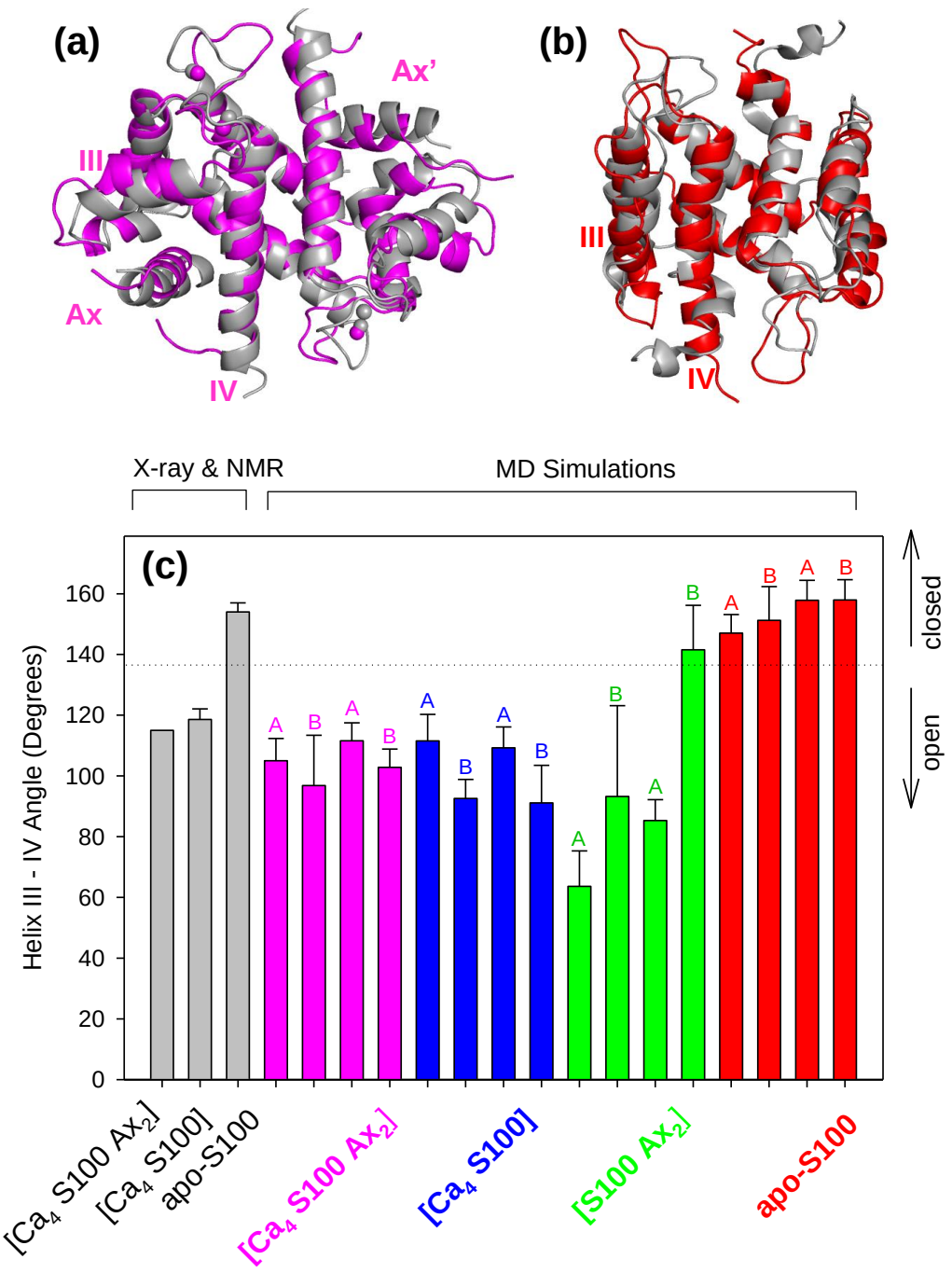


Figure 5. *(a) 1 μs MD structure of [Ca₄ S100A11 Ax₂] (magenta) and X-ray structure (gray, 1QLS).[39] (b) 1 μs MD apo-S100 structure (red) and NMR structure (gray, 1NSH #5).[46] (c) Helix III/IV and III'/IV' angles. Experimental values are from pdb files 1QLS,[39] 2LUC,[72] and 1NSH.[73] MD data were averaged over the final 0.5 μs windows of 1 μs simulation runs, and data are shown separately for subunits A and B from two independent runs for each condition. Error bars represent standard deviations.*

4.3.2 Details of MD Structures

Figure 6 provides a closer look at examples of t = 1 μs MD structures, using a layout consistent with that of Figure 1d. [Ca_4 S100 Ax_2] exhibited a globular fold with a well-developed hydrophobic core (Figure 6a). Both Ax peptides remained tightly associated with the hydrophobic target binding pockets. The EF-hand calcium binding loops at the protein surface were firmly rooted in the core via nonpolar anchor residues (L63 and L71). Each EF-hand was in close contact with an intricately packed "shoulder" consisting of D57/K32/M60/L71. In this shoulder D57 and K32 were connected by a salt bridge, while K32, M60, and L71 were linked by close hydrophobic contacts. The K32-D57 salt bridge as well as the adjacent packing contacts are highly conserved in S100 proteins.[74] It may seem counter-intuitive to have lysine (a "hydrophilic" residue) involved in hydrophobic packing, but motifs of this type are common, as governed by the aliphatic nature of the $C_\alpha H\text{-}CH_2\text{-}CH_2\text{-}CH_2\text{-}CH_2$- moiety in lysine.[75]

[S100 Ax_2] showed partial disruption of the D57/K32/M60/L71 shoulder (Figure 6b). For the structure exemplified in Figure 6, only one of the target binding sites retained tight Ax contacts. The second site adopted a closed structure, while the corresponding Ax peptide remained loosely associated via residual nonpolar contacts (colored in Figure 6b). Closure of this binding sites was a gradual process that extended over ~0.5 μs (Figure 7A (q)). The perturbed target interactions suggest that the corresponding Ax peptide is poised to separate from the complex, although dissociation did not take place on the time scale of our simulations. Still, the [S100 Ax_2] MD structure clearly represents an intermediate en route towards target-free apo-S100.

Except for the absence of Ax, the [Ca_4 S100] 1 μs structure was very similar to that of [Ca_4 S100 Ax_2]. Both possess open target binding sites and intact D57/K32/M60/L71 shoulders (Figure 6c). This is different from apo-S100, where target binding site closure in both subunits was associated with complete disintegration of the D57/K32/M60/L71 shoulders (Figure 6d).

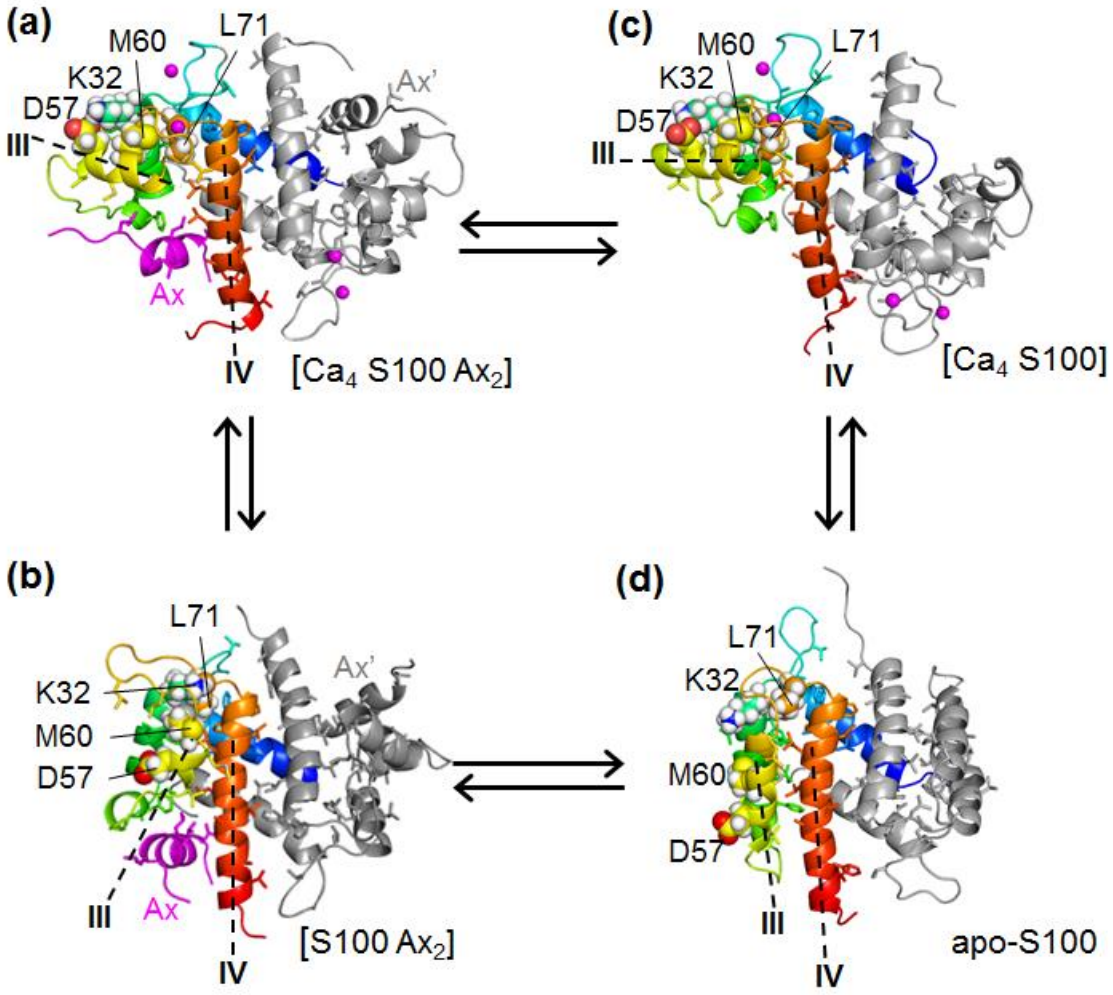


Figure 6. *1 μs MD structures for the four simulation conditions. Structures and equilibrium arrows are arranged in accordance with Figure 1d. One subunit in each panel is grayed out to reduce clutter. Hydrophobic side chains (I, V, L, F, M, A) are displayed as sticks. Spacefill representation was used for residues constituting the D57/K32/M60/L71 "shoulder", which is intact only in panels (a, c). Ca^{2+} and Ax are displayed in magenta. Dashed lines indicate the helix III/IV orientation, which defines whether target binding sites are open (a, c), or closed (b, d).*

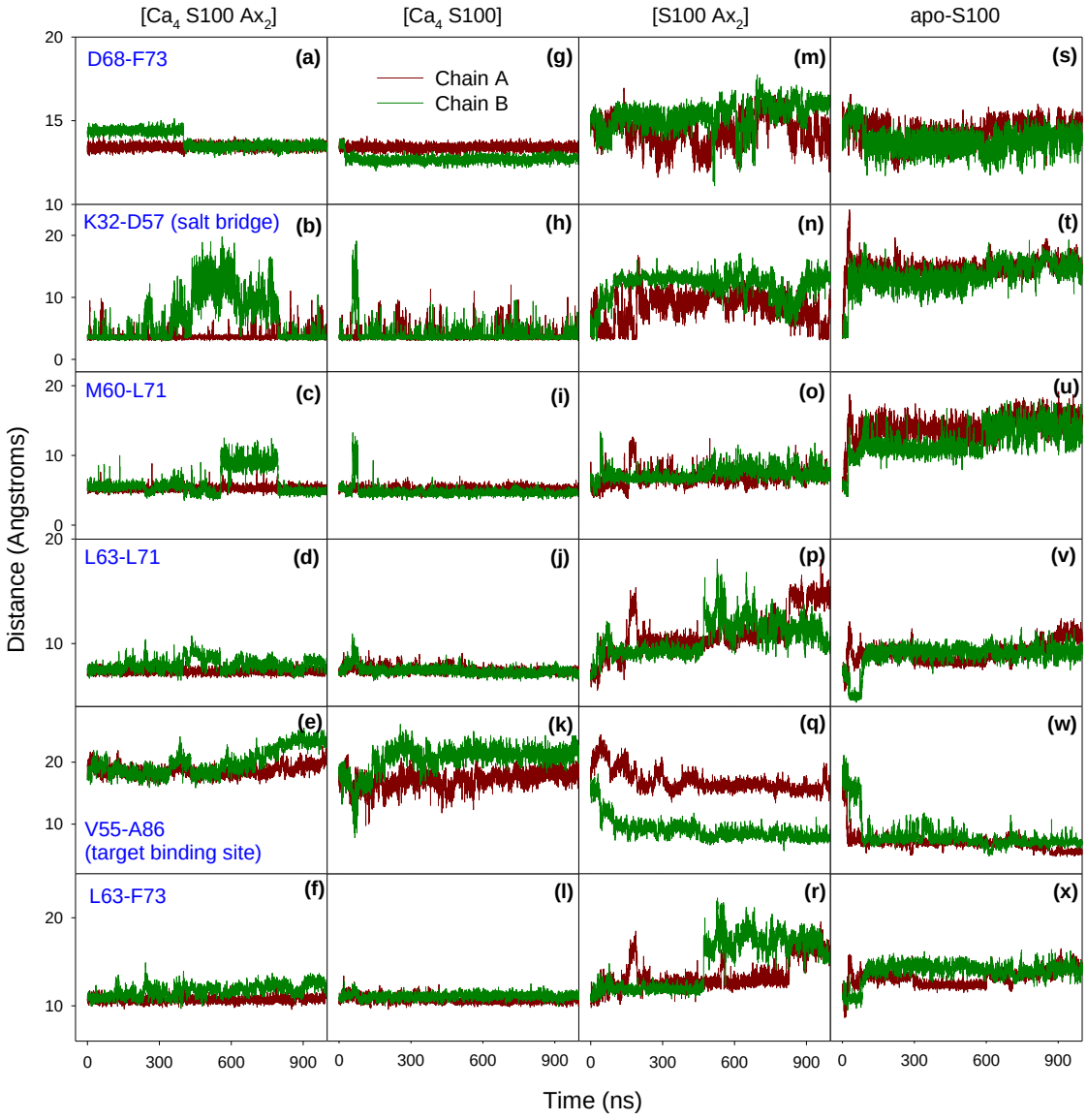


Figure 7(A). Complete 1 μs trajectory data, displaying inter-atomic distances for the four MD conditions, as noted along the top. Each panel displays data for chain A (dark red) and chain B (dark green) within the S100 dimer.

