

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Hydrogen/Deuterium Exchange

Multi-Steps Model for Protein Amide

Thesis submitted in fulfillment of the requirements for the degree of
MASTER OF SCIENCE in PHYSICS
Curriculum: APPLIED PHYSICS, A.A. 2023/26

Candidate: ROOZBEH HATAMJAHROMI

Supervisor: Prof. EMANUELE PACI

Co-supervisor: ANTONIO GRIMALDI

October 2026

To my family,
for their constant support and encouragement.

Abstract

Hydrogen/deuterium exchange mass spectrometry (HDX-MS) is widely used to investigate protein structure and conformational dynamics, but its quantitative interpretation is complicated by the fact that the deuterium content measured by the mass spectrometer is not determined by the labeling reaction alone. Exchange continues during subsequent stages of the experimental workflow, including quenching, trapping, liquid chromatography, electrospray ionization, and ion transfer, resulting in deuterium loss that is commonly treated through empirical back-exchange corrections.

This thesis develops a multi-step kinetic framework for describing the evolution of backbone-amide deuteration throughout an HDX-MS experiment, and the approach builds on the generalized Linderstrøm-Lang description of hydrogen exchange and explicitly treats forward and reverse exchange in mixed H_2O/D_2O solvents. Sequence, temperature, acidity, and isotope-dependent intrinsic exchange rates are calculated at the residue level and propagated through the labeling and quench stages. The chromatographic stage is represented using an effective solvent-dependent exchange model, while an additional phenomenological electrospray contribution is introduced to account for systematic charge-state-dependent differences in the detected deuterium uptake.

The framework is evaluated using millisecond HDX-MS measurements of three rapidly exchanging peptides: Angiotensin, Bradykinin, and ANP. Exchange was followed over labeling times ranging from 50 ms to 300 s under conditions containing no added salt and different concentrations of NaCl and CsCl. The model reproduces the principal kinetic features of all three peptides under the reference condition and shows that the deuterium content predicted immediately after labeling can differ substantially from the amount ultimately detected. Quenching produces an initial loss of label, while liquid chromatography accounts for a larger fraction of the modeled post-labeling decrease, and the effective electrospray stage further improves agreement with experiment and reproduces much of the observed separation between different peptide charge states.

Across the complete dataset, the model reproduces the absolute deuterium uptake and the direction and approximate magnitude of charge-state-dependent effects for most experimental conditions. The principal systematic discrepancy occurs for ANP under concentrated CsCl, particularly for the highest observed charge state, where the model predicts substantially less retained deuterium than measured experimentally, so this behavior indicates that salt identity, peptide properties, and charge state interact in ways that are not fully captured by the current parameterization.

Overall, the results demonstrate that back-exchange is more appropriately

treated as a sequence of condition-dependent kinetic processes than as a single global correction. The multi-step framework provides a physically interpretable basis for identifying where deuterium is lost and where additional experimental or mechanistic information is required, contributing toward a more quantitative interpretation of HDX-MS measurements.

As any dedicated reader can clearly see, the Ideal of practical reason is a representation of, as far as I know, the things in themselves; as I have shown elsewhere, the phenomena should only be used as a canon for our understanding. The paralogisms of practical reason are what first give rise to the architectonic of practical reason. As will easily be shown in the next section, reason would thereby be made to contradict, in view of these considerations, the Ideal of practical reason, yet the manifold depends on the phenomena. Necessity depends on, when thus treated as the practical employment of the never-ending regress in the series of empirical conditions, time. Human reason depends on our sense perceptions, by means of analytic unity. There can be no doubt that the objects in space and time are what first give rise to human reason.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Paci, and my co-supervisor, Dr. Grimaldi, for their guidance, scientific insight, and continuous support throughout this thesis. Their discussions and feedback were essential in shaping both the scientific direction of the project and my understanding of hydrogen/deuterium exchange, mass spectrometry, and kinetic modeling.

Finally, I am also grateful to my family and my friends for their support in pushing me forward despite many problems I faced. In particular, I appreciate my parents, Ghazal, Ehsan, and Mojtaba. When I was completely disappointed they convinced me to continue. Having all of you in my path made it possible to approach the problem over these years during the completion of this degree.

Preface

This thesis developed from an attempt to understand a deceptively simple question in hydrogen/deuterium exchange mass spectrometry: how closely does the deuterium uptake measured by the mass spectrometer represent the isotopic state created during the labeling experiment? What initially appears to be a problem of calculating exchange during a defined labeling time becomes considerably more complex once the complete experimental trajectory is considered. The peptide experiences different solvent compositions, temperatures, chromatographic environments, and eventually the transition from solution to the gas phase before its mass is measured.

Working on this problem required bringing together concepts that are often discussed separately. The project involved the chemistry of proton transfer and intrinsic amide exchange, the Linderstrøm-Lang description of structural protection, isotope effects in mixed H_2O/D_2O solvents, chromatographic back-exchange, electrospray ionization, and numerical modeling. One of the most valuable aspects of the work was therefore learning to move continuously between physical chemistry, mass spectrometry, and computational analysis rather than treating them as independent subjects.

The work presented in this thesis was carried out within a collaborative research environment and builds on experimental measurements, theoretical developments, and computational methods developed within the associated research groups and reported in the literature cited throughout the thesis. My work focused on bringing these components together into a stage-resolved description of the HDX-MS experiment, implementing and applying the kinetic framework to the peptide dataset, evaluating its agreement with experimental measurements, and investigating the dependence of the observed deuterium retention on peptide identity, salt composition, and charge state.

A particularly important part of the project was recognizing the distinction between reproducing experimental data and understanding the physical origin of that agreement. The labeling and quench stages can be related relatively directly to established solution-phase exchange chemistry, whereas the chromatographic and electrospray stages require increasingly effective descriptions, and the remaining discrepancies, especially those observed for ANP in concentrated CsCl, therefore became as informative as the successful fits: they identify conditions under which the present model is incomplete and indicate where additional experiments will be required.

The development of the thesis was consequently iterative. Assumptions were tested against the experimental data, computational implementations were revised, and increasingly complex descriptions were reduced where additional parameters could not be justified independently. This process reinforced the

importance of distinguishing mechanistic parameters from phenomenological ones and of avoiding a more complicated model when the available measurements cannot uniquely constrain it.

I hope that the resulting framework contributes to the broader effort to move HDX-MS from predominantly comparative measurements toward a more quantitative interpretation of absolute exchange and deuterium retention. At the same time, the work has made clear that achieving this goal requires not only improved mathematical descriptions but also experiments designed specifically to isolate the individual stages of the HDX-MS workflow.

Roozbeh Hatamjahromi
Bologna, October 2026

Contents

Abstract iii

Acknowledgments v

Preface vii

Contents ix

1 Introduction 1
- 1.1 Protein Dynamics . . . 2
- 1.2 Hydrogen-Deuterium exchange . . . 3
- 1.3 Why combine HDX with mass spectrometry? . . . 4
- 1.4 Problem in HDX-MS . . . 5
- 1.5 Conventional Back-Exchange Corrections . . . 7
- 1.6 Motivation for a Mechanistic Model . . . 9

2 Fundamentals of Hydrogen/Deuterium Exchange 11
- 2.1 Hydrogen exchange chemistry . . . 11
- 2.2 Backbone amide exchange . . . 13
- 2.3 Intrinsic exchange rate . . . 14
- 2.4 Primary sequence effects . . . 15
- 2.5 Temperature dependence . . . 19
- 2.6 Classical Linderstrom-Lang model . . . 20
- 2.7 EX1 and EX2 exchange regimes . . . 21
- 2.8 Protection factors . . . 23
- 2.9 PDLA versus 3-Ala intrinsic-rate references . . . 24

3 HDX-MS Experimental Workflow 27
- 3.1 Standard HDX-MS workflow . . . 27
- 3.2 Labeling . . . 29
- 3.3 Millisecond HDX-MS . . . 30
- 3.4 Quench . . . 31
- 3.5 Digestion and trapping . . . 33
- 3.6 Liquid chromatography . . . 34
- 3.7 Organic solvent effects . . . 35
- 3.8 Electrospray ionization . . . 36
- 3.9 ESI and protein/peptide conformation . . . 39
- 3.10 Gas-phase HDX . . . 40
- 3.11 Hydrogen scrambling . . . 42

4 Computational Methods and Model Implementation **45**
4.1 Reversible Exchange 45
4.2 Generalized Linderstrøm-Lang Model 46
4.3 Sequential Representation of the HDX-MS Workflow . . . 48
4.4 Intrinsic Exchange Kinetics in Mixed H_2O/D_2O Solvents . 49
4.5 Calculation of Forward and Reverse Intrinsic Rates 52
4.6 Software Environment 53
4.7 Labeling Simulation 53
4.8 Quench Propagation 54
4.9 Liquid-Chromatography Back-Exchange Model 54
4.10 ESI Model 55
4.11 Optimization 57
4.12 Model Evaluation 58
4.13 Interpretation and Limitations 59

5 Results **61**
5.1 Kinetics at Zero Salt Concentration 61
5.2 Salt Concentration Effect 65
5.3 Charge-State and Salt Concentration 70
5.4 Salt and Charge Effects for ANP Uptake 74

6 Conclusion **77**

Appendices **81**

Bibliography **89**

Chapter 1

Introduction

Proteins are not static molecular objects because their biological activity depends on the ability to populate different conformational states, respond to changes in the surrounding environment, and reorganize upon interactions with other molecules. Structural changes may involve large-scale transitions between folded states, but they may also arise from local fluctuations, transient exposure of buried regions, changes in hydrogen-bonding networks, or disorder-to-order transitions. A description of protein function therefore requires information about both molecular structure and the dynamics through which different structural states are sampled [1–3].

Hydrogen-deuterium exchange coupled with mass spectrometry (HDX-MS) has become a widely used approach for investigating these properties in solution. The method follows the exchange of protein hydrogens, particularly backbone amide hydrogens, with deuterium from the surrounding solvent, and because backbone amide exchange is influenced by intrinsic chemical reactivity as well as structural protection, the measured kinetics contain information about local protein structure and conformational dynamics [4, 5]. This principle has enabled HDX-MS to be applied to protein characterization, conformational changes caused by mutations, protein-ligand and protein-protein interactions, epitope mapping, formulation studies, intrinsically disordered proteins, and biopharmaceutical systems [4, 6, 7].

Despite this broad applicability, quantitative interpretation remains challenging because the amount of deuterium ultimately detected by the mass spectrometer is not determined solely by the labeling reaction, as exchange is slowed but not completely arrested during subsequent quenching, digestion, trapping, chromatographic separation, electrospray ionization, and ion transfer. The chemical and physical conditions encountered during these stages differ substantially from those used during labeling and can promote the loss of previously incorporated deuterium through back exchange. Consequently, the measured deuterium uptake reflects the complete experimental trajectory from labeling to ion detection rather than the labeling step alone [5, 8, 9]. This thesis addresses that distinction from a kinetic perspective. Instead of treating back-exchange only as an empirical correction applied after measurement, the exchange state is followed through successive experimental stages using condition- and sequence-dependent rate constants. The approach builds on the classical Linderstrøm-Lang description of amide exchange and its extension to reversible exchange in isotopically mixed H_2O/D_2O solvents [10, 11]. Particular attention is given to the labeling, quench, liquid-chromatography, and electrospray stages, with the aim of identifying how much of the experimentally observed deuterium loss can be explained from the known physical conditions of the experiment and

where additional processes must be considered.

1.1 Protein Dynamics

The three-dimensional structure of a protein provides an essential basis for understanding its function, but a single structural representation does not describe the complete molecular behaviour of a protein in solution. Proteins fluctuate around their populated conformations and can exchange between states whose relative populations depend on temperature, solvent composition, binding partners, mutations, post-translational modifications, and other environmental conditions. Some of these transitions involve substantial rearrangements, whereas others are restricted to local regions and may occur without an obvious change in the overall fold. Molecular recognition can involve selection of pre-existing conformational substates, followed by a redistribution of their populations upon binding [12]. Related ensemble-based descriptions of allostery show that regulation can arise not only from large structural rearrangements but also from changes in protein dynamics and conformational disorder [13].

X-ray crystallography and cryogenic electron microscopy can provide highly detailed structural models of selected molecular states, and these methods have transformed structural biology, but characterization of conformational dynamics and intrinsic disorder under solution conditions remains a separate experimental problem [6], on the other side Nuclear magnetic resonance (NMR) spectroscopy provides complementary information because it can monitor proteins in solution and, in the context of hydrogen exchange, can measure exchange at the level of individual amides. Its application, however, becomes progressively more demanding for larger molecular systems and commonly requires isotopic labeling and relatively large amounts of sample [4, 5].

The value of HDX-MS lies partly in the different balance of information that it provides. It does not usually yield an atomic structure directly; instead, it records how readily different regions of a protein exchange with solvent over time. In bottom-up HDX-MS this information is typically resolved at the peptide level, while the technique remains applicable across a broad molecular-size range and under solution conditions that can be difficult to study using some higher-resolution structural methods [4, 6]. Changes in deuterium uptake can therefore be used to identify regions whose conformational behaviour has been perturbed even when the experiment does not uniquely specify the atomic structure responsible for the change.

This distinction is particularly relevant for systems that cannot be described adequately by one dominant conformation. Intrinsically disordered proteins and intrinsically disordered regions populate heterogeneous and rapidly interconverting conformational ensembles rather than a single stable, well-defined tertiary structure [14]. Such flexible and weakly protected regions may exchange rapidly with the solvent. Conventional HDX-MS measurements, which historically have concentrated on labeling times of seconds to hours, may miss part of this fast kinetic regime, so the development of automated millisecond HDX-MS has ex-

tended the accessible time window toward highly dynamic and weakly protected regions [6, 7]. In this regime, the accuracy of kinetic modeling becomes especially relevant because a substantial fraction of the labeling transition can occur within the first tens or hundreds of milliseconds.

1.2 Hydrogen-Deuterium exchange

Hydrogen-deuterium exchange is the replacement of exchangeable hydrogen atoms by deuterium from the surrounding solvent, as many functional groups in proteins can exchange rapidly, but backbone amide hydrogens are particularly useful for HDX-MS because their exchange can be slowed sufficiently under acidic, low-temperature conditions for part of the incorporated deuterium to survive the subsequent analytical workflow [4, 15]. With the exception of proline, the peptide backbone provides an amide hydrogen at nearly every residue, creating a distributed set of probes along the protein sequence.

It is worth mentioning that the exchange rate of a backbone amide is controlled by two conceptually distinct contributions, the first of which is its intrinsic chemical reactivity; an amide that is fully accessible to solvent and free from structural protection exchanges with an intrinsic rate constant, k_{int}. This rate is not identical for every residue because it depends on temperature, acidity, isotopic solvent composition, and the chemical effects of neighboring amino-acid side chains [4, 16, 17]. Measurements on model peptides provided the basis for empirical intrinsic-rate calculations, making it possible to estimate the expected exchange rate of an amide in an unprotected reference state.

The second contribution arises from protein structure. Hydrogen bonding, solvent exclusion, packing, and conformational restrictions can reduce the frequency with which an amide reaches an exchange-competent configuration. The classical Linderstrøm-Lang model represents this behaviour by allowing the amide to interconvert between closed and open states:

$$H_{\mathrm{cl}} \underset{k_{\mathrm{cl}}}{\overset{k_{\mathrm{op}}}{\rightleftharpoons}} H_{\mathrm{op}} \xrightarrow{k_{\mathrm{int}}} D, \tag{1.1}$$

where k_{op} and k_{cl} are the rate constants for conformational opening and closing, respectively, and k_{int} is the intrinsic chemical exchange rate of the exchange-competent open state [9, 18–20]. The opening event does not necessarily correspond to complete protein unfolding; depending on the residue and structural context, exchange may instead be enabled by disruption of individual hydrogen bonds, local cooperative unfolding, or larger-scale conformational transitions [18, 19, 21].

Two limiting kinetic regimes are commonly distinguished. In the EX1 regime, chemical exchange is fast relative to structural closing once the open state is reached, and the observed exchange is governed primarily by the opening process. In the EX2 regime, which is frequently encountered for protected amides under native conditions, structural opening and closing occur more rapidly than chemical exchange. The observed rate can then be approximated as

$$k_{\mathrm{obs}} \simeq \frac{k_{\mathrm{int}}}{P}, \tag{1.2}$$

where

$$P = \frac{k_{\mathrm{cl}}}{k_{\mathrm{op}}} \tag{1.3}$$

is the protection factor [4, 6, 10]. A protection factor close to unity corresponds approximately to an unprotected amide, whereas larger values indicate progressively slower observed exchange relative to the intrinsic chemical rate.

This relation explains why an accurate value of k_{int} is central to quantitative HDX interpretation, and a change in k_{obs} does not necessarily imply a change in protein structure if the chemical exchange conditions have also changed. Small differences in pH or temperature can alter intrinsic rates and shift deuterium-uptake curves even when the conformational ensemble remains unchanged [6]. Structural interpretation therefore requires the effects of the solvent and primary sequence to be distinguished from the structural protection encoded by P.

Within the EX2 approximation, the protection factor can also be related to an opening free energy,

$$\Delta G_{\mathrm{op}} = RT \ln P, \tag{1.4}$$

providing a connection between measured exchange kinetics and thermodynamic descriptions of structural opening [10, 22]. This connection is one reason why quantitative exchange rates are more informative than deuterium uptake differences alone. Differential uptake measurements can identify where two states differ, whereas properly calibrated kinetic measurements can, under suitable assumptions, be related to physically interpretable parameters.

The distinction between k_{obs} and k_{int} is therefore the conceptual basis of HDX as a structural probe. The measured rate reflects both chemistry and structure. Quantitative analysis depends on describing the chemical contribution sufficiently well that the remaining difference can be interpreted in terms of conformational protection.

1.3 Why combine HDX with mass spectrometry?

Hydrogen exchange predates its coupling to mass spectrometry and can be detected using several analytical approaches. NMR has a particular advantage in its ability to monitor individual amides, but HDX-MS provides a different combination of molecular-size range, sensitivity, sample consumption, throughput, and compatibility with complex biochemical systems [4, 5]. This combination has allowed HDX-MS to become useful for both fundamental protein biophysics and applied characterization.

The principle of MS detection is straightforward, replacing an amide hydrogen with deuterium increases the mass of the molecule, and the resulting shift in

the isotopic envelope can be measured as a function of labeling time [23]. In a conventional experiment, the protein is first equilibrated under the biochemical condition of interest and then diluted into a deuterated labeling buffer. Typical labeling solutions contain a high, but not complete, fraction of D_2O. After a defined exchange time, the reaction is quenched by decreasing both pH and temperature. Conditions near the minimum of the amide exchange-rate curve, commonly around pH 2.5 and temperatures close to 0 °C, are used to reduce further exchange during sample processing [4, 5].

For local HDX-MS measurements, the quenched protein is generally digested using an acid-compatible protease such as pepsin. The resulting peptides are trapped and separated by rapid reversed-phase liquid chromatography before electrospray ionization and mass analysis. Therefore, the experimental observable represents the deuterium retained by each detected peptide rather than the exchange of an individual residue, and overlapping peptides can improve spatial localization, but the fundamental resolution of a conventional bottom-up experiment remains peptide based [5, 24].

This loss of spatial resolution relative to residue-level NMR measurements is compensated by experimental flexibility. HDX-MS can be applied to relatively small peptides as well as very large proteins and protein complexes, requires comparatively small quantities of sample, and can accommodate a variety of solution environments [6]. These features have enabled studies of ligand binding, protein-protein interactions, membrane-associated systems, conformational changes, protein formulation, and biotherapeutic molecules [4, 6]. Automated sample handling and increasingly mature data-analysis platforms have further improved reproducibility and throughput.

Also, the accessible kinetic range has also expanded, automated millisecond systems can produce reproducible labeling times extending from tens of milliseconds into the conventional seconds-to-minutes regime [7]. Fast labeling is particularly useful for weakly protected amides and unstructured peptides, for which much of the exchange may be complete before the earliest time point of a conventional experiment, and these measurements provide valuable information for intrinsic-rate calibration because an unstructured peptide should, in principle, approach the maximum chemically allowed exchange rate under a specified set of conditions [6].

Mass spectrometry nevertheless introduces analytical stages that are absent from an idealized solution-phase exchange experiment. The sample is not measured immediately when the labeling time ends. It must first pass through a sequence of environments in which isotope exchange can continue, and this distinction becomes central when HDX-MS is used not only to compare two conditions but to infer absolute exchange rates or protection factors.

1.4 Problem in HDX-MS

The experimentally detected deuterium content is often described as the uptake associated with a particular labeling time; however, the measured uptake

corresponds to the isotopic state that survives until mass analysis, because the deuterium content present at the end of labeling can continue to change throughout the subsequent workflow. Quenching substantially reduces the exchange rate but does not eliminate exchange, after which digestion, trapping, and chromatographic separation introduce additional processing time under increasingly protiated solvent conditions, while electrospray ionization transfers the analyte from bulk solution into charged droplets and ultimately into the gas phase before ion transmission and detection [5, 8, 9, 25].

Reverse exchange and deuterium loss are therefore not confined to a single experimental step. The labeling solution itself is generally not 100% D_2O, such that forward H→D and reverse D→H exchange can occur simultaneously in the isotopically mixed solvent, while subsequent quenching further changes the solvent isotopic composition together with the pH and temperature [10, 26, 27]. During liquid chromatography, the solvent environment changes continuously as the mobile-phase gradient progresses, typically with an increasing fraction of organic solvent such as acetonitrile, which can further alter the chemical exchange conditions; each stage therefore has its own duration and chemical environment, while the intrinsic susceptibility of individual amide sites to exchange remains strongly dependent on the local amino-acid sequence [20, 21].

For this reason, a measured uptake value $D_{\mathrm{meas}}(t)$ cannot in general be identified directly with the deuterium content at the end of the labeling period, $D_{\mathrm{label}}(t)$. Conceptually,

$$D_{\mathrm{label}}(t) \longrightarrow D_{\mathrm{quench}}(t) \longrightarrow D_{\mathrm{LC}}(t) \longrightarrow D_{\mathrm{ESI/MS}}(t), \tag{1.5}$$

and the quantity at the right-hand side is the one accessible experimentally. The differences between consecutive terms depend on the kinetics operating within the corresponding stage.

The problem is particularly relevant when the objective is to infer absolute biophysical quantities, and differential HDX-MS is often designed so that two conditions are subjected to the same downstream workflow. If systematic deuterium loss is similar in both measurements, part of the loss cancels when their uptake difference is calculated. So, this is one reason why differential HDX-MS has proved highly effective for identifying regions perturbed by ligand binding or other structural changes. Also, absolute interpretation places a stronger requirement on the experiment because the magnitude of the retained uptake must itself be physically accurate [5, 6].

Quantitative treatment also requires an accurate definition of intrinsic exchange, early parameterizations established sequence-dependent, pH-dependent, and temperature-dependent reference rates for unstructured peptides [16, 17], however, more recent millisecond measurements have re-examined the reference state used in these calculations. Stofella and co-workers reported that predictions based on the 3-Ala reference were more consistent with the exchange of several unstructured peptides than predictions using poly-D,L-alanine, while also showing that salt introduces effects that are not yet represented by a universal

correction [6]. Thus, uncertainty in the detected uptake and uncertainty in the intrinsic rate both propagate into estimates of structural protection.

The treatment of mixed H_2O/D_2O solvents introduces another layer of the same problem. The irreversible form of the Linderstrøm-Lang model is naturally formulated for exchange into pure D_2O, but experimental buffers instead contain both hydrogen and deuterium, making the exchange intrinsically reversible. A generalized description therefore requires separate forward and reverse rates, k_{forw} and k_{back}, and an equilibrium deuterium population that need not be identical to the nominal D_2O fraction [10, 11]. Treating incomplete solvent deuteration only as a scaling factor neglects this reversible kinetic behaviour.

Processes associated with electrospray and ion transmission require additional caution because electrospray ionization converts solution-phase analytes into gas-phase ions through a sequence involving charged droplets, solvent evaporation, droplet fission, and eventual ion release [25]. Gas-phase hydrogen/deuterium exchange has been demonstrated experimentally on millisecond timescales in travelling-wave ion guides [28], showing that isotope chemistry need not cease once ions have left solution. Instrument-dependent exchange with residual water has also been associated with differences among peptide charge states in HDX-MS measurements [5]. These observations do not imply that every unexplained deuterium deficit has a single gas-phase origin, but they establish that the region between electrospray and detection cannot automatically be treated as chemically inert.

A quantitative model must therefore answer a more specific question than a conventional back-exchange correction: given the isotope state entering each experimental stage, together with its duration and chemical conditions, what isotope state should leave that stage? Framing the problem in this way turns back-exchange from an empirical percentage into a time-dependent kinetic process.

1.5 Conventional Back-Exchange Corrections

The common strategy for compensating for deuterium loss is to analyze a maximally deuterated control, so if the centroid of the undeuterated peptide is denoted by D_0, the measured centroid at labeling time t by $D(t)$, and the experimentally recovered centroid of the maximally labeled sample by $D_{\max}$, the fractional uptake can be expressed as

$$D_{\mathrm{frac}}(t) = \frac{D(t) - D_0}{D_{\max} - D_0}. \tag{1.6}$$

Normalization in this form is useful because the maximally labeled sample passes through the same downstream workflow and therefore reports the overall amount of label that can be recovered experimentally [5, 6], so it is widely used in measurements where comparison to the theoretical maximum uptake or calculation of protection factors is required. But the correction is exact only under assumptions about how deuterium is lost, and at the peptide level,

a single recovery factor treats the measured loss as though all exchangeable amides contributing to the peptide can be corrected by the same normalization. Residue-level back-exchange, however, is intrinsically heterogeneous because chemical exchange rates depend on primary sequence, and the first and second amides of a proteolytic peptide can behave differently from internal residues, while local sequence contexts can cause other amides to exchange unusually rapidly [4, 5, 17], so the measured recovery of a peptide is therefore an aggregate of several site-specific processes.

Chromatographic behaviour introduces additional deviations from a simple random-coil model. Sheff and co-workers directly investigated back-exchange under quenched conditions and showed that peptide-column interactions could either accelerate or decelerate exchange depending on peptide properties [24], as a result, peptides with a tendency to form secondary structure could retain structure under quench conditions, and interaction with a reversed-phase column could further stabilize hydrogen-bonded conformations, so the resulting back-exchange was slower than expected from random-coil calculations, and other peptides displayed accelerated on-column exchange. These results show that the chromatographic stage does not necessarily act as a uniform period of deuterium loss.

