Towards Pharmaceutical Protein Stabilization:
Thermodynamic and Kinetic Property Database Generation

by

Katherine E. Lawson

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy, Chemical Engineering

Auburn, Alabama
August 5, 2023

Keywords: Chemical Reaction Engineering, DFT, Pharmaceutical Protein Stabilization,
Protein Degradation Kinetic and Thermodynamic Property Database Generation

Copyright 2023 by Katherine Elizabeth Lawson

Approved by

Dr. Andrew J. Adamczyk, Chair, Assistant Professor,
Department of Chemical Engineering
Dr. Elizabeth Lipke, Mary and John H. Sanders Professor,
Department of Chemical Engineering
Dr. Robert Pantazes, Assistant Professor, Department of Chemical Engineering
Dr. Robert Arnold, Professor, Department of Drug Discovery and Development
Dr. Alexey Petrov, Assistant Professor, Department of Biological Sciences
Dr. Steven Mansoorabadi, J. Milton Harris Associate Professor,
Department of Chemistry and Biochemistry

Abstract

Studies on peptide degradation mechanisms are of interest because of the high demand for protein-based therapeutics in the pharmaceutical industry. However, formulation efficiency is limited by the degradation of the desired products during multiple stages of the pharmaceutical protein manufacturing pipeline. While the use of additives and stabilizers has significantly reduced degradation, determining the most effective additives and stabilizers is accomplished through a "guess and check" approach, instead of understanding the underlying mechanism of the degradation pathway. Therefore, we sought to achieve a fundamental understanding of specific degradation mechanisms present in pharmaceutical proteins. Through the use of density functional theory (DFT), we examined peptide degradation at different points of the pharmaceutical protein manufacturing pipeline to generate the thermodynamic and kinetic property databases associated with the degradation mechanism of interest. Additionally, we classified the degradation mechanisms as either 'functional' or 'structural' based on the type of impact the reaction of interest had on therapeutic proteins. The propensity of the protein to be impacted by these forms of degradation was found to be dependent on the stage of the manufacturing pipeline, where fragmentation was more likely to occur than deamidation during downstream manufacturing and both deamidation and fragmentation occurred under ambient conditions. Furthermore, the propensity of the succinimide reaction pathway to occur under ambient environmental conditions indicated that the screening of cations to inhibit the reaction pathway was only applicable to the specified reaction pathway.

Acknowledgments

First, I would like to thank my advisor Dr. Andrew J. Adamczyk for his continued guidance and support throughout the course of my tenure as a member of his research group. In addition to his advice and guidance regarding the most impactful direction for my research, Dr. Adamczyk has inspired me to grow as a researcher and as a young professional. I am grateful for all of our discussions, for they have allowed me to see connections I might not have noted otherwise, and fully explore the diverse applications my research might have.

Next, I would like to thank my esteemed advisory committee for their guidance and suggestions during my studies. Dr. Elizabeth Lipke, Dr. Robert Pantazes, Dr. Robert Arnold, and Dr. Alexey Petrov are all well-established within the biomedical field, and I am fortunate to have the opportunity to learn from them. Additionally, I would like to thank Dr. Steven Mansoorabadi for acting as my dissertation reader.

As a strong community fosters success in research, I would like to acknowledge my groupmates: Dr. Ashraf Ali, Dr. Yeseul Choi, Dr. Tanzina Azad, and Dr. Siyuan Wu. The diversity of projects within our group has encouraged an environment of communication and mutual learning due to the interconnected nature of the multi-scale modeling methodologies our group implements. I would also like to thank my undergraduate researchers Joseph K. Dekle and Megan N. Evans for their dedication and hard work.

Additionally, I would like to acknowledge the three supercomputers our research group uses, the Alabama Super Computer (ASC), Auburn University Easley Cluster, and Auburn University Hopper Cluster, through which I was able to achieve the results presented in this study. I would also like to acknowledge the Department of Graduate Assistanceship in Areas of National Need (GAANN) fellowship, which funded three years of my degree and provided me with the opportunity to improve my communication abilities by guest lecturing and teaching recitations to undergraduate students. Furthermore, I have been fortunate to have published some of the works detailed in this study. The work detailed in Chapter 2 was reproduced from Ref 82 with permission from the Royal Society of Chemistry. The work detailed in Chapter 3 was reproduced from Ref 147 with permission from the American Chemical Society. The work detailed in Chapter 4 was reproduced from Ref 141 with permission from Elsevier.

Next, I would like to acknowledge the Auburn University Department of Chemical Engineering for all of their guidance and support during my time here as a graduate student. I am genuinely grateful for everything Elaine, Emma, and Naomi do to help our department.

I would also like to acknowledge Dr. Jeffrey Halpern of the University of New Hampshire for acting as a mentor to me, even after I was no longer his student. Without Dr. Halpern's encouragement, I would not have spoken to the Auburn University table at the 2017 AIChE Graduate School fair and would not have been able to obtain the experiences I did by pursuing my Ph.D. here.

As I have previously mentioned, I have been fortunate to experience the community that Auburn fosters and for that I am grateful to have my family and friends as a strong support system. My parents Bill and Debbie for all of their support throughout the course of my studies as a Ph.D. student in chemical engineering. My older brother Billy, his fiancée Tori, and our grandparents Anita, Oscar, and Alice for their encouragement and constant support. Last, but certainly not least, I would like to thank my friends for acting as the Auburn family during the course of my Ph.D.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

E_A	Activation Energy
Ala	Alanine
Arg	Arginine
Ã	Arrhenius pre-exponential factor
Asn	Asparagine
Asp	Aspartic Acid
BLA	Biologics License Application
BPCIA	Biologics Price Competition and Innovation Act
B3LYP	Becke 3-Parameter, Lee-Yang-Parr Functional
$\Delta G_{BINDING}$	Binding free energy
Ca(II)	Calcium
CHO	Chinese Hamster Ovary Cell
Co(II)	Cobalt (+2 oxidation state)
Co(III)	Cobalt (+3 oxidation state)
CAT	Committee for Advanced Therapies
HMPC	Committee on Herbal Medicinal Products
CHMP	Committee for Medicinal Products for Human Use
COMP	Committee for Orphan Medicinal Products
CVMP	The Committee for Veterinary Medicinal Products
CPCM	Conductor-like Polarizable Continuum Model
CQA	Critical Quality Attribute
Cys	Cysteine

ε	Dielectric
DFT	Density Functional Theory
ΔE	Energy Change
ΔH_{RXN}	Enthalpy of Reaction
EMA	European Medicines Agency
EP	Evans-Polanyi Correlation
FDA	Food and Drug Administration
$\hat{H}$	Hamiltonian Operator
k^F	Forward Rate Coefficient
ΔG_{RXN}	Gibbs Free Energy of Reaction
ΔG_{SF}	Gibbs Free Energy of Cation-Dipeptide Complex Succinimide Formation
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
GD3	Grimme D3 Empirical Dispersion Correction
His	Histidine
Fe(II)	Iron (+2 oxidation state)
Fe(III)	Iron (+3 oxidation state)
Ile	Isoleucine
K	Kelvin
Leu	Leucine
Lys	Lysine
Mg(II)	Magnesium

Mn(II)	Manganese (+2 oxidation state)
Mn(III)	Manganese (+3 oxidation state)
Met	Methionine
Ni(II)	Nickel (+2 oxidation state)
NMR	Nuclear Magnetic Spectroscopy
PDCO	The Paediatric Committee
ρ	Pearson Correlation Coefficient
PRAC	The Pharmacovigilance Risk Assessment Committee
Phe	Phenylalanine
PBS	Phosphate Buffer Saline
PES	Potential Energy Surface
PCA	Principal Component Analysis
Pro	Proline
PHSA	Public Health Service Act
RPD	Reaction Pathway Degeneracy
Ser	Serine
SGP	Same geometric plane
Thr	Threonine
TGE	Transient Gene Expression
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Ψ	Wave function

Zn(II) Zinc (+2 oxidation state)

Chapter 1: Introduction

1.1. The Impact of Pharmaceutical Proteins on the Global Biotechnology Market

The pharmaceutical protein market accounts for a significant portion of the global biotechnology market due to the diverse array of applications that the therapeutics can be used for; ranging from cancer treatments to acting to regulate functions within the body.[1, 2] A recent review by de la Torre and Albericio indicated that 15 therapeutic proteins were given federal approval in 2022, where nine of the approved therapeutics were monoclonal antibodies (mAbs).[2, 3] Accumulating a quarter of the current therapeutic market, biotechnology companies worldwide have increased the number of different monoclonal antibodies (mAbs) to clinical use significantly, where 27 mAbs were approved for the global market between 2015 and 2017.[4-7] The drastic increase in mAb production is likely due to recent advancements in understanding regarding disease states so the developed formulation is more specific towards disease targets, as well a wide range of applications within the biomedical industry.[4, 8, 9] This diversity is due to variations in the amino acid sequence of the complementary antigen-binding sites, which target specific epitopes and prevent antigen release.[10, 11] Figure 1.1.1. depicts the yearly number of novel therapeutics containing mAbs approved for by the Food and Drug Administration (FDA), relative to other therapeutics, since 2019.[3, 12-15]

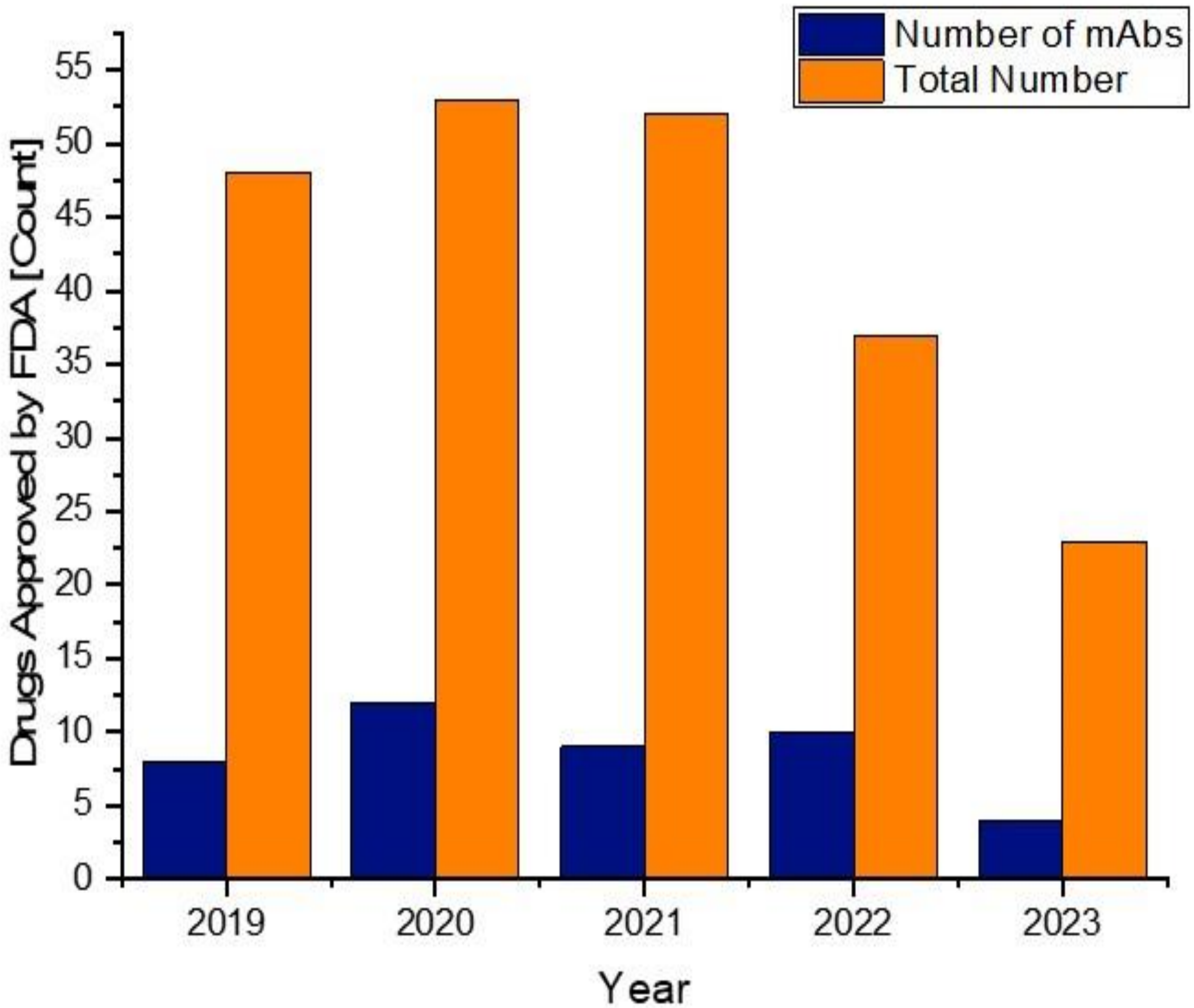

Figure 1.1.1. Bar chart of the number of novel monoclonal antibodies (mAb)-based therapeutics approved by the FDA per year relative to the total number of therapeutics approved. The total number of mAbs are the blue bars, while the total number therapeutics, including mAbs, are the orange bars.[3, 12-15]

1.2. Federal and Global Guidelines

There are multiple regulatory checkpoints that the pharmaceutical protein must pass through before the product may be brought to market. In the United States, the FDA is the leading authority on ensuring all food and drugs brought to a consumer market are safe.[16] Through the Pure Food and Drugs Act (est. 1906), the FDA validates all novel biotechnologies which enter the consumer market.[16] Similar to the FDA, the European Medicines Agency (EMA) is a regulatory agency that polices the quality of products with biological applications, and generated a network that connects medical regulations to the economic impact of the proposed product.[17] Through this coordination, the EMA has established seven advisory committees to determine the safety and efficacy of the proposed product.[17] These seven advisory committees are the Committee for Advanced Therapies (CAT), the Committee for Medicinal Products for Human Use (CHMP), the Committee on Herbal Medicinal Products (HMPC), the Committee for Orphan Medicinal Products (COMP), The Paediatric Committee (PDCO), the Pharmacovigilance Risk Assessment Committee (PRAC), and the Committee for Veterinary Medicinal Products (CVMP), which are directly involved with the development and regulation of medicine.[17-26] Of the seven committees, the CAT, CHMP, COMP, PDCO, and PRAC are of direct relevance to the works cited within this document.[19, 20, 22, 24, 25] The CHMP and PRAC work closely together to ensure that the adverse effects of approved therapeutics are reduced, while the PDCO evaluates if the proposed therapeutic can be administered to individuals between 0 and 17 years in age.[19, 24, 25] Additionally, the approval of advance therapeutics, new scientific developments, and the application of therapeutics for rare diseases via an orphan designation are accounted for by the CAT and the COMP.[20, 22] Given that the works cited in this document were generated to impact both the formulation and manufacturing of pharmaceutical proteins, specifically mAbs, it is likely that EMA approval would be contingent on these five committees.

For biologically derived products, such as pharmaceutical proteins, to become FDA-approved, they must account for the following stipulations. Via the Public Health Service Act (PHSA), all biologics must be licensed via the 351a pathway or the Biologics License Application (BLA).[27-30] Through this regulatory process and the multiple pre-clinical and clinical studies required, the approved biologics are deemed safe for human use and are effective against the given ailment.[28, 29, 31, 32] While Congress has amended for the expedition of generic chemicals found in pharmaceuticals, all biologics must pass each stipulation provided by the FDA, unless following the Biologics Price Competition and Innovation Act (BPCIA).[29, 31] First implemented by the EMA, the BPCIA allows the filing corporation to expedite approval if the proposed biologic is derived from a previously approved one.[33] However, if the biosimilar is provided this accelerated path towards approval, it must wait an additional 12 years to provide exclusivity to the already approved biologic it is similar to.[34, 35] For the development of pharmaceutical proteins, the EMA, FDA, and the World Health Organization (WHO) provide guidelines regarding both the current good manufacturing process (cGMP) and the clinical trials associated with the developed product, so that the approved product will be safe and effective when brought to market.[36-38]

1.3 The Pharmaceutical Protein Manufacturing Pipeline

The pharmaceutical protein manufacturing pipeline can be separated into five stages: protein engineering, upstream manufacturing, downstream manufacturing, the inclusion of additives and stabilizers to the formulation, and storage. Each stage represents a different aspect of the formulation, where the protein engineering step focuses on the design of the protein of interest, while the storage stage is dedicated to maintaining the stability and efficacy of the developed formulation. During the manufacturing stages (Figure 1.3.1), the formulation is exposed to perturbations within its environment that can impact its shelf-life, therefore it is necessary to understand each stage of the manufacturing pipeline and how it contributes to protein stability and efficacy. In the studies cited within this document, we evaluated the impact of the environmental changes observed during the downstream manufacturing, inclusion of additives and stabilizers, and storage stages of pharmaceutical protein manufacturing on the thermodynamic and kinetic properties associated with two prevalent protein degradation reaction pathways.

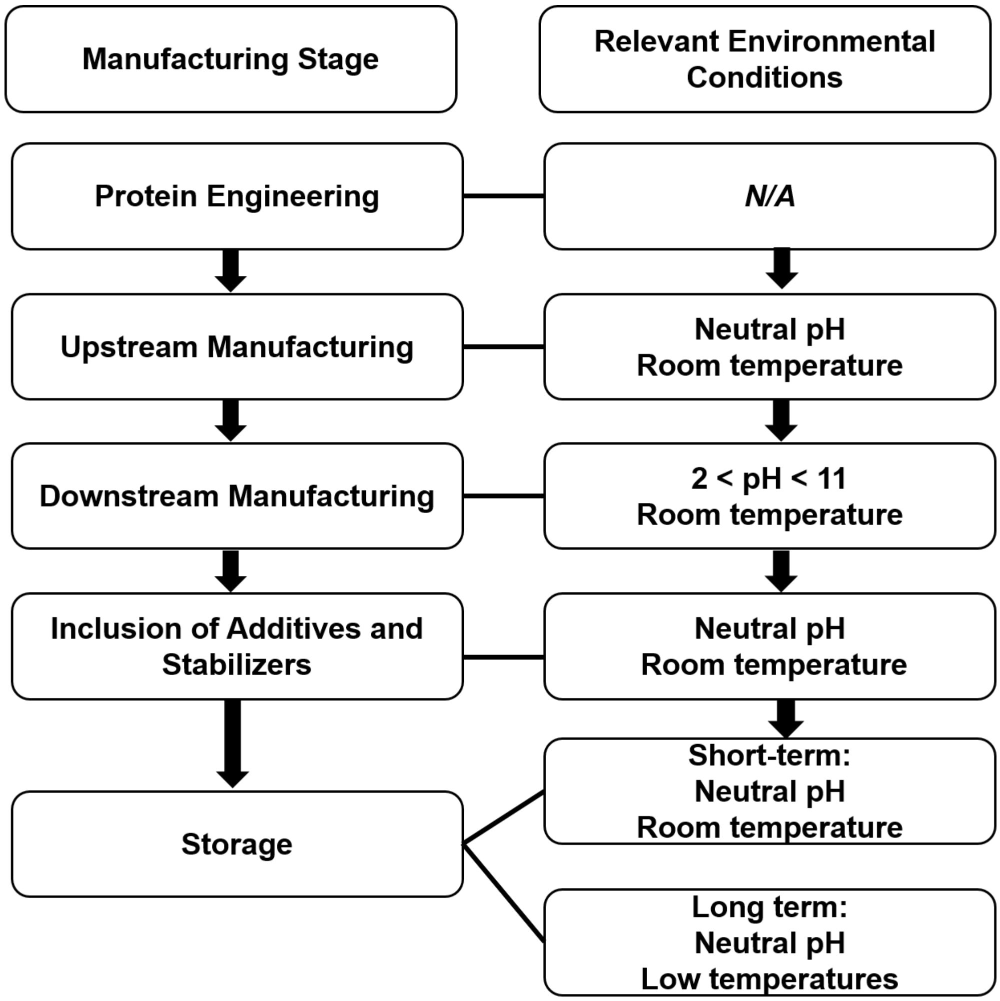

Figure 1.3.1. Flow diagram of the pharmaceutical protein manufacturing pipeline with the associated pH and temperature ranges.

1.3.i. *De Novo* Protein Engineering

The first stage of the pharmaceutical protein manufacturing pipeline is the design of the protein of interest. Approaches towards protein engineering include rational design, directed evolution, and computational approaches such as *de novo* design.[39-42] Directed evolution focuses on the development of protein mutants within a host, where screening of the generated mutants is applied to determine if the mutant possesses the desired functionality.[40-42] Described by Romero and Arnold as a landscape, where "regions of higher elevation represent more optimized proteins", directed evolution enables the expansion of protein functionality past the traits observed by wild type proteins.[40] However, the number of potential mutations available within the landscape makes characterization difficult.[40] Similar to directed evolution, rational design is an experimental technique that engineers the proteins of interest through the generation of a mutation library.[42] However, unlike directed evolution, the mutation site is limited to the sequence that maintains the "structure-function" relationship of interest.[43, 44] Therefore, to achieve fundamental understanding of the "structure-function relationship" associated with the protein of interest, protein engineering via *de novo* design is applied.[42] *De novo* design enables the evaluation of the protein design landscape through rational design, where target scaffolds can be generated by database motifs and used to rank the generated variants in regards of their stability and functionality.[43, 45] Once the functional proteins identified by *de novo* protein design have been validated through experimental methods, the sequences can be prepared for manufacturing of the therapeutic protein of interest.

1.3.ii. Upstream Manufacturing

Upstream manufacturing of pharmaceutical proteins focuses on the transfection of DNA and expression of the protein of interest within a host cell (Table 1.3.1). Optimization of the cell line used as a host for protein expression is contingent on multiple factors, where the growth rate of the cell and how well it can express the desired protein are among the design parameters.[46] For small scale production, transient gene expression (TGE) can be used to express proteins as single cell cultures via DNA transfer. The use of TGE has been noted by previous studies as an effective approach for developing pharmaceutical protein concentrations between 80 mg/L and 300 mg/L in mammalian Chinese Hamster Ovary (CHO) cell cultures, where limited studies are available for larger concentrations.[46-49] CHO cells are a common cell line used in pharmaceutical protein manufacturing for transfection processes involving both stable and transient gene expression.[46-50] The use of CHO cells as the popular cell line of interest is due to its high growth rate, adaptability within culture mediums, and high expression of proteins.[49, 51, 52] Expression of the proteins of interest on the cell is completed via the use of an expression vector, where the primary criteria for construction are maintenance of expression efficiency, independent expression from the genome site of integration, and correlated expression to the quantity of integrated genetic material.[50, 53] Once the cells have been transfected, they can then be seeded to generate the cell culture fed into the bioreactor to produce the desired proteins of interest.[51, 54]

The reactors associated with the mammalian cell culture of the therapeutic proteins include batch, continuous, fed-batch and perfusion-based systems.[46, 51, 55] Of the four types of reactors associated with this stage of the upstream manufacturing process, the perfusion-based system is the preferred method because it maintains the viable cell culture by removing waste and circulating nutrients.[55, 56] Furthermore, perfusion cell cultures can be used as a regulatory approach towards maintaining a consistent cell volume throughout the upstream manufacturing process.[57] Reactor scale up is contingent on multiple design parameters, where properties such as shear stress derived from mixing, carbon dioxide removal, and oxygen transfer are critical to maintain cell viability.[54, 58] Within mammalian cell cultures, proteins are secreted intracellularly and the cells are then lysed to form a lysate that can be used for harvesting.[59-61] The harvesting of the mAb has been classified as either the last step of upstream manufacturing or the first step of downstream manufacturing, where harvesting is accomplished via centrifugation and filtration to remove cells and debris to yield a harvested cell culture fluid that can be further purified during downstream manufacturing.[46, 52, 56]

Table 1.3.1. Processes associated with upstream manufacturing of pharmaceutical proteins.[46-48, 50-53, 55-57, 62, 63]

Process		Role
Host Cell Optimization	**Mammalian Cells**	Host cells through which proteins can be expressed
Protein Engineering	**TGE**	DNA transfer into single-cell cultures to produce recombinant proteins under short time durations
	Expression Vector	Enable production of mAbs at high and consistent rate
Mammalian Cell Culture	**Batch**	All essential nutrients in base medium
	Continuous	Nutrient feed and product culture removal is a continuous process
	Fed-batch	Essential nutrients fed into system
	Perfusion	Medium circulated through growing culture to supply nutrients and remove waste
Reactor scale-up		Takes cell-culture from bench scale to industrial scale
Centrifugation and Filtration		Remove cells and debris to yield harvested cell culture

1.3.iii. Downstream Manufacturing

Downstream manufacturing focuses on the extraction and purification of the protein of interest from the host cell (Table 1.3.2). Protein A chromatography is a popular and highly selective tool used to extract the expressed protein from its host cell culture.[64-66] Operating under an acidic pH, Protein A chromatography extracts the protein of interest from its host cell via charge-charge interactions.[65, 66] Specifically, by weakening the hydrophobic attraction exhibited by the protein and its host cell.[65, 66] Purification during Protein A chromatography, is achieved by adding an alkaline wash step to the chromatography process, where the protein is washed with an alkaline buffer prior to its extraction at an acidic pH.[67, 68] By including an alkaline wash stage to the Protein A chromatography workflow, irreversible aggregation via the chromatography is reduced and impurities are removed.[67, 68] Viral inactivation occurs between the extraction and secondary polishing stages associated with the downstream manufacturing stage of the pharmaceutical protein manufacturing pipeline, and is implemented to reduce the viral concentration generated from cell lines in accordance with federal guidelines.[55, 66, 69, 70]

After the protein is extracted from the host cell, purification and polishing is applied to the formulation to remove impurities from the system. Two types of approaches have been noted as practical for removing impurities from the generated formulation: chromatography and caprylic acid precipitation.[46, 66, 71-77] The use of chromatography methods such as anion-exchange, cation-exchange, ion-exchange, hydrophobic interaction, hydroxyapatite, and size exclusion are common for removing impurities and leached Protein A from the formulation.[46, 66] The chromatography approach used is representative of its desired function, where hydroxyapatite chromatography is primarily used to remove leached Protein A and cell and protein impurities are removed via multiple different types of chromatography.[66] Caprylic acid precipitation is an alternative approach to protein polishing where the impurities are extracted from the formulation instead of the protein of interest.[71, 72] Additionally, polishing via caprylic acid has been found to stabilize proteins against denaturation and inactivate viruses.[73-76] While polishing and purification of pharmaceutical proteins via caprylic acid precipitation occurs under acidic conditions, a wider pH range is associated with the different types of chromatography. Filtration is the final type of purification, where the type of filtration applied is in reference to the desired outcome.[66, 76, 78-80]

For example, ultrafiltration is technique used to transform the system pH from acidic to the neutral pH associated with the final formulation, while viral filtration removes viral impurities.[66, 76, 79]

Table 1.3.2. Associated roles of protein extraction and purification techniques included in downstream manufacturing of pharmaceutical proteins.[55, 65-68, 70-76, 78, 81, 82]

Process			Role
Extraction	Protein A chromatography		Extract protein from host batch cell
Viral inactivation			Reduce virus concentration to meet product safety standards
Polishing/ Purification	Alkaline wash		Remove cellular and DNA impurities and prevent irreversible aggregation post-Protein A chromatography
	Caprylic Acid Precipitation		Remove aggregates and host cell proteins
	Chromatography	Anion-exchange	Remove cell, protein, and viral impurities
		Cation-exchange	Remove aggregate, cell, protein, and viral impurities; Remove leached Protein A
		Hydrophobic Interaction	Remove aggregate, cell, protein, and viral impurities; Remove leached Protein A
		Hydroxyapatite	Remove DNA and viral impurities; Remove leached Protein A
		Ion-exchange	Remove aggregate, cell, protein, and viral impurities
		Size Exclusion	Remove aggregates and degradation products
Filtration	Absolute		Remove small impurities
	Ultra/Dia		Adjust for higher protein concentration; Exchange purification buffer to formulation buffer
	Viral		Remove viral impurities

1.3.iv. Inclusion of Additives and Stabilizers

Additives and stabilizers are included within pharmaceutical protein formulations to improve the shelf-life stability and efficacy of the generated product during storage. Classification of the additives and stabilizers (Table 1.3.2) is representative of the type of molecule included within the formulation, as well as its application towards improving the stability and efficacy of the developed product. Buffers are a common additive included within the developed formulation and act as a pH stabilizer to transition the formulation from the acidic and basic pH ranges associated with downstream manufacturing to the neutral pH of the final formulation, as well as reduce the electrostatic interactions between the solution and the residues on the protein surface.[83-92] Similarly, polymers are used to reduce protein self-interaction and the formation of aggregates within the formulation.[93-96] Conformational stability of the developed protein can be enhanced by residue-based additives and polyols, while sugars such as sucrose and dextrose are implemented as cryoprotectants during long term storage.[77, 83, 93-95, 97-103] The development of degradation-specific additives and stabilizers is an ongoing process, and can be supplemented by critical understanding regarding of the thermodynamic and kinetic parameters exhibited by the degradation reaction of interest.[92, 103, 104]

Table 1.3.3. Additive and Stabilizer Classifications associated with pharmaceutical protein formulations.[77, 83-85, 93-95, 97-106]

Classification	Use(s)
Residue-Based Additive	Reduce aggregation; Increase thermal stability of protein; Increase conformational stability of protein
Buffer	pH stabilizer
Cation-based Additive	Inhibit deamidation via cyclic intermediate reaction pathway
Cation-based Salt	Inhibit deamidation via cyclic intermediate reaction pathway
Polymer	Surface coating to inhibit undesirable protein interactions
Polyol	Enhance conformational stability of protein;
Sugar	Cryoprotectant for lyophilization; Reduce extent of denaturation during unfolding
Surfactant	Minimize aggregation

1.3.v. Storage

The storage approaches applied to therapeutic protein formulations are representative of tradeoffs exhibited by each approach. For example, storage under ambient conditions is attractive because the formulation is ready for immediate use, but its stability and efficacy is limited.[107-111] A patent filed by Lam (Document ID: US 20050158316 A1) indicated that the antibody formulation in their patent would have a shelf-life of about one month at temperatures around 30 ºC, while other studies at room temperature (T ~ 22 – 25 ºC) have indicated a significantly shorter shelf-life before chemical instability was observed within the formulation.[107-110] As a result, storage under ambient conditions would only be beneficial for formulations designed for immediate use. The current approach towards long-term storage of pharmaceutical proteins is through lyophilization, where the formulation is a freeze-dried solid and stored under low temperatures.[110-115] The lyophilized formulations shelf-lives have been found to be significantly longer than at room temperature.[116-119] However, prior studies have indicated that the stresses exhibited by the process reduce the long-term stability of the developed formulation.[120-127] As we previously noted, additives and stabilizers have been developed to enhance the efficacy and stability of the developed therapeutics, but their efficiency is dependent on critical understanding of the types of degradation that the protein is exposed to both during the manufacturing and storage stages. For this purpose, studies on protein degradation are of interest.

1.4. Functional versus Structural Degradation

Nowak et al. published a study in 2017 regarding the impact of forced degradation studies on the therapeutic protein market, where the insights derived could be applied to advance the field.[128] By understanding the prevalent protein degradation pathways that impact the therapeutic proteins, the methodologies associated with the pharmaceutical protein manufacturing pipeline can be optimized to reduce the observed degradation. Specifically, understanding the environmental factors and protein sequences which trigger a specific form of protein degradation can be applied to evaluation of the stages where the protein is exposed to the environment of interest. As a result, additives and stabilizers can be formulated to reduce the quantity of degradation observed at the specific point of the manufacturing pipeline. Furthermore, fundamental understanding of the type of degradation observed at specific areas of the protein can be applied to the expansion of the additive and stabilizer design space.

The degradation reactions observed in mAbs are commonly classified as either physical or chemical degradation, where physical degradation is representative of degradation within the protein structure that is not chemically induced, and chemical degradation causes a loss of stability and efficacy due to covalent modifications within the protein structure.[129-131] Aggregation, deamidation, and fragmentation are three types of protein degradation that have been found to impact both mAbs and physiological proteins.[132-134] Degradation within physiological proteins, which have been noted by prior studies to be highly stable under physiological conditions *in vivo*, can often lead to neurodegenerative diseases, while degradation within mAbs results in a loss of stability and function.[132-134]

Prior studies have identified aggregation, deamidation, and fragmentation as critical quality attributes (CQA) within mAb formulation and manufacturing due to the impact of the degradation mechanisms on the therapeutic efficacy, stability and safety for use *in vivo.* Specifically, aggregation and fragmentation have been noted to increase cytotoxicity and immunogenicity, while deamidation leads reduced efficacy due to charge heterogeneity.[135-137] For this purpose, quantification and characterization of the properties influencing these mechanisms of protein degradation is crucial for mAb formulation and manufacturing. Of the three degradation reactions, aggregation can occur as a result of either deamidation or peptide bond fragmentation; where a study by Alam et al noted that aggregation via deamidation was due to the promotion of "nonreducible cross-linked species", while a study by Perico et al indicated that the formation of fragments increased the rate of aggregate formation.[133, 134, 138-140] While the thermodynamic and kinetic parameters associated with aggregation are beyond the scope of this work, evaluation of the energetics associated non-enzymatic deamidation and peptide bond fragmentation yields valuable insights towards the CQAs associated with the formulation and manufacturing of mAbs.

To better understand the behavior of the deamidation and fragmentation reaction pathways, as well as the regions of the mAb they directly impact, we have classified the chemical degradation pathways evaluated within this work as either structural or functional degradation (Figure 1.4.1). We attributed structural degradation to the reaction pathways where the primary factor influencing the overall effectiveness of the protein was the stability of the polypeptide backbone. Specifically, where the degradation pathway of interest caused fragmentation within the backbone of the polypeptide chain.[128, 141-146] Similarly, functional degradation was attributed to the reaction pathways where the degradation-induced mutation of specific residues within the protein caused a loss of efficacy towards the epitope of interest.[80, 82, 147-155]

Non-enzymatic peptide bond fragmentation, where the solvated environment attacks the polypeptide chain to cleave the central amide bond, is an example of structural degradation.[128, 141-145] Furthermore, multiple studies have identified the hinge region, the flexible section of the pharmaceutical protein that allows it to adopt different conformations, to be the area of the protein most significantly impacted.[138, 145, 156-158] Deamidation, a form of degradation that converts net neutral Asparagine (Asn) and Glutamine (Gln) into their acidic counterparts Aspartic Acid (Asp) and Glutamic Acid (Glu), has been noted by multiple studies to decrease the efficacy of therapeutic proteins.[80, 82, 147-155] For this purpose, we sought to evaluate how these forms of functional and structural degradation were impacted by the changing environmental conditions associated with different point of the pharmaceutical protein manufacturing pipeline.

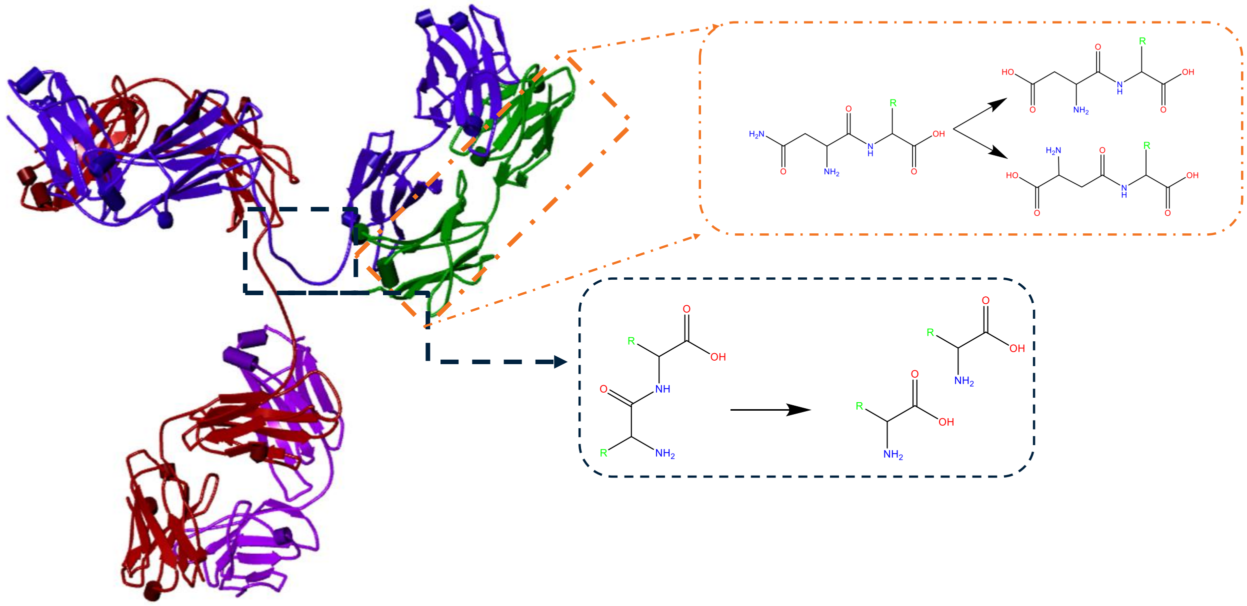

Figure 1.4.1. Graphical representation of the two types of protein degradation, structural and functional, analyzed in this work and the areas of the protein that they impact (PDB: 1IGT).[128, 159] Green chain represents Chain C, blue chain represents Chain D, purple chain represents Chain B, and red chain represents Chain A.

1.5. The Most Prevalent Amino Acid Residues in Pharmaceutical Proteins

The frequency of the 20 amino acids to occur within proteins cited by literature, ranges from computational and experimental analysis of mAbs (Flores, Gentiluomo, Ohta), to Globular and Transmembrane β-Barrel Proteins (TMBs).[160-163] Our evaluation of literature studies which defined the amino acid composition within the protein of interest indicated that residues with sulfuric sidechains, Cysteine (Cys) and Methionine (Met), were found to occur infrequently regardless of the protein evaluated.[160-163] Similarly, small residues such as Alanine (Ala), Glycine (Gly), and Serine (Ser) were found to be among the most frequently occurring residues observed within the protein of interest.[160-164] However, evaluation of the different types of protein degradation observed throughout the pharmaceutical protein manufacturing pipeline cannot be solely based on the limited studies that include the composition of the residues used to form the protein.

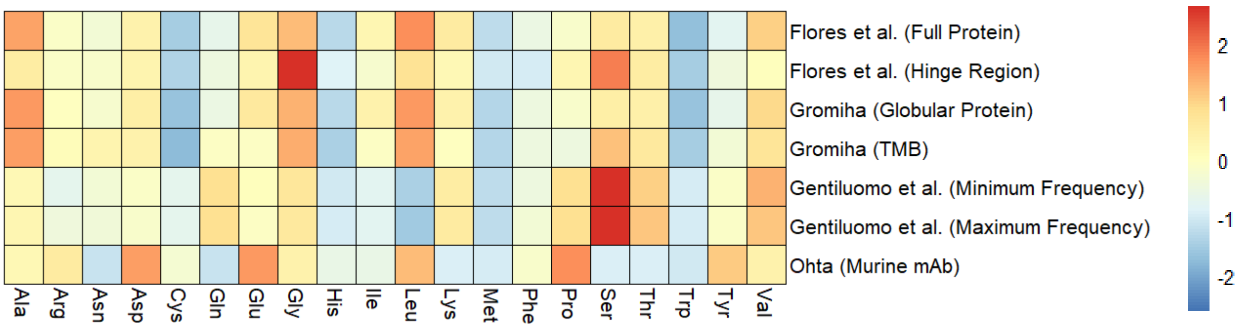

Figure 1.5.1. Generated heatmap of the composition of the 20 amino acid residues as denoted by literature protein structures.[160-163] Dark blue represents the residues occurring at the lowest frequency within the literature protein structure, while dark red represents the residues occurring at the highest frequency within the literature protein structure.

For example, the propensity of amino acids to occur within proteins that was tabulated through a bioinformatics study completed by Flores et al. was determined by analyzing a dataset of 1000 morphs (homologous protein structures), while the values given by Ohta were derived experimentally through an automated amino acid analyzer in reference to the protein of interest (Murine mAb).[162, 163] As a result, there was significant discrepancy observed between the amino acid compositions, where residues such as Lysine (Lys), Ser and Thr were found to occur at high frequencies according to Flores, but occurred at significantly lower frequencies in the Ohta study.[162, 163] Additionally, Proline (Pro) was found to occur at a much higher frequency by Gentiluomo (mAb) and Ohta (Murine mAb) relative to the findings of Flores (mAb) and Gromiha (Globular and TMB).[160-163] Therefore, it is necessary to model a diverse array of sidechain functionalities within the degradation studies to quantify their impact on reaction energetics. Automation of the X-ray crystallography structures to the RCSB Protein Data Bank (PDB), as well as the sequencing data associated with therapeutic protein development, would be necessary to generate the amino acid composition distributions as a function of the protein classification of interest, but is beyond the scope of this study.

1.6. Computational Approaches towards Improving Pharmaceutical Protein Stability

Previous studies have noted that the critical parameters associated with the different protein degradation reaction pathways are tedious to quantify and require insights towards the impact of structural effects on the reaction of interest.[144, 165-167] Individually, structure-based prediction of deamidation "hotspots" is achieved through half-life analysis based on model polypeptides, where hotspot prediction is based off the energetics observed in the modeled polypeptides.[151, 153, 167-169] Additionally, evaluation of the impact of the degradation reaction of interest on long-term formulation stability is either a time-consuming process or one that does not provide an accurate representation of the protein degradation due to the state of stress that the formulation is placed under during the accelerated degradation studies.[166, 170] For this purpose, computational models can be developed to predict the thermochemical and kinetic parameters associated with the degradation reaction of interest. Additionally, through the development of robust reaction classes, the impact of sidechain functionalities, electrostatic interactions, and the modeled environment can be correlated to the degradation reaction of interest. Furthermore, the energetics of specific degradation reactions can be generalized through the development of models based on the parameters associated with the reaction. As a result, fundamental insights regarding protein degradation can be derived and used to stabilize mAbs.

1.6.i. Quantum Chemistry and its Applicability to Protein Stability Studies

The electronic structures and energies associated with the structures used to model the different reactions associated with protein degradation can be achieved by solving the Schrödinger equation. First theorized in the 1920s and derived from Newtonian mechanics, the Schrödinger equation considers an electron a particle of mass in an external field.[171-173] To obtain the energy of a system, the Schrödinger equation is the baseline for computing the electronic structure of a given molecule (Equations 1.6.1).[171, 172]

$$\hat{H}\Psi = E\Psi \qquad (1.6.1)$$

Where $\hat{H}$ represents the Hamiltonian operator, a sum of the kinetic and potential energies within the system (Equation 1.6.2), E is the energy of the system, and Ψ is the wave function.[172]

$$\hat{H} = -\frac{\hbar}{2m}\nabla^2 + V(r,t) \qquad (1.6.2)$$

Where $-\frac{\hbar}{2m}\nabla^2$ is the kinetic energy operator and V(r,t) is the potential energy operator as a function of space and time.[172] For time-independent models, the potential energy operator is solely a function of space. For a many body problem, i.e. a model which has multiple particles, the Hamiltonian operator becomes Equation 1.6.3.[172, 174]

$$\sum_{i=1}^{N} \hat{H}_I \qquad (1.6.3)$$

As the hydrogen atom is the largest system through which the Schrödinger equation can be solved, multiple assumptions and approximations are required to obtain the electronic structures associated with the reaction of interest.[172] One such approach is the Born-Oppenheimer approximation, which assumes that the velocity of the electrons in a system is infinitely faster than that of the nuclei.[172, 175, 176] Via this approximation, the motion of the electrons and the nuclei within a system can be separated to calculate its energy via the electronic Schrödinger equation (Figure 1.6.4).[172, 175, 176]

$$\hat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \qquad (1.6.4)$$

Where $\hat{H}_{el}$ is the electronic Hamiltonian operator, Ψ_{el} is the electronic wavefunction, and E_{el} is the energy of the electrons within the system. Through this approximation, the electronic Hamiltonian becomes equation 1.6.5.[172, 175, 176]

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i} \nabla^2 - \sum_{j}^{n}\sum_{k}^{e}\frac{Zj}{\mathrm{Rjk}} + \sum_{l}^{e}\sum_{m>l}^{e}\frac{1}{Rlm} \tag{1.6.5}$$

Where $-\frac{1}{2}\sum_{i} \nabla^2$ is the kinetic energy operator, $\sum_{j}^{n}\sum_{k}^{e}\frac{Zj}{\mathrm{Rjk}}$ is nuclear attraction, and $\sum_{l}^{e}\sum_{m>l}^{e}\frac{1}{Rlm}$ is electron repulsion.[172, 175, 176]

Evaluation of the electronic structures associated with the reaction of interest is completed through their optimization. Specifically, the reactant and product structures associated with the reaction of interest are referred to as minima on the potential energy surface (PES), where the second derivatives associated with the optimization of the structures are positive.[172] Once the reactant structure has been optimized, it can be used to locate the transition state, which is represented as a first-order saddle point along the PES (second derivative is negative in one direction and positive in all others).[172]

Density Functional Theory (DFT) is a popular tool used to derive the electronic structures associated with modeled reactions due to its low computational cost and high accuracy relative to other *ab initio* approaches. Known as the "exact solution to the approximate Schrödinger equation, DFT assumes that the density of all molecules within a given system can be treated as a continuous charge density (1.6.6) and is reliant on two Hohenberg-Kohn theorems.[172, 177-180]

$$\rho(\vec{r}) = N\int d\vec{r}_2\, d\vec{r}_3 \ldots d\vec{r}_N \tag{1.6.6}$$

Where N is the number of electrons, $\vec{r}$ is the position of the electron in space, and $\rho(\vec{r})$ is the charge density of the system.[172]

The first Hohenberg- Kohn theorem states that the ground state energy of a system can be expressed as a functional of its electron density (Equation 1.6.7), while the second theorem states that the electron density which minimizes the functional is representative of the ground state structure (1.6.8).[172, 177, 181]

$$E_0 = E_0[\rho] \tag{1.6.7}$$

$$_{\min}(E[\rho]) \tag{1.6.8}$$

Minimization of the ground state functional is achieved through the use of Kohn-Sham (KS) orbitals (1.6.9).[172, 177, 181]

$$\left(-\frac{1}{2}\nabla_i^2 - \sum_\alpha \frac{Z\alpha}{|\overrightarrow{r1}-\overrightarrow{R\alpha}|} + \int \frac{\rho(\overrightarrow{r2})}{|\vec{r}1-\vec{r}2|}\,d\vec{r}_2 + V_{XC}(\vec{r})\right)\varphi_i^{KS}(\overrightarrow{r1}) = \varepsilon_i^{KS}\,\varphi_i^{KS}(\overrightarrow{r1}) \tag{1.6.9}$$

Where $-\frac{1}{2}\nabla_i^2 - \sum_\alpha \frac{Z\alpha}{|\overrightarrow{r1}-\overrightarrow{R\alpha}|}$ is the electronic Hamiltonian, $\int \frac{\rho(\overrightarrow{r2})}{|\vec{r}1-\vec{r}2|}$ is the Coulomb energy integral, and the exchange energy is given by Equation 1.6.10.[172, 177, 181]

$$V_{XC} = \frac{dExc}{d\rho} \tag{1.6.10}$$

Computation of the exchange correlation between electrons is dependent on the electron density functional selected.[172, 177, 181] The Becke-3-Parameter, Lee-Yang-Parr (B3LYP) functional is commonly used for protein studies and has been found to produce reliable results.[182-184] Known as a hybrid generalized gradient approximation (hybrid GGA), B3LYP includes one-third of the *ab initio* exchange energy (Equation 1.6.11).

$$E_{XC}^{B3LYP} = dE_X^{HF} + (1-\alpha)E_X^{B} + E_C^{LYP} \tag{1.6.11}$$

Where α is equal to one-third the exchange energy of the *ab initio* method Hartree Fock (HF), dE_X^{HF} is the derivative of the HF exchange energy, and E_X^B and E_C^{LYP} are the exchange and correlation energies of the GGA.[172, 177, 181, 185, 186] One of the limitations of the B3LYP functional is its ability to account for non-bonded interactions, such as van der Waals interactions, within the modeled system. For this purpose, the inclusion of an empirical dispersion correction, such as the Grimme D3 (GD3) empirical dispersion correction is included within the modeled level of theory.[187-190]

1.6.ii. Application of Statistical Thermodynamics and the Arrhenius Equation to Compute Critical Thermodynamic and Kinetic Parameters

Once the electronic structures of the reactant, transition state, and product structures associated with each modeled reaction have been optimized, critical reaction properties can be quantified using partition functions. To evaluate the thermodynamic and kinetic parameters associated with the reaction of interest, the microstates associated within the system are averaged to form a macrostate through the Boltzmann distribution law.[191-194] Known as the sum of probabilities that all particles within a system will exist in a particular arrangement, the Boltzmann distribution provides a relationship between the probability of a system's energetics and its temperature (Equation 1.6.12).[191, 192, 194, 195]

$$p_i = \frac{e^{-\frac{Ej}{kT}}}{\sum_j^m e^{-\frac{Ej}{kT}}} \qquad (1.6.12)$$

Where p_i denotes the probability of the system being in a microstate (j).[191, 192, 194, 195] In the numerator, E_j is the energy of each microstate present within the system, k is the Boltzmann constant (1.38 x 10^{-23} J/K), and T is the temperature of the system.[191, 192, 194, 195] The denominator is denoted by the canonical partition function (Q), where the energies are denoted as a function of the system's microstate (j to m).[191, 192, 194, 195] For a multi-body system, the canonical partition function is further reduced to only be dependent on the volume and temperature of the system, forming the molecular partition function (Equation 1.6.13).[191, 192]

$$q(V,T) = \sum_{i}^{m} e^{-\frac{\varepsilon i}{kT}} \qquad (1.6.13)$$

Where ε_i is a set of molecular energy states, k is the Boltzmann constant, and T is the temperature of the system.[191, 192] To analyze different contributions to the overall molecular energy, the molecular partition function can be specified by the type of energy analyzed (Equation 1.6.14).[191, 192]

$$q(V,T) = q_{translational} q_{rotational}\, q_{vibrational}\, q_{electronic} \qquad (1.6.14)$$

Where $q_{translational}$ is the partition function of translational energy states, $q_{rotational}$ is the partition function of rotational energy states, $q_{vibrational}$ is the partition function of vibrational energy states, and $q_{electronic}$ is the partition function of electronic energy states.[191, 192] As we previously mentioned, these partition functions are used to derive the energetics of the system, which in turn are used to compute key thermodynamic and kinetic properties such as the rate coefficients asssociated with the specified reaction.

Quantification of the rate coefficient yields critical insights towards the reaction analysis, such as definition of the direction of the reaction and its order. In this study, only first-order (unimolecular) and second-order (bimolecular) reactions are observed. Furthermore, the symmetry of the molecules used to define the structures evaluated in the forward and reverse rate coefficients is critical, because it prevents over counting observed in the molecular partition function by only attributing unique conformations to the rotational partition function.[196] Equation 1.6.15 defines the reaction path degeneracy, where the product of the chiral number and symmetry number associated with the reactant(s) (σ_R) is the numerator and the product of the chiral number and symmetry number associated with the reactant (σ_R).

$$rpd^F = \frac{\sigma TS * nchiral(TS)}{\sigma R * nchiral(R)} \qquad (1.6.15)$$

One approach towards determining rate coefficients within a kinetic model is conventional transition state theory (TST). TST is derived through the molecular partition function and assumes the reactant and transition state structures within a system are in quasi-equilibrium (Equation 1.6.16).[191, 192, 196, 197]

$$A + B \rightleftharpoons [AB^{\ddagger}] \longrightarrow P \tag{1.6.16}$$

Where A and B are the reactants, $[AB^{\ddagger}]$ is the transition state, and P represents the products of the given reaction. In addition to being in a pseudo-steady state with either the reactants or the products in a system, TST also assumes that the energy of all molecules within the system follow a Boltzmann distribution.[191, 192, 197] Through the assumptions used by TST, the partition functions of the transition state can be computed to yield the rate constant, which in turn can be used to compute the system activation barrier and Arrhenius pre-exponential through Arrhenius parameterization.

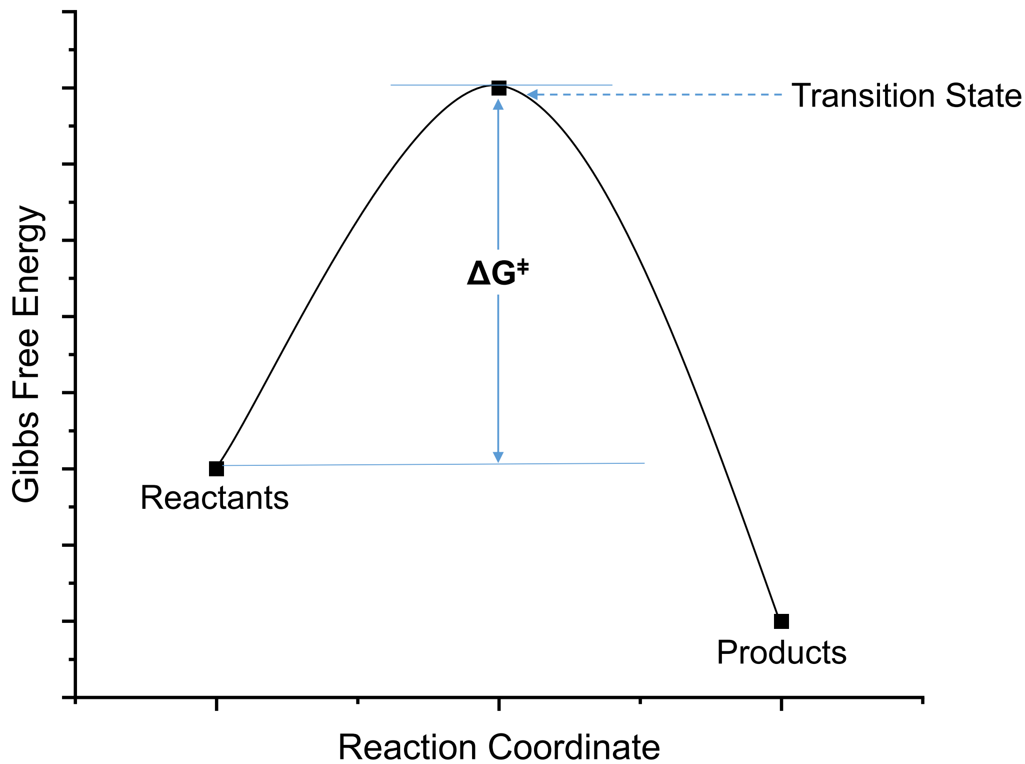

Figure 1.6.1. Graphical representation of TST.

Arrhenius parameterization is a common form of predicting reaction rates as a function of temperature. Specifically, due to its ability to generate a linear correlation between the natural log of the reaction rate ($\ln(k^F)$) and reciprocal temperature ($\frac{1}{T}$) (Equation 1.6.17).[198-200] Through this linear correlation, important metrics such as the molecular collision rate (A) and the activation barrier (E_A) can be quantified.[198-200] Quantification of these two values then provides insight towards the activation enthalpy and entropy of the system.[201, 202] While the activation barrier provides insight towards the enthalpy of the system, specifically the thermal energy required to shift the system from its reactants to its products, the collision rate indicates the average amount of disorder present when the reactants are converted to the transition state.[201, 202]

$$k = Ae^{-\frac{Ea}{RT}} \quad (1.6.17)$$

Where A is the collision rate or frequency factor of molecules, E_A is the energy necessary for the reaction to shift from the reactants to the products, R is the gas constant and T is the temperature.[198, 199] Analysis of if a chemical reaction follows Arrhenius behavior can be quantified by plotting ln(k) against the inverse temperature ($\frac{1}{T}$).[198-200, 203] If a linear trend line is present, the chemical reaction exhibits Arrhenius behavior, and the activation barrier (E_A) and the Arrhenius pre-exponential factor (log(A)) are defined as the slope and y-intercept, respectively.[198-200, 203]

1.6.iii. Chemical Model Development and Applied Solvation Models

For chemical systems where an aqueous system must be modeled, there are two types of solvation methods: implicit solvation and explicit solvation. Also known as the continuum solvation method, implicit solvation is a popular model due to the reduction of degrees of freedom present within the system.[204, 205] This reduction is completed by treating pre-determined aspects of the system as a continuum, i.e. an averaging of solvent properties to solve the Poisson distribution.[204] Explicit solvation is a more complex process that is typically used through either a Monte Carlo or a hybrid Quantum Mechanics/Molecular Dynamics (QM/MM) approach, where the solvent geometries are included in the model parameterization.[206, 207] Parameterization is achieved by first computing solvation energetics at the explicit QM level.[206] Through this calculation, the contributions of implicit and explicit solvation energetics are defined.[206] Once parameter definition is achieved, the MM force field is scaled to account for the explicit solvent molecules, allowing for the inclusion of solvation effects in the MM model.[206]

The definition of the chemical model is critical to accurately evaluate the reactions associated with protein degradation due to the tradeoff between model accuracy and complexity. For this purpose, the models developed in this study were developed using a baseline implicit solvation method, where the inclusion of an explicit water molecule to apply a hybrid solvation methodology was applied to improve the accuracy of the generated chemical models.[82, 141, 147] The conductor-like polarizable continuum model (CPCM) is popular in biological systems due to the distribution of polarization charges along the surface of the solute cavity, i.e. a continuous dielectric constant instead of a finite value.[179, 208, 209] Through this distribution, non-electrostatic interactions are taken into consideration, thus improving the computed electrostatic energy gradients.[208] Furthermore, modeling of the reactions under a variety of dielectrics can be achieved. While the dielectrics associated with the different regions of the protein, as well as the dielectrics associated with the different residue functional groups, have been studied extensively, the inclusion of a range of dielectrics to model the system is a recent advancement.[210-220]

A study conducted by Li et al specified how the electrostatic interactions exhibited by the modeled system are indicative of the sequence of amino acids and structure of the modeled region.[212] Specifically, recent modeling studies that analyze the energetics of regions of the proteins containing nonpolar residues or the energetics of reactions in the hydrophobic core are modeled using a relatively low dielectric constant ($\varepsilon = 4$).[212, 213, 221, 222] Similarly, surface protein studies, or studies that contain charged or polar residues, will have higher dielectric constants; $\varepsilon = 20$ for charged residues, and $\varepsilon = 40$ for polar residues, respectively.[212, 213, 216, 223] As hydrophobic, charged, and polar residues were used to generate the dipeptides modeled within our system, we determined it was necessary to analyze how the electrostatic interactions of the different residue functionalities influenced their reaction behavior. Therefore, we modeled the degradation reactions analyzed in this study under four different dielectrics: the three dielectrics associated with hydrophobic, charged, and polar residues, as well as the dielectric associated with pure water (ε = 80).[82, 213] By modeling the dielectrics associated with the different residue functional groups, and different sites within the mAbs, we were able to obtain an unbiased representation of how the behavior of these functionalities influences reaction energetics.[82, 141, 147]

1.6.iv. Application of Supervised and Unsupervised Learning Methods for Model Generalization

The model dipeptide is the smallest structure that can be used to evaluate the impact of sidechain functionalities on protein degradation. However, there are 400 possible dipeptide combinations that can be formulated. While certain forms of degradation are limited to specific residues, generalization of the reaction pathways where the full distribution of dipeptide combinations is possible is necessary. For this purpose, supervised and unsupervised learning methods can be applied to the degradation reaction of interest. An example of unsupervised learning is the Evans-Polanyi correlation, where the linear trend used to generate the model is based on the principle that more exothermic reactions will have lower activation energies, and model sensitivity analysis is achieved through the random removal and replacement of data points.[141, 224-232] Supervised learning methods, such as Principal Component Analysis (PCA), can be applied to reaction pathways where the parameter contributions to variability are not immediately recognizable or the identification of multiple significant parameters can be implemented to improve model linearity.[233-237]

1.6.v. Project Overview

The works presented in this study applied chemical engineering thermodynamic and reaction engineering principles to analyze how the changing environment of the pharmaceutical protein manufacturing pipeline influences structural and functional degradation. Using model dipeptides, we have identified points within the pharmaceutical manufacturing pipeline where one form of degradation would be more significantly impacted than the other, as well as areas where both forms of degradation are involved. Through this, we have built thermochemical and kinetic libraries using quantum chemistry and statistical thermodynamics to generate reaction networks, analyzed the competition between reaction pathways, identified linear correlations between thermochemical and kinetic parameters, correlated protein region propensity explored the impact of the additive and stabilizer design space to inhibit protein degradation, and correlated the propensity of specific regions towards degradation to the types of residues present within the area of interest (Figure 1.6.2).