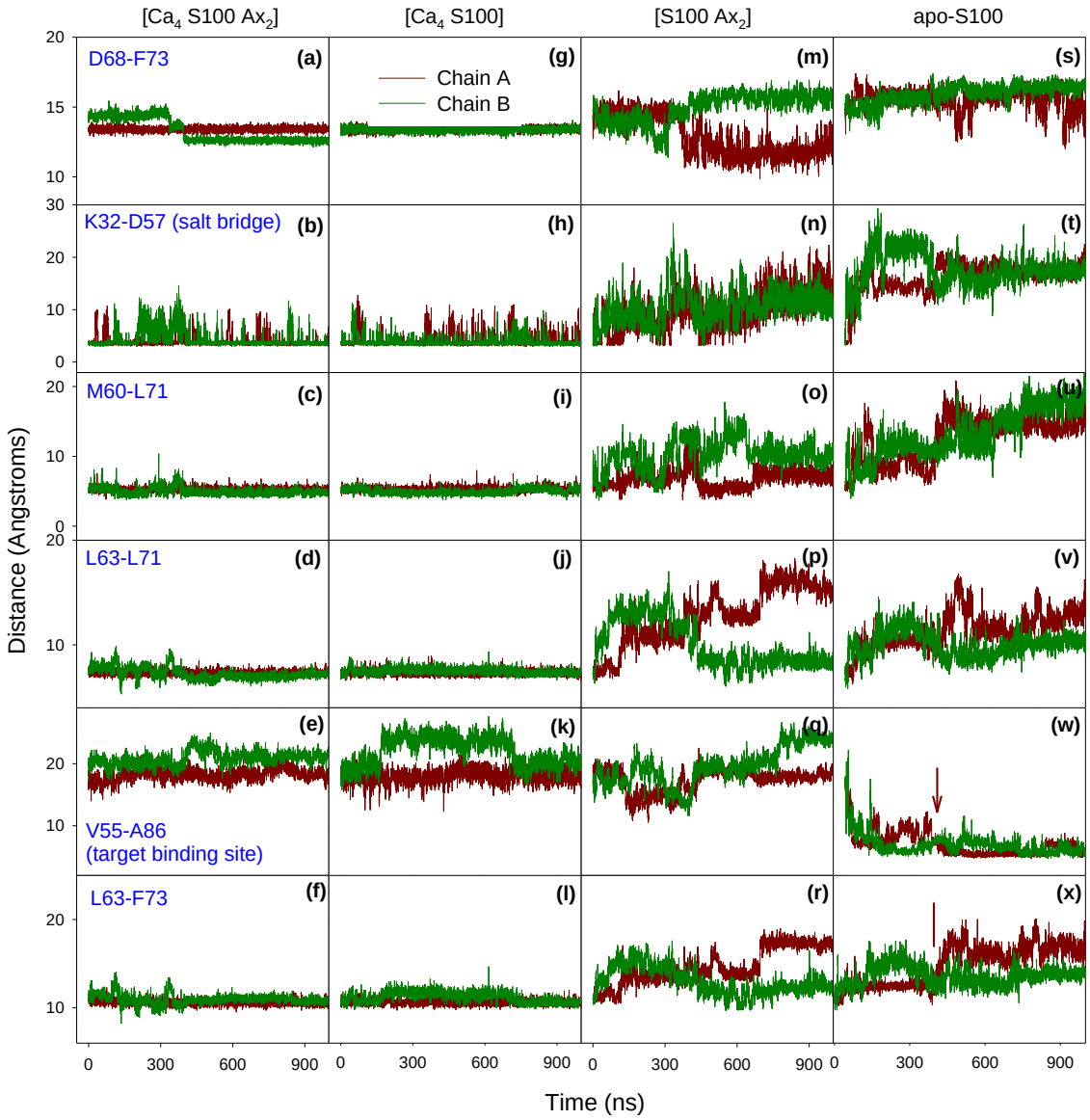


Figure 7(B). *Complete 1 μs trajectory data, as on the preceding page, but for an independent set of MD simulations. The significance of the arrows in (w), (x) at ~450 ns is discussed in the caption of Figure 10.*

4.3.3 Allosteric Control of Target Binding Sites

Our MD data capture a central aspect of calcium-mediated allosteric control, i.e., the fact that in apo-S100 the target binding sites collapse to a closed conformation that cannot bind Ax (Figure 6d). This is in contrast to [Ca_4 S100] where calcium maintains open target binding sites that are primed for Ax interactions (Figure 6c).

For deciphering the mechanism of allostery it is essential to identify the driving force for binding site closure. This event is not associated with the formation of stable salt bridges. Also, changes in H-bonding are minor (Table 1).[54,55] However, [Ca_4 S100] possesses numerous solvent exposed nonpolar residues in the target binding sites (Figure 8a). Binding site closure dramatically reduces the accessibility of these residues and allows them to form hydrophobic contacts (A45, F46, V55, M59, L63, A86, F91, Figure 8b). Hence, the main driving force for binding site closure is the hydrophobic effect (Figure 8a, b).[76-79] At the risk of sounding naïve, we point out parallels between this [Ca_4 S100] → apo-S100 closure (Figure 8a, b) and another "allosteric" process, i.e., a spring-loaded mouse trap that snaps close after a rodent pulls at the bait. Bait movements do *not* provide the driving force for the closure event. Instead, pulling at the bait triggers the release of energy stored in the spring. Analogously, Ca^{2+} loss does *not* provide the driving force for binding site closure. Instead, Ca^{2+} loss triggers the release of free energy associated with hydrophobic collapse of the binding site (Figure 8b).[77-79] Within this picture, the exposed nonpolar sites in [Ca_4 S100] serve a purpose analogous to that of the spring in the trap. This digression prompts two questions: (i) How does the presence of Ca^{2+} prevent binding site closure? (ii) How does the absence of Ca^{2+} trigger binding site closure? Both can be addressed by examining the temporal behavior of selected atom distances (Figure 8c), with primary focus on [Ca_4 S100] and apo-S100 (Figure 8d-u).

Table 1 (continued on subsequent pages): H-bonds derived from the final 0.5 μs of MD simulation runs. The first column lists the NH donor, the following columns list the main chain (m) and side chain (s) oxygen acceptors. Acceptor notation is as follows: 10Im(A97,B89) refers to residue 10, isoleucine, main chain, 97% H-bonded in chain A, 89% H-bonded in chain B. Data are shown only for NH sites that are H-bonded more than 20% in total; contributions below 2% were omitted. For the sake of simplicity, this table only refers to one set of MD replicates, corresponding to the runs highlighted in Figure 6.

NH	**[Ca_4 S100 Ax_2]**	**[Ca_4 S100]**	**[S100 Ax_2]**	**apo-S100**
2R		11Es(A0, B65)		
4T		2Rm(A30, B0)	2Rm(A7, B24) 7Es(A18, B0)	7Es(A21, B0)
5E	42Es(A100, B23) 2Rm(A0, B4)	42Es(A100, B22)	42Es(A40, B100) 5Es(A4, B0)	5Es(A88, B28)
6T	42Es(A100, B66)	42Es(A100, B96)	42Es(A78, B100)	
7E	4Ts(A0, B9) 7Es(A0, B31)	4Ts(A0, B26)	4Ts(A79, B81)	4Tm(A34, B21)
8R	4Tm(A0, B80)	4Tm(A42, B77) 5Em(A3, B0)	4Tm(A32, B57)	4Tm(A23, B20) 5Em(A3, B4)
9C	5Em(A90, B86) 6Tm(A0, B2)	5Em(A75, B82) 6Tm(A4, B4)	5Em(A88, B89)	5Em(A93, B81) 6Tm(A0, B2)
10I	6Tm(A98, B76) 7Em(A0, B2)	6Tm(A78, B84)	6Tm(A98, B84)	6Tm(A98, B89)
11E	7Em(A99, B86)	7Em(A99, B99)	7Em(A96, B100)	7Em(A97, B96)
12S	8Rm(A95, B94)	8Rm(A93, B93)	8Rm(A96, B89)	8Rm(A95, B94)
13L	9Cm(A97, B95)	9Cm(A91, B92)	9Cm(A96, B72) 10Im(A0, B3)	9Cm(A98, B94)
14I	10Im(A94, B77)	10Im(A97, B89)	10Im(A93, B92)	10Im(A97, B95)
15A	11Em(A97, B98)	11Em(A94, B96)	11Em(A99, B95)	11Em(A97, B97)
16V	12Sm(A94, B97)	12Sm(A95, B93)	12Sm(A85, B82)	12Sm(A99, B97)
17F	13Lm(A65, B62) 14Im(A3, B0)	13Lm(A97, B79)	13Lm(A58, B74)	13Lm(A95, B88)
18Q	14Im(A96, B97)	14Im(A94, B95)	14Im(A100, B97)	14Im(A100, B99)
19K	15Am(A30, B40) 16Vm(A25, B26)	15Am(A58, B58) 16Vm(A16, B13)	15Am(A63, B79) 16Vm(A6, B8)	15Am(A87, B76) 16Vm(A2, B4)
20Y	17Fm(A47, B49) 16Vm(A0, B6)	16Vm(A42, B22) 17Fm(A4, B14)	16Vm(A15, B79) 17Fm(A32, B0)	16Vm(A65, B56) 17Fm(A4, B13)
21A	18Qm(A76, B31) 17Fm(A0, B41)	17Fm(A25, B24) 18Qm(A35, B49)	17Fm(A74, B81) 18Qm(A6, B3)	17Fm(A91, B55) 18Qm(A0, B14)
22G	18Q(A0, B23) 34Es(A0, B32)	34Es(A90, B89)	18Qm(A2, B72) 29Tm(A77, B0) 19Km(A0, B5)	18Qm(A75, B18) 19Km(A3, B0) 29Tm(A0, B45)
23K				19Km(A35, B0) 21Am(A2, B0) 27Sm(A21, B0)

24D	21Am(A54, B17) 24Ds(A3, B10)	24Ds(A29, B28)	27Sm(A29, B0)	
25G				22Gm(A0, B14) 24Ds(A0, B22)
26H				23Km(A0, B13) 18Qm(A0, B14)
27S		18Qs(A0, B21) 24Dm(A0, B5)	24Dm(A20, B0)	21Am(A2, B12) 23Km(A17, B3) 22Gm(A0, B8) 24D(A0, B3)
28V				
29T	24Ds(A99, B0)		22Gm(A43, B0)	27Ss(A42, B37) 21Am(A22, B0)
30L	71Lm(A99, B92)	71Lm(A94, B86)	71Lm(A96, B20)	71Lm(A64, B73)
31S	34Es(A100, B76)	34Es(A99, B98)	20Ys(A0, B11) 34Es(A0, B22)	
32K			69Gm(A28, B89) 70Qs(A41, B0)	69Gm(A75, B81) 70Qs(A3, B0)
33T			31Ss(A0%, B4%) 68Ds(A0%, B2%) 68Dm(A0%, B14)	
34E		31Ss(A33, B22) 31Sm(A0, B2)	31Ss(A20, B31) 31Sm(A3, B5)	31Ss(A23, B0) 31Sm(A5, B0)
35F	31Sm(A99, B95)	31Sm(A99, B99)	31Sm(A98, B68) 32Km(A0, B12)	31Sm(A94, B93)
36L	32Km(A96, B94)	32Km(A97, B93)	32Km(A97, B75)	32Km(A95, B99)
37S	33Tm(A83, B76)	33Tm(A87, B74)	33Tm(A95, B81)	33Tm(A87, B92)
38F	34Em(A95, B85)	34Em(A88, B82)	34Em(A96, B96)	34Em(A89, B96)
39M	35Fm(A99, B94)	35Fm(A99, B99)	35Fm(A99, B100)	35Fm(A99, B98)
40N	36Lm(A79, B79) 37Sm(A7, B7)	36Lm(A98, B95)	36Lm(A99, B96)	36Lm(A90, B83) 37Sm(A0, B4)
41T	37Sm(A16, B18) 38Fm(A65, B52)	37Sm(A46, B30) 38Fm(A32, B41)	37Sm(A73, B68) 38Fm(A8, B11)	37Sm(A14, B15) 38Fm(A61, B57)
42E	38Fm(A31, B21) 41Ts(A5, B10)	38Fm(A70, B54) 41Ts(A0, B4)	38Fm(A99, B99)	38Fm(A28, B7) 41Ts(A9, B18)
43L	38Fm(A57, B63) 39Mm(A6, B4)	38Fm(A17, B32) 39Mm(A23, B17)	39Mm(A98, B96)	38Fm(A53, B76) 39Mm(A9, B0)
44A	41Tm(A35, B41) 42Em(A0, B3)	41Tm(A31, B0)	41Tm(A4, B0) 42Em(A26, B0)	41Tm(A26, B63) 39Mm(A6, B0)
45A		39Mm(A24, B0) 40Nm(A13, B0)		41Tm(A28, B0)
46F		39Mm(A26, B0) 43Lm(A2, B0)		42Em(A32, B0) 43Lm(A5, B0)
47T	43Lm(A93, B75) 44Am(A0, B8)	43Lm(A42, B92)	43Lm(A96, B34)	43Lm(A79, B85) 44Am(A2, B0)
48K	44Am(A88, B5) 45Am(A2, B47)	44Am(A36, B5) 45Am(A25, B45)	44Am(A89, B33) 45Am(A0, B22)	44Am(A49, B62) 45Am(A15, B7)