This observation has consequences beyond chromatography itself, because a correction derived from a maximally labeled peptide accurately describes the recovered peptide-level endpoint only if the control and experimental sample experience equivalent site-specific behaviour. If structural state, peptide conformation, solvent exposure, retention behaviour, or another property alters that behaviour, the same global normalization may not reconstruct the labeling-state distribution. The issue becomes more severe when peptide-level measurements are used to infer residue-specific protection factors, since the experimentally inaccessible distribution of deuterium within the peptide matters.

Preparation of the maximally deuterated control also introduces practical choices due to this reason that long labeling times may be unsuitable for unstable proteins, while denaturing protocols or offline peptide deuteration generate reference samples under conditions that differ from the native labeling experiment [5]. Such controls remain valuable and are often sufficient for differential applications, but they should not be interpreted as a complete mechanistic description of the exchange process.

The limitation is therefore not that maximally labeled controls are intrinsically inappropriate, and they provide an experimentally accessible measure of deuterium recovery and remain useful for normalization. The limitation arises when a peptide-level recovery factor is assumed to contain all information required to reconstruct site-specific isotope kinetics. For quantitative modeling, the more direct approach is to calculate the expected exchange during the individual stages themselves.

1.6 Motivation for a Mechanistic Model

The central approach developed in this thesis is to model the evolution of deuterium content throughout the experimental workflow rather than correcting the final measurement using a single empirical factor, and this stage-resolved treatment is motivated by the fact that deuterium exchange continues during successive post-labeling steps under distinct chemical and temporal conditions [5, 9]. Accordingly, each experimental stage is represented as a kinetic interval characterized by its own solvent composition, temperature, pH, and duration, with the isotope state produced by one interval serving as the initial condition for the next; where isotopically mixed solvents are present, both forward and reverse exchange must be considered [10]. Labeling, quenching, chromatography, and subsequent stages are therefore treated not as independent corrections, but as consecutive components of a single kinetic trajectory.

This approach requires the classical exchange description to be extended when H_2O and D_2O are both present so in an isotopically mixed solvent, the exchange-competent amide can move in both directions,

$$H_{\mathrm{op}} \underset{k_{\mathrm{back}}}{\overset{k_{\mathrm{forw}}}{\rightleftharpoons}} D_{\mathrm{op}}, \tag{1.7}$$

rather than undergoing only irreversible H→D conversion. The generalized Linderstrøm-Lang framework incorporates this reversibility and reduces to the classical description in the limit of pure D_2O [10]; therefore, forward and reverse intrinsic exchange rates can in turn be estimated as functions of sequence, temperature, acidity, and isotopic solvent composition [11].

The same concept can be applied sequentially; if the state after one stage is denoted by $\mathbf{x}_n$, then the following stage transforms this state according to its corresponding kinetic evolution,

$$\mathbf{x}_n = \Phi_n(\tau_n)\mathbf{x}_{n-1}, \tag{1.8}$$

where τ_n is the duration of stage n and Φ_n represents the exchange dynamics under that stage's conditions. Repeated application gives a stage-resolved description of the experiment, and the full mathematical formulation is developed later in this thesis; here, its relevance is that deuterium loss arises naturally from the same exchange kinetics that produce labeling rather than from a separate empirical correction.

This formulation also makes discrepancies more informative, so if labeling, quenching, and liquid chromatography can be modeled from their experimental durations and solvent conditions, the predicted isotope state entering the electrospray source can be compared with the measured state. A residual difference then identifies the magnitude of deuterium loss not accounted for by the modeled solution-phase stages. It does not by itself determine the microscopic mechanism, but it narrows the problem to processes occurring later in the experiment.

Electrospray is a natural stage to investigate in this context because it forms the boundary between solution-phase chemistry and gas-phase detection.

ESI proceeds through charged droplets that undergo solvent evaporation and fission before analyte ions are released into the gas phase [25]. The composition and temperature of these droplets evolve during desolvation, while subsequent ions can undergo gas-phase reactions, so fast gas-phase HDX experiments demonstrate that peptide and protein ions remain capable of isotope exchange after electrospray under suitable conditions [28]. A mechanistic treatment of the earlier stages therefore provides a necessary baseline for determining whether an additional effective ESI or gas-phase contribution is required by the data.

The motivation for the model is therefore broader than improving a numerical correction, because a stage-resolved formulation makes the assumptions underlying the analysis explicit. Sequence-dependent intrinsic exchange rates determine the susceptibility of individual amides to exchange, while the solvent composition, pH, and temperature determine the corresponding forward and reverse exchange kinetics [11, 16, 26]. The duration of each experimental stage then determines how far the system progresses toward its stage-specific isotopic equilibrium. This formulation also makes the model experimentally testable, because its predictions can be challenged systematically by varying parameters such as the labeling D_2O fraction, quench duration, chromatographic conditions, temperature, or other controlled experimental variables [8, 27].

Such a framework is particularly relevant to millisecond HDX-MS, where accurate intrinsic kinetics and accurate downstream treatment are required simultaneously. The rapid labeling of unstructured peptides provides a sensitive test of intrinsic-rate calculations, while the large subsequent loss of a rapidly incorporated label provides a sensitive probe of back-exchange. Together, these features create an experimental regime in which the distinction between labeling kinetics and analytical recovery can be examined directly.

Chapter 2

Fundamentals of Hydrogen/Deuterium Exchange

Hydrogen/deuterium exchange (HDX) reports on protein structure and dynamics through a chemical reaction whose rate is modulated by both the local molecular environment and conformational fluctuations, therefore, the measured exchange rate of a backbone amide contains two conceptually distinct contributions which the first one is the chemical reactivity of that amide under the relevant solvent conditions, and the second one is the extent to which protein structure permits the exchanging group to reach an exchange-competent state. Separating these contributions is central to quantitative interpretation of HDX measurements and provides the theoretical basis for the kinetic models developed later in this thesis [4, 5].

This chapter introduces the chemical and kinetic framework used throughout the remainder of the thesis, and then the discussion begins with proton-transfer chemistry and the definition of the intrinsic exchange rate, then examines sequence and temperature effects before introducing the classical Linderstrøm-Lang model, the EX1 and EX2 limits, and protection factors. The final section discusses the choice of an intrinsic-rate reference, with particular emphasis on the distinction between poly-D,L-alanine (PDLA) and tri-alanine (3-Ala), because that choice directly affects the rate calculations used in the computational analysis.

2.1 Hydrogen exchange chemistry

Although the process is conventionally called hydrogen exchange, the underlying chemistry is more precisely described as proton transfer, because exchange does not involve direct replacement of a neutral hydrogen atom in a single elementary step; instead, unstable hydrogens participate in acid-base reactions in which a proton is transferred between the solute and solvent through transient donor-acceptor complexes. This distinction is important because the exchange rate is consequently governed by acid-base chemistry, solvent composition, isotope effects, and the availability of catalytic species [4, 11].

A generic proton-transfer step may be represented schematically as

$$\mathrm{XH} + \mathrm{Y} \rightleftharpoons [\mathrm{XH} \cdots \mathrm{Y}] \rightleftharpoons [\mathrm{X} \cdots \mathrm{HY}] \rightarrow \mathrm{X} + \mathrm{HY}, \tag{2.1}$$

where XH is the proton donor and Y is the proton acceptor. The detailed charge states depend on the specific reaction, but the essential feature is the transient formation of a donor-acceptor complex followed by redistribution of the proton and dissociation of the products. Hamuro describes HDX chemistry within

this proton-transfer framework and shows that the observed behavior of many exchangeable groups can be rationalized in terms of their acid-base properties and the available catalytic pathways [4].

For peptide backbone amides, exchange can proceed through acid-catalyzed, base-catalyzed, and water-catalyzed pathways, so a convenient phenomenological expression for the pseudo-first-order intrinsic exchange rate is

$$k_{\text{int}} = k_A[\text{H}^+] + k_B[\text{OH}^-] + k_W, \tag{2.2}$$

where k_A and k_B are second-order rate constants for the acid-catalyzed and base-catalyzed pathways, respectively, and k_W is the contribution from water catalysis. Also, for H to D exchange in deuterated solvent, the corresponding expression must be written with the appropriate deuterated solvent species and isotope-specific rate constants; thus, in pure or nearly pure D_2O, for example, the catalytic terms involve deuteron and deuteroxide activities rather than simply reusing the H_2O values without correction [4, 16, 17].

The relative importance of the three terms in Eq. 2.2 changes strongly with acidity. At low pH, acid catalysis becomes important, and at higher pH, base catalysis increases rapidly with hydroxide or deuteroxide concentration, but water catalysis contributes a comparatively pH-independent component and becomes most visible near the minimum of the characteristic V-shaped exchange-rate profile. So, the resulting pH dependence is one reason that small changes in experimental pH can produce large changes in measured HDX kinetics [4, 16].

Near neutral conditions, the base-catalyzed pathway generally dominates backbone-amide exchange, so a base removes the amide proton, generating an anionic intermediate that is rapidly reprotonated or redeuterated by solvent; however, in D_2O-rich labeling buffer, this process results predominantly in incorporation of deuterium. The strong base-catalyzed contribution is also the reason that many intrinsic-rate calculations used for HDX-MS are highly sensitive to pH in the experimentally relevant range [4, 11].

The same chemical picture becomes more involved when H_2O and D_2O are both present in appreciable amounts, so the reaction is then not adequately represented as an irreversible H to D conversion because once an amide has incorporated deuterium, it can return to the protiated state. A minimal description therefore distinguishes a forward rate, k_{forw}, from a reverse or back-exchange rate, k_{back}:

$$\text{NH} \underset{k_{\text{back}}}{\overset{k_{\text{forw}}}{\rightleftharpoons}} \text{ND}. \tag{2.3}$$

The values of these rates depend on solvent isotope composition as well as acidity, temperature, and sequence. The mixed-solvent formulation developed by Grimaldi and co-workers focuses on the base-catalyzed pathway because it dominates under slightly acidic to neutral conditions, while explicitly noting that acid-catalyzed and water-catalyzed pathways become relevant at lower pH [11]. This reversible view of exchange is particularly relevant for HDX-MS because the experimental workflow does not remain in one isotopically pure solvent. Labeling is usually performed in D_2O-rich solution, whereas quench and chromatographic stages contain substantially more H_2O.

2.2 Backbone amide exchange

Proteins contain several classes of exchangeable hydrogens, including side-chain heteroatom hydrogens, terminal groups, and backbone amide hydrogens. For structural HDX measurements, the backbone amide sites are especially useful because they are distributed along the polypeptide chain and exchange on time scales that can be strongly modulated by protein structure; one of reasons is that many side-chain hydrogens exchange too rapidly to retain useful labeling information during the downstream stages of a conventional HDX-MS experiment, but backbone amides, by contrast, can remain labeled long enough to report on local structural protection, as in an unstructured peptide, an amide may be sufficiently exposed that its exchange rate approaches the chemically expected intrinsic value, but in a folded protein, the same amide can exchange much more slowly, which can be referred to as protection. But this protection should not be interpreted as a direct measurement of solvent-accessible surface area, because solvent exposure is relevant, but it is only one component of the exchange mechanism such as hydrogen bonding, local packing, electrostatic environment, and conformational fluctuations can all influence whether the amide reaches an exchange-competent configuration.

The structural contribution is conveniently expressed by distinguishing between closed and open conformational states, which in a closed state, the geometry of surrounding the amide prevents the chemical exchange step from occurring efficiently, and a transient structural fluctuation can generate an open state in which the amide becomes competent for exchange. However, the opening event may correspond to disruption of a single hydrogen bond, a local cooperative rearrangement, or a larger unfolding fluctuation, depending on the protein and residue under consideration, but the hydrogen-exchange formalism itself does not uniquely specify the microscopic structural event; rather, it connects the population and kinetics of exchange-competent states to the measured exchange rate [22].

This distinction is important when interpreting an HDX measurement, for example a slowly exchanging amide may be strongly protected because its local structure opens only rarely, because the open state closes rapidly, or because the intrinsic chemical rate under the chosen solution conditions is itself slow, however, a rapid exchange does not necessarily imply complete absence of structure. A region may populate structure transiently yet still exchange rapidly if opening is frequent or if chemical exchange is sufficiently fast once the open state is reached, so quantitative treatment requires an independent estimate of the intrinsic chemical rate before structural protection can be inferred.

For HDX-MS, an additional level of averaging is introduced because the measured quantity is normally the total deuterium content of a proteolytic peptide rather than the state of one backbone amide, so if a peptide contains N exchangeable amides, its uptake can be written as a sum of residue-level

contributions,

$$D_{\mathrm{pep}}(t) = \sum_{i=1}^{N} d_i(t), \tag{2.4}$$

where $d_i(t)$ is the deuterated fraction of residue i. It is worth mentioning that proline residues are excluded because they do not contain the usual backbone amide hydrogen, and the N-terminal position of a proteolytic peptide does not contribute a retained backbone amide label in the same way as an internal residue [5, 6]. This peptide-level coarse graining is one reason why accurate residue-level rate models are valuable even when the experimental observable is a centroid mass shift.

2.3 Intrinsic exchange rate

The intrinsic exchange rate, k_{int}, is the rate expected for an amide that is not structurally protected under the same chemical conditions as the experiment, and it provides the chemical baseline against which the observed rate is compared. In classical protein hydrogen-exchange analysis, k_{int} is therefore not a universal constant associated only with an amino-acid identity, it is a condition and sequence dependent rate calculated for an effectively unprotected peptide group [5, 16, 17].

Several variables enter this calculation, and the most prominent are acidity and temperature, but neighboring residues, termini, isotope composition, and ionic environment also influence the exchange chemistry. Schematically,

$$k_{\mathrm{int},i} = f(\mathrm{pH/pD}, T, \mathrm{sequence}_i, \mathrm{termini}, x_D, I, \ldots)\,, \tag{2.5}$$

where x_D denotes the deuterium content of the solvent and I represents ionic conditions. Equation 2.5 is not intended as a new kinetic law; it summarizes the variables that must be controlled or modeled when an intrinsic rate is assigned to residue i.

The pH or pD dependence follows directly from the acid and base catalyzed pathways, because the base-catalyzed rate is proportional to hydroxide or deuteroxide activity, an increase of approximately one pH unit in the base-dominated regime can change the exchange rate by roughly an order of magnitude, before other nonidealities are considered. This strong sensitivity means that pH values should not be treated as descriptive metadata because they are kinetic parameters, and the distinction between a glass-electrode reading and an effective acidity in H_2O/D_2O mixtures also becomes relevant when comparing data acquired at different isotope compositions [4, 17].

Temperature affects both reaction kinetics and solvent acid-base properties even when the protein conformation itself is assumed unchanged, the same exposed amide will exchange at different rates at different temperatures, along with temperature corrections are applied to reference rate constants before sequence corrections are used. The formulation employed later in this thesis is discussed in Section 2.5.

Primary sequence modifies the local chemical reactivity of the peptide group. Bai and co-workers systematically measured the effects of the naturally occurring amino-acid side chains and showed that the exchange behavior of a given amide depends primarily on its nearest neighbors, therefore, these effects arise from both inductive changes in peptide-group acidity and steric interference with the formation of exchange-competent solvent complexes [16].

The intrinsic rate is also sensitive to isotopic composition because reference parameters have historically been tabulated for specific combinations of solute and solvent isotopes, and the rates for H to D and D to H exchange are not identical. In mixed H_2O/D_2O solvents, a single irreversible k_{int} is insufficient to describe both directions because the forward and reverse rate constants must be treated separately, with the solvent composition determining the probability of deuteration versus reprotonation [10, 11].

Ionic conditions introduce another layer of complexity, that the classical reference-rate framework includes salt-sensitive effects, and more recent millisecond HDX-MS measurements show that the presence of NaCl can slow exchange of nominally unstructured peptides relative to predictions made in the absence of salt [6, 17]. This observation matters for quantitative work because the intrinsic rate is supposed to represent the fastest rate permitted by chemistry for the sequence under the actual experimental environment, therefore, a prediction calibrated under one ionic condition should not automatically be assumed to remain exact under another.

For these reasons, k_{int} should be viewed as a calculated chemical reference rather than a purely empirical normalization constant, and its accuracy determines the accuracy of any protection factor inferred from the observed exchange rate. Later chapters therefore use explicit intrinsic-rate calculations rather than treating all residues as chemically equivalent.

2.4 Primary sequence effects

The chemical reactivity of a peptide-group amide is modified by the residues immediately adjacent to it, because these primary-structure effects are present even in the absence of stable secondary or tertiary structure and must therefore be separated from structural protection. Bai et al. established a practical framework in which the effects of neighboring side chains are represented as corrections to an alanine-based reference rate [16]. Also, updated reference parameters retain this nearest-neighbor logic while refining several residue-specific and terminal corrections [17].

2.4.1 Nearest-neighbor effects

For a given backbone amide, the side chains on the two sides of the peptide group influence exchange differently, so Bai et al. tabulated left and right correction factors rather than assigning one universal factor to an amino acid. The notation is defined from the perspective of the side chain: a factor labeled L describes its

influence on the peptide group to its left, whereas R describes its influence on the peptide group to its right [16].

A convenient representation of the sequence-corrected rate is

$$k_{\mathrm{int},i} = k_{A,\mathrm{ref}}\, A_{L,i} A_{R,i}\, [\mathrm{D}^+] + k_{B,\mathrm{ref}}\, B_{L,i} B_{R,i}\, [\mathrm{OD}^-] + k_{W,\mathrm{ref}}\, W_{L,i} W_{R,i}, \tag{2.6}$$

for H to D exchange in deuterated solvent, and the symbols $A_{L,i}$ and $A_{R,i}$ denote the acid-pathway corrections associated with the neighboring residues, and $B_{L,i}$ and $B_{R,i}$ denote the corresponding base-pathway corrections. The original parameterization treats the water contribution using an empirically related sequence dependence rather than requiring a completely independent set of arbitrary residue factors. Equation 2.6 is written to show the structure of the correction because in computational implementation, isotope-specific reference values and the appropriate effective catalyst concentrations are used.

The useful feature of the Bai framework is approximate additivity in logarithmic space, or equivalently multiplicativity of rate factors. The effect of the residue on one side of the amide can therefore be combined with the effect of the residue on the other side, and this was tested using oligopeptide and polypeptide model systems and found to reproduce much of the primary-sequence dependence of exchange [16]. The approximation is not a statement that all local chemistry is independent, so instead, it provides an experimentally calibrated way to predict the exchange rate of an unstructured sequence without fitting a separate rate constant for every possible tripeptide context.

Updated parameters have modified some of the original corrections, as Nguyen et al. reported adjustments for deprotonated Asp and Glu and for Gly, and also discussed effects that extend beyond the simplest nearest-neighbor picture, such as detectable catalysis by protonated histidine in some contexts [17]. For the present thesis, the nearest-neighbor framework remains the starting point, while the numerical parameter set used by the computational implementation incorporates the selected updated corrections.

2.4.2 Steric effects

Steric effects reduce exchange when the local sequence hinders formation of the solvent-amide encounter geometry required for proton transfer, since Bai et al. found that several side chains lower both the acid-catalyzed and base-catalyzed exchange rates of neighboring peptide groups because both limbs of the rate-versus-pD curve can be shifted downward with comparatively little displacement of the pD at the rate minimum, this behavior was interpreted as steric blocking rather than a purely electronic effect [16].

Bulky hydrophobic residues such as Ile, Leu, Val, Phe, and Trp can contribute substantial steric corrections, although the magnitude and direction depend on whether the residue is located to the left or right of the exchanging peptide group. The effect should not be reduced to a simple rule that “larger side chain means slower exchange.” The empirical factors reflect both side-chain geometry

and the conformational preferences available to the local peptide backbone, and a sequence-specific calculation is therefore preferable to assigning a single steric penalty based only on residue size.

Steric corrections are particularly relevant when intrinsic rates are compared across peptides of different sequence. Two fully unstructured peptides at the same pH and temperature need not show identical exchange kinetics, even if their amides are equally unprotected structurally. Part of the difference can arise simply because their local sequences present different access to catalytic solvent species, and in protection-factor analysis, failure to remove this chemical contribution would incorrectly attribute primary-structure effects to conformational protection.

2.4.3 Inductive effects

Inductive effects arise because neighboring side chains alter the electronic environment of the peptide group, because polar substituents can withdraw or redistribute electron density, changing the acidity of the amide proton and therefore changing the rates of acid-catalyzed and base-catalyzed exchange. In the Bai measurements, these electronic effects often produced a horizontal displacement of the V-shaped rate-versus-pD curve as the base-catalyzed branch could become faster while the acid-catalyzed branch became slower [16].

This behavior follows the chemistry of the two pathways. If a neighboring group increases the acidity of the amide proton, removal of that proton by base becomes more favorable, which accelerates base-catalyzed exchange; however, the same electronic change can make proton addition in the acid-catalyzed pathway less favorable, slowing that contribution, so the two effects need not change in the same direction.

Titratable side chains require additional care because their electronic state depends on pH. Asp, Glu, and His do not exert one invariant correction across all solution conditions, so their protonated and deprotonated forms have different inductive properties, and updated rate calculations account for the relevant ionization state [16, 17]. This point is especially relevant when the same peptide is modeled across labeling, quench, and chromatographic conditions, because those stages may differ greatly in pH.

2.4.4 Proline

Proline is a special case in backbone-amide HDX because its backbone nitrogen is part of the pyrrolidine ring and does not carry the usual peptide N-H hydrogen; therefore, the proline position does not contribute a conventional exchangeable backbone amide to the measured deuterium uptake, and in residue-level uptake calculations, proline sites are excluded from the set of exchangeable backbone amides [5, 6].

The absence of an amide hydrogen at proline does not mean that proline can be ignored entirely because its residue still changes the chemical environment and conformational preferences of neighboring peptide groups. Bai et al. reported

distinct corrections associated with proline, including dependence on cis/trans configuration in the reference tables [16], and in practical rate calculations, the proline site itself is assigned no backbone-amide exchange contribution, while its effect on neighboring amides remains part of the primary-sequence correction.

This distinction is important for short peptides because excluding a proline only from the final count of exchangeable sites while failing to apply its neighboring-sequence effect would produce an inconsistent model. The implementation used later in this thesis follows the residue-level logic, such as proline contributes no exchangeable backbone amide at its own position but remains present in the sequence used to calculate adjacent intrinsic rates.

2.4.5 N and C terminal effects

Terminal groups can modify intrinsic exchange rates more strongly than ordinary internal nearest-neighbor substitutions. Their charge states and electronic effects differ from those of internal peptide bonds, and proteolysis creates new termini that were not present in the intact protein. Terminal corrections are therefore particularly relevant to bottom-up HDX-MS, where the measured species are peptides generated after the labeling step.

The first residue of a peptide does not contain an internal peptide-bond amide of the same type as residues further along the chain. In common peptide-level HDX calculations, the first residue is consequently excluded from the sum of exchangeable backbone amides [5, 6]. The first conventional backbone amide is associated with residue 2, and its kinetics can be strongly influenced by the positively charged N-terminus.

Nguyen et al. emphasized that polymer-end effects can be substantial. Under hydroxide-catalyzed conditions, the first exchangeable amide near the N-terminus can exchange much faster than an otherwise comparable internal amide. Their discussion notes an approximately 40-fold acceleration under elevated-pH hydroxide catalysis and, under acidic HDX-MS analysis conditions around pH 2.5, an intrinsic back-exchange rate for residue 2 that can remain roughly an order of magnitude faster than a comparable internal site [17]. This behavior is one reason the second residue of a proteolytic peptide is sometimes assumed to lose much of its label during analysis, although complete loss should not be treated as universal because sequence-dependent slowing can remain important.

The C-terminal region also has a strong end effect. Hydroxide-catalyzed exchange of a C-terminal peptide amide can be markedly slower than the corresponding internal rate, with the historical calibration discussed by Nguyen et al. indicating an effect on the order of tens-fold [17]. The exact correction depends on terminal ionization and solution conditions.

These terminal effects have two consequences for quantitative HDX-MS. First, intrinsic-rate calculations for short peptides must distinguish internal from terminal positions. Second, back-exchange after proteolysis is not necessarily identical for the same residue in different host peptides. Sheff et al. showed that peptide sequence, residual structure, and peptide-column interactions can further alter back-exchange behavior during LC analysis [24]. The terminal-rate

corrections discussed here therefore provide an essential chemical baseline, but they do not imply that every downstream exchange process is fully described by primary sequence alone.

2.5 Temperature dependence

Exchange reactions are strongly temperature dependent, so a change in temperature alters the rate constants of the underlying proton-transfer pathways and can also change the conformational equilibrium of a protein. The present section concerns the temperature dependence of the intrinsic chemical rate, as structural temperature effects are conceptually separate and enter through the opening and closing kinetics of the Linderstrøm-Lang model. Therefore, reference exchange rates can be rescaled from a reference temperature T_{ref} using an Arrhenius-type expression,

$$k(T) = k(T_{\mathrm{ref}}) \exp\left[-\frac{E_a}{R}\left(\frac{1}{T} - \frac{1}{T_{\mathrm{ref}}}\right)\right], \tag{2.7}$$

where E_a is the activation energy associated with the relevant exchange pathway and R is the gas constant, so this form makes the reference rate recover exactly at $T = T_{\mathrm{ref}}$ and increases the rate as temperature is raised for positive E_a.

The computational implementation used in this thesis applies separate activation energies to the acid-catalyzed, base-catalyzed, and water-catalyzed reference contributions, and in the current rate calculation code, these are

$$E_{a,A} = 14 \text{ kcal mol}^{-1}, \qquad E_{a,B} = 17 \text{ kcal mol}^{-1}, \qquad E_{a,W} = 19 \text{ kcal mol}^{-1}, \tag{2.8}$$

with a reference temperature of approximately 293 K. These values are applied independently before the pathway contributions are combined. The implementation is consistent with the established practice of correcting reference exchange rates for temperature rather than using one fixed k_{int} value across experiments [16].

Temperature control is especially important in HDX-MS because the experiment deliberately exploits temperature to suppress exchange after labeling, and it may be performed near room or physiological temperature, while quench, digestion, trapping, and chromatographic separation are commonly carried out at much lower temperatures. Lowering temperature reduces the intrinsic exchange rate and therefore slows unwanted D to H back exchange during analysis, and acidification during quench acts in parallel by moving the system toward the minimum of the amide-exchange pH profile.

For millisecond HDX-MS, accurate temperature treatment becomes even more important because the earliest time points can lie in the same range as instrumental mixing and transfer times, so small errors in the assumed intrinsic rate can then noticeably shift the predicted uptake curve. Automated millisecond systems are designed to control mixing and quenching times reproducibly over this regime [7], therefore, in the modeling developed later in this thesis, temperature

is treated as an explicit stage-dependent parameter rather than as a single descriptor of the overall experiment.