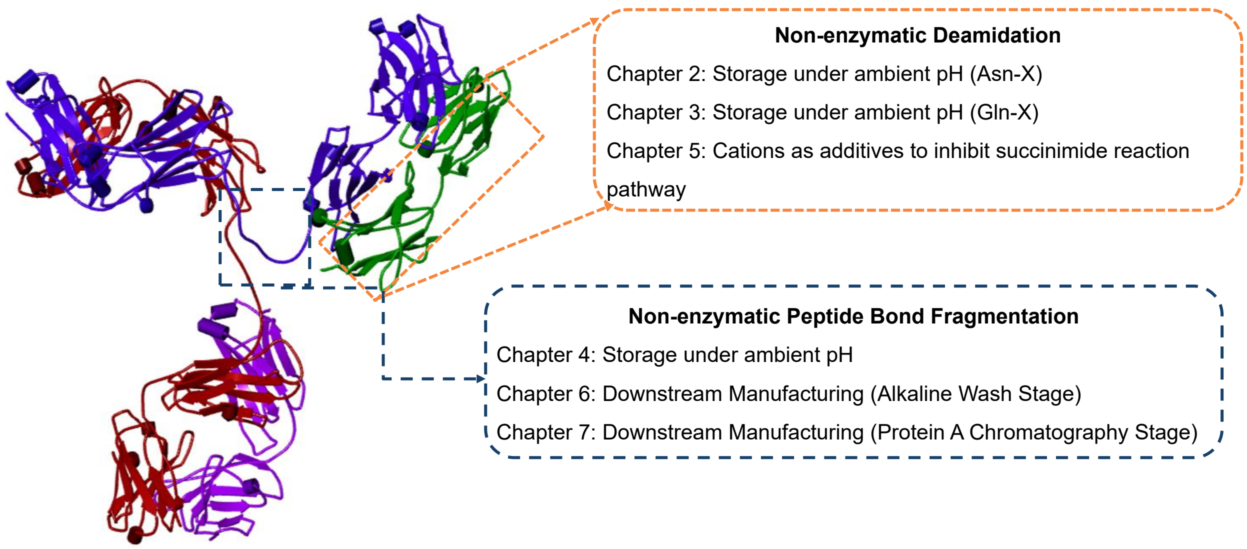

Figure 1.6.2. Graphical representation of the structural and functional protein degradation databases generated at different points of pharmaceutical protein manufacturing pipeline (PDB: 1IGT).[159] Green chain represents Chain C, blue chain represents Chain D, purple chain represents Chain B, and red chain represents Chain A.

Chapter 2: Towards Pharmaceutical Protein Stabilization: Mapping the Reaction Network of Asn-X Deamidation during Storage under Ambient Conditions

2.1. Background and Motivation

Peptide deamidation studies are of interest due to their impact on the formulation and efficacy of mAbs. Known as one of the major pathways for covalent peptide degradation, deamidation is more selective than other peptide degradation mechanisms precisely because it only impacts four of the 20 essential amino acids – Asparagine (Asn), Glutamine (Gln), Aspartic Acid (Asp) and Glutamic Acid (Glu).[80, 82, 128, 147, 149, 150, 155, 169, 184, 238, 239] Proteins which undergo deamidation are either converted from Asn to Asp or from Gln to Glu.[80, 82, 128, 147, 149, 150, 155, 169, 184, 238, 239] In addition to the selective conversion of the residues, deamidation is separated into two reaction networks due to its impact on mAbs.[82, 128, 147] Specifically, Asn deamidation is more readily observed than Gln deamidation within mAbs, where deamidation within the variable region of the protein results in a loss of efficacy (Figure 2.1.1).[82, 128, 129, 147, 149, 240]

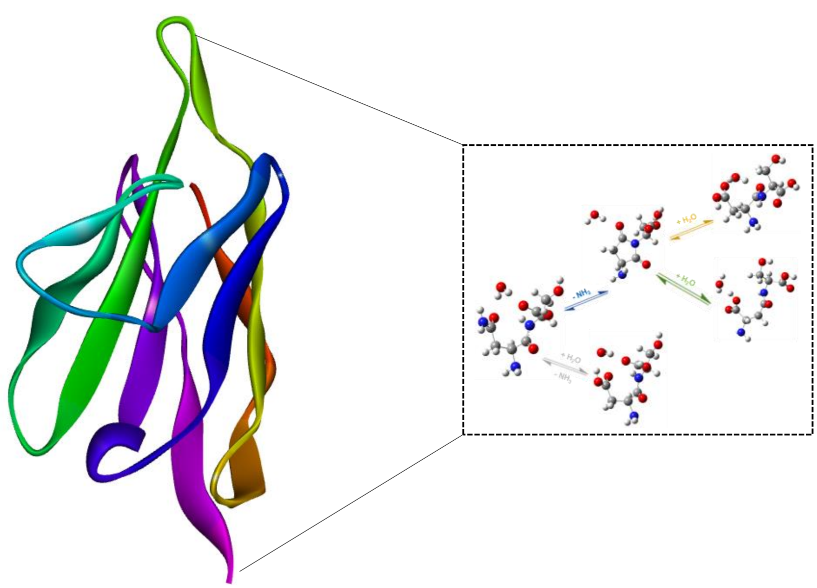

Figure 2.1.1. Graphical representation of the impact of Asn-X deamidation on the variable region of therapeutic proteins (PDB: 1LVE).[241]

Deamidation primarily occurs through catalysis via an acid.[80, 150, 155, 184] While less common, deamidation can also occur non-catalytically. Under ambient conditions, such as storage under ambient conditions, where the formulation is in its liquid phase, deamidation occurs through the elimination of Ammonia, where the neutral side chains of Asn and Gln residues are transformed into the acidic Asp and Glu amino acids, respectively.[80, 82, 128, 147, 149, 150, 155, 169, 184, 238] Similar to other peptide degradation mechanisms, experimental examination of peptide deamidation is well-established but tedious and difficult to quantify. Individually, structure-based prediction of deamidation "hotspots" is achieved through half-life analysis, while reaction kinetics are quantified through quantum chemical methodologies.[242] We attribute this difficulty to the relatively small quantities of degradation products in a free water system, as specified by Kahne and Still in an experimental hydrolysis study.[165, 242]

As we previously mentioned, the current approach towards understanding the mechanics behind peptide degradation is achieved through quantum chemical methodology.[169, 184, 242-245] However, current studies are limited by two factors. First, there are limited studies available to predict the thermochemical and kinetic properties of dipeptide deamidation at a neutral pH. Understanding the thermochemical and kinetic properties of dipeptide deamidation is fundamental in additive formulation for therapeutic proteins because it provides insight into the self-promoted hydrogen transfer from the amine group of the peptide bond to induce cyclization.[82, 169] Second, as few current studies fully map the reaction pathways of the deamidation networks, fundamental understanding of the mechanism is limited. By mapping the full network, the selectivity of a peptide to deamidate via direct hydrolysis or a cyclic intermediate can be quantified. Additionally, the computed thermochemical and kinetic parameters generated by mapping the full network can be applied to the prediction of rate-limiting steps and reaction selectivity through a microkinetic model as a function of pH.

For this purpose, we conducted a computational study, using a subset of amino acid dipeptides, to predict the thermochemical and kinetic parameters of each reaction in the Asn deamidation network. A total of 77 dipeptides (Appendix A) were analyzed over four solvent dielectrics (308 total reactions) in the conductor-like polarizability model (CPCM) solvated state, this paper presents the thermochemical and kinetic parameters, i.e. standard enthalpies of reaction, standard entropies of reaction, Gibbs free energies of reaction, Gibbs free energies of solvation, activation energies, and kinetic rate constants for each dipeptide. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic energy, and then statistical thermodynamics was applied to all reactions to determine temperature effects on the reaction.[82, 141, 147, 187-190, 246, 247] All thermochemical properties were calculated at standard pressure and temperature to reduce external variabilities. Overall, this study generated a database for the predicted thermochemical and kinetic parameters of the Asn-X deamidation reaction network, which can be applied future theoretical studies and as cross-validation with experimental studies.

2.2. Project Objectives

The objectives of this study were to develop a database of thermochemical and kinetic parameters associated with Asn-X deamidation during storage under ambient conditions. By mapping the reaction network, we were able to identify the correlation between different pathways and the environment of the dipeptide, as well as quantify the impact of residue functional groups on the propensity of Asn-X to deamidate. Furthermore, we have analyzed the competition between the two primary pathways of Asn-X deamidation, as well as the two products of the succinimide pathway, through the use of chemical reaction engineering principles. Overall, this study provides a self-consistent database of the thermochemical and kinetic parameters associated with the mechanisms of Asn-X deamidation under ambient conditions, as a function of the different dielectrics associated with amino acids, which can be expanded to account for other environmental effects in future studies.

2.3. Computational Methodology

2.3.i. Modeled Reaction Network and Chemical System

Transformation of residues is achieved through two primary pathways, direct hydrolysis and the succinimide reaction pathway (Figure 2.3.1).[80, 128, 149, 150, 155, 169, 184, 238]Direct hydrolysis of residues is achieved in a single step, and occurs through the hydrogen transfer of water to form free ammonia and the addition of a hydroxyl group to the peptide.[80, 128, 149, 150, 155, 169, 184, 238] The succinimide reaction pathway is a two-step process, where the proposed rate-determining step first forms the succinimide intermediate before it is hydrolyzed into either Asp-X or its isomeric form iso-Asp-X.[80, 128, 149, 150, 155, 169, 184, 238]Formation of the succinimide intermediate is referred to as a self-promoted process due to the nucleophilic attack by the amide Hydrogen to eliminate ammonia and spontaneously form the five-membered succinimide ring, and has been noted by prior studies to be the more competitive reaction pathway under ambient conditions.[80, 128, 149, 150, 155, 169, 184, 238] Evaluation of the theoretical isoimide reaction pathway was completed to determine if deamidation via the formation of the isoimide ring was competitive, and our results indicated that deamidation via the proposed reaction pathway was thermodynamically and kinetically unfavorable compared to the direct hydrolysis and succinimide reaction pathways.[82, 169] Therefore, only the competitive reaction pathways were evaluated within the results of our study.

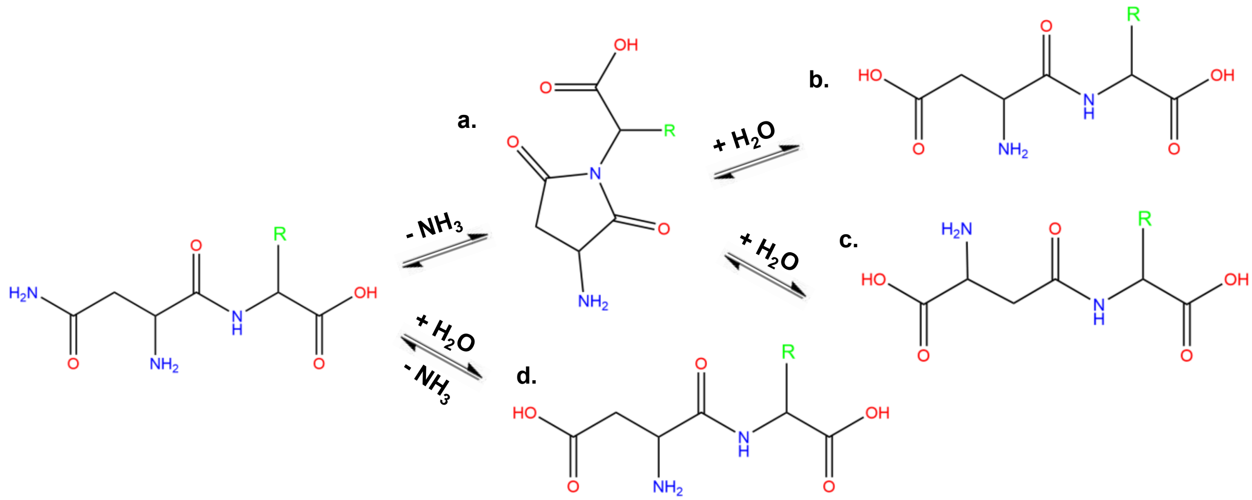

Figure 2.3.1. Reaction schematic of selected deamidation routes within the Asn-X deamidation reaction network. (a). Succinimide formation, (b). Succinimide hydrolysis into Asp-X, (c). Succinimide hydrolysis into iso-Asp-X, (d). Direct hydrolysis.

We modeled our reaction mechanism under two solvation models, purely implicit solvation through the Conductor-like Polarizable Continuum (CPCM) implicit solvation and a hybrid implicit/explicit solvation model that adds one explicit water molecule to interact with the N-terminus carbonyl of each dipeptide within the reaction network. The CPCM implicit solvation model is similar to the COSMO-RS solvation model which is commonly used in biological modeling, and was therefore used as the base solvation model for our studies.[209] Furthermore, we modeled our reaction network using four different dielectrics ($\varepsilon = 4$, 20, 40, and 80), to study how the solvent dielectric might impact the reaction energetics of the different dipeptides within our reaction network.[129, 248, 249]

The location of the water molecule relative to the dipeptide, as well as the number of explicit water molecules within our solvation model was analyzed by evaluating the impact of conformation and the number of waters on protein reaction energetics. Via this analysis, we determined that the difference in the forward rate coefficients as a function of what portion of the dipeptide the water interacts with, as well as the difference in rate coefficients between a hybrid solvation model with one and two water molecules, was not statistically significant. Therefore, our hybrid solvation model of one explicit water molecule interacting with the N-terminus carbonyl group of each dipeptide was expanded to each reaction within our deamidation network.

2.3.ii. Quantum Chemical Calculations and Statistical Thermodynamics

Quantum chemical calculations were performed with *Gaussian 16*.[82, 141, 147, 250, 251] All electronic energies for the protein dipeptides, intermediates and transition states were calculated using the B3LYP method.[82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized in the singlet state. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the reaction network.[82, 141, 147, 188-190, 246, 247, 253] To account for anharmonicity in the normal vibrational modes, as suggested by Scott and Random, the harmonic frequencies and zero-point energy (ZPE) were scaled by factors of 0.96 and 0.98 respectively.[82, 141, 147, 254] Following the work of previous reaction modeling completed by our group, partition functions based on the harmonic oscillator and rigid rotor approximations were used to calculate thermodynamic and kinetic properties as a function of temperature.[82, 141, 147, 255, 256] This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 257-263]

2.3.iii. Transition State Theory and Kinetic Parameters

Transition states for each specified reaction within the deamidation network were found using the Berny method.[82, 141, 147, 264-266] The imaginary frequency of each transition state was animated, and intrinsic reaction-coordinate calculations were performed to confirm that the normal vibrational mode pertained to the reaction coordinate of interest. A total of four reaction pathways

– direct hydrolysis, succinimide formation, succinimide hydrolysis into Asp, and succinimide hydrolysis into iso-Asp - were mapped for each chemical model used in this study. Conversion of intermediate structures to their final product form, as well as the ionization of N-terminus Asp residues under physiological conditions, were treated as equilibrated reactions and were not included in our computed rate coefficients.[239]

Conventional transition-state theory (TST) was then used to calculate rate coefficients at 1 atm, according to the macroscopic formulation in Equation (4.3.1).[82, 141, 147, 267]:

$$k^{TST}(T) = n_d \acute{\kappa} = n_d \Lambda \exp\frac{\Delta S^{\neq}}{R} \exp\left(\frac{-\Delta H^{\neq}}{RT}\right) \qquad (2.3.1)$$

Where Λ is defined in Equation (4.3.2):

$$\Lambda = \kappa(T)\frac{k_B T (V_m)^{-\Delta n}}{h} \qquad (2.3.2)$$

Here, $\acute{\kappa}$ is the single-event rate coefficient, κ(T) is the Wigner tunneling correction at temperature T, k_B is Boltzmann's constant, h is Planck's constant, V_m is the molar volume, R is the ideal gas constant, $\Delta S^{\neq}$ is the entropy of activation, $\Delta H^{\neq}$ is the enthalpy of activation, Δn is the change in the number of moles going from the reactant to the transition state (i.e., -1 for bimolecular reactions such as for the direct hydrolysis reaction in both the forward and reverse directions, the succinimide formation reaction in the reverse direction, and the succinimide hydrolysis reaction in the forward direction; and 0 for unimolecular reactions such as the succinimide formation reaction in the forward direction and the succinimide hydrolysis reaction in the reverse direction), and n_d is the reaction path degeneracy, or number of single events. The reaction path degeneracy follows Equation (2.3.3). $\Delta H^{\neq}$ and $\Delta S^{\neq}$ are calculated using standard formulae, and there were no internal rotation corrections needed within our system.[82, 141, 147, 196, 267]

The reaction path degeneracy, n_d, definition is based upon the symmetry of the reactant(s) and the transition state.[82, 141, 147, 196] Symmetry was determined by the number of chiral centers present in each molecule.[82, 141, 147, 196, 239, 256, 263] To achieve an unbiased representation of each modeled reaction, the maximum symmetry numbers were applied to each reaction.[82, 141, 147, 239] Analysis of the reaction path degeneracies used in our model, noted that the inclusion of chiral centers did not impact the forward reaction path degeneracy developed through the exclusion of chiral centers. Therefore, our model may be applied to both natural and synthetic proteins in the forward direction of each computed reaction.

$$n_{\mathrm{d}} = \frac{n_{\mathrm{chiral}}^{\neq}}{\prod_{\mathrm{j}} n_{\mathrm{chiral}}^{\mathrm{reactant,j}}} \times \frac{\prod_{\mathrm{j}} \sigma^{\mathrm{reactant,j}}}{\sigma^{\neq}} \tag{2.3.3}$$

Where n_{chiral} is the number of chiral isomers for the transition state and reactant(s); σ is a product of the external rotation symmetry numbers for the transition state and reactant(s). Internal symmetry numbers for non-cancelling internal rotations treated as hindered or free rotations were included in the hindered and free rotor partition functions, respectively.

The single-event parameters of the Arrhenius relationship, log(A) and E_A were obtained by fitting ln(k) versus T^{-1} over the temperature range 298 – 1500 K. This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 255, 257-263] Through linear regression analysis, we determined Arrhenius behavior was obeyed well for all reactions. The solvation impact of liquid phase kinetics is beyond the scope of this work but will be addressed in future studies following the work of Jalan et al, which computes liquid-phase kinetics in relation to both the solute and the solvent.[82, 141, 147, 268]

2.3.iv. Generalized Selectivity Model and Application to Peptide Deamidation

Reaction selectivity analysis was completed using chemical reaction engineering principles.[269] To obtain a robust analysis of our reaction network, we chose to compute the instantaneous selectivity for each reaction pathway within the deamidation reaction network. [269] To apply the instantaneous selectivity analysis to our reaction network, we first identified the proposed rate-determining step for the succinimide reaction pathway.[269] Once those rate-determining steps had been defined, we then applied Equation 2.3.4 to our analysis.[269]

$$\text{Instantaneous selectivity} = \frac{k'}{k''} \tag{2.3.4}$$

Where k' is the desired reaction and k" is the undesired reaction. Selectivity was analyzed from 298.15 K to 400 K, the range of temperatures cited in prior deamidation studies, to quantify the impact of temperature on reaction selectivity.[152, 153, 270] To determine if the results of our model agreed with prior studies regarding the propensity of Asn residues to deamidate via the succinimide pathway over the direct hydrolysis pathway at a neutral pH, the direct hydrolysis reaction pathway was treated as the desired reaction while the cyclic intermediate reaction pathways were treated as undesired.[271, 272] Additionally, to fully quantify the impact of the solvent dielectric on the competing pathways of succinimide hydrolysis, we also analyzed the selectivity between the succinimide hydrolysis into Asp-X and iso-Asp-X pathways to determine which dipeptides prefer to hydrolyze into either Asp or its more acidic counterpart, iso-Asp.[153, 273]

2.3.v. Sensitivity Analysis and Generalized Model Validation

Model accuracy is dependent on both the level of theory used and the chemical model chosen. In this study, we chose a level of theory that has been used to model Asn deamidation, B3LYP/6-31+G(2d,2p), with the inclusion of the GD3 empirical dispersion correction to account for van der Waals interactions within the reaction network.[187-190, 246, 247] Additionally, we modeled our reaction mechanism under two solvation models, purely implicit solvation through the Conductor-like Polarizable Continuum (CPCM) implicit solvation and a hybrid implicit/explicit solvation model that adds one explicit water molecule to interact with the N-terminus carbonyl of each dipeptide within the reaction network. For the purpose of validating our level of theory and identifying the chemical model that was more accurate, we first conducted a two-sample t-test for the difference in means between the forward rate coefficient ($\ln(k)^F$) for each reaction in the purely implicit solvation model and the forward rate coefficient ($\ln(k)^F$) for each reaction in the hybrid solvation model at 298.15 K and all dielectrics (Table 2.2.1).[274, 275] Through our statistical analysis we determined that only the difference in the means of the forward rate coefficients, at 298.15 K, was statistically insignificant for the direct hydrolysis, pathway.

Table 2.3.1. Two-sample t-test for the difference in means between the forward rate coefficient ($\ln(k)^F$) of the implicit chemical model and the forward rate coefficient ($\ln(k)^F$) implicit + 1 H_2O chemical model. Rate coefficients were taken at a temperature of 298.15 K for all dielectrics modeled in the reaction network.

Reaction	t-value	DOF	P-value
Direct Hydrolysis	-1.52	158	0.131
Succinimide Formation	-4.13	150	< 0.001
Succinimide Hydrolysis into Asp-X	-4.27	150	< 0.001
Succinimide Hydrolysis into iso-Asp-X	-6.15	150	< 0.001

The statistical insignificance of the inclusion of one explicit water molecule to the direct hydrolysis forward rate coefficients can be attributed to the linear nature of the transition state, where the interaction between the explicit water molecule and transition state is not as prevalent as that of the water molecule directly involved in the reaction. Specifically, unlike the transition states of the succinimide formation reaction, where the partial ring formation of the transition state interacts with the explicit water molecule, the direct hydrolysis transition states interact with the water molecule which attacks the N-terminus amine group to eliminate ammonia and form Asp-X. Therefore, due to the competition between the explicit water molecule, which does not directly interact with the reaction, and the reactive water, it is likely that the inclusion of the stabilizing water molecule would not influence of the reaction rate coefficient of the direct hydrolysis reaction.[276]

After establishing that the inclusion of one explicit water molecule within the solvation model significantly impacts the forward rate coefficient for three of the four reactions within our network, we sought to validate our level of theory and determine which chemical model provides a more accurate representation of Asn deamidation. For this purpose, we compared the values we computed for each chemical model to prior computational and experimental studies. Through this analysis, we determined that both chemical models predicted activation energies that were about 10 kcal.mol^{-1} smaller than those given by prior computational studies on direct hydrolysis and Succinimide formation reactions.[243, 277, 278]

However, when we compared our two chemical models to the experimentally quantified half-life of the Asparaginyl-Glycine (Asn-Gly) succinimide hydrolysis into Aspartic Acid-Glycine (Asp-Gly) reaction, we determined that the hybrid solvation model yielded a more accurate result (Table 2.2.2).[279] Specifically, $\Delta t_{1/2}$ of the implicit solvation model was four times greater than $\Delta t_{1/2}$ of the hybrid solvation model.[279] Therefore, we applied our analysis of the mapped Asn-X reaction network to the hybrid solvation model as a function of the four dielectric coefficients.

Table 2.3.2. Comparative analysis of the computed half-lives of Asn-Gly succinimide hydrolysis for the implicit and hybrid chemical models against the experimental half-life quantified by Geiger et al.[279] The half-lives were computed at 300 K and using a pseudo-first order reaction to match the parameters specified by Geiger (310 K, pseudo-first order reaction, half-life determined via analytical HPLC).[279]

Literature		Implicit Model		Implicit + 1 H_2O Model	
Reaction	**$t_{1/2}$**	**$t_{1/2}$ (Computed)**	**$\Delta t_{1/2}$ (Literature – Computed)**	**$t_{1/2}$ (Computed)**	**$\Delta t_{1/2}$ (Literature – Computed)**
Asn-Gly Succinimide Hydrolysis	2.30 hours	7.66 hours	5.36 hours	3.57 hours	1.27 hours

2.4. Results and Discussion

2.4.i. Structures

To achieve a complete analysis of each reaction pathway depicted within our reaction network, we optimized the following structures. Modeling of the direct hydrolysis reaction included two reactants (the dipeptide and water, infinitely separated), the transition state, and two products (the dipeptide and ammonia, infinitely separated). Formation of the cyclic intermediates were modeled as intramolecular processes through a single reactant, transition state, and two products (the cyclic intermediate and ammonia, infinitely separated). The cyclic intermediate hydrolysis pathways were modeled as single step reactions with two reactants (the cyclic intermediate and water, infinitely separated), the transition state, and one product were included in the modeled reaction pathway.[82, 280, 281] The dipeptide structures are labeled in the X-X format using the corresponding 3-letter identifiers of the amino acids used to synthesize the dipeptide and the ammonia and water structures are simply labeled as water and ammonia, respectively. Figure 2.4.1. depicts a ball and stick representation of sample reactions within each pathway present in the deamidation reaction network.

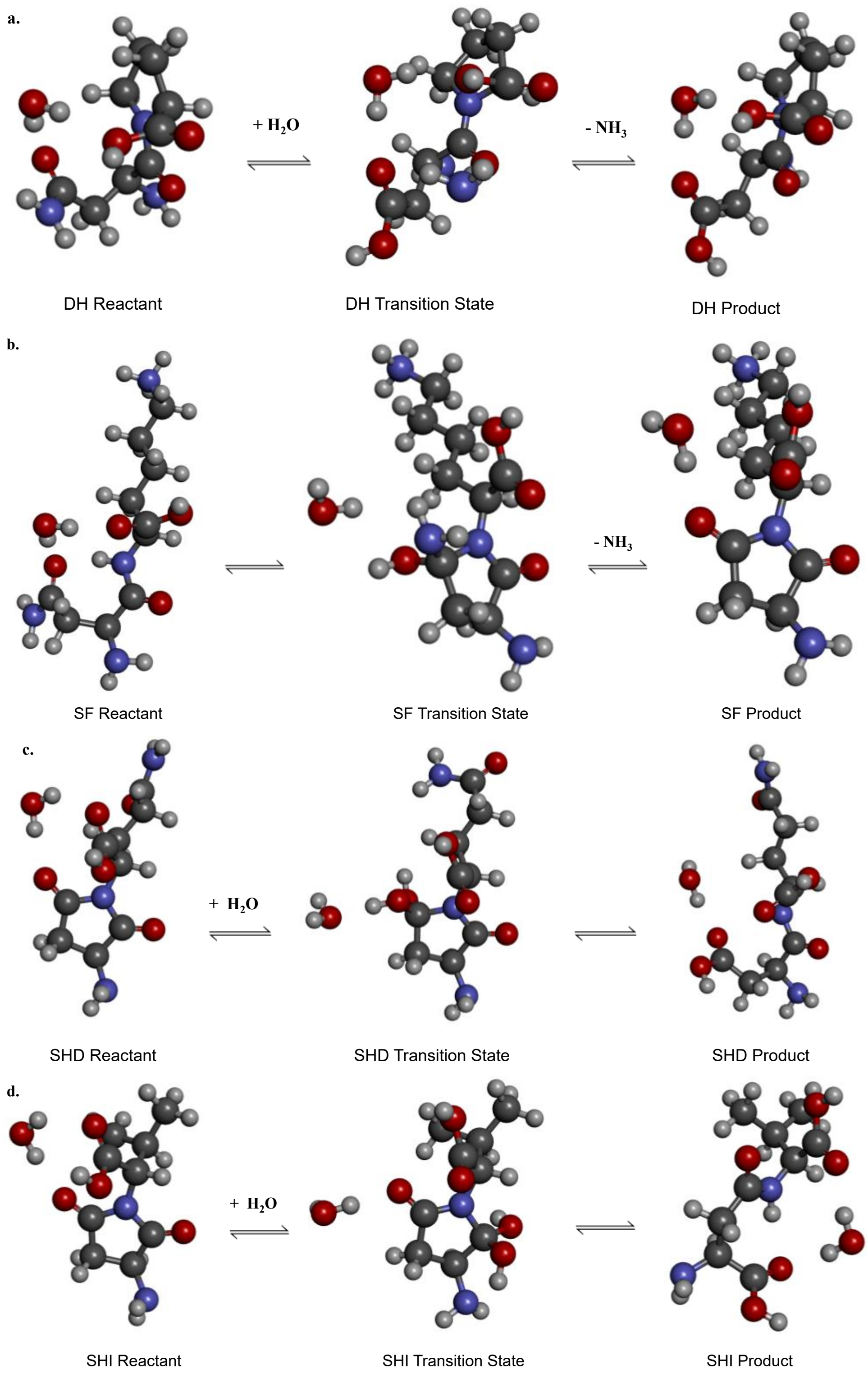

Figure 2.4.1. Representative ball and stick structures used to model the different reactions within the Asn-X deamidation reaction network during storage under ambient conditions.

The structures used in the direct hydrolysis pathway of Asn-X deamidation optimized in the expected form where the primary water-dipeptide interaction was through the nucleophilic attack of water to eliminate ammonia and form the acidic Asp-X dipeptide. The nucleophilic attack of water within the direct hydrolysis pathway occurred through the decomposition of the reactive water molecule to form a hydroxyl group (OH^-) and a Hydrogen atom (H^+). The hydrogen atom then attached to the N-terminus amine group to form ammonia, while the hydroxyl group attached to the N-terminus carbonyl group to convert the dipeptide from net neutral Asn-X to its acidic form. The direct hydrolysis reaction pathway was the only pathway within our reaction network that Asparaginyl-Proline (Asn-Pro) was considered in the analysis. The exclusion of Asn-Pro from the other reaction pathways is due to its inability to cyclize into succinimide due to the steric hindrances of the rotationally constrained rigid-ring structure of the Pro functional group.[151, 239, 282] According to a study completed by Stephenson and Clarke, succinimide formation within proteins is inhibited by steric hindrances of the interactions between the nitrogen of the central amide bond and the sidechain carbonyl of the deamidating Asn residue.[283] Therefore, it is physically improbable for Asn-Pro to deamidate via a pathway that is not direct hydrolysis.[283]

Analysis of the succinimide formation transition states indicated that all dipeptides partially cyclized in the transition state, and that the stabilizing water molecule in our hybrid solvation method interacts with both the N-terminus amine and amide bond through which the succinimide ring forms. As succinimide forms through the nucleophilic attack by the amide Hydrogen to eliminate ammonia and spontaneously form the five-membered succinimide ring, which is then hydrolyzed into either Asp-X or iso-Asp-X.[169, 243, 283] The hydrolysis of the succinimide intermediate was modeled as a single step process where the succinimide ring is opened through the nucleophilic attack of water. The nucleophilic attack of water forms a gemdiol-like transition state at the carbonyl group opposite the amide group of the cyclic ring.[243] Similar to the hydrolysis of succinimide dipeptides into Asp-X, the hydrolysis into iso-Asp-X dipeptides occurred as a single reaction. However, for the conversion into iso-Asp-X, the nucleophilic attack was directed towards the carbonyl adjoining the amide group of the cyclic ring. Via the formation of the gemdiol at this position, the isomeric form of Asp-X was formed. The structure of iso-Asp-X differs from Asp-X in the sense that the side-chain carbonyl group is moved to the peptide backbone, placing the amine and carbonyl groups in the same geometric plane.[273, 283, 284] In our hybrid solvation model, the stabilizing water directly interacts with the gemdiol functional group of the succinimide hydrolysis into Asp-X transition state, whereas it interacts with the N-terminus carbonyl group in the succinimide hydrolysis into iso-Asp-X pathway. The orientation of the stabilizing water relative to the Asp-X and iso-Asp-X gemdiol functional groups was defined to ensure self-consistency within our chemical model, where the assisting water primarily interacts with the N-terminus carbonyl group, which was defined by the succinimide intermediate.

2.4.ii. Thermochemistry and Arrhenius Parameterization

To determine the impact of the modeled dielectric on the reaction energetics of the different pathways within our reaction network, we categorized and analyzed the computed thermochemical and kinetic parameters for each modeled dielectric. In addition to the standard thermochemical values (Enthalpy of Reaction ($\Delta H_{RXN}^{298.15K}$), Gibbs Free Energy of Reaction ($\Delta G_{RXN}^{298.15K}$), and Entropy ($\Delta S_{RXN}^{298.15K}$)), Arrhenius parameters were computed in both directions. Analysis of the thermochemical and kinetic parameters associated with each reaction noted the following.[82]

2.4.ii.a. Direct Hydrolysis

The direct hydrolysis reaction was found to be both net endothermic and endergonic in nature, regardless of the modeled dielectric. As we have previously specified, the direct hydrolysis reaction pathway is least competitive under ambient conditions.[153] Therefore, the non-spontaneous nature of the direct hydrolysis reactions can be attributed to the likelihood of the reactions to occur at a neutral pH. The preference of peptide deamidation towards the succinimide pathway over the direct hydrolysis pathway at a neutral pH also suggests a need for energy input for the hydrolysis reactions.[153, 243, 285-287] Additionally, a significantly smaller entropic driving force was present in the direct hydrolysis reaction pathway compared to the computed entropies of the other reaction pathways in the Asn-X deamidation network. We attribute this deviance between reaction pathways to the cyclic nature of the other pathways within the network, as previous studies have noted the significance of entropy on the cyclization of protein structures.[288, 289] Additionally, we noted that the $\Delta G_{RXN}^{298.15K}$ was smallest when the modeled dielectric equal to 20 was applied to the reaction. Depicted in Table 2.4.1, the Gibbs free energy of the direct hydrolysis reaction at a dielectric equal to 20 is about 2 kcal.mol^{-1} smaller than the free energy of reaction associated with the other three dielectric coefficients. Previous computational studies have assigned a dielectric of 20 to charged residues, such as the product of the direct hydrolysis reaction Asp-X.[212, 213] The direct hydrolysis reaction has been noted in previous deamidation studies to be a more competitive pathway under either an acidic or a basic pH.[168] Therefore, due to the modeled dielectric mimicking an acidic or basic environment, it is likely that the direct hydrolysis reaction at ε =20 would be more spontaneous than at the other specified dielectric coefficients.

Table 2.4.1. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for direct hydrolysis at 298 K and 1 atm. Ã has units of L.mol^{-1}.s^{-1} for the elimination and addition reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of kcal.mol^{-1}.

ε	Forward Reaction			$\Delta H^{298.15}_{rxn}$	$\Delta G^{298.15}_{rxn}$	Reverse Reaction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	39.8	7.40	40.2	4.81	4.82	34.6	7.74	35.5
20	36.2	7.40	36.7	3.08	3.05	32.7	7.71	33.6
40	39.4	7.44	39.9	5.31	5.30	33.7	7.77	34.6
80	39.4	7.42	39.9	5.51	5.42	33.5	7.68	34.4

The phenomena observed at a dielectric of 20 for the Gibbs free energy of the direct hydrolysis reaction was also evident in the forward and reverse Arrhenius parameters. As depicted in Table 2.5.1, the forward and reverse Arrhenius parameters were consistent among dielectrics equal to 4, 40, and 80. For the reaction modeled using a dielectric coefficient equal to 4, the forward $E_0 = 39.8 \pm 7.12$ kcal.mol^{-1}, $E_A = 40.2 \pm 6.73$ kcal.mol^{-1}, and log(A) = 7.40 ± 1.42 s L.mol^{-1}.s^{-1}, while the reverse $E_0 = 34.6 \pm 7.62$ kcal.mol^{-1}, $E_A = 35.5 \pm 7.35$ kcal.mol^{-1}, and log(A) = 7.741 ± 1.14 L.mol^{-1}.s^{-1}, respectively. While the forward and reverse Arrhenius parameters computed for each of these dielectrics were within 1 kcal.mol^{-1} of each other, they were about 3 kcal.mol^{-1} greater than the Arrhenius parameters computed at ε = 20. Specifically, the forward Arrhenius parameters were computed as $E_0 = 36.2 \pm 6.83$ kcal.mol^{-1}, $E_A = 36.7 \pm 6.44$ kcal.mol^{-1}, and log(A) = 7.40 ± 1.20 L.mol^{-1}.s^{-1}, while the reverse Arrhenius parameters were computed as $E_0 = 32.7 \pm 7.31$ kcal.mol^{-1}, $E_A = 33.6 \pm 7.02$ kcal.mol^{-1}, and log(A) = 7.71 ± 1.17 L.mol^{-1}.s^{-1}, respectively.

The discrepancy in Arrhenius parameters between the dielectric associated with charged residues compared to the other three dielectrics analyzed coincides with the phenomena we observed in the analysis of the computed thermochemical parameters for the direct hydrolysis reaction. Further analysis of this trend suggests that the solvent dielectric would increase the favorability of the direct hydrolysis reaction within the deamidation reaction network. Specifically, that the solvent dielectric associated with charged residues is representative of the charged solvent environment in which the direct hydrolysis reaction is most thermodynamically and kinetically favorable within the deamidation reaction network. Via this analysis, we determined that the most suitable dielectric for modeling the direct hydrolysis reaction pathway in the Asn-X deamidation reaction network is the one associated with charged residues.

2.4.ii.b. Succinimide Formation

Analysis of the succinimide formation reaction indicated that, regardless of the solvent dielectric, the reaction was both endothermic and spontaneous in nature (Table 2.5.2), as was noted in previous studies.[243, 283] The spontaneity of the succinimide formation reaction can be attributed to the increased stability of cyclic protein structures compared to acyclic protein structures.[288, 289] This large entropy gain, which was observed in both cyclic intermediate formation reactions, is noted by literature as a factor influencing cyclization reactions.[290-292] Unlike the direct hydrolysis pathway, little variation was observed in the succinimide formation Arrhenius coefficients when the solvent dielectric was changed. The lowest activation barrier of succinimide formation was identified at the dielectric of pure water ($\varepsilon = 80$), and increased by about 0.20 $kcal.mol^{-1}$ from dielectrics equal to 40 and 20, respectively. The largest step change in activation energy between the solvent dielectrics occurred at the most hydrophobic dielectric ($\varepsilon = 4$). At a dielectric equal to 4, $E_0 = 40.5 \pm 6.64$ $kcal.mol^{-1}$, $E_A = 39.8 \pm 6.46$ $kcal.mol^{-1}$, and $\log(A) = 11.3 \pm 1.39$ s^{-1} in the forward direction, and $E_0 = 31.6 \pm 6.72$ $kcal.mol^{-1}$, $E_A = 31.6 \pm 6.48$ $kcal.mol^{-1}$, and $\log(A) = 6.31 \pm 1.30$ $L.mol^{-1}.s^{-1}$ in the reverse direction.

Table 2.4.2. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for succinimide formation at 298 K and 1 atm. Ã has units of s^{-1} and $L.mol^{-1}.s^{-1}$ for the elimination and addition reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Direction			$\Delta H^{298.15}_{rxn}$	$\Delta G^{298.15}_{rxn}$	Reverse Direction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	40.5	11.3	39.8	10.1	-0.80	31.6	6.31	31.6
20	39.3	11.6	38.8	9.11	-1.78	31.3	6.65	31.5
40	39.1	11.6	38.6	9.16	-1.62	31.1	6.72	31.3
80	39.0	11.5	38.5	9.12	-1.66	31.0	6.62	31.2

The inverse relationship between the solvent dielectric and the reaction activation barrier qualitatively matches prior findings that the reactivity of succinimide formation decreases with the solvent dielectric (Figure 2.4.3).[248, 293] Brennan and Clarke rationalize the reduced reactivity to the destabilization of the peptide bond Nitrogen, which facilitates the formation of the succinimide ring.[293] Specifically, that it would not be as labile as it is at higher solvent dielectrics.[293] While we are unable to model bond lability, we were able to qualitatively match the trend defined by Brennan and Clarke, where succinimide reactivity decreases with solvent dielectric (Figure 2.4.2).[248, 293] Therefore, it is likely that because there is no reactive water to drive the reaction, the impact of the solvent dielectric impacts the reactivity of the intramolecular reactions within our reaction network, such as succinimide formation.

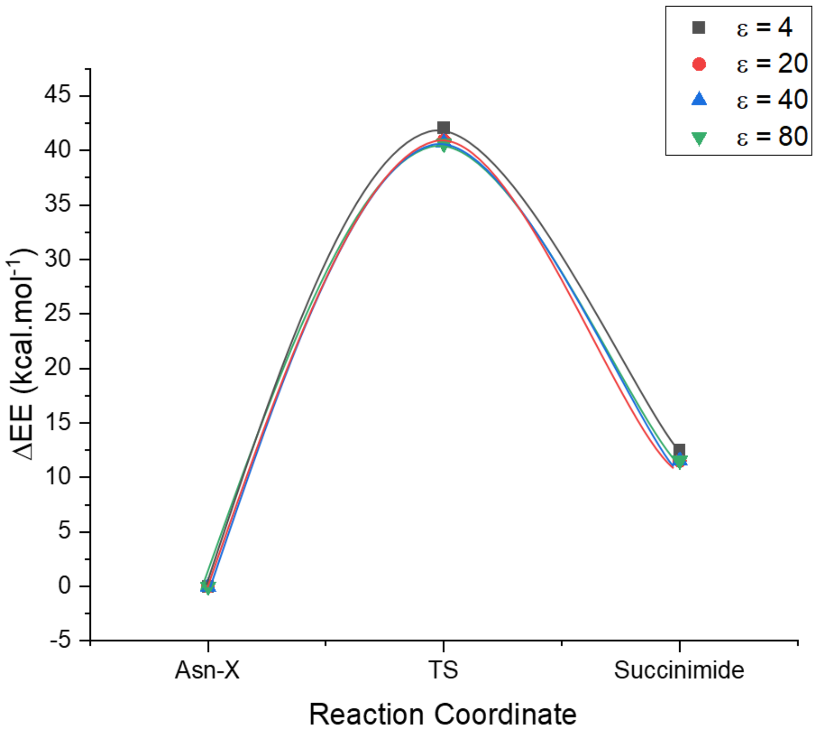

Figure 2.4.2. Diagram of the change in electronic energy observed during the succinimide formation reaction as a function of the modeled dielectric.

2.4.ii.c. Succinimide Hydrolysis into Asp-X

Analysis of the succinimide hydrolysis into Asp-X pathway (Table 2.4.3.) indicated that the reaction pathway was mildly exothermic and endergonic in nature. Analysis of the dipeptides within the reaction class indicated that nine of the 19 analyzed dipeptide reactions were endothermic in nature. The nine dipeptides found to be endothermic were Asparaginyl-Alanine (Asn-Ala), Asparaginyl-Glutamine (Asn-Gln), Asparaginyl-Glutamic Acid (Asn-Glu), Asn-Gly, Asparaginyl-Phenylalanine (Asn-Phe), Asparaginyl-Threonine (Asn-Thr), Asparaginyl-Tryptophan (Asn-Trp), and Asparaginyl-Valine (Asn-Val). Succinimide hydrolysis of Asn-Ala and Asn-Phe was found to be endothermic regardless of which product was formed. Asn-Ala and Asn-Phe are most likely endothermic regardless of the succinimide hydrolysis reaction because the C-terminus residues are amino acids commonly buried within a protein surface.[239, 294] Therefore, the two dipeptides would prefer a structure similar to that of the peptide backbone, such as the product of succinimide hydrolysis, over being exposed by the steric hindrance of the succinimide intermediate.[239, 294] As such, through Hammond's postulate, the succinimide hydrolysis of Asn-Ala and Asn-Phe would be endothermic.[295]

The endothermicity of Asn-Val, which was found to be endothermic for both succinimide hydrolysis reactions, can be attributed to its branched functional groups.[239, 245] Specifically, prior studies have noted that C-terminus residues with branched, hydrophobic side chains inhibit the formation of the succinimide intermediate.[245] Therefore, following Hammond's postulate, Asn-Val would prefer to hydrolyze into Asp-Val instead of remaining in the succinimide intermediate.[295] The propensity of Asn-Gln, Asn-Trp, and Asn-Tyr towards the product of succinimide hydrolysis is due to the role of the C-terminus residues as hydrogen bond donors. [239, 296] Our analysis of the three product structures indicated that the hydrogen-donating sidechains of each residue were in the same geometric plane as the hydrogen accepting Asp-X. Therefore, the products would be favored for the hydrolysis of these three dipeptides. The entropy of reaction (ΔS_{RXN}) was found to be negative for the hydrolysis of the succinimide intermediate, likely due to the entropy driving the formation of the cyclic intermediate.[290-292] Therefore, as the ring is broken, the system transitions from disordered to ordered resulting in a negative ΔS_{RXN} for all cyclic intermediate hydrolysis reactions.

Table 2.4.3. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for succinimide hydrolysis into Asp-X at 298 K and 1 atm. Ã has units of L.mol^{-1}.s^{-1} and s^{-1} for the hydrolysis and condensation reactions, respectively. E_0, E_A, and $\Delta H^{298.15}{}_{rxn}$ have units of kcal.mol^{-1}.

ε	Forward Direction			$\Delta H^{298.15}{}_{rxn}$	$\Delta G^{298.15}{}_{rxn}$	Reverse Direction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	13.5	6.80	13.9	-1.88	8.76	14.0	11.9	14.0
20	13.9	6.39	14.3	-0.06	10.5	12.6	11.8	12.6
40	13.8	6.86	14.2	-0.64	9.77	13.1	11.8	13.1
80	13.9	6.87	14.3	-0.40	9.91	13.0	11.7	13.0

Analysis of the Arrhenius parameters indicated that the solvent dielectric had little impact on the energetics of succinimide hydrolysis into Asp-X. The average activation energy of the reaction pathway was smallest at the most hydrophobic dielectric, and about 0.400 kcal.mol^{-1} greater at the other three solvent dielectrics. At $\varepsilon = 4$, $E_0 = 13.5 \pm 1.88$ kcal.mol^{-1}, $E_A = 13.9 \pm 1.94$ kcal.mol^{-1}, and log(A) = 6.80 ± 0.83 L.mol^{-1}.s^{-1} in the forward direction, while $E_0 = 14.0 \pm 3.49$ kcal.mol^{-1}, E_A =14.0 ± 3.60 kcal.mol^{-1}, and log(A) = 11.9 ± 1.05 s^{-1} in the reverse direction. Subsequently, at the dielectric of pure water ($\varepsilon = 80$), $E_0 = 13.9 \pm 1.45$ kcal.mol^{-1}, $E_A = 14.3 \pm 1.62$ kcal.mol^{-1}, and log(A) = 6.87 ± 1.01 L.mol^{-1}.s^{-1} in the forward direction, and $E_0 = 13.0 \pm 3.50$ kcal.mol^{-1}, $E_A = 13.0 \pm 3.68$ kcal.mol^{-1}, and log(A) = 11.7 ± 1.04 s^{-1} in the reverse direction. The negligible impact of the solvent dielectric on the succinimide hydrolysis pathway can be attributed to the formation of the succinimide intermediate being primarily impacted by the change in solvent dielectric. Specifically, because the lability of the peptide bond Nitrogen does not impact the reactivity of succinimide hydrolysis.[293] Therefore, the succinimide hydrolysis reaction pathway would not be as significantly impacted as the formation reaction. Additionally, we hypothesize that the interaction between the solvent and the dipeptide primarily impacts intramolecular reactions. As the hydrolysis of the succinimide intermediate occurs through the nucleophilic attack of water, the reactive water likely drives the reaction energetics for the reaction pathway, similar to the direct hydrolysis reaction.

2.4.ii.d. Succinimide Hydrolysis into iso-Asp-X

The secondary pathway of the succinimide reaction, hydrolysis into iso-Asp-X was also found to be exothermic and endergonic in nature (Table 2.4.4.). Similar to the hydrolysis into Asp-X reaction, the thermochemical and Arrhenius parameters were not significantly impacted by the change in solvent dielectric. Of the 19 dipeptides in the reaction pathway, about 11 of the reactions were exothermic. The four dipeptides with aliphatic C-terminus residues (Asn-Ala, Asparaginyl-Isoleucine (Asn-Ile), Asparaginyl-Leucine (Asn-Leu), and Asn-Val), as well as Asparaginyl-Asparagine (Asn-Asn), Asn-Gln, Asparaginyl-Histidine (Asn-His), and Asn-Phe were endothermic regardless of the solvent dielectric. Similar to Asn-Val, Asn-Ile and Asn-Leu both contain branched, hydrophobic residues. However, in the succinimide hydrolysis into Asp-X pathway, the hydrolysis of both dipeptides is exothermic. Therefore, the inhibition of the succinimide ring is not the primary cause. Instead, the orientation of the branched functional groups relative to the acidic N-terminus residue likely causes these two dipeptides to favor the formation of the product. When Asp-X is formed, the C-terminus Ile and Leu sidechains are in the same geometric plane, whereas they are in an opposing plane when iso-Asp-X is formed. The endothermicity of Asn-Asn, Asn-Gln, and Asn-His can be attributed to their role as hydrogen bond donors, where Asn-Gln favors both products because it is capable of donating more than one hydrogen atom with an sp^2 sidechain and is in the same geometric plane as both succinimide hydrolysis products.[239] The hydrogen bonding functional groups of these three dipeptides were found to be in the same geometric plane as the hydrogen bond accepting iso-Asp. Therefore, similar to the aliphatic dipeptides, the products would be more thermodynamically favorable than the reactants.

Table 2.4.4. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for succinimide hydrolysis into iso-Asp-X at 298 K and 1 atm. Ã has units of $L.mol^{-1}.s^{-1}$ and s^{-1} for the hydrolysis and condensation reactions, respectively. E_0, E_A, and $\Delta H^{298.15}{}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Direction			$\Delta H^{298.15}{}_{rxn}$	$\Delta G^{298.15}{}_{rxn}$	Reverse Direction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	9.15	6.55	9.36	-2.28	8.90	9.73	12.1	9.76
20	9.93	6.38	10.1	-0.53	10.7	8.81	11.9	8.78
40	9.78	6.51	9.90	-0.75	10.2	8.91	11.9	8.86
80	10.1	6.49	10.3	-0.50	10.7	8.78	11.9	8.76

Similar to the succinimide hydrolysis into Asp-X reaction pathway, there was little variability in the computed Arrhenius parameters as a function of the solvent dielectric, and the most hydrophobic solvent dielectric yielded the smallest average activation energy. At a dielectric equal to 4, $E_0 = 9.15 \pm 2.27$ $kcal.mol^{-1}$, $E_A = 9.36 \pm 2.46$ $kcal.mol^{-1}$, and $\log(A) = 6.55 \pm 1.19$ $L.mol^{-1}.s^{-1}$ in the forward direction, while $E_0 = 9.73 \pm 2.27$ $kcal.mol^{-1}$, $E_A = 9.76 \pm 2.44$ $kcal.mol^{-1}$, and $\log(A) = 12.1 \pm 0.89$ s^{-1} in the reverse direction. However, unlike the succinimide hydrolysis into Asp-X pathway, the dielectric coefficient associated with charged residues yielded the greatest activation energy. At a dielectric equal to 20, $E_0 = 9.93 \pm 1.60$ $kcal.mol^{-1}$, $E_A = 10.1 \pm 1.75$ $kcal.mol^{-1}$, and $\log(A) = 6.38 \pm 1.30$ $L.mol^{-1}.s^{-1}$ in the forward direction, and $E_0 = 8.78 \pm 2.87$ $kcal.mol^{-1}$, $E_A = 8.76 \pm 3.05$ $kcal.mol^{-1}$, and $\log(A) = 11.9 \pm 1.08$ s^{-1} in the reverse direction. The mild deviation in trends between the two succinimide hydrolysis pathways can be attributed to the higher sidechain acidity of iso-Asp compared to Asp.[297-299] Specifically, that the charged dielectric likely impacts the more acidic of the two succinimide hydrolysis products, and therefore influences reaction energetics.[212]

Regardless of the solvent dielectric used to model the succinimide hydrolysis into iso-Asp-X pathway, the average activation barrier of each dielectric was about 4 kcal.mol^{-1} less than the average activation energies computed for the succinimide hydrolysis into Asp-X pathway. The smaller activation energies associated with the iso-Asp-X pathway can be attributed to the preference of succinimide towards hydrolysis into the isomeric form of Asp under ambient conditions.[153, 273, 300, 301] According to previous deamidation studies, iso-Asp is the preferred deamidation product at a physiological pH due to the higher side chain acidity of iso-Asp compared to Asp.[299, 300] Therefore, as the transition state attacks the backbone carbonyl group, which opens into iso-Asp-X, it is more kinetically favorable than the reaction that occurs at the N-terminus carbonyl group to form Asp-X (Figure 2.4.3).

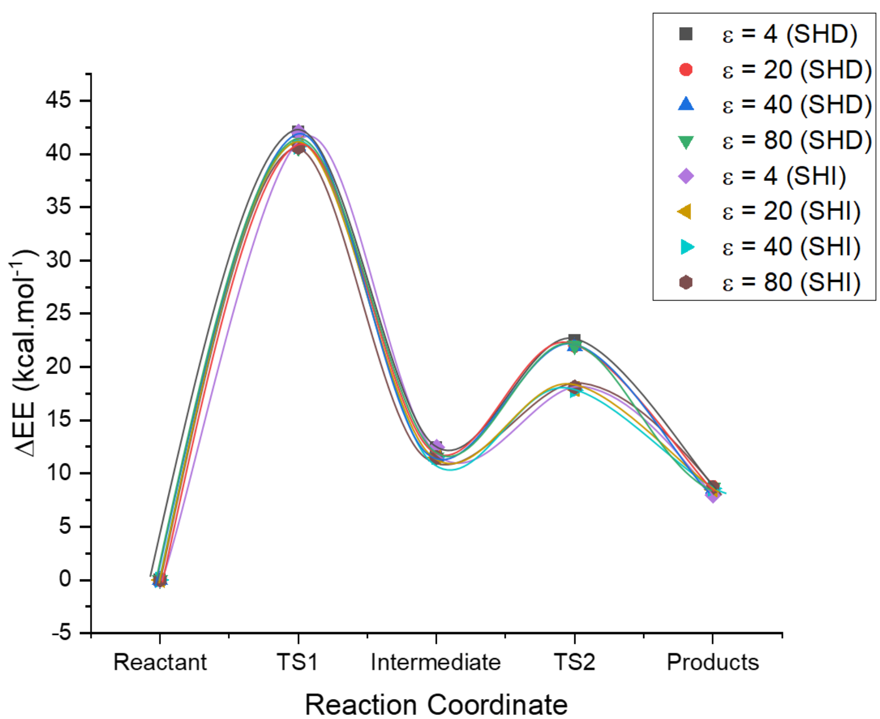

Figure 2.4.3. Diagram of the change in electronic energy and competition observed between the two products of the succinimide reaction pathways as a function of the modeled dielectric.