49N	45Am(A39, B0) 46Fm(A32, B0)	44Am(A0, B93) 45Am(A41, B0) 46Fm(A31, B0)	45Am(A76, B45) 46Fm(A10, B4) 11Ss(A0, B29)	45Am(A37, B56) 46Fm(A28, B21)
50Q	46Fm(A31, B0) 47Tm(A45, B0)	46Fm(A32, B0) 47Tm(A36, B12) 47Ts(A0, B17)	46Fm(A79, B20) 47Tm(A6, B29)	46Fm(A10, B18) 47Tm(A51, B34)
51K				90Ss(A36, B62)
52D				50Qs(A8, B0) 90Ss(A39, B0)
54G	52Ds(A48, B0)		57Ds(A0, B87)	52Ds(A60, B53)
55V	52Dm(A0, B27) 53Pm(A0, B3)			
56L		52Dm(A0, B97)		
57D		53Pm(A0, B82)		
58R	54Gm(A90, B48)	54Gm(A87, B86)	54Gm(A85, B71) 55Vm(A0, B6)	54Gm(A88, B91)
59M	55Vm(A97, B96)	55Vm(A97, B99)	55Vm(A91, B30) 56Lm(A0, B31)	55Vm(A72, B49) 56Lm(A3, B7)
60M	56Lm(A91, B89)	56Lm(A95, B98)	56Lm(A96, B46) 57Dm(A0, B23)	56Lm(A93, B93)
61K	57Dm(A83, B79) 58Rm(A0, B4)	57Dm(A84, B96)	57Dm(A94, B0)	57Dm(A84, B94) 58Rm(A2, B0)
62K	58Rm(A86, B82) 59Mm(A3, B4)	58Rm(A89, B85) 59Mm(A0, B3)	58Rm(A91, B0) 59Mm(A2, B0)	58Rm(A85, B80) 59Mm(A3, B3)
63L	59Mm(A88, B74) 60Mm(A3, B10)	59Mm(A92, B73) 60Mm(A2, B12)	59Mm(A87, B0) 60Mm(A3, B0)	59Mm(A90, B74) 60Mm(A0, B4)
64D	60Mm(A46, B34) 61Km(A12, B13)	60Mm(A53, B70) 61Km(A7, B5)	60Mm(A46, B0) 61Km(A17, B0)	60Mm(A80, B69) 61Km(A3, B8)
65L	75Es(A69, B56)	75Es(A63, B75)	60Mm(A57, B0) 64Ds(A27, B0)	61Km(A79, B60) 62Km(A7, B20)
66N	64Ds(A13, B29) 75Es(A82, B59)	64Ds(A13, B14) 75Es(A82, B85)	64Ds(A0, B37)	62Km(A72, B54) 63Lm(A13, B12)
67S	64Ds(A20, B22) 64Dm(A6, B6)	64Ds(A20, B30) 64Dm(A7, B0)	64Ds(A0, B2) 64Dm(A0, B33)	63Lm(A11, B0) 64Dm(A3, B0) 75Es(A7, B0)
68D	64Ds(A25, B30) 66Ns(A3, B5) 68Ds(A3, B3)	64Ds(A27, B37) 66Ns(A4, B6)		
69G	64Ds(A98, B85)	64Ds(A99, B99)	67Ss(A31, B18) 68Ds(A14, B6)	67Ss(A56, B0)
70Q	64Ds(A3, B5) 68Ds(A84, B85)	68Ds(A91, B0) 64Ds(A0, B88)		
71L	30Lm(A100, B94)	30Lm(A100, B96)	30Lm(A93, B93)	30Lm(A96, B98)
72D	75Es(A100, B99)	75Es(A100, B100)	75Es(A89, B59)	75Es(A63, B75)
73F	28Vm(A0, B43)	28Vm(A0, B26)		
74Q				
75E				
76F	72Dm(A96, B98)	72Dm(A98, B97)	72Dm(A98, B97)	72Dm(A97, B98)

77L	73Fm(A93, B99)	73Fm(A91, B96)	73Fm(A99, B97)	73Fm(A99, B99)
78N	74Qm(A15, B95) 75Em(A50, B0)	74Qm(A19, B86) 75Em(A36, B0)	74Qm(A93, B83)	74Qm(A90, B89)
79L	75Em(A5, B89) 76Fm(A52, B0)	75Em(A2, B87) 76Fm(A52, B0)	75Em(A91, B84) 76Fm(A0, B3)	75Em(A83, B83) 76Fm(A2, B2)
80I	76Fm(A34, B100)	76Fm(A40, B99)	76Fm(A100, B98)	76Fm(A99, B99)
81G	77Lm(A99, B96)	77Lm(A96, B87)	77Lm(A96, B94)	77Lm(A93, B88)
82G	78Nm(A89, B83) 79Lm(A0, B2)	78Nm(A73, B59) 79Lm(A4, B11)	78Nm(A77, B84) 79Lm(A4, B0)	78Nm(A73, B63) 79Lm(A5, B9)
83L	79Lm(A98, B95)	79Lm(A98, B90)	79Lm(A92, B90)	79Lm(A84, B65) 80Im(A0, B4)
84A	80Im(A97, B96)	80Im(A95, B91)	80Im(A92, B97)	80Im(A96, B94)
85V	81Gm(A91, B94)	81Gm(A83, B59) 82Gm(A0, B6)	81Gm(A93, B98)	81Gm(A98, B97)
86A	82Gm(A91, B91)	82Gm(A80, B39) 83Lm(A0, B3)	82Gm(A73, B94)	82Gm(A83, B89)
87C	83Lm(A98, B98)	83Lm(A97, B96)	83Lm(A88, B99) 84Am(A5, B0)	83Lm(A96, B96)
88H	84Am(A94, B94)	84Am(A81, B96) 85Vm(A9, B0)	84Am(A75, B95) 85Vm(A9, B0)	84Am(A94, B95)
89E	85Vm(A95, B91)	85Vm(A80, B95)	85Vm(A67, B97)	85Vm(A97, B97)
90S	86Am(A87, B76) 87Cm(A2, B0)	86Am(A71, B90) 87Cm(A2, B0)	86Am(A53, B83) 87Cm(A3, B0)	86Am(A93, B89)
91F	87Cm(A97, B93) 88Hm(A0, B2)	87Cm(A84, B88) 88Hm(A3, B0)	87Cm(A82, B93) 88Hm(A4, B3)	87Cm(A95, B88) 88Hm(A0, B4)
92V	88Hm(A62, B96)	88Hm(A84, B92)	88Hm(A67, B97) 89Em(A5, B0)	88Hm(A97, B87)
93K	89Em(A18, B86) 88Hm(A64, B0)	89Em(A75, B70) 90Sm(A9, B9)	89Em(A59, B81) 90Sm(A15, B0) 88Hm(A0, B2)	89Em(A95, B76) 88Hm(A0, B2)
94S	90Sm(A3, B41) 91Fm(A18, B14) 89Em(A5, B0)	90Sm(A17, B6) 91Fm(A32, B48)	90Sm(A30, B53) 91Fm(A34, B19)	90Sm(A72, B45) 91Fm(A11, B18)
95T	89Em(A62, B2) 90Sm(A16, B7) 91Fm(A0, B21) 92Vm(A0, B17)	90Sm(A56, B80) 91Fm(A2, B0)	91Fm(A11, B28) 92Vm(A11, B24) 90Sm(A26, B0)	91Fm(A48, B20) 92Vm(A20, B24)
96Q	92Vm(A0, B19) 93Km(A0, B3) 94Ss(A0, B3)			92Vm(A16, B14) 93Km(A19, B3) 94Ss(A0, B7)
97K				

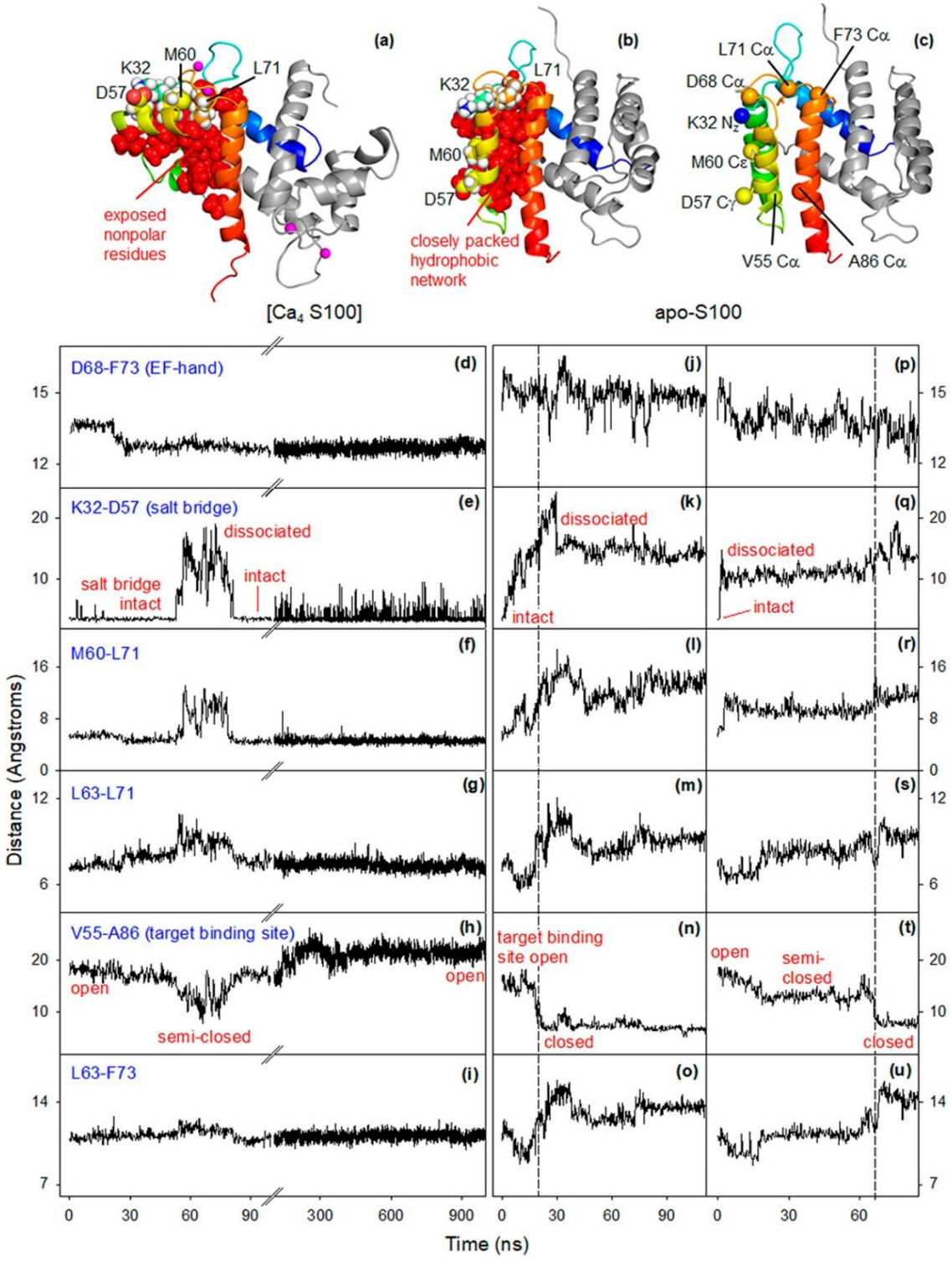


Figure 8. *(a) 1 μs MD structure of [Ca4 S100]. Exposed nonpolar residues in the target binding site are highlighted in red. (b) 1 μs MD structure of apo-S100. (c) Atoms used for tracking protein conformational dynamics. (d-i) Atom distances for [Ca4 S100]. Panels (j-o) and (p-u) display two apo-S100 runs. Dashed lines indicate target binding site closing midpoint (panels n, t).*

Our simulations reveal that the metalation sites became more dynamic after calcium loss, exemplified by the D68-F73 distance fluctuations in Figure 8d/j/p. F73 represents a quasi-stationary reference point, while D68 is in the center of the EF-hand. Importantly, the enhanced D68 dynamics are not directly correlated with target binding site closure events (Figure 8n, t). This lack of correlation reflects the fact that much of the EF-hand is only weakly coupled to the protein core (Figure 9a). The same is true for the pseudo EF-hand loops, which extend far into the solvent (Figure 8a, b). Thus, the problem has to be investigated from a broader perspective, by also examining events taking place elsewhere in the protein.

The egregious exposure of hydrophobic residues in the [Ca_4 S100] target binding sites can be maintained because of favorable nonpolar packing between helices II and III adjacent to the EF-hands. This hydrophobic cluster includes L63, it is capped off by the D57/K32/M60/L71 shoulder, and it suspends helix III in the open conformation (Figure 9a). Surprisingly, in [Ca_4 S100] the shoulder can undergo large perturbations *without* triggering irreversible binding site closure, exemplified by events between t = 50 and 80 ns (Figure 8e-i). During this time the K32-D57 salt bridge underwent dissociation, M60 moved away from L71, L63/L71 drifted apart, and the binding site began to close (evident from a decreasing V55-A86 distance, Figure 8h). A snapshot taken at t = 70 ns illustrates the severity of these perturbations, which even included disruption of D64 and N66 metal ligation in the EF-hand (Figure 9b). Remarkably, these events stalled at $t \approx 70$ ns. Instead of closing all the way, the binding site returned to the fully open state, and the D57/K32/M60/L71 shoulder underwent re-assembly (Figure 8e-h). The observed behavior suggests that a key event is not permitted to happen in [Ca_4 S100], which would otherwise allow binding site closure to go to completion.

Our data suggest that binding site closure in [Ca_4 S100] is prevented by a lack of L63-F73 distance fluctuations (more support for this assertion is provided in the next paragraph). As noted, L63 represents an EF-hand anchor. F73 remains virtually stationary under all conditions. In [Ca_4 S100] L63 and F73 are locked at a distance around 11 Å (Figure 8i). The invariability of this distance results from constraints imposed by calcium-protein contacts in the EF-hand (Figure 9a). In summary, in [Ca_4 S100] hydrophobic packing of

the shoulder/EF-hand region suspends helix III in the open state. This arrangement is stabilized by L63-F73 distance constraints that arise from Ca^{2+}-induced rigidification of adjacent to residues (Figures 1a, 9a). Thus, calcium ensures that perturbations in the shoulder get "blocked" as they propagate from K32 towards the core (note the gradually diminishing amplitude at $t \approx 70$ ns when going from Figure 8e to 4i). In this way calcium prevents irreversible binding site closure in [Ca_4 S100]. Another example of such a blocked propagation process, observed in an independent [Ca_4 S100] run, is highlighted in Figure 10.