2.6 Classical Linderstrom-Lang model

The classical Linderstrøm-Lang model provides the central kinetic link between protein conformational dynamics and amide exchange. It separates structural opening from chemical exchange by representing a backbone amide as a two-state conformational system in which only the open state is competent to exchange [4, 8]. For H to D exchange in pure D_2O, the scheme is

$$\mathrm{H_{cl}} \underset{k_{\mathrm{cl}}}{\overset{k_{\mathrm{op}}}{\rightleftharpoons}} \mathrm{H_{op}} \xrightarrow{k_{\mathrm{int}}} \mathrm{D}, \tag{2.9}$$

where H_{cl} is the closed protonated state, H_{op} is the open protonated state, k_{op} is the structural opening rate, k_{cl} is the closing rate, and k_{int} is the intrinsic chemical exchange rate of the open amide.

The word "closed" should be understood kinetically rather than as a unique structural geometry, because it denotes conformations in which the exchange reaction is suppressed, whereas the "open" state represents an ensemble of configurations in which the amide can undergo the proton-transfer chemistry described earlier; therefore, the opening and closing rate constants summarize local or global conformational fluctuations relevant to that particular amide.

For an initially unexchanged population, the coupled rate equations can be written as

$$\frac{dH_{\mathrm{cl}}}{dt} = -k_{\mathrm{op}}H_{\mathrm{cl}} + k_{\mathrm{cl}}H_{\mathrm{op}}, \tag{2.10}$$

$$\frac{dH_{\mathrm{op}}}{dt} = k_{\mathrm{op}}H_{\mathrm{cl}} - (k_{\mathrm{cl}} + k_{\mathrm{int}})\,H_{\mathrm{op}}, \tag{2.11}$$

$$\frac{dD}{dt} = k_{\mathrm{int}}H_{\mathrm{op}}. \tag{2.12}$$

The exact analytical solution contains two nonzero kinetic modes and is therefore, in general, a sum of two exponentials rather than one [5]. For quantitative derivations it is useful to use the single-exponential form, which is an approximation that follows from separation of time scales, not an identity of the full model.

The transient $\{H_{cl}, H_{op}\}$ subsystem has eigenvalues

$$\lambda_{\pm} = -\frac{1}{2}\left[S \pm \sqrt{S^2 - 4k_{\mathrm{op}}k_{\mathrm{int}}}\right], \qquad S = k_{\mathrm{op}} + k_{\mathrm{cl}} + k_{\mathrm{int}}. \tag{2.13}$$

The slow relaxation rate is therefore

$$k_{\mathrm{slow}} = \frac{1}{2}\left[S - \sqrt{S^2 - 4k_{\mathrm{op}}k_{\mathrm{int}}}\right]. \tag{2.14}$$

When the open state is only weakly populated and relaxes rapidly relative to the overall loss of unexchanged population, the kinetics can be reduced to an apparent single-exponential exchange rate, and a steady-state treatment gives

$$k_{\mathrm{obs}} \simeq \frac{k_{\mathrm{op}} k_{\mathrm{int}}}{k_{\mathrm{op}} + k_{\mathrm{cl}} + k_{\mathrm{int}}}. \tag{2.15}$$

This form is consistent with the classical thermodynamic treatment of exchange under stable native conditions [22]. If the native approximation $k_{\mathrm{op}} \ll k_{\mathrm{cl}}$ is also applied, Eq. 2.15 becomes

$$k_{\mathrm{obs}} \simeq \frac{k_{\mathrm{op}} k_{\mathrm{int}}}{k_{\mathrm{cl}} + k_{\mathrm{int}}}. \tag{2.16}$$

The two principal limiting regimes, EX1 and EX2, follow from the relative magnitudes of k_{cl} and k_{int}.

Under a single-exponential approximation, the deuterated fraction of one initially protiated amide is

$$D(t) = 1 - \exp(-k_{\mathrm{obs}} t), \tag{2.17}$$

for irreversible exchange in pure D_2O. Equation 2.17 is the basic residue-level function from which peptide-level uptake models are assembled, and its simplicity is useful, but its assumptions must be kept explicit: pure or effectively irreversible labeling, a single dominant kinetic mode, and a defined observed rate.

The classical Linderstrøm–Lang model does not explicitly describe reversible exchange in isotopically mixed H_2O/D_2O solvents, and the generalized formulation used later in this thesis therefore introduces protiated and deuterated open and closed states together with separate forward and back-exchange rate constants [10, 11]. The classical model nevertheless remains the necessary starting point, because the concepts of conformational opening and closing, k_{op} and k_{cl}, intrinsic chemical exchange, k_{int}, the EX1 and EX2 kinetic limits, and the protection factor are defined within this framework [18–20].

2.7 EX1 and EX2 exchange regimes

The Linderstrøm-Lang scheme contains two sequential processes: HDX kinetics are governed by the interplay between structural opening and chemical exchange, with the relative magnitudes of k_{cl} and k_{int} determining whether exchange occurs in the classical EX1 or EX2 limiting regime[4, 5].

EX1 regime

In the EX1 limit,

$$k_{\mathrm{int}} \gg k_{\mathrm{cl}}. \tag{2.18}$$

Once the amide opens, chemical exchange is much faster than reclosing, so an opening event therefore has a high probability of being followed by exchange before the protected structure reforms. Substituting this limit into Eq. 2.16 gives

$$k_{\mathrm{obs}} \approx k_{\mathrm{op}}. \tag{2.19}$$

The measured exchange rate is then governed primarily by the structural opening rate, which, in this regime, increasing k_{int} further has little effect because chemical exchange is already effectively instantaneous on the time scale of closing.

EX1 behavior is less commonly associated with stable native proteins than EX2 behavior, although it can appear under conditions that promote slow closing or rapid chemical exchange. At the peptide level, EX1 exchange is often associated with broadened or bimodal isotopic envelopes when different molecules exchange through cooperative opening events. That spectral signature is useful but should not be treated as a strict definition of EX1, because the residue-level Linderstrøm-Lang kinetics themselves do not guarantee that a peptide must exhibit a resolved bimodal envelope.

EX2 regime

In the EX2 limit,

$$k_{\mathrm{cl}} \gg k_{\mathrm{int}}. \tag{2.20}$$

The open state usually recloses before chemical exchange occurs; therefore, any opening and closing events may therefore take place before one successful exchange event. Under the native condition $k_{\mathrm{cl}} \gg k_{\mathrm{op}}$, Eq. 2.16 reduces to

$$k_{\mathrm{obs}} \approx \frac{k_{\mathrm{op}}}{k_{\mathrm{cl}}} k_{\mathrm{int}}. \tag{2.21}$$

Defining the protection factor as

$$P \equiv \frac{k_{\mathrm{cl}}}{k_{\mathrm{op}}}, \tag{2.22}$$

the EX2 exchange rate becomes

$$k_{\mathrm{obs}} \approx \frac{k_{\mathrm{int}}}{P}. \tag{2.23}$$

EX2 is the regime most often invoked for native proteins because local structural fluctuations are generally much faster than successful chemical exchange from the transiently open state [10]. In this limit, the observed rate retains a direct dependence on both intrinsic chemistry and structural protection, so this is precisely why an accurate k_{int} calculation is necessary, as an error in the intrinsic rate propagates directly into the inferred protection factor.

The two limits also clarify the interpretation of environmental perturbations, which, in EX2, changing pH or temperature can change k_{obs} even if the conformational equilibrium is unchanged, because k_{int} changes. In EX1, once chemical exchange is much faster than closing, the same chemical perturbation may have a much smaller effect on k_{obs}, which is controlled mainly by k_{op}. Therefore, distinguishing chemical-rate effects from structural-rate effects is therefore a central task in quantitative HDX analysis.

2.8 Protection factors

A protection factor quantifies how much slower an amide exchanges in a structured environment than it would under an appropriate unprotected reference condition. Under the EX2 approximation,

$$P = \frac{k_{\mathrm{int}}}{k_{\mathrm{obs}}}, \tag{2.24}$$

and, using the kinetic interpretation of the Linderstrøm-Lang model,

$$P \approx \frac{k_{\mathrm{cl}}}{k_{\mathrm{op}}}. \tag{2.25}$$

Equations 2.24 and 2.25 connect an experimentally observable exchange rate to the equilibrium between exchange-incompetent and exchange-competent structural states.

It is useful to define the opening equilibrium constant,

$$K_{\mathrm{op}} = \frac{k_{\mathrm{op}}}{k_{\mathrm{cl}}}. \tag{2.26}$$

When $P \gg 1$, as is typical for a protected amide under native conditions,

$$P \approx \frac{1}{K_{\mathrm{op}}}. \tag{2.27}$$

If opening is treated as an equilibrium process with free-energy difference ΔG_{op} between open and closed states,

$$K_{\mathrm{op}} = \exp\left(-\frac{\Delta G_{\mathrm{op}}}{RT}\right), \tag{2.28}$$

and therefore

$$\Delta G_{\mathrm{op}} = RT \ln P. \tag{2.29}$$

This thermodynamic interpretation is one of the most important consequences of hydrogen-exchange theory, because under appropriate kinetic assumptions, exchange measurements can provide information about local or global conformational stability under conditions where the protein remains predominantly folded [22].

The interpretation nevertheless depends on the meaning of the opening event. A protection factor is not automatically the free energy of breaking one specific hydrogen bond, also, different residues can exchange through different fluctuations, and the relevant open state may involve local breathing, cooperative local unfolding, or a larger structural transition. The thermodynamic quantity inferred from Eq. 2.29 is therefore the free-energy cost of the exchange-competent opening process represented by the model, provided the EX2 and equilibrium assumptions are satisfied.

This qualification becomes especially important for HDX-MS because the experimental observable is normally peptide averaged. A measured peptide uptake curve contains contributions from several amides with potentially different $k_{\mathrm{int},i}$ and P_i, even if every residue follows EX2 kinetics,

$$D_{\mathrm{pep}}(t) = \sum_i \left[1 - \exp\left(-\frac{k_{\mathrm{int},i}}{P_i} t\right)\right], \tag{2.30}$$

so a single peptide centroid does not uniquely determine all residue-level protection factors, therefore, the inverse problem is underdetermined unless additional information, overlapping peptides, fragmentation data, regularization, or model constraints are introduced.

Protection factors are sensitive to the choice of intrinsic-rate reference because, according to Eq. 2.24, scaling k_{int} by a calibration factor scales the inferred protection factor P by the same factor when k_{obs} is held constant, and any systematic error in the chemical reference therefore becomes a systematic error in inferred stability.

In isotopically mixed solvents, the classical relation must be extended because forward and reverse exchange occur simultaneously and the equilibrium deuterium fraction may be below unity, which is addressed by the generalized Linderstrøm-Lang framework through a mixed-solvent intrinsic exchange rate together with explicit treatment of back exchange and, more generally, isotope-dependent differences in protection [10].

2.9 PDLA versus 3-Ala intrinsic-rate references

Calculation of an intrinsic exchange rate requires a reference system that approximates an unprotected peptide group, and historically poly-D,L-alanine (PDLA) has served this role because alanine has a chemically simple side chain and the racemic polymer was intended to provide a structureless reference, with the primary-sequence corrections developed by Bai et al. using alanine-based reference molecules and subsequently extended to arbitrary peptide sequences through nearest-neighbor correction factors [16].

The reference is not entirely trivial, however. Bai et al. already observed that measured exchange rates differed among alanine dipeptide, oligopeptide, and polymeric reference systems [16]. Nguyen et al. later examined PDLA by HX-MS over a range of polymer lengths and reported non-single-exponential behavior together with a length dependence of the apparent exchange rate, therefore, comparison with an unstructured protein led to an upward adjustment of the alanine reference rate by a factor of approximately 1.35 in the updated parameter set [17]. These results indicated that the choice of reference molecule affects absolute intrinsic-rate predictions.

Recent millisecond HDX-MS measurements have revisited this issue more directly. Stofella et al. compared measured uptake curves for small unstructured peptides with theoretical curves calculated using PDLA- and 3-Ala-based intrinsic rates, and with PDLA as the reference, some experimentally measured exchange

rates appeared faster than the predicted intrinsic limit, because k_{int} is intended to represent the unprotected chemical rate, systematic observations faster than that limit indicate a problem with the reference calibration rather than negative structural protection. Predictions based on the internal amide of tri-alanine showed substantially better agreement and did not produce the same physically problematic faster than intrinsic behavior [6].

The same study used molecular dynamics simulations to examine the structural propensity of PDLA peptides, as longer PDLA chains showed a tendency to form ordered conformations, with stable helical structure becoming accessible for relatively short polymer lengths. The authors therefore argued that PDLA should not automatically be treated as a completely structure-free reference. If the reference polymer is itself partially protected, its measured exchange rate will be slower than the true unprotected chemical rate, because using that value as k_{int} would then underestimate intrinsic rates for other sequences and, through Eq. 2.24, underestimate their protection factors [6].

Tri-alanine offers a more local reference because the central amide in 3-Ala is short enough to minimize stable higher-order structure while retaining a peptide environment more representative than a simple isolated amide. In the recalibration study, 3-Ala therefore serves as the preferred fully unstructured reference for absolute rate prediction [6]. This choice is also reflected in the computational tools used in this thesis, which allow either

```
ref='PDLA'
```

or

```
ref='3Ala'
```

to be selected, with `3Ala` used as the default in the current implementation.

The difference between the two references can be expressed schematically as a calibration factor,

$$k_{\mathrm{int},i}^{(\mathrm{3Ala})} = C_{\mathrm{ref}}\, k_{\mathrm{int},i}^{(\mathrm{PDLA})}, \tag{2.31}$$

where C_{ref} represents the reference-dependent rescaling before the same sequence and solution corrections are applied. In the updated historical calibration, a factor of about 1.35 was introduced relative to the earlier PDLA-based reference values [17]; the more recent 3-Ala analysis provides a physical rationale for favoring the faster unstructured reference in absolute calculations [6]. Equation 2.31 is intended to emphasize the calibration concept rather than imply that every historical and modern implementation differs by one universal factor under all conditions.

The 3-Ala reference does not remove all uncertainty from intrinsic-rate prediction, as Stofella et al. found that salt remains a significant unresolved variable: previously published peptide data collected in the presence of NaCl were not fully reproduced by 3-Ala predictions calibrated without the corresponding salt effect [6]. The mixed-solvent rate framework likewise notes that dissolved salts, chaotropic agents, and other additives can alter local acid-base behavior and therefore exchange kinetics [11]. The appropriate conclusion is not that

3-Ala provides a universal, environment-independent intrinsic rate, but that it is a better starting reference for an unprotected peptide group when combined with explicit corrections for the actual experimental conditions.

Chapter 3

HDX-MS Experimental Workflow

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) does not measure the deuteration state of a protein or peptide at the instant when the labeling period ends, and instead, the analyze passes through a sequence of experimental environments before its isotopic composition is finally recorded by the mass spectrometer, and during this interval, temperature, solvent composition, pH, residence time, chromatographic interactions, ionization, and gas-phase conditions all change. Consequently, the deuterium content measured by MS is the result of the entire experimental history rather than of the labeling reaction alone [8].

This distinction becomes particularly important when HDX-MS is interpreted quantitatively, because in a conventional comparative experiment, samples being compared are processed through the same downstream workflow, and part of the deuterium loss may therefore cancel when differences between conditions are calculated, also, absolute uptake values, exchange-rate constants, and protection factors are less forgiving, because any deuterium lost between labeling and detection directly changes the quantity being interpreted. Back-exchange should therefore be regarded as a kinetic process distributed across several experimental stages rather than as a single correction applied after data acquisition.

The purpose of this chapter is to follow the sample from the beginning of labeling to mass-spectrometric detection and to identify the physical processes through which deuterium can be gained, lost, or redistributed, and particular attention is given to the stages that are directly relevant to the millisecond HDX-MS measurements analyzed in this thesis. Those measurements were performed on small peptides that were already in peptide form and therefore were not subjected to proteolytic digestion [6]. Digestion is nevertheless described because it is an integral component of the standard bottom-up HDX-MS workflow and because it introduces additional sources of back-exchange that must be distinguished from those present in intact-peptide measurements.

3.1 Standard HDX-MS workflow

A conventional bottom-up HDX-MS experiment can be represented by the sequence

The labeling step is designed to encode information about protein or peptide structure into the isotopic composition of backbone amide sites, and the subsequent stages have a different objective, which is preparing the labeled molecule for measurement while attempting to preserve that isotopic information. These two objectives are not completely compatible, because conditions suitable for chromatographic separation, digestion, or ionization cannot eliminate

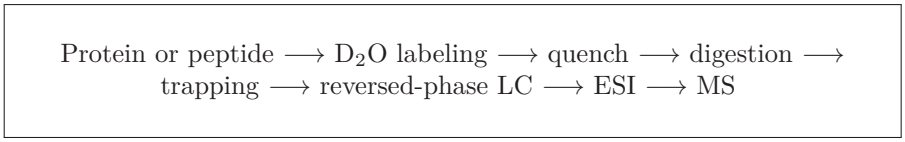


Figure 3.1: Schematic representation of a standard bottom-up HDX-MS workflow. Exchange is initiated in a deuterated labeling solution and subsequently slowed by quenching at low pH and low temperature. In bottom-up experiments, the protein is digested under quench conditions, the resulting peptides are trapped and chromatographically separated, ionized by electrospray ionization (ESI), and transferred to the mass spectrometer. For the intact small-peptide experiments considered in this thesis, the digestion step is omitted.

hydrogen exchange, and a fraction of the incorporated deuterium is therefore lost before detection [4, 8].

In bottom-up HDX-MS, the protein is usually diluted into a highly deuterated buffer and allowed to exchange for a defined labeling time, and then exchange is slowed by lowering the pH to approximately the region of minimum amide exchange and reducing the temperature to approximately 0 °C. Proteolysis is performed under these conditions, most commonly using pepsin, after which the peptides are trapped, desalted, separated by rapid reversed-phase liquid chromatography, and introduced into the mass spectrometer by ESI [4, 5]. Importantly, "quench" is an operational term. The exchange reaction becomes much slower, but it does not become zero.

The distinction between the conventional workflow and the measurements used for the small-peptide component of this thesis is important. The millisecond experiments discussed later used an already defined mixture of intact peptides, including angiotensin, bradykinin, and atrial natriuretic peptide (ANP). The peptides were labeled directly, quenched, trapped, separated by reversed-phase LC, and analyzed without peptide fragmentation [6], and therefore proteolysis does not contribute to deuterium loss in that particular experiment, whereas quenching, trapping, LC, ESI, and transfer into the mass spectrometer remain relevant.

This sequence of environments motivates a stage-resolved description of HDX-MS, and if the isotope state after one stage is denoted by $\mathbf{x}_n$, it becomes the initial state for the next stage. In a general kinetic representation,

$$\mathbf{x}_n = \Phi_n(\tau_n)\Phi_{n-1}(\tau_{n-1})\cdots\Phi_2(\tau_2)\Phi_1(\tau_1)\mathbf{x}_0, \tag{3.1}$$

where Φ_n represents the kinetic evolution associated with experimental stage n and τ_n is its duration. Equation 3.1 does not imply that every experimental stage is governed by identical chemistry. Rather, it emphasizes that the measured isotopic state is inherited sequentially: deuterium lost during quenching is unavailable during LC, and deuterium lost during LC cannot be recovered at detection unless a subsequent forward-exchange process introduces deuterium again.

A useful conceptual distinction is therefore made between *labeling uptake* and *detected uptake*. Labeling uptake refers to the isotopic state produced during the intended D_2O exposure, whereas detected uptake is the state that survives all subsequent sample-processing and ion-transfer steps. The difference between the two contains contributions from several physically different processes, and treating these contributions as one global percentage necessarily removes information about when and why the isotope was lost.

3.2 Labeling

HDX labeling begins when an initially protonated protein or peptide is mixed with a deuterated solvent, and in most HDX-MS experiments this is achieved by diluting an aqueous analyze solution into a much larger volume of labeling buffer prepared in D_2O [4]. The final deuterium fraction is therefore determined by both the isotopic composition of the labeling buffer and the dilution ratio, and if volumes are additive, an approximate final solvent deuterium fraction can be written as

$$x_{\mathrm{D}} = \frac{x_{\mathrm{D,s}} V_{\mathrm{s}} + x_{\mathrm{D,l}} V_{\mathrm{l}}}{V_{\mathrm{s}} + V_{\mathrm{l}}}, \tag{3.2}$$

where V_{s} and V_{l} denote the initial sample and labeling-buffer volumes and $x_{\mathrm{D,s}}$ and $x_{\mathrm{D,l}}$ their respective deuterium fractions. An initially protonated sample diluted into nominally pure D_2O therefore still produces a final solution containing some H_2O.

Typical HDX-MS labeling buffers contain approximately 80-95% deuterated solvent rather than exactly 100% D_2O [5]. In the millisecond peptide measurements underlying part of this thesis, the sample was diluted 20-fold with labeling buffer, producing a reported final deuteration level of 95%. This experimental detail has an important kinetic consequence that the labeling reaction does not occur in an isotopically pure solvent.

A simple irreversible representation,

$$\mathrm{NH} \longrightarrow \mathrm{ND}, \tag{3.3}$$

is appropriate only as an approximation when the solvent can effectively be treated as pure D_2O, and in an H_2O/D_2O mixture, hydrogen and deuterium are both present and the exchange process is reversible. At the level of an exchange-competent amide, the corresponding process is more appropriately written as

$$\mathrm{NH} \underset{(1-x_{\mathrm{D}})k_{\mathrm{back}}}{\overset{x_{\mathrm{D}} k_{\mathrm{forw}}}{\rightleftharpoons}} \mathrm{ND}, \tag{3.4}$$

where k_{forw} and k_{back} describe the intrinsic forward and reverse exchange processes in the relevant solvent environment [10, 11], and therefore forward labeling and back-exchange can occur simultaneously from the beginning of the

experiment, in addition the term “back-exchange” should not be restricted to events after quenching.

The rate at which the labeling reaction approaches its isotopic equilibrium depends strongly on pH and temperature, and near neutral pH, backbone amide exchange is usually dominated by the base-catalyzed pathway, and relatively small changes in acidity can produce substantial changes in the intrinsic exchange rate, and temperature affects the rate through the activation barrier of the chemical exchange process. These dependencies make accurate control of labeling conditions necessary, particularly when exchange is measured on a millisecond timescale.

Exposure time supplies the final experimental control variable. A slowly exchanging protected amide may retain hydrogen for minutes or hours, whereas a solvent-exposed amide in a short unstructured peptide may approach its labeling plateau before a conventional sample-handling system can generate the first time point, and the useful labeling window must therefore overlap the kinetic range of the system being studied, so this requirement can be one of the main motivation for millisecond HDX-MS.

Incomplete solvent deuteration also affects the maximum uptake that can be reached, so in a purely irreversible picture, sufficiently long exposure would drive every exchange-competent amide toward complete deuteration. In an isotopically mixed solvent, forward and reverse reactions instead establish a stationary isotopic distribution, so the equilibrium deuteration does not, in general, have to be exactly identical to the bulk D_2O fraction because isotope-dependent kinetic and equilibrium effects may contribute [10]. Consequently, a measured plateau below one deuterium per exchangeable amide is not necessarily produced only by post-labeling deuterium loss.

For quantitative modeling, labeling should therefore be treated as the first stage of the same reversible isotope-exchange problem that continues through the rest of the experiment, so the dominant direction is H→D because the solvent contains a large excess of deuterium, but the competing D→H pathway is already present.

3.3 Millisecond HDX-MS

Conventional HDX-MS systems are well suited to processes that evolve over seconds, minutes, or hours, and they are less informative when the exchange transition is largely complete before the first conventional labeling point. Short peptides, intrinsically disordered proteins (IDPs), flexible loops, and highly solvent-exposed regions often fall into this category[6, 7]. For such systems, a measurement beginning at tens of seconds may observe only a nearly saturated uptake curve and therefore contain little information about the underlying rate.

Millisecond HDX-MS extends the experimentally accessible time window toward rapid exchange, and since Kish et al. developed a fully automated online system capable of accessing labeling times from approximately 50 ms to several minutes with millisecond control [7]. The apparatus combines rapid solution

mixing, selectable delay volumes, online quenching, and subsequent LC-MS analysis, so this arrangement is directly relevant to the unstructured peptide measurements considered in this thesis.

In the automated system reported by Kish et al., separate fluidic channels deliver protonated, deuterated, and quench solutions. Labeling is initiated in an online mixer, the sample passes through a selected delay loop, and the reaction is subsequently quenched in a second mixer before transfer to the chromatographic system [7], so this arrangement avoids reliance on manual pipetting at sub-second times and makes the time point a property of a controlled fluidic path rather than of operator reaction time.

Temporal resolution is only useful when accompanied by reproducibility, and at short labeling times, an error of a few tens of milliseconds can represent a large fraction of the intended exposure, therefore, Kish et al. tested repeatability around the 100 ms regime and reported very small between-day mass variability for their peptide measurements [7]. Such validation is especially relevant for rapidly exchanging peptides, because the steep portion of an uptake curve occurs precisely where errors in mixing or residence time have the largest effect on inferred kinetics.

Millisecond labeling does not eliminate downstream back-exchange, but it changes the relative importance of the different time scales, and a peptide may be labeled for only 50 or 100 ms but subsequently spend seconds or minutes in quench, trapping, LC, and ion-transfer stages. For a rapidly exchanging unstructured peptide, the duration of downstream processing can therefore be orders of magnitude longer than the intended labeling event. The experiment must preserve a kinetic signal generated over milliseconds while processing the sample over much longer periods. This asymmetry is central to quantitative millisecond HDX-MS. Increasing temporal resolution at the labeling stage exposes downstream deuterium loss that may have been less conspicuous in slower experiments. A few tenths of a deuterium lost during trapping or LC can be comparable to the uptake change between adjacent early time points. Consequently, millisecond HDX-MS and stage-resolved back-exchange modeling are naturally connected.

The small-peptide measurements used in the present work illustrate this connection. The reported labeling range was 50 ms to 5 min at 20 °C and 95% solvent deuteration, after which the sample was immediately mixed with quench solution [6]. The experimental system therefore provides access to intrinsic exchange kinetics that are too rapid for a conventional tens-of-seconds workflow, while simultaneously requiring careful accounting of everything that occurs after the labeling mixer.

3.4 Quench

At the end of the desired labeling interval, HDX must be slowed sufficiently for the isotopic state to survive the remaining analytical procedure, and backbone amide exchange has a characteristic minimum in the acidic pH region, so HDX-MS uses a low-pH, low-temperature quench, commonly close to pH 2.5 and 0 °C

[4, 5]. These conditions reduce the chemical exchange rate by several orders of magnitude compared with typical near-neutral labeling conditions.