2.4.iii. Reaction Selectivity during Storage under Ambient Conditions

Selectivity analysis of the Asn-X deamidation network was achieved using instantaneous selectivity through a proposed rate-determining step.[82, 269] Analysis of the steps within the succinimide reaction pathway identified the formation of the cyclic intermediate to be the rate-determining, and were therefore analyzed against the forward rate coefficient of the direct hydrolysis reaction. The presence of the succinimide formation reaction as the rate-determining step agrees with both experimental and computational deamidation studies, and was not found to be influenced by the modeled dielectric.[82, 153, 184, 245, 277, 279] In addition to the identification of the rate- determining step, we also sought to analyze the competition between the two products of the succinimide pathway –Asp-X and iso-Asp-X, where the hydrolysis of succinimide into Asp-X was treated as the desired product. Lastly, we sought to determine if Asn-X deamidation reaction selectivity was influenced by temperature. Analysis of the temperature range for mAb manufacturing, where lyophilization does not occur, (298 K < T < 400 K) indicated that temperature did not influence reaction selectivity, an observation which was supported by the results of a study conducted by Bhanuramanand et al., where deamidation reactions were influenced at temperatures greater than 350 K.[302, 303]

2.4.iii.a. Direct Hydrolysis versus Succinimide Formation

As we previously mentioned, Asn-Pro is only capable of deamidating through the direct hydrolysis pathway due to the steric hindrance of the Pro functional group.[151, 239, 282] Through our analysis of instantaneous selectivity as a function of solvent dielectric (Table 2.4.5), we found that Asn-Cys and Asn-Ser also preferred deamidation through the direct hydrolysis reaction pathway. The Ser functional group is highly polar and capable of hydrogen bonding.[239, 304] Therefore, it likely acts as both a hydrogen bond donor and acceptor to the reactive water within the direct hydrolysis reaction.[239, 304] While other dipeptides within the reaction have similar hydrogen bonding capabilities, Ser is defined as a very small residue, and as such, its proximity to the water molecule initiating the direct hydrolysis reaction would be more influential than any of the other polar C-terminus residues.[164] Asn-Cys likely prefers direct hydrolysis to the formation of the succinimide intermediate due to prior findings that the deamidation of Asn-Cys dipeptides tended towards a cleavage reaction.[293] As the direct hydrolysis reaction would be better categorized as a cleavage reaction compared to the succinimide formation reaction, it is likely that Asn-Cys deamidation prefers the elimination of ammonia through the nucleophilic attack of water than through cyclization.[293]

Table 2.4.5. Reaction selectivity of the two primary routes of Asn-X deamidation during storage under ambient conditions as a function of the modeled dielectric.

ε	Direct hydrolysis	Succinimide Formation
4	Asn-Asp, Asn-Cys, Asn-Pro, Asn-Ser, Asn-Val	Asn-Ala, Asn-Arg, Asn-Asn, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp
20	Asn-Asp, Asn-Cys, Asn-Gly, Asn-Leu, Asn-Pro, Asn-Ser, Asn-Tyr, Asn-Val	Asn-Ala, Asn-Arg, Asn-Asn, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp
40	Asn-Cys, Asn-Pro, Asn-Ser	Asn-Ala, Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val
80	Asn-Cys, Asn-Pro, Asn-Ser	Asn-Ala, Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val

Analysis of reaction selectivity at the lower dielectrics within our model indicated that the Asparaginyl-Aspartic Acid (Asn-Asp) and Asn-Val dipeptides were found to prefer the direct hydrolysis pathway in our model. We attribute the propensity of Asn-Val towards direct hydrolysis over the succinimide pathway to the findings of Wakankar et al., which noted how more hydrophobic dielectrics led to decreased reactivity in the formation of the cyclic intermediate.[248]

Therefore, it is likely that, at a lower dielectric, the impact of the steric hindrance caused by the branched Val sidechain on succinimide formation would become more prevalent.[245] The preference of Asn-Asp towards direct hydrolysis is likely due to how, as Brennan and Clarke specified, the lower solvent dielectrics would result in side chain protonation and a preference towards direct hydrolysis.[293] While the side chain protonation was not directly modeled within our reaction network, a similar trend likely promoted the preference of Asn-Asp towards direct hydrolysis over succinimide at the hydrophobic dielectric.[293]

At a solvent $\varepsilon = 20$, eight of the 20 dipeptides were found to prefer direct hydrolysis over succinimide formation. As we previously specified, a prior study on the impact of solvent dielectrics on Asn deamidation noted that the reactivity of the cyclic intermediate formation decreased with ε.[248] Additionally, through our thermochemical and Arrhenius parameterization, the direct hydrolysis reaction was found to be most stable at a dielectric similar to that of charged residues.[212, 213] Of the two dipeptides found to prefer direct hydrolysis solely at $\varepsilon = 20$, Asparaginyl-Tyrosine (Asn-Tyr) contains a functional group capable of hydrogen bonding.[239] Because Asn-Tyr is capable of acting as both a hydrogen bond acceptor and donor, it is likely that the 'acidic' environment of the solvent dielectric promoted hydrogen bonding between the N-terminus Asn and C-terminus Tyr.[212, 213, 239] The propensity of Asn-Leu towards direct hydrolysis over the formation of the succinimide intermediate is likely due to the similarity of the Leu sidechain to the Val sidechain, where the branched functional group inhibits the formation of the succinimide intermediate at a solvent dielectric equal to the dielectric of charged residues.[245] Therefore, due to this factor as well as the solvent environment in which direct hydrolysis is a more competitive reaction mechanism, Asn-Leu deamidates via direct hydrolysis over the formation of the succinimide intermediate.[212, 213]

2.4.iii.b. Succinimide Hydrolysis into Asp-X versus Succinimide Hydrolysis into iso-Asp-X

In addition to our analysis of the competition between the two major pathways within our deamidation reaction network, direct hydrolysis and the succinimide intermediate, we also quantified the competition between the two products of succinimide hydrolysis (Table 2.4.6). Of the dipeptides which did not deamidate through the direct hydrolysis pathway, only the Asn-Ala succinimide intermediate was found to prefer hydrolysis into Asp-Ala at each dielectric. The propensity of Asn-Ala towards hydrolysis into Asp-Ala instead of iso-Asp-Ala can be attributed to the structural properties of the C-terminus residue. Ala is a hydrophobic, aliphatic, and very small molecule.[164, 239] Therefore, because prior studies have noted that the side chain of iso-Asp is more acidic than Asp under ambient conditions, it is likely that the Asn-Ala succinimide intermediate would prefer to hydrolyze into the least polar of the two product structures regardless of the solvent dielectric.[239, 300]

Table 2.4.6. Selectivity analysis of the two reaction pathways associated with the hydrolysis of the succinimide intermediate as a function of modeled dielectric.

ε	Asp-X	Iso-Asp-X
4	Asn-Ala, Asn-His	Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val
20	Asn-Ala	Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val
40	Asn-Ala	Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val
80	Asn-Ala	Asn-Arg, Asn-Asn, Asn-Asp, Asn-Gln, Asn-Glu, Asn-Gly, Asn-His, Asn-Ile, Asn-Leu, Asn-Lys, Asn-Met, Asn-Phe, Asn-Thr, Asn-Trp, Asn-Tyr, Asn-Val

At the most hydrophobic dielectric coefficient within our reaction network (ε = 4), Asn-His was also found to hydrolyze into Asp-His instead of iso-Asp-His. The preference of the Asn-

His dipeptide towards hydrolyzing into Asp-His is likely due the previously observed phenomena of succinimide reactivity decreasing with lower dielectric coefficients.[248] His has been noted in prior studies as a residue that promotes the formation of iso-Asp.[168, 301] Therefore, following the established trend that more hydrophobic dielectric coefficients influence succinimide reactivity, it is likely that it would promote hydrolysis into Asp-X at a dielectric equal to 4.

2.5. Conclusions and Future Directions

In summary, a theoretical investigation on mapping the competitive pathways of Asn deamidation was achieved for all Asn-X dipeptides over four solvent dielectrics. Rate coefficients and Arrhenius parameters for dipeptide deamidation were calculated for 308 (77 reactions per model dielectric) reactions using B3LYP and statistical thermodynamics.[82] Selectivity analysis indicated the formation of the succinimide intermediate as the proposed rate-limiting step, which was noted in experimental literature, and validated by the propensity of cyclization to act as a stabilizing force within folded proteins.[184, 305] Additionally, aside from the dielectric associated with charged residues, formation of the succinimide intermediate was greatly preferred over direct hydrolysis. At a dielectric equal to 20, the preference of dipeptides towards the direct hydrolysis pathway increased to about 40 percent, where the preference of protein deamidation towards direct hydrolysis under acidic or basic conditions likely influenced the reaction selectivity.[243] Furthermore, the most hydrophobic dielectric coefficient was found to influence the electronic energies of the succinimide formation reaction pathway, suggesting that the dielectric of the solvent influences the reactivity of the intramolecular reaction within the deamidation reaction network. Similar to the rate-limiting step, these observed trends matched with previous findings in experimental literature.[243] Future studies will evaluate deamidation at the secondary and tertiary structure, where prior studies have indicated that the secondary structure acts as a stabilizer against deamidation.[155, 167, 306-308] For this purpose, our findings observed at the primary structure would need to be evaluated in regards to higher order protein structures so that reaction selectivity could be correlated within different levels of the protein hierarchy.

Chapter 3: Towards Pharmaceutical Protein Stabilization: Mapping the Reaction Network of Gln-X Deamidation during Storage under Ambient Conditions

3.1. Background and Motivation

Protein degradation studies are of interest due to the wide array of applications proteins can be used for within the mAb industry, as well as in better understanding the mechanics behind protein-based diseases.[4, 82, 147, 309, 310] Therefore, it is necessary to quantify how changes in the protein environment may impact its stability during formulation, as well as within the human body. For particularly unstable amino acids, such as Asparagine (Asn) and Glutamine (Gln), where the net neutral residues are spontaneously converted into their acidic counterparts, Aspartic Acid (Asp) and Glutamic Acid (Glu), this form of degradation has been found to be prevalent in both of the previously specified scenarios.[80, 82, 147, 149, 151, 240, 244, 311, 312] Specifically, that the presence of ionized residues in the place of net neutral amino acids, the protein can either lose functionality or stability (Figure 3.1.1).[82, 128, 147, 240, 244, 311, 312]

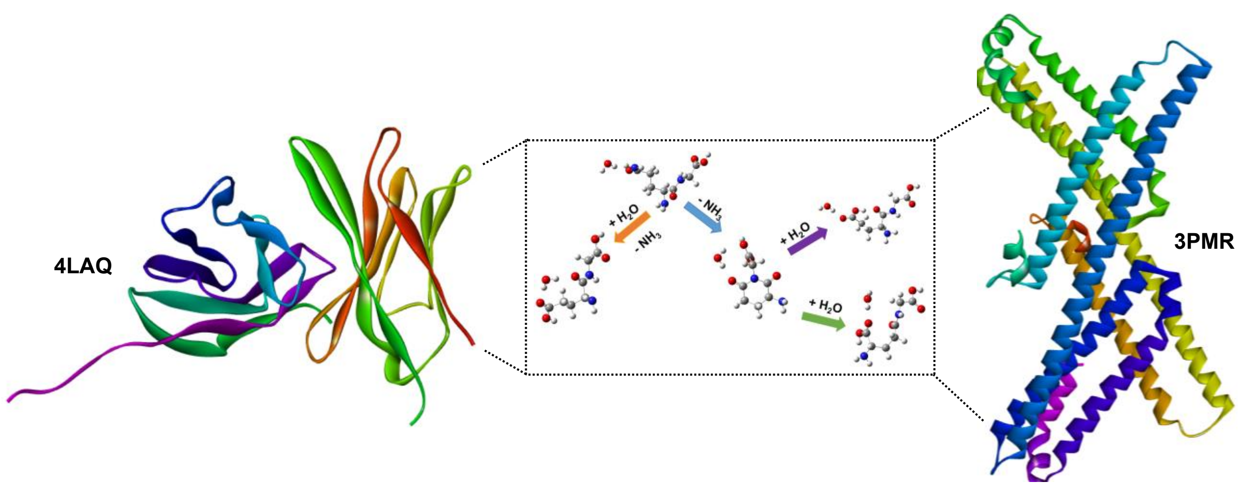

Figure 3.1.1. Graphical representation of the impact of Gln-X deamidation on therapeutic (4LAQ) and physiological (3PMR) proteins.[313, 314]

In therapeutic proteins, this often results in a loss of binding efficacy because deamidation reactions are most often observed in the region of the protein which targets and binds to a specific epitope.[82, 128, 129, 147, 244, 311, 312] Deamidation of proteins within the body often leads to protein misfolding which causes neurodegenerative diseases such as Huntington's and Parkinson's Disease.[82, 128, 147, 240, 244, 311, 312, 315] In particular, deamidation of Gln residues has been attributed to age-related ailments such as the formation of cataracts and collagen breakdown in bones, where the long-lived proteins are impacted.[311, 315] Therefore, it is critical to achieve a complete and unbiased understanding of the atomistic interactions within protein deamidation, as well as how those interactions deviate between the two residues.

For this purpose, we conducted a computational study, using Gln-X dipeptides, to predict the thermochemical and kinetic parameters of each reaction within the Gln deamidation reaction network, and compare those results against our generated Asn deamidation reaction network.[82, 147] To do so, a total of 77 (Appendix B) dipeptides were analyzed over four modeled dielectrics (308 total reactions) using the conductor-like polarizable continuum model (CPCM) to compute the thermochemical and kinetic parameters, i.e. standard enthalpies of reaction, standard entropies of reaction, Gibbs free energies of reaction, Gibbs free energies of solvation, activation energies, and kinetic rate constants for each dipeptide. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic energy, and then statistical thermodynamics was applied to all reactions to determine temperature effects on the reaction.[82, 141, 147, 188-190, 247, 253] All thermochemical properties were calculated at standard pressure and temperature to reduce external variabilities. Through this mechanistic study we have generated a self-consistent, unbiased database to predict the thermochemical and kinetic parameters of Gln-X deamidation, which can be analyzed against our previously generated Asn-X reaction network and expanded to account for other solvent-protein interactions in future theoretical studies.[82, 147] Furthermore, we have generated our database under four different modeled dielectrics, so that the reaction energetics and reaction selectivity could be studied as a function of the electrostatics within the modeled system.

3.2. Project Objectives

As we previously mentioned, degradation via deamidation significantly influences the binding strength of the developed therapeutic during formulation and storage. Additionally, by analyzing Gln-X deamidation in tandem with Asn-X deamidation, we can obtain insights towards the degradation behavior of long-lived therapeutic proteins.[316-318] For this purpose, we sought to develop a reaction network that would be representative of the different reaction pathways associated with deamidation. Furthermore, as the competition between reaction pathways can be used to optimizing the manufacturing pipeline GMP, we applied chemical engineering principles to identify the primary pathways of Gln-X deamidation. Overall, this study provides a self-consistent database of the thermochemical and kinetic parameters associated with the mechanisms of Gln-X deamidation under ambient conditions, as a function of the different dielectrics associated with amino acids, which can be expanded to account for other environmental effects in future studies.

3.3. Computational Methodology

3.3.i. Modeled Reaction Network and Chemical System

Development of our reaction network was modeled after the Asn-X reaction network under ambient storage conditions to reduce external variabilities within our analysis, and agreed with prior experimental studies.[80, 82, 128, 147, 149-151, 155, 169, 184, 240, 311, 319] As we have previously specified, deamidation is the spontaneous conversion of net neutral amino acids into their acidic counterparts.[80, 149, 151, 240, 244, 311, 312, 319] Additionally, protein deamidation occurs through two primary pathways, direct hydrolysis and the formation of a cyclic intermediate.[80, 128, 149-151, 155, 169, 184, 240, 311, 319] Of the two deamidation reaction networks, the understanding of Gln deamidation has been primarily studied in terms of the phenomena associated with Asn deamidation.[82, 147, 240, 244] While the two deamidation reaction networks share pathways and are both prone towards increased reaction rates at either an acidic or a basic pH, there is a lack in fundamental understanding towards why the rate of deamidation of Gln residues has been noted in prior studies to be slower and less reactive than Asn.[320]

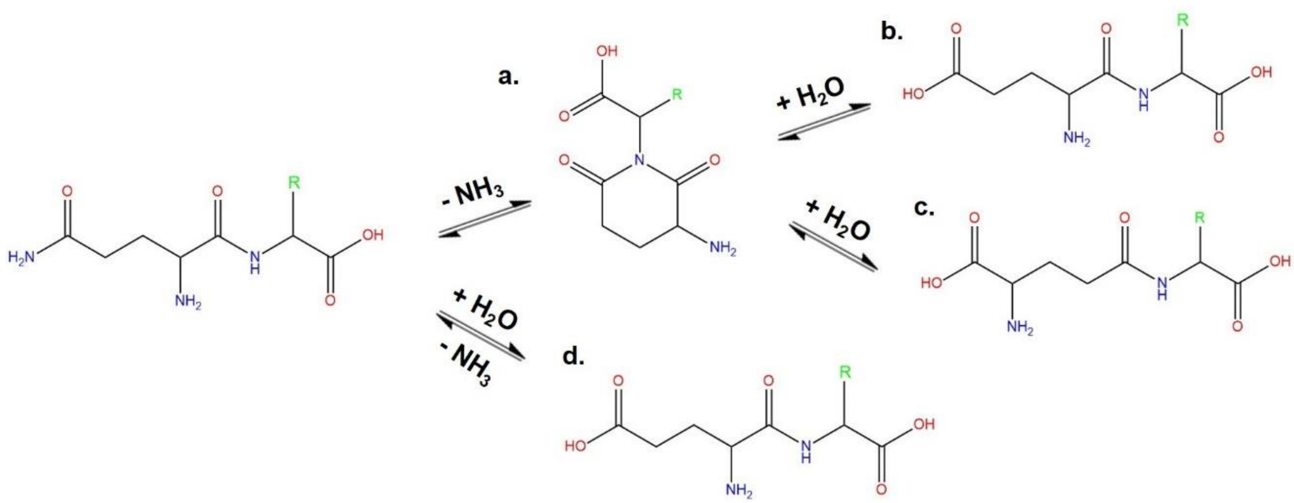

Figure 3.3.1. Reaction schematic of selected deamidation routes within the Gln-X deamidation reaction network. (a). Glutarimide formation, (b). Glutarimide hydrolysis into Glu-X, (c). Glutarimide hydrolysis into iso-Glu-X, (d). Direct hydrolysis.

Similar to the generation of the Asn-X deamidation reaction network, we modeled our reactions using a hybrid solvation model where one explicit water molecule acts as a passive stabilizer to our system.[82, 147] The implicit component of our chemical model applies the CPCM implicit model to the modeled reaction.[82, 141, 147, 209] The impact of the interactions between the different residue types and the solvated environment is reflected through the use of four different modeled dielectrics, ε = 4, 20, 40 and 80.[212, 213] Where the dielectrics equal to 4, 20, and 40 are associated with hydrophobic, charged and polar residues, respectively, while the modeled dielectric equal to 80 is that of pure water.[212, 213] By modeling these four modeled dielectrics, we were able to analyze how the interactions between the dipeptides and their environment impact the reaction energetics of deamidation. Similarly, through our mapping of the Asn-X deamidation reaction network, we determined that the conformation of the explicit water molecule relative to the dipeptide, as well as the number of water molecules present within the reaction system, would not significantly impact the energetics associated with Gln-X deamidation.[82, 147]

3.3.ii. Quantum Chemical Calculations and Statistical Thermodynamics

Quantum chemical calculations were performed with *Gaussian 16*.[82, 141, 147, 250, 251] All electronic energies for the protein dipeptides, intermediates and transition states were calculated using the B3LYP method.[82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized in the singlet state. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the reaction network.[82, 141, 147, 188-190, 246, 247, 253] To account for anharmonicity in the normal vibrational modes, as suggested by Scott and Random, the harmonic frequencies and zero-point energy (ZPE) were scaled by factors of 0.96 and 0.98 respectively.[82, 141, 147, 254] Following the work of previous reaction modeling completed by our group, partition functions based on the harmonic oscillator and rigid rotor approximations were used to calculate thermodynamic and kinetic properties as a function of temperature.[82, 141, 147, 255, 256] This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 257-263]

3.3.iii. Transition State Theory and Kinetic Parameters

Transition states for each specified reaction within the deamidation network were found using the Berny method.[82, 141, 147, 264-266] The imaginary frequency of each transition state was animated, and intrinsic reaction-coordinate calculations were performed to confirm that the normal vibrational mode pertained to the reaction coordinate of interest. A total of four reaction pathways – direct hydrolysis, glutarimide formation, glutarimide hydrolysis into Glu, and glutarimide hydrolysis into iso-Glu - were mapped for each chemical model used in this study. Conversion of intermediate structures to their final product form, as well as the ionization of N-terminus Glu residues under physiological conditions, were treated as equilibrated reactions and were not included in our computed rate coefficients.[239]

Conventional transition-state theory (TST) was then used to calculate rate coefficients at 1 atm, according to the macroscopic formulation in Equation (3.3.1).[82, 141, 147, 267]:

$$k^{TST}(T) = n_d \acute{k} = n_d \Lambda \exp\frac{\Delta S^{\neq}}{R} \exp(\frac{-\Delta H^{\neq}}{RT}) \qquad (3.3.1)$$

Where Λ is defined in Equation (3.3.2):

$$\Lambda = \kappa(T)\frac{k_B T (V_m)^{-\Delta n}}{h} \qquad (3.3.2)$$

Here, $\acute{k}$ is the single-event rate coefficient, $\kappa(T)$ is the Wigner tunneling correction at temperature T, k_B is Boltzmann's constant, h is Planck's constant, V_m is the molar volume, R is the ideal gas constant, $\Delta S^{\neq}$ is the entropy of activation, $\Delta H^{\neq}$ is the enthalpy of activation, Δn is the change in the number of moles going from the reactant to the transition state (i.e., -1 for bimolecular reactions such as for the direct hydrolysis reaction in both the forward and reverse directions, the glutarimide formation reaction in the reverse direction, and the glutarimie hydrolysis reaction in the forward direction; and 0 for unimolecular reactions such as the glutarimide formation reaction in the forward direction and the glutarimide hydrolysis reaction in the reverse direction), and n_d is the reaction path degeneracy, or number of single events. The reaction path degeneracy follows Equation (3.3.3). $\Delta H^{\neq}$ and $\Delta S^{\neq}$ are calculated using standard formulae, and there were no internal rotation corrections needed within our system.[82, 141, 147, 196, 267]

The reaction path degeneracy, n_d, definition is based upon the symmetry of the reactant(s) and the transition state.[82, 141, 147, 196] Symmetry was determined by the number of chiral centers present in each molecule.[82, 141, 147, 196, 239, 256, 263] To achieve an unbiased representation of each modeled reaction, the maximum symmetry numbers were applied to each reaction.[82, 141, 147, 239] Analysis of the reaction path degeneracies used in our model, noted that the inclusion of chiral centers did not impact the forward reaction path degeneracy developed through the exclusion of chiral centers. Therefore, our model may be applied to both natural and synthetic proteins in the forward direction of each computed reaction.

$$n_{\mathrm{d}} = \frac{n_{\mathrm{chiral}}^{\neq}}{\prod_{\mathrm{j}} n_{\mathrm{chiral}}^{\mathrm{reactant,j}}} \times \frac{\prod_{\mathrm{j}} \sigma^{\mathrm{reactant,j}}}{\sigma^{\neq}} \tag{3.3.3}$$

Where n_{chiral} is the number of chiral isomers for the transition state and reactant(s); σ is a product of the external rotation symmetry numbers for the transition state and reactant(s). Internal symmetry numbers for non-cancelling internal rotations treated as hindered or free rotations were included in the hindered and free rotor partition functions, respectively.

The single-event parameters of the Arrhenius relationship, log(A) and E_A were obtained by fitting ln(k) versus T^{-1} over the temperature range 298 – 1500 K. This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 255, 257-263] Through linear regression analysis, we determined

Arrhenius behavior was obeyed well for all reactions. The solvation impact of liquid phase kinetics is beyond the scope of this work but will be addressed in future studies following the work of Jalan et al, which computes liquid-phase kinetics in relation to both the solute and the solvent.[82, 141, 147, 268]

3.3.iv. Generalized Selectivity Model and Application to Peptide Deamidation

Elucidation of the difference in energetics between the Asn-X and Gln-X deamidation reaction network was completed using a two-sample t-test for the difference in means of the forward rate coefficients as a function of the modeled dielectric.[275] Statistical analysis was completed via a code developed by our group through the R statistical coding language.[274] By analyzing the forward rate coefficients of each modeled reaction within the reaction networks as a function of the modeled dielectric, so that trends shared between the two reaction networks could be studied. Additionally, by comparing the forward rate coefficients, we obtained insights regarding which reaction pathways were statistically significant from one another.

Reaction selectivity analysis was completed using chemical reaction engineering principles.[269] To obtain a robust analysis of our reaction network, we chose to compute the instantaneous selectivity for each reaction pathway within the deamidation reaction network.[269] To apply the instantaneous selectivity analysis to our reaction network, we first identified the proposed rate-determining step for each of the multi-step reaction pathways.[269] Once those rate-determining steps had been defined, we then applied Equation 3.3.4 to our analysis.[269]

$$\text{Instantaneous selectivity} = \frac{k'}{k''} \qquad (3.3.4)$$

Where k' is the desired reaction and k" is the undesired reaction. Selectivity was analyzed from 298.15 K to 400 K, the range of temperatures cited in prior deamidation studies, to quantify the impact of temperature on reaction selectivity.[152, 153] To determine if the results of our Gln-X deamidation model agreed with our Asn-X deamidation reaction network, as well as prior experimental findings, we treated the direct hydrolysis reaction pathway as the desired reaction while the glutarimide formation reaction was treated as undesired.[82, 147, 271, 272] Additionally, to fully quantify the impact of the modeled dielectric on the competing pathways of glutarimide hydrolysis, we conducted a secondary selectivity analysis upon the two products of the glutarimide pathway: Glu-X and iso-Glu-X.

3.3.v. Sensitivity Analysis and Generalized Model Validation

As we previously mentioned, the purpose of this study is to elucidate the differences between the Gln-X deamidation reaction network and the more widely studied Asn-X deamidation reaction network.[82, 147] Therefore, to reduce the bias associated between the two studies, we validated our computational methodology in terms of both Asn and Gln deamidation. Through prior studies, we had determined that the use of the B3LYP functional with the GD3 empirical dispersion correction was suitable for modeling our generated reaction networks.[82, 141, 147] Therefore, to maintain a self-consistent methodology through which the database of thermochemical and kinetic parameters associated with Gln-X deamidation could be analyzed against those generated for the Asn-X deamidation reaction network, we analyzed our modeled system under the same level of theory and chemical model as our prior studies.[82, 141, 147]

To further validate our proposed chemical model, we compared the computed values of the direct hydrolysis and glutarimide formation reactions for our proposed chemical model against two theoretical studies (Table 3.3.1).[244, 321] One of the direct hydrolysis reactions modeled by Halim et al. depicts the nucleophilic attack of water upon Gln to form its acidic counterpart, Glu and free ammonia.[244] Therefore, it was determined to be an acceptable metric to compare the average activation energies of direct hydrolysis at the modeled dielectric of pure water for our proposed chemical model.[244]

Table 3.3.1. Comparative analysis of theoretical literature activation barriers of direct hydrolysis and uncatalyzed glutarimide formation against the average activation barriers computed via the implicit + 1 H_2O solvation model at the modeled dielectric equal to pure water ($\varepsilon = 80$).[244, 321] Glutarimide formation activation energies for the solvation models were analyzed against the rate-determining step of the stepwise mechanism proposed by Kato et al.[321]

Reaction	Literature E_A (Theoretical)	Computed E_A (computed chemical model)	Difference (Theoretical – computed chemical model)
Units	kcal.mol^{-1}	kcal.mol^{-1}	kcal.mol^{-1}
DH	41.6	40.5	1.11
GF (RD)	48.8	50.6	-2.23

Our analysis of the activation barrier predicted by Halim et al. compared to our proposed chemical model indicated that the average activation barrier at $\varepsilon = 80$ was about 1.107 kcal.mol^{-1} lower than the barrier given by Halim et al. [244] The Kato study analyzes glutarimide formation as a stepwise mechanism, where the reaction can occur via uncatalyzed, water-catalyzed, phosphate-catalyzed, and carbonate-catalyzed approach.[321] While our study models the formation of the glutarimide intermediate as a concerted approach, which agrees well with prior experimental studies, the computational study completed by Kato et al. is the only known study to both model the formation of the glutarimide intermediate and analyze its reaction energetics via an uncatalyzed mechanism.[149, 151, 240, 279, 321] Therefore, we compared the average activation energy of glutarimide formation via our chemical model against the rate-determining step of the uncatalyzed glutarimide formation reaction in the Kato study (Table 3.3.1).[321] The average activation barrier of our concerted model at $\varepsilon = 80$ was found to be 2.23 kcal.mol^{-1} greater than the rate-determining barrier of the glutarimide formation mechanism provided by Kato.[321] As we previously noted, the glutarimide formation reaction in our study was modeled as a concerted mechanism. Therefore, due to concerted mechanisms generally yielding higher activation energies, it is likely that the small deviation between our computed barrier and the one computed by Kato is a result of the mechanism and not our chemical model.[321] Via this analysis, we determined that our computed activation barriers agree with the prior computational findings for uncatalyzed direct hydrolysis and glutarimide formation.[244, 321] Therefore, to reduce potential variabilities within our analysis, we confirmed that the chemical model used in our Asn-X deamidation study would be acceptable for modeling the Gln-X deamidation reaction pathway.[82, 147]

3.4. Results and Discussion

3.4.i. Structures

As we previously specified, four reaction pathways were analyzed to map the Gln-X deamidation reaction network. The naming conventions of the dipeptides followed the format of the Asn-X deamidation study, where Gln-Xaa denotes N-terminus Gln covalently bound to each of the amino acids analyzed in the reaction class. Additionally, only C-terminus residues that are net charged under ambient conditions- Arginine (Arg), Aspartic Acid (Asp), Glu, and Lysine (Lys)- were modeled as such.[239] However, to maintain the charge equilibrium between the intermediate and product states, N-terminus Glu was not treated as an ionizable residue and its ionization is beyond the scope of this study.[147, 239] Modeling of the direct hydrolysis reaction included two reactants (Gln-X and water, infinitely separated), the transition state, and two products (Glu-X and ammonia, infinitely separated), while formation of the glutarimide intermediate was modeled using the Gln-X dipeptide, the reaction transition state, and the two products (glutarimide-X and ammonia, infinitely separated). Lastly, hydrolysis of glutarimide occurred through two reactants (glutarimide-X and water, infinitely separated), the transition state, and either Glu-X or its isomeric form, Iso-Glu-X.

The direct hydrolysis reaction of Gln-X occurred via the same mechanism as the direct hydrolysis of Asn-X, where the nucleophilic attack of water eliminates ammonia from the Gln sidechain and converts the net neutral residue to its acidic counterpart, Glu-X.[82, 244] As depicted in Figure 3.4.1.a, the transition state structure depicts a decomposed water molecule, where the hydroxyl group is oriented to bond with the N-terminus Gln sidechain carbonyl and the hydrogen atom interacts with the amine to form free ammonia. Cyclization of Gln-X into the six-membered glutarimide ring occurs via the nucleophilic attack of the amide bond hydrogen upon the Gln sidechain to eliminate ammonia and form the cyclic intermediate. Of the 20 Gln-X dipeptides proposed within each reaction class, Glutaminyl-Proline (Gln-Pro) was found to be incapable of deamidation via the cyclic intermediate pathway.[151] The inability of the dipeptide to cyclize into glutarimide is similar to the Asn-X deamidation study, where the steric hindrance of the C-terminus Pro residue inhibits formation of the cyclic ring.[151]

Analysis of the reaction transition state structures (Figure 3.4.1.b), indicates that the glutarimide formation transition state peptide bond exists in the trans conformation compared to the cis conformation of the succinimide formation transition state.[322] As determined by Clarke, the conformation of the dipeptide indicates its favorability towards the formation of the cyclic intermediate.[273] Specifically, the distance between the sidechain carbonyl and the peptide bond Nitrogen, where a distance of 4.89Å was found to be kinetically unfavorable for the formation of the succinimide intermediate.[273] Our analysis of the transition state structure geometries indicated that the average distance between the Gln sidechain carbonyl and the peptide bond nitrogen was 4.730Å, a distance similar to the one determined to be energetically unfavorable by Clarke. Similarly, the average distance for the succinimide formation transition states was determined to be 2.513Å, a value closer to the distance found to be energetically favorable (1.890Å), where the deviation from the experimental values is likely due to the influence of environmental factors on the experimental peptide and the range of residues modeled within our study.[273] However, the ratio of computed distances between the two cyclic intermediate transition states was found to be about 2.2:1, while the distance of unfavorable to favorable distances in the Clarke study was 2.5:1.[273] As a result, it is likely that the conformation of the glutarimide formation transition state relative to that of the succinimide transition state will influence its reaction energetics.

Hydrolysis of the glutarimide intermediate occurs through the nucleophilic attack of water upon the glutarimide intermediate, where the hydrogen atom attaches to the carbonyl oxygen and the hydroxyl group attacks the carbonyl carbon (Figures 3.4.1.c and 3.4.1.d).[240, 321] Similar to the Asn-X deamidation reaction network, the transition states of both cyclic intermediate hydrolysis pathways were found to be gemdiol-like in nature, where the position of the gemdiol, relative to the amide bond, is dependent on the reaction pathway. Hydrolysis into Glu-X yields a transition state where the gemdiol forms at the N-terminus carbonyl, while hydrolysis into iso-Glu-X generates a gemdiol at the backbone carbonyl. As a result, the glutarimide ring opens to form either Glu-X or its isomeric form, iso-Glu-X.[240, 321]

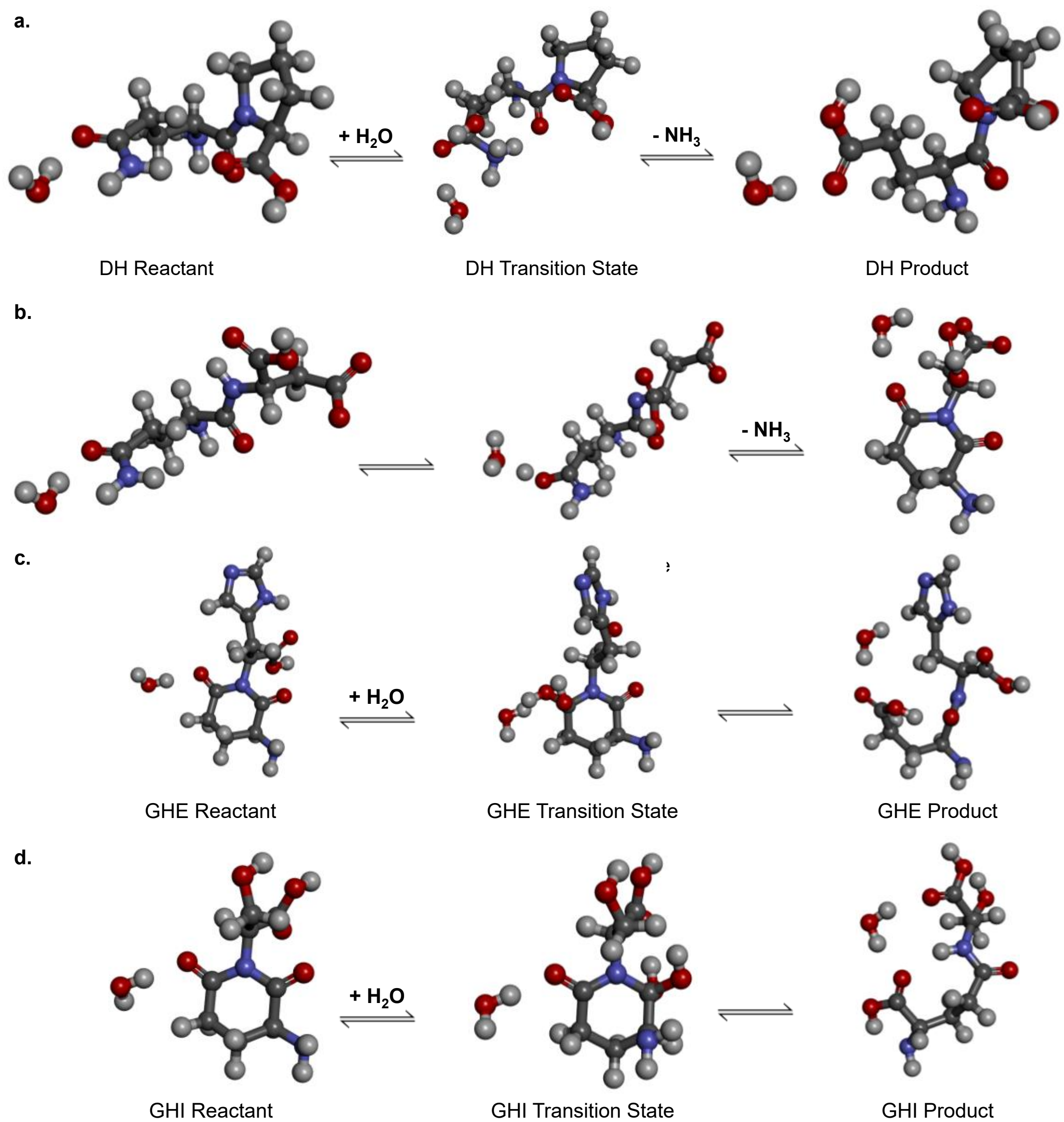

Figure 3.4.1. Representative ball and stick structures of the reactions used to generate the reaction network of Gln-X deamidation during storage under ambient conditions. (a). Direct hydrolysis reaction, (b). Glutarimide formation reaction, (c). Glutarimide hydrolysis into Glu-X reaction, (d). Glutarimide hydrolysis into iso-Glu-X reaction.

3.4.ii. Thermochemistry and Arrhenius Parameterization

To elucidate the differences between the reaction energetics of Asn-X and Gln-X deamidation reaction networks, we categorized and analyzed the computed thermochemical and kinetic parameters for each modeled reaction as a function of the modeled dielectric. In addition to the standard thermochemical values (Enthalpy of Reaction ($\Delta H_{RXN}^{298.15K}$), Gibbs Free Energy of Reaction ($\Delta G_{RXN}^{298.15K}$), and Entropy ($\Delta S_{RXN}^{298.15K}$)), Arrhenius parameters were computed in both directions. Analysis of the thermochemical and kinetic parameters associated with each reaction noted the following.[147]

3.4.ii.a. Direct Hydrolysis

The reaction energetics of the direct hydrolysis of Gln-X into Glu-X resembles that of Asn-X direct hydrolysis (Table 3.4.1). As we previously specified, the direct hydrolysis reaction of Gln-X deamidation occurs through the same reaction mechanism as Asn-X deamidation.[151, 323] Additionally, previous studies have hypothesized that the energetics associated with the direct hydrolysis of Gln residues would follow the trends determined for the mechanism associated with Asn residues.[149, 151, 240] However, as noted by Robinson et al., the competition between the two preliminary pathways is more prevalent in Gln-X deamidation compared to Asn-X deamidation.[80, 149] In our study, the increased favorability of the direct hydrolysis pathway is likely due to the impact of modeled dielectric instead of the associated thermochemical parameters. Where the modeled dielectric associated with charged residues, a system representative of the acidic and basic environments through which direct hydrolysis is the preferred reaction mechanism, is the most spontaneous.[80, 82, 147, 149, 168, 212, 240, 324]

Table 3.4.1. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for direct hydrolysis at 298 K and 1 atm. $\tilde{A}$ has units of $L.mol^{-1}.s^{-1}$ for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	**Forward Reaction**			**$\Delta H^{298.15}_{rxn}$**	**$\Delta G^{298.15}_{rxn}$**	**Reverse Reaction**		
	E_0	**Log $\tilde{A}$**	**E_A**			**E_0**	**Log $\tilde{A}$**	**E_A**
4	40.9	7.14	41.4	4.75	4.42	35.7	7.24	36.7
20	36.6	7.10	37.2	2.65	2.54	33.5	7.36	34.6
40	39.9	7.20	40.5	4.98	4.85	34.5	7.44	35.6
80	39.8	7.25	40.5	5.01	4.87	34.4	7.48	35.5

As depicted by Figure 3.4.2, the forward rate coefficient associated with the direct hydrolysis reaction is most thermodynamically and kinetically favorable at ε = 20. The forward Arrhenius parameters associated with the charged dielectric are $E_0 = 36.6 \pm 1.38$ $kcal.mol^{-1}$, $E_A = 37.2 \pm 1.58$ $kcal.mol^{-1}$, and log(A) = 7.09 ± 0.93 $L.mol^{-1}.s^{-1}$, while the reverse Arrhenius parameters were computed as $E_0 = 33.5 \pm 2.15$ $kcal.mol^{-1}$, $E_A = 34.6 \pm 2.43$ $kcal.mol^{-1}$, and log(A) = 7.36 ± 1.01 $L.mol^{-1}.s^{-1}$, respectively. The greatest Arrhenius parameters were computed at the most hydrophobic modeled dielectric where $E_0 = 40.9 \pm 1.52$ $kcal.mol^{-1}$, $E_A = 41.4 \pm 1.65$ $kcal.mol^{-1}$, and log(A) = 7.14 ± 1.05 $L.mol^{-1}.s^{-1}$ in the forward direction, while the reverse Arrhenius parameters were computed as $E_0 = 35.7 \pm 2.97$ $kcal.mol^{-1}$, $E_A = 36.7 \pm 3.26$ $kcal.mol^{-1}$, and log(A) = 7.24 ± 1.22 $L.mol^{-1}.s^{-1}$, in the reverse direction respectively.

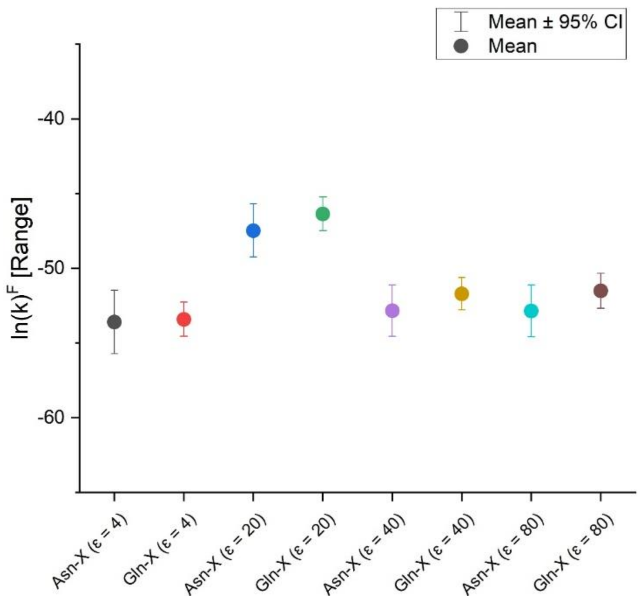

Figure 3.4.2. Interval plot of the difference in forward rate coefficients of the direct hydrolysis reaction associated with Gln-X deamidation relative to Asn-X deamidation.

Statistical analysis of the forward rate coefficients of the direct hydrolysis reaction determined that, while direct hydrolysis of Gln-X dipeptides was found to occur at a quicker rate relative to Asn-X dipeptides, the difference in means was found to be statistically insignificant.[82, 147] The increased forward rate coefficient of Gln-X direct hydrolysis relative to Asn-X direct hydrolysis is likely an artifact of the lack of competition between the two primary reaction pathways within the reaction network. Specifically, studies conducted by Capasso and Robinson suggested that the competition between the direct hydrolysis and cyclic intermediate formation reactions was more pronounced when Gln deamidates.[151, 320] The examination of Asn and Gln polypeptides by Robinson indicated that, while 95 percent of the analyzed Asn polypeptides deamidated into succinimide, only 82 percent of the Gln polypeptides deamidated into glutarimide.[151] Furthermore, the impact of the residues surrounding the Gln-X active site of deamidation could not be differentiated between the propensity of the polypeptides towards forming glutarimide versus deamidating via direct hydrolysis.[151] As a result, we hypothesized that a larger quantity of our 20 modeled dipeptide reactions would be more likely to deamidate via the direct hydrolysis reaction compared to the selectivity analyzed in our Asn-X reaction network.[82, 147]

3.4.ii.b. Glutarimide Formation

As we previously specified, the transition state structures of glutarimide formation relative to succinimide formation indicate that formation of the glutarimide intermediate would be kinetically unfavorable due to the distance between the N-terminus carbonyl and the amide bond nitrogen, agreeing with prior studies.[273, 319, 325] However, the thermochemical and kinetic parameterization regarding the unfavorability of the reaction is limited. Analysis of the modeled reaction indicated that, unlike succinimide formation, glutarimide formation is non-spontaneous in nature. The non-spontaneity of the reaction can be attributed to the stability of the glutarimide intermediate relative to that of succinimide, as denoted by Capasso et al.[320] Specifically, the impact of ring strain on the formation of the glutarimide intermediate relative to the succinimide intermediate. As denoted by Baeyer theory, the ideal bond angle of a cyclic ring is 108°, where deviation from the ideal bond angle results in a less stable intermediate.[326-328] While the average bond angle of succinimide formation is 108°, the bond angles of the glutarimide intermediate were found to be 118°, about 10° greater than the ideal bond angle. Therefore, formation of the glutarimide intermediate would not be as thermodynamically favorable compared to succinimide. Additionally, Glutaminyl-Isoleucine (Gln-Ile), Glutaminyl-Threonine (Gln-Thr), and Glutaminyl-Valine (Gln-Val) were found to be the most non-spontaneous reactions, regardless of modeled dielectric. The non-spontaneous nature of these dipeptides is likely due to a similar phenomena that was observed in the succinimide formation reaction, where the branched sidechains of the C-terminus residues causes a steric hindrance to the formation of the cyclic intermediate.[329]

A study conducted by Brennan and Clarke noted that the instability of the anionic peptide bond nitrogen at hydrophobic modeled dielectrics resulted in reduced favorability for the formation of the succinimide intermediate.[293] Our examination of the glutarimide formation reaction noted that the reaction became more thermodynamically and kinetically unfavorable as the modeled dielectric decreased (Table 3.4.2).[82, 147] Additionally, at $\varepsilon = 40$ and $\varepsilon = 80$, Glutaminyl-Asparagine (Gln-Asn) and Glutaminyl-Glutamine (Gln-Gln) were found to be the most kinetically favorable reactions. Prior studies have noted that Asn and Gln are relatively unstable compared to other residues under physiological conditions.[149, 151, 240] The instability of these residues suggests that they would be more prone towards the self-promoted hydrogen transfer of the amide bond nitrogen to eliminate ammonia and form the cyclic ring.[82, 147]

Table 3.4.2. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for glutarimide formation at 298 K and 1 atm. Ã has units of s^{-1} and $L.mol^{-1}.s^{-1}$ for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}{}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Reaction			$\Delta H^{298.15}{}_{rxn}$	$\Delta G^{298.15}{}_{rxn}$	Reverse Reaction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	54.8	10.2	54.1	10.3	2.23	44.6	7.40	45.3
20	51.7	12.8	52.0	9.66	2.04	42.0	10.3	43.7
40	50.8	12.7	51.2	9.71	1.99	41.1	10.1	42.8
80	50.6	12.8	51.0	9.69	1.98	40.9	10.2	42.7

Analysis of the forward and reverse Arrhenius parameters associated with glutarimide formation indicated that the impact of the modeled dielectric on the reaction followed similar energetics to succinimide formation.[82, 147, 293] At the modeled dielectric associated with pure water, $E_0 = 50.6 \pm 3.57$ $kcal.mol^{-1}$, $E_A = 51.0 \pm 4.05$ $kcal.mol^{-1}$, and $\log(A) = 12.8 \pm 1.19$ s^{-1}, while the reverse $E_0 = 40.9 \pm 4.40$ $kcal.mol^{-1}$, $E_A = 42.7 \pm 4.80$ $kcal.mol^{-1}$, and $\log(A) = 10.2 \pm 1.20$ $L.mol^{-1}.s^{-1}$, respectively. Consequently, at $\varepsilon = 4$, the forward Arrhenius parameters are computed at $E_0 = 54.8 \pm 6.79$ $kcal.mol^{-1}$, $E_A = 54.1 \pm 6.91$ $kcal.mol^{-1}$, and $\log(A) = 10.2 \pm 1.55$ s^{-1}, while the reverse Arrhenius parameters were computed as $E_0 = 44.6 \pm 8.17$ $kcal.mol^{-1}$, $E_A = 45.3 \pm 8.23$ $kcal.mol^{-1}$, and $\log(A) = 7.40 \pm 1.28$ $L.mol^{-1}.s^{-1}$, respectively. The larger deviation between the energetics associated with the modeled dielectric of pure water and the most hydrophobic modeled dielectric, relative to the succinimide formation reaction, is likely due to the proposed instability of the glutarimide intermediate (Figure 3.4.3).[82, 147] As Capasso noted, formation of the six-membered ring was found to be difficult to achieve, suggesting that the intermediate would be less stable compared to succinimide.[320] Therefore, it is likely that the stability of the glutarimide intermediate would be more significantly impacted than succinimide under hydrophobic modeled dielectrics.[82, 147, 293, 320]

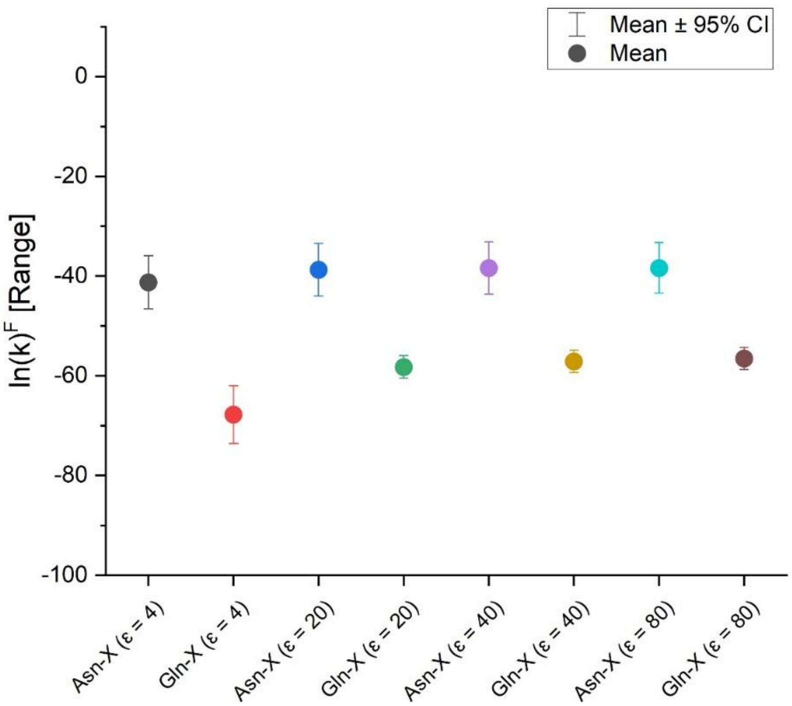

Figure 3.4.3. Interval plot of the difference in forward rate coefficients of the cyclic intermediate formation reaction associated with Gln-X deamidation relative to Asn-X deamidation.

Unlike the direct hydrolysis reaction, where the forward rate coefficients were found to be statistically insignificant of one another, formation of the succinimide intermediate was significantly quicker than the formation of the glutarimide intermediate regardless of the modeled dielectric. Additionally, both intermediate formation reactions were found to exhibit similar behavior in regards to the modeled dielectrics, where the forward rate coefficient was relatively similar from $\varepsilon = 80$ to $\varepsilon = 20$, and then significantly decreases at $\varepsilon = 4$.[82, 147] However, the forward rate coefficient associated with glutarimide formation at $\varepsilon = 4$ was found to decrease significantly relative to the other modeled dielectrics, suggesting that the combined energetic unfavorability of the formation of the six-membered ring and instability of the anionic nitrogen under low dielectrics, impacts glutarimide formation more significantly than succinimide formation.[82, 147, 293, 320]

Out of the four reaction pathways analyzed within each deamidation reaction network, the cyclic intermediate formation reaction was the only pathway found to exhibit statistically significant energetics between Asn-X and Gln-X.[82, 147] This phenomena is likely a consequence of the energetic unfavorability of the formation of the glutarimide intermediate, as first referenced by Capasso et al.[320] According to Capasso, as well as a study conducted by Robinson to compare the rates of Asn and Gln deamidation, deamidation was found to be less prevalent when Gln residues were present in the N-terminus relative to Asn residues.[151, 320] As we previously specified, the ratio of the relative distance between the sidechain carbonyl carbon of Gln and the peptide bond nitrogen between the Gln-X and Asn-X cyclic intermediate formation reactions was found to be 2.2:1, where the ratio identified by Clarke et al. was 2.5:1.[273] Furthermore, Robinson et al. determined that the formation of the cyclic intermediate through the deamidation of Gln-X polypeptides was less prevalent than that of Asn-X polypeptides.[149, 151] Therefore, it is likely that the non-spontaneous nature, as well as the higher activation barrier, of the glutarimide formation reaction relative to the formation of the succinimide intermediate is an artifact of the distance between the sidechain carbonyl carbon and the peptide bond nitrogen atoms.

3.4.ii.c. Glutarimide Hydrolysis into Glu-X

The hydrolysis of glutarimide into Glu-X was found to be exothermic and non-spontaneous in nature (Table 3.4.3), similar to the hydrolysis of Succinimide into Asp-X.[82, 147] Like the succinimide hydrolysis reaction, the associated entropy loss can be attributed to the transition of the system from disordered, formation of the cyclic ring, to ordered as the ring opens.[290-292] Hydrolysis of the glutarimide intermediate into Glu-X was found to be about 6 $kcal.mol^{-1}$ more exothermic and exergonic compared to succinimide hydrolysis into Asp-X. The increased exothermicity of the hydrolysis reaction relative to succinimide hydrolysis into Asp-X is likely due to the proposed instability of the glutarimide intermediate. As a result, the dipeptides with C-terminus residues found to inhibit the formation of the glutarimide intermediate –Gln-Ile, Gln-Thr, and Gln-Val- were found to be largely exergonic and exothermic relative to the other dipeptides within the reaction class. Similarly, hydrolysis of the Gln-Asn glutarimide intermediate was the only reaction found to be endothermic in nature. As we previously specified, glutarimide formation was most kinetically favorable when Asn was covalently bound to Gln. Therefore, it is likely that a greater amount of energy would be necessary to hydrolyze the intermediate into its final form.

Table 3.4.3. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for glutarimide hydrolysis into Glu-X at 298 K and 1 atm. Ã has units of $L.mol^{-1}.s^{-1}$ and s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Reaction			$\Delta H^{298.15}_{rxn}$	$\Delta G^{298.15}_{rxn}$	Reverse Reaction		
	E_0	Log Ã	E_A			E_0	Log Ã	E_A
4	15.2	7.68	15.9	-7.30	2.80	21.2	12.4	21.6
20	16.0	7.85	16.9	-6.18	3.59	21.1	12.3	21.4
40	16.0	7.76	16.9	-6.18	3.64	21.1	12.3	21.4
80	16.0	7.78	16.9	-6.12	3.64	21.1	12.2	21.4

The Arrhenius parameters associated with glutarimide hydrolysis into Glu-X were similar to those associated with succinimide hydrolysis into Asp, where the reaction was found to be most thermodynamically and kinetically favorable at the most hydrophobic modeled dielectric by about 1 $kcal.mol^{-1}$ and the energetics associated with other three modeled dielectrics were negligible (Figure 3.4.4).[82, 147] At ε = 20, $E_0 = 16.0 \pm 2.45$ $kcal.mol^{-1}$, $E_A = 16.9 \pm 2.59$ $kcal.mol^{-1}$, and log(A) = 7.85 ± 0.61 $L.mol^{-1}.s^{-1}$ in the forward direction, while $E_0 = 21.1 \pm 3.90$ $kcal.mol^{-1}$, $E_A = 21.4 \pm 4.21$ $kcal.mol^{-1}$, and log(A) = 12.3 ± 1.25 s^{-1} in the reverse direction. At the most hydrophobic modeled dielectric, $E_0 = 15.2 \pm 2.03$ $kcal.mol^{-1}$, $E_A = 15.9 \pm 2.05$ $kcal.mol^{-1}$, and log(A) = 7.68 ±

0.47 L.mol^{-1}.s^{-1} in the forward direction, while E_0 = 21.2 ± 4.23 kcal.mol^{-1}, E_A =21.6 ± 4.49 kcal.mol^{-1}, and log(A) = 12.4 ± 1.10 s^{-1} in the reverse direction, respectively. The larger deviation between reaction energetics at the most hydrophobic modeled dielectric and the other three modeled dielectrics can be attributed to the relative stability of the glutarimide intermediate compared to the succinimide intermediate, where the instability of the anionic peptide bond nitrogen at ε = 4 would further promote the hydrolysis of the glutarimide intermediate.[293, 329]

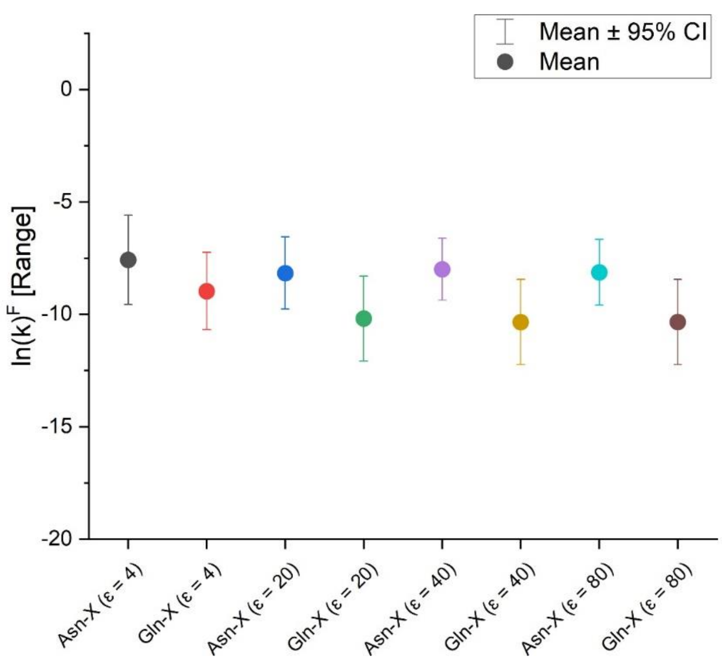

Figure 3.4.4. Interval plot of the difference in forward rate coefficients of the cyclic intermediate hydrolysis reaction into the acidic products associated with Gln-X deamidation relative to Asn-X deamidation.