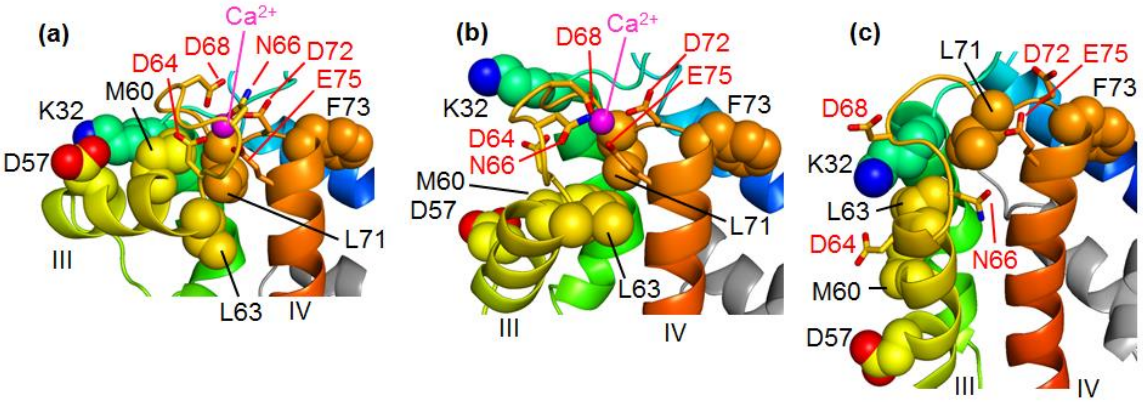


Figure 9. *Close-up views of the EF-hand calcium binding region during MD runs. (a) [Ca_4 S100] at 1 μs. (b) [Ca_4 S100] while undergoing a structural perturbation at $t \approx 70$ ns. (c) apo-S100. Side chains comprising the D57/K32/M60/L71 shoulder, as well as those of L63 and F73 are shown as spheres. EF-hand side chains involved in calcium binding are shown as sticks, and labeled in red.*

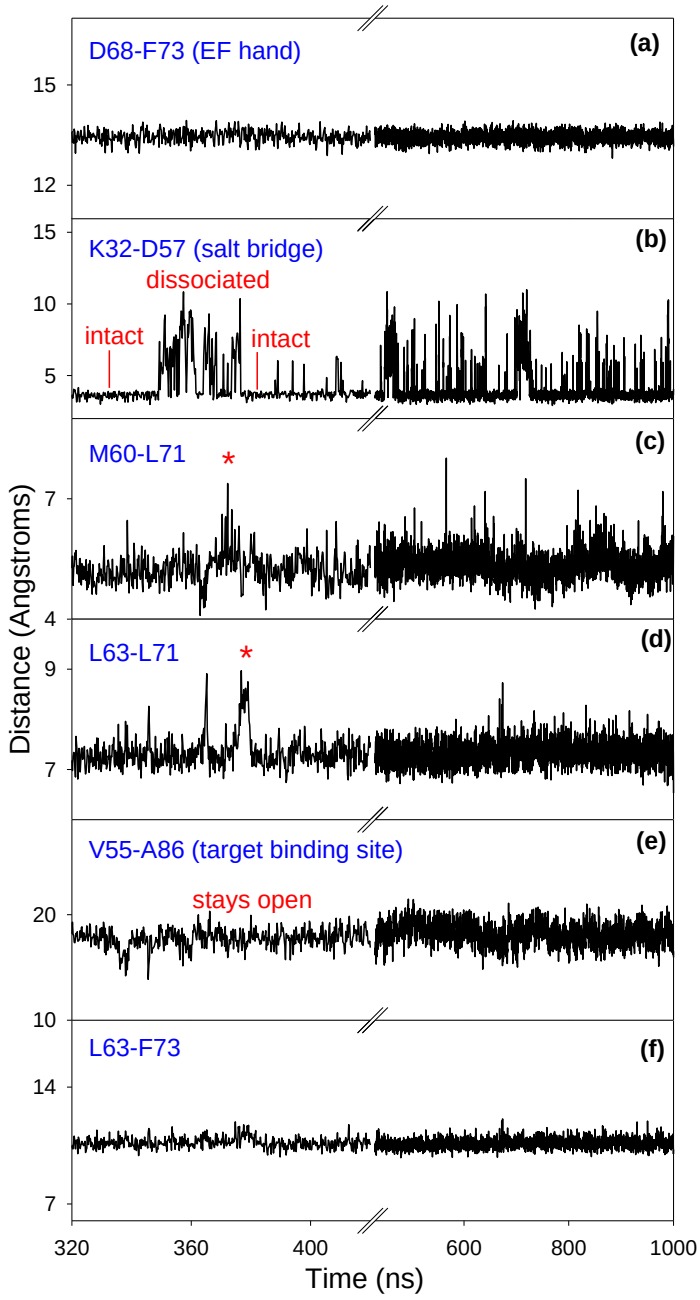


Figure 10. *MD simulation data generated in a [Ca_4 S100] run. This figure illustrates the blocked propagation of a structural perturbation, resembling the events illustrated in Figure 4d-j of the main text (the data shown here are from an independent replicate). The K32-D57 salt bridge temporarily dissociates at t ≈ 350 ns (panel b). This is followed by the transient disruption of M60-L71 contacts, with a subsequent fluctuation in L63-L71 packing (indicated by asterisks in panels c, d). Further propagation of the perturbation is blocked, as the L63-F73 distance remains almost constant because of Ca^{2+}-induced rigidification. As a result, the target binding site stays open instead of undergoing hydrophobically-driven closure.*

In apo-S100 all closing transitions were irreversible. Although different MD runs showed slightly different sequences, each closing event was associated with major L63-F73 distance fluctuations and a subsequent L63-F73 distance increase from ~11 Å to ~14 Å. We will highlight two runs in more detail. In the first example disassembly of the shoulder took place gradually over ~40 ns (Figure 8k-m). During this time the K32-D57 salt bridge dissociated and M60 moved away from L71. The L63-L71 distance underwent a major fluctuation which extended to L63-F73. The latter event triggered closure of the target binding site at $t \approx 25$ ns (dashed line, Figure 8n). In another apo-S100 run (Figure 8q-u) dissociation of the K32-D57 salt bridge took place within the first few nanoseconds, followed by repositioning of M60. Alterations in L63-L71 distance were minor, but a change in L63-F73 distance within the initial 10 ns allowed the target binding site to move into a semi-closed state. At $t \approx 60$ ns the binding site closed completely, while simultaneously the L63-F73 distance increased to ~14 Å. Our data suggest that L63-F73 distance *fluctuations* are the key prerequisite for target binding site closure, rather than the L63-F73 distance increase from 11 Å to 14 Å. This is evident from the fact that at the transition midpoints of Figure 8 o,u the L63-F73 distances are still relatively small, around 12 Å. Also, after the binding sites had closed, the L63-F73 distance could temporarily return from 14 Å back to 12 Å (Figure 7B, panel x, chain A, 300-600 ns). Numerous other distances were scrutinized, but no events were correlated with binding site closure as clearly as the L63/F73 behavior.

We conclude that target binding site closure in apo-S100 is the result of a domino cascade that starts with disintegration of the D57/K32/M60/L71 shoulder. The labile K32-D57 salt bridge represents the primary "agitator". Dissociation of this salt bridge is followed by M60 swinging away from L71. In the absence of Ca^{2+} the hydrophobic packing of residues next to L63 and F73 is not stabilized by EF-hand mediated rigidification, allowing the formation of extensive new hydrophobic contacts as the target binding sites close (Figure 8b). Complete 1 μs trajectories for all four conditions are provided in Figure 7. Those data confirm a high propensity of the K32-D57 salt bridge to undergo reversible dissociation even for [Ca_4 S100 Ax_2], in line with the labile nature of this contact in [Ca_4 S100].

The apo-S100 closing cascades highlighted in Figure 8 refer to collapse of the first binding site in the dimer. In both instances the collapse of the second site took place ca. 70 ns after the first one. Those subsequent closure events also required L63-F73 distance fluctuations (Supporting Figure S7). In a third apo-S100 run both sites underwent initial closure roughly simultaneously around 25 ns, but chain A settled into a fully relaxed closed state only after ~450 ns (Figures 7, 11). Taken together, these observations suggest that closure events affecting the two binding sites within a S100 dimer are not directly coupled to one another.

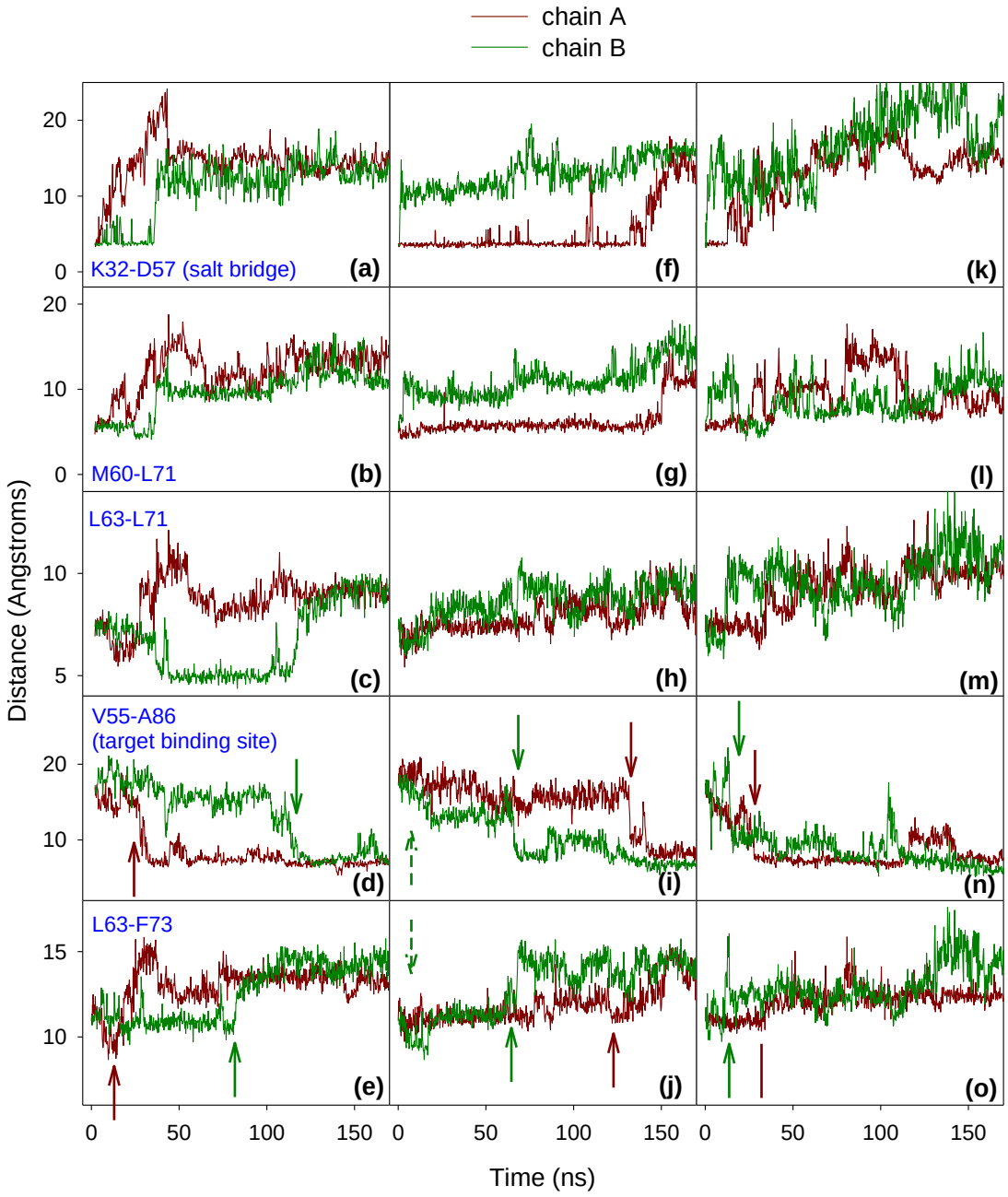


Figure 11: *Initial 170 ns time segments of three different apo-S100 MD runs, displaying inter-atomic distances for the two S100 subunits, chain A (dark red) and chain B (dark green). Arrows in (d, i, n) indicate target binding site closure events. Arrows in (e, j, o) highlight L63-F73 distance fluctuations/alterations that precede these closing events or occur quasi-simultaneously with them. The chain A target binding site in (n) fully settles into its irreversibly closed state only after ~450 ns; consistent with all other apo-S100 runs, this closure event is associated with a major L63-F73 fluctuation (see arrows in Supporting Figure 7, panels w, x)*

4.3.4 Probing Calcium and Target Binding by HDX/MS

Our MD simulations successfully captured ns - μs events associated with allosteric control. However, conformational fluctuations in solution can extend to time scales that are orders of magnitude longer,[80] stretching into realms that are inaccessible by MD techniques.[68,81] A comprehensive characterization of protein dynamics thus requires the application of complementary approaches that report on dynamics taking place on a wider time range. Continuous labeling HDX/MS is well suited for this purpose. This technique monitors the deuteration of backbone amides. HDX/MS reports on protein dynamics because NH deuteration is mediated by H-bond fluctuations that are coupled to backbone motions. EX2 rate constants can be expressed as $k_{HDX} = (k_{op}/k_{cl})\ k_{ch}$, where k_{op} and k_{cl} are H-bond opening/closing rate constants, while k_{ch} reflects deuteration of unprotected sites.[82] Deuteration kinetics thus provide a cumulative picture of H-bond fluctuations that occur on time scales of sub-microseconds to seconds and beyond.[80] To be clear, the use of HDX/MS in our experiments does not imply that changes in H-bonding constitute the mechanistic foundation of S100 allostery. Rather, as noted throughout the text, rearrangements of hydrophobic moieties were identified as the primary factor for signal propagation and binding site closure (Figure 8). HDX/MS nonetheless provides a window into allosteric events, because different allosteric switching states exhibit different HDX fingerprints.