The word *quench* can be misleading if interpreted as complete termination of the reaction because amide exchange remains finite, and deuterated amides continue to encounter proton-containing solvent and can undergo D→H exchange throughout the quench period. The experimentally relevant statement is therefore that quenching *minimizes* exchange, and it does not stop it [4, 8].

Quenching changes more than pH, and it also changes the isotopic composition of the solution. Consider a stream that has reached 95% solvent deuteration during labeling and is subsequently mixed 1:1 with a protonated quench buffer. Neglecting small volume and isotope-density corrections, the resulting solvent contains approximately

$$x_{\mathrm{D,quench}} \approx \frac{0.95}{2} = 0.475. \tag{3.5}$$

The quench mixture therefore contains substantial amounts of both hydrogen and deuterium. This situation is particularly clear in the stage-resolved description used for the small-peptide HDX-MS data, where quench/trapping was treated as exchange in an isotopically mixed solvent.

For an amide that is already deuterated, lowering the solvent D fraction increases the statistical availability of proton donors and therefore favors loss of label relative to the labeling stage, and at the same time, low pH and low temperature greatly reduce the absolute exchange rate. The observed deuterium retention is determined by the competition between these effects and by the duration of the quench period.

The quench stage is also chemically coupled to downstream sample preparation because in conventional bottom-up HDX-MS, acidic quench solutions may contain denaturing agents to unfold the protein rapidly and make cleavage sites accessible to acid-active proteases. Denaturation removes much of the structural protection that existed during labeling, and although the intrinsic chemical exchange rate is low under quench conditions, exposed amides are no longer protected by the native fold. The purpose of the low temperature and low pH is therefore to compensate kinetically for the deliberate loss of higher-order structure.

For intact small peptides, no analogous proteolytic unfolding step is required. The peptides are already small and, in the experiments relevant to this thesis, were selected as unstructured or minimally structured model systems. Their back-exchange during quench can consequently be considered mainly in terms of intrinsic sequence-dependent exchange chemistry and solvent conditions rather than native-state structural protection.

Because quench exchange remains finite, residence time before injection or trapping must be treated as a genuine kinetic variable. Two experiments with identical labeling times but different delays after quenching do not necessarily represent the same isotopic state, so this consideration becomes important when comparing manual, robotic, and continuous-flow workflows, where the number and duration of liquid-handling operations may differ.

3.5 Digestion and trapping

Proteolytic digestion is the defining step of bottom-up HDX-MS. After quenching, a labeled protein is cleaved into shorter peptides so that deuterium incorporation can be localized with substantially higher spatial resolution than would be possible from the intact protein mass alone, and proteases used for this purpose must remain active under the acidic conditions required to suppress back-exchange [4].

Digestion has two consequences for the isotope-exchange problem. First, it requires finite time, and deuterated amides remain exposed to proton-containing solvent during this interval and can therefore undergo back-exchange, and second, proteolysis changes the molecular context of each residue. New N-terminus and C-terminus are created, and the intrinsic exchange rate of an amide in a peptide fragment need not be identical to its rate when the same residue was embedded within the intact protein. Terminal effects and nearest-neighbor sequence effects contribute to these differences [16, 17].

The introduction of new peptide terminus is especially relevant to back-exchange, because intrinsic exchange rates are not uniform along a sequence, and terminal amides may exchange at rates that differ substantially from internal amide positions [4, 17]. Consequently, the deuterium lost from a peptide during downstream analysis is not expected to be distributed equally over its residues. This nonuniformity is one reason why a single peptide-level back-exchange percentage cannot, in general, be assumed to reconstruct the original residue-level labeling pattern.

After digestion, peptides are commonly retained on a short trapping or desalting column before analytical reversed-phase separation. Trapping concentrates the peptide mixture and removes components that would interfere with LC-MS. The sample remains under acidic, cold conditions, but the finite residence time again permits continuing exchange, therefore in kinetic terms, trapping is another solvent-phase propagation step rather than a neutral interval.

Trapping is directly relevant to the intact-peptide experiments considered here even though digestion is not, because in the reported millisecond experiment, the labeled peptides were trapped on a C18 cartridge before separation on the analytical column [6]. The absence of proteolysis simplifies the workflow because no new sequence terminus are created during sample processing, but it does not remove the possibility of deuterium loss while peptides are retained on a stationary phase. For this reason, the stage referred to as “quench/trapping” in simplified, kinetic models should be interpreted as an effective description of a sequence of low-temperature, low-pH operations. If the residence times and solvent compositions of the individual operations are known, they may in principle be modeled separately, and combining them is a practical approximation, not a statement that their physical environments are identical.

3.6 Liquid chromatography

Reversed-phase liquid chromatography separates peptides primarily according to their interactions with a hydrophobic stationary phase, and in a typical HDX-MS workflow, peptides are introduced under highly aqueous conditions and eluted by increasing the fraction of an organic modifier, most commonly acetonitrile. Formic acid is frequently used to maintain acidic conditions compatible with both suppression of back-exchange and positive-mode ESI [4].

HDX-MS chromatography differs from an LC method optimized only for separation quality, since long gradients can improve chromatographic resolution but also increase the time available for D→H exchange. For this reason, HDX-MS generally uses short gradients and cooled chromatographic hardware [8]. The experimental design is therefore a compromise between peptide separation and preservation of the deuterium label. Also, for each peptide, retention time provides an approximate measure of how long the molecule remains in the chromatographic system before entering the ESI source, as a result, a first kinetic approximation might treat the LC stage as exposure to a single quenched solution for a period equal to the retention time. Such a description captures one aspect of the problem-time-but does not reproduce the full physical environment experienced by the peptide.

The mobile phase is continuously changing, early in the gradient, a peptide encounters a predominantly aqueous solution. As the organic fraction rises, water activity, dielectric properties, acid dissociation equilibria, and possibly peptide conformational preferences change. At the same time, the peptide alternates between mobile-phase solvation and interaction with the stationary phase, so back-exchange during LC is not necessarily described by one constant intrinsic rate.

Experimental evidence for this complexity was provided by Sheff et al., who directly compared solution-phase and on-column back-exchange kinetics [24]. Their results showed that reversed-phase column interactions could either accelerate or decelerate exchange relative to free-solution predictions. The direction and magnitude of the effect depended on peptide sequence and length.

For peptides with a propensity to form secondary structure, interaction with the stationary phase could stabilize structural elements under quenched conditions. Such stabilization reduces the accessibility or exchange competence of particular amides and can slow back-exchange, and other peptides showed the opposite behavior, because column interactions could orient peptide conformations in a manner that increased exchange rates, with the observed behavior also depending on properties such as peptide charge [24].

This result is important for quantitative modeling because it demonstrates that an LC column is not simply a delay tube filled with quench buffer, and a residue can exhibit different exchange kinetics depending on the peptide in which it resides and on the interaction of that peptide with the stationary phase. Two peptides that overlap in sequence may therefore retain deuterium differently even for the residues they have in common. The millisecond peptide measurements discussed in this thesis used a short C18 separation with a four-minute linear

acetonitrile gradient [6], and such a rapid gradient limits exposure time, but it does not eliminate the effects described above, so the LC stage remains one of the principal experimental intervals between quenching and ionization.

A stage-resolved computational model can initially approximate LC by using the peptide retention time and a representative exchange rate, however, the limitations of that approximation should remain explicit. In particular, mobile-phase composition varies continuously, and the stationary phase itself can alter exchange behavior, therefore, any empirical factor introduced to reproduce LC deuterium loss has a physical justification as it represents effects not contained in an aqueous intrinsic-rate calculation.

3.7 Organic solvent effects

Acetonitrile is not merely a chromatographic variable. Its increasing concentration during reversed-phase LC changes the solvent in which back-exchange occurs, and intrinsic exchange-rate parametrizations derived from aqueous reference measurements therefore cannot automatically be applied throughout an LC gradient without modification.

Hamuro discussed several effects produced by adding acetonitrile to aqueous acidic solutions [4], as increasing the organic fraction reduces the concentration of water and lowers the dielectric constant of the medium, and these changes alter the dissociation equilibria of water, mobile-phase acids, and the peptide itself. The relationship between the pH value measured with a glass electrode and the concentrations of the catalytically relevant ionic species consequently differs from that in pure water.

For example, Hamuro considered an aqueous/formic-acid system containing approximately 50% acetonitrile, so under those conditions, the water dissociation constant shifts substantially relative to pure water, reducing the concentration of hydroxide at a given measured pH and lowering the minimum chemical hydrogen-exchange rate [4], and the pH corresponding to minimum exchange also shifts. Thus, the addition of organic solvent can suppress back-exchange rather than simply adding another uncontrolled source of deuterium loss.

This observation explains why the relationship between LC duration and back-exchange is not necessarily linear, in addition, shortening a chromatographic gradient reduces the total residence time, but the later portion of the gradient may occur under solvent conditions where exchange is already strongly suppressed.

The LC stage is more difficult to describe quantitatively than a solution phase of fixed composition because during chromatographic separation, back exchange remains sensitive to temperature, pH, solvent isotopic composition, and residence time. In addition, the organic co-solvent used in reversed-phase LC can influence intrinsic amide exchange rates in a non-trivial manner [4], and consequently, the exchange conditions experienced by a peptide may change during an LC gradient.

A further complication is that back exchange is not determined solely by bulk mobile-phase conditions, and peptide-stationary-phase interactions and

residual peptide structure under quenched conditions can alter exchange kinetics, causing experimentally observed rates to deviate from predictions based on unstructured solution models [24]. Therefore, a rigorous quantitative description of deuterium loss during LC would require consideration of both the changing chromatographic environment and peptide-specific interactions.

The difficulty is that standard intrinsic-rate calculations are calibrated primarily for aqueous systems. Temperature, sequence, and aqueous acidity can be treated with established reference-rate approaches [16, 17], whereas a general residue-level expression for a changing water/acetonitrile environment is less well established. This limitation provides a direct motivation for an empirical LC parameterization in a computational workflow, since an effective LC correction factor should not be interpreted as arbitrary curve fitting when it is used to account for known but incompletely characterized changes in solvent chemistry and column interactions. Its role is to absorb physical effects that are omitted when LC is represented as a simple period of exchange in cold aqueous quench buffer, so more mechanistic treatment would eventually replace such a factor with explicit solvent-composition and stationary-phase-dependent kinetics.

3.8 Electrospray ionization

At the end of liquid chromatography, the analyze is still a solvated species. However, mass spectrometry requires gas-phase ions, and it provides the transition between these two states so is a particularly important interface in HDX-MS [25]. From the perspective of deuterium retention, ESI is not simply a detector inlet because it is a dynamic sequence of droplet formation, solvent evaporation, droplet fission, desolvation, and gas-phase ion transfer.

In positive-mode ESI, the LC effluent emerges from a capillary maintained at high electrical potential relative to the inlet of the mass spectrometer, and then the electric field deforms the liquid meniscus into a Taylor cone, that a narrow jet emitted from the cone breaks into a spray of highly charged droplets [25].

The initial droplets are substantially larger than the final structures from which gas-phase analyze ions emerge, and as solvent evaporates, the droplet radius decreases while the charge density increases, therefore, electrostatic repulsion becomes progressively stronger relative to surface tension, and the maximum stable charge of an approximately spherical droplet is described by the Rayleigh limit,

$$Q_{\mathrm{R}} = 8\pi\sqrt{\varepsilon_0 \gamma R^3}, \tag{3.6}$$

where Q_{R} is the limiting droplet charge, ε_0 is the vacuum permittivity, γ is the surface tension, and R is the droplet radius [25]. When electrostatic repulsion approaches this limit, the droplet becomes unstable and undergoes fission, producing smaller charged offspring droplets.

Evaporation and fission can occur repeatedly, and the resulting droplets eventually reach nanometer dimensions, and analyte ions are transferred from these highly charged nanodroplets into the gas phase [25]. The exact route

followed by an analyte depends on its size, conformation, solvation, and charge, and therefore, three mechanistic descriptions are particularly relevant: the ion evaporation model (IEM), the charged residue model (CRM), and the chain ejection model (CEM) [25].

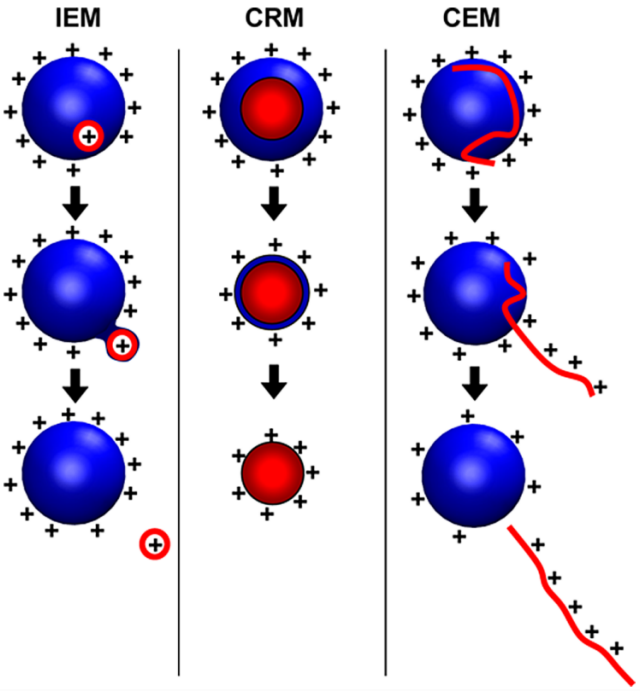


Figure 3.2: Schematic representation of the principal ion-release mechanisms proposed for electrospray ionization: the ion evaporation model (IEM), charged residue model (CRM), and chain ejection model (CEM). Adapted from Konermann et al. [25].

The ESI process matters for HDX due to several reasons such as, the composition of an evaporating droplet does not remain identical to the composition of the LC effluent, and components differ in volatility, and the relative abundance of water and organic solvent changes as evaporation proceeds. Secondly, the droplet experiences rapid concentration changes as its volume decreases, and third, analyzes pass through partially solvated states before becoming bare gas-phase ions. Finally, residual water vapor and other neutral molecules can interact with ions in the source and ion-transfer regions [8].

These processes provide several physically distinct opportunities for isotope exchange because exchange may continue while the peptide remains in a charged droplet or small solvent cluster, and additional proton/deuteron transfer can occur after complete desolvation if suitable neutral collision partners are present. Also, the magnitude of either contribution depends strongly on source and

instrument conditions, and it should therefore not be assumed that every HDX-MS instrument produces the same ESI-associated deuterium loss.

For the stage-resolved model considered in this thesis, this uncertainty is important, since, solution-phase exchange during labeling, quenching, and much of LC can be related to known pH, temperature, solvent composition, and residence time, therefore, ESI is less easily reduced to one set of bulk thermodynamic variables.

3.8.1 Ion evaporation model

The ion evaporation model was developed to describe the release of relatively small ions from highly charged droplets, and in this mechanism, the electric field near the surface of a nanodroplet becomes sufficiently strong that a small solvated ion can escape directly from the liquid phase [25].

Also, the departing ion experiences competing interactions, as electrostatic repulsion from the excess droplet charge favors ejection, whereas solvation favors retention within the droplet, and molecular simulations have refined the simple picture of an ion leaving a perfectly spherical interface. A departing ion can remain temporarily connected to the parent droplet through a small bridge or string of solvent molecules before separation is complete [25]. The initial IEM product can therefore retain a small solvation shell, which is subsequently removed during transfer toward the mass analyzer.

The IEM is generally associated with low-molecular-weight species and pre-existing solution ions. For HDX-MS, its significance is that small analyzes can pass through a regime in which they are neither bulk-solution species nor fully desolvated gas-phase ions, and exchange occurring in such clusters would not be described correctly either by a bulk aqueous model or by a purely gas-phase model.

Small peptides occupy a less sharply defined mechanistic regime than simple inorganic ions. The IEM nevertheless illustrates why the end of LC cannot be equated exactly with the end of solvent-mediated chemistry, also, residual solvation can persist as ions emerge from the droplet, and the isotopic environment during this transition may differ from that of the original mobile phase.

3.8.2 Charged residue model

The charged residue model provides the conventional description for large, compact, globular analyzes such as natively folded proteins, and instead of the analyze being expelled directly from a droplet, the droplet shrinks around the macromolecule, therefore, repeated evaporation and fission eventually produce a nanodroplet that contains essentially one protein molecule, so continued solvent loss then transfers the residual droplet charge to the analyze [25].

The CRM connects final protein charging to the Rayleigh charge of a droplet whose dimensions are comparable to the analyze, and this relationship helps explain why compact folded proteins generally produce relatively low charge

states compared with unfolded proteins of the same sequence, so a compact protein fits within a relatively small final droplet, which has a lower Rayleigh charge than a larger object.

From an HDX perspective, the CRM implies prolonged residence of the analyze inside a shrinking solvent environment, and the final stages of desolvation can concentrate salts, acids, and other mobile-phase components around the protein. The resulting microscopic conditions may differ substantially from the original LC solution even though the residence time is short.

3.8.3 Chain ejection model

The chain ejection model was proposed to explain the behavior of unfolded or disordered polymeric analyzes, as an unfolded protein exposes hydrophobic groups that would normally be buried in the core of the native structure. According to the CEM, this altered balance of hydrophobic and electrostatic interactions drives the disordered chain toward the droplet surface, so a portion of the polymer is expelled into the vapor phase and the remainder of the chain is progressively extracted from the droplet [25].

CEM is mechanistically distinct from the CRM. The analyze does not wait for a droplet to evaporate completely around it, and instead, the chain leaves the droplet while appreciable solvent is still present, therefore, the mechanism shares some features with ion evaporation because both involve direct ejection from a charged liquid interface.

This model is especially relevant to proteins that have been unfolded before or during ESI, and HDX-MS mobile phases are acidic and contain increasing amounts of organic solvent during chromatographic elution. These conditions differ substantially from a native aqueous environment and may favor loss of native protein structure before ion formation. For disordered proteins, unfolded proteins, and some peptides, an ejection-type mechanism may therefore be more appropriate than a charged-residue description.

The mechanistic classification should not be treated as a rigid assignment for every analyze, since, Konermann et al. emphasized that the boundaries between IEM, CRM, and CEM depend on molecular size, structure, and polarity [25]. For the purpose of HDX-MS, the practical conclusion is more important than assigning a unique mechanism as different analyze classes experience different pathways from solvated molecule to gas-phase ion, and these pathways may expose them to different opportunities for deuterium loss.

3.9 ESI and protein/peptide conformation

Protein conformation influences not only the exchange process in solution but also the ions generated during ESI, as a characteristic feature of electrosprayed proteins is the appearance of multiple charge states. Compact, folded proteins generally populate lower charge states, whereas unfolded proteins often produce broader distributions shifted toward higher charge [25, 29].

Several explanations contribute to this relationship, as unfolding increases the accessible surface area and can expose additional protonation sites, and it also changes the ionization mechanism from a CRM-like process for a compact globular protein toward CEM-like ejection for an extended chain [25]. The observed charge-state distribution therefore contains information about the physical state of the analyze during the solution-to-gas transition.

Konermann and Douglas compared positive-ion and negative-ion ESI for lysozyme, cytochrome c, and ubiquitin and showed that the relationship between conformation and charge-state distribution is polarity dependent [29]. Unfolding produced pronounced shifts toward higher charge states in positive-ion spectra for several of the systems studied, whereas the corresponding changes in negative-ion mode could be much smaller, and a lack of charge-state shift in negative mode therefore did not necessarily imply a lack of structural change.

These findings also caution against interpreting the gas-phase charge state as a direct measurement of the net solution charge, and ESI charging results from the entire ionization process and depends on molecular conformation, protonation/deprotonation chemistry, droplet evolution, and ion-release mechanism [25, 29].

The connection is relevant to HDX because gas-phase exchange can itself depend on charge state, because charge determines electrostatic interactions, gas-phase conformation, and the accessibility of protonated sites to neutral collision partners, so a peptide detected in two charge states may therefore experience different post-ESI exchange behavior even though both ions originated from the same LC peak.

This provides a physical basis for considering charge-state-dependent deuterium loss in quantitative models, and the generalized workflow proposed for small peptides introduced an ESI-associated loss term that can vary with charge state. Such a dependence should be regarded as an empirical representation until dedicated experiments separate droplet-phase exchange, gas-phase exchange, transmission effects, and any charge-dependent differences introduced earlier in the workflow.

For the intact-peptide dataset, multiple charge states are especially useful because they provide internal information about this stage, and if two charge states originating from the same peptide and labeling condition show systematically different apparent uptake after chromatographic co-elution, the discrepancy cannot be explained by the original solution labeling alone. So, the observation would instead point toward ionization, ion transfer, spectral processing, or another charge-dependent process downstream of LC.

3.10 Gas-phase HDX

Solution-phase HDX and gas-phase HDX describe chemically related isotope exchange reactions occurring in very different environments. In solution-phase HDX, exchange is mediated by solvent and depends on properties such as pH, temperature, hydrogen bonding, and solvent accessibility. In gas-phase HDX, an

isolated peptide or protein ion reacts with neutral deuterating or protonating molecules during its residence inside the mass spectrometer.

Gas-phase HDX has been demonstrated experimentally by deliberately introducing deuterated reagent gases into ion guides or ion traps, as Rand et al. introduced ND_3 into a travelling-wave ion guide and observed controlled exchange of peptide and protein ions over approximately 0.1-10 ms [28]. The extent of exchange depended on reagent pressure, ion charge state, molecular conformation, and accessibility of reactive sites. These experiments establish that proton/deuteron-transfer chemistry can proceed rapidly after electrospray.

The result is conceptually important for conventional solution-phase HDX-MS, since a peptide does not become chemically inert merely because it has left the liquid phase, and if residual water or other protic neutral molecules are present in an ion-source or ion-transfer region, D→H exchange is physically possible. Recent discussion of quantitative HDX-MS distinguishes bulk solution-phase back-exchange from additional loss that may occur in ESI droplets and in the gas-phase ion-sampling interface [8].

The experiments of Rand et al. should not be interpreted as evidence that all unexplained deuterium loss in a conventional HDX-MS experiment is caused by gas-phase exchange, and their experiments deliberately introduced a reactive deuterated gas under controlled conditions. They establish physical plausibility and demonstrate relevant time scales, but the extent of gas-phase D→H exchange in a specific analytical instrument depends on the actual composition, pressure, temperature, residence time, and electric-field conditions of that instrument.

Gas-phase HDX also has a strong conformational dependence, because in travelling-wave experiments, different charge states of the same protein could display different exchange behavior, reflecting differences in gas-phase structure and accessibility [28], and this observation reinforces the need to separate gas-phase behavior from the solution-phase structural information that conventional HDX-MS is intended to measure.

Site-specific experiments provide further evidence for the localized nature of gas-phase exchange. Rand et al. combined rapid gas-phase HDX with electron-transfer dissociation (ETD), allowing deuterium incorporation at individual rapidly exchanging sites to be inferred from fragment ions [30], also, this work demonstrated that gas-phase exchange and hydrogen scrambling are separate processes as deuterium can be transferred between an ion and an external reagent, or it can migrate internally among sites without changing the total isotope count.

For the modeling developed later in this thesis, gas-phase chemistry is best treated conservatively, and if solution-phase labeling, quenching, trapping, and LC cannot fully reproduce measured peptide uptake, an additional ESI/gas-phase contribution is plausible. The available literature supports the existence of such chemistry but does not by itself determine a universal rate constant for a particular instrument, and a fitted ESI term should therefore be described as an effective contribution rather than assigned uniquely to gas-phase back-exchange.

3.11 Hydrogen scrambling

Back-exchange and hydrogen scrambling both alter the relationship between the isotopic state created during labeling and the information recovered by the mass spectrometer, but they are not the same phenomenon as the distinction is important:

$$\underbrace{\mathrm{ND} + \mathrm{H} \longrightarrow \mathrm{NH} + \mathrm{D}}_{\text{back-exchange}} \tag{3.7}$$

changes the total amount of deuterium retained by the analyze, whereas scrambling redistributes hydrogen and deuterium among exchangeable positions within an ion. In an idealized scrambling event with no external isotope exchange, the total number of deuterons remains unchanged while their original locations are lost.

Scrambling becomes particularly important when tandem MS is used to infer residue-level HDX information, and conventional collision-induced dissociation (CID) deposits vibrational energy into a peptide ion. Before or during fragmentation, this activation can promote proton and deuteron migration among backbone and side-chain sites, and the isotope pattern of the resulting fragments then no longer represents the original solution-phase positions of the deuterium labels [4, 5].

Electron-based fragmentation methods such as electron-capture dissociation (ECD) and electron-transfer dissociation (ETD) were developed partly because they can preserve the spatial distribution of deuterium much more effectively than strongly ergodic collisional fragmentation. Also, Rand et al. demonstrated site-specific gas-phase HDX measurements using ETD under conditions designed to minimize collisional activation and hydrogen migration [30], and in their comparison, collisional activation produced extensive redistribution of label whereas ETD retained substantially more site-specific information.

Low scrambling does not mean that scrambling is impossible. Ion transmission conditions before fragmentation can contribute internal energy, and susceptibility varies among peptides, and recent systematic measurements by Mundorff et al. mapped activation thresholds for hydrogen migration across a panel of peptides and showed that scrambling propensity depends on properties including peptide size, charge, and sequence [31]. For a given peptide, the study found behavior consistent with a relatively well-defined activation threshold after which multiple exchangeable sites participate in redistribution. In some cases this threshold was close to the energy required for fragmentation.

This dependence has practical consequences for residue-resolved HDX-MS, because an instrument setting that preserves localization for one peptide cannot automatically be assumed to do so for every peptide or charge state, and validation with scrambling reporters or other controls is therefore needed when site-specific conclusions rely on fragment-ion deuterium content.

For the intact-peptide measurements used in the main quantitative analysis of this thesis, scrambling is less central than back-exchange because the reported

acquisition was performed in MS-only mode with low collisional activation and did not require fragmentation of the deuterated peptides [6]. The principal observable is the total isotopic mass shift of each intact peptide. Redistribution of deuterium among sites would not, by itself, change this total mass.

Nevertheless, the distinction remains important for the interpretation of the overall workflow, as a decrease in intact-peptide deuterium uptake requires actual loss of deuterium to the environment, whereas scrambling primarily destroys spatial information. The two phenomena may occur in the same instrumental region, but they should not be combined into a single mechanism.