Analysis of the forward rate coefficients of cyclic intermediate hydrolysis into Asp-X and Glu-X indicated that, similar to the cyclic intermediate formation reaction, the succinimide hydrolysis into Asp-X and glutarimide hydrolysis into Glu-X reactions were quickest at ε = 4.[82, 147] Furthermore, the mean difference in forward rate coefficients found that the difference in means between the forward rate coefficients was only statistically significant at ε = 40 (p = 0.042).

Of the four dielectrics modeled within our study, the dielectric associated with polar residues is most representative of the hydrolysis of cyclic intermediates. Specifically, due to the interactions of the water molecule with the carbonyl functional group of the cyclic intermediate. Therefore, it is likely that the increased forward rate coefficient of succinimide hydrolysis relative to glutarimide hydrolysis would be most prevalent at this modeled dielectric. Furthermore, the difference in computed means was found to be least statistically significant at the modeled dielectric associated with hydrophobic residues, and is likely an artifact of the instability of the cyclic intermediate at that dielectric. As a result, the propensity of glutarimide towards hydrolysis into Glu-X would exhibit similar reaction energetics to the hydrolysis of succinimide into Asp-X.[82, 147]

3.4.iii.d. Glutarimide Hydrolysis into iso-Glu-X

The hydrolysis of glutarimide into its isomeric form, iso-Glu-X, was found to exhibit similar thermochemical characteristics as the glutarimide hydrolysis into Glu-X pathway, where the most exergonic and exothermic reactions were observed for the dipeptides with Ile, Thr, and Val in the C-terminus and only Gln-Asn was found to be endothermic in nature (Table 3.4.4, Figure 3.4.5). Our analysis of the Arrhenius parameters associated with glutarimide hydrolysis into iso-Glu-X indicated that the reaction was most thermodynamically and kinetically favorable at $\varepsilon = 4$ relative to the other modeled dielectrics.

Table 3.4.4. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for glutarimide hydrolysis into iso-Glu-X at 298 K and 1 atm. $\tilde{A}$ has units of $L.mol^{-1}.s^{-1}$ and s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Reaction			$\Delta H^{298.15}_{rxn}$	$\Delta G^{298.15}_{rxn}$	Reverse Reaction		
	E_0	Log $\tilde{A}$	E_A			E_0	Log $\tilde{A}$	E_A
4	11.2	6.47	11.5	-6.70	3.41	16.6	11.2	16.5
20	12.1	6.69	12.5	-5.93	3.80	16.9	11.1	16.7
40	12.2	6.63	12.5	-5.96	3.81	16.9	11.1	16.8
80	12.2	6.60	12.4	-5.95	3.78	16.9	11.0	16.8

At the polar modeled dielectric, $E_0 = 12.2 \pm 4.46$ $kcal.mol^{-1}$, $E_A = 12.5 \pm 4.37$ $kcal.mol^{-1}$, and $\log(A) = 6.63 \pm 0.72$ $L.mol^{-1}.s^{-1}$ in the forward direction, while $E_0 = 16.9 \pm 4.92$ $kcal.mol^{-1}$, $E_A = 16.8 \pm 4.98$ $kcal.mol^{-1}$, and $\log(A) = 11.1 \pm 0.95$ s^{-1} in the reverse direction. Subsequently, the activation barrier associated with $\varepsilon = 4$ were found to be about 1 $kcal.mol^{-1}$ less than the values calculated at the other three modeled dielectrics, similar to the energetics exhibited for the glutarimide hydrolysis into Glu-X pathway. In the forward direction, $E_0 = 11.2 \pm 4.31$ $kcal.mol^{-1}$, $E_A = 11.5 \pm 4.18$ $kcal.mol^{-1}$, and $\log(A) = 6.47 \pm 0.52$ $L.mol^{-1}.s^{-1}$, while $E_0 = 16.6 \pm 4.20$ $kcal.mol^{-1}$, $E_A = 16.5 \pm 4.18$ $kcal.mol^{-1}$, and $\log(A) = 11.2 \pm 0.62$ s^{-1} in the reverse direction. Similar to the glutarimide hydrolysis into Glu-X reaction pathway, reaction favorability was greatest at $\varepsilon = 4$. Additionally, the activation energies of glutarimide hydrolysis were smaller when iso-Glu-X was the final product, regardless of modeled dielectric. Therefore, it is likely that the propensity of glutarimide towards iso-Glu-X mirrors that of succinimide towards iso-Asp-X.[82, 147, 240] Similar to the hydrolysis into Asp-X and Glu-X reaction pathway, our analysis of the forward rate coefficients indicated that hydrolysis was quickest at $\varepsilon = 4$ and statistically significant at $\varepsilon = 40$ (Figure 3.5.4). However, unlike the hydrolysis of the cyclic intermediate into the acidic form of Asn-X and Gln-X pathways, analysis of all modeled dielectrics indicated that the difference in means between the forward rate coefficients was more prevalent when the isomeric product was formed, regardless of the modeled dielectric. Specifically, it is likely that the propensity of both reaction networks towards the formation of iso-Asp-X and iso-Glu-X would resulted in a quicker rate coefficient associated with the Asn-X pathway relative to the Gln-X pathway, regardless of the modeled dielectric.[82, 147, 240, 320]

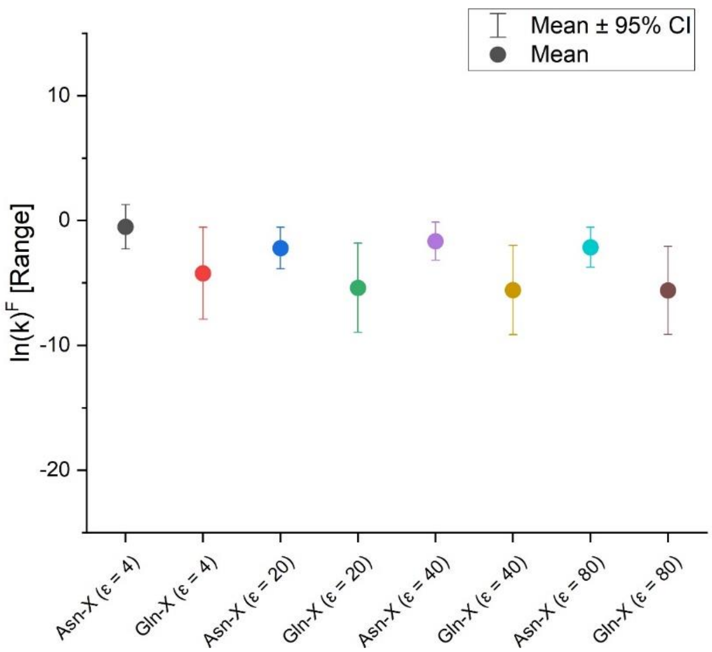

Figure 3.4.5. Interval plot of the difference in forward rate coefficients of the cyclic intermediate hydrolysis reaction into the isomeric products associated with Gln-X deamidation relative to Asn-X deamidation.

3.4.iv. Reaction Selectivity during Storage under Ambient Conditions

Analysis of reaction selectivity was achieved via instantaneous selectivity, where the proposed rate-determining step, glutarimide formation, was compared against the direct hydrolysis reaction.[147, 269] Secondary sensitivity analysis was applied to the glutarimide hydrolysis reaction to quantify the selectivity of glutarimide towards hydrolysis into either Glu-X or iso-Glu-X when only the cyclic intermediate pathway was considered. Analysis of the accepted temperature range for mAb manufacturing, where lyophilization does not occur, (298 K < T < 400 K) indicated that temperature did not influence reaction selectivity, an observation which was supported by the results of a study conducted by Bhanuramanand et al., where deamidation reactions were influenced at temperatures greater than 350 K.[302, 303] By analyzing reaction selectivity as a function of modeled dielectric, we were able to determine how the electrostatic interactions between the dipeptide and its solvated environment impacted its selectivity towards a particular reaction pathway.

Gln-X deamidation via the direct hydrolysis reaction pathway was found to be significantly more competitive than the glutarimide formation pathway regardless of modeled dielectric (Table 3.4.5). The propensity of the dipeptides towards direct hydrolysis agrees with prior findings and is likely due to the difficulty in forming the six-membered glutarimide ring as noted by Capasso, where the linear nature of the direct hydrolysis transition state would be more kinetically favorable.[240, 320] Favorability towards glutarimide formation over direct hydrolysis was observed for one dipeptide at $\varepsilon = 4$, the most hydrophobic modeled dielectric, and two dipeptides at $\varepsilon = 40$ and $\varepsilon = 80$, the modeled dielectrics associated with polar residues and pure water. Glutaminyl-Glutamic Acid (Gln-Glu) was found to prefer glutarimide formation over direct hydrolysis at the most hydrophobic modeled dielectric. Analysis of the transition state structure at this modeled dielectric indicates that the reaction most resembles the cyclic glutarimide at $\varepsilon = 4$, and as a result would be more kinetically favorable than the direct hydrolysis reaction of the dipeptide. At $\varepsilon = 40$ and $\varepsilon = 80$, Gln-Asn and Gln-Gln were found to prefer to form the glutarimide intermediate. As we previously noted, Asn and Gln are relatively unstable under ambient conditions compared to other amino acids.[149, 151, 240] Therefore, it is likely that these residues would be more prone towards the intramolecular hydrogen transfer at the amide bond nitrogen compared to other amino acids in the C-terminus position.[149, 151, 240] However, at the modeled dielectrics associated with hydrophobic and charged residues glutarimide formation would no longer be kinetically favorable for these dipeptides due to the instability of the anionic nitrogen at a hydrophobic modeled and the increased favorability of the direct hydrolysis reaction at $\varepsilon = 20$.[293]

Table 3.4.5. Selectivity analysis of the two primary pathways of Gln-X deamidation as a function of the modeled dielectric.

ε	Direct hydrolysis	Glutarimide formation
4	Gln-Ala, Gln-Arg, Gln-Asn, Gln-Asp, Gln-Cys, Gln-Gln, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Lys, Gln-Met, Gln-Phe, Gln-Pro, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val	Gln-Glu
20	Gln-Ala, Gln-Arg, Gln-Asn, Gln-Asp, Gln-Cys, Gln-Gln, Gln-Glu, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Lys, Gln-Met, Gln-Phe, Gln-Pro, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val	N/A
40	Gln-Ala, Gln-Arg, Gln-Asp, Gln-Cys, Gln-Glu, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Lys, Gln-Met, Gln-Phe, Gln-Pro, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val	Gln-Asn, Gln-Gln
80	Gln-Ala, Gln-Arg, Gln-Asp, Gln-Cys, Gln-Glu, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Lys, Gln-Met, Gln-Phe, Gln-Pro, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val	Gln-Asn, Gln-Gln

The preference of Glutaminyl-Glycine (Gln-Gly) towards the direct hydrolysis reaction was found to be contradictory towards prior findings, where the dipeptide was found to prefer cyclization into the glutarimide intermediate.[293] The deviation from experimental findings is likely due to environmental impacts that are beyond the scope of this study. Specifically, that the termination of our model system at two amino acids likely caused an increase in conformational flexibility relative to the peptides analyzed by Riggs and Robinson.[82, 147, 149, 151, 240] Therefore, cyclization of the dipeptide into glutarimide would not be influenced by its neighbors and likely reduces the modeled reaction favorability.[82, 147, 149, 151, 240]

Analysis of glutarimide hydrolysis selectivity indicated that the preferred product of glutarimide hydrolysis, regardless of the modeled dielectric, is iso-Glu-X (Table 3.4.6). At the modeled dielectric equal to pure water, the ratio of iso-Glu-X to Glu-X was about 3.8:1; agreeing with the findings of Riggs et al.[240] The preference of succinimide to hydrolysis into iso-Asp-X over Asp-X was attributed to the higher sidechain acidity of the isomeric product, therefore it is likely that a similar phenomenon occurs within the Gln-X deamidation reaction network.[82, 147, 298] Additionally, as the modeled dielectric decreased, the preference of glutarimide to hydrolyze into iso-Glu-X increased to about 4.8:1; where the selectivity of Gln-Ser shifted at $\varepsilon = 4$ and 20. The shift of the dipeptide towards iso-Glu-X under the hydrophobic and charged modeled dielectrics is likely due to the orientation of hydrolytic functional group towards the gemdiol transition state of the glutarimide hydrolysis into Glu-X reaction pathway under the modeled dielectrics. While the hydroxyl of the Ser functional group is oriented away from the gemdiol at $\varepsilon = 40$ and 80, the transition states at the hydrophobic and charged dielectrics indicate that the functional group is oriented towards the gemdiol. As a result, the dipole moment of the reaction at $\varepsilon = 4$ is about 2.76 times the magnitude of the reaction at $\varepsilon = 80$. Therefore, it is likely that the increased magnitude of the produced dipole moment shifts the selectivity of the reaction from preferring to hydrolyze into Glu-Ser to hydrolyzing into iso-Glu-Ser.[330]

Table 3.4.6. Selectivity analysis of the two reaction pathways associated with the hydrolysis of the glutarimide intermediate as a function of modeled dielectric.

ε	Glutarimide hydrolysis into Glu-X	Glutarimide hydrolysis into iso-Glu-X
4	Gln-Arg, Gln-Cys, Gln-Glu, Gln-Lys	Gln-Ala, Gln-Asn, Gln-Asp, Gln-Gln, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Met, Gln-Phe, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val
20	Gln-Arg, Gln-Cys, Gln-Glu, Gln-Lys	Gln-Ala, Gln-Asn, Gln-Asp, Gln-Gln, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Met, Gln-Phe, Gln-Ser, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val
40	Gln-Arg, Gln-Cys, Gln-Glu, Gln-Lys, Gln-Ser	Gln-Ala, Gln-Asn, Gln-Asp, Gln-Gln, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Met, Gln-Phe, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val
80	Gln-Arg, Gln-Cys, Gln-Glu, Gln-Lys, Gln-Ser	Gln-Ala, Gln-Asn, Gln-Asp, Gln-Gln, Gln-Gly, Gln-His, Gln-Ile, Gln-Leu, Gln-Met, Gln-Phe, Gln-Thr, Gln-Trp, Gln-Tyr, Gln-Val

3.5. Conclusions and Future Directions

In summary, a theoretical investigation on mapping the four selected reaction pathways of Gln deamidation was achieved for all Gln-X dipeptides over four modeled dielectrics. Rate coefficients and Arrhenius parameters for dipeptide deamidation were calculated for 308 (77 reactions per model dielectric) reactions using B3LYP and statistical thermodynamics.[147] Analysis of the reaction energetics relative to our prior study on Asn-X deamidation indicated that only the direct hydrolysis reaction exhibited similar reaction energetics, and the forward rate coefficients of the cyclic intermediate formation reaction were statistically significant from one another at all modeled dielectrics.[82, 147] Formation of the glutarimide intermediate was found to be thermodynamically and kinetically unfavorable relative to succinimide, where both reactions were found to be most unfavorable at the most hydrophobic solvent dielectric.[82, 147, 293] The larger activation barrier of glutarimide formation and its subsequent hydrolysis relative to the succinimide reaction pathway indicates that deamidation of Gln residues via this pathway is likely more prevalent in long-lived proteins; which is likely why Gln deamidation has primarily been connected to age-related diseases.[311, 315]

Lastly, selectivity analysis using chemical reaction principles indicates that Gln prefers to hydrolyze into Glu over cyclizing into glutarimide under ambient conditions. The preference of the dipeptides towards direct hydrolysis over glutarimide formation can be attributed to the distance between the N-terminus carbonyl and the peptide bond nitrogen, where prior studies had found formation of the cyclic intermediate to be kinetically unfavorable when the distance was near 4.89Å, where the conformation of the dipeptide inhibits the formation of the cyclic intermediate.[273] Additionally, the cyclic intermediate hydrolysis pathway was found to agree with prior experimental studies and our findings within the Asn-X deamidation reaction network, where hydrolysis into the isomeric form of Glu is preferred over Glu.[82, 147, 240] Similar to the reaction network developed in regards to Asn-X deamidation, the Gln-X deamidation reaction network needs to be evaluated at the secondary and tertiary levels of the protein hierarchy to evaluate the impact of conformation and interactions between sidechains on the selectivity of the modeled reactions towards deamidation via a specific reaction pathway

Chapter 4: Towards Pharmaceutical Protein Stabilization: Generalizing the Energetics of Non-Enzymatic Peptide Bond Fragmentation during Storage under Ambient Conditions

4.1. Background and Motivation

As we previously mentioned, studies on protein degradation under ambient conditions are of interest due to the global market share of mAbs. The drastic increase in the production of mAbs is likely due to their specificity for disease targets and a wide range of applications within the biomedical industry.[4-6, 8, 159, 331] However, the formulation efficiency of mAbs, is limited by the degradation of the desired products during multiple points of the biomanufacturing process.[332, 333] While the use of additives has significantly reduced degradation, determining the most effective additives is accomplished through a "guess and check" approach instead of understanding the underlying mechanism of the degradation pathway. One such mechanism of degradation is hydrolysis (Figure 4.1.1). Known as a primary route of covalent degradation of peptides, hydrolysis primarily occurs through catalysis by either an acid or base.[143, 334-338] While less common, hydrolysis can also occur non-catalytically. At a neutral pH, uncatalyzed polyamide hydrolysis occurs via the fragmentation of polypeptides into individual amino acids through the nucleophilic attack of water.[141, 143, 144, 165, 281, 336, 339] Specifically, the amide bond of the polypeptide backbone is targeted.[141, 338, 339] Fragmentation of the polypeptide often results in a lack of efficacy and stability, where the hinge region of the mAb is the primary site of degradation.[128, 141, 158, 337, 340] Therefore, it is critical to achieve a fundamental understanding the reaction energetics associated with non-enzymatic peptide bond fragmentation and how the dielectrics associated with the varying amino acid functionalities might influence those reaction energetics.

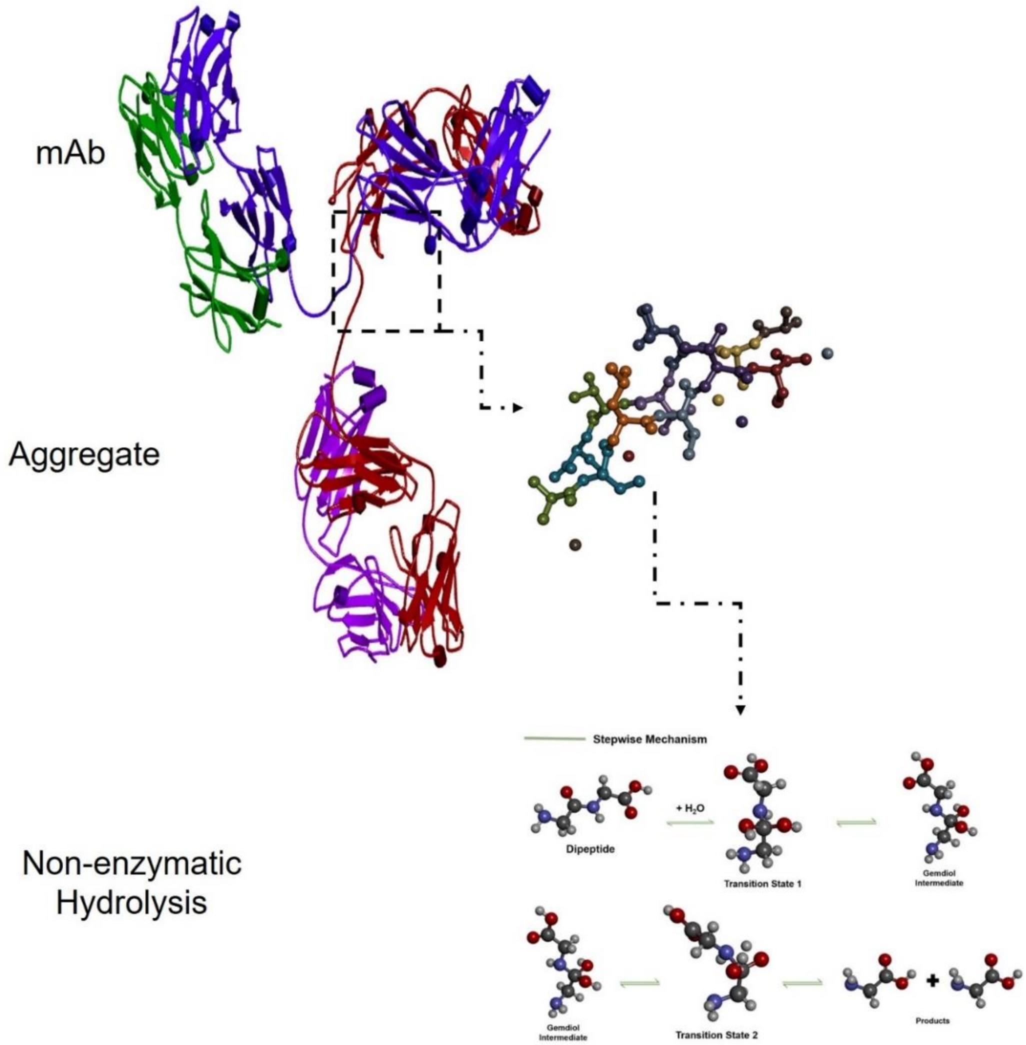

Figure 4.1.1. Multi-scale representation of the impact of non-enzymatic dipeptide hydrolysis upon mAbs using the 1IGT monoclonal antibody and the 6DJ0 Amyloid fibril.[128, 159, 341-343] Aggregation is a form of protein degradation that can be a result of other protein degradation mechanisms, such as non-enzymatic peptide bond fragmentation under ambient conditions, where the result is the formation of amorphous or amyloid-like fibrils as depicted by 6DJ0 that reduce the stability and function of the mAb.[343, 344]

For this purpose, we conducted a computational study, using a subset of amino acid dipeptides, to predict the thermochemistry and kinetic parameters of the different hydrolysis degradation mechanisms. This paper presents the thermochemical and kinetic properties of 168 (42 reactions (Appendix C) per modeled dielectric) non-enzymatic hydrolysis reactions in the conductor-like polarizability model (CPCM) solvated state, i.e. standard enthalpies of reaction, standard entropies of reaction, Gibbs free energies of reaction, activation energies, and kinetic rate constants. The subset of dimers used in this study was selected to provide a representation of how all amino acid functional groups would be affected by the hydrolysis reaction, as well as the modeled dielectric. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic energy, and then statistical thermodynamics was applied to all hydrolysis reactions to determine temperature effects on the reaction.[187-190, 246, 247] All thermochemical properties were calculated at standard pressure and temperature to reduce external variabilities.

4.2. Study Objectives

As we previously mentioned, degradation via fragmentation significantly influences the structural integrity of the developed therapeutic during manufacturing and storage. Additionally, similar to our deamidation studies under ambient conditions, we sought to predict the thermochemical and kinetic parameters associated with non-enzymatic peptide bond fragmentation in its most stable environment. For this purpose, we sought to develop a chemical model, of fragmentation via hydrolysis of the peptide bond, which would be representative of the kinetics provided by experimental literature. Furthermore, as there are 400 possible dipeptide combinations that can be subjected to peptide bond hydrolysis, we applied chemical engineering principles and statistical learning techniques to generalize the energetics of the proposed rate-determining step of the reaction. Overall, this study provides a self-consistent database of the thermochemical and kinetic parameters associated with non-enzymatic peptide bond fragmentation under ambient conditions, as a function of the different dielectrics associated with amino acids, which can be expanded to account for other environmental effects in future studies. Additionally, we have generated a statistically significant linear model through which the activation energies of non-enzymatic dipeptide hydrolysis may be predicted as a function of their enthalpies of reaction.

4.3. Computational Methodology

To reduce variability observed between the developed databases, the base methodology applied to non-enzymatic peptide bond fragmentation was self-consistent with the methodology used by the deamidation studies.[82, 141, 147, 251]

4.3.i. Modeled Mechanism and Chemical System

There is an ongoing debate regarding the correct mechanism of non-enzymatic polyamide hydrolysis under ambient conditions.[143, 144, 281, 338] Specifically, if fragmentation of the polyamide bond occurs via the nucleophilic attack of water or a hydroxide-catalyzed reaction under ambient conditions.[143, 144, 281, 338, 339] In this study, we modeled the uncatalyzed mechanism, where water acts as a reactant and is consumed during the reaction, to establish the reaction family concept via a neutral reaction pathway.[281, 339] Specifically, via a stepwise approach where the formation of the intermediate is the proposed rate-determining step (Figure 4.1.2).[281, 345, 346]

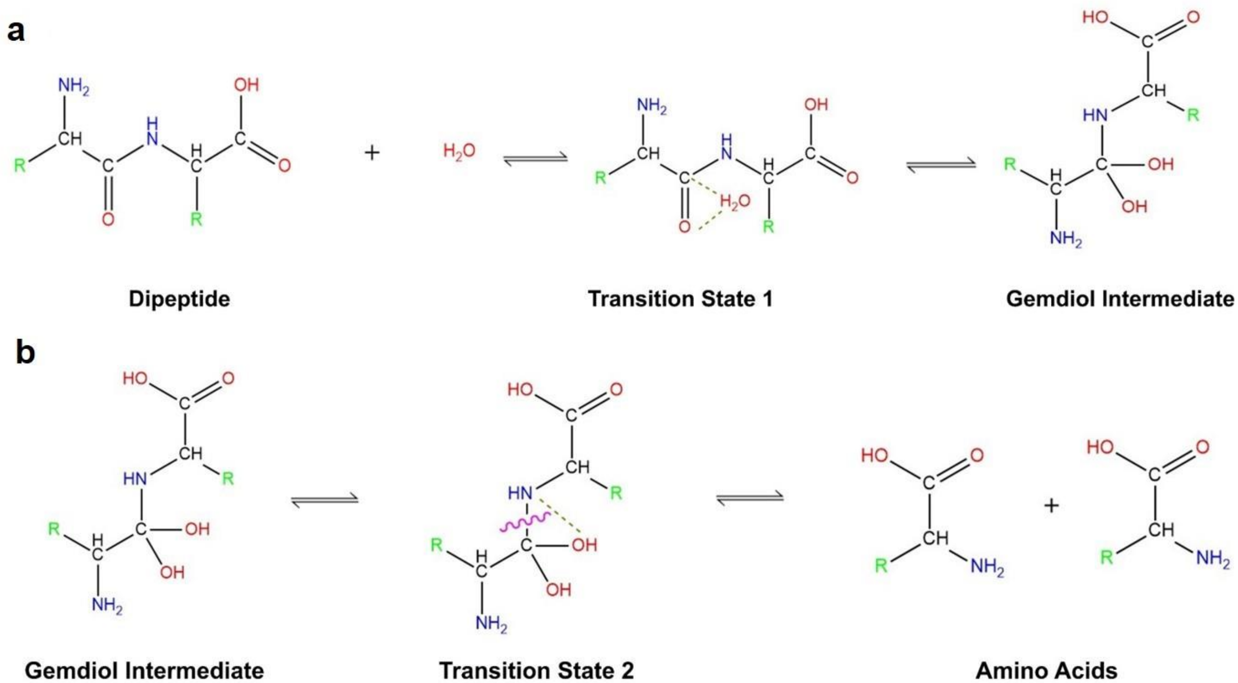

Figure 4.3.1. Reaction schematic of non-enzymatic dipeptide hydrolysis at a neutral pH via the stepwise mechanism. (a). The gemdiol formation step, (b). The proton transfer step.

Modeling of a neutral pH within our system was achieved by defining water as the solvent within the CPCM implicit solvation method, a solvation model that is similar to the COSMO-RS solvation model, as well as modeling the dipeptide net charge as a function of the residue functionalities.[82, 141, 147, 209, 239, 347] Specifically, the analyzed residues contained sidechains that were classified as either ionized or non-ionized at a pH = 7.[239, 347] In our modeled system, we analyzed the impact of uncatalyzed amide bond hydrolysis on dipeptides with cationic, neutral, or anionic sidechains to determine how the residue functional groups influenced the reaction energetics. Additionally, as we previously stated, we sought to determine how the energetics of non-enzymatic dipeptide hydrolysis changed with the modeled dielectric.

We modeled our system under four different dielectrics to account for the behavior of different residue sidechain functionalities, as well as the different regions of the protein.[82, 141, 147, 213] Modeling the reactions by using a spectrum of dielectrics allows for analysis of reaction energetics in relation to the residue functionalities and the portions of the protein where the reaction might occur. Specifically, recent modeling studies that analyze the energetics of regions of the proteins containing nonpolar residues or the energetics of reactions in the hydrophobic core are modeled using a relatively low dielectric constant ($\varepsilon = 4$).[212, 213, 221, 222] Similarly, surface protein studies, or studies that contain charged or polar residues, will have higher dielectric constants; $\varepsilon = 20$ for charged residues, and $\varepsilon = 40$ for polar residues, respectively.[212, 213, 216, 223] By modeling the dielectrics associated with the different residue functional groups, and different sites within the mAbs, we were able to obtain an unbiased representation of how the behavior of these functionalities influences non-enzymatic peptide bond fragmentation under ambient conditions.

4.3.ii. Quantum Chemical Calculations and Statistical Thermodynamics

Quantum chemical calculations were performed with *Gaussian 16*.[82, 141, 147, 250, 251] All electronic energies for the protein dipeptides, free amino acids, intermediates and transition states were calculated using the B3LYP method.[82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized in the singlet state. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the modeled reactions.[82, 141, 147, 188-190, 246, 247, 253] To account for anharmonicity in the normal vibrational modes, as suggested by Scott and Random, the harmonic frequencies and zero-point energy (ZPE) were scaled by factors of 0.96 and 0.98 respectively.[82, 141, 147, 254] Following the work of previous reaction modeling completed by our group, partition functions based on the harmonic oscillator and rigid rotor approximations were used to calculate thermodynamic and kinetic properties as a function of temperature.[82, 141, 147, 255, 256] This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 257-263]

4.3.iii. Transition State Theory and Kinetic Parameters

Transition states associated with both steps of the reaction mechanism were found using the Berny method.[82, 141, 147, 264-266] The imaginary frequency of each transition state was animated, and intrinsic reaction-coordinate calculations were performed to confirm that the normal vibrational mode pertained to the reaction coordinate of interest. Conventional transition-state theory (TST) was then used to calculate rate coefficients at 1 atm, according to the macroscopic formulation in Equation (4.3.1).[82, 141, 147, 267]

$$k^{TST}(T)= n_d\acute{\kappa} = n_d\Lambda \exp\frac{\Delta S^{\neq}}{R} \exp(\frac{-\Delta H^{\neq}}{RT}) \quad (4.3.1)$$

Where Λ is defined in Equation (4.3.2):

$$\Lambda = \kappa(T)\frac{k_B T(V_m)^{-\Delta n}}{h} \quad (4.3.2)$$

Here, $\acute{\kappa}$ is the single-event rate coefficient, κ(T) is the Wigner tunneling correction at temperature T, k_B is Boltzmann's constant, h is Planck's constant, V_m is the molar volume, R is the ideal gas constant, $\Delta S^{\neq}$ is the entropy of activation, $\Delta H^{\neq}$ is the enthalpy of activation, Δn is the change in the number of moles going from the reactant to the transition state (i.e., -1 for bimolecular reactions such as for the gemdiol formation step in the forward direction and the proton transfer step in the reverse direction; and 0 for unimolecular reactions such as the gemdiol formation step in the reverse direction and the proton transfer step in the forward direction), and n_d is the reaction path degeneracy, or number of single events. The reaction path degeneracy follows Equation (4.3.3). $\Delta H^{\neq}$ and $\Delta S^{\neq}$ are calculated using standard formulae, and there were no internal rotation corrections needed within our system.[82, 141, 147, 196, 267]

The reaction path degeneracy, n_d, definition is based upon the symmetry of the reactant(s) and the transition state.[82, 141, 147, 196] Symmetry was determined by the number of chiral centers present in each molecule.[82, 141, 147, 196, 239, 256, 263] To achieve an unbiased representation of each modeled reaction, the maximum symmetry numbers were applied to each reaction.[82, 141, 147, 239] Analysis of the reaction path degeneracies used in our model, noted that the inclusion of chiral centers did not impact the forward reaction path degeneracy developed through the exclusion of chiral centers. Therefore, our model may be applied to both natural and synthetic proteins in the forward direction of each computed reaction.

$$n_{\mathrm{d}} = \frac{n_{\mathrm{chiral}}^{\neq}}{\prod_{\mathrm{j}} n_{\mathrm{chiral}}^{\mathrm{reactant,j}}} \times \frac{\prod_{\mathrm{j}} \sigma^{\mathrm{reactant,j}}}{\sigma^{\neq}} \tag{4.3.3}$$

Where n_{chiral} is the number of chiral isomers for the transition state and reactant(s); σ is a product of the external rotation symmetry numbers for the transition state and reactant(s). Internal symmetry numbers for non-cancelling internal rotations treated as hindered or free rotations were included in the hindered and free rotor partition functions, respectively.

The single-event parameters of the Arrhenius relationship, log(A) and E_A were obtained by fitting ln(k) versus T^{-1} over the temperature range 298 – 1500 K. This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 255, 257-263] Through linear regression analysis, we determined Arrhenius behavior was obeyed well for all reactions. The solvation impact of liquid phase kinetics is beyond the scope of this work but will be addressed in future studies following the work of Jalan et al, which computes liquid-phase kinetics in relation to both the solute and the solvent.[82, 141, 147, 268]

4.3.iv. Generalized Model Validation

Model accuracy is dependent on both the level of theory and the chemical model applied to non-enzymatic peptide bond fragmentation associated with storage under ambient conditions. Confirmation of the chemical model that would most accurately represent non-enzymatic dipeptide hydrolysis at a neutral pH was completed through the comparative analysis of the energetics exhibited by the proposed rate-determining step of the stepwise mechanism against the metrics of uncatalyzed Glycyl-Glycine (Gly-Gly) hydrolysis given by Radzicka et al. at the dielectric of pure water ($\varepsilon = 80$).[143] Depicted in Table 4.3.1, the gemdiol formation reaction, which was found to be rate-determining for Gly-Gly hydrolysis, predicted the forward rate coefficient within one order of magnitude of the value presented by Radzicka et al.[143] Furthermore, the stepwise mechanism is consistent with experimental studies on non-enzymatic hydrolysis at acidic and basic pH ranges, suggesting that the mechanism of non-enzymatic peptide hydrolysis is conserved over pH.[142]

Table 4.3.1. Comparative analysis of the computed Arrhenius parameters and forward rate coefficients of non-enzymatic Gly-Gly hydrolysis of the proposed rate-determining step of the stepwise mechanism against the parameters quantified by Radzicka et al. [143] The rate coefficients were computed at 298 K and using a pseudo-first order reaction to match the parameters specified by Radzicka et al. (298 K, pseudo-first order reaction, rate coefficient determined via high-field NMR spectroscopy). [143] Evaluation of the Arrhenius pre-exponential factor, log(A), was derived via the Arrhenius equation by using the activation barrier and forward rate coefficient given by Radzicka et al. [143] While the correct units of the computed Arrhenius pre-exponential are $L.mol^{-1}.s^{-1}$, the pseudo-first order approximation was applied to reduce variability.

k^F [Experimental] (s^{-1})	**k^F [Computed] (s^{-1})**	**ΔError (s^{-1})**
6.30×10^{-11}	1.61×10^{-10}	3.92×10^{-01}
E_A [Experimental] ($kcal.mol^{-1}$)	**E_A [Computed] ($kcal.mol^{-1}$)**	**ΔError ($kcal.mol^{-1}$)**
23.0	22.0	1.00
log(A) [Experimental] (s^{-1})	**log(A) [Computed] (s^{-1})**	**ΔError (s^{-1})**
6.67	6.21	0.46

Furthermore, the use of the B3LYP functional is common among protein studies and was found to be reliable for modeling protein reaction mechanisms.[182-184] To validate the reliability of our computational methodology, we analyzed our computed forward rate coefficients of Gly-Gly hydrolysis at the B3LYP level of theory against a secondary study on non-enzymatic peptide bond hydrolysis at a neutral pH. A recent study completed by Sun et al. predicted the forward rate coefficients of Gly-Gly hydrolysis as a function of pH and at a higher temperature than our model system (T = 368 K).[142] By interpolating the forward rate coefficients provided by our Arrhenius trend at T = 300 K and T = 400 K, we were able to compute the forward rate coefficient of Gly-Gly hydrolysis to be 8.90×10^{-6} s^{-1}. The comparative analysis of our computed rate coefficient against the one provided by Sun et al. ($k^F = 1.40 \times 10^{-7}$ s^{-1}), indicates that modeling non-enzymatic peptide bond hydrolysis under this level of theory will yield values within one order of magnitude of the experimental parameter.[142] Therefore, to maintain the self-consistency of our base computational methodology, we accepted the DFT/B3LYP/6-31+G(2d,2p) with GD3 empirical dispersion correction as the level of theory through which the stepwise mechanism would be modeled.[82, 141, 147, 251]

4.3.v. Generalized Reaction Models and Applications towards Non-Enzymatic Peptide Bond Fragmentation

In addition to predicting the thermochemical and kinetic parameters associated with the stepwise mechanism of non-enzymatic dipeptide hydrolysis at a neutral pH, we also sought to determine if a linear correlation could be used to generalize the reaction energetics associated with each reaction in the mechanism. The Evans-Polanyi correlation is a common model which generalizes a linear relationship between the activation barrier (E_A) and the reaction enthalpy (ΔH_{RXN}).[141, 224] From this correlation, thermochemical parameters of reactions within a specified reaction mechanism can be estimated. Specifically, that the pre-exponential factor of the Arrhenius equation (log(A)) is the same for all reactions in a given reaction mechanism.[141, 224] Generalization was achieved by plotting the enthalpy of reaction (ΔH_{RXN}) versus the activation barrier (E_A). For the two reactions associated with the stepwise mechanism, we sought to achieve an Evans-Polanyi correlation. Using these correlations, dipeptide activation energies can be predicted using Equation 4.3.4:

$$E_A = E_0 + \alpha \Delta H_{RXN} \qquad (4.3.4)$$

To examine the strength of our Evans-Polanyi correlation, statistical bootstrapping was then applied to the correlation using an R script developed by our group.[141, 274] Bootstrapping is a statistical sensitivity method that is used to examine the accuracy of a given model by randomly resampling the data set and then generating a new linear correlation from the smaller, resampled representation of the original correlation.[141, 225-227, 230-232, 348] The newly generated correlation can then be used to produce replacement values for the randomly, removed data points, which in turn are implemented to analyze the prediction error associated with the model. [141, 228, 229, 231] Specifically, the difference between the replacement and initial data points is indicative of the prediction error associated with the model, where the data points that are randomly removed from the sample to generate the bootstrapped correlation act as the population the correlation would be applied to.[141, 231] While the size of the bootstrapped population can vary based on the analyzed system, we chose to randomly remove 20 percent of the data points present in our initial correlations to generate the new trends that would allow us to analyze the goodness of fit within our models.[141, 225-227, 232] These correlations were then used to predict the values of the removed data points, where the associated error between the predicted and actual values indicates the predictiveness of the correlation.[141, 225-227, 232] Through this analysis, we were able to achieve a statistically significant Evans-Polanyi correlation and apply statistical sensitivity methods to test the strength of our generated model.

4.4. Results and Discussion

4.4.i. Structures

A subset of 21 dipeptides was used to model the reaction behavior of different dipeptide functionalities upon non-enzymatic peptide bond fragmentation under ambient conditions via the stepwise mechanism.[141] Specifically, small and large residues, aromatic residues, hydrolytic residues, charged residues, and aliphatic residues were included within our reaction subset. By including each of these residue types in both the N and C-terminus positions of our modeled dipeptide, we were able to analyze how the amino acid functionalities impacted the reaction energetics of each hydrolysis mechanism, and how the change in modeled dielectric influenced the different dipeptide functionalities.

The stepwise mechanism was modeled using two reactants, the dipeptide and water infinitely separated, the transition state used to form the gemdiol intermediate, the gemdiol intermediate, the transition state used to cleave the central amide bond through a concerted proton transfer, and the two individual amino acids that comprised the reactant dipeptide, infinitely separated. All dipeptides exhibited the same transition state behavior, regardless of the modeled dielectric. The first transition state depicts the nucleophilic attack of water to hydrolyze the dipeptide, where the attacking water decomposes so that the hydroxyl group attaches to the carbonyl carbon of the peptide bond and the free-floating hydrogen attaches to the carbonyl oxygen, thus forming the gemdiol intermediate. The gemdiol intermediate is then subjected to a concerted proton transfer, where the proton exists halfway between the donating hydroxyl functionality of the gemdiol intermediate and the peptide bond amine group through which the central amide bond is cleaved. Once the central amide bond has been cleaved, the two amino acids that had comprised the dipeptide exist separate from each other within the solution. A representative ball and stick schematic of the stepwise mechanism is present in Figure 4.4.1.

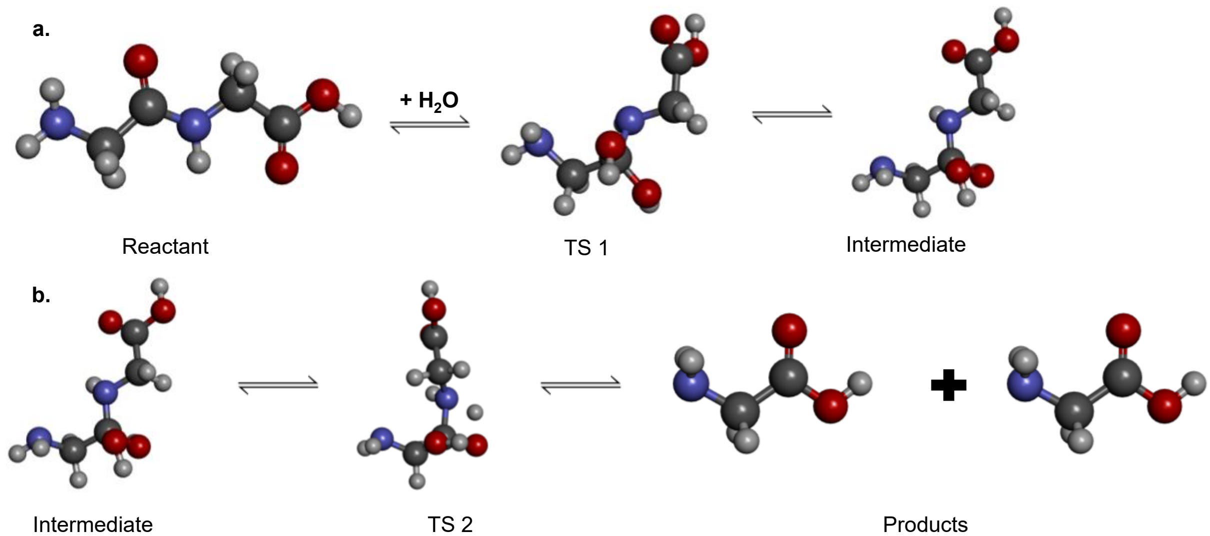

Figure 4.4.1. Ball and stick representation of the structures used to model the stepwise mechanism of non-enzymatic dipeptide hydrolysis (a) gemdiol formation step, (b). proton transfer step.

4.4.1.a. Sequence Impact on Reaction Energetics

The sequence of the two residues which comprise the dipeptides modeled within this study can greatly impact its energetics due to the associated steric hindrances of sidechain functionalities within the transition state structure, as well as the dipeptide status as either physiological or synthetic. Of the 21 modeled dipeptides within this study, three were found to be synthetic. Classification as either physiological or synthetic was determined by examining the sequences of five pharmaceutical protein structures.[159, 342, 349-352] Dipeptides which occurred at least once within each structure were classified as physiological, while dipeptides that did not appear within the structures were labeled as synthetic. Via this analysis, it was determined that Arginyl-Lysine (Arg-Lys), Asparatyl-Arginine (Asp-Arg), and Tyrosyl-Glycine (Tyr-Gly) were synthetic. The classification of Asp-Arg and Tyr-Gly as synthetic indicates the importance of sequence structure on reaction energetics because their counterparts, Argininyl-Aspartic Acid (Arg-Asp) and Glycyl-Tyrosine (Gly-Tyr), were found to be physiological. As a result, the energetics of the two dipeptides are likely to deviate from one another.

Similarly, the transition state structures of each of the modeled dipeptides will differ based on which residue is in the N-terminus position compared to the C-terminus position. As we previously mentioned, water either decomposes or attaches to the central amide bond based on the modeled mechanism. For dipeptides where small, flexible residues are in the N-terminus position, the transition state is likely to be more kinetically favorable compared to dipeptides with large, bulky functionalities due to potential steric hindrances.

4.4.ii. Thermochemistry and Arrhenius Parameterization

We sought to analyze how the gemdiol formation and proton transfer reactions behaved as a function of dielectric.[141] Through this analysis, we noted that the stepwise mechanism was found to be both net endothermic and net endergonic in nature, where the endothermicity and non-spontaneous nature of the reaction mechanism can be attributed to its proposed rate-determining step, the formation of the gemdiol intermediate. The presence of the gemdiol intermediate as the proposed rate-determining step agrees well with other studies on amide hydrolysis, where energy is expended upon the system via the nucleophilic attack of water to form the gemdiol intermediate prior to the subsequent, spontaneous proton transfer that cleaves the dipeptide into its associated amino acid substituents.[281, 353] The spontaneity of the proton transfer step can be attributed to the entropic penalty exhibited by the formation of the gemdiol intermediate, which is reflected in its non-spontaneous nature[353] Specifically, that the formation of the gemdiol intermediate through the nucleophilic attack of water likely reduces the configurations of the dipeptide available for hydrogen bonding within the system, and as a result causes a loss of entropy.[354-357] At the dielectric associated with hydrophobic residues ($\varepsilon = 4$), this phenomena is not as significant, likely due to the hydrophobic interactions between the dipeptide and modeled dielectric compared to interactions at other dielectrics.[358] However, when the transfer of a proton from the gemdiol to cleave the central amide bond is initiated, entropy is gained due to an increase in conformational flexibility within the system as a result of the fragmented peptide bond.

4.4.ii.a. Gemdiol Formation Step

The average thermochemical and Arrhenius parameters of the gemdiol formation step of the stepwise mechanism are given in Table 4.4.1 as a function of the modeled dielectric. Analysis of the average Arrhenius parameters indicated that the reaction was least kinetically favorable at the dielectric associated with polar residues. At $\varepsilon = 40$, $E_0 = 21.4 \pm 2.73$ kcal.mol^{-1}, $E_A = 21.9 \pm 2.87$ kcal.mol^{-1}, and log(A) = 6.27 ± 1.02 L.mol^{-1}.s^{-1}, in the forward direction, and $E_0 = 3.13 \pm 1.77$ kcal.mol^{-1}, $E_A = 3.43 \pm 1.86$ kcal.mol^{-1}, and log(A) = 12.7 ± 0.74 s^{-1}, in the reverse direction. The 0.200 kcal.mol^{-1} is likely a result of the change in activation energy exhibited by Arg-Asp and Tyr-Gly as a function of the modeled dielectric. While the activation barrier of Arg-Asp was found to increase as the modeled dielectric decreased, the proportional decrease in the activation energy of Tyr-Gly gemdiol formation was not exhibited at $\varepsilon = 40$. Specifically, the activation barrier decreased by 0.11 kcal.mol^{-1} at $\varepsilon = 40$, while it decreased by 5.30 kcal.mol^{-1} at $\varepsilon = 20$ and 1.30 kcal.mol^{-1} at $\varepsilon = 4$. Therefore, the increase in activation energy of Arg-Asp gemdiol formation would not be as pronounced at $\varepsilon = 4$ and $\varepsilon = 20$.

Table 4.4.1. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for the gemdiol formation reaction of non-enzymatic dipeptide hydrolysis at 298 K and 1 atm. Ã has units of L.mol^{-1}.s^{-1} and s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of kcal.mol^{-1}.

ε	Forward Reaction			ΔH_{RXN}	ΔG_{RXN}	Reverse Reaction		
	E_0	log(Ã)	E_A			E_0	log(Ã)	E_A
4	21.2	6.13	21.5	15.6	28.2	3.34	12.5	3.54
20	21.1	6.35	21.5	16.0	28.6	2.87	12.8	3.19
40	21.4	6.27	21.8	16.1	28.6	3.13	12.7	3.43
80	21.1	6.13	21.5	15.7	28.3	3.14	12.6	3.44

Our analysis of the overall trends associated with the average forward Arrhenius parameters indicated that as the modeled dielectric decreased, the activation energy also decreased. Of the 21 dipeptides modeled within the reaction, three reactions were found to deviate from the observed behavior – Arg-Asp, Gly-Tyr, and Tyr-Gly. The deviation of Gly-Tyr was observed at $\varepsilon = 40$, while the Tyr-Gly activation barrier was found to be smallest at $\varepsilon = 20$. The decreased reaction favorability of Gly-Tyr at $\varepsilon = 40$ is likely a result of polar-aromatic interactions.[359] A study conducted by Pace and Gao indicates that polar-aromatic interactions resulted in more favorable protein folding.[359] As the transition state of Gly-Tyr gemdiol formation at $\varepsilon = 40$ is relatively linear than nature, it is likely that the polar-aromatic interactions exhibited at the model dielectric would result in a less favorable reaction. The decrease in activation energy exhibited by Tyr-Gly at the dielectric associated with charged residues is likely representative of the increased stability of the transition state through the interactions of the N-terminus functional group with its environment. Specifically, the dielectric of $\varepsilon = 20$ resembles the dielectric associated with charged residues, and cation-π interactions have been noted as stabilizing electrostatic interactions between charged and aromatic residues.[212, 213, 360, 361]. Therefore, the reaction energetics of Tyr-Gly gemdiol formation at $\varepsilon = 20$ would resemble the reaction favorability of Gly-Tyr, where the flexible residue Gly is in close proximity to the reactive center.

An inverse relationship between the modeled dielectric and the activation barrier, as well as the Gibbs Free Energy and Enthalpy of reaction, was observed for Arg-Asp gemdiol formation. The decrease in reaction favorability with modeled dielectric is likely a consequence of the electrostatic interactions the dipeptide experienced, as well as the position of the residues. Specifically, Asp-Arg, which contains the same residue functional groups as Arg-Asp but contains the acidic Asp in the N-terminus position, was found to observe the overall trend. Additionally, as the modeled dielectric decreased, the formation of the Asp-Arg gemdiol was found to be more exothermic in nature, the inverse of the energetics exhibited by Arg-Asp. The deviance between the reaction energetics of these two dipeptide reactions is likely due to the proximity of the N-terminus functional group to the reactive center. Specifically, a study conducted by Xie et al. indicated that strong salt-bridge formation between the Arg functional group and a carboxylic acid was observed at distances less than 2.00Å.[361] In our modeled transition states the minimum distance between the reactive center and the Arg functional group was observed at 4.763 Å at the dielectric associated with hydrophobic residues. Therefore, it is likely that the increase in endothermicity and activation energy at $\varepsilon = 20$ and $\varepsilon = 40$ is a result of the residues interacting with the environment instead of stabilizing the transition state.

Furthermore, the significant decrease in reaction favorability exhibited by Arg-Asp at $\varepsilon = 4$ is likely a consequence of the geometry of its transition state in addition to the electrostatic interactions the dipeptide experiences. The backbone dihedral angle of the Arg-Asp transition state at $\varepsilon = 4$ was found to deviate significantly from the backbone dihedral angles computed at the other 3 modeled dielectrics. Similarly, while the transition states at $\varepsilon = 20$, 40, and 80 were found to be relatively linear, where the Arg and Asp functionalities were oriented away from each other, the folded conformation at $\varepsilon = 4$ brought the functionalities within 2.67 Å of each other. Analysis of the Arg-Lys, Asp-Arg, Aspartyl-Glutamate (Asp-Glu), and Glutamyl-Aspartate (Glu-Asp) transition states at $\varepsilon = 4$ indicated that the distance between the charged functionalities was greater than 7.60 Å. Therefore, it is likely that the proximity of the charged functionalities of Arg-Asp at $\varepsilon = 4$ is representative of the buried, hydrophobic portions of the protein, while the sequence is only physiological at the protein surface, and as a result the dipeptide would be more energetically unfavorable at $\varepsilon = 4$.[212, 213]

4.4.ii.b. Proton Transfer Step

Analysis of the Arrhenius parameters associated with the proton transfer step of the stepwise mechanism (Table 4.4.2) indicated that the reaction was least energetically favorable at the most hydrophobic solvent dielectric ($\varepsilon = 4$). At $\varepsilon = 4$, $E_0 = 27.3 \pm 3.04$ kcal.mol^{-1}, $E_A = 26.7 \pm 3.22$ kcal.mol^{-1}, and $\log(A) = 12.6 \pm 0.82$ s^{-1}, in the forward direction, and $E_0 = 34.1 \pm 7.31$ kcal.mol^{-1}, $E_A = 35.0 \pm 7.41$ kcal.mol^{-1}, and $\log(A) = 5.30 \pm 0.78$ s^{-1}, in the reverse direction. As we have previously specified, the interactions between the dipeptide and its solvated environment are representative of the energetics it experiences. As prior studies have indicated, the propensity for proton transfer reactions to occur quickly under polar, charged, and pure water environments.[358] Therefore, it is likely that the improbability of a proton transfer reaction to occur at or near hydrophobic residues causes the 0.24 kcal.mol^{-1} increase in the computed Arrhenius parameters compared to those computed at ε = 20, 40, and 80.

Table 4.4.2. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for the proton transfer reaction of non-enzymatic dipeptide hydrolysis at 298 K and 1 atm. $\tilde{A}$ has units of s^{-1} and L.mol^{-1}.s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of kcal.mol^{-1}.

ε	**Forward Reaction**			**ΔH**				

ε	**Forward Reaction** E_0	**Forward Reaction** $\log(\tilde{A})$	**Forward Reaction** E_A	ΔH_{RXN}	ΔG_{RXN}	**Reverse Reaction** E_0	**Reverse Reaction** $\log(\tilde{A})$	**Reverse Reaction** E_A
4	27.3	12.6	26.7	-6.48	-20.6	34.1	5.30	35.0
20	27.0	12.6	26.4	-7.10	-21.3	34.4	5.31	35.3
40	27.1	12.6	26.5	-7.16	-21.2	34.5	5.33	35.4
80	27.1	12.5	26.5	-7.18	-21.2	34.6	5.26	35.4

Of the 21 modeled proton transfer reactions, two dipeptides were found to exhibit energetics that denoted proton transfer as the rate-determining step regardless of modeled dielectric – Arg-Lys and Gly-Tyr. Analysis of the forward rate coefficients associated with the gemdiol formation reaction of these modeled dipeptides indicated that Gly-Tyr exhibited the largest forward rate coefficient regardless the dielectric, while the rate coefficient of Arg-Lys ranged from the third to seventh largest. However, when the forward rate coefficients of Arg-Lys and Gly-Tyr proton transfer were analyzed, the rate coefficients of the two reactions were ranked as the sixth and eighth smallest rate coefficients, respectively. Therefore, it is likely that the disparity in rate coefficient ranking between the gemdiol formation and proton transfer reactions would result in a shift in the proposed rate determining step.

As previous studies have indicated, proton transfer is more likely to occur when charged or polar residues are present within the system.[359] However, our analysis of the computed rate-coefficients of proton transfer that contained Lys in the C-terminus were slower relative to the rate-coefficients of reactions with Arg in the C-terminus. The discrepancy in the computed rates between the dipeptides containing basic residues in the C-terminus is likely due to the distance between the residue functional groups and the transition state reactive centers. The distance between the Arg guandinium functional group and the reactive center was found to be at least 1.200Å less than the distance between the Lys amine functional group and the reactive center. Therefore, it is likely that the difference in proximity would result in a quicker rate of proton transfer for dipeptide reactions that contain Arg in the C-terminus.

Analysis of the ranked forward rate coefficients of the Ala-Tyr and Gly-Tyr gemdiol formation reactions indicated that Ala-Tyr, like Gly-Tyr, exhibited a large rate coefficient. The propensity of these dipeptides towards hydrolysis is likely a result of the frequency of occurrence of Ala, Gly, and Tyr within the mAb hinge region. A bioinformatics study conducted by Flores et al lists how frequently each of the 20 amino acids appears in the different portions of the mAb, as well as the most prevalent residues within the hinge region.[163] From this, we determined that Tyr was the residue least likely to be found in the hinge region.[163] The reduced presence of Tyr within the mAb hinge region is likely due to the bulky nature of its functional group, which would reduce the flexibility of the hinge region.[362-364] Subsequently, Gly was the residue that frequently occurred within the hinge region [163] The propensity of Gly to occur in the hinge region is likely due to its characteristic as the smallest and most flexible amino acid.[239, 364, 365] Therefore, it is likely that hydrolysis of Gly-Tyr would be energetically favorable, as indicated by the computed rate coefficient. However, as Gly is a nonpolar residue, the subsequent proton transfer would not be as energetically favorable compared to dipeptides with either polar or charged functional groups. [358] As a result, the rate-determining step of Gly-Tyr hydrolysis would shift from gemdiol formation to proton transfer.

In addition to Arg-Lys and Gly-Tyr, the proton transfer reactions of Arg-Asp at $\varepsilon = 80$, Glu-Asp at $\varepsilon = 4$, 40, and 80, and Valyl-Alanine (Val-Ala) at $\varepsilon = 4$ were found to be rate-determining. The deviation of the proposed rate-determining step of Val-Ala at $\varepsilon = 4$ is likely a result of the increased favorability of the gemdiol formation reaction under the dielectric associated with hydrophobic residues. The rate of proton transfer for Val-Ala was observed to be the slowest reaction, regardless of the modeled dielectric. However, at $\varepsilon = 4$, the rate of Val-Ala gemdiol formation was ranked as the fifth slowest reaction, while it was ranked as the third slowest reaction at $\varepsilon = 20$, and the second slowest reaction at $\varepsilon = 40$ and 80. The increased rate of reaction at $\varepsilon = 4$ is likely a result of the transition state being most stable under the dielectric associated with nonpolar residues. Furthermore, as the rate of Val-Ala proton transfer is the slowest out of all the dipeptides, it is likely that the increased favorability of Val-Ala gemdiol formation at $\varepsilon = 4$ would shift the proposed rate-determining step.