HDX/MS conditions were designed to mimic those of the MD investigations. Comparison of the deuteration kinetics for [Ca_4 S100 Ax_2], [Ca_4 S100], [S100 Ax_2] and apo-S100 reveals that the addition of calcium and/or Ax solicits highly complex responses throughout the protein (Figures 12, 13). These range from stabilizing effects to local destabilization. Similarly convoluted HDX patterns have been reported for other allosteric proteins.[83-86] The HDX/MS data are also consistent with NMR-based HDX experiments on the apo and Ca_4 forms of S100B.[87]

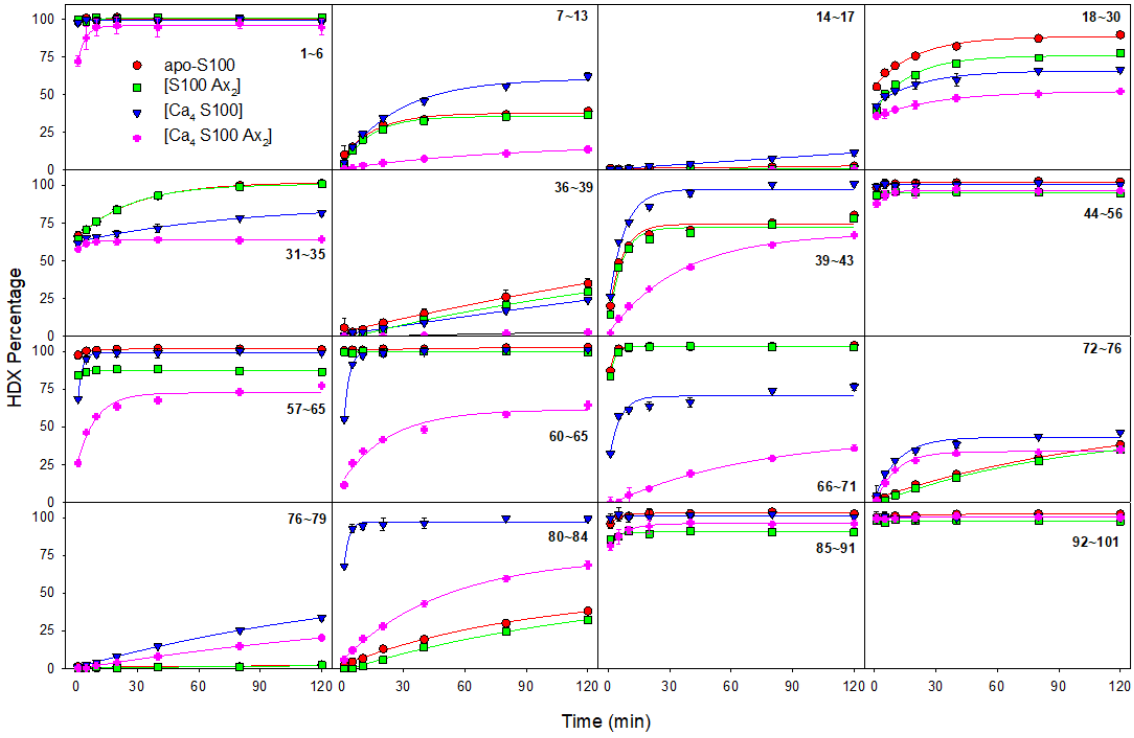


Figure 12. *Deuteration kinetics of S100A11 segments (complete data set). Each panel contains data for all four experimental conditions. Data points represent the average of three independent measurements. Standard deviations are shown as error bars. Biexponential fits (solid lines) were included for illustrative purposes only.*

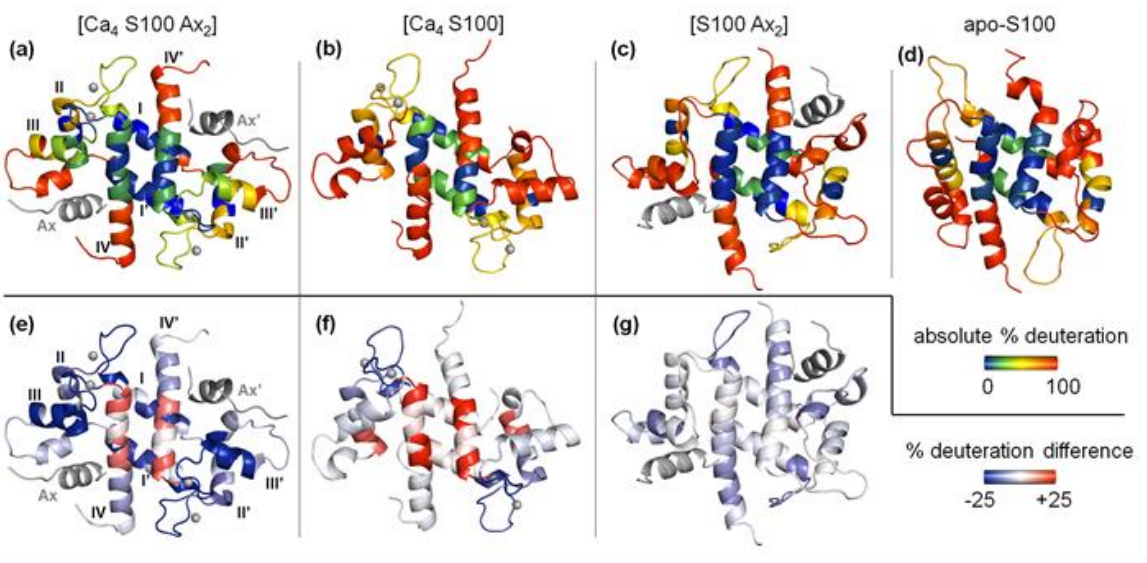


Figure 13. *Graphic representation of the experimental deuteration percentages for t = 10 min. The structures shown represent 1 μs MD conformations. Top row (a-d): absolute deuteration percentages. Bottom row (e-g): deuteration difference relative to apo-S100. For these difference maps, red represents increased deuteration compared to the apo-protein, blue represents reduced deuteration.*

The most pertinent HDX/MS results are summarized in Figure 14. Some regions were highly protected under all conditions, exemplified by peptide 14-17 which is part of the tightly packed core (Figure 14a). Other segments, such as the N- and C-termini were completely deuterated already at the earliest time point, attesting to their flexible nature without stable H-bonds (Figure 14d). Calcium and Ax reduced the deuteration of the pseudo EF-hand in an additive fashion, demonstrating that both binding partners stabilize the H-bonding network in this area (Figure 14b). The EF-hand showed pronounced synergy between calcium and Ax-mediated stabilization. This region exhibited complete deuteration in apo-S100, some protection in [Ca_4 S100], and dramatically reduced HDX in [Ca_4 S100 Ax_2] (Figure 14c). Unlike for the pseudo EF-hand, Ax alone did not cause significant changes in the EF-hand (Figure 14b and 14c).

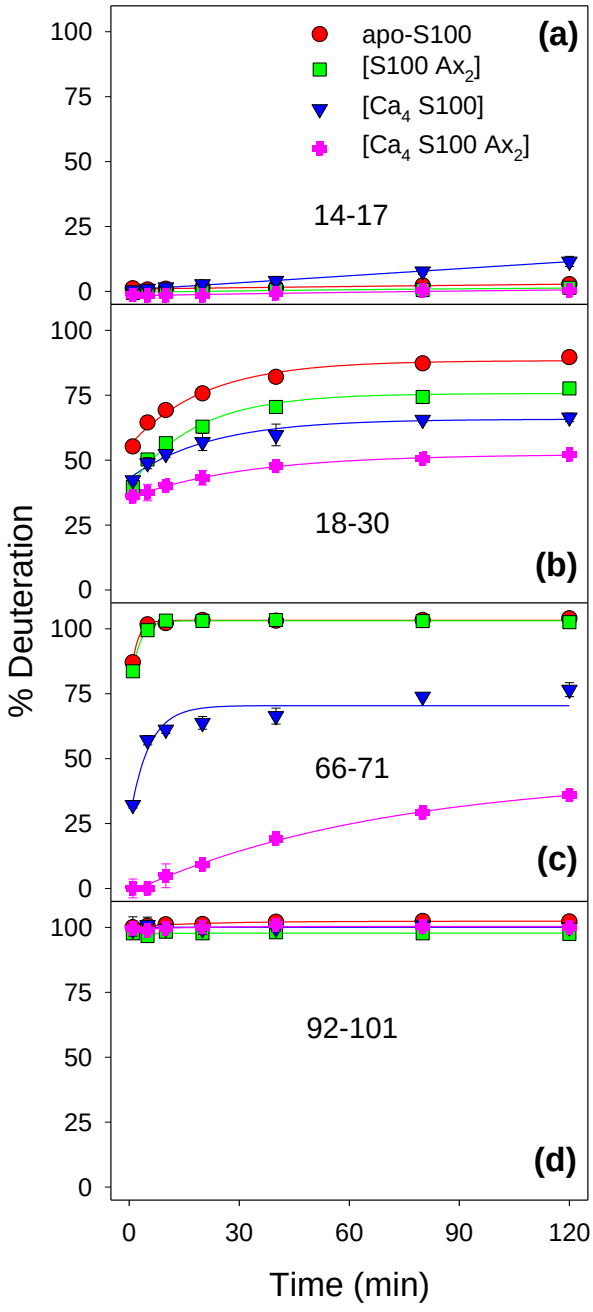


Figure 14. *HDX/MS kinetics of selected segments. (a) Residues 14-17 in helix I, (b) residues 18-30 in the pseudo EF-hand, (c) residues 66-71 in the EF-hand calcium binding site, (d) residues 92-101, representing the C-terminus. Each panel contains results for four experimental conditions. Data points represent the average of three independent measurements. Standard deviations are shown as error bars. Bi-exponential fits (solid lines) were included for illustrative purposes.*

The discussion of our MD data in the preceding sections implicitly focused on allosteric control along steps 1 and 2 of the thermodynamic cycle of Figure 1d, where Ca^{2+} facilitates Ax binding by preventing closure of the target binding site. The calcium-mediated Ax affinity enhancement is $\Delta\Delta G° = -RT \ln C$. The HDX protection patterns of Figure 14b and 14c reflect the flipside of allosteric control, i.e., the fact that Ax governs the calcium affinity in accordance with steps 3 and 4 of Figure 1d.[48-50] Deuteration level changes in [Ca_4 S100 Ax_2] relative to [Ca_4 S100] are consistent with Ax-mediated calcium binding enhancement, as Ax allosterically stabilizes the H-bonding network in all four calcium-binding sites. This is evident from the observation that HDX levels of the metalation sites are lower when Ax and Ca^{2+} are present together, than when only Ca^{2+} is bound (Figure 14b, 14c). Our MD data reveal the basis of this allosteric reciprocity.[50] The presence of Ax in [S100 Ax_2] ensures that several residues close to the calcium binding region maintain a "holo-like" orientation, e.g., the D57/K32/M60/L71 shoulder does not get disrupted to the same extent as in apo-S100 (Figure 6b and 6d). Ax enhances the calcium binding affinity by $\Delta\Delta G° = -RT \ln C$. Earlier work suggests that $C \approx 10$,[46,51-53] such that the allosteric stabilization associated with either pathway (1/2 or 3/4 in Figure 1d) is roughly 6 kJ mol^{-1}.

4.3.5 HDX Experiments and MD-Derived H-Bond Patterns

Classical models of backbone NH deuteration envision that the transient disruption of amide H-bonds is the sole determinant of HDX rates.[88] Recent studies suggest that other factors such as solvent accessibility and electrostatics may also exert some influence, but the central role of H-bond fluctuations remains undisputed.[80,81,89,90] It is interesting to compare experimental deuteration values with MD-derived H-bonding patterns. Figure 15 shows HDX percentages measured after 10 minutes of labeling. Also shown are MD data reflecting the fraction of time that each backbone amide is H-bonded. This H-bonding analysis takes into account all possible protein acceptor sites, including carboxylates which participate in NH contacts in the calcium binding loops (Figure 7). Readers are reminded

that the MD data only reflect a 1 μs time window, while the experimental data report on fluctuations that extend to seconds and beyond.[80] HDX/MS can therefore pinpoint regions undergoing slow conformational dynamics that go undetected in MD simulations.

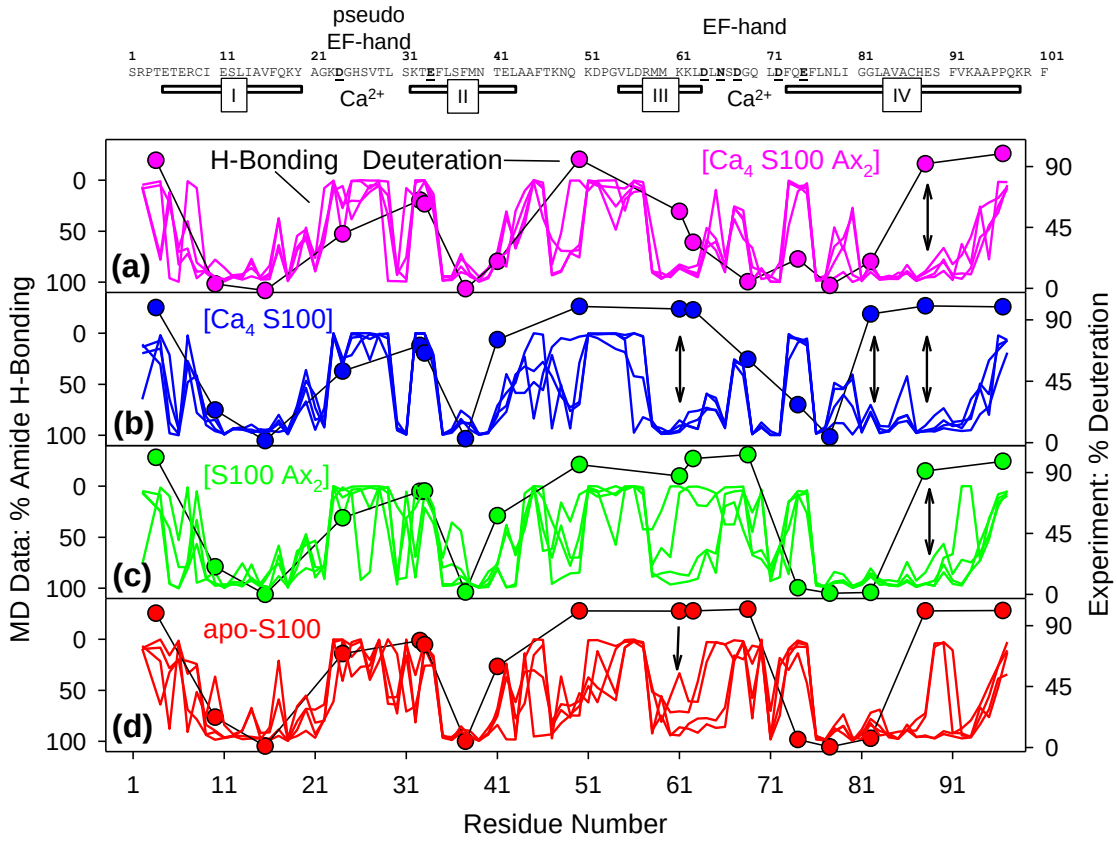


Figure 15. *Comparison of HDX/MS data at t = 10 min (round symbols, plotted vs. peptide midpoint) and MD H-bonding patterns for (a) [Ca₄ S100 Ax₂], (b) [Ca₄ S100], (c) [S100 Ax₂], and (d) apo-S100. Solid lines represent the time fraction during which backbone NH sites are H-bonded during the final 0.5 μs of the MD simulations (data are shown for both subunits and for two independent MD runs). The H-bonding scale is inverted, such that high HDX levels are aligned with low H-bond percentages.*

Simulated H-bond percentages and experimental HDX levels in Figure 15 are remarkably consistent throughout much of the sequence range. All four conditions resulted in weak H-bonding (high HDX levels) at the termini and inter-helical loops. H-bonds in helical regions tended to be more stable (with lower HDX levels). In the EF-hand around residue 70 simulated H-bonds were well developed (with low HDX values) for [Ca_4 S100 Ax_2] and [Ca_4 S100]. Less pronounced H-bonding (and elevated HDX levels) were encountered in this region for [S100 Ax_2] and apo-S100. Thus, both experiments and MD simulations revealed that calcium significantly stabilizes H-bonds in the vicinity of the EF-hands.