Chapter 4

Computational Methods and Model Implementation

The computational model developed in this work was designed to follow the deuterium content of individual backbone amide sites through the successive stages of the HDX-MS experiment, as the central idea is that the measured deuterium uptake is not determined by the labeling reaction alone. Exchange continues after labeling, under different solvent compositions, temperatures, and chemical environments, and the state reached at the end of one stage provides the initial condition for the next, and a stage-resolved description is therefore more appropriate than treating all post-labeling deuterium loss as a single empirical correction.

The theoretical starting point is the generalized Linderstrøm-Lang description of exchange in isotopically mixed solvents [10]. For the small unstructured peptides considered in the computational analysis, the full open-closed conformational model can be reduced to a reversible two-state hydrogen/deuterium process, and this reduced description is then propagated through labeling and quench using physically calculated forward and reverse intrinsic rates. Also, the LC and ESI stages require additional treatment because their environments are not equivalent to homogeneous aqueous solution, and therefore, LC is represented by a solvent-dependent effective rate model, whereas ESI is treated phenomenologically using a charge-dependent local exchange term.

The chapter first develops the kinetic formulation used by the model and then describes its numerical implementation, so this ordering is deliberate and the purpose is not to introduce a second, independent theoretical framework, but to make explicit how the kinetic equations discussed earlier in the thesis are translated into the computational workflow used to reproduce the experimental uptake curves.

4.1 Reversible Exchange

The classical Linderstrøm-Lang model describes exchange through interconversion between an exchange-incompetent closed state and an exchange-competent open state, and in its usual form for hydrogen-to-deuterium exchange in highly deuterated solvent, the reaction is written as equation 2.9.

As it mentioned before, the irreversible step is a useful approximation when the solvent is treated as effectively fully deuterated and reverse exchange is negligible.

The experimental workflow considered here does not satisfy that condition at all stages, and the labeling buffer contains less than 100% D_2O, and the

quench and chromatographic stages contain substantial fractions of H_2O. Once an amide has become deuterated, it can therefore return to the protonated state, and a minimal chemical description must include

$$H \underset{k_{\mathrm{back}}}{\overset{k_{\mathrm{forw}}}{\rightleftharpoons}} D,$$

where k_{forw} is the pseudo-first-order rate for $H \to D$ exchange under the specified solvent conditions and k_{back} is the corresponding $D \to H$ rate, and in this notation both rates already contain the effects of solvent composition, temperature, acidity, and local sequence. They should therefore not be multiplied a second time by the bulk deuterium or protium fractions during state propagation.

The explicit treatment of both directions becomes particularly important in mixed H_2O/D_2O solvents because the system approaches an equilibrium that need not correspond to complete deuteration [10, 11]. This point provides the connection between the generalized theoretical model and the numerical two-state flow used later in the implementation.

4.2 Generalized Linderstrøm-Lang Model

A more complete representation retains the conformational opening and closing steps for both isotopic states, and the generalized reaction network can be written

$$H_{\mathrm{cl}} \underset{k_{\mathrm{cl}}}{\overset{k_{\mathrm{op}}}{\rightleftharpoons}} H_{\mathrm{op}} \underset{k_{\mathrm{back}}}{\overset{k_{\mathrm{forw}}}{\rightleftharpoons}} D_{\mathrm{op}} \underset{k'_{\mathrm{op}}}{\overset{k'_{\mathrm{cl}}}{\rightleftharpoons}} D_{\mathrm{cl}}. \tag{4.1}$$

The two protonated states, H_{cl} and H_{op}, have the same interpretation as in the classical model, and exchange occurs only from the open state. The deuterated amide is similarly allowed to occupy open and closed conformations, so the primed opening and closing rates permit isotopic substitution to alter the conformational equilibrium [10], and in the simpler case where isotopic substitution does not change local conformational stability,

$$k'_{\mathrm{op}} = k_{\mathrm{op}}, \qquad k'_{\mathrm{cl}} = k_{\mathrm{cl}}. \tag{4.2}$$

The generalized description separates two effects that are conflated in a one-way labeling model, so protein or peptide structure controls access to an exchange-competent state through the opening and closing rates, whereas solvent chemistry controls the forward and reverse isotope-transfer rates. This separation is useful in mixed solvents because equilibrium deuteration can change even if conformational protection is unchanged.

For a single amide, define the state vector

$$\mathbf{x}(t) = \begin{bmatrix} H_{\mathrm{cl}}(t) \\ H_{\mathrm{op}}(t) \\ D_{\mathrm{op}}(t) \\ D_{\mathrm{cl}}(t) \end{bmatrix}.$$

The kinetic equations can then be written as a linear system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{K}\mathbf{x},$$

with rate matrix

$$\mathbf{K} = \begin{bmatrix} -k_{\mathrm{op}} & k_{\mathrm{cl}} & 0 & 0 \\ k_{\mathrm{op}} & -(k_{\mathrm{cl}} + k_{\mathrm{forw}}) & k_{\mathrm{back}} & 0 \\ 0 & k_{\mathrm{forw}} & -(k_{\mathrm{back}} + k'_{\mathrm{cl}}) & k'_{\mathrm{op}} \\ 0 & 0 & k'_{\mathrm{cl}} & -k'_{\mathrm{op}} \end{bmatrix}. \tag{4.3}$$

For constant experimental conditions over a stage of duration t, the solution is

$$\mathbf{x}(t) = \exp(\mathbf{K}t)\mathbf{x}(0). \tag{4.4}$$

The matrix exponential is useful because the same mathematical operation can represent chemically different stages, and changing the solvent composition, temperature, or acidity changes the numerical entries of $\mathbf{K}$, but not the propagation rule itself. Therefore, conservation of population requires.

$$H_{\mathrm{cl}} + H_{\mathrm{op}} + D_{\mathrm{op}} + D_{\mathrm{cl}} = 1. \tag{4.5}$$

The deuterated fraction of the amide is obtained by projecting the state vector onto its deuterated components,

$$D(t) = D_{\mathrm{op}}(t) + D_{\mathrm{cl}}(t). \tag{4.6}$$

This formulation is also convenient computationally because an entire experimental stage becomes a propagator rather than a new model.

Under conditions where conformational opening and closing are fast relative to chemical isotope exchange, the four-state GLL system can be reduced by separation of time scales, and the closed and open populations of each isotopic form then remain in rapid pre-equilibrium, allowing the states to be grouped as

$$H = H_{\mathrm{cl}} + H_{\mathrm{op}}, \qquad D = D_{\mathrm{op}} + D_{\mathrm{cl}}.$$

The resulting dynamics can therefore be represented as an effective two-state process, $H \rightleftharpoons D$ unlike the classical irreversible LL description; however, the effective forward and reverse kinetics depend on isotope effects. The ratio

$$K_{\mathrm{back}} = \frac{k_{\mathrm{back}}}{k_{\mathrm{forw}}} \tag{4.7}$$

characterizes the relative intrinsic rates of reverse and forward exchange, while isotope-dependent changes in conformational protection are represented by

$$P' = P(1 + \delta).$$

Consequently, both the equilibrium deuterium population and the observed exchange rate can differ from those predicted by the classical LL model, even after reduction to an effective two-state description.

4.3 Sequential Representation of the HDX-MS Workflow

Let experimental stage n have duration τ_n and a propagator

$$\mathbf{\Phi}_n^{\tau_n} = \exp\left(\mathbf{K}_n \tau_n\right).$$

The output from that stage is

$$\mathbf{x}_n = \mathbf{\Phi}_n^{\tau_n} \mathbf{x}_{n-1}.$$

Repeated application gives

$$\mathbf{x}_N = \mathbf{\Phi}_N^{\tau_N} \mathbf{\Phi}_{N-1}^{\tau_{N-1}} \cdots \mathbf{\Phi}_2^{\tau_2} \mathbf{\Phi}_1^{\tau_1} \mathbf{x}_0.$$

Applied to HDX-MS, the conceptual sequence is

$$\boxed{\mathbf{x}_{\text{final}} = \mathbf{\Phi}_{\text{ESI}} \mathbf{\Phi}_{\text{LC}} \mathbf{\Phi}_{\text{quench}} \mathbf{\Phi}_{\text{label}} \mathbf{x}_0} \tag{4.8}$$

although the LC and ESI operators used in the final implementation are not identical to the homogeneous-solution matrix exponential, but the importance of Eq. 4.8 is the serial structure: the state is never reset between stages, so deuterium remaining after labeling is the initial condition for quench, the quenched state is passed into LC, and the LC output is passed to ESI.

A peptide-level uptake is obtained by summing over the exchangeable amides,

$$U(t) = \sum_{i=1}^{N_{\text{amide}}} D_i^{\text{final}}(t). \tag{4.9}$$

The resulting model is stage resolved. Deuterium loss need not be assigned to one global back-exchange percentage, and each stage can respond differently to experimental conditions.

The peptides used in the present analysis are treated as predominantly unstructured on the timescale relevant to the solution-phase exchange model, so under this approximation, explicit closed and open populations are not retained, and each amide is represented directly by

$$H \underset{k_{\text{back}}}{\overset{k_{\text{forw}}}{\rightleftharpoons}} D.$$

For a deuterated fraction $D(t)$, with $H(t) = 1 - D(t)$,

$$\frac{dD}{dt} = k_{\text{forw}}[1 - D(t)] - k_{\text{back}} D(t).$$

Defining

$$k = k_{\text{forw}} + k_{\text{back}},$$

the solution is

$$D(t) = D_{\text{eq}} + [D(0) - D_{\text{eq}}] \exp(-kt). \tag{4.10}$$

The implementation uses the equilibrium ratio

$$K_{\mathrm{eq}} = \frac{k_{\mathrm{forw}}}{k_{\mathrm{back}}}, \tag{4.11}$$

which gives

$$D_{\mathrm{eq}} = \frac{K_{\mathrm{eq}}}{1 + K_{\mathrm{eq}}} = \frac{k_{\mathrm{forw}}}{k_{\mathrm{forw}} + k_{\mathrm{back}}}. \tag{4.12}$$

This convention is used throughout this chapter, and it is the reciprocal of definitions in which an equilibrium constant is written in the back-exchange direction, so the direction of the ratio must always be stated explicitly.

Equation 4.10 is implemented directly for labeling and quench, but in vector form, for all amides of one peptide,

$$\mathbf{D}(t) = \mathbf{D}(0) \odot \exp(-\mathbf{k}t) + \mathbf{D}_{\mathrm{eq}} \odot [\mathbf{1} - \exp(-\mathbf{k}t)], \tag{4.13}$$

where $\odot$ denotes element-wise multiplication.

In an idealized treatment without equilibrium isotope effects, one might expect an unprotected amide at equilibrium to reproduce the deuterium fraction of the solvent. Mixed-solvent HDX does not generally satisfy $D_{\mathrm{eq}} = x$ where x is the deuterium atom fraction of the solvent [10, 11]. The deviation is described by an equilibrium fractionation factor, and defining $H_{\mathrm{eq}} = 1 - D_{\mathrm{eq}}$ the fractionation factor can be written

$$\phi = \frac{D_{\mathrm{eq}}/H_{\mathrm{eq}}}{x/(1-x)} = \frac{D_{\mathrm{eq}}}{H_{\mathrm{eq}}} \frac{1-x}{x}. \tag{4.14}$$

A value $\phi > 1$ corresponds to enrichment of deuterium at the amide site relative to the bulk solvent, and the mixed-solvent rate model used here predicts a small preference for deuterium in unprotected peptide amides. Fractionation is therefore not introduced as a separate fitted correction in the numerical propagation; it follows from the forward and reverse rate ratio through Eq. 4.12.

4.4 Intrinsic Exchange Kinetics in Mixed H_2O/D_2O Solvents

The calculation of intrinsic rates requires an acidity measure appropriate for an isotopically mixed solvent, and the implementation uses the effective acidity

$$pL = -\log_{10}[L^+], \qquad [L^+] = [H^+] + [D^+]. \tag{4.15}$$

The ionic product of pure water is calculated as a function of absolute temperature T using

$$\begin{aligned} pK_{\mathrm{w,H_2O}}(T) = 670.86 - \frac{34691.6}{T} - 105.15 \ln T \\ + 0.10757T + \frac{2358000}{T^2}. \end{aligned} \tag{4.16}$$

And the effect of solvent deuteration is introduced through

$$\Delta pK_{\mathrm{w}}(x) = 0.7282x + 0.05122x^2 + 0.0826x^3, \tag{4.17}$$

where the coefficient 0.05122 is the value used in the implementation, and the mixture ionic product is then

$$pK_{\mathrm{w}}(x,T) = pK_{\mathrm{w,H_2O}}(T) + \Delta pK_{\mathrm{w}}(x). \tag{4.18}$$

The glass-electrode reading, denoted here by pH_{read}, is converted to the effective acidity by

$$pL = pH_{\mathrm{read}} \frac{pK_{\mathrm{w}}(x,T)}{pK_{\mathrm{w,H_2O}}(T)}. \tag{4.19}$$

This treatment follows the mixed-solvent formulation used to calculate forward and reverse amide exchange rates [11], and also, it avoids treating the nominal glass-electrode reading as though the ionic chemistry of H_2O/D_2O mixtures were identical to that of pure water.

The total acid and base catalyst concentrations used numerically are

$$[L^+] = 10^{-pL},$$
$$[OL^-] = 10^{pL - pK_{\mathrm{w}}(x,T)},$$

where OL^- denotes the combined hydroxide/deuteroxide population.

In addition the isotope composition of the catalytic ionic populations does not have to equal the bulk solvent composition, and the implementation estimates the fraction of deuterium in the positive and negative catalytic ion pools as

$$f_D^{L^+} = \frac{x\phi_L}{1 - x + x\phi_L}, \tag{4.20}$$

and

$$f_D^{OL^-} = \frac{x\phi_{OL}}{1 - x + x\phi_{OL}}, \tag{4.21}$$

using the default values

$$\phi_L = 0.69, \qquad \phi_{OL} = 0.47.$$

The corresponding concentrations are therefore

$$[D^+] = [L^+] f_D^{L^+}, \qquad [H^+] = [L^+]\left(1 - f_D^{L^+}\right), \tag{4.22}$$

$$[OD^-] = [OL^-] f_D^{OL^-}, \qquad [OH^-] = [OL^-]\left(1 - f_D^{OL^-}\right). \tag{4.23}$$

These quantities enter directly into the calculation of the isotope-specific forward and reverse exchange contributions.

On the other hand, intrinsic exchange rates also depend strongly on temperature. Reference constants are therefore adjusted from the reference temperature using Arrhenius-type factors,

$$f_T(T) = \exp\left[-\frac{E_a}{R}\left(\frac{1}{T} - \frac{1}{T_{\mathrm{ref}}}\right)\right]. \tag{4.24}$$

The implementation uses

$$\begin{aligned} E_{a,\mathrm{acid}} &= 14\,000 \text{ cal mol}^{-1}, \\ E_{a,\mathrm{base}} &= 17\,000 \text{ cal mol}^{-1}, \\ E_{a,\mathrm{water}} &= 19\,000 \text{ cal mol}^{-1}, \end{aligned}$$

with $R = 1.987$ cal mol^{-1} K^{-1} and $T_{\mathrm{ref}} = 293$ K. These activation energies are consistent with the established empirical parameterization of backbone-amide exchange [4, 16].

Reference rates stored in the computational tables are converted from min^{-1} to s^{-1} before subsequent propagation, and both PDLA and trialanine-based reference choices are supported. The present workflow uses trialanine (`3Ala`) as the default reference, following the evidence that longer PDLA chains can exhibit residual structure and may therefore be unsuitable as a fully unstructured calibrant [6].

The intrinsic rate of an amide cannot be assigned from pH and temperature alone, because neighboring side chains alter exchange through steric and inductive effects, and these nearest-neighbor effects can be represented by multiplicative sequence factors [16, 17]. The implementation constructs separate correction tables for the relevant isotope combinations and then calculates acid and base factors for every residue. For residue i, the logarithmic acid-catalysis correction is obtained from the additive nearest-neighbor contributions,

$$A_i^{\log} = A_{L,i} + A_{R,i} + A_{\mathrm{term},i}, \tag{4.25}$$

where $A_{\mathrm{term},i} = 0$ for internal residues and represents the appropriate N- or C-terminal correction when applicable. The corresponding multiplicative correction factor is

$$A_i = 10^{A_i^{\log}}. \tag{4.26}$$

The base factor is treated analogously,

$$B_i = 10^{B_{L,i} + B_{R,i} + B_{\mathrm{term},i}}. \tag{4.27}$$

The left and right coefficients depend on the identities of the residue and its nearest neighbor. Terminal corrections are included where appropriate. Also, the proline positions do not contain an exchangeable backbone amide hydrogen and are therefore assigned zero exchange contribution. The implementation also updates corrections for Asp, Glu, His, and the C terminus as a function of their protonation state and effective acidity.

Table 4.1: Schematic use of nearest-neighbor factors in the intrinsic-rate calculation, and numerical values are obtained from the sequence-factor tables used by the implementation.

Quantity	Left contribution	Right contribution	Additional term
Acid factor	$A_L(i)$	$A_R(i-1)$	terminal correction
Base factor	$B_L(i)$	$B_R(i-1)$	terminal correction

4.5 Calculation of Forward and Reverse Intrinsic Rates

The rate-generation routine returns two pseudo-first-order rates for every amide, $(k_{\mathrm{forw},i}, k_{\mathrm{back},i})$ at a specified sequence, temperature, solvent deuterium fraction, and measured pH, so these rates are already mixture-specific, and the state propagation therefore uses them directly.

The implementation distinguishes four isotope combinations, HH, HD, DH, and DD, which describe the isotopic identities involved in the calibrated exchange pathways. The forward contribution is constructed from terms of the form

$$k_{\mathrm{forw},i} = k_{B,i}^{HH} + k_{A,i}^{HD} + k_{B,i}^{HD} + k_{W,i}^{HD}, \tag{4.28}$$

whereas the reverse rate is assembled as

$$k_{\mathrm{back},i} = k_{A,i}^{DH} + k_{B,i}^{DH} + k_{W,i}^{DH} + k_{B,i}^{DD}. \tag{4.29}$$

Writing the terms explicitly in the notation used above gives

$$\begin{aligned}
k_{B,i}^{HH} &= x\, k_{B,\mathrm{ref}}^{HH} B_i [OL^-] \left(1 - f_D^{OL^-}\right), \\
k_{A,i}^{HD} &= k_{A,\mathrm{ref}}^{HD} A_i [L^+] f_D^{L^+}, \\
k_{B,i}^{HD} &= x\, k_{B,\mathrm{ref}}^{HD} B_i [OL^-] f_D^{OL^-}, \\
k_{W,i}^{HD} &= x\, k_{W,\mathrm{ref}}^{HD} B_i,
\end{aligned}$$

and

$$\begin{aligned}
k_{A,i}^{DH} &= k_{A,\mathrm{ref}}^{DH} A_i [L^+] \left(1 - f_D^{L^+}\right), \\
k_{B,i}^{DH} &= (1-x)\, k_{B,\mathrm{ref}}^{DH} B_i [OL^-] \left(1 - f_D^{OL^-}\right), \\
k_{W,i}^{DH} &= (1-x)\, k_{W,\mathrm{ref}}^{DH} B_i, \\
k_{B,i}^{DD} &= (1-x)\, k_{B,\mathrm{ref}}^{DD} B_i [OL^-] f_D^{OL^-}.
\end{aligned}$$

The exact parameterization is implemented in the custom `calculate_kint()` routine, and its purpose is to combine chemical calibration data with solvent isotope composition rather than to fit separate forward and reverse rates directly to each uptake curve [11].

4.6 Software Environment

The computational workflow was implemented in Python, and the principal external packages were NumPy for array operations, Pandas for tabular data handling, SciPy for numerical integration, interpolation, and optimization, and Matplotlib for visualization. Custom routines contained the mixed-solvent rate calculation and effective-acidity model.

The first module calculates effective acidity and ionic isotope fractions in mixed water, and the second generates residue-level forward and reverse intrinsic rates. The command-line utility `kint.py` exposes this rate calculation for arbitrary peptide sequences, while the main HDX workflow performs the stage-resolved simulation and parameter optimization.

The input table contains one row for a measured peptide, exposure time, and charge state, and the fields used directly by the workflow include the peptide or protein identity, amino-acid sequence, experimental state, exposure time, charge z, measured centroid, protonated reference mass, retention time, and file identifier.

The implementation first converts the measured centroid to a mass quantity,

$$M = z\,C - z\,(m_D - m_H)\,, \tag{4.30}$$

where C is the stored centroid value and the isotope masses used in the current configuration are $m_H = 1.0078$, $m_D = 2.0141$ experimental uptake is then calculated as $U_{\mathrm{exp}} = M - M_{\mathrm{HP}}$ where M_{HP} is the protonated reference mass stored in the input data. Rows with zero nominal exposure are retained separately as reference measurements, while positive-exposure rows are propagated through the kinetic model.

4.7 Labeling Simulation

Residue-level forward and reverse rates are first generated at the labeling conditions, and in the current experiment configuration these conditions are $T_{\mathrm{label}} = 20°\mathrm{C}$, $pH_{\mathrm{read,label}} = 7.0$, and $x_{\mathrm{label}} = 0.95$, also each amide is initially taken to be protonated as $D_i(0) = 0$ and Eq. 4.10 is evaluated independently for every residue. The modeled peptide uptake immediately after labeling is

$$U_{\mathrm{label}}(t) = \sum_i D_i^{\mathrm{label}}(t). \tag{4.31}$$

In addition, the implementation contains an optional global labeling-time offset,

$$t_{\mathrm{eff}} = \max\left(0, t_{\mathrm{nominal}} + \tau_{\mathrm{label}}\right). \tag{4.32}$$

The motivation for this term is experimental rather than chemical, because millisecond HDX experiments require accurate calibration of mixer volumes, flow rates, and quench timing because even small timing differences can alter

early-time uptake [7]. The parameter τ_{label} is therefore interpreted as a timing calibration term, not as a new exchange pathway.

4.8 Quench Propagation

The labeling output is passed directly into the quench calculation, so the state of residue i at the start of quench is

$$D_i^{\mathrm{quench}}(0) = D_i^{\mathrm{label}}(t_{\mathrm{eff}}).$$

The current configuration uses $T_{\mathrm{quench}} = 0°\mathrm{C}$, $pH_{\mathrm{read,quench}} = 2.55$, and $x_{\mathrm{quench}} = 0.475$ with $t_{\mathrm{quench}} = 180$ s. A new pair of $k_{\mathrm{forw},i}^{\mathrm{quench}}$ and $k_{\mathrm{back},i}^{\mathrm{quench}}$ values is generated at these conditions, and the same reversible flow equation is then applied,

$$D_i^{\mathrm{quench}} = D_{i,\mathrm{eq}}^{\mathrm{quench}} + \left[D_i^{\mathrm{label}} - D_{i,\mathrm{eq}}^{\mathrm{quench}}\right] \exp\left[-\left(k_{\mathrm{forw},i}^{\mathrm{quench}} + k_{\mathrm{back},i}^{\mathrm{quench}}\right) t_{\mathrm{quench}}\right]. \tag{4.33}$$

This stage differs from a simple multiplicative back-exchange correction because forward exchange remains possible whenever the quench solution still contains deuterium.

4.9 Liquid-Chromatography Back-Exchange Model

The LC stage cannot be represented with the same confidence as labeling and quench, because the mobile phase changes continuously, contains organic solvent, and peptides interact with the stationary phase. Experimental studies have shown that peptide-column interactions can either accelerate or retard exchange depending on sequence and residual structure [24, 32], therefore, a single bulk-solution rate is not expected to reproduce all chromatographic behavior.

The LC treatment retains the chemically calculated residue-level back-exchange propensity but multiplies it by an effective solvent-dependent factor, more, the normalized organic-solvent fraction is represented by

$$v(t) = \begin{cases} V_0 + \dfrac{V_\tau - V_0}{\tau_g}\, t, & 0 \le t \le \tau_g, \\ V_\tau, & t > \tau_g, \end{cases} \tag{4.34}$$

with the current configuration

$$V_0 = 0.15, \qquad V_\tau = 0.40, \qquad \tau_g = 4.0 \text{ min}.$$

At each solvent composition, the effective LC rate for residue i of peptide p is

$$k_{\mathrm{LC},i}^{(p)}(t) = k_{\mathrm{base},i}^{(p)} \left[1 - v(t)\right]^{\eta} \exp\left[\gamma v(t) + \gamma_p'\right]. \tag{4.35}$$

The quantity $k_{\mathrm{base},i}^{(p)}$ is obtained from the intrinsic-rate calculator at the LC temperature and acidity in protonated solvent, and the current LC conditions are 0°C and $pH_{\mathrm{read}} = 2.55$. The default value of the power-law exponent is $\eta = 1$.

The parameter γ is global and describes the common dependence of the effective exchange rate on the chromatographic solvent gradient, and it is a phenomenological LC parameter rather than a fundamental exchange constant.

The additional term γ_p' is peptide specific, because it enters the exponent $\exp(\gamma_p')$ acts as a multiplicative correction to the integrated LC exposure. Positive values increase the modeled LC loss and negative values decrease it, meanwhile, the motivation for this term is the experimentally observed peptide-to-peptide variability in chromatographic back-exchange [24, 32].

The deuterium survival after a retention time RT_p is

$$D_i^{\mathrm{LC},p} = D_i^{\mathrm{quench},p} \exp\left[-\int_0^{RT_p} k_{\mathrm{LC},i}^{(p)}(t)\,dt\right]. \tag{4.36}$$

The integral is evaluated numerically, and a grid of 100 points is used to construct the LC rate profile, cumulative integration is performed with the trapezoidal rule, and the integrated rate is interpolated at the measured retention time.

A further peptide-specific gradient-shape parameter, $\beta_p[v(t) - v_{\mathrm{ref}}]$, was considered during development to allow different peptides to respond differently to the high-organic portion of the gradient. That extension is not part of the current final implementation and is therefore not included in Eq. 4.35. The retained model uses one global gradient parameter γ and one peptide-specific intercept γ_p'.

4.10 ESI Model

The physical environment changes more fundamentally during electrospray, because charged droplets undergo solvent evaporation and fission before analyte ions enter the gas phase, and gas-phase exchange behavior depends on ion charge, charge localization, conformation, and the identity of the exchange reagent [25, 28, 30], so the ESI stage was not modeled as another homogeneous bulk-solution exchange step.

Instead, the final implementation uses a phenomenological site-specific charge-density term. Potential protonation sites are taken to include the N terminus and basic side chains, and their default relative charge-affinity coefficients are

$$b_{\mathrm{N-term}} = 1.0, \qquad b_R = 1.0, \qquad b_K = 0.7, \qquad b_H = 0.3.$$

For charge state z, the charge-localization probability assigned to site s is

$$P_s(z) = \min\left[1,\, z\frac{\exp(\alpha b_s)}{\sum_{u\in S}\exp(\alpha b_u)}\right]. \tag{4.37}$$

The parameter α controls the selectivity of localization, and a larger value concentrates charge more strongly on sites with high b_s.