Similarly, as we previously mentioned, Arg-Asp gemdiol formation was most energetically favorable at $\varepsilon = 80$, where no electrostatic interactions were evident. As a result, the rate coefficients of the gemdiol formation and proton transfer reactions were nearly identical. The Glu-

Asp hydrolysis reaction also exhibited rate coefficients of gemdiol formation and proton transfer that were within the same order of magnitude at $\varepsilon = 4$, 40, and 80. Unlike Aspartyl-Glutamic Acid (Asp-Glu), where the acidic functional group of the C-terminus is oriented away from the reactive center, the Asp functional group is in the same geometric plane as the reactive center. Therefore, it is likely that the interactions between the free proton and the functional group destabilizes the transition state. Subsequently, at the dielectric associated with charged residues, the gemdiol formation reaction becomes rate-determining, suggesting that the Asp functional group is stabilized by its environment.

4.4.iii. Evans-Polanyi Correlations as a Function of Modeled Dielectric

As we previously noted, the Evans-Polanyi correlation is a linear relationship between the enthalpy of a given reaction and its activation barrier, where more exothermic reactions will have lower activation energies.[141, 224, 366, 367] Through this correlation, the activation energy of an unknown reaction within a reaction class can be predicted from its reaction enthalpy. Therefore, to accurately predict the reaction activation barrier from the generalized linear model, a statistically significant Evans-Polanyi correlation must be achieved. In our study, we generated individual Evans-Polanyi correlations for the gemdiol formation and proton transfer steps of the stepwise mechanisms. Through this analysis, we determined that the Evans-Polanyi correlation could only be used to generalize the gemdiol formation reaction. Additionally, we were able to achieve statistically significant linear trends for the overall data set, as well as a function of the modeled dielectric. Through this, we were able to examine how the impact of the modeled dielectrics impacted the energetics associated with the Evans-Polanyi correlation.

As depicted in Figure 4.4.2 and Table 4.4.3, a statistically significant Evans-Polanyi correlation was achieved for the gemdiol formation step of the stepwise mechanism for the overall data set, as well as at each of the four modeled dielectrics. The strongest correlation was observed at $\varepsilon = 4$, while the weakest correlation was exhibited at $\varepsilon = 80$. The strength of the linear trend at the most hydrophobic solvent dielectric is likely due to the presence of Arg-Asp, which is acts as an outlier and maxima upon our linear trend.

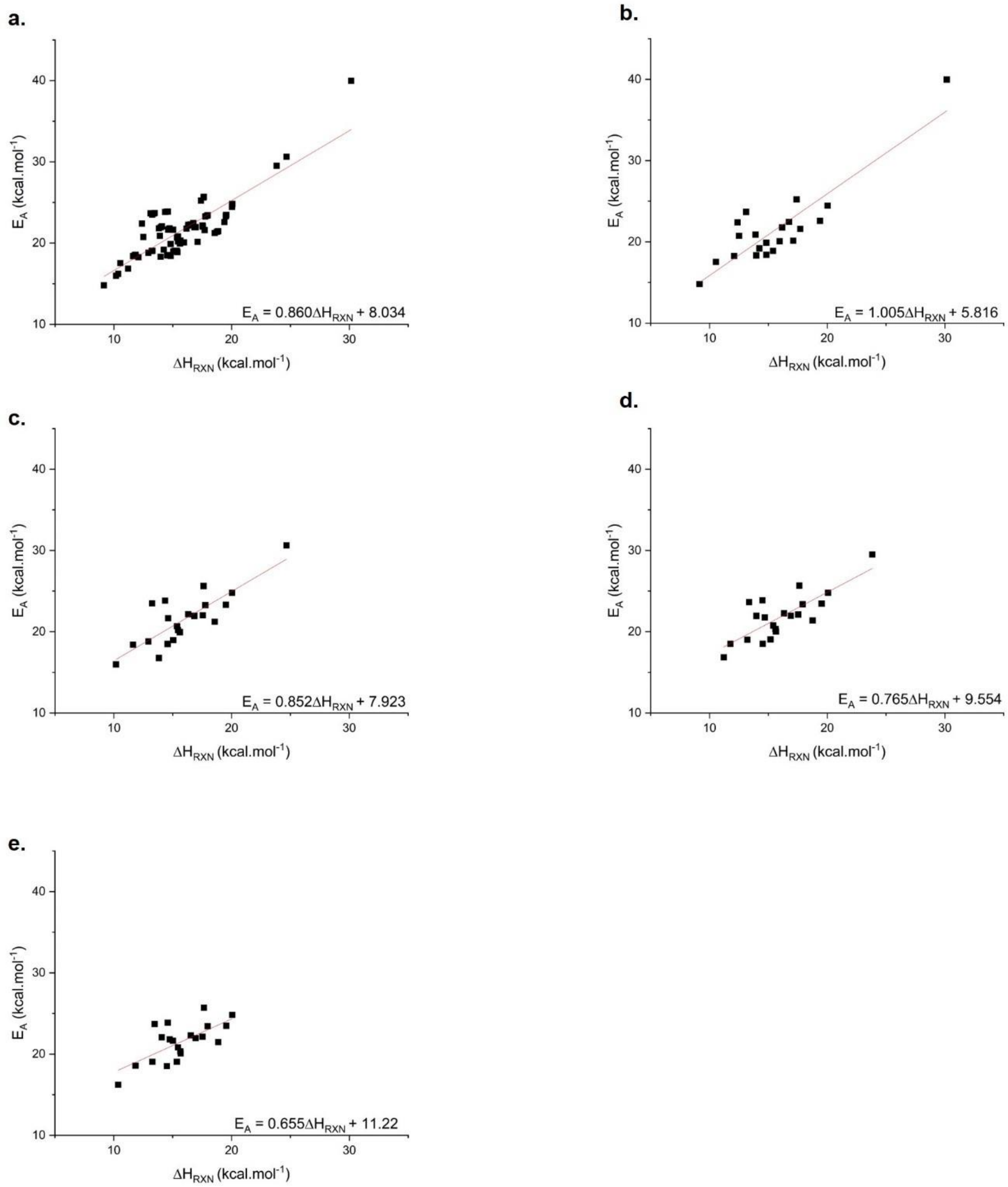

Figure 4.4.2. Evans-Polanyi correlation of the gemdiol formation step of the stepwise mechanism of non-enzymatic dipeptide hydrolysis. (a). Overall correlation, (b). $\varepsilon = 4$, (c). $\varepsilon = 20$, (d). $\varepsilon = 40$, (e). $\varepsilon = 80$. (Initial correlations).

As we previously noted, the activation energy and enthalpy of reaction for the hydrolysis of Arg-Asp is the most energetically unfavorable at $\varepsilon = 4$. Additionally, the presence of the dipeptide upon the trendline indicates that it follows the Evans-Polanyi principle that more endothermic reactions will have higher activation energies. [224, 366, 367] Therefore, because there is a greater than 10 kcal.mol^{-1} difference between Arg-Asp and the next data point, it is likely that Arg-Asp acts as an influential point upon the linear trendline and skews the strength of the generated correlation.[368, 369]

Table 4.4.3. Statistical analysis of the overall Evans-Polanyi correlation, as well as the correlations generated as a function of dielectric, for the gemdiol formation reaction (Initial correlations).

Correlation	α	E_0	R^2		Pearson's r	
	Value	Value	Value	p-value	Value	p-value
Overall	0.86	8.03	0.68	< 0.001	0.83	< 0.001
4	1.01	5.82	0.78	< 0.001	0.88	< 0.001
20	0.85	7.92	0.67	< 0.001	0.82	< 0.001
40	0.77	9.55	0.62	< 0.001	0.79	< 0.001
80	0.66	11.2	0.47	0.00	0.68	0.00

The Evans-Polanyi correlations at $\varepsilon = 20$ and $\varepsilon = 40$ were found to have strong linear correlations that were not as significantly impacted as the linear trends generated at $\varepsilon = 4$ and $\varepsilon = 80$. The correlation strength at these two dielectrics can be attributed to the interactions of the modeled dipeptides with the solvated environment. As we have previously specified, $\varepsilon = 20$ and $\varepsilon = 40$ are the dielectrics associated with charged and polar residues within a protein.[212, 213] Therefore, due to the constraints associated with our solvation model, as well as the large number of polar and charged residues within our modeled reaction class, it is likely that the electrostatic interactions between the dipeptides and the solvated environment are more significant at the modeled dielectrics associated with charged and polar residues.[208, 209, 212, 213, 370] This phenomenon is likely why the weakest correlation was observed at $\varepsilon = 80$. As we have previously specified, $\varepsilon = 80$ is the dielectric associated with pure water. As we have previously specified, a diverse array of residue functional groups was taken into account within our modeled reaction class to model the electrostatic interactions of the different functionalities with the changing modeled dielectric. Therefore, it is likely that the moderate strength of the Evans-Polanyi correlation generated at $\varepsilon = 80$ is less than at the other modeled dielectrics due to the electrostatic interactions between the dipeptides and their solvated environment.[208, 209, 370, 371]

The generated correlation for the full data set was found to be moderately strong ($R^2 = 0.68$ $p < 0.001$).[371, 372] While the moderate strength of the correlation coefficient for the Evans-Polanyi trend generated for the full data set indicates that it would be a more accurate tool compared to the individual correlation generated at $\varepsilon = 80$, there are also trade-offs associated with using this correlation. Specifically, by lumping the energetics associated with the gemdiol formation step at each modeled dielectric together, insight towards how the dielectric impacts the generated trend is lost. Additionally, while the overall trend has a significantly stronger correlation coefficient than the one associated with the $\varepsilon = 80$ Evans-Polanyi correlation, the strength of the correlations at $\varepsilon = 4$, 20 and 40 all fall in the same range as the overall correlation. Therefore, it is likely that the predictions derived from the individual correlations will have similar associated errors as those computed using the overall correlation.

4.4.iii.a. Initial Correlations and Reactions that do not Exhibit Evans-Polanyi-like Behavior

Of the 21 modeled reactions, Alanyl-Lysine (Ala-Lys), Glutaminyl-Glycine (Glu-Gly) and Glycyl-Proline (Gly-Pro) did not exhibit Evans-Polanyi behavior, regardless of the modeled dielectric. Removal of these three dipeptide reactions from the data set (Figure 4.4.3., Table 4.4.4) resulted in the linear strength of all correlations aside from at $\varepsilon = 80$ shifting into the upper echelon of the strong categorization ($R^2 > 0.83$).[371, 372] At $\varepsilon = 80$, the generated correlation was found to shift from moderate to strong ($R^2 = 0.73$).[371, 372]

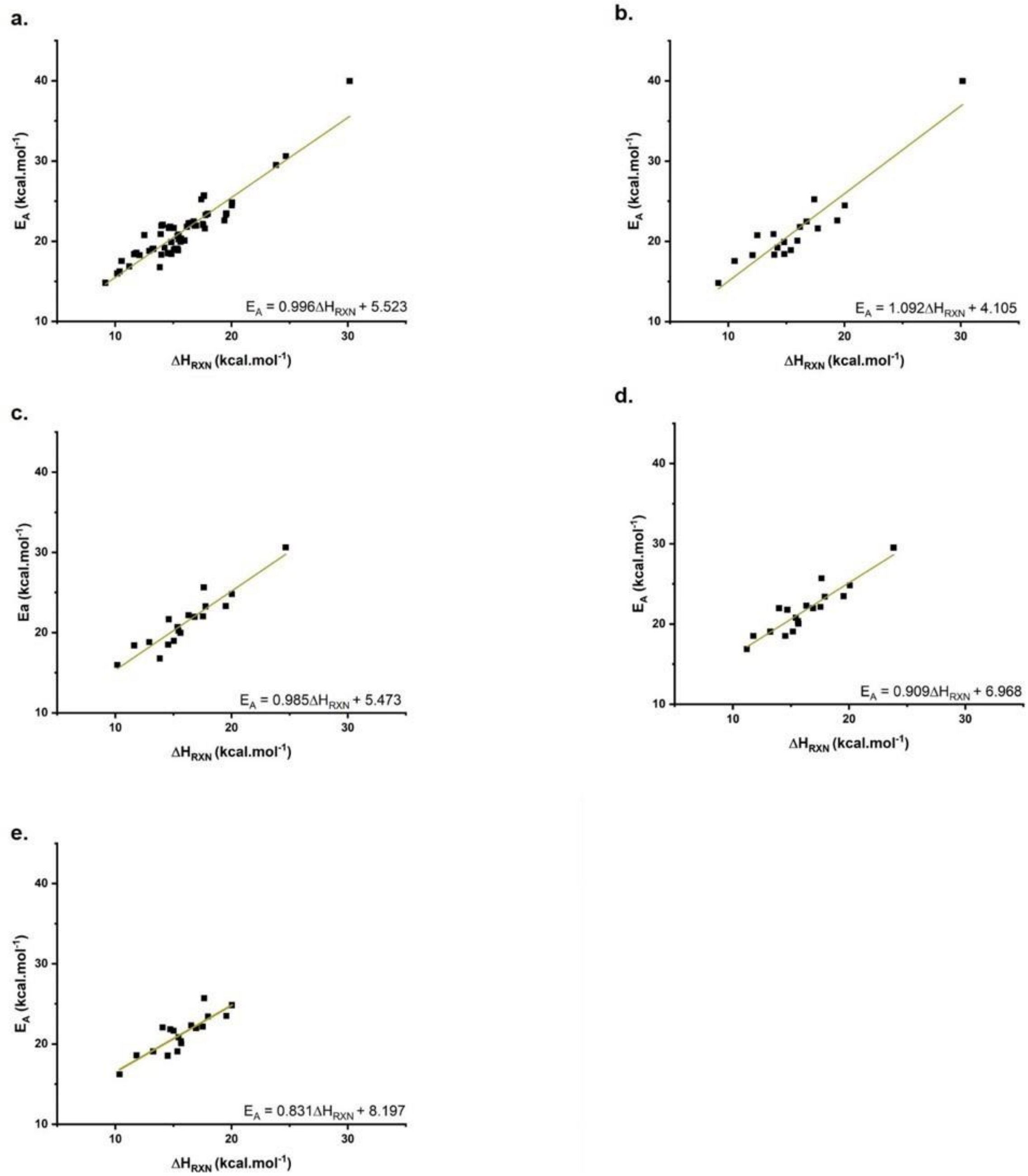

Figure 4.4.3. Evans-Polanyi correlation of the gemdiol formation step of the stepwise mechanism of non-enzymatic dipeptide hydrolysis. (a). Overall correlation, (b). $\varepsilon = 4$, (c). $\varepsilon = 20$, (d). $\varepsilon = 40$, (e). $\varepsilon = 80$. (Outlier removal correlations).

Evaluation of the gemdiol formation reactions indicated that the activation barriers of Ala-Lys and Glu-Gly were higher than their associated enthalpies of reaction would suggest, similarly the opposite trend was observed for Gly-Pro gemdiol formation. The discrepancy between the activation barrier and the enthalpy of reaction for Gly-Pro gemdiol formation is likely a result of the difference in conformation between the reactant and the gemdiol intermediate. Specifically, prior studies have indicated that amide bonds that contain Pro in the C-terminus position are more prevalent and more energetically favorable when in the cis conformation.[373-376] Evaluation of the omega (ω) dihedral angle of the Gly-Pro gemdiol relative to its reactant indicated that the gemdiol intermediate, similar to the reactant structure, had a ω angle that was representative of the trans conformation. Analysis of the transition state ω angle indicated a conformation that was about 45° closer to cis than the reactant structure, suggesting that the lower activation barrier is representative of the shift of the transition state into a more stable conformation.[373-376] As a result, the activation barrier of Gly-Pro gemdiol formation would not be representative of its enthalpy of reaction.

Similar to Gly-Pro, the non-Evans-Polanyi behavior of Ala-Lys can be attributed to its conformation. The ω angles of the reactant and gemdiol intermediate for Ala-Lys gemdiol formation were observed to be within six degrees of the ideal trans conformation, while the transition state ω angle was more representative of a cis conformation at 68.5°. Therefore, as amide bonds without Pro in the C-terminus are more stable in the trans conformation, the cis conformation of the transition state would result in a higher activation barrier than what is represented by the enthalpy of reaction.[373-376] Unlike Ala-Lys and Gly-Pro, the deviation observed in Glu-Gly gemdiol formation was not a consequence of its backbone conformation. The reactant, transition state, and gemdiol intermediate ω angles were found to be representative of the trans conformation.

However, comparison of the Glu-Gly transition state against its gemdiol intermediate, as well as Aspartyl-Serine (Asp-Ser) indicated that the interactions between the reactive center and the C-terminus carboxyl group might influence energetics. While the carboxyl groups of Asp-Ser transition state and the Glu-Gly gemdiol intermediate were found to be in the same geometric plane as the reactive center, the carboxyl group of the Glu-Gly transition state was oriented away from the reactive center. Therefore, it is unlikely that the alcohol functional group of the reactive center gemdiol would be able to form hydrogen bonds with the C-terminus carboxylic group and further stabilize the transition state.[377] As a result, the Glu-Gly gemdiol formation reaction would not exhibit Evans-Polanyi behavior. To reduce model biases, the generated correlations which included these outliers were implemented in our sensitivity analysis. Therefore, it is likely that the activation barriers predicted for these reactions will be less accurate compared to other reactions within the data set.

Table 4.4.4. Statistical analysis of the overall Evans-Polanyi correlation, as well as the correlations generated as a function of dielectric, for the gemdiol formation reaction (Outlier removal correlations).

Correlation	α	E_0	R^2	
	Value	Value	Value	p-value
Overall	1.00	5.53	0.85	< 0.001
4	1.09	4.11	0.89	< 0.001
20	0.99	5.47	0.87	< 0.001
40	0.91	6.97	0.83	< 0.001
80	0.83	8.20	0.73	< 0.001

4.4.iii.b. Impact of Bootstrapping on Model Linearity

The goodness of fit of each generated linear Evans-Polanyi correlation was analyzed through the use of statistical bootstrapping (Figure 4.4.4, Table 4.4.5). As we previously mentioned, the statistical bootstrapping applied to our generated correlations was achieved through the random and unbiased removal of 20 percent of data points. The bootstrapped samples were then used to generate a new linear correlation and predict the activation barriers of the removed reactions as a function of their given enthalpy of reaction. To achieve a robust analysis of the accuracy of our generated correlations, five iterations of bootstrapping were performed on each Evans-Polanyi correlation.

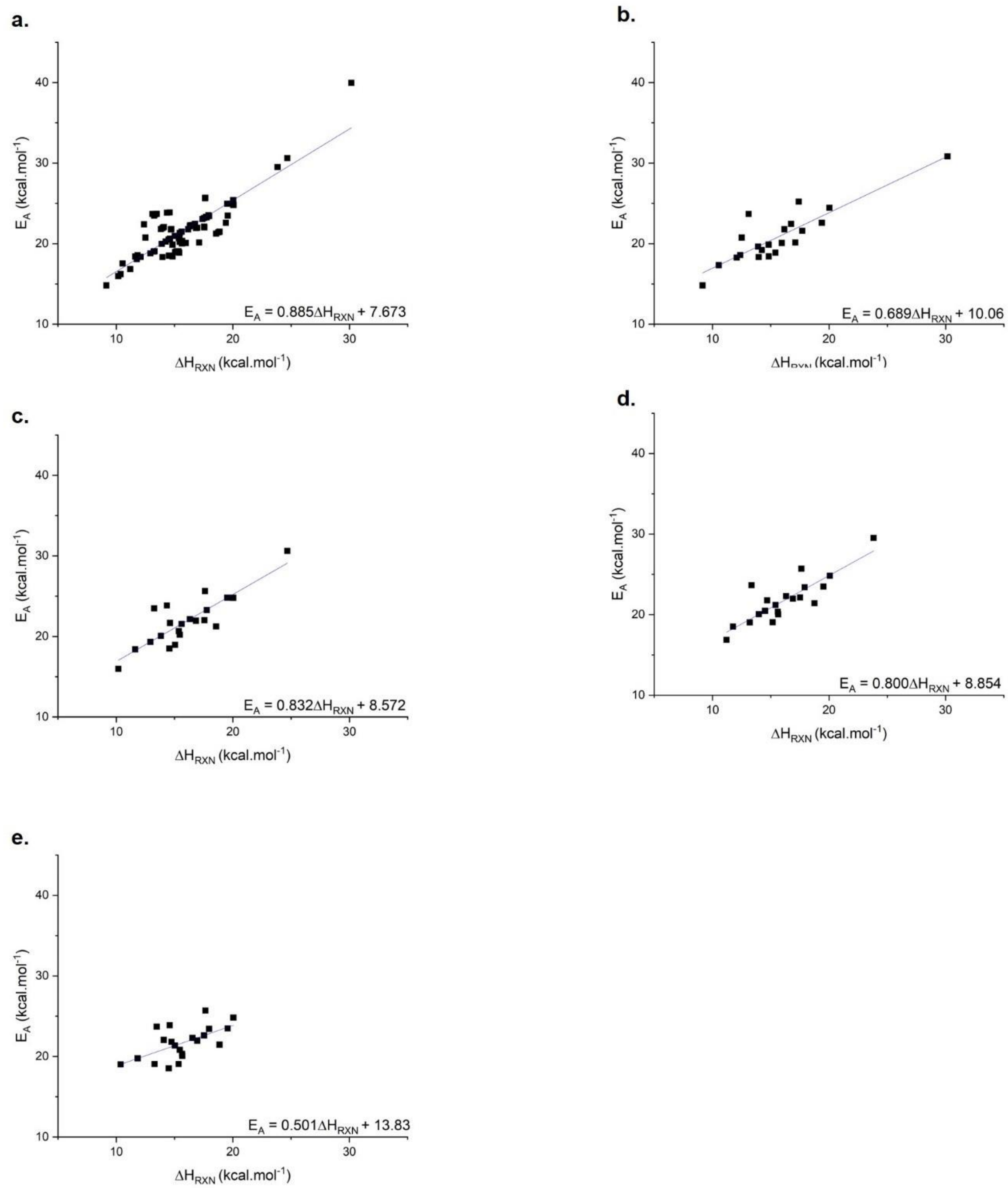

Figure 4.4.4. Sample bootstrapped Evans-Polanyi correlations of the gemdiol formation step. (a). Overall correlation, (b). ε = 4, (c). ε = 20, (d). ε = 40, (e). ε = 80.

Evaluation of the bootstrapping sensitivity analysis on the overall Evans-Polanyi correlation indicated that all five iterations yielded correlations that could be classified as either moderate or strong ($0.66 < R^2 < 0.73$).[372] However, of the five iterations, only the fifth iteration yielded an R^2 value that was less than 0.7. While the R^2 value associated with the fifth bootstrapped correlation still falls within the category of moderate linear strength, we sought to identify the cause of the decrease in linear strength of this correlation compared to the other bootstrapped iterations. Through our analysis, we determined that the fifth bootstrapped iteration was the only trend generated through the random removal of the outlier Arg-Asp at the dielectric associated with hydrophobic residues. As we previously mentioned, the activation barrier of Arg-Asp at $\varepsilon = 4$ was found to be a significant outlier within our modeled system and likely influences the strength of the generated correlations. Therefore, the R^2 value of the bootstrapped iteration that randomly removed Arg-Asp at $\varepsilon = 4$ would likely be smaller than the iterations which were generated using Arg-Asp at $\varepsilon = 4$ as a data point.

Table 4.4.5. Statistical analysis of the overall Evans-Polanyi correlation, as well as the correlations generated as a function of dielectric, for the gemdiol formation reaction given by sample bootstrapped correlations.

Correlation	α	E_0	R^2		Pearson's r	
	Value	Value	Value	p-value	Value	p-value
Overall	0.89	7.67	0.72	< 0.001	0.85	< 0.001
4	0.69	10.1	0.77	< 0.001	0.88	0.001
20	0.83	8.57	0.71	< 0.001	0.84	< 0.001
40	0.80	8.85	0.72	< 0.001	0.85	< 0.001
80	0.50	13.8	0.36	0.00	0.60	0.04

In addition to analyzing the impact of the bootstrapping sensitivity method on the overall Evans-Polanyi correlation, we evaluated the dielectric-specific correlations to analyze how the linear strength of each modeled generalization was impacted by bootstrapping. Through this analysis, we determined that, similar to the initial Evans-Polanyi correlations, the linear trend at the dielectric associated with pure water was the weakest of the five generated correlations. As we previously mentioned, the variability associated with the enthalpy of reaction and activation barrier is lowest at ε =80, suggesting that the trend is not dependent on the value of one particular data point. Therefore, while the linear strength of the correlations ranged from weak ($R^2 = 0.36$, $p = 0.00$) to moderate ($R^2 = 0.50$, $p < 0.001$), it is likely that the predictiveness of the model will not be as strongly impacted by the replacement values of the randomly removed data points.[371, 372]

Similarly, our hypothesis that the linear strength of the Evans-Polanyi correlation generated at ε = 4 was a consequence of the outlier Arg-Asp was supported by the trends achieved by each bootstrapping iteration. A correlation was found to exist between the presence of Arg-Asp as a removed data point and the linear strength of the bootstrapped correlation, similar to the decrease in linear strength observed in the fifth bootstrapped iteration of the overall correlation. Specifically, correlations that did not randomly replace Arg-Asp were found to exhibit strong linearity ($R^2 > 0.84$, $p < 0.001$), while correlations that did randomly replace Arg-Asp decreased in linear strength ($0.739 < R^2 < 0.77$, $p < 0.001$).[371, 372] Therefore, it is likely that the presence of Arg-Asp, in the correlation causes an over prediction of the slope.

Of the four dielectric-dependent Evans-Polanyi correlations, the bootstrapped trends at ε = 20 and ε = 40 were found to both exhibit moderately strong to strong linearity and not be dependent on a singular data point within the correlation.[371, 372] As a result, we hypothesized that the prediction error of the bootstrapped iterations at these dielectrics would be similar in nature to the prediction error exhibited by the overall data set.

4.4.iii.c. Model Predictiveness

As we previously indicated, the data points that are randomly removed from the bootstrapped correlation are used to represent the population through which the correlation would be applied to, and the prediction error is the difference between the original value and the replacement value generated by the bootstrapped correlation.[231] Therefore, to analyze the predictiveness of our generated Evans-Polanyi correlations we computed the average, absolute error between the activation energies predicted through the bootstrapped correlation and the activation energies computed by *CalcK* (Figure 4.4.6). As we previously noted, the five bootstrapped iterations of the Evans-Polanyi correlation for the overall data set exhibited moderately strong linearity. As a result, the prediction error associated with the overall correlation was found to be less than 2.000 kcal.mol^{-1} regardless of the bootstrapping iteration.

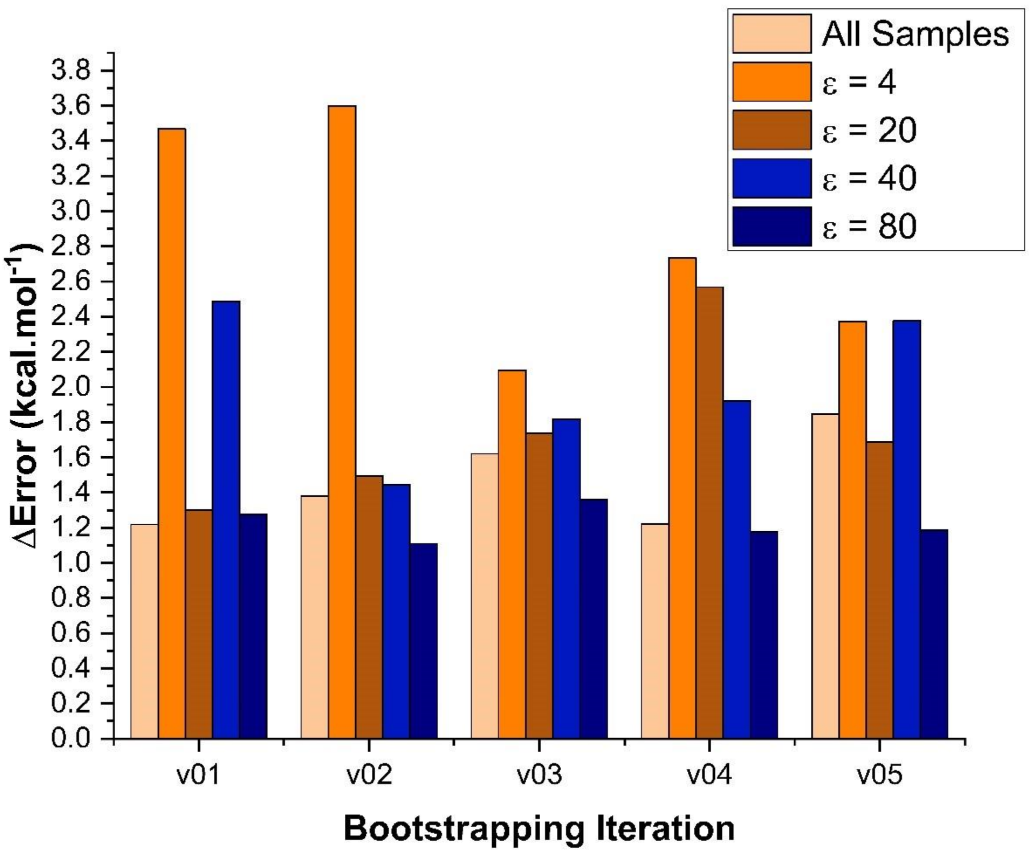

Figure 4.4.6. Bar chart of the average and absolute value of ΔError ($E_{A,predicted}$ - $E_{A,computed}$) as a function of the bootstrapped iteration. $E_{A,predicted}$ refers to the activation barriers predicted by the boostrapped correlations, while $E_{A,computed}$ is the value calculated from quantum chemical calculations and statistical thermodynamics.

Of the four modeled dielectric Evans-Polanyi correlations, the weakest bootstrapped trends exhibited the lowest prediction error. As we previously mentioned, the standard deviation associated with the enthalpy of reaction and activation barrier was lowest at $\varepsilon = 80$. Therefore, it is likely that there were no data points within the correlation that the prediction accuracy was dependent on. Similarly, the bootstrapped correlations at $\varepsilon = 4$ exhibited the strongest linearity, but the greatest prediction error. The greater prediction error associated with the five bootstrapped trends at $\varepsilon = 4$ is likely a consequence of the dependency of the Evans-Polanyi correlation on the Arg-Asp outlier. As we previously noted, the correlations that randomly replaced Arg-Asp had lower R^2 values relative to the correlations that did not resample the outlier. However, the prediction accuracy at $\varepsilon = 4$ was found to be lower relative to the other correlations regardless of the presence of Arg-Asp as a replaced data point. Specifically, correlations that randomly removed Arg-Asp significantly under predicted the activation barrier of the reaction, but were more accurate in predicting the values of the other data points. Similarly, the correlations that included Arg-Asp resulted in activation barriers that were over predicted by about 3.000 kcal.mol^{-1}. Therefore, it is likely that the inaccuracy of these correlations relative to the other generated Evans-Polanyi trends is representative of the dependence on the Arg-Asp outlier at $\varepsilon = 4$.

Four of the five bootstrapped iterations at $\varepsilon = 20$ were found to yield an average prediction error that was similar to the prediction error exhibited by the correlations at $\varepsilon = 80$ and for the overall data set. The fourth bootstrapped iteration yielded an average error that was about double the prediction error of the other bootstrapped iterations, and can be attributed to under predicting the activation barrier of a dipeptide that does not exhibit Evans-Polanyi energetics (Glu-Gly) by 4.887 kcal.mol^{-1}. Similarly, the greatest average prediction error at $\varepsilon = 40$ was observed for a correlation that randomly replaced the Ala-Lys and Glu-Gly activation energies. The fifth bootstrapping iteration at $\varepsilon = 40$, which also exhibited a high average prediction error relative to other correlations, was the only trend that did not randomly remove a reaction containing a nonpolar residue in the C-terminus. Therefore, it is likely that the electrostatic interactions of the C-terminus residue are not represented by the Evans-Polanyi correlation.

Through our analysis of the average and absolute value of the prediction error associated with each bootstrapped iteration of our generated Evans-Polanyi correlations, we determined that the reduced accuracy was a result of either replacing the activation barrier of the reactions found to not follow Evans-Polanyi behavior, or by replacing the Arg-Asp outlier at $\varepsilon = 4$. Therefore, we determined that our bootstrapped correlations were accurate in predicting the activation barriers of reactions which follow the Evans-Polanyi principle and are not outliers within our data set. [224, 366, 367]

4.5. Conclusions and Future Directions

In summary, a robust analysis of the impact of the modeled dielectric upon non-enzymatic amide hydrolysis via the uncatalyzed, nucleophilic attack of water was completed for the stepwise reaction mechanism. Rate coefficients and Arrhenius parameters for non-enzymatic dipeptide hydrolysis at a neutral pH was computed for 168 (42 reactions per modeled dielectric) total reactions using B3LYP and statistical thermodynamics.[141] The predicted rate coefficient of our proposed rate-determining step was benchmarked against an experimental rate-coefficient to validate our state of the art computational methodology and solvation model.[141, 143] As a result, we determined that the gemdiol formation step of the stepwise mechanism estimated a rate coefficient that was within one order of magnitude of the value given by Radzicka et al.[143] Additionally, we found the propensity of the proton transfer reaction to occur via charged or polar residues, as well as bulk water, to be exhibited by our generated model.[358]

Lastly, a statistically significant linear correlation was generated for the gemdiol formation step of the stepwise mechanism for the overall data set, as well as for each modeled dielectric. Via the modeled Evans-Polanyi correlations, we determined that the activation energies associated with the gemdiol formation step of non-enzymatic amide hydrolysis could be generalized into linear trends. Furthermore, five iterations of statistical bootstrapping analysis was applied to the overall correlation, as well as the correlations generated at each modeled dielectric. While the linear strength of the modeled correlations ranged from weak to strong, the average associated error in predicted reaction activation barriers was consistently found to be less than 3.000 kcal.mol^{-1}, and the higher associated error for the two bootstrapped iterations at $\varepsilon = 4$ was a consequence of the random removal of the Arg-Asp outlier. Future studies will automate the Evans-Polanyi correlation so that generalization proposed rate-determining step of non-enzymatic peptide bond fragmentation can be applied for larger polypeptide structures, and if necessary, a correction factor can be derived so that the predicted values generated from the correlation can be included in the generation of a microkinetic model of protein degradation.

Chapter 5: Towards Pharmaceutical Protein Additive and Stabilizer Screening: A Theoretical Study on the Effectiveness of Cations towards the Inhibition of Deamidation via the Succinimide Reaction Pathway

5.1. Background and Motivation

Studies on the impact of additives and stabilizers towards inhibiting protein degradation are of interest due to the current shelf-life of therapeutic proteins at room temperature. At room temperature (T ~ 22 – 25 ºC), the formulations are more susceptible towards degradation and rapidly lose stability and efficacy.[107-110] As a result, the current approach towards extending the stability of the formulation is through lyophilization and long-term storage at low temperatures (-80 ºC < T < -20 ºC).[110, 112-115] Lyophilized formulations have been noted by previous studies to increase the shelf-life of mAbs from weeks to years.[116-119] However, prior studies have also indicated that the approach can be detrimental to long-term stability due to the stresses lyophilization places on the polypeptide formulation.[120-127]

For this purpose, the inclusion of additives and stabilizers has been implemented to reduce protein degradation within mAb formulations.[77, 83-85, 93-95, 97-104] The compounds included as additives and stabilizers within mAb formulations are classified based on their role in improving the stability and efficacy of the manufactured product. Sugars such as sucrose and dextrose are implemented as cryoprotectants during lyophilization of the formulation, while polyols enhance the conformational stability of the formulation.[83, 93-95, 97-102] Polymers are also used to reduce protein self-interaction and the formation of aggregates within the formulation.[93-96] Buffers are also common stabilizers included in mAb formulations and act to stabilize the formulation pH at different points of the manufacturing process.[83-90] Furthermore, buffers can also be used to stabilize the polypeptide by reducing the electrostatic interactions of charged residues on the surface.[91, 92]

However, in the case of Asparagine deamidation, degradation via the succinimide intermediate reaction pathway is predominant under ambient conditions.[80, 82, 110, 150, 151, 153, 168, 169, 273, 279, 378, 379] Specifically, the formation of the succinimide intermediate is a self-promoted, spontaneous process that first forms a cyclic intermediate that subsequently hydrolyzes into either Aspartic Acid (Asp) or iso-Aspartic Acid (iso-Asp); where iso-Asp is the preferred product under ambient conditions.[82, 110, 147, 155, 169, 279, 297, 300, 301, 379] Prior studies have indicated that the presence of iso-Asp within the polypeptide formulation reduces the efficacy of the mAb.[104, 380-385] To combat this issue, a patent was filed by Rehder in 2010 (Document ID: US 20100158908 A1), where cations were proposed as additives to inhibit the deamidation of Asn residues into iso-Asp within the mAb formulation.[104] Furthermore, the results highlighted by the patent indicated that the inclusion of $CaCl_2$ to formulations reduced the rate of succinimide formation.[104] However, fundamental understanding regarding the impact of the cations on the spontaneity of the succinimide formation reaction was not analyzed (Figure 5.1.1).

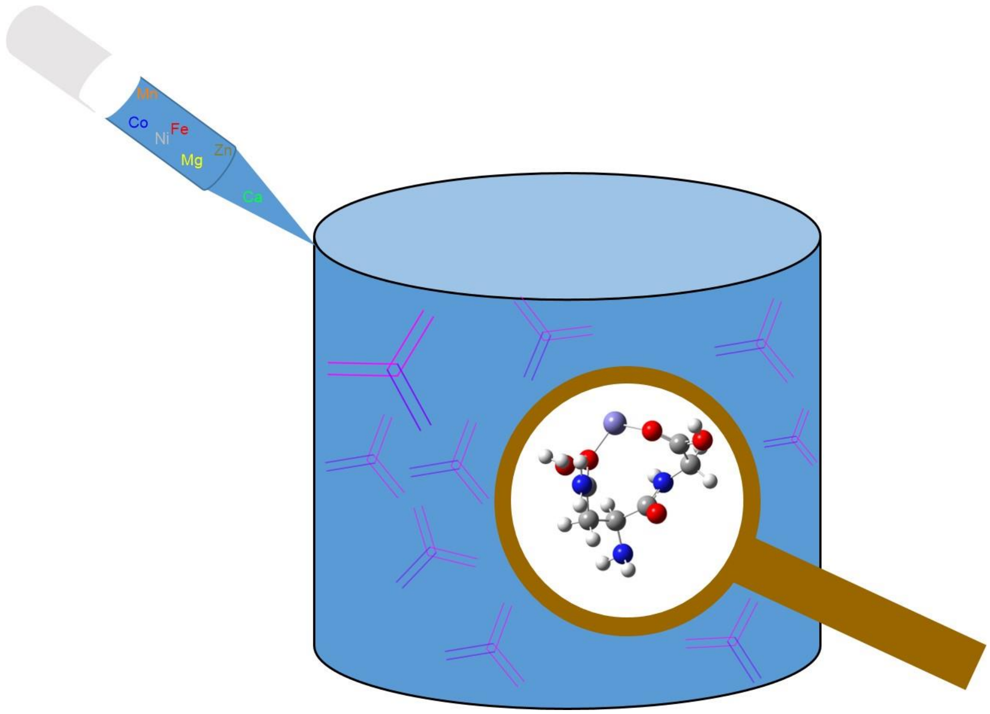

Figure 5.1.1. Multi-scale representation of the impact of cations on inhibiting deamidation via the succinimide reaction pathway.

For this purpose, we conducted a computational study, using a model dipeptide complexed with a single cation, to quantify the impact of cations on the spontaneity of the succinimide formation reaction. This study presents the most probable spin states of 48 cation-dipeptide complexes of Asparaginyl-Glycine (Asn-Gly) succinimide formation in the conductor-like polarizability model (CPCM) solvated state. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic and Gibbs free energies at 298 K.[187-190, 246, 247] Selective screening of the effective cation additives as inhibitors of the succinimide reaction pathway was evaluated via the computed binding free energies and free energies of inhibition of the site-specific complexes analyzed in this study.

5.2. Study Objectives

This study provides a first-principles analysis of the effectiveness of cations as inhibitors towards the formation of the succinimide intermediate. Specifically, due to the presence of transition metals within our analyzed complexes, we sought to identify the most energetically favorable and realistic complexes formed at each proposed site of inhibition. In addition to identifying the most probable spin states as a function of the cation oxidation state and site of inhibition, we also sought to validate our proposed model against experimental data. After verification of the chemical model and computational methodology was completed, we analyzed how the site of inhibition impacted the binding strength and effectiveness of the cation-dipeptide complexes towards the inhibition of Asn-Gly succinimide formation. Through this analysis, we successfully screened seven cations in terms of the effectiveness towards the inhibition of the succinimide intermediate as a function of the inhibition site of interest.

5.3. Computational Methodology

As we previously mentioned, this study builds upon the foundations of the developed reaction networks of Asn-X and Gln-X deamidation under ambient conditions. For this purpose, we implemented a similar base methodology for this study to reduce variability within our analysis and maintain a self-consistent methodology amongst our developed databases.[82, 141, 147, 251]

5.3.i. Cluster Development and Proposed Cation-Dipeptide Interactions

Cluster development was completed through the hypothesis that only the cation acts as an additive against deamidation via the succinimide reaction pathway, while the anion acts as charge stabilizer by forming a salt with the cation. Furthermore, when the formulation is in a solvated state, the salt dissociates such that the cation exists freely within solution. Of the 19 Asn-X dipeptides that can participate in deamidation via the succinimide reaction pathway, Asn-Gly has been noted to have a higher propensity towards deamidation due to the small, flexible nature of the C-terminus Gly.[151, 169, 283, 386, 387] Therefore, to analyze the effectiveness of cations towards reducing the spontaneity of the succinimide formation reaction under ambient conditions, all complexes were formed using the Asn-Gly structures generated in our Asn-X deamidation reaction network study.[82]

All complexes were modeled via interactions between the cation and the carbonyl functional group of the amide of interest in accordance to previous studies.[388-391] Analysis of the site-specific impact of cation inhibition on Asn-Gly succinimide formation was completed via the location of each complex. As we previously noted, deamidation via the succinimide pathway can form either Asp-Gly or its more acidic isomer, iso-Asp-Gly; where iso-Asp-Gly is the preferred product under ambient conditions.[82, 147, 155, 169, 279, 297, 300, 301] Formation of the two products occurs via the hydrolysis of the Asn-Gly succinimide intermediate either at the sidechain carbonyl (Asp-Gly) or the backbone carbonyl (iso-Asp-Gly).[82, 147, 155, 169, 279, 297, 300, 301]

Additionally, we sought to identify if the conformation of the backbone impacted the effectiveness of the cations modeled in this study. Specifically, by manipulating the dipeptide backbone such that the backbone and sidechain carbonyls are in the same geometric plane. For site-specific inhibition of Asn-Gly succinimide formation where the reactant structure was not manipulated, the torsional angle of the uninhibited reactant structure is 65.47^{o} (C_9-C_5-C_3-C_6), while the uninhibited reactant structure of the site-specific inhibition involving the presence of the sidechain and backbone carbonyl groups within the same geometric plane was computed at 73.07^{o}.

Table 5.3.1. Classification of cations used to form cation-amide complexes for the inhibition of Asn-Gly succinimide formation.

Cation	Classification
Calcium	Alkaline Earth Metal
Cobalt	First-Row Transition Metal
Iron	First-Row Transition Metal
Magnesium	Alkaline Earth Metal
Manganese	First-Row Transition Metal
Nickel	First-Row Transition Metal
Zinc	First Row Transition Metal

5.3.2. Quantum Chemical Calculations

Quantum chemical calculations were performed with *Gaussian 16*.[250] All electronic energies for the protein dipeptides were calculated using the B3LYP method. [82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized as a function of the oxidation and spin states of the cation of interest.. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the modeled reactions.[82, 141, 147, 188-190, 246, 247, 253] Similar to the Asn-X and Gln-X deamidation reaction networks, a hybrid solvation method was implemented using implicit solvation, through the conductor-like polarizable continuum model (CPCM), and a single stabilizing water molecule interacting with the sidechain oxygen.[82, 147, 209, 370] Due to the tedious nature of the spin multiplicity calculations, only the dielectric associated with pure water ($\varepsilon = 80$) was included within our methodology, where future studies will include the electrostatic impact of the other three modeled dielectrics on the inhibition of the succinimide formation reaction.[82, 141, 147, 251] Evaluation of the binding free energies of the modeled complexes was completed using equation 5.3.1, where $\Delta G_{COMPLEX}$ represents the free energy associated with the complex formed between the Asn-Gly reactant structure and the cation of interest and ΔG_{Cation} is the free energy of the free cation within the CPCM solvated environment.

$$\Delta G_{BINDING} = \Delta G_{COMPLEX} - \Delta G_{Asn\text{-}Gly\_Reactant} - \Delta G_{Cation} \qquad (5.3.1)$$

The impact of the cations upon inhibiting succinimide formation was studied through equation 5.3.2, where the inhibition of Asn-Gly succinimide formation is evaluated by the perturbation of the reaction free energy. Specifically, a comparison of the free energy difference of the inhibited product and reactant structures relative to the reaction free energy of uninhibited Asn-Gly succinimide formation.

$$\Delta G_{SF,INHIBITED} = \Delta G_{Product\_Complex} + \Delta G_{NH3} - \Delta G_{Reactant\_Complex} \qquad (5.3.2)$$

5.3.iii. Oxidation States and Spin Multiplicity

As five of the seven cations of interest are classified as transition metals (Table 5.3.1), we sought to determine the impact of their oxidation and spin states upon the electronic energy of our modeled system, as well as which oxidation and spin states were accurately representative of the formulation conditions. Rehder defines the oxidation numbers of the cations analyzed within the patent to all be divalent in nature.[104] Evaluation of the transition metals used in biological systems indicated that Cobalt (Co), Iron (Fe), and Manganese (Mn) were found to frequently exist in both the +2 and +3 oxidation states, while Nickel (Ni) and Zinc (Zn) were usually found in a +2 oxidation state.[392-400] Therefore, to achieve a robust analysis of the impact of transition metals on inhibiting the succinimide intermediate, Co, Fe, and Mn were modeled in their +2 and +3 oxidation states.

Similar to the oxidation states associated with each of the transition metals analyzed in our study, the impact of the spin state on the thermodynamics of the succinimide formation reaction pathway must also be considered. A study by Halcrow noted that the spin state in biological systems was representative of the associated ligand.[401] Prior studies regarding the impact of metals on biological systems have classified as either 'A', 'B', or 'borderline' based on the ligand type the metal is most likely to form a complex with.[402, 403] Alkaline earth metals, such as Calcium (Ca) and Magnesium (Mg) are referred to as class A metals due to their preference towards ligands that contain oxygen, while transition metals are classified as 'borderline' due to their ability to bind with either Class A or Class B ligands.[402-404]

As we previously mentioned, this study analyzes how a subset of cations cited by Rehder impacted the thermochemistry of the cyclic intermediate reaction pathway of deamidation. For this purpose, while our developed clusters only include our modeled dipeptide and the cation of interest, the salts formed via the interactions of the cations with their respective ligands can also be used to indicate the spin state that would be most representative of the physiological system. For this purpose, we sought to identify the spin states of the cations as a function of the cluster, where the cation is treated as dissociated from its ligand, as well as a function of its respective salt. Rehder cited that salts could be formed within the formulation via the ionic interactions of the cations with Cl^-, SO_4^{2-}, CH_3COO^-, and PO_4^{3-}, where the listed anionic compounds are considered weak field ligands.[104, 405-408] Therefore, the salt form of the divalent cations is likely to prefer a high spin state, while the dissociation of the salt within the solution would be more energetically favorable in a low spin state.[406] To achieve a robust analysis of the impact of the spin states on the electronic energies of the Asn-Gly succinimide formation reactant and product structures, the transition metal complexes were modeled at their low and high spin states; allowing us to identify the most probable spin state of each cation at each site of inhibition analyzed in this study. Figure 5.3.1. provides the electron configurations of the low and high spin states of the Co(II) and Co(III) complexes analyzed in this study.

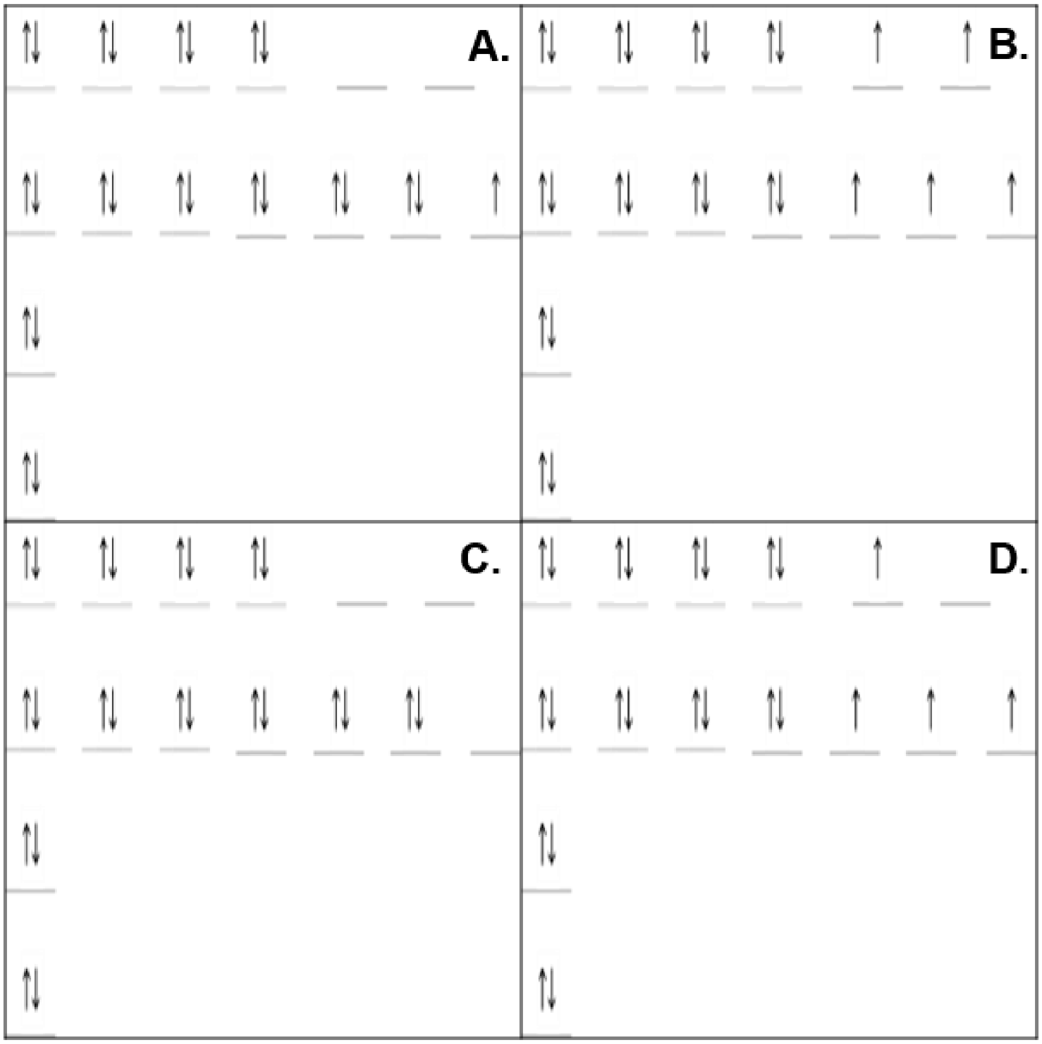

Figure 5.3.1. Example electron configurations of the low and high spin states of the transition metals as a function of their respective oxidation states. (a). Co(II) (low spin), (b). Co(II) (high spin), (c). Co(III) (low spin), (d). Co(III) (high spin).

The multiplicity associated with each structure is given by Equation 5.3.3:

$$\text{Spin multiplicity} = (n+1) \qquad (5.3.3)$$

Where n is the number of unpaired electrons. A single unpaired electron was present in the electron configurations of Co(II) and Mn(II), indicating that the two divalent cations would exist in the doublet spin state, while Fe(II), Ni(II), and Zn(II) were found to exist in the singlet spin state when the cations were treated as dissociated within the system. Evaluation of the spin states for the cation salts indicated that Co(II) exists in a sextet state, Fe(II) and Ni(II) in a quintet spin state, and Mn^{2+} exists in the quartet spin state. Prior studies indicated that Zn(II) exists solely within a singlet spin state, and as a result only the low spin state was considered.[409] Furthermore, a study by Bestgen in 2018 indicated that Zn(II) would behavior similarly to an alkaline earth metal due to the closed shell nature of its electron configuration.[410] As we previously specified, Co, Fe, and Mn were found to participate in biological processes while in their +3 oxidation states, therefore we also needed to consider their associated high and low spin multiplicities. At low spin, Co(III) and Mn(III) exist in the singlet state while Fe(III) exists in the doublet state. Analysis of the high spin states indicated the associated high spin multiplicity of Co(III) is the quintet state, the associated high spin multiplicity of Fe(III) is the quartet state, and the associated high spin multiplicity of Mn(III) is the triplet spin state.

5.3.iv. Level of Theory and Chemical Model Validation

As we have previously noted, the base methodology implemented in this study was used to ensure self-consistency within the developed databases and reduce the variability associated with the comparison of the spontaneity of the cation-dipeptide complexes and the uninhibited formation of the Asn-Gly succinimide intermediate. However, similar to prior studies, we sought to validate that the level of theory and chemical model used in this study yielded results that were an accurate representation of the physical system. For this purpose, we validated our model against both an experimental and a computational study. An experimental study by Okur et al. provides the IR frequency associated with the bond stretch of an amide backbone carbonyl complexed with Ca(II), while the Pluharova study gives an indication of the acceptable error range associated with the experimental - *ab initio* comparison.[389, 411] To reduce any variability within our analysis, only complexes where Ca(II) inhibits the backbone carbonyl were considered.

Through this analysis (Table 5.3.2), we determined that the maximum associated error was present in the reactant structure, where the carbonyl bond stretch was found to be 30.19 cm^{-1} greater than the bond stretch reported by Okur et al.[389] Evaluation of the disparities between the *ab initio* frequencies reported by Pluharova et al. relative to the experimental values given by Okur et al. indicated that the associated error was about 23 cm^{-1}.[389, 411] Therefore, as amide-Ca(II) complexes analyzed by the Okur and Pluharova studies was not Asn-Gly, it is likely that the additional 7 cm^{-1} of deviation is a result of the modeled system and not the chemical model of level of theory implemented in this study. For this purpose, we determined that our current methodology would yield appropriate results of the impact of cations on the inhibition of the Asn-Gly succinimide formation reaction.

Table 5.3.2. Comparative analysis of the computed and literature IR frequencies of the backbone carbonyl bond stretches of amide-Ca(II).[389]

Computed		**Literature**		**ΔError (cm^{-1})**
Complex	**C=O bond stretch (cm^{-1})**	**Complex**	**C=O bond stretch (cm^{-1})**	
Asn-Gly- Ca(II) SF Product (Cation inhibits backbone carbonyl)	1660	Amide-Ca(II)	1645	15.0
Asn-Gly- Ca(II) SF Reactant (Cation inhibits backbone carbonyl)	1670	Amide-Ca(II)	1645	25.0
Asn-Gly- Ca(II) SF Reactant (SGP Conformation)	1660	Amide-Ca(II)	1645	15.0

Validation of our most probable spin state analysis was completed via a qualitative comparison of the $\Delta E_{HS\text{-}LS}$ of Fe(II) complexed with a strong field ligand, as given by the computational study by Buijs et al.[406] As we previously noted, carbonyl functional groups are defined as strong field ligands and would therefore favor the low spin state.[412] Evaluation of the energetics of the associated Fe-dipeptide complexes indicated that the low spin state was more energetically favorable, and was validated against the energetics of Fe(II) complexed with BPY (2,2'-bipyridine), a strong field ligand.[406]

5.3.v. Most Probable Spin State

Determination of the most probable electronic spin multiplicity for each of our transition metals was completed after validation of both our computational methodology and chemical model. To further validate our findings regarding the impact of different cations upon the inhibition of Asn-Gly succinimide formation, we conducted a level of theory analysis to identify the most probable spin state of the transition metal complexes analyzed in this study. Evaluation of the most probable spin state was conducted for each modeled structure so that the impact of the site of inhibition on the energetic favorability could be quantified. Specifically, to determine if the site of cation-ligand binding influenced the spin multiplicity that yielded the most stable complex. For the purpose of this study, the most stable spin state was defined as the spin multiplicity that did not cause fragmentation in either the reactant or product structure, and exists in the lowest energy state.

Table 5.3.3. depicts the most favorable spin state of each transition metal as a function of its oxidation state and its site of inhibition. Fe was found to prefer the low spin state regardless of its oxidized form and its site of inhibition, while the other transition metals favored both their low and high spin states. Co(II) and Ni(II) were found to be more energetically favorable in their high spin state at all inhibition sites, but, for the complexes where the cation of interest interacts with the sidechain carbonyl, fragmentation was observed in both the reactant and product structures. Analysis of the complex structures indicated that fragmentation of the high spin structures at the dipeptide backbone carbon-carbon bonds was representative of the system attempting to form

more sigma bonds as an approach towards stabilizing the unpaired electrons present when the structures were optimized in the high spin state. Therefore, due to the structural instability observed at the high spin state of these complexes, the low spin state was determined to be more favorable.