Diverging behavior between simulated H-bond patterns and experimental HDX data were observed in a few instances. The double-headed arrows in Figure 15 highlight segments where simulations indicate well-developed H-bonds, while the experimental HDX data show a lack of protection. This behavior was observed around residue 88 close to the center of helix IV for all conditions, with the exception of apo-S100 (Figure 15d). Similar effects were also encountered for [Ca_4 S100] and apo-S100 in the vicinity of residue 62 (Figure 15b, 15d). Thus, regions marked with double-headed arrows undergo H-bond opening/closing transitions that take place on time scales much slower than the μs regime explored in our simulations. Overall, Figure 15 nonetheless demonstrates a high level of consistency between simulated and measured H-bond properties.

4.4 Conclusions

Allosteric proteins possess communication pathways that allow the transfer of signals between remote sites along dynamically coupled residues.[1,6,9,11,15-21,23-34,91] Previous studies implicitly assumed that the starting point of any given communication pathway coincides with an effector binding site. Hence, it is commonly thought that effector binding (or loss) constitutes the initial trigger event that subsequently propagates via a domino-like cascade, culminating in conformational changes elsewhere in the protein. Figure 16a schematically illustrates this classical paradigm, assuming a scenario where the bound effector promotes

an open target binding site. Loss of the effector (magenta) triggers a cascade of events along an allosteric pathway (blue), ultimately causing closure of the target binding site (red).

Figure 16a represents a reasonable description of allosteric regulation for many proteins, as suggested by similar cartoons in numerous reviews and biochemistry texts.[8,92,93] However, calcium-mediated S100 control follows a different mechanism, as uncovered in this work. Closure of the target binding site is the result of an allosteric cascade that is *not* triggered by the effector. Instead, the K32-D57 salt bridge acts as an "agitator" that is labile and destabilizes packing interactions of adjacent residues. Random thermal fluctuations of this agitator represent the initial trigger of the allosteric cascade. The agitator is coupled to the target binding site by dynamically linked residues that allow the initial perturbation to propagate, resulting in closure of the binding site (Figure 16b). The effector (calcium) acts by blocking signals emanating from the agitator, i.e., by interfering with the transmission of allosteric signals (Figure 16c). This blockage is achieved by damping the fluctuations of residues involved in the allosteric cascade.

Figure 16d summarizes details of the signal blocking mechanism uncovered in this study. In [Ca_4 S100] helix III is suspended in the open conformation by packing interactions within the D57/K32/M60/L71 shoulder. These interactions extend to adjacent hydrophobic sites such as L63 and L71, and to helix II which is positioned behind helix III (Figure 16d, left hand side). The top center panel of Figure 16d illustrates events taking place upon calcium loss from [Ca_4 S100]. The K32/D57 salt bridge is labile and undergoes dissociation/re-association transitions which represent the initial trigger event (marked as "1" in the allosteric cascade of Figure 16d). This agitation propagates via disruption of K32/M60 ("2") and M60/L71 ("3") interactions. The breakdown of packing interactions becomes irreversible once L63/F73 undergo a major distance fluctuation ("4"), allowing helix III to swing downward such that the target binding site closes ("5"). This closing event is driven by the burial of hydrophobic sites between helices III and IV. If calcium is present (Figure 16d, bottom center) structural disturbances generated at the K32-D57 agitator cannot proceed beyond L71, because metal-protein contacts in the EF-hand

rigidify the L63-F73 distance. Hence, the allosteric cascade comes to a halt, and the target binding site remains open.

There is prior evidence for scenarios where allosteric pathways are affected by interactions with other communication channels. For example, the outcome of effector binding can be altered by affinity modulators, efficacy modulators, agonists, or antagonists.[8,92,93] The current work further expands the spectrum of possible branched signaling scenarios. We uncover a mechanism where a constitutively active agitator communicates with a remote target binding site. Only in the absence of calcium does the resulting signal cascade cause binding site closure, while bound calcium dampens the motion of key residues along the allosteric pathway. Thus, the effector (calcium) works by blocking the information flow between agitator and target binding site.

In general terms, we propose that an allosteric agitator can be defined as follows: (i) An agitator is a structurally labile element that undergoes incessant opening/closing fluctuations. (ii) The open state of the agitator destabilizes adjacent noncovalent contacts. (iii) Coupling of the agitator to dynamically linked residues allows the propagation of conformational perturbations in a domino-like fashion, ultimately affecting the structure/dynamics of a remote protein region (e.g., a target binding site). (iv) Allosteric effectors can block signals emanating from the agitator by stabilization (rigidification) of elements along the propagation pathway.

In future work, it will be interesting to determine if similar agitator/blocking scenarios also apply to other classes of allosteric proteins. The combined MD and HDX/MS strategy applied here should be well suited for endeavors of this type.

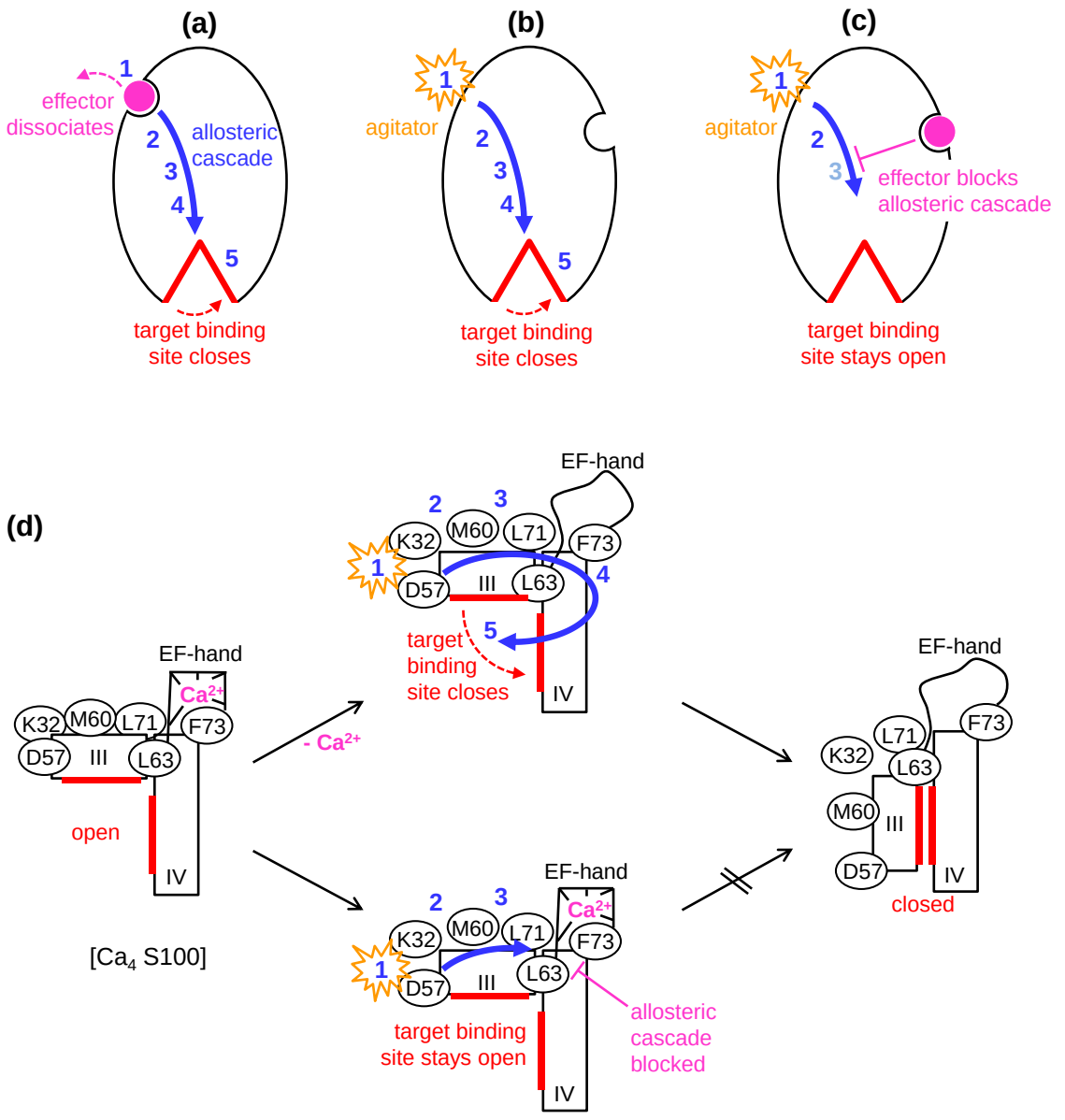


Figure 16. *(a) Conventional allosteric mechanism, where effector loss triggers closing of a remote target binding site. Numbers 1, 2, ... refer to structural events associated with the signaling cascade. Panels (b) and (c): Generic representation of the mechanism uncovered in this study. (b) A fluctuating agitator triggers an allosteric cascade that causes binding site closure. (c) The presence of an effector blocks the cascade, such that the binding site stays open. (d) Cartoon summary of the calcium-mediated blocking mechanism in S100A11. The upper pathway results in binding site closure. In the lower pathway calcium in the EF-hand blocks signal transfer, information flow stalls, and the binding site stays open.*

4.5 References

(1) Perutz, M. F. *Annu. Rev. Physiol.* **1990**, *52*, 1.

(2) Schirmer, T.; Evans, P. R. *Nature* **1990**, *343*, 140.

(3) Schwartz, T. W.; Hubbell, W. L. *Nature* **2008**, *455*, 473.

(4) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.

(5) Koshland, D. E.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *5*, 365.

(6) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295.

(7) Gui, D. Y.; Lewis, C. A.; Vander Heiden, M. G. *Sci. Signal.* **2013**, *6*.

(8) De Smet, F.; Christopoulos, A.; Carmeliet, P. *Nat. Biotechnol.* **2014**, *32*, 1113.

(9) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; Baell, J. B.; Sexton, P. M.; Christopoulos, A.; Shaw, D. E. *Nature* **2013**, *503*, 295.

(10) Cooper, A.; Dryden, D. T. F. *Eur. Biophys. J. Biophys. Lett.* **1984**, *11*, 103.

(11) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331.

(12) Sowole, M. A.; Simpson, S. A.; Skovpen, Y. V.; Palmer, D. R. J.; Konermann, L. *Biochemistry* **2016**, *55*, 5413.

(13) Kern, D.; Zuiderweg, E. R. P. *Curr. Op. Struct. Biol.* **2003**, *13*, 748.

(14) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789.

(15) Nussinov, R.; Tsai, C. J. *Trends Pharmacol. Sci.* **2014**, *35*, 256.

(16) Boulton, S.; Akimoto, M.; Selvaratnam, R.; Bashiri, A.; Melacini, G. *Sci Rep* **2014**, *4*, 7306.

(17) Malherbe, P.; Kratochwil, N.; Knoflach, F.; Zenner, M. T.; Kew, J. N. C.; Kratzeisen, C.; Maerki, H. P.; Adam, G.; Mutel, V. *J. Biol. Chem.* **2003**, *278*, 8340.

(18) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 59.

(19) Villali, J.; Pontiggia, F.; Clarkson, M. W.; Hagan, M. F.; Kern, D. *J. Mol. Biol.* **2014**, *426*, 1554.

(20) Shi, L.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 2140.

(21) Rivalta, I.; Sultan, M. M.; Lee, N. S.; Manley, G. A.; Loria, J. P.; Batista, V. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1428.

(22) Kumar, A.; Aguirre, J. D.; Condos, T. E. C.; Martinez-Torres, R. J.; Chaugule, V. K.; Toth, R.; Sundaramoorthy, R.; Mercier, P.; Knebel, A.; Spratt, D. E.; Barber, K. R.; Shaw, G. S.; Walden, H. *Embo J.* **2015**, *34*, 2506.

(23) Henry, E. R.; Mozzarelli, A.; Viappiani, C.; Abbruzzetti, S.; Bettati, S.; Ronda, L.; Bruno, S.; Eaton, W. A. *Biophys. J.* **2015**, *109*, 1264.

(24) Jurica, M. S.; Mesecar, A.; Heath, P. J.; Shi, W. X.; Nowak, T.; Stoddard, B. L. *Structure* **1998**, *6*, 195.

(25) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. *J. Chem. Theory Comput.* **2009**, *5*, 2486.

(26) Lin, M. M. *J. Am. Chem. Soc.* **2016**, *138*, 5036.

(27) Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 4735.

(28) Manley, G.; Rivalta, I.; Loria, J. P. *J. Phys. Chem. B* **2013**, *117*, 3063.

(29) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R. L.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 21216.

(30) Rajasekaran, N.; Suresh, S.; Gopi, S.; Raman, K.; Naganathan, A. N. *Biochemistry* **2017**, *56*, 294.

(31) Deredge, D.; Li, J. W.; Johnson, K. A.; Wintrode, P. L. *J. Biol. Chem.* **2016**, *291*, 10078.