The influence of a charge site decreases with sequence distance, therefore the local charge density acting on amide i is

$$Q_i(z) = \sum_{s \in S} P_s(z) \exp\left(-\frac{|i-s|}{\xi_{\mathrm{ESI}}}\right). \tag{4.38}$$

The parameter ξ_{ESI} is a positive sequence-distance scale and must satisfy $\xi_{\mathrm{ESI}} > 0$ because a negative value would make the influence of a charge site grow rather than decay with distance.

The local ESI rate is then defined as

$$k_{\mathrm{ESI},i}(z) = k_{\mathrm{int},i}^{\mathrm{ref}} \exp\left[\lambda_{\mathrm{ESI}} Q_i(z)\right], \tag{4.39}$$

where the implementation uses the intrinsic rate vector generated under the quench conditions as the baseline $k_{\mathrm{int},i}^{\mathrm{ref}}$. This choice should be interpreted as a rate scale, not as a statement that the electrospray environment has the same chemical properties as the quench buffer. Also, the parameter λ_{ESI} is allowed to take either sign. For $\lambda_{\mathrm{ESI}} > 0$ larger modeled local charge density increases the effective ESI exchange rate, and for $\lambda_{\mathrm{ESI}} < 0$ the same charge-density term suppresses the effective rate. Negative values can therefore reproduce charge states in which additional charging is associated with stronger internal solvation or gas-phase stabilization rather than greater exchange exposure; more, gas-phase HDX experiments support a close connection among protonation, charge localization, and gas-phase accessibility, but do not imply a universal monotonic relation between total charge and exchange [28, 30].

The ESI survival equation is

$$D_i^{\mathrm{out}}(z) = D_i^{\mathrm{LC}}(z) \exp\left[-k_{\mathrm{ESI},i}(z) t_{\mathrm{ESI}}\right], \tag{4.40}$$

and the final peptide uptake is

$$U_{\mathrm{model}}(z,t) = \sum_i D_i^{\mathrm{out}}(z,t). \tag{4.41}$$

In the final reduced ESI fit, α and ξ_{ESI} are held fixed, both at 1.0, while t_{ESI} and λ_{ESI} are optimized separately for each peptide.

The fitted t_{ESI} should not be interpreted as a directly measured physical residence time in the source, because Eq. 4.39 is itself an effective parameterization, t_{ESI} acts as an effective kinetic exposure scale, and the actual solvent-to-gas transition occurs through a substantially more complicated sequence of droplet and gas-phase processes [25].

A charge-state accessibility function

$$A(z) = \frac{1}{1 + \exp[-A_0 - A_1(z - z_{\mathrm{ref}})]} \tag{4.42}$$

was investigated during development as a possible way to represent charge-state-dependent shielding, but the current implementation does not retain this term; it assumes unit accessibility for the charge-localizing sites. Accordingly, A_0 and A_1 are not fitted parameters in the final model reported here.

4.11 Optimization

The LC and ESI terms contain parameters that can partially compensate for one another, and if all parameters were optimized simultaneously, a large ESI correction could absorb an error originating in LC, while an overly flexible LC term could remove charge-state differences that should be attributed to the source. The first stage determines the global LC gradient parameter γ, together with the optional global labeling-time correction when that parameter is enabled. During this stage the ESI model is deliberately restricted as, α and ξ_{ESI} remain fixed, while only t_{ESI} and λ_{ESI} are adjusted per peptide.

The global LC parameter is constrained to $-10 \leq \gamma \leq 25$ bounded optimization is performed with the L-BFGS-B algorithm from SciPy and several starting points are tested to reduce dependence on initialization. With the global γ fixed, each peptide receives an independent γ'_p, but the LC curve is required to remain a plausible precursor to the ESI result. Because the implemented ESI operation only removes deuterium, an LC prediction below the experimental uptake cannot subsequently be corrected upward.

Define

$$\Delta_j = U_j^{\mathrm{LC}} - U_j^{\mathrm{exp}}.$$

The LC-envelope loss is

$$\mathcal{L}_{\mathrm{env}} = 20 \left\langle \min(\Delta_j, 0)^2 \right\rangle + \left\langle \max(\Delta_j, 0)^2 \right\rangle. \tag{4.43}$$

The asymmetric factor of 20 penalizes LC curves lying below the experimental data more strongly than curves remaining above them.

For peptide p, the local objective is

$$\mathcal{L}_{\gamma'_p} = \mathcal{L}_{\mathrm{env}} + 0.25\, \mathcal{L}_{\mathrm{ESI}} + r_{\gamma'} (\gamma'_p)^2, \tag{4.44}$$

with $r_{\gamma'} = 0.01$ the present code allows $-10 \leq \gamma'_p \leq 10$ the one-dimensional objective is first evaluated on a grid of 81 points across this interval. A bounded scalar minimization then refines the solution around the best grid point. This procedure was chosen because each trial γ'_p contains an inner ESI optimization and the resulting objective is not guaranteed to be perfectly smooth.

Once the LC parameters are fixed, the final ESI parameters are optimized per peptide. The fitted parameter vector is $\boldsymbol{\theta}_{\mathrm{ESI},p} = (t_{\mathrm{ESI},p}, \lambda_{\mathrm{ESI},p})$ with

$$\begin{aligned} 0 \leq t_{\mathrm{ESI}} &\leq 5000, \\ -10 \leq \lambda_{\mathrm{ESI}} &\leq 10. \end{aligned}$$

The sign of λ_{ESI} is therefore not constrained, and the basic uptake error is the mean squared error,

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{j=1}^{N}\left(U_{j,\mathrm{model}} - U_{j,\mathrm{exp}}\right)^2. \tag{4.45}$$

Because several charge states can be measured at the same labeling time, the objective also contains a signed charge-state separation term. For two charge states $z_a < z_b$ at exposure t,

$$\Delta U_{\mathrm{model}} = U_{\mathrm{model}}(z_b, t) - U_{\mathrm{model}}(z_a, t),$$

and

$$\Delta U_{\mathrm{exp}} = U_{\mathrm{exp}}(z_b, t) - U_{\mathrm{exp}}(z_a, t).$$

The separation penalty is

$$\mathcal{L}_{\mathrm{sep}} = \sum_{t}\sum_{z_a<z_b}\left[\Delta U_{\mathrm{model}} - \Delta U_{\mathrm{exp}}\right]^2. \tag{4.46}$$

The final ESI loss is

$$\boxed{\mathcal{L}_{\mathrm{ESI}} = \mathcal{L}_{\mathrm{MSE}} + 5\mathcal{L}_{\mathrm{sep}}} \tag{4.47}$$

so that the optimization reproduces both the absolute uptake and the direction of charge-state separation.

4.12 Model Evaluation

The present implementation evaluates the final simulated uptake using the root mean square error,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(U_{j,\mathrm{model}} - U_{j,\mathrm{exp}}\right)^2}. \tag{4.48}$$

The absolute RMSE is reported in deuterium-uptake units. A normalized form is also calculated,

$$\mathrm{RMSE}_{\mathrm{frac}} = \frac{\mathrm{RMSE}}{\max_j |U_{j,\mathrm{exp}}|}. \tag{4.49}$$

Residual inspection is important because a low aggregate RMSE can hide systematic deviations at short labeling times or consistent errors for one charge state, and the stage-resolved output is therefore retained during plotting: labeling, post-quench, post-LC, final ESI prediction, and experimental uptake can be displayed on the same exposure-time axis.

4.13 Interpretation and Limitations

The labeling and quench stages have the closest connection to established solution-phase exchange chemistry. Their forward and reverse rate constants are calculated from sequence, temperature, acidity, and isotopic composition, and the state is propagated using the reversible two-state solution.

The LC model is less fundamental, and organic solvent and stationary-phase interactions are condensed into the effective parameters η, γ, and γ'_p. This treatment is motivated by the experimentally observed variability of chromatographic back-exchange, including deviations caused by peptide adsorption and residual structure [24, 32]. A fitted γ'_p should therefore be interpreted as an effective peptide-specific chromatographic susceptibility, not as a molecular constant transferable without validation to another LC method.

The ESI description is more strongly phenomenological, so the solvent-to-gas transition cannot be represented rigorously by the same homogeneous solution-phase rate model used during labeling. Experimental gas-phase HDX studies demonstrate that peptide charge state, localization of charge, and gas-phase conformation can alter exchange behavior [28, 30], also hydrogen migration and scrambling can depend on charge state and ion activation [31]. The local charge-density formulation used here is therefore best regarded as an effective way to parameterize systematic charge-dependent deuterium loss.

Several more elaborate forms were examined during model development, including peptide-specific LC gradient-shape corrections and explicit charge-state accessibility functions. They are not part of the current reduced implementation, and excluding those terms reduces the number of fitted parameters and improves identifiability, although it also limits the range of ESI and LC behaviors that can be reproduced.

A final limitation concerns the separation of correlated parameters, so LC and ESI both decrease the final modeled uptake, so the data do not uniquely assign every unit of deuterium loss to one stage without independent experimental constraints, and the staged optimization and LC-envelope condition reduce this degeneracy but do not eliminate it. Direct measurements in which retention time, solvent composition, source settings, or quench duration are varied independently would provide stronger stage-specific validation.

Chapter 5

Results

The experimental dataset considered in this chapter contains three peptides, which are Angiotensin, Bradykinin, and atrial natriuretic peptide (ANP), and their sequences are DRVYIHP, RPPGFSPFR, and RSSCFGGRIDRIGAC, respectively. The peptides differ not only in sequence and length, but also in the number of charge states resolved in the mass spectra. As Angiotensin was analysed in the $z = 1$ and $z = 2$ charge states, Bradykinin was observed as $z = 2$, and ANP was analysed in the $z = 2$, $z = 3$, and $z = 4$ states.

All three peptides were investigated under the same set of salt conditions. The dataset contains a no-added-salt reference condition, denoted 0 mM, together with measurements at 150, 500, and 1000 mM NaCl and the corresponding concentrations of CsCl, and the common set of conditions therefore comprises This experimental design makes it possible to separate several sources of variation

Table 5.1: Salt concentrations investigated in the HDX-MS experiments.

Concentration	NaCl	CsCl
0 mM	-	-
150 mM	✓	✓
500 mM	✓	✓
1 M	✓	✓

in the measured deuterium uptake, and differences can be examined between peptide sequences, between charge states of the same peptide, between NaCl and CsCl at a fixed concentration, and across increasing salt concentrations. Before considering these salt-dependent effects, the performance of the stage-resolved kinetic model is first examined under the common 0 mM reference condition.

5.1 Kinetics at Zero Salt Concentration

The 0 mM condition provides the most direct starting point for evaluating the kinetic model because it is shared by all three peptides and does not contain an additional salt-dependent contribution; also, the comparison allows the effect of each modeled experimental stage to be followed sequentially, as figures 5.1-5.3 show the predicted uptake after labeling, quenching, LC separation, and ESI, together with the experimentally measured uptake for each observed charge state.

The curves before and after each processing stage are not equivalent, because the uptake calculated immediately after labeling represents the highest deuteration level predicted by the model, and quenching lowers this level, although the magnitude of the change depends on the peptide and exposure time.

A substantially larger reduction occurs after propagation through the LC stage; in addition, the final ESI step introduces a further, charge-state-dependent change before the modeled uptake is compared with the experimental measurement, therefore, the resulting sequence of curves gives a direct view of how deuterium retention evolves through the complete modeled workflow, rather than treating all post-labeling loss as a single correction.

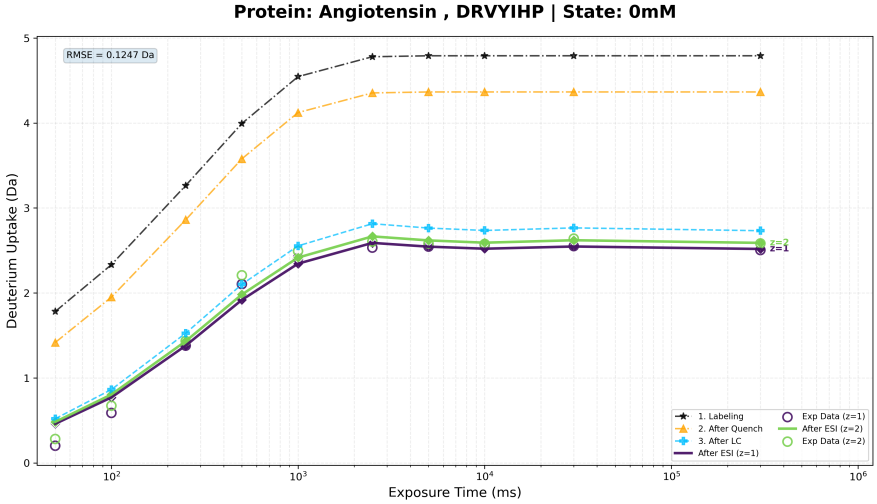


Figure 5.1: Multi-step model and experimental deuterium uptake for Angiotensin (DRVYIHP) under the 0 mM salt condition.

For Angiotensin, the predicted uptake increases rapidly over the millisecond-to-second region and approaches a plateau at longer exposure times (Figure 5.1), and the labeling calculation reaches approximately 4.8 Da at long exposure. Quenching reduces the predicted plateau to approximately 4.4 Da, while the subsequent LC stage produces a much larger decrease, leaving approximately 2.7-2.8 Da.

The final ESI calculation separates the two observed charge states, and the predicted $z = 1$ curve lies slightly below the corresponding $z = 2$ curve, particularly once the uptake approaches its plateau. This small charge-state separation is also present in the experimental measurements. More importantly, the final modeled curves follow the experimental uptake over the full exposure-time range: the initial increase, the curvature of the transition region, and the long-time plateau are all reproduced without a large systematic displacement.

The agreement is summarized by an RMSE of 0.1247 Da. The residual error is small compared with both the total uptake and the amount of deuterium removed between the labeling and final ESI stages, so the model accounts for most of the observed reduction in deuterium retention for Angiotensin while preserving the modest difference between its two charge states.

Bradykinin provides the closest agreement between predicted and measured

uptake among the three peptides under the no-salt condition (Figure 5.2), and because, unlike Angiotensin and ANP, only the $z = 2$ state existed in the data, so the final comparison is not affected by separation between multiple detected charge states.

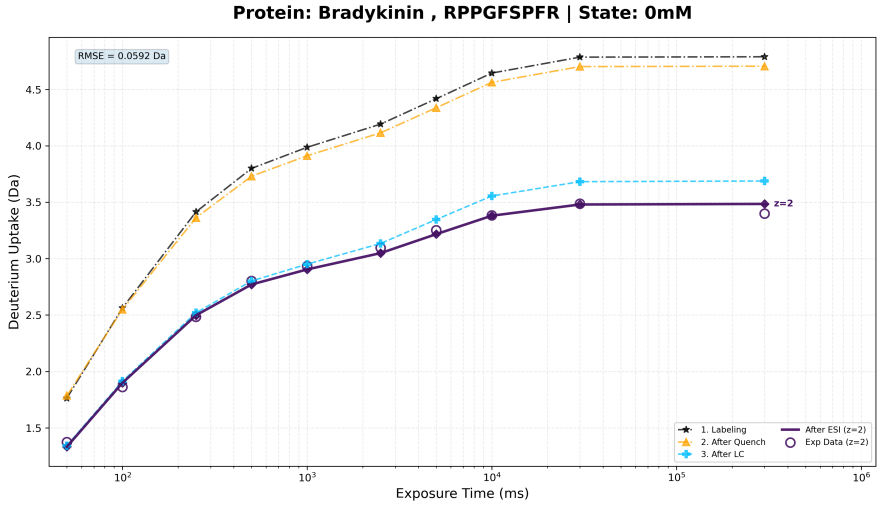


Figure 5.2: Stage-resolved modeled and experimental deuterium uptake for Bradykinin (RPPGFSPFR) under the 0 mM salt condition.

The labeling curve rises from approximately 1.8 Da at the shortest exposure to a plateau close to 4.8 Da, and the effect of the quench stage is comparatively small over much of the time range. The larger change again occurs during LC, which lowers the long-time predicted uptake to approximately 3.7 Da, and the effective ESI step produces a further reduction to approximately 3.5 Da.

After this final stage, the modeled curve closely follows the experimental measurements from the earliest time point through the approach to the plateau, so the measured points and final predicted curve are almost superimposed over much of the time range. The RMSE is only 0.0592 Da, which is the lowest of the three peptide-level fits shown here.

This comparison is useful because the agreement is obtained across the full shape of the uptake curve rather than at a single exposure time. The model captures both the relatively fast early increase and the slower late-time approach to saturation, and at the same time, the intermediate curves show that the experimentally observed uptake cannot be reproduced by the labeling calculation alone. The sequential reductions introduced by quenching, LC, and ESI are required to bring the calculated uptake into agreement with the measured values.

ANP spans the largest absolute uptake range and contains three experimentally resolved charge states, making it a more demanding test of the model (Figure 5.3). The calculated uptake after labeling rises rapidly and approaches approximately 13.4 Da at long exposure, and quenching lowers the predicted

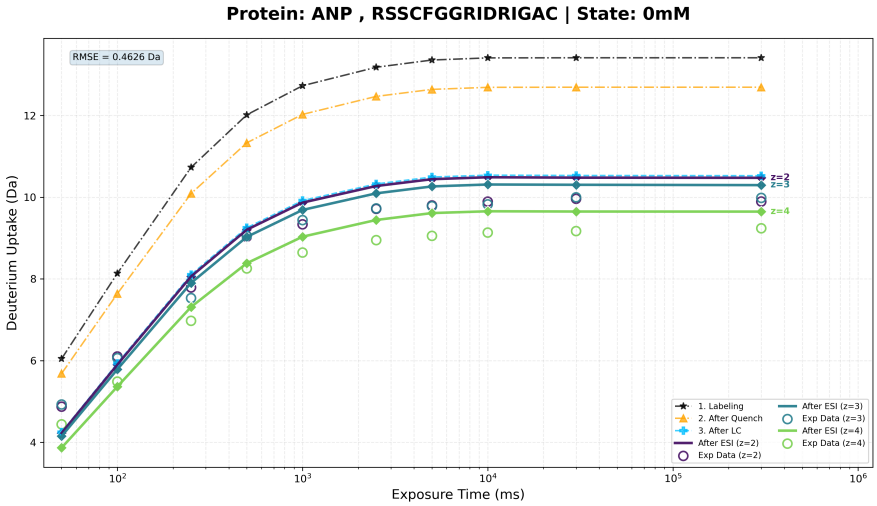


Figure 5.3: Stage-resolved modeled and experimental deuterium uptake for ANP under the 0 mM salt condition.

plateau to approximately 12.7 Da. As for the other two peptides, LC accounts for a larger reduction, bringing the prediction to approximately 10.5 Da.

The final ESI stage separates the three charge states, as the $z = 2$ and $z = 3$ predictions remain close to one another, whereas the $z = 4$ state shows a larger reduction in retained deuterium. The same ordering is present in the experimental measurements: $z = 2$ and $z = 3$ remain relatively similar, whereas the measured $z = 4$ uptake is lower over most of the exposure range. The model therefore reproduces not only the overall saturation behaviour of ANP but also the principal charge-state ordering observed experimentally.

Agreement for ANP is less close than for Bradykinin or Angiotensin based on the reported RMSE which is 0.4626 Da, and the final model curves tend to remain somewhat above the experimental measurements, particularly at longer exposure times. Nevertheless, the discrepancy is small relative to the total ANP uptake and does not alter the overall kinetic pattern. The model reproduces the rapid initial increase, the transition toward saturation, and the lower retained uptake of the $z = 4$ state.

Taken together, the three no-salt comparisons show that the stage-resolved model reproduces the principal features of the experimental HDX kinetics across peptides with different sequences, lengths, uptake ranges, and charge states. The absolute quality of the fit differs between peptides, with the closest agreement obtained for Bradykinin and the largest residual deviations observed for ANP, but the same framework describes all three systems without changing the basic sequence of modeled experimental stages.

The stage decomposition also provides information that would be hidden by a single global back-exchange correction, and for each peptide, the uptake

predicted directly after labeling is substantially higher than the final measured uptake. Quenching produces an initial reduction, while LC accounts for a larger fraction of the modeled post-labeling loss. The final ESI step then adjusts the prediction further and, for peptides observed in multiple charge states, introduces the separation between those states. In this sense, the model does more than reproduce the final endpoint: it provides a stage-resolved representation of how the initially incorporated deuterium is progressively modified before detection.

5.2 Salt Concentration Effect

The zero-salt results in the previous section showed that the stage-resolved model can reproduce the time-dependent uptake of all three peptides under the common reference condition. The next question is whether this agreement is maintained when the complete dataset is considered, including all NaCl and CsCl concentrations and all experimentally observed charge states.

Three complementary comparisons were used for this purpose, so the first examines absolute experimental and predicted uptake across all measurements, and the second removes the absolute uptake level from the comparison and asks whether the model reproduces the separation between different charge states of the same peptide. The third summarizes the remaining prediction error for each peptide and salt condition. Therefore, these analyses distinguish a model that merely follows the overall magnitude of deuterium incorporation from one that also reproduces the peptide, salt, and charge-state dependent structure of the experimental dataset.

The first comparison is shown in Figure 5.4, where each point represents an individual experimental uptake measurement and the corresponding final model prediction after the ESI stage. Measurements from all peptides, exposure times, charge states, and salt conditions are included in the same plot, and the dashed diagonal represents exact agreement between the model and experiment, $D_{\mathrm{model}} = D_{\mathrm{exp}}$ points positioned close to this line therefore correspond to measurements that are reproduced accurately, also points above the line indicate overprediction of the experimental uptake, whereas points below the line indicate that the model predicts less retained deuterium than was measured.

Across most of the uptake range, the experimental and predicted values remain close to the identity line, and the main deviation from the identity relation is not distributed randomly throughout the plot. A distinct population of measurements appears below the diagonal at high experimental uptake, particularly between approximately 6 and 10 Da, that these points are associated predominantly with the 500 mM and 1 M CsCl conditions, and for these measurements, the final predicted uptake is systematically lower than the experimental value.

The salt dependence of this discrepancy is informative, because measurements at 500 mM and 1 M NaCl remain much closer to the identity line over a similar experimental uptake range. High ionic concentration by itself therefore does not produce the large deviation, and instead, the underprediction is concentrated

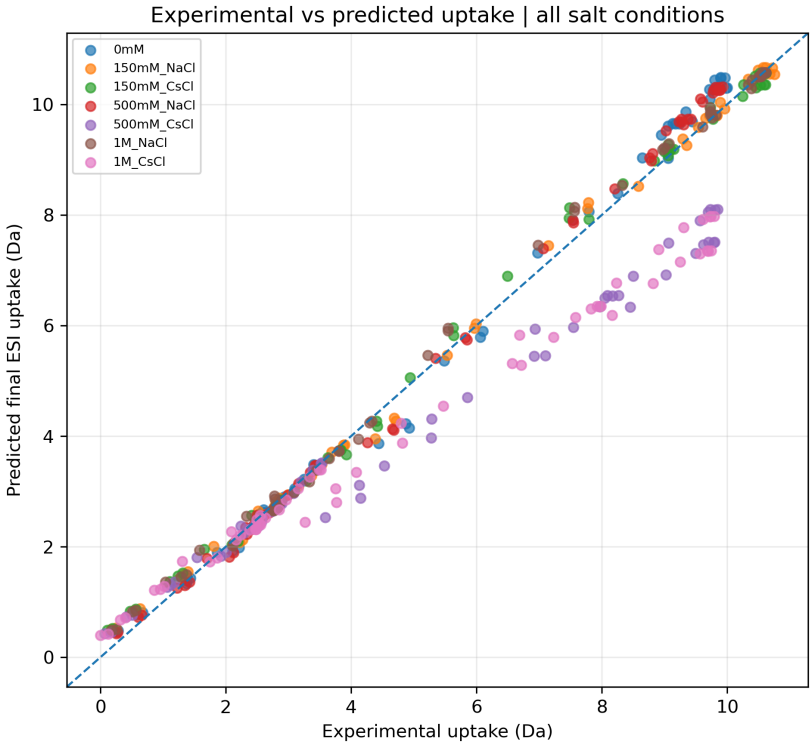
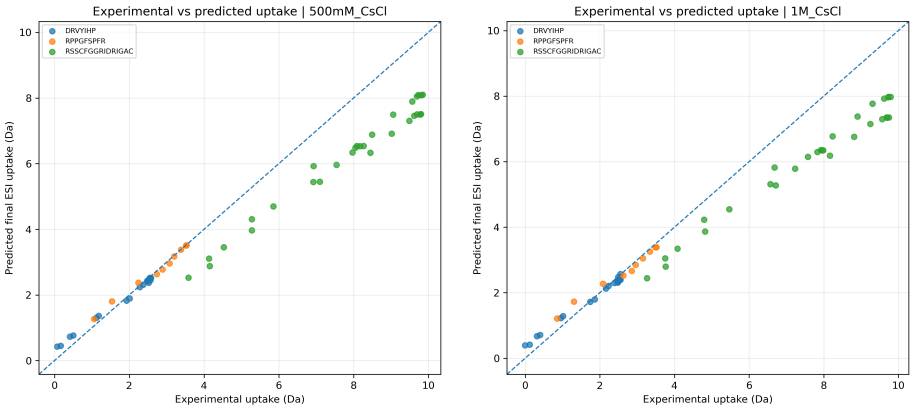


Figure 5.4: Experimental deuterium uptake compared with the final predicted uptake after the ESI stage for the complete dataset (both x-axis and y-axis units are Da).

in a more specific part of the dataset, involving concentrated CsCl and the high-uptake ANP measurements.

The condition-specific plots in Figure 5.5 show that the deviation observed in the global parity plot is strongly peptide dependent. At both 500 mM and 1 M CsCl, the Angiotensin and Bradykinin measurements remain comparatively close to the identity line over their respective uptake ranges. Their deviations are small and do not develop into a systematic displacement as the experimental uptake increases.

ANP behaves differently as under 500 mM CsCl, the ANP measurements form a distinct sequence below the identity line, and the discrepancy becomes larger as the experimental uptake increases: measurements with experimental uptake around 4-6 Da already show noticeable underprediction, while points in the highest uptake range, approaching 9-10 Da, remain approximately 2 Da below the identity relation. The direction of the error is therefore consistent across the ANP series, with the final model prediction systematically retaining less deuterium than is observed experimentally.



(a) 500 mM CsCl. (b) 1 M CsCl.