Similarly, the complex formed between Co(III) and the backbone carbonyl of the Asn-Gly succinimide formation product structure was found to fragment. However, unlike the structural instability observed by Co(II) and Ni(II), the fragmentation was not consistent in both the reactant and product structures. Therefore, due to the conflicting spin state favorability between the reactant and product structures, we concluded that Co(III) would not be an appropriate additive to inhibit the formation of iso-Asp-Gly. Similarly, as the structural instability of the Co(II) and Ni(II)-sidechain carbonyl complexes was observed in the product structures, the inconsistency of the spin states between the product structures and the complex formed when the sidechain and backbone carbonyls were present in the same geometric plane eliminated those complexes from consideration.

Lastly, the Co(III) and Fe(III) complexes formed at the sidechain carbonyl were eliminated from our analyzed data set due to the observed fragmentation in both the low and high spin states. Fragmentation of the Co(III) complex was only observed in the reactant structure, while the Fe(III) complex fragmentation was observed at both the reactant and product structures. Evaluation of the Co(III) sidechain carbonyl complex showed interactions between the cation and the stabilizing water within the system, suggesting that the oxide formed stabilized the structure against fragmentation. Prior studies had highlighted the propensity of Fe(III) towards fragmentation of the protein backbone.[394] Furthermore, evaluation of the low and high spin structures indicated that no oxides were formed to stabilize against the unpaired electrons present in both structures. Therefore, Fe(III) was eliminated from consideration of effective cations for the inhibition of the succinimide intermediate at the sidechain carbonyl.

Table.5.3.3. The most probable spin state of the transition metal-dipeptide complexes modeled as a function of transition metal oxidation state and site of inhibition. ǂDenotes favorability determined via structural stability instead of via $\Delta E_{HS\text{-}LS}$. łDenotes structural instability observed in both the low and high spin states. §Denotes conflicting spin state between reactant and product structures due to structural instability.

Cation	Inhibits	Favors
Co(II)	Reactant in cis conformation	High spin
Co(II)	Backbone carbonyl	High spin
Co(II)ǂ	Sidechain carbonyl	Low spin
Co(III)	Reactant in cis conformation	Low spin
Co(III)§	Backbone carbonyl	Low spin
Co(III)ł	Sidechain carbonyl	Low spin
Fe(II)	Reactant in cis conformation	Low spin
Fe(II)	Backbone carbonyl	Low spin
Fe(II)	Sidechain carbonyl	Low spin
Fe(III)	Reactant in cis conformation	Low spin
Fe(III)	Backbone carbonyl	Low spin
Fe(III)ł	Sidechain carbonyl	Low spin
Mn(II)	Reactant in cis conformation	Low spin
Mn(II)	Backbone carbonyl	High spin
Mn(II)	Sidechain carbonyl	Low spin
Mn(III)	Reactant in cis conformation	Low spin
Mn(III)	Backbone carbonyl	Low spin
Mn(III)	Sidechain carbonyl	Low spin
Ni(II)	Reactant in cis conformation	High spin
Ni(II)	Backbone carbonyl	High spin
Ni(II)ǂ	Sidechain carbonyl	Low spin

5.4. Results and Discussion

5.4.i. Structures

All cation-dipeptide complexes were modeled using the same hybrid solvation methodology as our prior studies on deamidation to ensure that the Gibbs free energies of our modeled complexes were consistent with those observed in the uninhibited Asn-Gly succinimide formation reaction. Therefore, all structures included a single water acting as a passive stabilizer at the N-terminus Asn carbonyl. The position of the cation of interest relative to Asn-Gly was representative of the site of inhibition, where three types of complexes are formed. The first modeled complex prohibits the formation of Asp-Gly through the succinimide reaction pathway, where the cation of interest interacts with the sidechain carbonyl of the N-terminus Asn (Figure 5.4.1.a). Similarly, to inhibit the formation of iso-Asp-Gly via the succinimide reaction pathway, the cation of interest forms a complex with the backbone carbonyl (Figure 5.4.1.b). The last cation-dipeptide complex situates the cation between the sidechain and backbone carbonyls, where the reactant structure has been manipulated such that the sidechain and backbone carbonyls are in the same geometric plane (SGP), and the product structures are determined via the inhibition site of interest (Figures 5.4.1.c and 5.4.1.d). The bond lengths observed between the cation of interest and the corresponding carbonyl oxygen were found to be consistent with previously reported values (ΔError < 0.2Å).[413-421] Of the structures analyzed, the complex formed between Ni(II) and the backbone carbonyl was found to have a significantly different value relative to the literature, and can be attributed to the stability of the structure in its high spin state.[420]

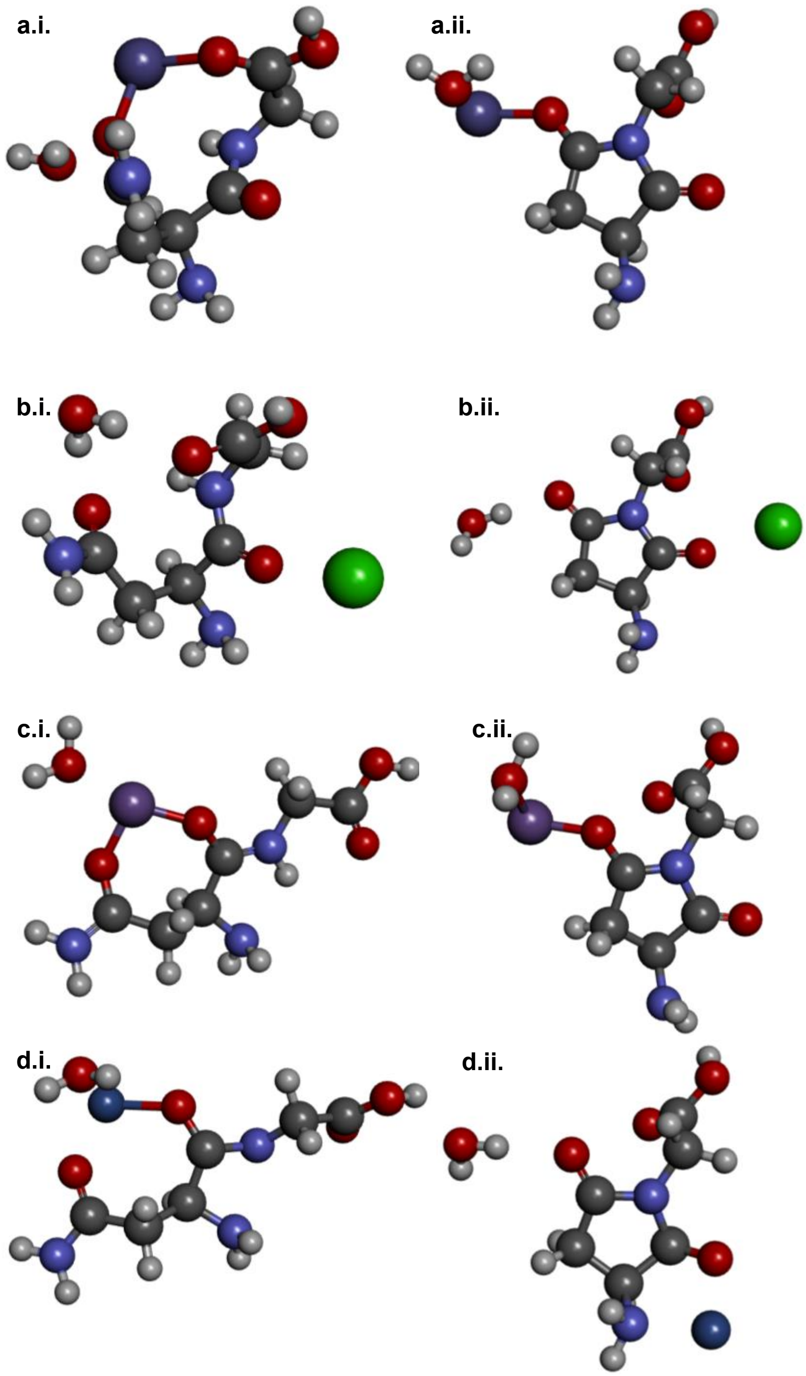

Figure 5.4.1. Sample ball and stick representation of the three different types of cation-dipeptide interactions to inhibit Asn-Gly succinimide formation. (A). Cation inhibits sidechain carbonyl: (i). Reactant structure, (ii). Product structure; (B). Cation inhibits backbone carbonyl: (i). Reactant structure, (ii). Product structure; (C). Cation inhibits sidechain carbonyl, cis conformation: (i). Reactant structure (cis conformation), (ii). Product structure; (D). Cation inhibits backbone carbonyl, cis conformation: (i). Reactant structure (cis conformation), (ii). Product structure.

Evaluation of the reactant structures of the three different types of site-specific inhibition analyzed within this study indicated that the two of the three typical types of cation coordination, noted in prior studies, are observed.[416, 422-425] As denoted by Deacon and Phillips in 1980, the three types of cation-amide coordination are “unidentate, bidentate, and bridging”, where the type of coordination is denoted by the number of oxygen atoms interacting with the cation as well as the positioning of the formed complex.[416, 422-425] Unidentate coordination is observed at all Asn-Gly succinimide formation product structures regardless of the inhibition site of interest, where the distance between the cation of interest and the carboxylic acid of the C-terminus Gly residue is greater than 3.7Å. Examination of the complexes where bidentate coordination was observed indicated that 1.80 Å < Cation-COO^- < 2.37 Å, suggesting that the observed coordination of the cation-amide complex formation in the reactant structures is determinant on cation proximity to multiple carbonyls. Inverse bridging is observed in the Asn-Gly succinimide product structure of the complexes formed at the sidechain carbonyl oxygen, where, instead of a ligand forming a complex with two cations, the cation of interest forms a complex with the sidechain carbonyl oxygen and the stabilizing water molecule within the system.[422, 424-426] The interactions observed between the stabilizing water and the cation-carbonyl complex suggests further stabilization within the system, where the cation of interest attempts to form a metal oxide with the stabilizing water.[427-429]

5.4.ii. Site-Specific Binding Free Energies

The binding free energies of the site-specific complexes evaluated in this study are present in Table 5.4.2. Prior studies have indicated that the associated free energies of protein binding are spontaneous, where the affinity of the cation towards forming a complex with the carbonyl of interest is reflected in its spontaneity.[430-440] Furthermore, previous studies have indicated that irreversible binding was often found in strongly exergonic reactions, while weaker binding was found to be reversible.[441, 442] The irreversibility of inhibitors complexed with proteins has been found to induce adverse effects among administration, as well as reduce efficacy due to the stability of the formed complex.[442-445] Evaluation of the associated $\Delta G_{BINDIING}$ of protein-ligand complexes in the literature indicated that a free energies of binding were not greater than -30 kcal.mol^{-1}, therefore we considered all binding energies greater than this value to be irreversible.[430-440]

Table 5.4.2. Computed binding free energies of the cation-dipeptide complexes formed for the site-specific inhibition of Asn-Gly succinimide formation.

Sidechain Carbonyl		**Backbone Carbonyl**		**Backbone and Sidechain Carbonyl in Same Geometric Plane**	
Cation	**$\Delta G_{BINDING}$ (kcal.mol^{-1})**	**Cation**	**$\Delta G_{BINDING}$ (kcal.mol^{-1})**	**Cation**	**$\Delta G_{BINDING}$ (kcal.mol^{-1})**
Ca(II)	-3.03	Ca(II)	-3.41	Ca(II)	-7.45
Co(II)	-50.3	Co(II)	-75.5	Co(II)	-84.3
Fe(II)	-33.6	Fe(II)	-17.5	Co(III)	-187
Mg(II)	-20.5	Fe(III)	-181	Fe(II)	-71.3
Mn(II)	-60.1	Mg(II)	-13.1	Fe(III)	-184
Mn(III)	-125	Mn(II)	-19.5	Mg(II)	-34.6
Ni(II)	-64.1	Mn(III)	-144	Mn(II)	-65.2
Zn(II)	-21.6	Ni(II)	-87.1	Mn(III)	-158
		Zn(II)	-12.8	Ni(II)	-114
				Zn(II)	-35.8

Evaluation of the three sites of inhibition indicated that the weakest complexes were formed by the interactions of Ca(II) and the Asn-Gly reactant structure, regardless of the site of inhibition. Prior studies have indicated that Ca(II) forms weak bonds with amide carbonyls.[389, 411, 446] Furthermore, it was the only cation to be cited by Rehder to be effective in reducing the rate of succinimide formation, suggesting that the binding of the additive to the carbonyl of interest is reversible.[104] Evaluation of complexes formed between the cations of interest and the dipeptide functional groups at each of three specified sites indicated that the binding free energies of the cation-dipeptide complexes could be split into two groups: cations classified as or exhibiting behavior similar to alkaline earth metals, and cations that are classified as transition metals. As we previously noted, Zn(II) has been noted by Bestgen to exhibit behavior similar to alkaline earth metals due to the closed shell nature of its electron configuration.[410] Additionally, an experimental spectroscopy study by Okur et al. in 2013 evaluated the binding of alkaline earth metals to the peptide backbone, where weak binding was exhibited between the cation of interest and the peptide backbone.[389] Furthermore, a study by Song indicated that the open shell nature of the d-orbital of transition metals would result in more stable binding relative to the complexes formed by other metals and proteins.[447] As a result, stronger binding free energies would be observed by the complexes formed by transition metals.

Analysis of the complexes formed at the backbone and sidechain carbonyls indicated that both the oxidation number and cation coordination impacted the binding free energies. Specifically, the binding free energies of the +3 oxidative species were found to double the binding free energies of the +2 oxidative species, and can be attributed to the electrostatic forces exhibited by the higher oxidation number.[448]

As we previously noted, the complexes formed at the sidechain carbonyl were found to observe bidentate coordination in the reactant structure of Asn-Gly succinimide formation, and observed a greater affinity for binding relative to the complex formed at the backbone carbonyl. Prior studies have indicated that bidentate coordination is more favorable than unidentate coordination, and this was observed for four of the seven complexes with an oxidation number of +2.[449] The three complexes that were found to deviate from the trend were Ca(II), Co(II) and Ni(II), where Ca(II) observed unidentate coordination at both binding sites and was found to exhibit similar binding free energies. The higher affinity for complex formation observed by Co(II) and Ni(II) can be attributed to our evaluation of the most probable spin state, where the structural instability of the complexes at their respective high spin states resulted in the low spin state energies being used in our analysis.

Evaluation of the binding free energies at this site suggests that the bidentate coordination observed by all cations improves the binding affinity of the complexes because of the proximity of both carbonyls to the cation of interest when present in the same geometric plane.[449] The Ca(II) complex formed when the backbone and sidechain carbonyls are in the same geometric plane was found to be twice as strong as the Ca(II) complexes formed at the backbone and sidechain carbonyls. Additionally, it was significantly weaker than the complexes formed by the other cations analyzed at the specified site of inhibition. Prior studies have indicated that complexes formed by Ca(II) and neutral groups are more energetically stable through unidentate coordination compared to bidentate coordination.[450] Therefore, it would not have as strong an affinity as the other cation complexes formed when the backbone and sidechain carbonyls are in the same geometric plane.

5.4.iii. Impact of Site-Specific Cation Amide Complexes for the Inhibition of Asn-Gly Succinimide Formation

Rehder determined that the inclusion of $CaCl_2$ to the formulation was found to successfully inhibit the formation of iso-Asp via the succinimide formation pathway.[104] Specifically, the rate of succinimide formation was slowed after the inclusion of the additive within the formulation.[437] As we previously noted, the driving force of the succinimide reaction pathway is its spontaneity under ambient conditions.[82, 273, 279] Therefore, it is likely that the inclusion of the cation to the mAb formulation impacts the spontaneity of the succinimide formation reaction and, as a result, reduces the kinetics associated with the formation of the intermediate.[104] For this purpose, evaluation of the screened cations as effective inhibitors of Asn-Gly succinimide formation was completed by comparing the inhibited and uninhibited free energies of succinimide formation.

5.4.iii.a. Inhibition of Asp-Gly Formation via the Succinimide Reaction Pathway

Of the eight complexes where the cation of interest inhibits the sidechain carbonyl, only Fe(II) was found to be ineffective in reducing the spontaneity of the formation of the Asn-Gly succinimide intermediate (Table 5.4.3, Figure 5.4.2). The Gibbs free energy of succinimide formation associated with Asn-Gly complexed with Fe(II) was found to be about twice as spontaneous as the uninhibited reaction, suggesting that the cation acts as a promoter instead of an inhibitor at this site. A study by Gangwar et al. had noted that the presence of Fe(II), Mn(II), and Ni(II) in the media of the cultured CHO cells resulted in the formation of more acidic species within the formulation, such as Asp-Gly and iso-Asp-Gly.[451] Furthermore, previous studies have indicated that Asp is a strong ligand for iron binding in physiological processes.[452-454] Therefore, it is likely that Fe(II) would promote the formation of a product that it can form a strong complex with, Asp-Gly. Similar to the observed free energies of binding between the carbonyl of interest and the cation, the free energy associated with the formation of the succinimide intermediate at the +3 complexes doubled the free energies observed at the complexes formed at the +2 complexes. As we previously noted, the +3 oxidation state experiences a stronger electronegative force compared to the +2 oxidation state.[448] Therefore, it is likely that the stability of the complex formed by the reactant structure inhibits the self-promoted formation of the succinimide intermediate.

Table 5.4.3. Ranked $\Delta G_{INHIBITION}$ of cations inhibiting Asn-Gly succinimide formation (cation inhibits sidechain carbonyl).

Cation	ΔG_{SF} (kcal.mol^{-1})	$\Delta G_{INHIBITED} > \Delta G_{UNINHIBITED}$ (kcal.mol^{-1}) >	Ranking [Strongest to weakest inhibition]
Uninhibited	-4.70	N/A	N/A
Mn(III)	11.0	Yes	1
Mn(II)	6.66	Yes	2
Ni(II)	2.60	Yes	3
Co(II)	1.64	Yes	4
Mg(II)	0.62	Yes	5
Ca(II)	-1.94	Yes	6
Zn(II)	-3.99	Yes	7
Fe(II)	-9.64	No	8

Evaluation of the Cation^{2+} complexes indicated that all complexes, aside from Fe(II), exhibited weak inhibition of succinimide formation. Furthermore, only the Ca(II) and Zn(II) complexes were found to yield a spontaneous ΔG_{SF}. As we previously noted, prior studies have noted that Ca(II) binds weakly to amide carbonyls.[389, 411, 446] Therefore, it would not be as effective an inhibitor as the other cations complexed with the sidechain carbonyl. Zn has been identified as a cofactor for the deaminase enzyme class, a class of enzymes that produce ammonia from amides via the Zn coordinating with water to hydrolyze the amide.[455, 456] Given that the complex formed at the sidechain carbonyl would inhibit the formation of Asp, the only product of deamidation via the direct hydrolysis reaction pathway, it is likely that Zn would be a less effective inhibitor at this site.[80, 82, 129, 147, 149-153, 242-244, 277] Furthermore, evaluation of the ΔG_{SF} associated with Ca(II), Mg(II), and Zn(II) relative to the ΔG_{SF} associated with the transition metal-SC complexes indicated that the trends observed in the binding free energies impacted the effectiveness of the cation as an inhibitor of Asp-Gly formation. Specifically, the stability of the transition metal dipeptide complexes, relative to the complexes formed by alkaline earth metals, resulted in a greater impact on the spontaneity of the Asn-Gly succinimide formation reaction.

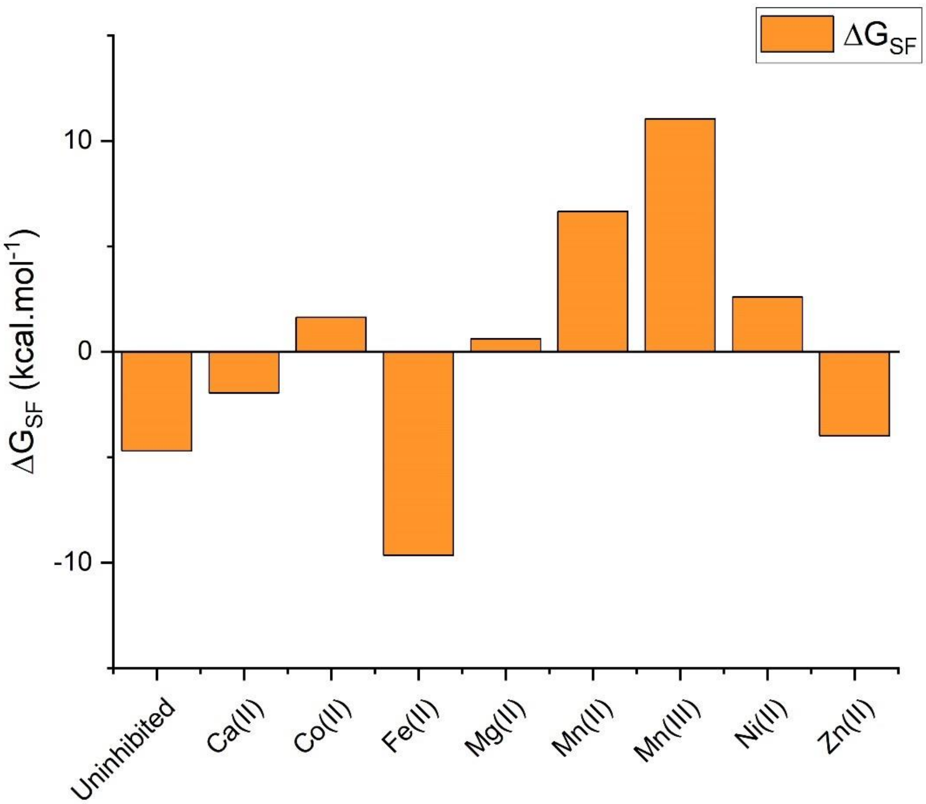

Figure 5.4.2. Bar plot of $\Delta G_{INHIBITION}$ of cations inhibiting Asn-Gly succinimide formation (cation inhibits sidechain carbonyl).

5.4.iii.b. Inhibition of iso-Asp-Gly Formation via the Succinimide Reaction Pathway

Of the nine complexes where the cation of interest inhibits the backbone carbonyl, only Ni(II) was found to be ineffective in reducing the spontaneity of the formation of the Asn-Gly succinimide intermediate (Table 5.4.4, Figure 5.4.3). As we previously noted, Ni(II) was cited as a metal within the CHO cell culture medium that promoted the formation of acidic species during mAb manufacturing.[451] Furthermore, unidentate coordination was observed in both the reactant and product structures of the complexes formed at the backbone carbonyl. As a result, it is likely that the cation would promote succinimide formation when interacting at the backbone carbonyl.[449]

Table 5.4.4. Ranked $\Delta G_{INHIBITION}$ of cations inhibiting Asn-Gly succinimide formation (cation inhibits backbone carbonyl).

Cation	ΔG_{SF} (kcal.mol^{-1})	$\Delta G_{INHIBITED}$ > $\Delta G_{UNINHIBITED}$ (kcal.mol^{-1}) >	Ranking [Strongest to weakest inhibition]
Uninhibited	-4.70	N/A	N/A
Mn(III)	47.1	Yes	1
Fe(III)	18.1	Yes	2
Co(II)	8.23	Yes	3
Fe(II)	3.99	Yes	4
Mn(II)	2.51	Yes	5
Mg(II)	2.36	Yes	6
Zn(II)	1.90	Yes	7
Ca(II)	-1.14	Yes	8
Ni(II)	-7.47	No	9

Evaluation of the complexes formed at the backbone carbonyl indicated that similar trends were observed at both the backbone and sidechain carbonyl complexes. Specifically, the complexes formed by cations with a +3 oxidation state were twice as inhibitive as those with a +2 oxidation state. Furthermore, Zn(II) and Ca(II) were found to be the weakest inhibitors of succinimide formation when complexed with the backbone carbonyl. However, the inhibition of the Zn(II) complex was found to be endergonic instead of exergonic, suggesting that the Zn(II) would be more effective for inhibiting the formation of iso-Asp-Gly due to its role as a cofactor for hydrolytic removal of ammonia.[455, 456] Similar to the complexes formed between the cation of interest and the sidechain carbonyl, the weak effectiveness of the complexes formed by Ca(II), Mg(II), and Zn(II) relative to the transition metal complexes indicates that the behavior of the cation of interest directly impacts its effectiveness in inhibiting iso-Asp-Gly formation.

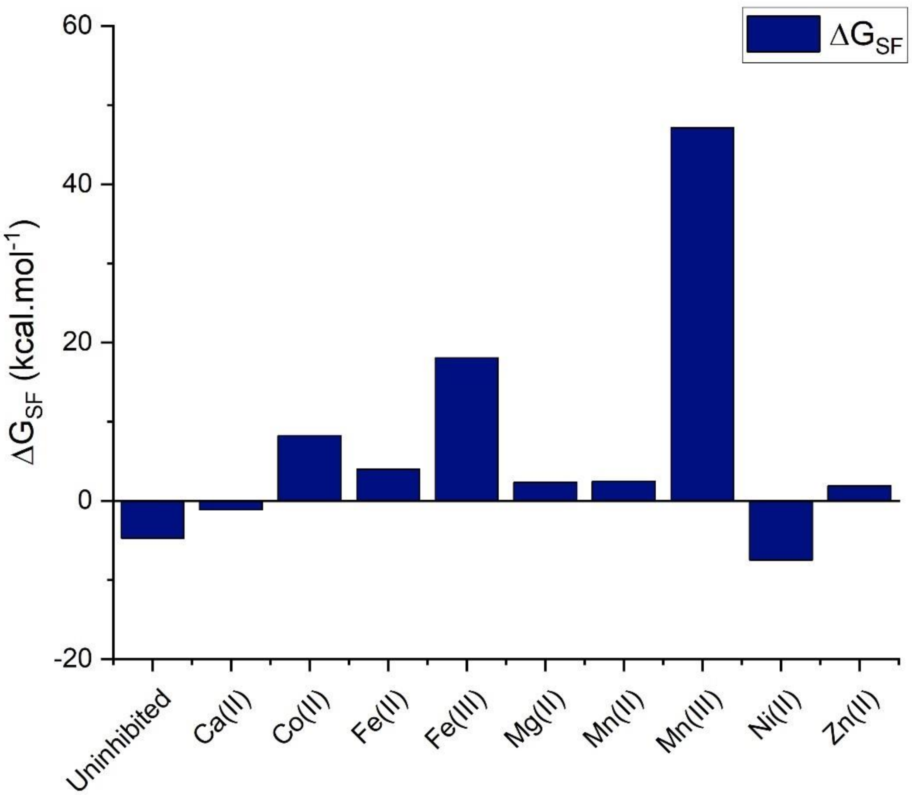

Figure 5.4.3. Bar plot of $\Delta G_{INHIBITION}$ of cations inhibiting Asn-Gly succinimide formation (cation inhibits backbone carbonyl).

5.4.iii.c. Inhibition of Asp-Gly and iso-Asp-Gly Formation via the Succinimide Reaction Pathway when the Backbone and Sidechain Carbonyls are in the Same Geometric Plane

Evaluation of the structural impact of manipulating the Asn-Gly reactant structure so that both the sidechain and backbone carbonyls are in the same geometric plane indicated that the uninhibited reactant structure promoted succinimide formation (Table 5.4.5, Figure 5.4.4). A study by Liu et al. indicated that succinimide formation was more readily observed in protein splicing when the sidechain carbonyl of a neighboring Asn residue was in a position to interact with the backbone carbonyl, suggesting that the conformation of the interacting residues impacted the rate of succinimide formation.[457] Given that the distance between the carbonyls decreases by over 1Å when the dipeptide backbone is manipulated to put both functional groups in the same geometric plane, it is likely that the sidechain carbonyl promotes succinimide formation when a cation inhibitor is not present within the system.

Table 5.4.5. Ranked $\Delta G_{INHIBITION}$ of cations inhibiting Asn-Gly and iso-Asp-Gly succinimide formation, backbone and sidechain carbonyls in same geometric plane in reactant structure).

Cation	**ΔG_{SF} (kcal.mol^{-1})**	**$\Delta G_{INHIBITED} > \Delta G_{UNINHIBITED}$ (kcal.mol^{-1})**	**Ranking [Strongest to weakest inhibition]**
Uninhibited	-4.70	N/A	N/A
Mn(III) [iso-Asp]	60.6	Yes	1
Fe(II) [iso-Asp]	57.2	Yes	2
Mn(III) [Asp]	44.0	Yes	3
Fe(II) [Asp]	27.5	Yes	4
Zn(II) [iso-Asp]	24.4	Yes	5
Mg(II) [iso-Asp]	23.3	Yes	6
Fe(III) [iso-Asp]	21.1	Yes	7
Co(III) [Asp]	20.5	Yes	8
Ni(II) [iso-Asp]	18.8	Yes	9
Co(II) [iso-Asp]	16.5	Yes	10
Mg(II) [Asp]	14.2	Yes	11
Mn(II) [Asp]	11.2	Yes	12
Zn(II) [Asp]	9.68	Yes	13
Ca(II) [iso-Asp]	2.33	Yes	14
Ca(II) [Asp]	1.90	Yes	15
Uninhibited, SGP	-5.27	No	16

Evaluation of the cation-amide complexes formed when the dipeptide backbone and sidechain carbonyls are in the same geometric plane indicated that all free energies were endergonic and greater than the free energies observed by the complexes formed at the backbone and the sidechain carbonyls. As we previously noted, the SGP complexes observed greater binding affinity due to bidentate coordination, and the coordination was found to be more stable due to the closer proximity of the carbonyls to the cation of interest.[449] Therefore, it is likely that the enhanced stability of the complexes formed in the reactant structure deters Asn-Gly from spontaneously cyclizing into the succinimide intermediate.

Similar to the observed binding free energies, ΔG_{SF} of the observed Ca(II) complexes were significantly smaller compared to the other complexes formed at the specified site. As we previously noted, Ca(II) is the only cation cited by Rehder for successfully reducing the rate of succinimide formation, and we hypothesized that this was due to the ability of the cation to bind weakly to the amide carbonyl.[104] Furthermore, prior studies have indicated that complexes formed by Ca(II) and neutral groups are more energetically stable through unidentate coordination compared to bidentate coordination.[450] Therefore, the formation of the succinimide intermediate at these complexes would not be non-spontaneous as those formed by the other cations.

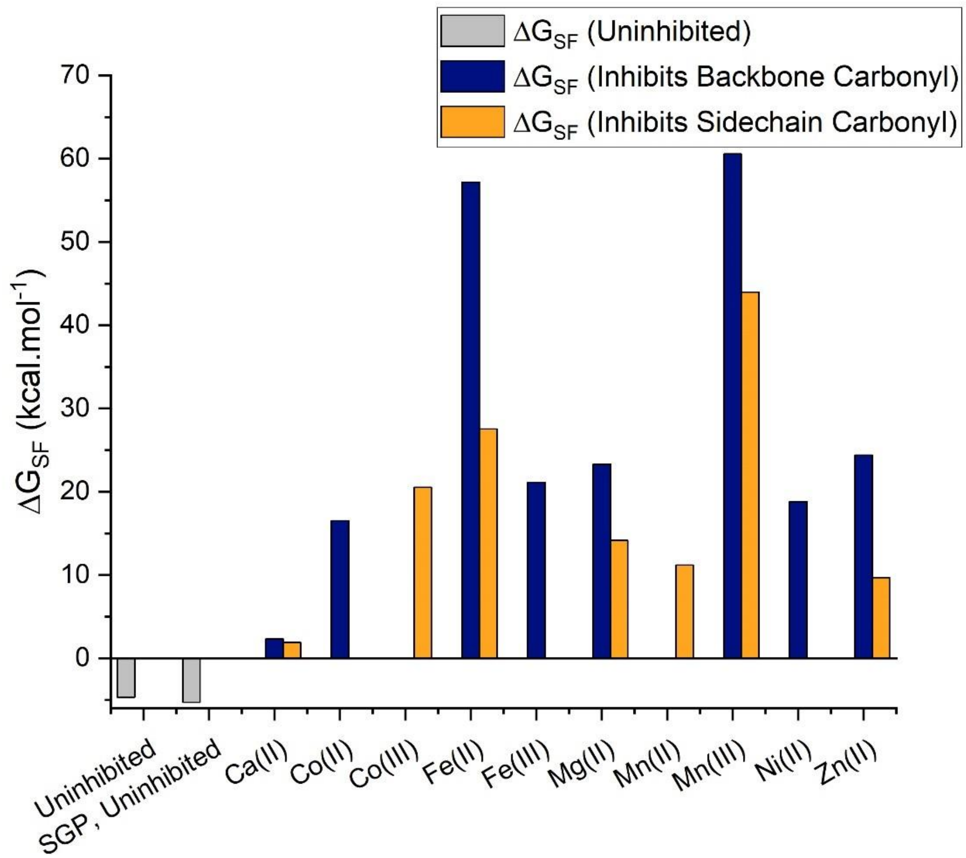

Figure 5.4.4. Bar plot of ΔG$_{INHIBITION}$ of cations inhibiting Asp-Gly and iso-Asp-Gly succinimide formation (backbone and sidechain carbonyls in same geometric plane in reactant structure).

5.4.iv. Screening of Cations for Successful Inhibition of Asn-Gly Succinimide Formation

Screening of the cations noted by Rehder was contingent on two metrics – probability of reversible binding and effectiveness in reducing the spontaneity of the Asn-Gly succinimide formation reaction. Through the development of three different complexes, we achieved a site-specific analysis of the effectiveness of each cation. For the inhibition of Asp-Gly formation via the succinimide reaction pathway, the complexes formed by Ca(II), Mg(II), and Zn(II) were the only complexes that met both criteria. In addition to these three cations, the complexes formed by Fe(II) and Mn(II) were found to successfully inhibit iso-Asp-Gly formation.

Evaluation of the differences between the two complexes suggests that the type of cation coordinate observed between the two sites was most impactful, where prior studies have indicated Fe and Mn are more stable in the bidentate coordination, as represented by their respective binding free energies.[458, 459] Furthermore, the role of Zn(II) as a cofactor for deaminases indicated that it would be a more effective inhibitor of iso-Asp formation relative to Asp formation.[455, 456] Lastly, the manipulation of the backbone such that the backbone and sidechain carbonyls were present in the same geometric plane increased reaction spontaneity while uninhibited. However, the bidentate coordination observed between the backbone and sidechain carbonyls increased binding affinity such that only the complex between Ca(II) and the dipeptide would be reversible. Therefore, the inhibition of the succinimide intermediate at this site would be undesired.

5.5. Conclusions and Future Directions

In summary, a theoretical investigation on the effectiveness of cations as inhibitors of Asn-Gly deamidation via the succinimide reaction pathway was completed, and a database of cation-dipeptide complexes was generated at the modeled dielectric associated with bulk water. Evaluation of the impact of the cations on reducing the spontaneity of Asn-Gly succinimide formation was achieved through a site-specific approach, where the 48 cation-dipeptide complexes were representative of the three major sites of inhibition: the sidechain carbonyl, the backbone carbonyl, and when the two carbonyls are present in the same geometric plane. Model validation was achieved by comparing the bond stretch of the Ca(II)-backbone carbonyl to experimental results, as well as the most probable spin state of the Fe(II) complexes.[389, 406] Evaluation of the site-specific binding free energies indicated that the strongest energy was observed for complexes including cations with a +3 oxidation state, while bidentate coordination was found to improve the binding affinity of the cation-dipeptide complexes, and similar energetic trends were observed in the inhibition of the succinimide intermediate.[448, 449] Additionally, the complexes formed with transition metals were found to bind more strongly and have a greater impact on reducing the spontaneity of the reaction than the complexes exhibiting alkaline earth metal behaviour; where the observed trends are representative of more stable complexes formed between the transition metals and proteins.[447]

Lastly, screening of cation effectiveness towards reducing the formation of the succinimide intermediate indicated that three parameters impacted if the cation of interest was likely to bind reversibly and effectively reduce the spontaneity of Asn-Gly succinimide formation. The complexes formed at the backbone carbonyl were found to be more effective than those formed at the sidechain carbonyl due to the observed cation coordination in the reactant structure, while the +3 oxidation state was found to produce the strongest inhibition at both complexes. Additionally, the complexes formed using Ca(II) were found to be the weakest effective inhibitors regardless of the site of inhibition, and can be attributed to the propensity of the cation towards unidentate coordination.[450] Future studies will expand the generated database to include the other three dielectrics modeled in prior studies completed by our group so that the electrostatic impact can be analyzed. Additionally, complexes with other C-terminus amino acids will be modeled to determine if cation effectiveness is contingent on the sidechain functional group of the C-terminus residue.

Chapter 6: Towards Pharmaceutical Protein Stabilization: Generalizing the Energetics of Non-enzymatic Peptide Bond Fragmentation during the Alkaline Wash Step of Downstream Manufacturing

6.1. Background and Motivation

pH-dependent studies on protein degradation are of interest due to the environmental factors associated during the downstream manufacturing portion of the pharmaceutical protein manufacturing pipeline. Downstream manufacturing of mAbs focuses on the extraction and purification of the mAb from its host cell culture, and can be broken up into three criteria – extraction, purification, and polishing.[52, 460-464] Extraction and purification of the protein of interest from its host cell culture occurs under both acidic and basic conditions.[65-67, 70-72, 74, 76, 81, 465-475] Specifically, through a methodology patented by Breece et al. in 2003 (Patent ID: EP2048154).[465] Specifically, a wash step is included to the accepted protein elution technique, protein A chromatography, to purify the mAb from any impurities prior to its elution from the column.[465] This methodology was then expanded to more alkaline conditions, and used as a technique for reducing aggregation and protein self-association during protein extraction, where the wash buffer pH ranged from $9 < pH < 12$ (Patent IDs: SG10201908027Q (Wang) and WO2016031932 (Yamanouchi)).[67, 68, 470]

While it was determined that the inclusion of the alkaline wash step within protein A chromatography was favorable for reducing aggregation, multiple studies have found perturbations in the system pH to negatively impact protein stability and increase the rate of protein degradation.[138, 142, 145, 470, 476-481] Within the mAb, the hinge region is an area of the protein that is most significantly impacted by changes in pH (Figure 6.1.1).[138, 145, 476] Furthermore, the hinge region is prone towards structural degradation, where fragmentation of peptide bonds within the polypeptide chain leads to structural instability.[128, 138, 142-145, 476] During the alkaline washing stage of downstream manufacturing, peptide bond fragmentation occurs through the electrophilic attack of OH^-, and the propensity of the polypeptide towards fragmentation under basic conditions is likely representative of the electrostatic interactions observed within the reaction.[142-144, 335]

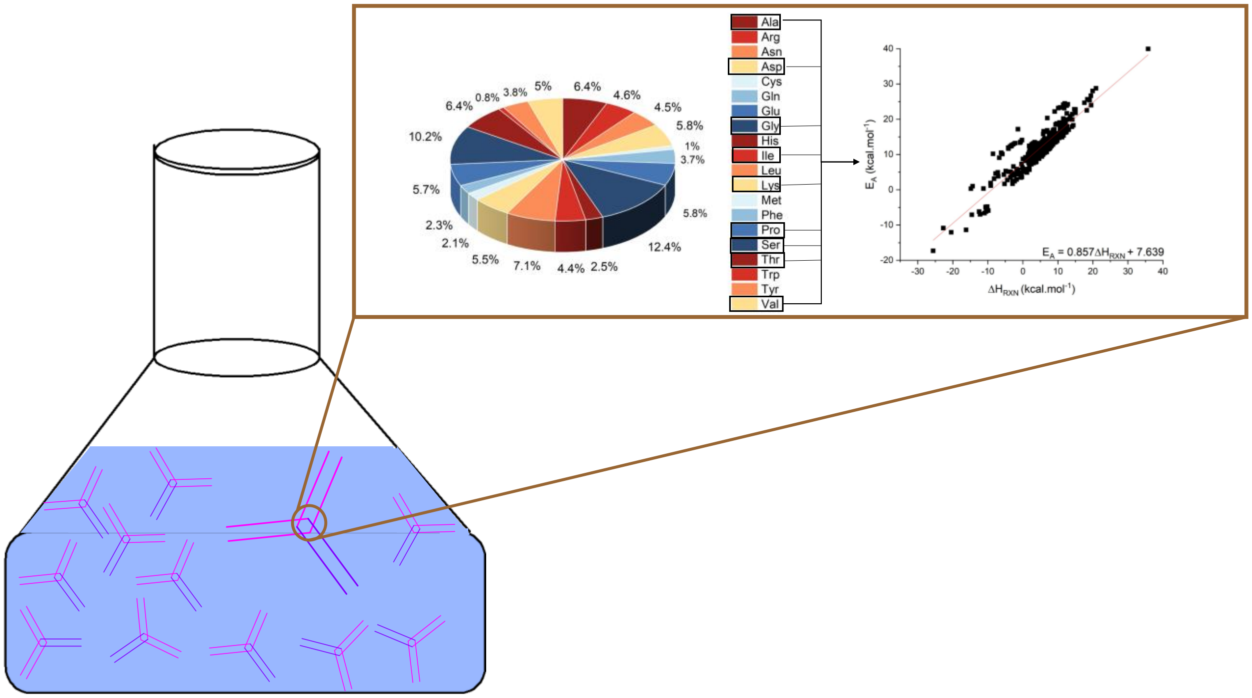

Figure 6.1.1. Multi-scale representation of non-enzymatic peptide bond fragmentation on hinge region fragmentation as a function of the pH ranges associated with the alkaline wash stage of mAb downstream manufacturing.

A study by Kumar and Nussinov classifies the interactions within a protein structure as either “specific” or “non-specific”, where electrostatic interactions are defined as the first category and can contribute towards protein stability.[482] Specifically through the salt bridges and hydrogen bonds that polar or charged residues can form with each other and with the solvated environment.[138, 482-484] However, the electrostatic interactions observed within the polypeptide structure can also be destabilizing, such as the repulsive nature between two like-charged residues.[485, 486] Therefore, it is critical to account for the impact of the different residue functionalities upon reaction energetics. The advancement of bioinformatics have provided new insights towards how the structural integrity of the hinge region can be affected by the amino acid residues most prevalent within the hinge region.[163] As a result, the impact of electrostatic interactions on non-enzymatic peptide bond fragmentation under basic conditions can be modeled using amino acids that are most prevalent within the hinge region. Furthermore, the thermochemical and kinetic parameters predicted for those reactions can be used to generalize the energetics associated with the reaction.

For this purpose, we conducted a computational study, using a subset of dipeptides, to predict the thermochemistry and kinetic parameters of the acidic and basic non-enzymatic peptide bond fragmentation mechanisms. This paper presents the thermochemical and kinetic properties non-enzymatic peptide bond fragmentation under basic conditions (180 reactions (Appendix D) per modeled dielectric, 720 total reactions) in the conductor-like polarizability model (CPCM) solvated state, i.e. standard enthalpies of reaction, standard entropies of reaction, Gibbs free energies of reaction, activation energies, and kinetic rate constants. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic energy, and then statistical thermodynamics was applied to all fragmentation reactions to determine temperature effects on the reaction.[82, 141, 147, 187-190, 246, 247] All thermochemical properties were calculated at standard pressure and temperature to reduce external variabilities. Furthermore, a statistically significant generalization was achieved so that the activation energy of the proposed rate-determining step of the dipeptide fragmentation reaction of interest could be predicted from its reaction enthalpy.

6.2. Study Objectives

In this study, the impact of the non-enzymatic peptide bond fragmentation on the stability of the hinge region of the mAb was analyzed relative as a function of the alkaline wash stage of downstream manufacturing. As we previously specified, the purpose of the compiled works is to analyze the impact of different stages of the mAb manufacturing pipeline on functional and structural degradation. During downstream manufacturing, where the proteins are subjected to perturbations in pH, structural degradation is the more prevalent than functional degradation. Especially in the hinge region, where fragmentation of the peptide bond within the polypeptide chains results in structural instability. For this purpose, we leveraged our understanding of peptide bond fragmentation, from our study under ambient conditions, to determine if the energetics of the proposed rate-determining step of peptide bond fragmentation could be generalized under the environmental conditions associated with the alkaline wash step of the mAb downstream manufacturing. Furthermore, by modeling dipeptide reactions that only contained the nine amino acids that occur at the highest frequency within the hinge region, we were able to analyze the impact of these amino acids on reaction energetics. Specifically, we sought to determine how the residue functional groups and the position of the residue of interest within our modeled dipeptide influenced the rate coefficients of peptide bond fragmentation. Overall, this study examines the reaction energetics of the basic mechanism of peptide bond fragmentation and identifies key correlations that can be used to generalize the reaction behavior of these mechanisms.

6.3. Computational Methodology

As we previously mentioned, this study builds upon the foundations of the developed reaction network of non-enzymatic peptide bond fragmentation under ambient conditions. For this purpose, we implemented a similar base methodology for this study to reduce variability within our analysis and maintain a self-consistent methodology amongst our developed databases.[82, 141, 147, 251]

6.3.i. Proposed Reaction Mechanism and Reaction Class

Modeling of non-enzymatic peptide bond fragmentation under basic conditions was completed via a stepwise approach, where the electrophilic attack of OH^- resulted in the formation of a five-membered intermediate centered around the peptide bond, and was hypothesized as the proposed rate-determining step.[142] Once the intermediate was formed, a subsequent hydrogen transfer cleaved the central amide bond to form the free amino acids.[142] This mechanism agrees with experimental studies on peptide bond fragmentation under basic conditions and is depicted in Figure 6.3.1.

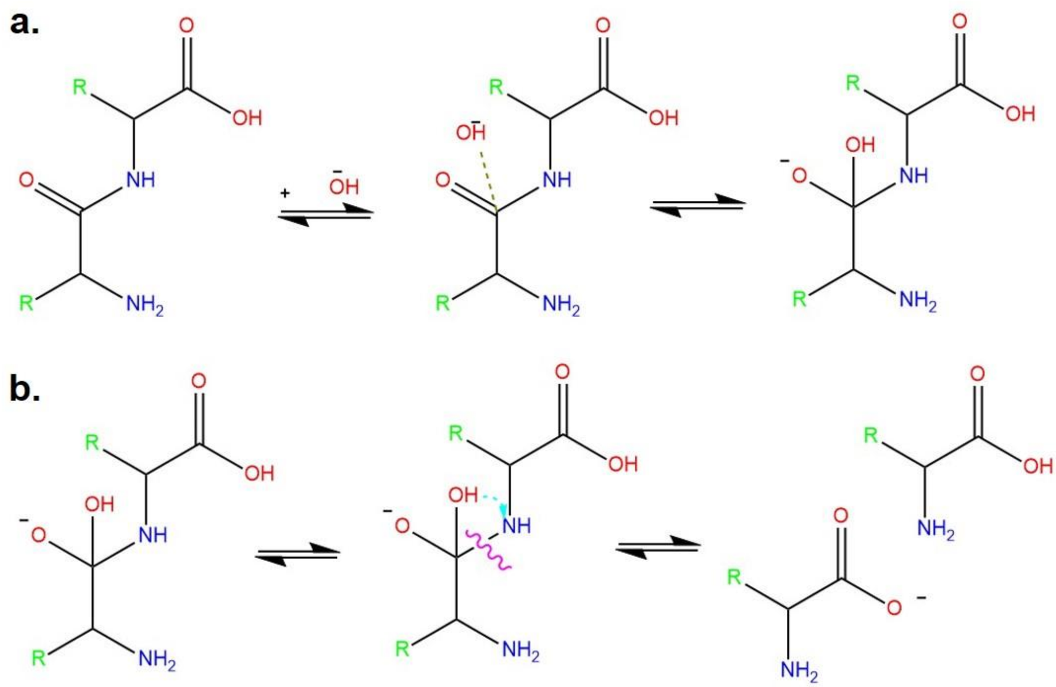

Figure 6.3.1. Proposed mechanism of non-enzymatic peptide bond fragmentation under basic conditions. (a). The electrophilic attack of OH^- to form an intermediate (Proposed rate-determining). (b). Proton transfer to cleave the central amide bond and form free amino acids in solution.

As we have previously mentioned, the dipeptide is the smallest model system through which polyamide peptide bond hydrolysis can be studied. Furthermore, there are 400 possible dipeptide combinations that can be modeled within the system. To reduce our training set to a more manageable number of reactions, as well as to model reactions that are more likely to occur within the hinge region of the mAb, we derived our model dipeptides from the most frequently occurring amino acids present in the hinge region. A bioinformatics study conducted by Flores et al. identified how often each amino acid appeared within the hinge region, where small, flexible residues such as Glycine (Gly) and Serine (Ser) were the most common occurrences.[163, 164] Through this study, we were able to identify the 10 amino acids that were most likely to appear within the hinge region of the protein.[163] Of those 10 amino acids, Aspartic Acid (Asp) and Glutamic Acid (Glu) were found to occur at the same frequency.[163] Therefore, as Asp and Glu share the sidechain classification of 'acidic', we hypothesized that the reaction energetics exhibited by the pH-dependent hydrolysis of Asp-containing dipeptides would be representative of the energetics of reactions containing Glu.[239] Through this approximation, we generated a reaction class of 81 model dipeptides through which pH-dependent peptide bond hydrolysis could be analyzed. Furthermore, the impact of sidechain functionalities on reaction energetics was taken into consideration by modeling all but two classifications –thiol-containing and aromatic.[239] The exclusion of these two functional group classifications was a result of their occurrence within the hinge region, where Flores et al. determined that aromatic and thiol-containing amino acids were ranked among the residues least likely to be present.[163]

Evaluation of residue position on reaction energetics was evaluated through the development of our reaction class. The basic mechanism of peptide bond fragmentation were modeled using 81 dipeptides; where each dipeptide contained residues that were cited by Flores et al. as the most frequent to occur within the hinge region in either the N or C-terminus position of the dipeptide.[163] By modeling the interactions of the residues with one another, as well as the impact of their position in either the N or C-terminus, we achieved a robust understanding of how the interactions between different functionalities impacted the energetics of peptide bond fragmentation. Furthermore, as depicted in Table 6.3.1, the ionization states of each residue modeled within our reaction class was analyzed as a function of the modeled mechanism. Of the nine amino acids used to generate our reaction class, only Asp and Lysine (Lys) contain ionizable sidechains where Asp is an acidic residue and Lys is a basic residue.[239] Furthermore, while the titration point of Asp does not fall within the pH range associated with the techniques implemented to polish and purify the mAb during manufacturing, the sidechain of Lys does shift from cationic to neutral at a pH of 10.4.[67, 68, 470, 487, 488]

Table 6.3.1. Ionization states of amino acid sidechains under the basic conditions associated with downstream manufacturing.[67, 68, 470, 487, 488]

Amino Acid	Basic pH			
	Ionizable sidechain $8.0 < pH < 10.4$	Charge	Ionizable sidechain $10.4 < pH < 11.5$	Charge
Ala	No	0	No	0
Asp	Yes	-1	No	0
Gly	No	0	No	0
Leu	No	0	No	0
Lys	Yes	1	No	0
Pro	No	0	No	0
Ser	No	0	No	0
Thr	No	0	No	0
Val	No	0	No	0

Therefore, to fully correlate the impact of the pH range associated with the alkaline wash stage of the downstream manufacturing of mAbs, the titratable nature of the Lys sidechain must be taken into consideration. For this purpose, the reaction classes were expanded to 90 reactions within the dataset, where the nine additional reactions were representative of how the neutral Lys sidechain impacts the reaction energetics of peptide bond fragmentation under basic conditions.

Furthermore, we have accounted for the electrostatic impact upon reaction energetics by modeling the reactions within the dataset under four dielectrics (ε = 4, 20, 40, and 80).[210-220, 489] As we noted in our previous studies, a robust analysis of the factors influencing protein degradation includes modeling the reactions under the dielectrics associated with the different sites of the protein.[82, 141, 147] Specifically, the smallest dielectric used to model non-enzymatic peptide bond fragmentation under alkaline conditions (ε = 4) is representative of the hydrophobic core, as well interactions formed by residues with hydrophobic sidechains.[82, 141, 147, 212, 213, 221, 222, 489] Similarly, modeled reactions under the dielectrics of ε = 20 and ε = 40 are representative of the electrostatic interactions associated with charged and polar residues, as well as a representation of the reaction behavior on the protein surface.[82, 141, 147, 212, 213, 216, 223] Lastly, the dielectric associated with bulk water (ε = 80) provides insights towards the behavior of polypeptide chains floating freely in solution, thus giving us a full analysis into how the behavior of the electrostatic interactions observed at each of the modeled dielectrics influences reaction energetics.[82, 141, 147, 213]

6.3.ii. Quantum Chemical Calculations and Statistical Thermodynamics

Expansion of our previously developed non-enzymatic peptide bond fragmentation database was completed through the use of a self-consistent methodology.[82, 141, 147, 251] Quantum chemical calculations were performed with *Gaussian 16*.[82, 141, 147, 250, 251] All electronic energies for the protein dipeptides, free amino acids, intermediates and transition states were calculated using the B3LYP method.[82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized in the singlet state. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the modeled reactions.[82, 141, 147, 188-190, 246, 247, 253] To account for anharmonicity in the normal vibrational modes, as suggested by Scott and Random, the harmonic frequencies and zero-point energy (ZPE) were scaled by factors of 0.96 and 0.98 respectively.[82, 141, 147, 254] Following the work of previous reaction modeling completed by our group, partition functions based on the harmonic oscillator and rigid rotor approximations were used to calculate thermodynamic and kinetic properties as a function of temperature.[82, 141, 147, 255, 256] This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 257-263]

6.3.iii. Transition State Theory and Kinetic Parameters

Transition states associated with the basic mechanism of peptide bond fragmentation were found using the Berny method.[264-266] The imaginary frequency of each transition state was animated, and intrinsic reaction-coordinate calculations were performed to confirm that the normal vibrational mode pertained to the reaction coordinate of interest. Conventional transition-state theory (TST) was then used to calculate rate coefficients at 1 atm, according to the macroscopic formulation in Equation (6.3.1).[267]

$$k^{TST}(T) = n_d \acute{k} = n_d \Lambda \exp\frac{\Delta S^{\neq}}{R} \exp\left(\frac{-\Delta H^{\neq}}{RT}\right) \tag{6.31}$$

Where Λ is defined in Equation (6.3.2):

$$\Lambda = \kappa(T)\frac{k_B T (V_m)^{-\Delta n}}{h} \tag{6.3.2}$$

Here, $\acute{k}$ is the single-event rate coefficient, κ(T) is the Wigner tunneling correction at temperature T, k_B is Boltzmann's constant, h is Planck's constant, V_m is the molar volume, R is the ideal gas constant, $\Delta S^{\neq}$ is the entropy of activation, $\Delta H^{\neq}$ is the enthalpy of activation, Δn is the change in the number of moles going from the reactant to the transition state (i.e., -1 for bimolecular reactions such as the intermediate formation step in the forward direction, and the proton transfer step in the reverse direction; and 0 for unimolecular reactions such as the intermediate formation step in the reverse direction and the proton transfer step in the forward direction), and n_d is the reaction path degeneracy, or number of single events. The reaction path degeneracy follows Equation (6.3.3). $\Delta H^{\neq}$ and $\Delta S^{\neq}$ are calculated using standard formulae, and there were no internal rotation corrections needed within our system.[82, 141, 147, 196, 267]

The reaction path degeneracy, n_d, definition is based upon the symmetry of the reactant(s) and the transition state.[82, 141, 147, 196] Symmetry was determined by the number of chiral centers present in each molecule.[82, 141, 147, 196, 239, 256, 263] To achieve an unbiased representation of each modeled reaction, the maximum symmetry numbers were applied to each reaction.[82, 141, 147, 239] Analysis of the reaction path degeneracies used in our model, noted that the inclusion of chiral centers did not impact the forward reaction path degeneracy developed through the exclusion of chiral centers. Therefore, our model may be applied to both natural and synthetic proteins in the forward direction of each computed reaction. The reaction path degeneracies for this study follow Equation (6.3.3):

$$n_{\mathrm{d}} = \frac{n_{\mathrm{chiral}}^{\neq}}{\prod_{\mathrm{j}} n_{\mathrm{chiral}}^{\mathrm{reactant,j}}} \times \frac{\prod_{\mathrm{j}} \sigma^{\mathrm{reactant,j}}}{\sigma^{\neq}} \tag{6.3.3}$$

Where n_{chiral} is the number of chiral isomers for the transition state and reactant(s); σ is a product of the external rotation symmetry numbers for the transition state and reactant(s). Internal symmetry numbers for non-cancelling internal rotations treated as hindered or free rotations were included in the hindered and free rotor partition functions, respectively.

The single-event parameters of the Arrhenius relationship, log(A) and E_A were obtained by fitting ln(k) versus T^{-1} over the temperature range 298 – 1500 K. This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 255, 257-263] Through linear regression analysis, we determined Arrhenius behavior was obeyed well for the majority of the reactions modeled within this study, where non-Arrhenius behavior was observed by the reactions which contained negative apparent activation barriers. The solvation impact of liquid phase kinetics is beyond the scope of this work but will be addressed in future studies following the work of Jalan et al, which computes liquid-phase kinetics in relation to both the solute and the solvent.[268]

6.3.iv. Generalized Model Validation

To maintain the self-consistent methodology associated with non-enzymatic peptide bond fragmentation, we applied the computational methodology and chemical model of our peptide fragmentation study under ambient conditions to this study.[82, 141, 147] As we previously noted, there is experimental agreement that non-enzymatic peptide bond fragmentation under basic conditions is a stepwise mechanism.[142] Therefore, to validate our methodology and chemical model, we compared the forward rate coefficient of the proposed rate determining step of the mechanism against experimental values provided by Sun and by Smith and Hansen (Table 6.3.1).[142, 144]

Evaluation of the associated error between our computed rate coefficients and the experimental values was completed via linear interpolation to reduce variability within the analysis.[142, 144] Additionally, we found that the pseudo-first order approximation we had previously applied to our fragmentation study under ambient conditions was not appropriate for the basic mechanism. Therefore, to convert the rate coefficients associated with the proposed rate-determining step of Gly-Gly and Gly-Val peptide bond fragmentation into first-order rate coefficients, we multiplied the concentration of H^+ [= 10^{-9} M (pH = 9), 10^{-10} M (pH = 10)] with the rate coefficients derived from our Arrhenius parameterization at the modeled dielectric associated with pure water. Through this we determined that the proposed rate-determining step of Glycyl-Valine (Gly-Val) and Gly-Gly peptide bond fragmentation could be computed within two orders of magnitude of the values provided experimentally.[142, 144] Therefore, we determined that both our chemical model and computational methodology would accurately reflection the reaction energetics of peptide bond fragmentation under basic conditions.