(32) Donovan, K. A.; Zhu, S. L.; Liuni, P.; Peng, F.; Kessans, S. A.; Wilson, D. J. *J. Biol Chem.* **2016**, *291*, 9244.

(33) Underbakke, E. S.; Iavarone, A. T.; Chalmers, M. J.; Pascal, B. D.; Novick, S.; Griffin, P. R.; Marletta, M. A. *Structure* **2014**, *22*, 602.

(34) Sheff, J. G.; Farshidfar, F.; Bathe, O. F.; Kopciuk, K.; Gentile, F.; Tuszynski, J.; Barakat, K.; Schriemer, D. C. *Mol. Cell. Proteomics* **2017**, *16*, 428.

(35) Gifford, J. L.; Walsh, M. P.; Vogel, H. J. *Biochem. J.* **2007**, *405*, 199.

(36) Gross, S. R.; Sin, C. G. T.; Barraclough, R.; Rudland, P. S. *Cell. Mol. Life Sci.* **2014**, *71*, 1551.

(37) Bresnick, A. R.; Weber, D. J.; Zimmer, D. B. *Nat. Rev. Cancer* **2015**, *15*, 96.

(38) Santamaria-Kisiel, L.; Rintala-Dempsey, A. C.; Shaw, G. S. *Biochem. J.* **2006**, *396*, 201.

(39) Rety, S.; Osterloh, D.; Arie, J.-P.; Tabaries, S.; Seeman, J.; Russo-Marie, F.; Gerke, V.; Lewit-Bentley, A. *Structure* **2000**, *8*, 175.

(40) Huang, Y. K.; Chou, R. H.; Yu, C. *J. Biol. Chem.* **2016**, *291*, 14300.

(41) Hermann, A.; Donato, R.; Weiger, T. M.; Chazin, W. J. *Front. Pharmacol.* **2012**, *3*, 67.

(42) Gabel, M.; Chasserot-Golaz, S. *J. Neurochem.* **2016**, *137*, 890.

(43) Jaiswal, J. K.; Lauritzen, S. P.; Scheffer, L.; Sakaguchi, M.; Bunkenborg, J.; Simon, S. M.; Kallunki, T.; Jäättelä, M.; Nylandsted, J. *Nat. Commun.* **2014**, *5*, 4795.

(44) Rezvanpour, A.; Santamaria-Kisiel, L.; Shaw, G. S. *J. Biol. Chem.* **2011**, *286*, 40174.

(45) Rintala-Dempsey, A.; Rezvanpour, A.; Shaw, G. S. *FEBS* **2008**, *275*, 4956.

(46) Dempsey, A. C.; Walsh, M. P.; Shaw, G. S. *Structure* **2003**, *11*, 887.

(47) Capozzi, F.; Casadei, F.; Luchinat, C. *J. Biol. Inorg. Chem.* **2006**, *11*, 949.

(48) Williamson, J. R. *Nat. Chem. Biol.* **2008**, *4*, 458.

(49) Nandigrami, P.; Portman, J. J. *J. Chem. Phys.* **2016**, *144*, 105101.

(50) Fenton, A. W. *Trends Biochem. Sci.* **2008**, *33*, 420.

(51) Liriano, M. A.; Varney, K. M.; Wright, N. T.; Hoffman, C. L.; Toth, E. A.; Ishima, R.; Weber, D. J. *J. Mol. Biol.* **2012**, *423*, 365.

(52) Allen, B. G.; Durussel, I.; Walsh, M. P.; Cox, J. A. *Biochem. Cell Biol.* **1996**, *74*, 687.

(53) Martin, S. R.; Bayley, P. M.; Brown, S. E.; Porumb, T.; Zhang, M.; Ikura, M. *Biochemistry* **1996**, *35*, 3508.

(54) Grabarek, Z. *J. Mol. Biol.* **2006**, *359*, 509.

(55) Dupuis, L.; Mousseau, N. *J. Chem. Phys.* **2012**, *136*, 035101.

(56) Scotts, C. E.; Kekenes-Huskey, P. M. *Biophys. J.* **2016**, *110*, 1052.

(57) McDowell, C.; Chen, J. L.; Chen, J. H. *J. Mol. Biol.* **2013**, *425*, 999.

(58) Markowitz, J.; Rustandi, R. R.; Varney, K. M.; Wilder, P. T.; Udan, R.; Wu, S. L.; Horrocks, W. D.; Weber, D. J. *Biochemistry* **2005**, *44*, 7305.

(59) Rintala, A. C.; Schöekess, B. O.; Walsh, M. P.; Shaw, G. S. *J. Biomol. NMR* **2002**, *22*, 191.

(60) Smith, S. P.; Barber, K. R.; Dunn, S. D.; Shaw, G. S. *Biochemistry* **1996**, *35*, 8805.

(61) Rintala-Dempsey, A.; Santamaria-Kisiel, L.; Liao, Y.; Lajoie, G.; Shaw, G. S. *Biochemistry* **2006**, *45*, 14695.

(62) Naka, M.; Qing, Z. X.; Sasaki, T.; Kise, H.; Tawara, I.; Hamaguchi, S.; Tanaka, T. *Biochim. Biophys. Acta* **1994**, *1223*, 348.

(63) Wales, T. E.; Fadgen, K. E.; Gerhardt, G. C.; Engen, J. R. *Anal. Chem.* **2008**, *80*, 6815.

(64) Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'Cuk, A. I.; Kaltashov, I. A. *Protein Sci.* **2005**, *14*, 543.

(65) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1–2*, 19.

(66) Huang, J.; MacKerell, A. D. *J. Comput. Chem.* **2013**, *34*, 2135.

(67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(68) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915.

(69) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(70) Hess, B.; Henk, B.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.

(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(72) Hung, K.-W.; Chang, Y.-M.; Yu, C. *J. Biomol. NMR* **2012**, *54*, 211.

(73) Lee, H. W.; Wylie, G.; Bansal, S.; Wang, X.; Barb, A. W.; Macnaughtan, M. A.; Ertekin, A.; Montelione, G. T.; Prestegard, J. H. *Protein Sci.* **2010**, *19*, 1673.

(74) Malik, S.; Revington, M.; Smith, S. P.; Shaw, G. S. *Proteins* **2008**, *73*, 28.

(75) Creighton, T. E. *Proteins*; W. H. Freeman & Co: New York, 1993.

(76) Kyte, J.; Doolittle, R. *J. Mol. Biol.* **1982**, *157*, 105.

(77) Baldwin, R. L.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12462.

(78) Grdadolnik, J.; Merzel, F.; Avbelj, F. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 322.

(79) Southall, N. T.; Dill, K. A.; Haymett, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521.

(80) Skinner, J. J.; Lim, W. K.; Bedard, S.; Black, B. E.; Englander, S. W. *Protein Sci.* **2012**, *21*, 996.

(81) Persson, F.; Halle, B. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 10383.

(82) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75.

(83) Bobst, C. E.; Zhang, M.; Kaltashov, I. A. *J. Mol. Biol.* **2009**, *388*, 954.

(84) Sowole, M. A.; Alexopoulos, J. A.; Cheng, Y.-Q.; Ortega, J.; Konermann, L. *J. Mol. Biol.* **2013**, *425*, 4508.

(85) Burke, J. E.; Babakhani, A.; Gorfe, A. A.; Kokotos, G.; Li, S.; Woods, V. L.; McCammon, J. A.; Dennis, E. A. *J. Am. Chem. Soc.* **2009**, *131*, 8083.

(86) Shukla, A. K.; Westfield, G. H.; Xiao, K.; Reis, R. I.; Huang, L.-Y.; Tripathi-Shukla, P.; Qian, J.; Li, S.; Blanc, A.; Oleskie, A. N.; Dosey, A. M.; Su, M.; Liang, C.-R.; Gu, L.-L.; Shan, J.-M.; Chen, X.; Hanna, R.; Choi, M.; Yao, X. J.; Klink, B. U.; Kahsai, A. W.; Sidhu, S. S.; Koide, S.; Penczek, P. A.; Kossiakoff, A. A.; Woods Jr, V. L.; Kobilka, B. K.; Skiniotis, G.; Lefkowitz, R. J. *Nature* **2014**, *512*, 218.

(87) Marlatt, N. M.; Shaw, G. S. *Biochemistry* **2007**, *46*, 7478.

(88) Hvidt, A.; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287.

(89) McAllister, R. G.; Konermann, L. *Biochemistry* **2015**, *54*, 2683.

(90) Anderson, J. S.; Hernandez, G.; LeMaster, D. M. *Biophys. Chem.* **2013**, *171*, 63.

(91) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295.

(92) Laskowski, R. A.; Gerick, F.; Thornton, J. M. *FEBS Lett.* **2009**, *583*, 1692.

(93) Conn, P. J.; Christopoulos, A.; Lindsley, C. W. *Nat. Rev. Drug Discov.* **2009**, *8*, 41.

Chapter 5. Conclusions

5.1 Summary

The work in this thesis mainly describes how measurements of protein dynamics can be used to extract information related to the general behavior of proteins in different biophysical/biochemical contexts. In general, experimental methods to study protein dynamics are still relatively limited. Many traditional protein analysis methods can either only examine protein under particular conditions (X-ray crystallography, Cryo-EM) or provide limited information (UV/CD spectrometry). This situation makes HDX-MS a very attractive experimental method for probing protein dynamics, even with limited background information. HDX/MS relies on the principle that unfolded and/or highly dynamic regions undergo faster deuteration than tightly folded segments. The broad usability and ease of operation get HDX-MS more and more attention in various aspects, evident from the steadily growing number of laboratories who have adopted this technique. Here we present three different protein systems that have been investigated by HDX-MS. These studies cover three different perspectives: protein folding, protein-ligand interactions and allosteric effects. Also, MD simulations are used as a parallel tool in the last project. MD simulations are capable of providing detailed information at the atomic level which well complements the peptidic resolution in bottom-up HDX-MS.

In Chapter 2, a common phenomenon that creates trouble in the pharmaceutical industry was explored. Gas/water interfaces (in the form of air bubbles or foam) are detrimental to the stability of proteins, often causing aggregation. This represents a potential problem for industrial processes, e.g., the production and handling of protein drugs.[1-3] We use HDX-MS for probing the conformational dynamics of the model protein Mb in the presence of N_2 bubbles. The conformational dynamics of the model protein myoglobin under N_2 sparging condition is explored by HDX-MS, and an interesting EX1 behavior is observed. This event reflects the interconversion of the native state with conformers that are globally

unfolded and long-lived. In contrast to this, EX2 dynamics take place in the bubble-free environment and suggest the occurrence of short-lived excursions to partially unfolded conformers. A simple model is built to describe the observation: "semi-unfolded" ↔ "native" ↔ "globally unfolded" → "aggregated". To our knowledge, this is the first time HDX-MS is applied to explore the surface denaturation of proteins.

One of the most widely used applications of HDX-MS is to probe how protein dynamics change as a result of protein-ligand interactions. Chapter 3 conducts a detailed analysis of a vital protein system that regulates osteoclastogenesis: OPG/RANKL/HS.[4,5] In brief, OPG exist simultaneously as monomer (OPG^{SH}) and covalently linked dimer form (OPG_2^{SS}). Both forms are capable of blocking RANKL from deactivating bone resorption [6,7] and the dimerization can be regulated by HS.[8] HDX-MS data reveal the dynamics changes of OPG oligomers under various ligand conditions and shows RANKL and HS have different effects on OPG^{SH} and OPG_2^{SS}. The effect induced by RANKL on OPG^{SH} is rather weak but can be strengthened with HS. We also proposed a model to explain the RANKL and HS binding mechanism on OPG. One of the challenges in this project is the absence of structural information on full-length (untruncated) OPG. This Chapter particularly highlights the advantages of HDX-MS as providing structural information that was not available before.

Allosteric proteins possess dynamically coupled residues for the propagation of input signals to distant target binding sites. The input signals usually correspond to "effector is present" or "effector is not present". Many aspects of allosteric regulation remain incompletely understood. Chapter 4 focuses on S100A11, a dimeric EF-hand protein with two hydrophobic target binding sites. An annexin peptide (Ax) served as the target. Target binding is allosterically controlled by Ca^{2+} which promotes the formation of a [Ca_4 S100 Ax_2] complex. Ax peptides are accommodated between helices III/IV and III'/IV'. Without Ca^{2+} these binding sites are closed, precluding interactions with Ax. The allosteric mechanism was probed by microsecond MD simulations in explicit water, complemented by HDX-MS. Consistent with experimental data, MD runs in the absence of Ca^{2+} and Ax culminated in target binding site closure. In simulations on [Ca_4 S100] the target binding sites remained open. These results capture the essence of allosteric control, revealing how

Ca^{2+} prevents binding site closure. Both HDX-MS and MD data showed that the metalation sites become more dynamic after Ca^{2+} loss. However, these enhanced dynamics do not represent the primary trigger of the allosteric cascade. Instead, a labile salt bridge (K32-D57) acts as an incessantly active "agitator" that destabilizes the packing of adjacent residues, causing a domino chain of events that culminates in target binding site closure. This agitator represents the starting point of the allosteric signal propagation pathway. Ca^{2+} binding rigidifies elements along this pathway, thereby blocking signal transmission. This newly discovered mechanism differs from the conventional view that allosteric effects always start from the effector binding site. It remains to be seen in how far this novel paradigm can be extended to other allosteric regulation events.

5.2 Future Directions

5.2.1 HDX-MS on Membrane Proteins

Membrane proteins are fascinating to study since they carry critical function in living organisms. They are the channels for cells and organelles to link outside and inside environments. They are also the binding sites for over 50% of modern drugs.[9] However, membrane proteins continue to pose challenges to biochemistry and biology due to their hydrophobic nature. The 3D structures of many membrane proteins are still not available. That makes HDX-MS a promising tool to study membrane protein since there is no prerequisite for structural information.

Recent attempts to apply HDX-MS to membrane proteins have achieved success to a certain extent.[10] Many detergents (e.g., n-dodecyl-β-D-maltopyranoside, n-Octyl-β-d-glucopyranoside) are used to enhance solubility and digestion of membrane proteins. But to date, no systematic research has been done on how to apply detergents in HDX-MS properly. Guidelines and tips and are scattered across various papers or kept as "secrets" in

different laboratories. A universal protocol needs to be established for high throughput investigations.