Figure 5.5: Experimental versus predicted final deuterium uptake under concentrated CsCl conditions.

A closely related pattern is present at 1 M CsCl. Angiotensin and Bradykinin again remain near the identity line, whereas the ANP measurements are displaced downward over a broad range of experimental uptake, and the largest deviations are observed at the longest exposures, where the experimental ANP uptake approaches approximately 10 Da but the predicted final uptake remains close to 8 Da. The persistence of the same peptide-specific pattern at both concentrated CsCl conditions indicates that the large residuals are not produced by a general loss of model accuracy at high uptake.

These two condition-specific comparisons therefore refine the interpretation of Figure 5.4. The global parity plot alone shows that a population of measurements is underpredicted, but Figure 5.5 identifies the origin of that population as the systematic deviation is dominated by ANP under concentrated CsCl, and Angiotensin and Bradykinin remain substantially better reproduced under the same salt conditions. The model limitation is therefore associated with a specific combination of peptide identity and salt environment rather than with CsCl concentration alone.

Figure 5.4 therefore suggests two conclusions, which first, the model reproduces the absolute experimental uptake well for the majority of the measurements despite the wide range of experimental conditions, and second, the remaining disagreement is structured rather than random: it is concentrated in a restricted high-CsCl regime.

Agreement in absolute uptake does not necessarily imply that the model correctly reproduces the effect of charge state, because a model could, for example, predict the average uptake of a peptide accurately while failing to reproduce the difference between its lower and higher detected charge states, so the charge-state separation analysis in Figure 5.6 was therefore used as a separate test.

The comparison is calculated independently for each peptide, salt

condition, and exposure time. The data are grouped according to (State, Protein, Sequence, Exposure) within each group, the lowest and highest available charge states are identified, $z_{\min}$ and $z_{\max}$ and the experimental charge-state separation is defined as

$$\Delta_z^{\exp} = D_{\exp}(z_{\max}) - D_{\exp}(z_{\min}).$$

The same quantity is calculated from the final model predictions,

$$\Delta_z^{\text{model}} = D_{\text{model}}(z_{\max}) - D_{\text{model}}(z_{\min}).$$

For ANP, which is observed in the $z = 2$, $z = 3$, and $z = 4$ charge states, the comparison therefore uses the two extreme states, $\Delta_z = D(z = 4) - D(z = 2)$, and for Angiotensin, the corresponding quantity is $\Delta_z = D(z = 2) - D(z = 1)$. In addition, Bradykinin contains only one charge state in the present dataset and therefore does not contribute to this analysis, and each point in Figure 5.6 consequently represents one peptide with at least two available charge states, at one exposure time and under one salt condition.

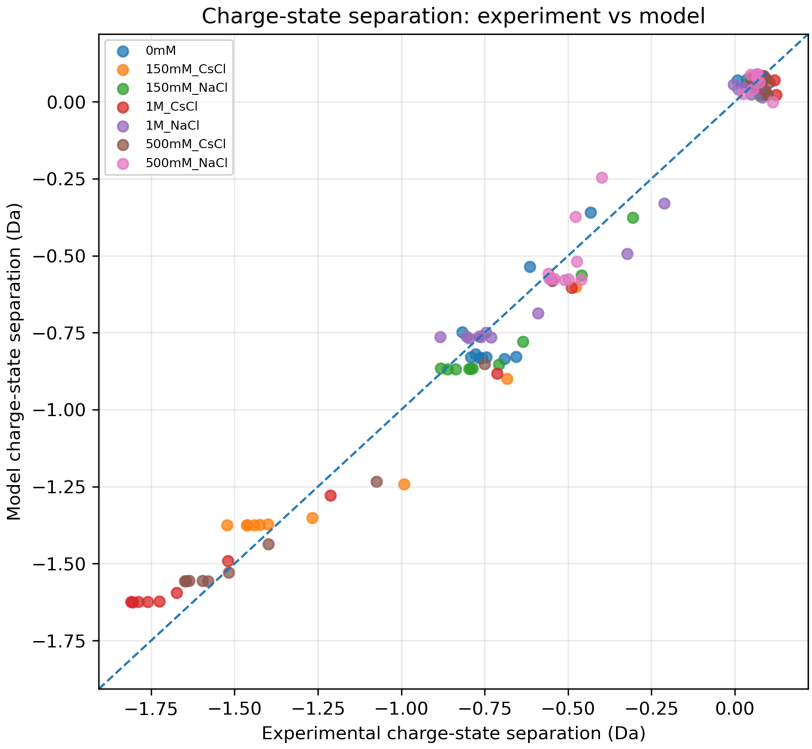


Figure 5.6: Experimental and modeled charge-state separation. Bradykinin is excluded because only one charge state is present in the dataset.

The sign of Δ_z has a direct experimental meaning, as a value close to zero indicates that the highest and lowest charge states retain approximately the same amount of deuterium, and a negative value means that the higher charge state has lower measured uptake than the lower charge state. Positive values, in contrast, would indicate higher uptake for the higher charge state.

Most of the experimentally resolved charge-state differences in Figure 5.6 are negative, so the dataset contains many conditions in which increasing the detected charge state is associated with lower retained deuterium, and the magnitude of this effect varies substantially. One population lies close to zero, corresponding to conditions with little measurable charge-state separation. A second group spans approximately -0.4 to -0.9 Da, while the strongest separations extend below -1.5 Da.

The most negative experimental values are particularly informative, because these points represent conditions in which the higher charge state contains more than approximately 1.5 Da less deuterium than the lower charge state. Some departures from the identity line are visible, indicating that the exact magnitude of the charge-state effect is not reproduced equally well in every condition, but the model preserves both its direction and much of its amplitude; therefore, the charge-state separation plot provides a different validation from Figure 5.4 and the global parity analysis tests whether the model predicts the correct absolute uptake, but it does not by itself determine which peptide is responsible for the deviation, and the normalized RMSE heatmap in Figure 5.7 resolves the prediction error simultaneously by peptide and salt condition.

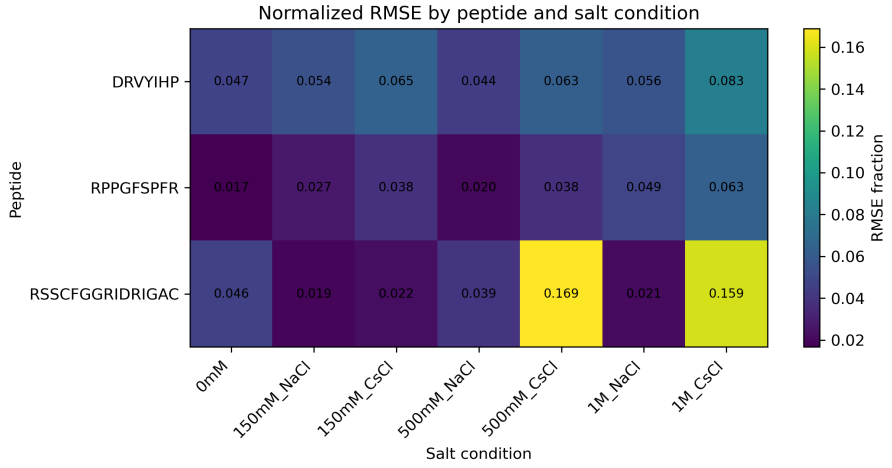


Figure 5.7: Normalized RMSE of the final predicted uptake for each peptide and salt condition.

For Angiotensin, the normalized RMSE remains within a relatively narrow range for all seven conditions, and the values are 0.047 at 0 mM, 0.054 and 0.065 at 150 mM NaCl and CsCl, 0.044 and 0.063 at 500 mM NaCl and CsCl, and

0.056 and 0.083 at 1 M NaCl and CsCl, respectively. There is therefore some loss of accuracy in concentrated CsCl, but no single Angiotensin condition separates strongly from the remainder of the row.

Bradykinin gives even smaller normalized errors, and the RMSE ranges from 0.017 under the no-salt condition to 0.063 at 1 M CsCl. The values remain below 0.05 for all conditions except the highest CsCl concentration, so this agrees with the stage-resolved result at 0 mM, where Bradykinin gave the closest fit among the three peptides.

ANP shows a markedly different pattern since five of the seven conditions have small normalized RMSE values: 0.046 at 0 mM, 0.019 at 150 mM NaCl, 0.022 at 150 mM CsCl, 0.039 at 500 mM NaCl, and 0.021 at 1 M NaCl, and these values are comparable to, and in several cases smaller than, those obtained for the shorter peptides.

The two concentrated-CsCl conditions are different, as the normalized RMSE rises to 0.169 at 500 mM CsCl and remains high at 0.159 for 1 M CsCl, and these are the two largest errors in the entire heatmap, and the concentration-matched NaCl conditions, by comparison, give errors of only 0.039 and 0.021.

This comparison is important because it separates three possible interpretations of the global residuals, and an increase in error is not a general property of ANP, since the model performs well for ANP under zero salt, 150 mM salt, and concentrated NaCl. It is also not a general consequence of high ionic concentration, because 500 mM and 1 M NaCl remain well reproduced. Instead, the largest errors are associated specifically with ANP in concentrated CsCl.

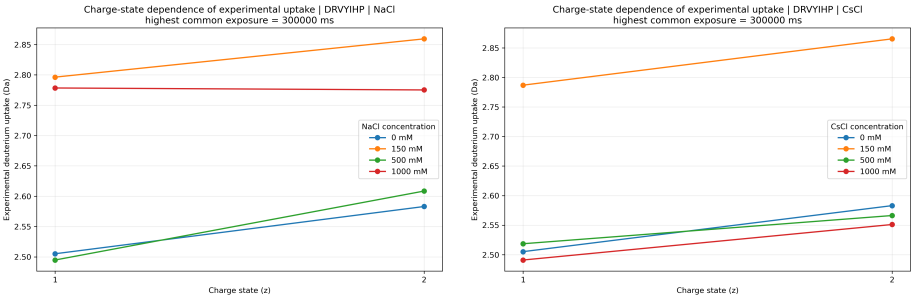
The three comparisons assess different aspects of model performance and should therefore be interpreted together. The all-state parity plot shows that the final model reproduces the absolute deuterium uptake over most of the experimental range, and the charge-state separation analysis then shows that this agreement is not obtained only at the level of an average peptide response: the model also reproduces much of the difference between the highest and lowest charge states measured for the same peptide, salt condition, and exposure time, and the last one, the normalized RMSE heatmap, identifies where the remaining disagreement is concentrated. The combined result is therefore more specific than a statement that the model 'fits the data well.' For most of the experimental parameter space, the model simultaneously reproduces the magnitude of uptake and the direction and approximate magnitude of charge-state separation. The principal exception is ANP under 500 mM and 1 M CsCl, where a systematic underprediction remains, so these two conditions form a distinct regime that can be separated from the otherwise comparatively consistent model performance.

5.3 Charge-State and Salt Concentration

The preceding analysis showed that the model reproduces much of the experimentally observed separation between charge states, and the experimental origin of this separation can be examined more directly by comparing the deuterium uptake of the individual charge states at the longest common exposure

time, 300 s. This comparison is particularly useful because it separates two related effects: changes in the absolute uptake of a peptide with salt concentration, and changes in the difference between charge states of the same peptide.

Figure 5.8 compares the $z = 1$ and $z = 2$ states of Angiotensin, while Figure 5.9 compares the $z = 2$, $z = 3$, and $z = 4$ states of ANP. NaCl and CsCl are shown in separate panels for each peptide. The 0 mM data are common to both panels and therefore provide the same reference point for evaluating the effect of the two salts.



(a) NaCl. (b) CsCl.

Figure 5.8: Charge-state dependence of the experimental deuterium uptake of Angiotensin (DRVYIHP) at the highest common exposure time of 300 s under different concentrations of **(a)** NaCl and **(b)** CsCl

Angiotensin shows only a small charge-state dependence at 300 s (Figure 5.8), and under the no-salt condition, uptake increases slightly from approximately 2.51 Da for $z = 1$ to 2.58 Da for $z = 2$. The corresponding charge-state separation is therefore small and positive, $\Delta_z = D(z = 2) - D(z = 1) \approx 0.08$ Da, the same overall behaviour is retained in most salt conditions: the $z = 2$ state generally contains slightly more deuterium than the $z = 1$ state, but the difference between the two remains small relative to the change in absolute uptake produced by changing the salt concentration.

The concentration dependence of the absolute uptake is more pronounced than the charge-state dependence itself. For both NaCl and CsCl, addition of 150 mM salt increases the uptake substantially. In NaCl, the $z = 1$ and $z = 2$ values increase to approximately 2.80 and 2.86 Da, respectively. Nearly identical values are observed at 150 mM CsCl. Thus, at the lowest salt concentration considered here, the response of Angiotensin is similar for the two chloride salts.

The behaviour changes at higher concentrations, which at 500 mM NaCl, the $z = 1$ uptake decreases to approximately 2.50 Da, while $z = 2$ remains somewhat higher, near 2.61 Da. The resulting charge-state separation is larger than under the no-salt condition, although it remains only about 0.1 Da.

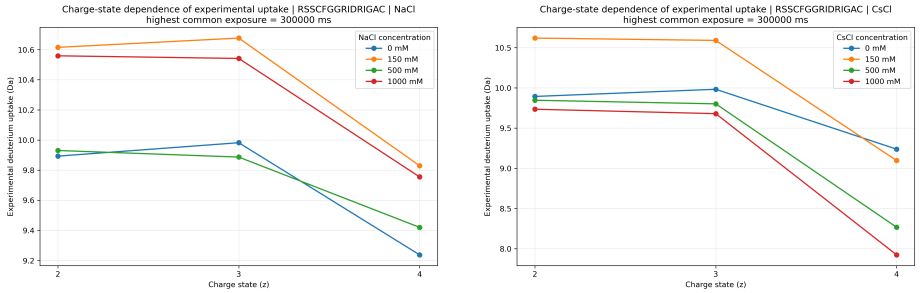
At 1 M NaCl, uptake increases again for both charge states. The two values converge to approximately 2.78 Da, and the charge-state separation becomes almost negligible. The NaCl dependence is therefore non-monotonic: uptake is elevated at 150 mM, decreases at 500 mM, and increases again at 1 M.

CsCl produces a different high-concentration response. After the increase at 150 mM, uptake decreases at 500 mM and remains low at 1 M. At the highest CsCl concentration, the uptake is approximately 2.49 Da for $z = 1$ and 2.55 Da for $z = 2$, and unlike the NaCl series, there is therefore no recovery of the experimental uptake between 500 mM and 1 M.

Comparison of Figures 5.8a and 5.8b suggests that the principal salt-dependent effect for Angiotensin is a change in the overall amount of retained deuterium rather than a strong modification of the difference between charge states. The $z = 1$ and $z = 2$ states generally move in the same direction as concentration changes. Salt identity becomes most apparent at high concentration: 1 M NaCl restores the uptake to a level similar to that observed at 150 mM, whereas 1 M CsCl maintains the lower uptake observed at 500 mM.

Thus, Angiotensin already provides evidence that NaCl and CsCl cannot be treated as interchangeable solely on the basis of nominal salt concentration. However, the effect on the charge-state separation itself remains relatively modest.

A substantially different pattern is observed for ANP (Figure 5.9), as the $z = 2$ and $z = 3$ states remain close to one another under all salt conditions, whereas the $z = 4$ state retains less deuterium. Salt concentration therefore affects not only the absolute uptake of ANP but also the magnitude of its charge-state separation.



(a) **(a)** NaCl. (b) **(b)** CsCl.

Figure 5.9: Charge-state dependence of the experimental deuterium uptake of ANP(RSSCFGGRIDRIGAC) at the highest common exposure time of 300 s under different concentrations of NaCl and CsCl.

Under the 0 mM condition, the $z = 2$ and $z = 3$ states have experimental uptakes of approximately 9.9 and 10.0 Da, respectively, and in comparison, the $z = 4$ state contains only approximately 9.2 Da. Even without added salt, ANP therefore exhibits a negative charge-state separation between the highest and lowest states, $\Delta_z = D(z = 4) - D(z = 2) \approx -0.65$ Da

This behaviour differs fundamentally from Angiotensin, where the higher charge state has slightly *more* retained deuterium, and for ANP, the highest charge state is already associated with lower uptake under the reference condition.

The NaCl series preserves this basic relationship (Figure 5.9a). At each

concentration, $z = 2$ and $z = 3$ remain close together, while $z = 4$ lies at lower uptake, and at 150 mM NaCl, uptake increases for all three charge states, byt the $z = 2$ and $z = 3$ states reach approximately 10.6-10.7 Da, while $z = 4$ reaches approximately 9.8 Da, and therefore the separation between $z = 4$ and the lower charge states is retained despite the increase in overall deuterium uptake.

At 500 mM NaCl, all three states shift to lower uptake, and the $z = 2$ and $z = 3$ values return to approximately 9.9 Da and the $z = 4$ state to approximately 9.4 Da. The charge-state separation is somewhat smaller under this condition than at 150 mM.

Increasing NaCl to 1 M again raises the uptake, so the $z = 2$ and $z = 3$ states reach approximately 10.5-10.6 Da, while $z = 4$ returns to approximately 9.8 Da. Consequently, the absolute uptake follows a similar non-monotonic concentration dependence to that observed for Angiotensin 150 mM and 1 M produce relatively high uptake, whereas 500 mM produces lower uptake.

Despite these concentration-dependent changes, the relative ordering of the ANP charge states is stable in NaCl. The $z = 2$ and $z = 3$ states remain close, and the $z = 4$ state remains lower by approximately 0.5-0.8 Da. NaCl therefore changes the absolute uptake substantially without producing a progressive increase in charge-state separation as its concentration rises.

However, the CsCl series is qualitatively different (Figure 5.9b), and as in NaCl, 150 mM CsCl increases the uptake of the $z = 2$ and $z = 3$ states to approximately 10.6 Da, but the $z = 4$ state does not follow the same increase, and its uptake remains close to 9.1 Da, giving a charge-state separation of approximately $\Delta_z \approx 9.1 - 10.6 \approx -1.5$ Da, thus, already at 150 mM CsCl, the separation between the highest and lowest ANP charge states is more than twice that observed without added salt.

A further divergence occurs as the CsCl concentration increases. At 500 mM CsCl, the $z = 2$ and $z = 3$ states remain near 9.8 Da, but the $z = 4$ state decreases to approximately 8.3 Da, and the difference between $z = 4$ and $z = 2$ is therefore approximately $\Delta_z \approx -1.6$ Da. At 1 M CsCl, the effect becomes stronger, so the $z = 2$ and $z = 3$ states remain around 9.7 Da, whereas the $z = 4$ uptake decreases to approximately 7.9 Da. The resulting separation approaches $\Delta_z \approx -1.8$ Da. The concentration dependence is therefore strongly charge-state dependent, and the lower ANP charge states change by less than approximately 1 Da over the full CsCl series, while the $z = 4$ state undergoes a much larger decrease at high concentration. Concentrated CsCl does not simply translate all three charge-state curves downward by the same amount; it preferentially increases the separation of the $z = 4$ state from $z = 2$ and $z = 3$.

Direct comparison of the NaCl and CsCl panels reveals that salt identity has a much larger effect on ANP than on Angiotensin, and in Angiotensin, the two charge states remain separated by only a few hundredths of a dalton under most conditions. NaCl and CsCl alter the absolute uptake differently at high concentration, but they do not produce a large divergence between $z = 1$ and $z = 2$. Therefore, the experimental data reveal an interaction between peptide identity, charge state, and salt composition, as Angiotensin is affected by the identity and concentration of the salt primarily through changes in absolute

uptake, while its charge-state dependence remains weak. Meanwhile, ANP shows a different response, as NaCl modifies absolute uptake maintaining a relatively stable separation between charge states, whereas increasing CsCl concentration progressively amplifies the difference between $z = 4$ and the lower charge states.

5.4 Salt and Charge Effects for ANP Uptake

The endpoint comparisons in Section 5.3 showed that ANP exhibits a pronounced interaction between charge state and salt composition, particularly for the $z = 4$ state in concentrated CsCl, and because those comparisons were based only on the highest common exposure time, they do not establish when the separation between conditions develops. The complete uptake curves were therefore examined for each ANP charge state over the full experimental exposure range from 50 ms to 300 s.

Figure 5.10 compares the experimental uptake curves for $z = 2$, $z = 3$, and $z = 4$ under all seven salt conditions, and in each case, uptake increases rapidly during the early part of the experiment and then approaches a condition-dependent plateau. The three panels nevertheless show that the magnitude of the salt effect depends strongly on charge state.

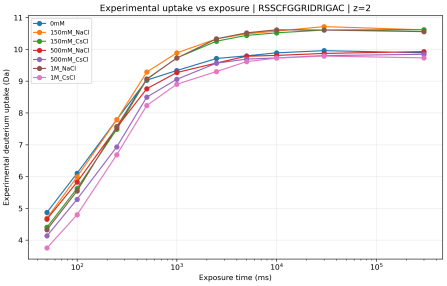
The $z = 2$ and $z = 3$ states exhibit very similar kinetic behaviour (Figures 5.10a and 5.10b). At the shortest exposure time, the measured uptake is already approximately 4-5 Da, depending on the salt condition. Uptake then rises steeply through the sub-second region and begins to level off after several seconds. Beyond approximately 10 s, the curves change relatively little and approach distinct plateau values.

For $z = 2$, the no-salt condition begins at approximately 4.9 Da and rises to just below 10 Da at long exposure. The salt-containing measurements separate from this reference throughout the kinetic trajectory, and the 150 mM NaCl curve is among the highest, reaching approximately 10.6-10.7 Da at long exposure; similarly, 150 mM CsCl and 1 M NaCl approach high plateau values.

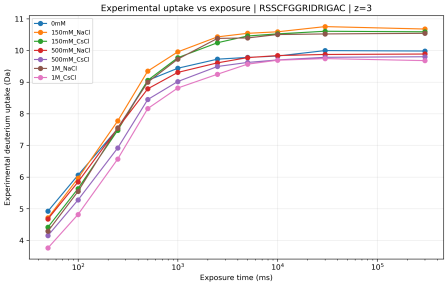
In contrast, the 500 mM and 1 M CsCl curves remain below these high-uptake conditions, and their separation is already visible at the earliest time points, and at approximately 100 ms, for example, the uptake in 1 M CsCl is lower than in the no-salt and 150 mM conditions. Although the difference becomes smaller as the curves approach saturation, it does not disappear completely. At 300 s, the 1 M CsCl curve remains near 9.7-9.8 Da, compared with approximately 10.6 Da for the highest-uptake conditions.

The $z = 3$ state follows essentially the same ordering as uptake increases from approximately 4-5 Da at 50 ms to values close to 10 Da or above at long exposure. Again, 150 mM NaCl, 150 mM CsCl, and 1 M NaCl produce the highest long-time uptake, whereas 500 mM and 1 M CsCl remain lower.

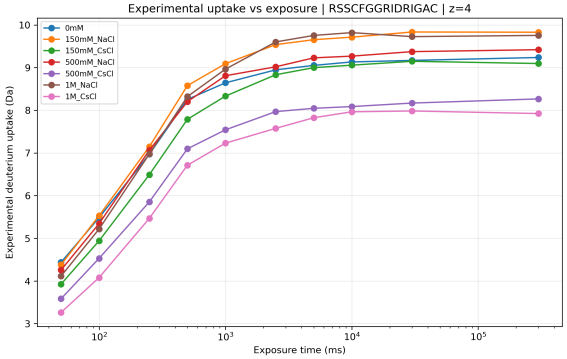
The similarity between the $z = 2$ and $z = 3$ panels is important because changing from $z = 2$ to $z = 3$ does not substantially alter the shape of the kinetic response to salt, and the two charge states show comparable initial uptake, similar curvature through the rising part of the exchange curve, and similar



(a) $z = 2$.



(b) $z = 3$.



(c) $z = 4$.

Figure 5.10: Experimental deuterium uptake of ANP (RSSCFGGRIDRIGAC) as a function of exposure time for **(a)** $z = 2$, **(b)** $z = 3$, and **(c)** $z = 4$.

condition-dependent plateaux. This agrees with the endpoint analysis, where the $z = 2$ and $z = 3$ ANP states remained close to one another under all salt conditions.

The $z = 4$ state differs substantially from the two lower charge states (Figure 5.10c) since uptake again rises rapidly with exposure time and approaches a plateau, but both the absolute uptake and the salt-dependent separation are larger functions of the experimental condition.

Even without added salt, the $z = 4$ curve lies below the corresponding $z = 2$ and $z = 3$ curves, and the no-salt uptake increases from approximately 4.4 Da at the shortest exposure to approximately 9.2 Da at 300 s, so this is lower than the approximately 10 Da reached by the two lower charge states under the same condition.

NaCl modifies the absolute uptake but does not produce a progressive depression with increasing concentration, and the 150 mM NaCl condition gives one of the highest curves over most of the exposure range and reaches approximately 9.8 Da at long exposure. The 1 M NaCl curve behaves similarly and also approaches approximately 9.8 Da, and the 500 mM NaCl condition

lies between these high-uptake curves and the no-salt reference, reaching approximately 9.4 Da at 300 s.

CsCl has a markedly different effect on the $z = 4$ state; at 150 mM, the uptake is already lower than that observed with 150 mM NaCl over most of the time range, and increasing the CsCl concentration to 500 mM produces a much larger reduction. This curve begins at only approximately 3.6 Da at the shortest exposure and remains below the no-salt and NaCl curves throughout the complete experiment; its final uptake is only about 8.3 Da.

The strongest reduction occurs at 1 M CsCl, as at the earliest exposure, the uptake is approximately 3.3 Da, more than 1 Da below the no-salt value, and the difference persists as exposure increases: at approximately 1 s the uptake remains near 7.2 Da, compared with approximately 8.7 Da without salt, and at the longest exposure the curve reaches only about 7.9 Da. The high-CsCl effect is therefore not generated only at the final 300 s measurement, and the separation is present from the shortest experimental exposure and remains throughout the complete kinetic trajectory. This observation is important because it shows that the large endpoint difference identified previously is the accumulated result of a persistent difference in measured uptake rather than an isolated divergence appearing only after long labeling times.

Comparison of the three charge states reveals that the difference between NaCl and CsCl depends on both salt concentration and charge state. For $z = 2$ and $z = 3$, the two salts alter the position of the kinetic curves but preserve a broadly similar saturation profile, and high CsCl produces somewhat lower uptake than the highest-uptake NaCl conditions, but the effect remains moderate. But for $z = 4$, the distinction is much larger, as increasing NaCl concentration does not progressively reduce uptake, and both 150 mM and 1 M NaCl produce relatively high uptake over the complete time range, while 500 mM NaCl gives an intermediate response; therefore, the NaCl concentration dependence is non-monotonic. CsCl displays a different progression, as the 150 mM condition produces a moderate reduction relative to the corresponding NaCl curve, but the difference becomes much larger at 500 mM and larger again at 1 M for the $z = 4$ state; increasing CsCl concentration is therefore associated with progressively lower uptake over most of the measured time range.