Table 6.3.2. Comparative analysis of the computed Arrhenius parameters and forward rate coefficients of Gly-Gly and Gly-Val peptide bond fragmentation under basic conditions against the experimental rate coefficients.[142, 144] The rate coefficients of the proposed rate-determining step was computed via interpolation of the rate coefficients derived from Arrhenius parameterization, so that temperature variability could be removed from the analysis. Furthermore, the comparative rate coefficients of Gly-Gly and Gly-Val peptide bond fragmentation were computed using a second order approximation was applied to the basic mechanisms by multiplying the concentration of H^+ at pH = 9.0 and pH = 10.0 to the Arrhenius-predicted rate coefficients to convert the rate coefficient into the first order.

Study	**Dipeptide Reaction**	**pH**	**Temperature (K)**	**k (s^{-1})**	**Computed k (s^{-1})**	**Δk^F (Literature – Computed)**
Sun et al. (2019)	Gly-Gly	10.0	368	8.4×10^{-7}	2.73×10^{-8}	3.08×10^{1}
Smith and Hansen (1998)	Gly-Val	9.0	310	2.33×10^{-10}	6.76×10^{-9}	3.44×10^{-2}

6.3.v. Generalized Reaction Models and Applications towards Hinge Region Fragmentation

Similar to our work completed under ambient conditions, we sought to determine what correlations could be applied to non-enzymatic peptide bond hydrolysis under basic conditions. The first correlation applied to our models was the Evans-Polanyi correlation, a linear correlation between the reaction activation barrier (E_A) and its associated reaction enthalpy.[141, 224] Through this, the activation barrier of dipeptide hydrolysis under the fragmentation mechanism of interest can be predicted by using its associated enthalpy of reaction (Equation 6.3.4).[141, 224] Furthermore, we can examine the impact of each modeled dielectric on reaction energetics by generating Evans-Polanyi correlations as a function of the modeled dielectric, as well as for the overall dataset.

$$E_A = E_0 + \alpha \Delta H_{RXN} \tag{6.3.4}$$

To maintain the self-consistency of our methodology, as well as determine the robustness and predictiveness of each generated correlation, statistical sensitivity analysis was applied to all statistically significant Evans-Polanyi correlations through a script developed in R by our group.[141, 274] As we previously mentioned, bootstrapping is a statistical sensitivity method that evaluates a model through the random removal of points within the data set.[141, 225-227, 230-232, 348] After the random removal of 20 percent of the data points, a new correlation is generated to predict the replacement values of the removed observations.[141, 225-227, 230-232, 348] Through this approach, the linear strength of the replacement trend, as well as the prediction accuracy of the replacement data points can be evaluated. The use of 5 iterations of random resampling evaluates the model robustness to ensure that the linear strength and prediction accuracy is not a byproduct of specific data points being removed and replaced through the bootstrap method.

Furthermore, principle component analysis (PCA) was implemented to ensure that any clustering observed in the initial Evans-Correlations did not bias the generated models. Ringnér describes PCA as a "mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the dataset."[236] Specifically, by combining variables to form new linear variable, known as principal components.[236] The first principal component (PC 1) represents the direction of the largest source of variation, while the second principal component (PC 2) describes the largest source of variation orthogonal to PC 1.[236, 237, 490] Through this analysis, the variables associated within the proposed rate-determining reaction of the basic mechanism was analyzed to identify if additional parameters were necessary to generate a statistically significant and robust Evans-Polanyi correlation. PCA was conducted through the use of an R script that was adapted from PCA tutorial code on Github.[491] The inclusion of additional parameters within the Evans-Polanyi correlation was completed through the use of regression with a binary descriptor ($\beta = 0$ or $\beta = 1$). Acting as a dummy variable, the binary descriptor acts as an additional source of input within the correlation.[492-494] Similar to the PCA code, the regression of the Evans-Polanyi correlation with a binary descriptor was completed through the use of an R script that was adapted from the Statology tutorial.[495]

6.4. Results and Discussion

6.4.i. Structures

As we previously specified, a subset of 90 dipeptides was used to model the basic reaction mechanisms of non-enzymatic peptide bond hydrolysis in accordance with the 10 most frequently occurring amino acid residues present within the mAb hinge region, as well as the titratable nature of the Lys sidechain.[163, 488, 496] The impact of the functional groups on reaction energetics as a function of the pH range that was modeled was accounted for via its position at the N and C-terminus. Furthermore, the impact of the modeled dielectric on reaction energetics was observed via the conformational behavior of each dipeptide.

The mechanism of non-enzymatic peptide bond fragmentation under basic conditions was modeled as a stepwise process, where the formation of the intermediate was the proposed rate-determining step. The dipeptide and OH^- were treated as infinitely separated reactants, where OH^- decomposed in the transition state structure to attack the peptide bond carbonyl of the dipeptide, and then formed a 5-membered intermediate. Once the intermediate has been formed, the neutral hydroxyl group donated its proton to the peptide bond nitrogen, cleaving the peptide bond and forming an anionic N-terminus residue and a net neutral C-terminus residue, where the final products exist separately from each other in the solution. A representative ball and stick schematic of the stepwise mechanism is present in Figure 6.4.1.

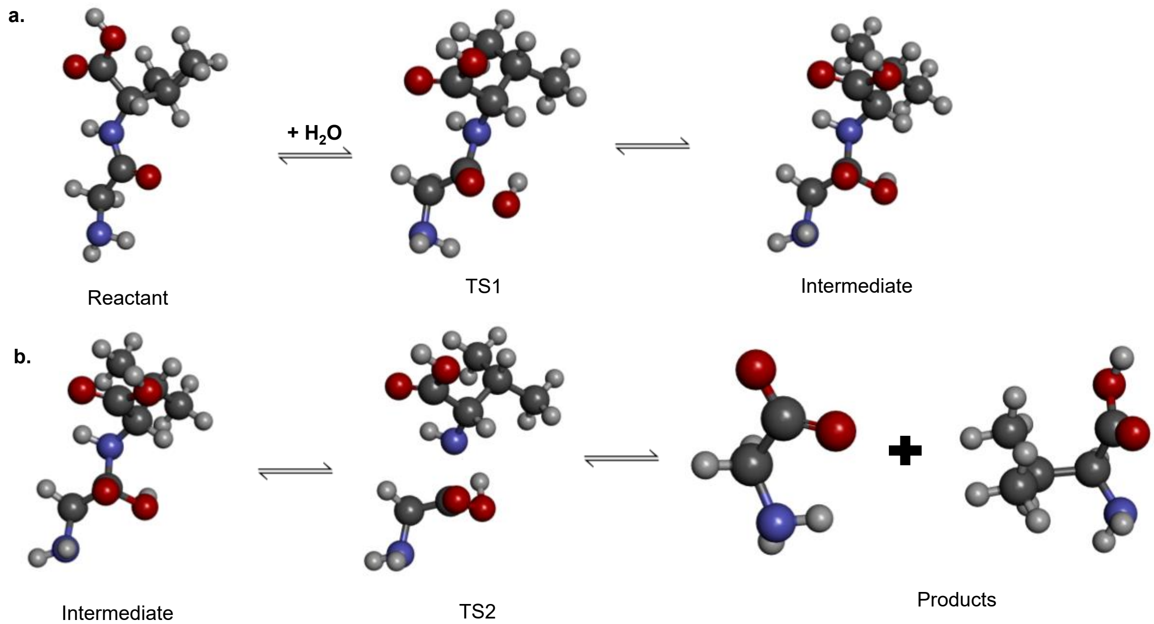

Figure 6.4.1. Ball and stick representation of non-enzymatic peptide bond fragmentation under basic conditions. (a). Intermediate formation step, (b). Proton transfer step.

6.4.ii. Thermochemistry and Arrhenius Parameterization

The overall reaction was found to be net exothermic and spontaneous in nature at all modeled dipeptides, where the proposed rate-determining step was found to be endothermic and endergonic at ε = 20, 40, and 80. At the modeled dielectric associated with hydrophobic residues, the majority of the reactions within the dataset were found to be more exothermic than at the other three dielectrics; where the reactions that were computed to be more endothermic contained anionic Asp.

6.4.ii.a. Non-enzymatic peptide bond fragmentation under basic pH (Intermediate formation reaction)

Analysis of the proposed rate-determining step of the basic mechanism indicated that the reaction became more energetically favorable as the modeled dielectric decreased (Table 6.4.1). At the modeled dielectric associated with hydrophobic residues, the average ΔH_{RXN} and G_{RXN} were found to be the most exothermic and spontaneous compared to the other modeled dielectrics. The computed Arrhenius parameters were similar at the modeled dielectrics associated with charged and polar residues, as well as pure water, but significantly lower at the modeled dielectric associated with hydrophobic residues. At the modeled dielectric associated with polar residues, $E_0 = 10.7 \pm 3.30$ kcal.mol^{-1}, $E_A = 13.4 \pm 3.33$ kcal.mol^{-1}, and log(Ã) = 8.89 ± 0.44 L.mol^{-1}.s^{-1} in the forward direction, while $E_0 = 1.99 \pm 2.54$ kcal.mol^{-1}, $E_A = 2.96 \pm 2.68$ kcal.mol^{-1}, and log(Ã) = 13.8 ± 0.56 s^{-1}, respectively. At $\varepsilon = 4$, the value of the activation barrier is about 5.50 kcal.mol^{-1} lower, suggesting the reaction is most energetically favorable under hydrophobic conditions.

Table 6.4.1. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for the intermediate formation step of the non-enzymatic peptide bond fragmentation reaction under basic conditions at 298 K and 1 atm. Ã has units of L.mol^{-1}.s^{-1} and s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of kcal.mol^{-1}.

ε	Forward Reaction			ΔH_{RXN}	ΔG_{RXN}	Reverse Reaction		
	E_0	log(A)	E_A			E_0	log(A)	E_A
4	4.54	8.95	7.28	-0.25	11.4	2.62	14.0	3.71
20	10.1	8.86	12.8	6.10	17.6	1.95	13.8	2.92
40	10.7	8.89	13.4	6.66	18.2	1.99	13.8	2.96
80	11.0	8.91	13.7	6.95	18.5	2.00	13.8	2.70

The lower energetics exhibited at $\varepsilon = 4$ can be partially attributed to the presence of negative apparent activation energies within the forward reaction. The presence of the negative E_A suggests that the presence of 'barrierless' reactions, where the transition state is present within the potential energy well of the reactant adduct.[497] Of the 90 dipeptides analyzed, 13 were computed with a negative forward E_A. Furthermore, each of the 13 dipeptides contained a positively-charged Lys residue in either the N or C-terminus position. Through the generation of the adduct structures, we determined that the negative E_A was relative, as we had hypothesized. Additionally, we identified a position-dependent hydrogen transfer present in 12 of the 13 adduct structures. When Lys was observed in the N-terminus position, the amide bond nitrogen acted as the hydrogen donor, while the amino group donated the hydrogen atom when Lys was present in the C-terminus position. Of the 13 modeled adducts, only Prolyl-Lysine (Pro-Lys) was not found to follow the observed behavior. Instead, the amide bond nitrogen was found to be the hydrogen donor, which is likely a consequence of the steric hindrance of the cyclic Pro backbone.

Examination of prior literature indicated the propensity of the Lys sidechain towards deprotonation under a hydrophobic environment.[498] A study conducted by Takayama et al., under a mildly basic environment, analyzed the behavior of Lys within the hydrophobic core of the protein.[498] Specifically, they determined that the deprotonation of the buried sidechain to donate its hydrogen atom to bulk water was a near instantaneous process.[498] Furthermore, it was evident of that penetration of water within the hydrophobic core would result in a hydrogen exchange between the buried Lys sidechain and the solvent.[498] Similarly, prior studies have also noted the propensity of the amide bond nitrogen to participate in hydrogen exchanges is catalyzed by the presence of OH^- atoms, as well as the neighboring effect of residues close to the amide group.[499-501] In the study conducted by Molday et al., Lys was among the residues noted to promote OH^- catalyzed hydrogen exchange under hydrophobic conditions.[501] Therefore, it is likely that the negative apparent activation energies observed for the observed reactions within our dataset are indicative of the dipeptides attempting hydrogen exchange within their environment prior to the fragmentation of the peptide bond, and would not be observed in larger structures where the cationic Lys sidechain would be able to form salt bridges to stabilize the structure.

While reactions containing Lys were found to have negative activation energies under the most hydrophobic modeled dielectric, the activation energies of the reactions containing the anionic Asp were found to increase as the modeled dielectric decreased. Prior studies have indicated that the presence of ionized residues within the hydrophobic core of the protein are destabilizing.[502] Furthermore, as Asp is an anionic residue, the repulsive nature of the like-like charges within the hydrophobic core would further destabilize the reaction.[503] Therefore, it is likely that these two factors increased the thermodynamic cost associated with the proposed rate-determining step at the modeled dielectric.

6.4.ii.b. Proton Transfer Step

Evaluation of the average thermochemical and Arrhenius parameters (Table 6.4.2.) associated with the proton transfer step of the basic mechanism indicated that the reaction was both exothermic and spontaneous in nature. Furthermore, as the modeled dielectric decreased, the energetic favorability of the reaction increased. At $\varepsilon = 4$, $E_0 = 12.9 \pm 2.80$ kcal.mol^{-1}, $E_A = 13.7 \pm 2.91$ kcal.mol^{-1}, and $\log(\tilde{A}) = 14.0 \pm 0.63$ s^{-1} in the forward direction, while $E_0 = 36.2 \pm 7.74$ kcal.mol^{-1}, $E_A = 38.7 \pm 7.77$ kcal.mol^{-1}, and $\log(\tilde{A}) = 6.88 \pm 0.57$ L.mol^{-1}.s^{-1}, respectively. A study conducted by Isom et al. discusses the paradox that is associated with the presence of charged residues within the hydrophobic core of proteins.[502] Specifically, while the presence of ionized amino acids within the core is considered to be energetically unfavorable, it has also been noted that these residues promote proton transfer.[358, 502] The rationale behind the propensity of ionized residues towards the increased favorability of the hydrogen transfer reaction is due to the prior finding that charged and polar residues promote the formation of a hydrogen bonding network.[358]

Through this network, proton transfer reactions are more energetically favorable because the capacity of ionized and polar residues to act as hydrogen acceptors and donators stabilizes the structures involved in the reaction.[358, 504] Therefore, the presence of anionic Asp at the modeled dielectric associated with hydrophobic residues will be energetically favorable due to the salt bridge formed between the proton and the negatively-charged Asp residue.[358]

Table 6.4.2. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for the proton transfer step of the non-enzymatic peptide bond fragmentation reaction under basic conditions at 298 K and 1 atm. Ã has units of s^{-1} and $L.mol^{-1}.s^{-1}$ for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of $kcal.mol^{-1}$.

ε	Forward Reaction			ΔH_{RXN}	ΔG_{RXN}	Reverse Reaction		
	E_0	log(A)	E_A			E_0	log(A)	E_A
4	12.9	14.0	13.7	-23.1	-37.2	36.2	6.88	38.7
20	13.3	14.0	14.1	-25.9	-39.9	39.4	6.88	41.9
40	13.5	13.9	14.3	-26.0	-39.9	39.6	6.91	42.1
80	13.6	14.0	14.4	-26.0	-39.9	39.7	6.94	42.3

However, we determined that, at the modeled dielectric associated with hydrophobic residues, our hypothesis regarding the rate-determining step was proven to be incorrect. Our proposal that the formation of the intermediate would be rate-determining was supported by prior studies on peptide bond fragmentation, and was proven to be an accurate hypothesis for the majority of the reactions at ε = 20, 40, and 80.[141, 505, 506] However, at ε = 4, 50% of the reactions found the proton transfer step to be rate-determining. Of the reactions that found the proton transfer reaction to be rate-determining, 13 dipeptides were found to have negative apparent activation energies during the formation of the intermediate and therefore had negative temperature dependence.[497, 507-509] Therefore, these reactions were not considered in our evaluation of the rate-determining step at ε = 4.

The 34 dipeptides that found the intermediate formation step of the basic mechanism to be rate-determining at $\varepsilon = 4$ likely exhibited this behavior due to electrostatic and steric factors. When Asp and Leu are in the N-terminus position, the formation of the intermediate is rate-determining. Similarly, eight out of the nine reactions that contained Asp and Pro in the C-terminus position found the proton transfer reaction to occur at a quicker rate relative to the intermediate formation reaction. As we previously specified, the presence of ionized residues within the hydrophobic core of the protein was found to be both energetically unfavorable and a facilitator for reactions within the hydrophobic core, such as proton transfer reactions.[503] Therefore, reactions containing Asp would be more likely to find the proton transfer reaction energetically favorable relative to the intermediate formation reaction.

6.4.iii. Generalizing non-enzymatic peptide bond fragmentation under basic conditions via the Evans-Polanyi correlation

Similar to the non-enzymatic peptide bond fragmentation under storage, we sought to determine if the Evans-Polanyi correlation could be applied to the proposed rate-determining steps of the basic mechanism. As we previously specified, the dipeptide is the smallest system that can be used to analyze peptide bond fragmentation within the hinge region, and there are 400 possible dipeptides that can be formed. Within the modeled reaction mechanism, generalization of the proposed rate-determining step would provide the most insight towards reaction energetics. Therefore, we analyzed the intermediate formation reaction as an overall correlation, as well as a function of each of the four modeled dielectrics, to determine if a statistically significant correlation could be achieved.

6.4.iii.a. Initial Correlations

Evaluation of the correlations associated with the proposed rate-determining step of the basic mechanism indicated that the activation energy could be generalized through a statistically significant linear trend (Table 6.4.3, Figure 6.4.2). The overall correlation and the correlation generated at the dielectric associated with hydrophobic residues were found to exhibit strong linearity, while the correlations at the dielectrics associated with charged residues, polar residues, and bulk water were found to exhibit moderate linearity.[371, 372] As we previously noted, negative apparent activation energies were observed for 13 dipeptide reactions within the ε = 4 dataset. However, those reactions were also found to be exothermic in nature. As a result, the Evans-Polanyi principle was observed where the R^2 of the correlation was found to be strong, while the correlations at the other three modeled dielectrics were found to be moderate in comparison.[371, 372] Similarly, the R^2 value associated with the full dataset is representative of the energetics exhibited at each modeled dielectric, where it exhibits stronger linearity than the correlations at ε = 20, 40, and 80, but a weaker linearity than the trend observed at ε = 4.

Table 6.4.3. Statistical parameters associated with the Evans-Polanyi correlations of the proposed rate-determining step of the basic mechanism of non-enzymatic peptide bond fragmentation as a function of the modeled dielectric [Initial Correlation].

Correlation	E_0	α	R^2	p-value	ρ	p-value
All	7.64	0.86	0.83	< 0.001	0.91	< 0.001
ε = 4	7.51	0.92	0.91	< 0.001	0.96	< 0.001
ε = 20	8.65	0.68	0.56	< 0.001	0.75	< 0.001
ε = 40	9.11	0.65	0.49	< 0.001	0.70	< 0.001
ε = 80	9.39	0.62	0.45	< 0.001	0.67	< 0.001

Figure 6.4.2 depicts the overall Evans-Polanyi correlation, as well as the modeled dielectric-specific correlations, generated for the proposed rate-determining step of the basic mechanism. As we previously noted, a statistically significant Evans-Polanyi correlation was achieved for the overall correlation, as well as at each modeled dielectric. Analysis of the five correlations the majority of data points were centered around the linear trendline for the ε = 20, 40, and 80 correlations, while the overall and ε = 4 correlations were more linearly distributed.

However, the reduced spread of the data points used to generate the $\varepsilon = 20$, 40, and 80 correlations, relative to the overall and $\varepsilon = 4$ correlations, negatively impacted their linear strength. While the reduced spread observed in each correlation appeared to be indicative of the low standard deviation present in each reaction class, we sought to evaluate its impact on the linear strength and predictiveness of our model. Specifically to see if the linearity and predictiveness of the previously noted correlations could be improved via the inclusion of additional parameters. For this purpose, we included supervised learning methods into our study.

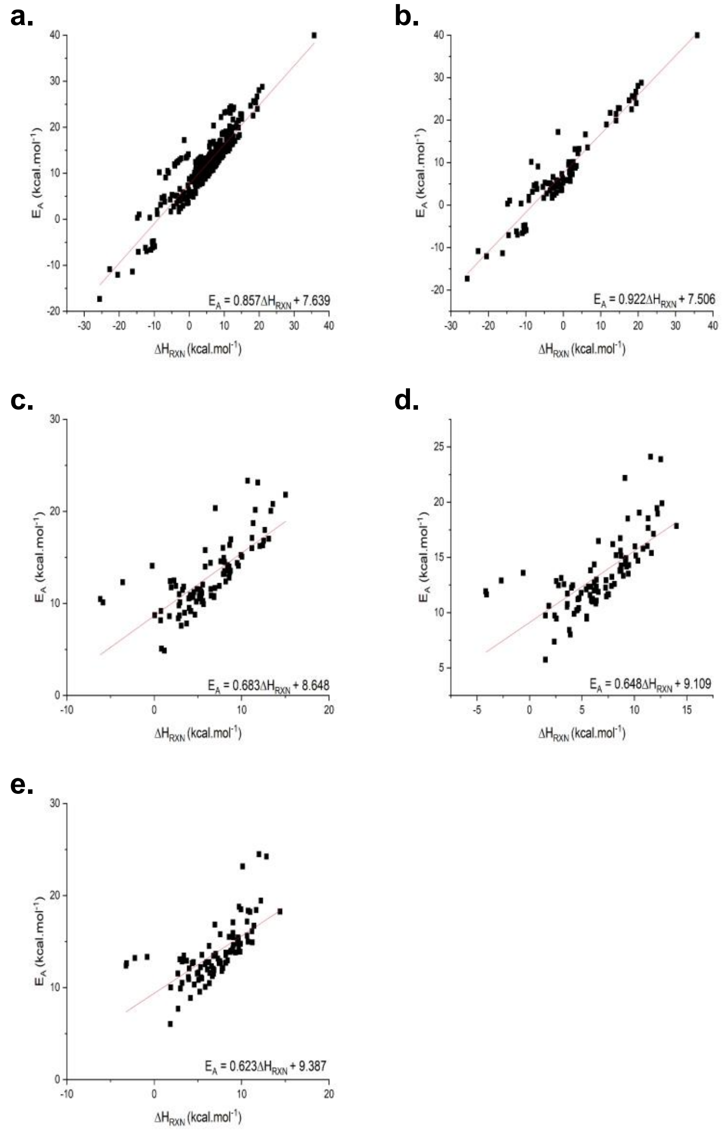

Figure 6.4.2. Evans-Polanyi correlations as a function of modeled dielectric [Initial correlation]. (a). Full dataset, (b). ε = 4, (c). ε = 20, (d). ε = 40, (e). ε = 80.

6.4.iii.b. Supervised Learning through Principal Component Analysis (PCA) and its Impact on the Evans-Polanyi Correlation

As we previously specified, five statistically significant Evans-Polanyi correlations were developed for the proposed rate-determining step of the basic mechanism of peptide bond fragmentation. However, the presence of clustering within the developed models suggested the possibility of model bias within each correlation. For this purpose, we applied Principal

Component Analysis (PCA) to identify if the prediction accuracy of the developed Evans-Polanyi correlations could be increased through the inclusion of additional descriptors. Using the full dataset, five parameters were analyzed and their respective contributions to variability were computed (Table 6.4.4). The most significant variation observed within the dataset is present in the PC 1 direction, while the direction orthogonal to PC 1 (PC 2) accounts for observed variability, which does not contribute as significantly.[233, 234] Of the five parameters, the two parameters associated with the derivation of the Evans-Polanyi correlation, ΔH_{RXN} and E_A, contributed the most significantly to the variation observed along the PC 1 direction. The contribution of ΔH_{RXN} and E_A to the variance observed in the PC 1 direction was computed to be slightly more than the next contributing parameter – the associated net charge of the reaction. Additionally, the contributions of the propensity of the residues to be found in the hinge region were found to contribute significantly in the PC2 direction, while the reaction net charge, ΔH_{RXN} and E_A were found to be negligible in comparison. Evaluation of the propensity of the N and C terminus residues to occur within the hinge region was not found to be negligible relative to the variability associated with the reaction net charge, ΔH_{RXN} and E_A. Therefore, evaluation of the contributions of the two parameters to reaction variability would not be appropriate for inclusion within our developed Evans-Polanyi models. As a result, through PCA we determined that, while our initial Evans-Polanyi correlations accounted for a significant amount of the variability within our dataset, the inclusion of a third parameter would likely improve the correlation strength and predictiveness of our generated models.

Table 6.4.4. Parameter contributions to variability within the intermediate formation reaction of non-enzymatic peptide bond fragmentation under basic conditions [Full dataset].

PC 1 Direction		PC 2 Direction	
Parameter	**Contribution to Variability**	**Parameter**	**Contribution to Variability**
ΔH_{RXN}	33.8	ΔH_{RXN}	0.72
E_A	33.3	E_A	0.06
Net Charge	31.3	Net Charge	0.22
N-terminus residue rank	0.14	N-terminus residue rank	47.8
C-terminus residue rank	1.46	C-terminus residue rank	51.2

Evaluation of the net charges within the dataset indicated that the datasets could be split into three categories: cationic, neutral, and anionic. For the basic mechanism, the presence of net neutral reactions is indicative of the pH range where the ionization state of the Lys sidechain counteracts the electrophilic attack of attack of OH^- upon the peptide bond. Additionally, the cationic net charge within the reaction class is representative of Lys-Lys peptide bond fragmentation, where the sidechains of both Lys residues are cationic (pH < 10.4).[487, 488] For this purpose, we hypothesized that the clustering observed in our initial correlations could be accounted for via the development of charge-driven correlations. To test this hypothesis, we generated three Evans-Polanyi correlations – the initial correlation, a PCA-driven correlation that includes the associated net charge of the reaction as a binary descriptor, and a PCA-driven correlation that splits the associated net charge of the reaction into two correlations (anionic and charge ≥ 0), and then evaluates the statistically significant correlation.

The impact of a charge-driven correlation was analyzed by splitting the dataset into anionic and cationic or neutral correlations, where only the anionic trend was found to be statistically significant.[371, 372] Through this correlation, we developed an alternative approach towards addressing the variability indicated by PCA. Evaluation of the linear strength of each anionic correlation indicated that the linear strength of the correlations observed at $\varepsilon = 20$, 40, and 80 increased by 0.05, while the overall correlation and the $\varepsilon = 4$ correlations decreased by 0.03. The observed perturbations suggest that, at $\varepsilon = 20$, 40, and 80, the reactions containing cationic Lys did not exhibit Evans-Polanyi-like behavior, and therefore decreased the linearity of the modeled correlations. Similarly, for the $\varepsilon = 4$ correlations where the cationic Lys reactions exhibited strong Evans-Polanyi-like behavior, the linearity of the generated correlation would be very strong relative to other generated correlations. Clustering was observed at the $\varepsilon = 4$ correlation, where two separate clusters of data points were observed in the correlation. Evaluation of the anionic dataset at the modeled dielectric associated with hydrophobic residues indicated that the one cluster was representative of the reactions that did not include Asp ($R^2 = 0.37$), while the other cluster only contained reactions that included Asp ($R^2 = 0.97$). As we previously indicated, the presence of anionic Asp in the hydrophobic core was noted to be destabilizing and thermodynamically unfavorable.[502, 503] Therefore, it is likely that the presence of the energetically unfavorable reactions drove the increased linearity of the $\varepsilon = 4$ correlation.

As we previously specified, the net charge of the reactions within our dataset drove variability in the direction orthogonal to the variability associated with the ΔH_{RXN} and E_A of our dataset. Furthermore, the reaction class can be split into anionic dipeptides and dipeptides that are either net neutral or cationic. For this purpose, we associated the dummy variable within our correlation with the net charge of the reaction; where 1 represents anionic reactions and 0 represents the neutral and cationic reactions within the dataset. Initial analysis indicates that the binary correlation would not improve model predictiveness because the inclusion of a dummy variable within the Evans-Polanyi correlation did not significantly impact its linear strength.

6.4.iii.c. Model Sensitivity and Robustness: Unsupervised versus Supervised Correlations

As we noted in our previous study on peptide bond fragmentation under ambient conditions, validation of our generalized Evans-Polanyi correlations was achieved through statistical sensitivity analysis.[141] By applying five iterations of statistical bootstrapping to each of the previously developed Evans-Polanyi correlations, we evaluated both the linear strength and robustness of the generated trends.[141] Evaluation of statistical sensitivity on the initial Evans-Polanyi correlation indicated that the model linearity and robustness of the bootstrapped correlations exhibited similar trends to the original correlation (Table 6.4.5, Figure 6.4.3).

Table 6.4.5. Sample statistical parameters associated with the bootstrapped with replacement Evans-Polanyi correlations as a function of modeled dielectric [Initial Correlation].

Correlation	E_0	α	R^2	p-value	ρ	p-value
All	7.45	0.87	0.87	< 0.001	0.93	< 0.001
ε = 4	7.55	0.92	0.92	< 0.001	0.96	< 0.001
ε = 20	7.50	0.83	0.74	< 0.001	0.86	< 0.001
ε = 40	8.96	0.66	0.54	< 0.001	0.74	< 0.001
ε = 80	10.1	0.52	0.43	< 0.001	0.66	< 0.001

Specifically, the five bootstrapped correlations observed by the overall and ε = 4 datasets were found to exhibit strong linearity and minimal variance ($\sigma_{wrp} < 0.01$), while the ε = 20, 40, and 80 datasets demonstrated weaker linearity and larger variance ($0.06 < \sigma_{wrp} < 0.11$).[371, 372] The trends observed among the bootstrapped iterations of the initial Evans-Polanyi correlations indicated that replacing the values of the activation barriers of the reactions randomly removed from the dataset did little to impact model linearity. Furthermore, the negligible variance observed among the five iterations of bootstrapping indicated that the generated correlations were not reliant on the energetics of specific peptide bond fragmentation reactions.[368] Therefore, we validated the linearity and robustness of the initial Evans-Polanyi correlations regardless of the modeled dataset.

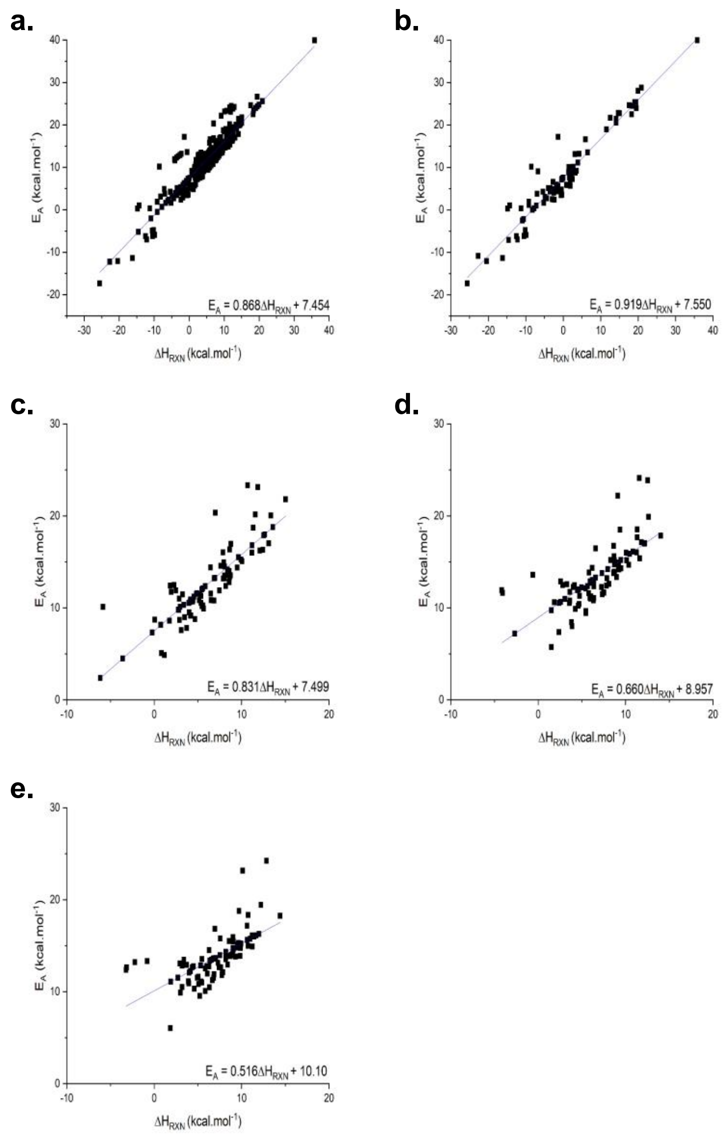

Figure 6.4.3. Sample bootstrapped with replacement Evans-Polanyi correlations as a function of modeled dielectric [Initial correlation]. (a). Full dataset, (b). ε = 4, (c). ε = 20, (d). ε = 40, (e). ε = 80.

The trends observed by the initial Evans-Polanyi correlations were also observed in the PCA-driven correlations, where both correlations were found to slightly increase model linearity. As we previously specified, the associated R^2 of the binary Evans-Polanyi correlations was not found to be significantly different from the initial correlation. Therefore, it is likely that the charge-based parameter is an extraneous metric within our generated trend, and does little to improve the accuracy of the Evans-Polanyi correlation for the proposed rate-determining step of peptide bond fragmentation under basic conditions.

6.4.iii.d. Prediction Accuracy of Generated Correlations: Unsupervised versus Supervised Correlations

The average associated error between the computed activation barriers and the predicted activation barriers was found to fall within the accepted error range ($\Delta E_{RROR} < 4$ kcal.mol^{-1}) for each of the three types of Evans-Polanyi correlations (Figure 6.4.4). Across the five iterations of bootstrapping, the binary correlation was found to be the least accurate of the three Evans-Polanyi correlations regardless of the modeled dataset. As we previously indicated, the inclusion of a binary descriptor for the associated net charge of the reaction within the Evans-Polanyi correlation did not significantly impact the linear strength of the modeled correlations, suggesting that the inclusion of the dummy variable as an additional parameter within the correlation would cause overfitting.[510-512] A study by Hawkins indicated that, while the overfit correlation yields strong linearity, the inclusion of the unnecessary parameter results in reduced model predictiveness.[510] As we previously noted, the inclusion of the charge-based dummy variable as an additional parameter within the binary correlation was not found to improve the linearity of our generated Evans-Polanyi correlations. Therefore, it is likely that the binary correlation is overfit and the most inaccurate of the three types of Evans-Polanyi correlations.[510-512]

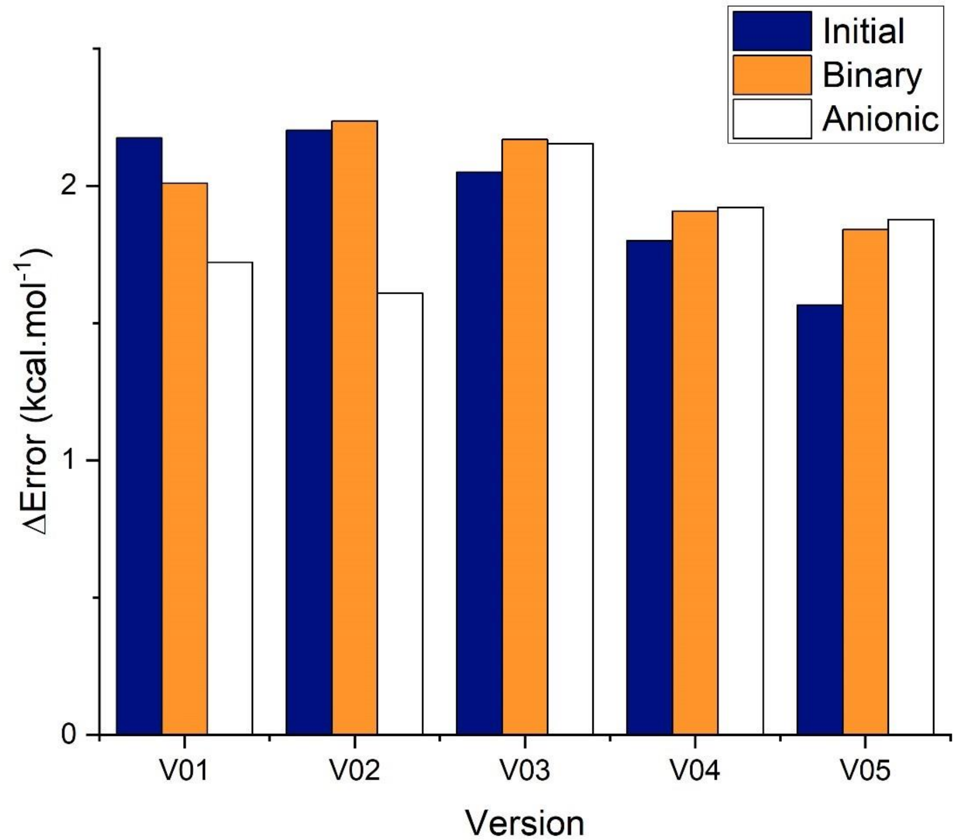

Figure 6.4.4. Bar plot of the average, absolute error associated with the initial, binary, and anionic correlations for the overall dataset.

Evaluation of the associated prediction error of the modeled dielectric-specific correlations indicated that the initial and anionic correlations were found to be least accurate at the modeled dielectric associated with hydrophobic residues, while the binary correlation was least accurate at the modeled dielectric associated with charged residues. As we previously indicated, the electrostatic interactions associated with the different modeled dielectrics impacted the reaction energetics of the modeled reaction. Furthermore, the presence of the cationic Lys reactions within the $\varepsilon = 4$ dataset was found to skew the linearity of the initial correlation. As a result, the predictiveness of the bootstrapped correlation would be less accurate relative to the correlations at the other modeled dielectrics. In the binary correlation, the dummy variable is defined by either a '1' for anionic reactions or a '0' for reactions that have a net charge > -1. Within the dataset, the reaction energetics of the dipeptides containing cationic Lys were found to shift more from $\varepsilon = 40$ to $\varepsilon = 20$ relative to other dipeptides. Furthermore, the binary correlation was most accurate in predicting the average E_A at the dielectric associated with polar residues. Therefore, it is likely that the impact of the electrostatic interactions observed in the $\varepsilon = 20$ dataset is detracted by the inclusion of the charged-based parameter in the correlation.

6.4.iv. Insights towards mAb Hinge Region Stability under Alkaline Conditions

Evaluation of the computed forward rate coefficients indicated that, similar to the generated Evans-Polanyi correlations, the associated net charge of the proposed rate-determining step influenced its propensity towards non-enzymatic peptide bond fragmentation under basic conditions. Specifically, the reactions that contained anionic Asp were found to have small forward rate coefficients relative to the other reactions analyzed within our generated dataset. As we previously specified, the presence of like-like charges within a protein structure were found to be destabilizing due to electrostatic repulsion.[503, 513-516] As a result, the propensity of the specified reactions towards non-enzymatic peptide bond fragmentation under basic conditions would be less energetically favorable relative to the dipeptides comprised of unionized or cationic residues. While Lys was found by Flores et al. to occur less frequently in the hinge region relative to all other amino acids, aside from Val, within our generated dataset, 11 of the reactions containing Lys were listed among the top third of computed forward rate coefficients.[163] Of those 11 reactions, only three dipeptides contained Lys residues with unionized sidechains. Therefore, it is likely that electrostatic attraction was observed between the cationic Lys sidechain and OH^- acted to stabilize the transition state structure, resulting in larger forward rate coefficients.[503, 513, 514, 517, 518]

The frequency of the residues to occur within the hinge region were found to impact the propensity of the reactions towards non-enzymatic peptide bond fragmentation under basic conditions. Specifically, out of the 30 reactions computed to have the largest k^F, 13 contained Gly. As we previously noted, Gly was highlighted by Flores to occur the most frequently within the hinge region.[163] The propensity of Gly towards occurring within the hinge region, as well as its favorability towards non-enzymatic peptide bond fragmentation under basic conditions, is likely due to the flexible nature of the residue. Prior studies have indicated that the susceptibility of the hinge region towards fragmentation was due to its flexible nature.[138, 157, 158, 480, 519-522] Similar to the reactions containing Gly, the non-enzymatic peptide bond fragmentation reactions containing Ala (fourth most occurring) and Ser (second most occurring) were also found to be the most energetically favorable. Of the 20 amino acids, Gly, Ala, and Ser are all classified as small and flexible residues.[164, 239, 365, 523, 524] Therefore, it is highly probable that the flexibility of the residues, similar to the flexibility of the hinge region, contributed to reaction favorability.

Similarly, dipeptides containing N-terminus Val and C-terminus Pro were found to be among the least energetically favorable reactions observed within the generated dataset. A study by Huang and Nau ranked the 20 amino acids based on their flexibility, where Val and Pro were determined to be the third least flexible and least flexible residues, respectively.[365] Therefore, it is likely that the rigid nature of the respective residues influenced their propensity towards non-enzymatic peptide bond fragmentation under basic conditions, and their frequency of occurrence within the hinge region. Additionally, prior studies on polypeptide fragmentation indicated that the “Proline effect”, where C-terminus Pro promotes fragmentation, was least energetically favorable under a basic environment, such as during the alkaline wash stage of downstream manufacturing.[525-529] As a result, the forward rate coefficients associated with Xaa-Pro non-enzymatic peptide bond fragmentation analyzed within our study would not be representative of the "Proline effect".[525-529]

6.5. Conclusions and Future Directions

In summary, a theoretical study on non-enzymatic peptide bond fragmentation under basic conditions was completed to analyze the impact of the degradation mechanism on the stability of the mAb hinge region during the alkaline wash stage of mAb downstream manufacturing. Rate coefficients and Arrhenius parameters for the stepwise reaction mechanism were computed for 720 total reactions (180 reactions per modeled dielectric) using B3LYP and statistical thermodynamics. The forward rate coefficient of the proposed rate-determining step of the reaction mechanism was benchmarked against two experimental studies to validate our computational methodology and proposed chemical model.[142, 144] Furthermore, the impact of the insights derived from our study upon the stability of the mAb hinge region was evaluated through the development of our modeled reaction class, where the residues most likely to be found in the area of interest, as well as the ionization states of the Lys sidechain, were accounted for.[163, 487, 488] Evaluation of the reaction energetics under the four modeled dielectrics indicated that reactions containing Asp become more kinetically unfavorable as the modeled dielectric decreased, while the other reactions modeled within the reaction class were computed to be more kinetically favorable as the modeled dielectric decreased. Furthermore, the Lys sidechain was found to deprotonate at the modeled dielectric associated with hydrophobic residues, resulting in a negative apparent activation barrier.[498]

A statistically significant Evans-Polanyi correlation was achieved for the proposed rate-determining step of non-enzymatic peptide bond fragmentation under basic conditions. Through the application of PCA, three different types of Evans-Polanyi correlations were developed and analyzed through statistical sensitivity methods. Evaluation of the model robustness and predictiveness for the overall correlation, as well as the correlations developed at each modeled dielectric, indicated that the binary Evans-Polanyi correlation was the least accurate of the three types of generalizations. Furthermore, while the anionic and initial correlations yielded similar prediction accuracies, the overall reaction net charge associated with the anionic correlation was found to limit its effectiveness. For this purpose, we concluded that the initial correlation was the most accurate for predicting the activation barriers of the intermediate formation reaction as a function of the enthalpy of reaction of the dipeptide of interest.

Lastly, evaluation of the forward rate coefficients associated with the proposed rate-determining step indicated that charge-charge interactions and residue flexibility contributed to reaction propensity towards non-enzymatic peptide bond fragmentation.[365, 503, 513-516, 523, 524] Furthermore, the propensity of the flexible residues to promote fragmentation under alkaline conditions indicated that the tendency of the hinge region towards fragmentation was a product of the residues most likely to occur within the hinge region, where three out of the four most frequently occurring residues were all ranked to be small and highly flexible.[157, 158, 163, 365, 519-524] Future studies will automate the Evans-Polanyi correlation to improve its predictive accuracy. Additionally, the thermochemical and kinetic parameters computed by our developed database will be used to generate a robust microkinetic model of protein degradation.

Chapter 7: Towards Pharmaceutical Protein Stabilization: Data-Driven Analysis of Non-enzymatic Peptide Bond Fragmentation during the Protein A Chromatography Stage of Downstream Manufacturing

7.1. Background and Motivation

As we previously noted, the downstream manufacturing stage of the pharmaceutical protein pipeline is pH-dependent, where the protein is exposed to both acidic and basic environments during its extraction and purification. Protein A chromatography, which is a popular technique used to separate the protein of interest from its host cell, is completed under acidic conditions.[64-66] A study conducted by Amritkar indicated that binding between the complimentary binding region of the protein and the cell is preferred under neutral conditions, but the shift in environment pH to acidic weakens the observed hydrophobic interactions via the repulsion of like charges.[65] As a result, the protein dissociates from the cell to be extracted from the column.[64-66] While the perturbation in system pH is a necessary step within the manufacturing pipeline, multiple studies have identified the hinge region to be highly susceptible to fragmentation via the pH-dependent fragmentation of peptide bonds.[138, 142, 145, 470, 476-481] Under acidic conditions, fragmentation is observed via the electrophilic attack of H_3O^+ on the peptide bond to cleave the polypeptide chain into two fragments.[142, 144]

As we previously noted, the effectiveness of Protein A chromatography is observed via the electrostatic repulsion observed between two like-charged residues at the binding site.[65] Therefore, it is necessary to take the impact of different residue functionalities on the reaction energetics of non-enzymatic peptide bond fragmentation under acidic conditions into account. Furthermore, the correlation between the frequency of residues to occur within the hinge region, the functionality of their sidechains, and the propensity of the hinge region towards fragmentation must be taken into consideration (Figure 7.1.1).[163]

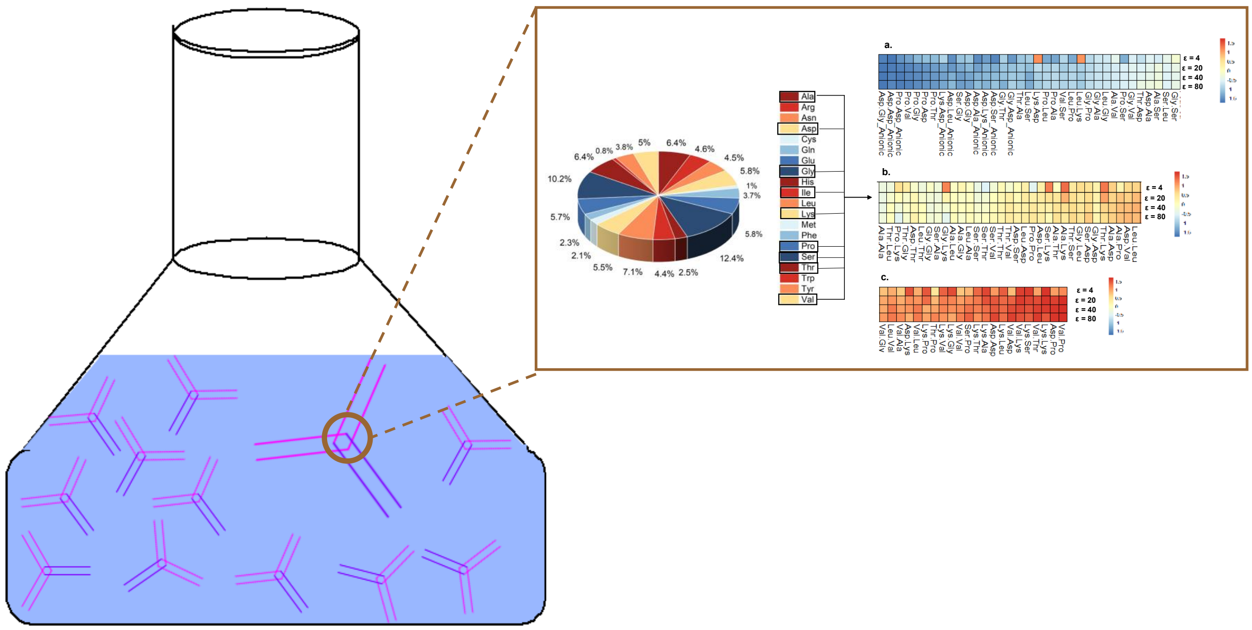

Figure 7.1.1. Multi-scale representation of non-enzymatic peptide bond fragmentation on hinge region fragmentation as a function of the pH ranges associated with the Protein A Chromatography stage of mAb downstream manufacturing.

For this purpose, we conducted a computational study, using a subset of dipeptides, to predict the thermochemistry and kinetic parameters of the acidic non-enzymatic peptide bond fragmentation mechanism. This paper presents the thermochemical and kinetic properties non-enzymatic peptide bond fragmentation under basic conditions (90 reactions (Appendix E) per modeled dielectric, 360 total reactions) in the conductor-like polarizability model (CPCM) solvated state, i.e. standard enthalpies of reaction, standard entropies of reaction, Gibbs free energies of reaction, activation energies, and kinetic rate constants. Density functional theory (DFT) at the B3LYP/6-31+G(2d,2p) level of theory with the GD3 empirical dispersion correction was used to calculate the electronic energy, and then statistical thermodynamics was applied to all fragmentation reactions to determine temperature effects on the reaction.[82, 141, 147, 187-190, 246, 247] All thermochemical properties were calculated at standard pressure and temperature to reduce external variabilities. Furthermore, supervised learning methods were used to identify the correlation between the parameters associated with the modeled reaction, and those learnings were applied to determine the reactions most susceptible to non-enzymatic peptide bond fragmentation under acidic conditions.

7.2. Objectives

The purpose of this study was to separately analyze the perturbations in the system pH associated with the downstream manufacturing of mAbs, where the reactions analyzed in this study are representative of non-enzymatic peptide bond fragmentation as a function of the Protein A chromatography stage of downstream manufacturing. By analyzing the Protein A chromatography and alkaline wash stages as separate studies, we were able to evaluate the similarities observed between the two pH-endpoints associated with the therapeutic protein manufacturing pipeline, and correlate those finds to mAb hinge region stability. Furthermore, we sought to determine what correlations were conserved between the pH-dependent fragmentation mechanisms in regards to reaction favorability and generalization of reaction energetics. Overall, this study examines the reaction energetics of the acidic mechanism of peptide bond fragmentation and identifies key correlations that can be used to evaluate the propensity of the hinge region towards non-enzymatic peptide bond fragmentation under acidic conditions.

7.3. Computational Methodology

7.3.i. Modeled Mechanism and Reaction Class

As we previously noted, the impact of the perturbations in system pH on the stability of the mAb hinge region during the downstream manufacturing of mAbs was treated as two separate studies: one to analyze the impact of the Protein A chromatography stage of downstream manufacturing (acidic conditions) and one under basic conditions to study the impact of the alkaline wash stage of downstream manufacturing on hinge region stability. For this purpose, we developed a similar reaction class to analyze the energetics of non-enzymatic peptide bond fragmentation for each stage of downstream manufacturing, where the only deviation was in regards to taking the ionization state of Aspartic Acid (Asp) into account. Similar to the Lysine (Lys) sidechain shifting from cationic to neutral during the alkaline wash stage of downstream manufacturing, the sidechain of Asp transitions from neutral to anionic during the Protein A chromatography stage of downstream manufacturing (Table 7.3.1).[65, 66, 71, 72, 74, 81, 515, 530, 531] Therefore, to achieve a robust analysis of the electrostatic interactions observed by the non-enzymatic peptide bond fragmentation reaction, we augmented our reaction class to include nine reactions containing anionic Asp.

Table 7.3.1. Ionization states of amino acid sidechains under the acidic conditions associated with downstream manufacturing.[65, 66, 71, 72, 74, 81, 515, 530, 531]

Amino Acid	**Acidic pH**			
	Ionizable sidechain 3.0 < pH < 3.90	**Charge**	**Ionizable sidechain 3.90 < pH < 5.0**	**Charge**
Ala	No	0	No	0
Asp	No	0	Yes	-1
Gly	No	0	No	0
Leu	No	0	No	0
Lys	Yes	1	Yes	1
Pro	No	0	No	0
Ser	No	0	No	0
Thr	No	0	No	0
Val	No	0	No	0

The mechanism of non-enzymatic peptide bond fragmentation under acidic conditions was adapted from the mechanism proposed by Sun et al. at a system pH of 3.0, where H NMR Spectra was implemented to compute the pseudo-first order kinetics associated with the reaction.[142] Sun proposed that the non-enzymatic fragmentation of Glycyl-Glycine (Gly-Gly) under acidic conditions occurred through proton coordination between H_3O^+ and the peptide bond carbonyl oxygen prior to the nucleophilic attack of H_2O to form an intermediate where the carbonyl carbon contains a hydroxyl and a protonated hydroxyl functional group(HO-C-OH^{2+}).[142] The intermediate was then proposed to return to a net neutral form (HO-C-OH) prior to its interaction with H^+ to cleave the amide bond, resulting in two Gly amino acids within the solution. [142] Preliminary modeling of the reaction mechanism proposed by Sun et al. indicated that the coordination between the carbonyl oxygen and H_3O^+ was equilibrated. Furthermore, without an explicit solvent molecule to interact with, the hydrogen atom would be attracted to the more electrophilic nitrogen atom of the peptide bond, instead of protonating one of the two hydroxyl functional groups.[532, 533] When the amide bond nitrogen was attacked by the proton instead of the hydroxyl oxygen, the generated intermediate structure was found to be fragmented in nature, suggesting that a concerted proton transfer occurred to cleave the amide bond during the electrophilic attack of H_3O^+. Furthermore, evaluation of the potential energy surface indicated that the reaction would continue to roll "downhill" from the fragmented intermediate to the infinitely separated products, suggesting that the change in protonation from the N-terminus residue to the C-terminus residue is equilibrated. For this purpose, we modeled the non-enzymatic peptide bond fragmentation under acidic conditions as a single reaction (Figure 7.3.1), where the equilibrated reactions are accounted for in the reactant and product species.

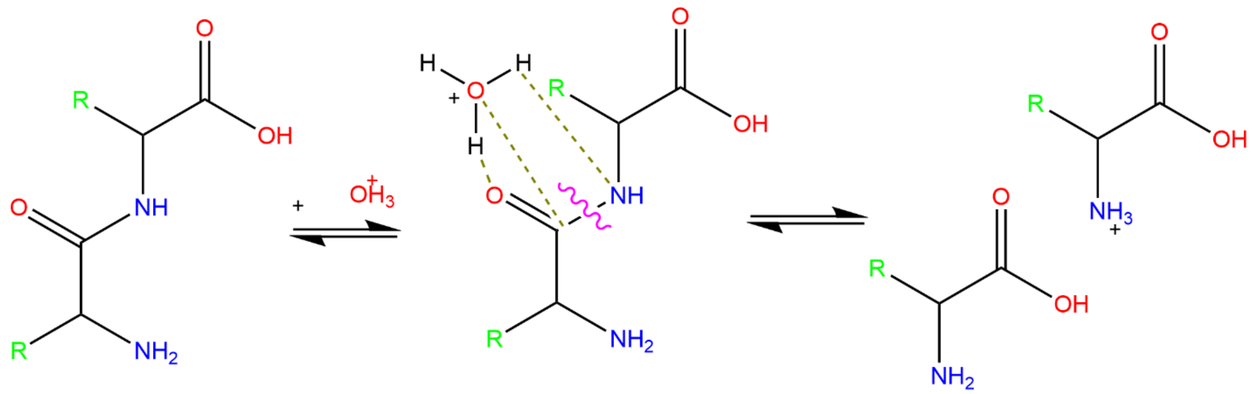

Figure 7.3.1. Proposed mechanism of non-enzymatic peptide bond fragmentation under acidic conditions.

All modeled reactions were generated using four modeled dielectrics to maintain our self-consistent methodology and achieve a robust analysis of the impact of electrostatic interactions on reaction behavior.[210-220, 489] The inclusion of the four different modeled dielectrics (ε = 4, 20, 40, and 80) is representative of the different sidechain functionalities modeled in this study, as well as different sites where the protein degradation might occur.[210-220, 489] Specifically, the reaction energetics exhibited by small polypeptides floating freely in solution would likely be represented by ε = 80, while the ionic and charged interactions observed at the protein surface would be represented by ε = 40 and ε =20, respectively.[82, 141, 147, 212, 213, 216, 223] Lastly, ε = 4 was used to mimic the reaction energetics of non-enzymatic peptide bond fragmentation at the hydrophobic core of the protein.[82, 141, 147, 212, 213, 221, 222, 489] Through this analysis, we achieved an in-depth understanding regarding the impact of electrostatics on reaction favorability towards non-enzymatic peptide bond fragmentation under acidic conditions.