Also, how detergents affect protein structure and conformations is still a question that remains quite controversial. Generally, nonionic detergents are considered to be “mild” for proteins. But how “mild” these detergents are and whether these detergents would cause artifacts is still unknown. It would be interesting to use HDX-MS to examine the dynamic changes induced by detergents on one or several model proteins, such as bacteriorhodopsin. The results of such studies can benefit HDX field and guide future investigations on membrane proteins.

5.2.2 HDX-MS on the OPG/RANKL/RANK System

Chapter 3 in this dissertation concentrates on dynamics changes of OPG under different circumstances. However, we must keep it in mind that OPG is only part of the osteoclastogenesis regulatory system. The dynamics changes on RANKL in OPG-RANKL were not investigated. Further, it would be interesting to check the dynamics difference on RANKL between two complexes: OPG-RANKL and RANK-RANKL. Such experiments could potentially provide significant insights into bone metabolism, and they could help therapeutic developments for osteoporosis and osteoproliferation.

5.2.3 Computational Simulation of the HDX Process

The actual exchange between hydrogen and deuterium at amide backbone sites still can not be simulated by MD methods or related computational tools. This much limits the

capability of MD simulations to reappear reproduce or predict the result of HDX experiments. Classical MD simulations are unsuitable for modeling HDX events at an atomistic level because typical force fields do not allow the dissociation or formation of covalent bonds. However, with hybrid quantum-mechanical/molecular-dynamics (QM/MD) techniques atomistic simulations of HDX experiments may become achievable.[11] In this case, heavy atoms of the protein will still be treated by classical MD simulations. Simultaneously, exchange events between hydrogens on the protein and deuterium in the aqueous/gas phase will be simulated by quantum mechanical or density-functional theory (DFT)-based methods. It is foreseeable that the development of such methods will be a time-consuming process. However, the application prospects of such method are enormous, as they would provide a truly comprehensive view of protein HDX events, perhaps resulting in novel paradigms for the interpretation of experimental HDX data.

5.3 References

(1) Graham, D. E.; Phillips, M. C. *J. Colloid Interface Sci.* **1979**, *70*, 403.

(2) Donaldson, T. L.; Boonstra, E. F.; Hammond, J. M. *J. Colloid Interface Sci.* **1980**, *74*, 441.

(3) Clarkson, J. R.; Cui, Z. F.; Darton, R. C.; Clarkson, J. R. *J. Colloid Interface Sci.* **1999**, *215*, 323.

(4) Simonet, W. S.; Lacey, D. L.; Dunstan, C. R.; Kelley, M.; Chang, M. S.; Luthy, R.; Nguyen, H. Q.; Wooden, S.; Bennett, L.; Boone, T.; Shimamoto, G.; DeRose, M.; Elliott, R.; Colombero, A.; Tan, H. L.; Trail, G.; Sullivan, J.; Davy, E.; Bucay, N.; Renshaw-Gegg, L.; Hughes, T. M.; Hill, D.; Pattison, W.; Campbell, P.; Sander, S.; Van, G.; Tarpley, J.; Derby, P.; Lee, R.; Boyle, W. J. *Cell* **1997**, *89*, 309.

(5) Udagawa, N.; Takahashi, N.; Yasuda, H.; Mizuno, A.; Itoh, K.; Ueno, Y.; Shinki, T.; Gillespie, M. T.; Martin, T. J.; Higashio, K.; Suda, T. *Endocrinology* **2000**, *141*, 3478.

(6) Boyce, B. F.; Xing, L. *Arthritis Research & Therapy* **2007**, *9*, S1.

(7) Boyle, W. J.; Simonet, W. S.; Lacey, D. L. *Nature* **2003**, *423*, 337.

(8) Li, M.; Yang, S.; Xu, D. *J. Biol. Chem.* **2016**, *291*, 24160.

(9) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. *Nat. Rev. Drug Discovery* **2006**, *5*, 993.

(10) Zhang, X.; Chien, E. Y. T.; Chalmers, M. J.; Pascal, B. D.; Gatchalian, J.; Stevens, R. C.; Griffin, P. R. *Anal. Chem.* **2010**, *82*, 1100.

(11) Andreoni, W.; Curioni, A.; Mordasini, T. *IBM Journal of Research and Development* **2001**, *45*, 397.

Appendix I-Permissions

Copyright Clearance Center RightsLink® Home Account Info Help

Title: Protein structural dynamics at the gas/water interface examined by hydrogen exchange mass spectrometry
Author: Yiming Xiao, Lars Konermann
Publication: Protein Science
Publisher: John Wiley and Sons
Date: Apr 2, 2015
Copyright © 2015, John Wiley and Sons

Logged in as:
Yiming Xiao
University of Western Ontario
Account #:
3001315627
LOGOUT

Order Completed

Thank you for your order.

This Agreement between University of Western Ontario -- Yiming Xiao ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4400590376692
License date	Aug 02, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Protein Science
Licensed Content Title	Protein structural dynamics at the gas/water interface examined by hydrogen exchange mass spectrometry
Licensed Content Author	Yiming Xiao, Lars Konermann
Licensed Content Date	Apr 2, 2015
Licensed Content Volume	24
Licensed Content Issue	8
Licensed Content Pages	10
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Hydrogen-Deuterium Exchange Mass Spectrometry and Molecular Dynamics Simulations for Studying Protein Structure and Dynamics
Expected completion date	Aug 2018
Expected size (number of pages)	200
Requestor Location	University of Western Ontario 1151 Richmond Street London, ON N6A 3K7 Canada Attn: University of Western Ontario
Publisher Tax ID	EU826007151
Total	0.00 USD

Chapter: [27] Circular dichroism and optical rotatory dispersion of proteins and polypeptides
Book: Methods in Enzymology
Author: Alice J Adler,Norma J Greenfield,Gerald D Fasman
Publisher: Elsevier
Date: 1973

Order Completed

Thank you for your order.

This Agreement between University of Western Ontario -- Yiming Xiao ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4399091120822
License date	Jul 30, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Methods in Enzymology
Licensed Content Author	Alice J Adler,Norma J Greenfield,Gerald D Fasman
Licensed Content Date	Jan 1, 1973
Licensed Content Pages	61
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No
Original figure numbers	Figure 7
Title of your thesis/dissertation	Hydrogen-Deuterium Exchange Mass Spectrometry and Molecular Dynamics Simulations for Studying Protein Structure and Dynamics
Expected completion date	Aug 2018
Estimated size (number of pages)	200
Requestor Location	University of Western Ontario 1151 Richmond Street London, ON N6A 3K7 Canada Attn: University of Western Ontario
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

ORDER MORE CLOSE WINDOW

Title: On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization

Author: Lars Konermann, Antony D. Rodriguez, Jiangjiang Liu

Publication: Analytical Chemistry

Publisher: American Chemical Society

Date: Aug 1, 2012

Logged in as:
Yiming Xiao
University of Western Ontario
Account #:
3001315627

LOGOUT

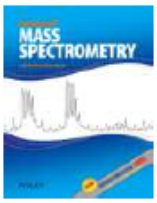
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & 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 request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

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

BACK CLOSE WINDOW



Title:	Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches
Author:	Lars Konermann, Xin Tong, Yan Pan
Publication:	Journal of Mass Spectrometry
Publisher:	John Wiley and Sons
Date:	Jun 3, 2008

Order Completed

Thank you for your order.

This Agreement between University of Western Ontario -- Yiming Xiao ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4399150152112
License date	Jul 31, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Mass Spectrometry
Licensed Content Title	Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches
Licensed Content Author	Lars Konermann, Xin Tong, Yan Pan
Licensed Content Date	Jun 3, 2008
Licensed Content Volume	43
Licensed Content Issue	8
Licensed Content Pages	16
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Hydrogen-Deuterium Exchange Mass Spectrometry and Molecular Dynamics Simulations for Studying Protein Structure and Dynamics
Expected completion date	Aug 2018
Expected size (number of pages)	200
Requestor Location	University of Western Ontario 1151 Richmond Street London, ON N6A 3K7 Canada Attn: University of Western Ontario

Publisher Tax ID	EU826007151
Total	0,00 USD

Would you like to purchase the full text of this article? If so, please continue on to the content ordering system located here: Purchase PDF

If you click on the buttons below or close this window, you will not be able to return to the content ordering system.

ORDER MORE CLOSE WINDOW

Order Completed

Thank you for your order.

This Agreement between University of Western Ontario -- Yiming Xiao ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4399151499264
License date	Jul 31, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Wiley Books
Licensed Content Title	Hydrogen Exchange Mass Spectrometry of Proteins: Fundamentals, Methods, and Applications
Licensed Content Author	David D. Weis
Licensed Content Date	Mar 1, 2016
Licensed Content Pages	1
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1.5
Will you be translating?	No
Title of your thesis / dissertation	Hydrogen-Deuterium Exchange Mass Spectrometry and Molecular Dynamics Simulations for Studying Protein Structure and Dynamics
Expected completion date	Aug 2018
Expected size (number of pages)	200
Requestor Location	University of Western Ontario 1151 Richmond Street London, ON N6A 3K7 Canada Attn: University of Western Ontario
Publisher Tax ID	EU826007151
Total	0,00 USD

ORDER MORE CLOSE WINDOW

Hi,

Thank you for contacting ACS Publications Support.

Your permission request is granted and there is no fee for this reuse. In your planned reuse, you must cite the ACS article as the source, add this direct link https://pubs.acs.org/doi/abs/10.1021/jacs.7b04380 , and include a notice to readers that further permissions related to the material excerpted should be directed to the ACS.

If you need further assistance, please let me know.

Sincerely,

Ashley Gibson

ACS Publications

Customer Services & Information

Website: https://help.acs.org

Incident Information:

Incident #: 2136647
Date Created: 2018-08-02T01:20:45
Priority: 3
Customer: Yiming Xiao
Title: For permission for resuing the article "Calcium-Mediated Control of S100 Proteins: Allosteric Communication via an Agitator/Signal Blocking
Description: Dear sir or madam,

I would like to reuse "Calcium-Mediated Control of S100 Proteins: Allosteric Communication via an Agitator/Signal Blocking Mechanism" as part of my Ph.D. thesis.

link: https://pubs.acs.org/doi/abs/10.1021/jacs.7b04380

part for reusing: entire article (as author)

This article will be used in my graduation thesis.

Thank for your work!

Yiming Xiao
Ph.D. Candidate in Analytical/Physical Chemistry
Lars Konermann Group
University of Western Ontario

Appendix II-Curriculum Vitae

Name:	Yiming Xiao
Post-secondary Education and Degrees:	Ph.D. Physical/Analytical in Chemistry, in Progress The University of Western Ontario London, Ontario, Canada 2013-present B.Sc. Honors Specialization in Biochemistry The University of Windsor Windsor, Ontario, Canada 2009-2013
Honors and Awards:	Province of Ontario Graduate Scholarship 2017-2018 Dean's Honor List (The University of Windsor) 2012-2013 President's Honor Roll (The University of Windsor) 2009-2010 Recognition of Achievement Bursary (The University of Windsor) 2010, 2012
Related Work Experience	Teaching Assistant The University of Western Ontario 2012-2017

Peer-Reviewed Publications:

Xiao, Y.; Li, M.; Larocque, R.; Zhang, F.; Malhotra, A.; Linhardt, R.J.; Konermann, L.; Xu, D., Dimerization interface of osteoprotegerin revealed by hydrogen-deuterium exchange mass spectrometry. Manuscript submitted.

Condos, T.EC.; Dunkerley, K.M.; Freeman, E.A.; Barber, K.; Aguirre, J.; Chaugule, V.; Xiao, Y.; Konermann, L.; Walden, H.; Shaw, G.S., Synergistic Recruitment of UbcH7~Ub and Phosphorylated Ubl Domain Triggers Parkin Activation. Manuscript submitted.

Spratt, D.E.; Kathryn R. Barber, K.R.; Marlatt, N.M.; Ngo, V.; Macklin, J.A.; Xiao, Y.; Konermann, L.; Duennwald, M.L; Shaw, G.S., Co-expression Systems to Identify and

Quantify a Preference for S100 Protein Heterodimerization. Manuscript submitted.
Xiao, Y.; Shaw, G. S.; Konermann, L., Calcium-Mediated Control of S100 Proteins: Allosteric Communication via an Agitator/Signal Blocking Mechanism. *J. Am. Chem. Soc.* 2017, 139 (33), 11460-11470.

Mandacaru, S. C.; do Vale, L. H. F.; Vahidi, S.; Xiao, Y.; Skinner, O. S.; Ricart, C. A. O.; Kelleher, N. L.; de Sousa, M. V.; Konermann. L., Characterizing the Structure and Oligomerization of Major Royal Jelly Protein 1 (MRJP1) by Mass Spectrometry and Complementary Biophysical Tools. *Biochemistry* 56, 1645–1655 (2017)

Xiao, Y.; Konermann, L., Protein structural dynamics at the gas/water interface examined by hydrogen exchange mass spectrometry. *Protein Sci.* 2015, 24 (8), 1247-1256. Reproduced with permission © 2015, John Wiley and Sons.

Conference Presentations:

Xiao, Y.; Xu, D.; Konermann, L., Mass Spectrometry Insights into Bone Remodeling: Probing Ligand Interactions of OPG Monomers and Dimers by H/D Exchange. Poster. *Proceedings of the 66th ASMS Conference on Mass Spectrometry and Allied* Topics; San Diego, CA, (2018).

Xiao, Y.; Shaw, G.S.; Konermann, L., Calcium-Mediated Control of S100 Proteins: Allosteric Communication via a Signal Blocking Mechanism. Poster. *Proceedings of the 65th ASMS Conference on Mass Spectrometry and Allied Topics*; Indianapolis, IN, (2017).

Xiao, Y.; Shaw, G.S.; Konermann, L., Calcium-Mediated Interactions of S100A11 with Annexin Peptides Probed by HDX-MS and MD Simulations. Poster. *Proceedings of the 64th ASMS Conference on Mass Spectrometry and Allied Topics*; San Antonio, TX, (2016).

Xiao, Y.; Konermann, L., Protein Structural Dynamics at the Gas/Water Interface Examined by H/D Exchange Mass Spectrometry. Poster. *Proceedings of the 63th ASMS Conference on Mass Spectrometry and Allied Topics*; St. Louis, Texas, MO, (2015).