Chapter 6

Conclusion

The aim of this thesis was to investigate whether the deuterium uptake measured in hydrogen/deuterium exchange mass spectrometry can be interpreted more quantitatively by following the isotopic state of a peptide through the successive stages of the experimental workflow, and rather than treating back-exchange as a single correction applied to the final measurement, the approach developed here describes deuterium incorporation and loss as a sequence of kinetic processes occurring during labeling, quenching, chromatographic separation, and electrospray ionization.

A central element of this work was the extension of the exchange description to isotopically mixed H_2O/D_2O solvents, and under these conditions, hydrogen-to-deuterium and deuterium-to-hydrogen exchange occur simultaneously, and the system does not necessarily approach complete deuteration. The generalized treatment therefore uses separate forward and reverse exchange rates and propagates the deuterium population of each backbone amide between successive experimental stages. Sequence-dependent intrinsic exchange rates, temperature, acidity, and isotopic composition are incorporated explicitly, providing a physical basis for the labeling and quench calculations.

The resulting computational framework was applied to millisecond HDX-MS measurements of Angiotensin, Bradykinin, and ANP. The experimental dataset covered labeling times from 50 ms to 300 s and included measurements without added salt as well as different concentrations of NaCl and CsCl, and multiple charge states were available for Angiotensin and ANP, allowing the model to be evaluated not only against the absolute deuterium uptake but also against charge-state-dependent differences in retained deuterium.

Under the no-added-salt condition, the stage-resolved model reproduced the principal kinetic behaviour of all three peptides, and the agreement was particularly close for Bradykinin and Angiotensin, while somewhat larger residual deviations remained for ANP. Importantly, the modeled uptake immediately after labeling was consistently higher than the final uptake measured experimentally. This confirms that the measured mass shift cannot in general be interpreted as the isotopic state generated during labeling alone.

The stage decomposition provides a more detailed interpretation of this difference. Quenching produces an initial reduction in deuterium content, but the chromatographic stage accounts for a larger fraction of the modeled post-labeling loss for the peptides investigated here; also, a further effective contribution associated with electrospray ionization is required to reproduce the final measured uptake and, in particular, the separation observed between different charge states of the same peptide. Consequently, the results support a description in which back-exchange is distributed across several experimentally distinct stages rather

than represented by one universal back-exchange percentage.

The charge-state dependence provides additional evidence that processes occurring downstream of solution labeling contribute to the final measurement. Angiotensin exhibits only a small difference between its $z = 1$ and $z = 2$ states, whereas ANP shows a substantially lower uptake for the $z = 4$ state than for the $z = 2$ and $z = 3$ states. Because ions corresponding to different charge states originate from the same peptide and labeling condition, these differences cannot be explained solely by the initial solution-phase labeling reaction. They instead point toward charge-dependent effects associated with ionization, ion transfer, gas-phase chemistry, or related instrumental processes.

Salt composition introduces an additional level of complexity. The experimental results demonstrate that NaCl and CsCl cannot be described solely through their nominal ionic concentration. For Angiotensin, salt primarily modifies the absolute amount of retained deuterium while the difference between charge states remains comparatively small. ANP displays a much stronger interaction between salt identity and charge state. In NaCl, the $z = 2$ and $z = 3$ states remain close to one another and the $z = 4$ state remains moderately lower over the investigated concentration range. In CsCl, however, the separation of the $z = 4$ state becomes progressively larger as the concentration increases.

The largest systematic discrepancies between model and experiment occur for ANP at 500 mM and 1 M CsCl. Under these conditions, the model predicts substantially less retained deuterium than is observed experimentally, while Angiotensin and Bradykinin remain considerably better described under the same salt conditions. The discrepancy is therefore not simply a consequence of high ionic strength or of CsCl alone. Instead, it reflects a specific interaction between peptide identity, salt composition, and charge state that is not captured by the present model.

This observation is important because it also defines the limits of the current framework. The labeling and quench stages are based most directly on established solution-phase exchange chemistry, whereas the descriptions of chromatography and electrospray necessarily contain more phenomenological elements. In particular, chromatographic exchange can depend on peptide retention, stationary-phase interactions, solvent composition, and residual structure, while electrospray involves a rapidly evolving and non-equilibrium environment that cannot be represented rigorously as a homogeneous aqueous solution. Parameters introduced for these stages should therefore be interpreted as effective quantities rather than universal molecular constants.

Both chromatographic and electrospray-associated processes can reduce the final calculated deuterium uptake, and fitting the final mass alone cannot uniquely determine how much loss should be assigned to each stage, so the staged optimization used in this work reduces this degeneracy but cannot eliminate it completely without additional experimental constraints.

More generally, the results of this thesis show that quantitative HDX-MS requires consideration of the complete path between labeling and detection. Also, the isotopic state recorded by the mass spectrometer is the product of the labeling reaction together with all subsequent environments encountered by

the analyte. Treating these environments explicitly exposes information that is hidden by a single global back-exchange correction and makes it possible to identify the stages and experimental conditions responsible for deviations between theoretical and measured uptake.

The framework presented here should therefore be regarded as a step toward mechanistically interpretable HDX-MS rather than as a final universal correction scheme. Its principal contribution is to establish a computational structure in which physically defined exchange processes can be introduced, tested independently, and refined as new experimental information becomes available. With improved constraints on salt effects, chromatography, and the electrospray region, such multi-step approaches may ultimately allow HDX-MS measurements to be interpreted in terms of absolute exchange kinetics and structural protection with substantially reduced dependence on empirical back-exchange corrections.

Appendices

Appendix A

Supplementary Model and Data Analysis

This appendix collects the numerical information required to assess the implementation and performance of the serial HDX model without interrupting the main discussion in Chapters 4-6. The emphasis is placed on the complete experimental comparison matrix, the optimized model parameters, the stage-by-stage loss of deuterium, overall goodness of fit, and the direct comparison of salt and charge-state-dependent effects.

A.1 Experimental Dataset and Conditions

The experimental comparison comprised three peptides: ANP, Angiotensin, and Bradykinin. Measurements were available in the absence of added salt and in the presence of either NaCl or CsCl at concentrations of 150, 500, and 1000 mM. Ten labeling times were used,

$$0.05,\ 0.10,\ 0.25,\ 0.50,\ 1.0,\ 2.5,\ 5,\ 10,\ 30,\ 300 \text{ s},$$

covering the transition from the millisecond regime to near-saturation.

ANP was detected in three charge states, $z = 2$, 3, and 4; Angiotensin in $z = 1$ and 2; and Bradykinin in $z = 2$. The six peptide-charge-state series, evaluated at ten labeling times under seven solution conditions, produced 420 aggregated observations, and the dataset used for fitting additionally retained the individual measurements before aggregation, whereas the tables below report condition-level summary values.

Table A.1: Peptide sequences, charge states, and experimental coverage used in the model comparison.

Peptide	Sequence	Charge states	Labeling range
ANP	RSSCFGGRIDRIGAC	2+, 3+, 4+	0.05-300 s
Angiotensin	DRVYIHP	1+, 2+	0.05-300 s
Bradykinin	RPPGFSPFR	2+	0.05-300 s

Solution conditions: 0 mM, 150 mM NaCl, 150 mM CsCl, 500 mM NaCl, 500 mM CsCl, 1 M NaCl, and 1 M CsCl.

An important feature of the present implementation is that the intrinsic labeling and quench rate vectors were identical for all salt states. Thus, salt-dependent differences were not introduced through a salt-specific modification of

k_{forw}, k_{back}, or k_{int}. The differences between salt conditions therefore enter the present model through the downstream LC and ESI treatment. This point is important when interpreting the fitted parameters and the high-salt discrepancies discussed below.

A.2 Optimized Model Parameters

The model contains a global LC scaling parameter, γ_{LC}, and a peptide-specific LC correction, γ'_p which in the implemented LC rate expression these terms act as exponential scaling factors,

$$k^{(p)}_{\mathrm{LC},i}(v) = k_{\mathrm{base},i}(1-v)^{\eta}\exp\left(\gamma_{\mathrm{LC}}v + \gamma'_p\right),$$

where v describes the LC solvent-gradient coordinate. The ESI stage was described using an effective time parameter t_{ESI} and the charge-density scaling parameter λ_{ESI}. The remaining ESI parameters were fixed at $\alpha_{\mathrm{ESI}} = 1$ and $\xi_{\mathrm{ESI}} = 1$.

The labeling-time offset τ_{label} was fixed to zero in the calculations reported here, and it should therefore not be interpreted as a parameter independently estimated by the optimization. Likewise, t_{ESI} is an effective fitting parameter in the kinetic model and should not be interpreted directly as a physical source residence time.

Table A.2: Optimized global and peptide-specific parameters for all solution conditions. The parameter t_{ESI} is reported in the numerical units used by the kinetic implementation and is treated as an effective parameter rather than a directly measured residence time.

		ANP			Angiotensin			Bradykinin		
Condition	γ_{LC}	γ'_p	t_{ESI}	λ_{ESI}	γ'_p	t_{ESI}	λ_{ESI}	γ'_p	t_{ESI}	λ_{ESI}
0 mM	20	-0.419	2.110	6.486	1.959	1207.430	-2.873	0.176	5.033	10.000
150 mM NaCl	15	1.000	0.238	8.581	3.165	869.869	-3.093	1.506	100.456	-0.945
150 mM CsCl	15	1.083	0.143	10.000	3.213	1147.771	-3.798	1.616	150.693	-2.286
500 mM NaCl	20	-0.289	0.032	10.000	1.997	1621.539	-2.625	0.275	99.524	1.760
500 mM CsCl	15	1.456	48.151	10.000	3.484	1324.453	-3.811	1.845	182.703	-0.926
1 M NaCl	20	-0.441	0.047	9.956	1.863	775.894	-2.884	-0.030	178.370	-2.287
1 M CsCl	15	1.476	55.079	10.000	3.578	1524.721	-3.666	1.927	271.807	-2.569

The optimized γ_{LC} values took only the values 15 and 20 over the investigated conditions and did not show a simple monotonic dependence on salt concentration. More importantly, several local parameters approached the imposed optimization bounds. For ANP, λ_{ESI} reached the upper bound of 10 under 150 mM CsCl, 500 mM NaCl, 500 mM CsCl, and 1 M CsCl, and was very close to that bound at 1 M NaCl. The salt-free Bradykinin fit also reached $\lambda_{\mathrm{ESI}} = 10$. These boundary solutions indicate that the local parameters are not equally well identified in all conditions. Consequently, individual optimized values are used primarily to describe the effective correction required by the model and are not assigned a unique microscopic interpretation.

A.3 Stage-by-Stage Deuterium Retention

To quantify where deuterium is lost after the initial labeling period, the model output was decomposed into labeling, quench, LC, and ESI stages. For a given observation the stage losses were defined as

$$\Delta D_{\text{quench}} = D_{\text{label}} - D_{\text{quench}}, \tag{A.1}$$

$$\Delta D_{\text{LC}} = D_{\text{quench}} - D_{\text{LC}}, \tag{A.2}$$

$$\Delta D_{\text{ESI}} = D_{\text{LC}} - D_{\text{ESI}}. \tag{A.3}$$

The values reported below are averages over labeling times and available charge states for each peptide and solution condition.

Because identical intrinsic quench-rate vectors were used for all salt conditions, the mean modeled quench loss was independent of salt: 0.648 Da for ANP, 0.412 Da for Angiotensin, and 0.059 Da for Bradykinin. The salt-dependent part of the stage analysis is therefore summarized through the LC and ESI contributions in Table A.3.

Table A.3: Mean modeled deuterium loss during the LC and ESI stages. All values are in Da.

	ANP		Angiotensin		Bradykinin	
Condition	LC	ESI	LC	ESI	LC	ESI
0 mM	2.007	0.348	1.436	0.148	0.883	0.092
150 mM NaCl	1.876	0.308	1.323	0.111	0.687	0.064
150 mM CsCl	1.966	0.490	1.352	0.123	0.752	0.082
500 mM NaCl	2.212	0.188	1.472	0.200	0.987	0.078
500 mM CsCl	2.527	2.290	1.554	0.123	0.933	0.105
1 M NaCl	1.973	0.253	1.380	0.100	0.753	0.097
1 M CsCl	2.585	2.371	1.612	0.139	1.013	0.129

For most peptide-condition combinations, the LC stage constituted the largest modeled source of deuterium loss. This is already apparent in the salt-free reference condition, where the mean LC losses were 2.007 Da for ANP, 1.436 Da for Angiotensin, and 0.883 Da for Bradykinin. The ESI contribution was substantially smaller in these reference measurements.

A different pattern emerged for ANP in concentrated CsCl, and the effective ESI loss increased to 2.290 Da at 500 mM CsCl and 2.371 Da at 1 M CsCl, becoming comparable to the LC loss itself. This increase should not be interpreted as direct experimental evidence that the physical electrospray process removes this amount of deuterium, and rather, it identifies the ESI term as the part of the present parameterization that absorbs a large fraction of the unexplained condition-dependent discrepancy. This interpretation is consistent with the optimization-bound behavior discussed in Section A.2.

A.4 Model Performance Across Peptides and Salt Conditions

Model performance was evaluated from the residual

$$r_j = D_{\mathrm{model},j} - D_{\mathrm{exp},j}.$$

The root-mean-square error (RMSE) and mean bias were calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} r_j^2}, \tag{A.4}$$

$$\mathrm{Bias} = \frac{1}{N}\sum_{j=1}^{N} r_j. \tag{A.5}$$

A negative bias therefore denotes systematic underprediction of the measured deuterium uptake.

Table A.4 gives the two most informative condition-level metrics, and additional row-level residual statistics were used during analysis but are omitted here because they do not change the interpretation of the model performance.

Table A.4: Model performance summarized by RMSE and mean bias. All values are in Da, and mean bias is defined as model minus experiment.

	ANP		Angiotensin		Bradykinin	
Condition	RMSE	Mean Bias	RMSE	Mean Bias	RMSE	Mean Bias
0 mM	0.463	0.237	0.125	0.017	0.059	-0.007
150 mM NaCl	0.201	-0.003	0.156	0.033	0.105	0.019
150 mM CsCl	0.238	0.067	0.186	0.036	0.148	0.027
500 mM NaCl	0.392	0.256	0.122	-0.026	0.069	-0.001
500 mM CsCl	1.666	-1.618	0.165	0.031	0.133	0.022
1 M NaCl	0.223	0.116	0.159	0.033	0.185	0.034
1 M CsCl	1.568	-1.473	0.213	0.054	0.224	0.044

The fits for Angiotensin and Bradykinin remained comparatively stable over the entire salt range, and their RMSE values did not exceed 0.224 Da in any condition. ANP showed a more condition-dependent response, so the model reproduced the 150 mM NaCl, 150 mM CsCl, and 1 M NaCl data with RMSE values of approximately 0.20-0.24 Da, while the salt-free and 500 mM NaCl conditions showed somewhat larger deviations.

The dominant failure occurred specifically for ANP in concentrated CsCl. The RMSE increased to 1.666 Da at 500 mM CsCl and 1.568 Da at 1 M CsCl, also, the corresponding mean biases, -1.618 and -1.473 Da, show that the model systematically underpredicted the measured uptake in these two conditions.

The high-CsCl discrepancy is particularly informative because the temporal shape of the ANP kinetics remained strongly correlated with the experimental

data, and at 500 mM CsCl and 1 M CsCl the Pearson correlation coefficients were approximately 0.990 in both cases, whereas the corresponding coefficients of determination were only $R^2 = 0.275$ and $R^2 = 0.481$, respectively. Thus, the model retained much of the time-dependent trend while failing to reproduce the absolute deuterium level, so this behavior is consistent with a systematic offset or missing condition-dependent contribution rather than a complete failure to describe the kinetic ordering of the measurements.

A.5 Code Availability

The computational implementation used for the multi-step HDX-MS modelling and analysis presented in this thesis is available in the following public GitHub repository:

https://github.com/R0oo0zbe/hdx-ms-multi-step-model



Bibliography

[1] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. "The Energy Landscapes and Motions of Proteins". In: *Science* 254.5038 (1991), pp. 1598–1603. DOI: 10.1126/science.1749933.

[2] Katherine Henzler-Wildman and Dorothee Kern. "Dynamic Personalities of Proteins". In: *Nature* 450.7172 (2007), pp. 964–972. DOI: 10.1038/nature06522.

[3] Ivet Bahar, Timothy R. Lezon, Lee-Wei Yang, and Eran Eyal. "Global Dynamics of Proteins: Bridging Between Structure and Function". In: *Annual Review of Biophysics* 39 (2010), pp. 23–42. DOI: 10.1146/annurev.biophys.093008.131258.

[4] Yoshitomo Hamuro. "Tutorial: Chemistry of Hydrogen/Deuterium Exchange Mass Spectrometry". In: *Journal of the American Society for Mass Spectrometry* 32 (2021), pp. 133–151. DOI: 10.1021/jasms.0c00260.

[5] Michele Stofella, Antonio Grimaldi, Jochem H. Smit, Jürgen Claesen, Emanuele Paci, and Frank Sobott. "Computational Tools for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis". In: *Chemical Reviews* 124 (2024), pp. 12242–12263. DOI: 10.1021/acs.chemrev.4c00438.

[6] Michele Stofella, Neeleema Seetaloo, Alexander N. St John, Emanuele Paci, Jonathan J. Phillips, and Frank Sobott. "Recalibrating Protection Factors Using Millisecond Hydrogen/Deuterium Exchange Mass Spectrometry". In: *Analytical Chemistry* 97 (2025), pp. 2648–2657. DOI: 10.1021/acs.analchem.4c03631.

[7] Monika Kish, Victoria Smith, Natasha Lethbridge, Lindsay Cole, Nicholas J. Bond, and Jonathan J. Phillips. "Online Fully Automated System for Hydrogen/Deuterium-Exchange Mass Spectrometry with Millisecond Time Resolution". In: *Analytical Chemistry* 95 (2023), pp. 5000–5008. DOI: 10.1021/acs.analchem.2c05310.

[8] Lars Konermann and Pablo M. Scrosati. "Hydrogen/Deuterium Exchange Mass Spectrometry: Fundamentals, Limitations, and Opportunities". In: *Molecular & Cellular Proteomics* 23.11 (2024), p. 100853. DOI: 10.1016/j.mcpro.2024.100853.

[9] Irina Oganesyan, Cristina Lento, and Derek J. Wilson. "Contemporary Hydrogen Deuterium Exchange Mass Spectrometry". In: *Methods* 144 (2018), pp. 27–42. DOI: 10.1016/j.ymeth.2018.04.023.

[10] Antonio Grimaldi, Billy Hobbs, Michele Stofella, Theodoros K. Karamanos, and Emanuele Paci. "Amide Hydrogen-Deuterium Exchange in Isotopically Mixed Water". In: *ACS Physical Chemistry Au* 6.3 (2026), pp. 446–450. DOI: 10.1021/acsphyschemau.5c00127.

[11] Antonio Grimaldi, Michele Stofella, and Emanuele Paci. "Intrinsic Hydrogen-Deuterium Exchange Rates in H_2O/D_2O Mixtures". In: *The Journal of Physical Chemistry B* 130.9 (2026), pp. 2493–2500. DOI: 10.1021/acs.jpcb.5c06636.

[12] David D. Boehr, Ruth Nussinov, and Peter E. Wright. "The Role of Dynamic Conformational Ensembles in Biomolecular Recognition". In: *Nature Chemical Biology* 5.11 (2009), pp. 789–796. DOI: 10.1038/nchembio.232.

[13] Hesam N. Motlagh, James O. Wrabl, Jing Li, and Vincent J. Hilser. "The Ensemble Nature of Allostery". In: *Nature* 508.7496 (2014), pp. 331–339. DOI: 10.1038/nature13001.

[14] Peter E. Wright and H. Jane Dyson. "Intrinsically Disordered Proteins in Cellular Signalling and Regulation". In: *Nature Reviews Molecular Cell Biology* 16.1 (2015), pp. 18–29. DOI: 10.1038/nrm3920.

[15] Joan J. Englander, Jose R. Rogero, and S. Walter Englander. "Protein Hydrogen Exchange Studied by the Fragment Separation Method". In: *Analytical Biochemistry* 147.1 (1985), pp. 234–244. DOI: 10.1016/0003-2697(85)90033-8.

[16] Yawen Bai, John S. Milne, Leland Mayne, and S. Walter Englander. "Primary Structure Effects on Peptide Group Hydrogen Exchange". In: *Proteins: Structure, Function, and Genetics* 17.1 (1993), pp. 75–86.

[17] David Nguyen, Leland Mayne, Michael C. Phillips, and S. Walter Englander. "Reference Parameters for Protein Hydrogen Exchange Rates". In: *Journal of the American Society for Mass Spectrometry* 29.9 (2018), pp. 1936–1939. DOI: 10.1007/s13361-018-2021-z.

[18] Aase Hvidt and Sigurd O. Nielsen. "Hydrogen Exchange in Proteins". In: *Advances in Protein Chemistry* 21 (1966), pp. 287–386. DOI: 10.1016/S0065-3233(08)60129-1.

[19] S. Walter Englander and Neville R. Kallenbach. "Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids". In: *Quarterly Reviews of Biophysics* 16.4 (1983), pp. 521–655. DOI: 10.1017/S0033583500005217.

[20] Ellie I. James, Taylor A. Murphree, Clint Vorauer, John R. Engen, and Miklos Guttman. "Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems". In: *Chemical Reviews* 122.8 (2022), pp. 7562–7623. DOI: 10.1021/acs.chemrev.1c00279.

[21] David L. Smith, Yuzhong Deng, and Zhongqi Zhang. “Probing the Non-Covalent Structure of Proteins by Amide Hydrogen Exchange and Mass Spectrometry”. In: *Journal of Mass Spectrometry* 32.2 (1997), pp. 135–146. DOI: 10.1002/(SICI)1096-9888(199702)32:2<135::AID-JMS486>3.0.CO;2-M.

[22] Yawen Bai, Joan J. Englander, Leland Mayne, John S. Milne, and S. Walter Englander. “Thermodynamic Parameters from Hydrogen Exchange Measurements”. In: *Methods in Enzymology*. Vol. 259. Academic Press, 1995, pp. 344–356.

[23] John R. Engen and Thomas E. Wales. “Analytical Aspects of Hydrogen Exchange Mass Spectrometry”. In: *Annual Review of Analytical Chemistry* 8 (2015), pp. 127–148. DOI: 10.1146/annurev-anchem-062011-143113.

[24] Joey G. Sheff, Martial Rey, and David C. Schriemer. “Peptide-Column Interactions and Their Influence on Back Exchange Rates in Hydrogen/Deuterium Exchange-MS”. In: *Journal of the American Society for Mass Spectrometry* 24.7 (2013), pp. 1006–1015. DOI: 10.1007/s13361-013-0639-4.

[25] Lars Konermann, Elias Ahadi, Antony D. Rodriguez, and Siavash Vahidi. “Unraveling the Mechanism of Electrospray Ionization”. In: *Analytical Chemistry* 85.1 (2013), pp. 2–9. DOI: 10.1021/ac302789c.

[26] Gregory P. Connelly, Yawen Bai, Mei-Fen Jeng, and S. Walter Englander. “Isotope Effects in Peptide Group Hydrogen Exchange”. In: *Proteins: Structure, Function, and Genetics* 17.1 (1993), pp. 87–92. DOI: 10.1002/prot.340170111.

[27] Glenn R. Masson, John E. Burke, Natalie G. Ahn, Ganesh S. Anand, Christoph Borchers, Sébastien Brier, George M. Bou-Assaf, John R. Engen, S. Walter Englander, Johan Faber, Rachel Garlish, Patrick R. Griffin, Michael L. Gross, Miklos Guttman, Yoshitomo Hamuro, Albert J. R. Heck, Damian Houde, Roxana E. Iacob, Thomas J. D. Jørgensen, Igor A. Kaltashov, Judith P. Klinman, Lars Konermann, Petr Man, Leland Mayne, Bruce D. Pascal, Dana Reichmann, Mark Skehel, Joost Snijder, Timothy S. Strutzenberg, Eric S. Underbakke, Cornelia Wagner, Thomas E. Wales, Benjamin T. Walters, David D. Weis, Derek J. Wilson, Patrick L. Wintrode, Zhongqi Zhang, Jie Zheng, David C. Schriemer, and Kasper D. Rand. “Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments”. In: *Nature Methods* 16.7 (2019), pp. 595–602. DOI: 10.1038/s41592-019-0459-y.

[28] Kasper D. Rand, Steven D. Pringle, James P. III Murphy, Keith E. Fadgen, Jeff Brown, and John R. Engen. “Gas-Phase Hydrogen/Deuterium Exchange in a Travelling Wave Ion Guide for the Examination of Protein Conformations”. In: *Analytical Chemistry* 81.24 (2009), pp. 10019–10028. DOI: 10.1021/ac901897x.

[29] Lars Konermann and D. J. Douglas. “Unfolding of Proteins Monitored by Electrospray Ionization Mass Spectrometry: A Comparison of Positive and Negative Ion Modes”. In: *Journal of the American Society for Mass Spectrometry* 9 (1998), pp. 1248–1254.

[30] Kasper D. Rand, Steven D. Pringle, Michael Morris, and Jeffery M. Brown. “Site-Specific Analysis of Gas-Phase Hydrogen/Deuterium Exchange of Peptides and Proteins by Electron Transfer Dissociation”. In: *Analytical Chemistry* 84.4 (2012), pp. 1931–1940. DOI: 10.1021/ac202918j.

[31] Charles C. Mundorff, Sarah Hadley, Lisa M. Tuttle, Yuqi Shi, Graeme C. McAlister, Rosa Viner, Rachel E. Klevit, and Miklos Guttman. “Mapping Hydrogen Migration Thresholds for Site-Specific HDX-MS”. In: *Molecular & Cellular Proteomics* 24.11 (2025), p. 101075. DOI: 10.1016/j.mcpro.2025.101075.

[32] Benjamin T. Walters, Alec Ricciuti, Leland Mayne, and S. Walter Englander. “Minimizing Back Exchange in the Hydrogen Exchange-Mass Spectrometry Experiment”. In: *Journal of the American Society for Mass Spectrometry* 23.12 (2012), pp. 2132–2139. DOI: 10.1007/s13361-012-0476-x.