7.3.ii. Quantum Chemical and Statistical Thermodynamics Calculations

Expansion of our previously developed non-enzymatic peptide bond fragmentation database was completed through the use of a self-consistent methodology.[82, 141, 147, 251] Quantum chemical calculations were performed with *Gaussian 16*.[82, 141, 147, 250, 251] All electronic energies for the protein dipeptides, free amino acids, intermediates and transition states were calculated using the B3LYP method.[82, 141, 147, 246, 247, 251, 252] The electronic wave functions for all structures were optimized in the singlet state. Geometries and harmonic frequencies of the lowest energy conformers were determined at the B3LYP/6-31+G(2d, 2p) level with a Grimme D3 (GD3) empirical dispersion correction to account for van der Waals forces within the modeled reactions.[82, 141, 147, 188-190, 246, 247, 253] To account for anharmonicity in the normal vibrational modes, as suggested by Scott and Random, the harmonic frequencies and zero-point energy (ZPE) were scaled by factors of 0.96 and 0.98 respectively.[82, 141, 147, 254] Following the work of previous reaction modeling completed by our group, partition functions based on the harmonic oscillator and rigid rotor approximations were used to calculate thermodynamic and kinetic properties as a function of temperature.[82, 141, 147, 255, 256] This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 257-263]

7.3.iii. Transition State Theory and Kinetic Parameters

Transition states associated with the acidic mechanism of peptide bond fragmentation were found using the Berny method.[264-266] The imaginary frequency of each transition state was animated, and intrinsic reaction-coordinate calculations were performed to confirm that the normal vibrational mode pertained to the reaction coordinate of interest. Conventional transition-state theory (TST) was then used to calculate rate coefficients at 1 atm, according to the macroscopic formulation in Equation (7.3.1).[267]

$$k^{TST}(T) = n_d \acute{k} = n_d \Lambda \exp\frac{\Delta S^{\neq}}{R} \exp(\frac{-\Delta H^{\neq}}{RT}) \quad (7.3.1)$$

Where Λ is defined in Equation (7.3.2):

$$\Lambda = \kappa(T)\frac{k_B T (V_m)^{-\Delta n}}{h} \quad (7.3.2)$$

Here, $\acute{\kappa}$ is the single-event rate coefficient, κ(T) is the Wigner tunneling correction at temperature T, k_B is Boltzmann's constant, h is Planck's constant, V_m is the molar volume, R is the ideal gas constant, $\Delta S^{\neq}$ is the entropy of activation, $\Delta H^{\neq}$ is the enthalpy of activation, Δn is the change in the number of moles going from the reactant to the transition state (i.e., -1 for bimolecular reactions, the forward and reverse reactions), and n_d is the reaction path degeneracy, or number of single events. The reaction path degeneracy follows Equation (7.3.3). $\Delta H^{\neq}$ and $\Delta S^{\neq}$ are calculated using standard formulae, and there were no internal rotation corrections needed within our system.[82, 141, 147, 196, 267]

The reaction path degeneracy, n_d, definition is based upon the symmetry of the reactant(s) and the transition state.[82, 141, 147, 196] Symmetry was determined by the number of chiral centers present in each molecule.[82, 141, 147, 196, 239, 256, 263] To achieve an unbiased representation of each modeled reaction, the maximum symmetry numbers were applied to each reaction.[82, 141, 147, 239] Analysis of the reaction path degeneracies used in our model, noted that the inclusion of chiral centers did not impact the forward reaction path degeneracy developed through the exclusion of chiral centers. Therefore, our model may be applied to both natural and synthetic proteins in the forward direction of each computed reaction. The reaction path degeneracies for this study follow Equation (7.3.3):

$$n_{\mathrm{d}} = \frac{n_{\mathrm{chiral}}^{\neq}}{\prod_{\mathrm{j}} n_{\mathrm{chiral}}^{\mathrm{reactant,j}}} \times \frac{\prod_{\mathrm{j}} \sigma^{\mathrm{reactant,j}}}{\sigma^{\neq}} \tag{7.3.3}$$

Where n_{chiral} is the number of chiral isomers for the transition state and reactant(s); σ is a product of the external rotation symmetry numbers for the transition state and reactant(s). Internal symmetry numbers for non-cancelling internal rotations treated as hindered or free rotations were included in the hindered and free rotor partition functions, respectively.

The single-event parameters of the Arrhenius relationship, log(A) and E_A were obtained by fitting ln(k) versus T^{-1} over the temperature range 298 - 1500 K. This procedure was performed automatically using a modified version of the *CalcK* script previously employed by our group for reaction kinetics analysis.[82, 141, 147, 255, 257-263] Through linear regression analysis, we determined Arrhenius behavior was obeyed well for the majority of the reactions modeled within this study, where non-Arrhenius behavior was observed by the reactions which contained negative apparent activation barriers. The solvation impact of liquid phase kinetics is beyond the scope of this work but will be addressed in future studies following the work of Jalan et al, which computes liquid-phase kinetics in relation to both the solute and the solvent.[268]

7.3.iv. Chemical Model and Level of Theory Validation

Model validation was completed by comparing the forward rate coefficients of Glycyl-Glycine (Gly-Gly) and Glycyl-Valine (Gly-Val) non-enzymatic peptide bond fragmentation at a pH of 3.0 to the literature values provided by Sun et al. and by Smith and Hansen (Table 7.3.2). [142, 144] Validation was completed using the values computed at the modeled dielectric associated with bulk water to reduce variability within the analyzed system. Similarly, as the forward rate coefficient given by Sun et al. falls between two temperature ranges of the *CalcK* output, the value used as a validation point was achieved via linear interpolation. Furthermore, analysis of our modeled system suggested that, similar to our study under ambient conditions, a pseudo-first-order approximation would be an appropriate metric to apply to our analysis. Through this, we determined the forward rate coefficient of Gly-Val non-enzymatic peptide bond fragmentation was within one order of magnitude of the value given by Smith and Hansen.[144] Evaluation of the forward rate coefficient of Gly-Gly non-enzymatic peptide bond fragmentation indicated that a greater margin of error was observed between the computed and literature values, and can be attributed to the linear interpolation applied to the generated value.[142] Therefore, we determined that both our self-consistent chemical model and computational methodology would accurately reflect the reaction energetics of peptide bond fragmentation under acidic conditions.

Table 7.3.2. Comparative analysis of the computed Arrhenius parameters and forward rate coefficients of Gly-Gly and Gly-Val peptide bond fragmentation under acidic conditions against the experimental rate coefficients [ε = 80].[142, 144] The rate coefficients of the value compared against the Sun study was computed via interpolation of the rate coefficients derived from Arrhenius parameterization, so that temperature variability could be removed from the analysis. Furthermore, the comparative rate coefficients of Gly-Gly and Gly-Val peptide bond fragmentation were computed using a pseudo-first order approximation.

Study	Dipeptide Reaction	pH	Temperature (K)	k (s^{-1})	Computed k (s^{-1})	Δk^F (Literature – Computed)
Sun et al. (2019)	Gly-Gly	3.0	368	4.80×10^{-7}	9.30×10^{-6}	5.16×10^{-2}
Smith and Hansen (1998)	Gly-Val	3.0	310	1.35×10^{-9}	1.01×10^{-8}	1.34×10^{-1}

7.3.v. Supervised Learning Methods and Applications towards Hinge Region Stability

Evaluation of the Evans-Polanyi correlations generated for the full dataset, as well as at each modeled dielectric indicated that zero statistically significant correlations could be achieved by generalizing the relationship between the enthalpy of reaction and the activation barrier. As we previously noted, the fundamental principle of the Evans-Polanyi correlation is that the more exothermic reactions will have lower activation barriers.[224] Analysis of the thermochemical and kinetic parameters associated with the non-enzymatic peptide bond fragmentation reaction under acidic conditions indicated that the computed activation energies did not scale linearly with the exothermicity of the modeled reactions, suggesting that the Evans-Polanyi correlation would be an inappropriate model to apply to the developed reaction class.

To confirm our hypothesis that the Evans-Polanyi correlation would be inappropriate for our modeled system, we applied principal component analysis (PCA) to our generated dataset. PCA was conducted through the use of an R script that was adapted from PCA tutorial codes and applied to the full dataset associated with the non-enzymatic peptide bond fragmentation reaction under acidic conditions.[491, 534] As we previously noted, PCA provides valuable insight towards identifying the parameters within a dataset that contribute the most significantly to the observed variation.[236, 237, 490] Additionally, the calculation of the correlation coefficients between parameters can be implemented to identify the most significant relationships.[534] Therefore, through the application of supervised learning methods, we were able to identify the parameters that contributed to non-enzymatic peptide bond fragmentation under acidic conditions and obtain insights towards their impact on mAb hinge region stability.

7.4. Results and Discussion

7.4.i. Structures

As we previously noted, analysis of the stages associated with downstream manufacturing of mAbs was completed using similar reaction classes, where the nine deviating structures were representative of the residue ionization states of interest. To model the reaction energetics of non-enzymatic peptide bond fragmentation under the conditions associated with the Protein A chromatography stage of downstream manufacturing, nine reactions including anionic Asp were appended to the 81 reactions used to generate our initial reaction class. Similar to the reactions used to model non-enzymatic peptide bond fragmentation under the environmental conditions associated with the alkaline wash stage of downstream manufacturing, the initial reaction class was developed by modeling the nine most frequently occurring residues within the hinge region in both the N and C-terminus position.[163]

The mechanism of non-enzymatic peptide bond fragmentation under acidic conditions was modeled as a single reaction, where the proton coordination between H_3O^+ and the peptide bond carbonyl oxygen was found to be instantaneous and lumped into the rate-determining step. Similarly, initial analysis of the modeled mechanism indicated that peptide bond fragmentation occurred during the electrophilic attack of H_3O^+ on the peptide bond. The dipeptide and H_3O^+ were treated as infinitely separated reactants, where H_3O^+ decomposed in the transition state structure to cleave the peptide bond and form a net neutral N-terminus residue and a cationic C-terminus residue that were treated as infinitely separated products. A representative ball and stick schematic of the stepwise mechanism is present in Figure 7.4.1.

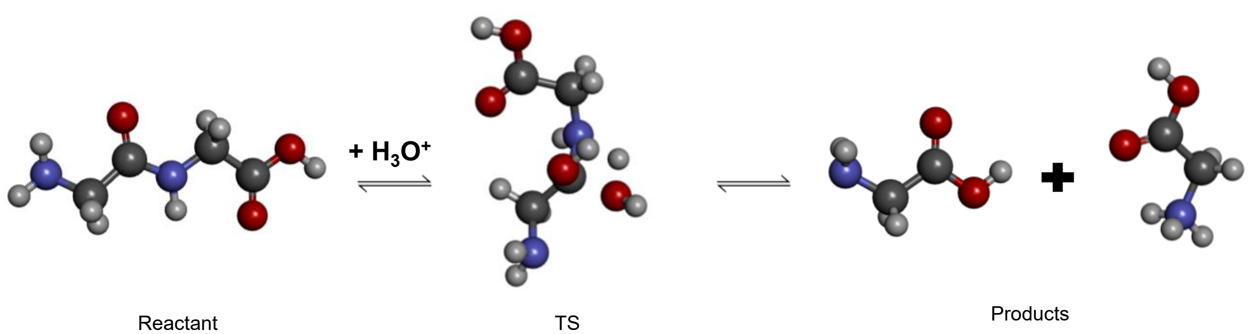

Figure 7.4.1. Ball and stick representation of non-enzymatic peptide bond fragmentation under acidic conditions.

7.4.ii. Thermochemistry and Arrhenius Parameterization

Non-enzymatic peptide bond fragmentation under acidic conditions was found to be both net exothermic and spontaneous in nature where, similar to non-enzymatic peptide bond fragmentation under basic conditions, reaction favorability increased as the modeled dielectric decreased, where the largest increase in reaction favorability was observed between the modeled dielectrics associated with hydrophobic and charged residues (Table 7.4.1). At the modeled dielectric associated with charged residues, $E_0 = 18.9 \pm 4.78$ kcal.mol^{-1}, $E_A = 18.4 \pm 4.89$ kcal.mol^{-1}, and log(Ã) = 5.85 ± 0.47 L.mol^{-1}.s^{-1} in the forward direction, while $E_0 = 44.2 \pm 5.48$ kcal.mol^{-1}, $E_A = 45.6 \pm 5.64$ kcal.mol^{-1}, and log(Ã) = 5.49 ± 0.45 s^{-1}, respectively. At ε = 4, $E_0 = 11.4 \pm 10.7$ kcal.mol^{-1}, $E_A = 11.2 \pm 10.9$ kcal.mol^{-1}, and log(Ã) = 5.84 ± 0.61 L.mol^{-1}.s^{-1} in the forward direction, while $E_0 = 40.1 \pm 10.4$ kcal.mol^{-1}, $E_A = 41.6 \pm 10.6$ kcal.mol^{-1}, and log(Ã) = 5.58 ± 0.61 s^{-1}, respectively.

Table 7.4.1. Average B3LYP Arrhenius parameters, zero point energy corrected barriers, enthalpy of reaction, and Gibbs free energy of reaction for the intermediate formation step of the non-enzymatic peptide bond fragmentation reaction under basic conditions at 298 K and 1 atm. Ã has units of L.mol^{-1}.s^{-1} for the forward and reverse reactions, respectively. E_0, E_A, and $\Delta H^{298.15}_{rxn}$ have units of kcal.mol^{-1}.

ε	Forward Reaction			ΔH_{RXN}	ΔG_{RXN}	Reverse Reaction		
	E_0	log(A)	E_A			E_0	log(A)	E_A
4	11.4	5.84	11.2	-30.9	-31.5	40.1	5.58	41.6
20	18.6	5.85	18.4	-27.7	-28.4	44.2	5.49	45.6
40	19.5	5.84	19.2	-27.3	-28.1	44.7	5.43	46.0
80	19.9	5.83	19.6	-27.1	-27.9	44.9	5.42	46.2

Evaluation of the reactions modeled under the most hydrophobic dielectric indicated that the reactions containing anionic Asp yielded a similar phenomena to the trend observed by cationic Lys under hydrophobic conditions associated with non-enzymatic peptide bond fragmentation under alkaline conditions. Specifically, negative apparent activation barriers were observed by the reactions containing cationic Asp at ε = 4. Prior studies on hydrogen exchange have indicated that the phenomenon is catalyzed by either strong acids or strong bases.[535-537] Prior studies have highlighted anionic Asp as a residue that is prone towards promoting hydrogen exchange under acidic conditions, where Asp was computed as the residue to promote the highest rate of hydrogen exchange in both the N and C-terminus positions.[500, 501, 538] Additionally, it is likely that the polar interactions that would be exhibited between the residue and the modeled dielectrics at ε =20, 40, and 80 would stabilize it against the phenomenon. Therefore, as the reactions modeled within this study do not account for the salt bridges and other electrostatic interactions typically observed within the hydrophobic core to protect the acidic sidechain from hydrogen exchange, it is likely that the observed energetics would not be physiological in larger structures.

7.4.iii. Application of Supervised Learning Methods towards Identifying Contributing Factors of Non-enzymatic Peptide Bond Fragmentation under Acidic Conditions

As we previously indicated, one of the primary motivators of this study was to identify the parameters that contributed the most significantly to the energetics associated with non-enzymatic peptide bond fragmentation under acidic conditions, and correlate those findings to the stability of the mAb hinge region. Comparative analysis of the correlation heatmaps generated for the acidic and basic datasets (Figure 7.4.2) yielded significant insight towards why the Evans-Polanyi correlation would be an appropriate generalization of the energetics associated with the proposed rate-determining step of non-enzymatic peptide bond fragmentation under basic conditions, but not under acidic conditions. As indicated by the color legends on the right hand side of each heatmap, the correlation observed between the parameters associated with the respective datasets ranged from strongly positive (dark red) to strongly negative (dark blue).[491, 534, 539] Evaluation of the basic correlation heatmap (Figure 7.4.2.b) indicated evidence of a strong positive correlation between the enthalpy of reaction and the activation barrier. Similarly, both ΔH_{RXN} and E_A were found to observe strong negative correlation when paired with the overall net charge of the reaction. Furthermore, evaluation of the pair correlation coefficients indicated that the correlation observed between ΔH_{RXN} and E_A was extremely strong ($\rho = 0.91$), while the correlations observed between the reaction net charge and ΔH_{RXN} ($\rho = -0.54$), and the reaction net charge and E_A ($\rho = -0.54$) were computed to be moderate in comparison.[371, 372] Therefore, the stronger correlation observed between the enthalpy of reaction and activation energy indicated that the Evans-Polanyi correlation would be suitable for the basic dataset, and the inclusion of the reaction net charge parameter could be implemented to improve model linearity and predictiveness.

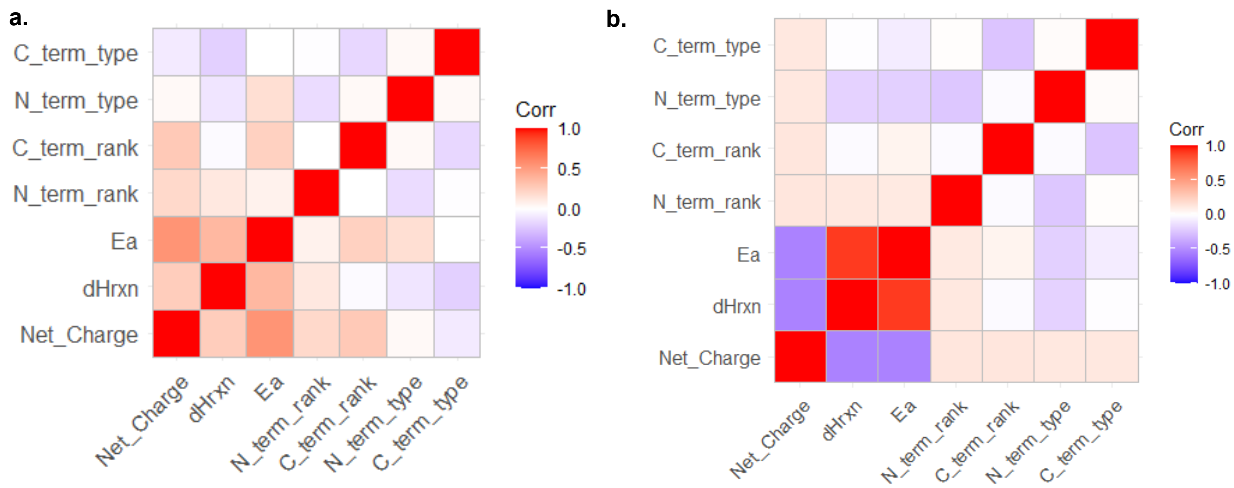

Figure 7.4.2. Heatmap of the correlation observed between the parameters associated with non-enzymatic peptide bond fragmentation [full dataset]. (a). Fragmentation under acidic conditions, (b). Fragmentation under basic conditions.

Conversely, the strongest correlation observed in the acidic dataset was found in the correlation associated with the reaction net charge and the activation barrier ($\rho = 0.55$). Evaluation of the correlation observed between ΔH_{RXN} and E_A indicated a weaker correlation ($\rho = 0.37$) was observed relative to the strongest correlation, suggesting that the associated net charge of the reaction was more responsible for the variability observed in the computed activation barriers relative to the associated enthalpy of reactions. Furthermore, comparative analysis of the correlation coefficients observed for the acidic and basic datasets indicated that a greater number of parameters contributed to the variability observed in the acidic dataset. Specifically, evaluation of the correlation coefficients in the basic dataset indicated that the difference in the correlation observed between the reaction net charge and ΔH_{RXN} was significantly greater than the fourth strongest correlation (C-terminus Rank: C-terminus Type; $\rho = -0.25$). Additionally, the correlation observed between ΔH_{RXN} and the other parameters, or E_A and the other parameters, was computed to be weak or negligible. [371, 372] In the acidic dataset, the observed decrease in correlation strength was not as significant, suggesting that the variability observed within the acidic dataset could be accounted for through the inclusion of more parameters relative to the basic dataset.

As we previously noted, the Evans-Polanyi correlation was not applicable for non-enzymatic peptide bond fragmentation under acidic conditions due to a moderate linear correlation being observed between the associated reaction net charge and the activation barrier, while relatively weak correlations were observed between the net charge and the enthalpy of reaction and between the enthalpy of reaction and the activation energy. For this purpose, we expanded the descriptors used within our dataset to obtain a robust analysis of the parameters contributing to the variation observed within our generated dataset. The flexibility of the hinge region has been identified as a factor that influences its propensity towards fragmentation.[138, 157, 158, 480, 519-522] Furthermore, three out of the four residues determined by Flores to be the most frequently occurring within the hinge region are classified as small via the Sigma Aldrich datasheet on amino acid molecular weights.[163, 164] For this purpose, we expanded our analysis of the parameters included in PCA to evaluate the impact of the residue size in either the N or C-terminus position, as well as the flexibility of the residues in accordance to the scale developed by Huang and Nao.[365]

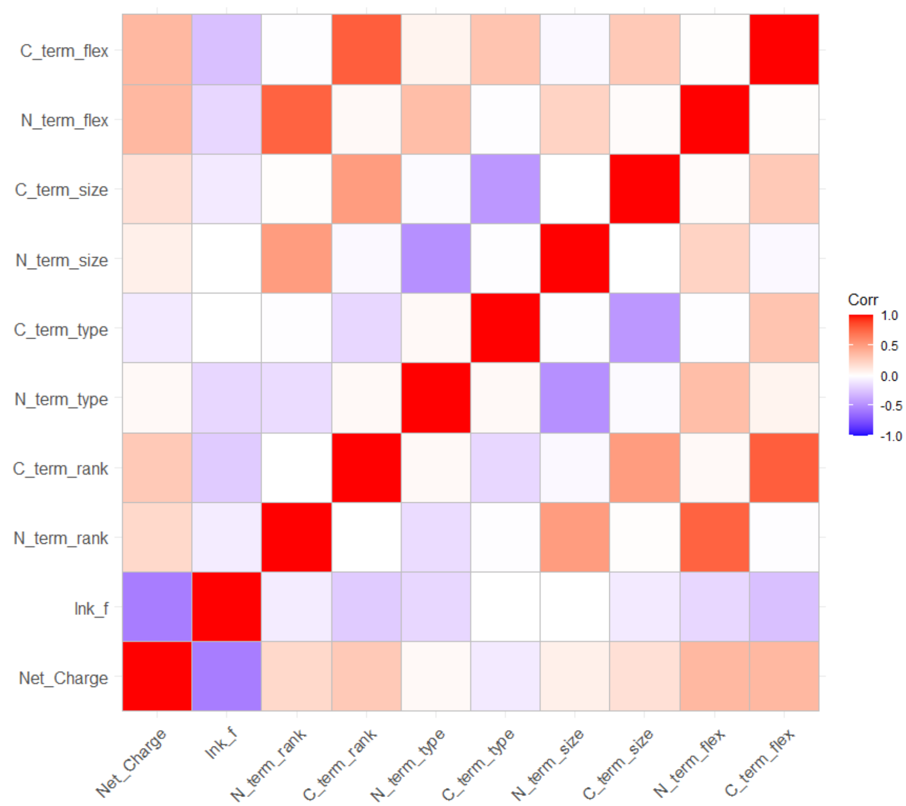

Figure 7.4.3. Heatmap of the correlation observed between the parameters associated with non-enzymatic peptide bond fragmentation [full dataset] relative to $\ln(k)^F$.

Figure 7.4.3. depicts the heatmap of the correlation between the parameters observed within the dataset, where the parameters associated with the enthalpy of reaction and activation barrier were replaced with $\ln(k)^F$ at 298 K. Evaluation of the correlation between $\ln(k)^F$ and the other parameters observed within the dataset indicated that the most significant linear correlation was observed between $\ln(k)^F$ and the associated reaction net charge (ρ = -0.56). Additionally, weakly negative linear correlation was observed between $\ln(k)^F$ and the following parameters: the propensity of the C-terminus residue to occur in the hinge region (ρ = -0.22), the N-terminus residue sidechain type (ρ = -0.17), the flexibility of the N-terminus residue (ρ = -0.17), and the flexibility of the C-terminus residue (ρ = -0.27).[371, 372] For this purpose, only the previously noted parameters were included within PCA.

7.4.iv. Identification of Residues Promoting Non-enzymatic Peptide Bond Fragmentation and Insights towards mAb Hinge Region Stability under Acidic Conditions

Table 7.4.2 gives the parameter contributions to variability relative to the direction associated with the greatest amount of variance (PC 1), and the direction orthogonal to it (PC 2). Evaluation of the parameter contributions observed in the PC 1 direction indicated that $\ln(k)^F$, the associated reaction net charge, and the flexibility of the C-terminus residue were the three most significant contributors to variability.[235, 236] In the direction orthogonal to PC 1, the flexibility of the N-terminus residue towards fragmentation, functionality of the N-terminus sidechain, and the propensity of the C-terminus residue to occur within the hinge region were found to contribute the most significantly to variability. Therefore, we hypothesized that the impact of the previously noted parameters on the favorability of the reaction would be ranked as follows: associated reaction net charge, C-terminus residue flexibility, N-terminus residue flexibility, N-terminus sidechain functionality, propensity of the C-terminus residue to occur within the hinge region. Via this analysis, we examined the propensity of the modeled reactions towards non-enzymatic peptide bond fragmentation under acidic conditions via the ranking of their forward rate coefficients as a function of the modeled dielectric.

Table 7.4.2. Non-enzymatic peptide bond fragmentation under acidic conditions parameter contributions to variability [Full Dataset].

PC 1		PC 2	
Parameter	**Contribution to Variability**	**Parameter**	**Contribution to Variability**
Net Charge	25.5	Net Charge	6.08
$\ln(k)^F$	33.1	$\ln(k)^F$	9.03
C-terminus residue ranking	18.3	C-terminus residue ranking	17.1
N-terminus residue sidechain type	0.01	N-terminus residue sidechain type	20.1
N-terminus residue flexibility	0.97	N-terminus residue flexibility	33.0
C-terminus residue flexibility	22.1	C-terminus residue flexibility	14.8

7.4.iv.a. Impact of Net Charge on Reaction Favorability

Evaluation of the ranked forward rate coefficients by their propensity towards non-enzymatic peptide bond fragmentation under acidic conditions indicated that the associated net charge of the reaction impacted reaction favorability. Figure 7.4.4 depicts the reaction favorability as a function of the modeled dielectric, where the color scale ranges from dark blue (most kinetically favorable reactions) to dark red (least kinetically favorable reactions), and is split into the three reaction net charges associated with non-enzymatic peptide bond fragmentation under acidic conditions.

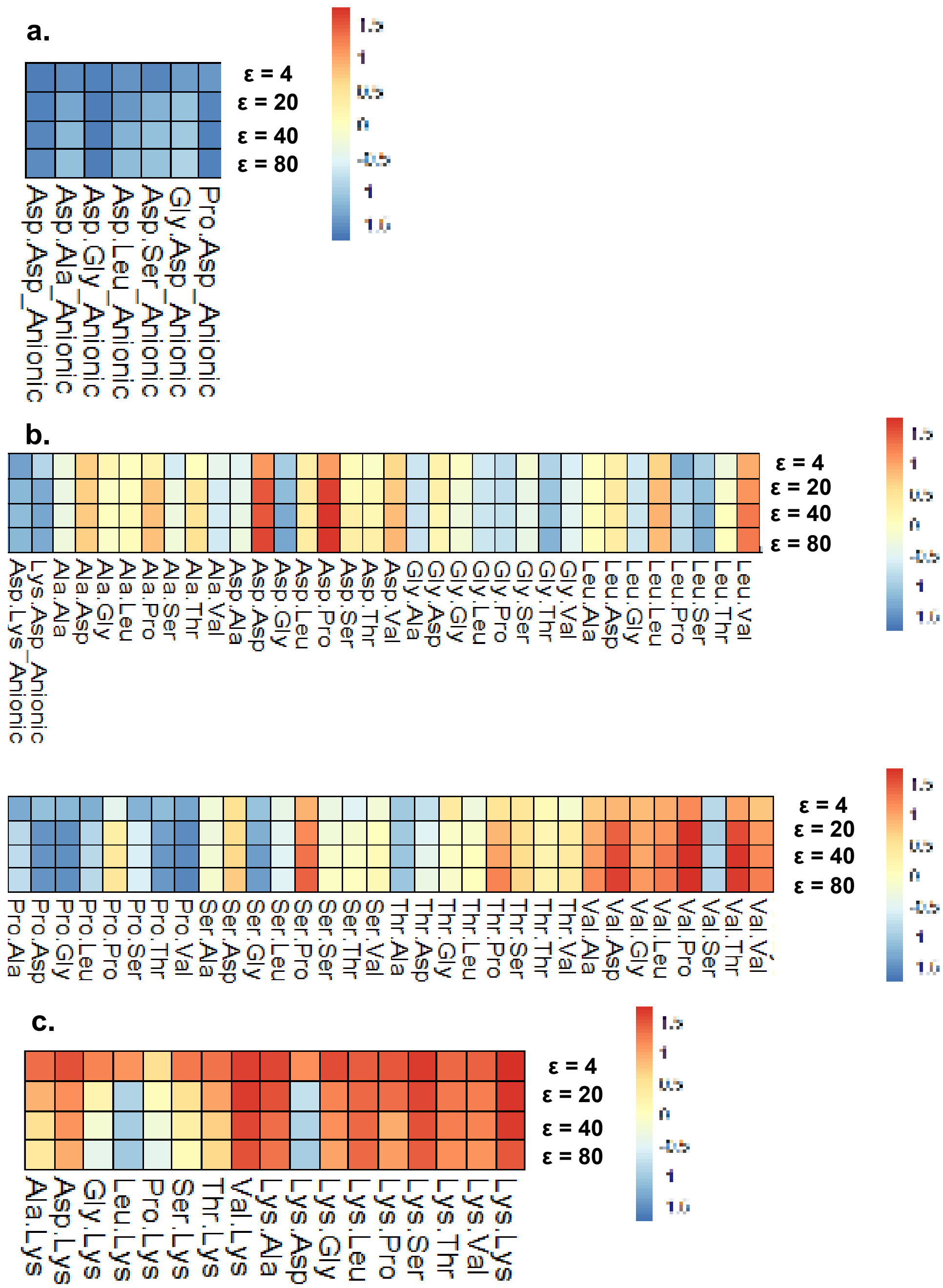

Figure 7.4.4. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed kF. Dipeptides are arranged via the associated net charge of reaction. (a). Anionic and net neutral reactions, (b). Cationic reactions (charge = 1), (c). Cationic reactions (charge ≥ 2).

Analysis of the heatmap indicated that reactions containing Anionic Asp generated the quickest forward rate coefficients, while reactions containing Lys in the N-terminus were among the slowest. As we previously noted, all parameters that were found to be weakly correlated to the forward rate coefficient exhibited negative correlation. In regards to the relationship between k^F and associated reaction net charge, the negative correlation indicated that as the reaction would become more kinetically unfavorable as it became more cationic.[372] This trend is representative of the electrostatic repulsion exhibited by like charges, where the electrophilic attack of H_3O^+ would be repulsed by the cationic Lys sidechain.[503, 513-515] Similarly, reactions containing Asp in its ionized state (net anionic) would be attracted to the cationic H_3O^+, and therefore the forward rate coefficient would be increased by the electrostatic attraction.[503, 513, 514, 517, 518]

However, the upper and lower bounds of reaction favorability were not limited to the reactions where charge-charge interactions were exhibited between the modeled dipeptide and H_3O^+. Evaluation of Figure 7.4.4.b indicated that the k^F associated with the reactions containing net neutral residues ranged from large forward rate coefficients to small coefficients. The range of forward rate coefficients, when the net charge of the reaction is 1, indicates that the propensity of dipeptides towards non-enzymatic peptide bond fragmentation is influenced by parameters other than the associated net charge of the reaction. As we previously specified, the propensity of the C-terminus residue to occur in the hinge region and the functionality of the N-terminus sidechain were also found to exhibit weakly linear correlation to k^F. Therefore, it is likely that the size of the forward rate coefficient associated with the computed reactions is dependent on all three parameters.

7.4.iv.b. Impact of Residue Flexibility on Reaction Favorability

As we previously noted, the flexibility of the N and C-terminus residues were found to observe weak linear correlation the forward rate coefficient. For this purpose, we generated heatmaps of the reaction favorability as a function of the modeled dielectric (Figure 7.4.5). Of the

10 reactions containing Gly in the C-terminus, only two reactions fell within the least kinetically favorable range (red) [Lys-Gly and Val-Gly]. As we previously noted, reactions containing cationic Lys in the N-terminus were found to be kinetically unfavorable due to the electrostatic repulsion of like-like charges.[503, 513-515] Additionally, Val is classified as one of the least flexible residues within our dataset.[365] Unlike Val-Ser, which is ranked among one of the more kinetically favorable reactions within our dataset, the generated heatmap indicates that Val-Gly fragmentation under acidic conditions is kinetically unfavorable. The favorability of Val-Ser towards non-enzymatic peptide bond fragmentation under acidic conditions is likely due to the polar interactions observed between the hydroxyl of the Ser functional group and the reactive center, where the distance is 1Å less than the distance between the Asp functional group and the reaction center.[239, 503] Furthermore, the hydroxyl group of the Thr sidechain was oriented away from the reaction center, suggesting that the sidechain would not be able to interact with the reaction center. As a result, only the Val-Ser reaction would be kinetically favorable.

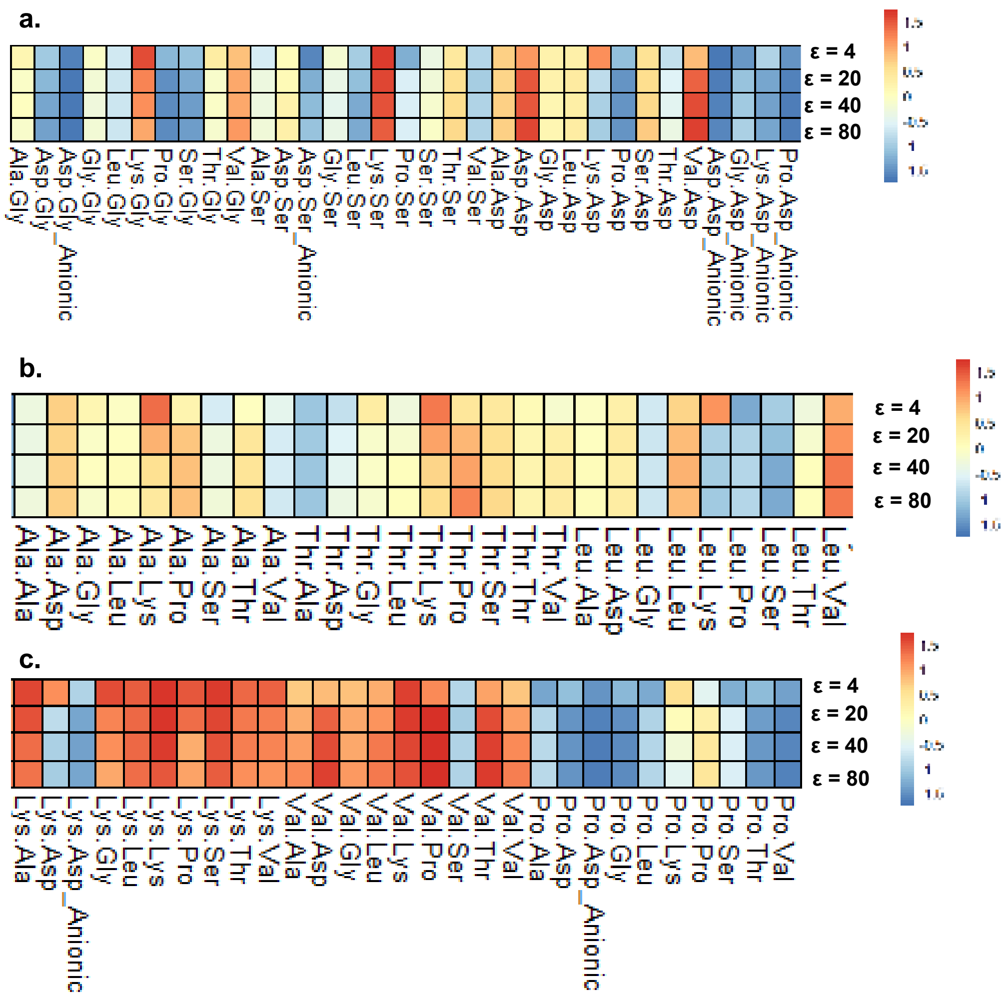

Figure 7.4.5. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed k^F. Dipeptides are arranged via the flexibility of the C-terminus residue. (a). Most flexible, (b). Medium flexibility relative to other residues in dataset, (c). Least flexible.

Reactions containing C-terminus Ser indicated that the most kinetically favorable reactions occurred when Ser was covalently bound to inflexible residues in the N-terminus or anionic Asp. As we previously noted, polar interactions between the Ser sidechain and the reactive center would increase reaction favorability.[239, 503, 540] The reduced favorability of the reactions containing small, flexible residues covalently bound to one another was observed by all reactions aside for Gly-Ala and Ser-Gly (Figure 7.4.6), where all nine reactions were found to fall within the top half of reaction favorability. Furthermore, the reduced favorability of the reactions is likely a result of the other factors influencing the reaction favorability of non-enzymatic peptide bond fragmentation under acidic conditions.

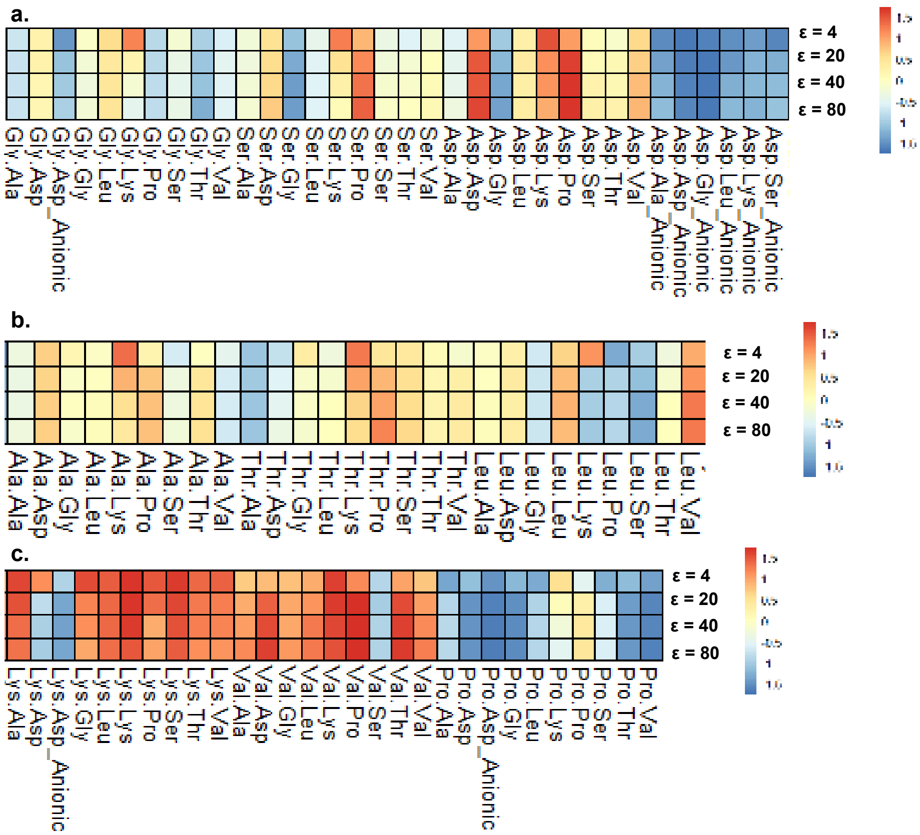

Figure 7.4.6. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed k^F. Dipeptides are arranged via the flexibility of the N-terminus residue. (a). Most flexible, (b). Medium flexibility relative to other residues in dataset, (c). Least flexible.

7.4.iv.c. Impact of Residue Sidechain Functionalities on Reaction Favorability

The correlations associated with the residue sidechain functionality and k^F and the propensity of the C-terminus residue to occur in the hinge region and k^F were found to exhibit weaker linear correlation relative to the relationship observed by the associated reaction net charge and k^F. However, similar to the associated reaction net charge, the functionality of the residue sidechain was found to not be position dependent at ε = 4. As depicted in Figure 7.4.7, the rate coefficients associated with dipeptides containing Lys were found to decrease at ε = 4, regardless of the residue position in the dipeptide. Similarly, the reactions containing anionic Asp were more significantly impacted than the reactions containing Asp in its unionized form. Prior studies on charged residues in the hydrophobic core of proteins found the presence of charged residues that do not form salt bridges to be destabilizing.[502, 515, 541-545] While the sidechain of Lys is cationic in nature, the reactions containing Asp were modeled using the ionized and neutral form of the acidic sidechain. Therefore, the net charge associated with the reaction would either be net neutral, where the charges are canceled by the electrostatic attraction between the anionic sidechain of Asp and the cationic H_3O^+, while the reactions containing neutral Asp behaved similarly to the reactions containing unionized residues. As a result, only the Lys-containing reactions would be impacted by the energetic cost of buried residues within the hydrophobic core of the protein.[502, 515, 541-545]

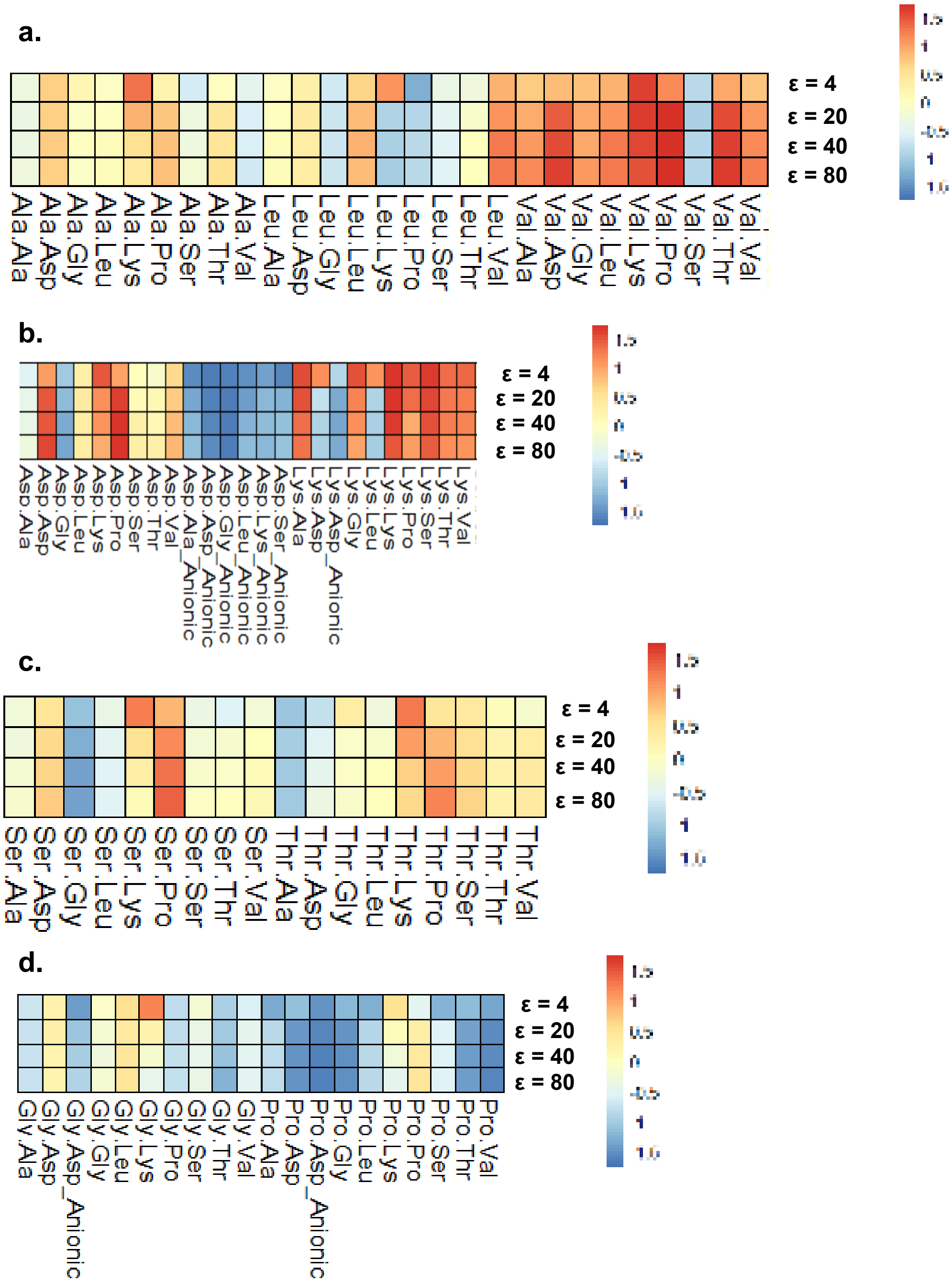

Figure 7.4.7. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed k^F. Dipeptides are arranged via the functionality of the N-terminus residue sidechain. (a). Aliphatic, (b). Charged, (c). Hydrolytic, (d). Special.

Evaluation of the trends exhibited in Figure 7.4.6 indicated that the non-enzymatic peptide bond fragmentation was more kinetically favorable when Pro was the N-terminus residue. The propensity of Pro-Ybb towards fragmentation was found to contradict prior experimental findings, where the "Proline effect" indicated that the presence of Pro in the C-terminus position would promote the fragmentation of polypeptides.[515, 525-529] The rate coefficients associated with our study on the energetics of non-enzymatic peptide bond fragmentation during the alkaline wash stage of downstream manufacturing indicated that the reactions containing Pro in the C-terminus position were among the least kinetically favorable, and agree with the finding by Loo et al. that the proximity of the residue to a basic environment reduced its propensity towards fragmentation.[527] Furthermore, prior studies on the "Proline effect" indicated that polypeptides containing Pro were more likely to fragment when the polypeptide was protonated.[528, 529, 546] Previous studies indicated that the "Proline effect" was representative of the affinity of the amine of the Pro ring to accept protons relative to other amino acids.[547-549] Evaluation of the transition state structures indicated that the proton closest to the Pro ring amine was oriented towards the functional group, while it was oriented towards the peptide bond in the structures of the other N-terminus residues. Therefore, it is likely that the N-terminus Pro acts as a proton acceptor during the electrophilic attack of H_3O^+, thus increasing the propensity of the dipeptides towards non-enzymatic fragmentation under acidic conditions. Only Pro-Lys and Pro-Pro were found to deviate from this trend, where the basic nature of the Lys sidechain and the steric hindrance of the C-terminus Pro likely reduce the favorability of the reaction.[138, 527]

7.4.iv.d. Impact of Residue Frequency on Reaction Energetics

As we previously specified, the correlation observed between the propensity of the C-terminus residue to occur in the hinge region and k^F was found to be weaker than the correlation associated with the associated net charge of reaction. Additionally, at the modeled dielectric associated with hydrophobic residues, negligible correlation was observed. Figure 7.4.8 depicts the heatmap of k^F as a function of the modeled dielectric, where the x-axis represents the propensity of the C-terminus residue to occur in the hinge region. Evaluation of the generated heatmap indicated that the rate coefficients associated with both Gly and Ser were larger when the residues were in the C-terminus position of the dipeptide relative to the N-terminus position. While the reactions containing Gly in the N-terminus position were found to be among the more kinetically favorable, the associated rate coefficients of Gly covalently bound to a bulkier residue indicated that the flexibility of the smallest amino acid promoted non-enzymatic peptide bond fragmentation.[138, 239] Similarly, the reactions containing N-terminus Ser towards non-enzymatic peptide bond fragmentation were most kinetically favorable when the residue was covalently bound to either the first (Gly) or third (Leu) most frequently occurring residues in the hinge region. Furthermore, the three most frequently occurring amino acids in the hinge region, according to Flores et al., were found to promote non-enzymatic peptide bond fragmentation when covalently bound to one another. Of the nine dipeptides, only Leu-Leu was found to not observe this trend; where the reaction unfavorability can be attributed to the bulky nature of the aliphatic sidechains.[163] As a result, we determined that, similar to the associated net charge of the reaction and the classification of the residue sidechain, the propensity of residues to occur in the hinge region influenced non-enzymatic peptide bond fragmentation.

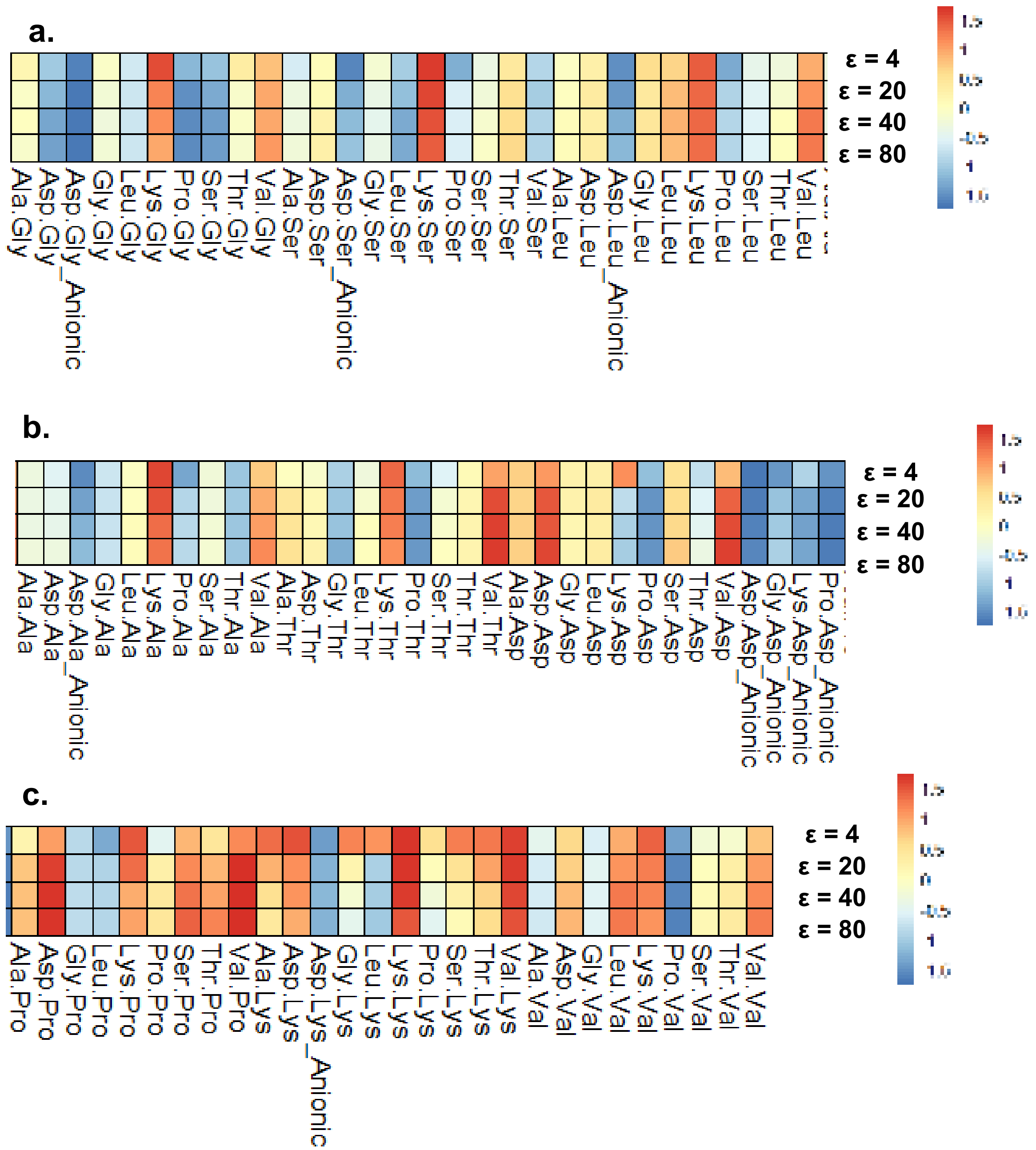

Figure 7.4.8. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed k^F. Dipeptides are arranged via the frequency of the C-terminus residue to appear in the hinge region. (a). Reactions containing Gly, Ser, or Leu in the C-terminus, (b). Reactions containing Ala, Thr, or Asp in the C-terminus, (c). Reactions containing Pro, Lys, or Val in the C-terminus.

7.4.7. Overall Insights towards mAb Hinge Region Stability

Figure 7.4.9 depicts the spectrum of reaction favorability of the generated reaction class towards non-enzymatic peptide bond fragmentation under acidic conditions. As we previously noted, multiple factors influenced the propensity of the modeled dipeptides towards non-enzymatic peptide bond fragmentation under acidic conditions; where the impact of the reaction net charge is most evident at $\varepsilon = 4$. The smaller forward rate coefficients associated with the Lys-containing residues can be attributed to both its net charge and its respective ranking in terms of frequency within the hinge region. Similarly, Val was the residue least likely to occur in the hinge region within our modeled dataset, and promoted fragmentation when in the C-terminus position and covalently bound to small residues. Val-Ser fragmentation was also found to be kinetically favorable, and can be attributed to both the frequency Ser occurs in the hinge region and its functional group.[163, 239] Specifically, the ability of the hydrolytic functional group to form hydrogen bonds with H_3O^+ to stabilize the reaction.[239, 503, 540]

Dipeptide reactions containing Gly, Ser, and Ala were found to promote non-enzymatic peptide bond fragmentation under acidic conditions, indicating that the propensity of small, flexible residues to occur within the hinge region and promote peptide bond fragmentation was conserved across pH-dependent environments. Conservation of the impact of reaction net charge upon reaction energetics was also observed, where reactions containing anionic Asp yielded large forward rate coefficients while the electrostatic repulsion between the cationic Lys sidechain and H_3O^+ significantly decreased reaction favorability.

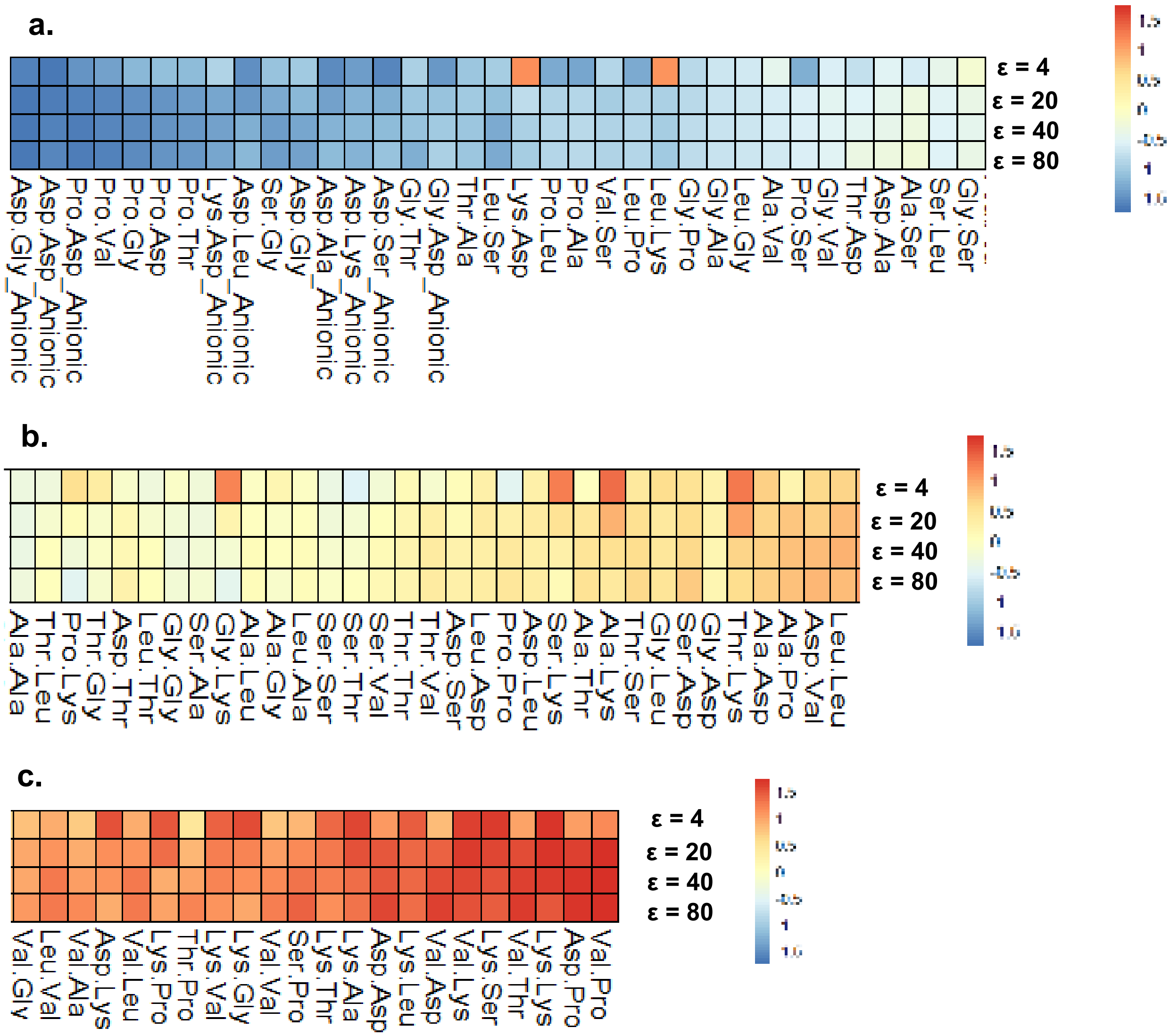

Figure 7.4.9. Heatmap of the computed k^F associated with non-enzymatic peptide bond fragmentation under acidic conditions as a function of the modeled dielectric. Dark blue represents the dipeptide reactions with the largest computed k^F, while dark red is associated with the reactions with the smallest computed computed k^F. (a). Most kinetically favorable reactions (b). Midpoint of reaction favorability, (c). Least kinetically favorable reactions.

Evaluation of the parameters influencing k^F indicated the following. First, that unlike the energetics exhibited by the fragmentation reactions modeled under the environmental conditions associated with the alkaline wash stage of downstream manufacturing, propensity towards fragmentation was not driven only by the presence of flexible residues within the modeled reactions. Instead, the conformationally rigid Pro was found to promote non-enzymatic peptide bond fragmentation when present in the N-terminus of the dipeptide and is representative of analytical methods such as mass spectrometry for protein analysis.[365, 546, 549, 550] As we previously specified, the "Proline effect" is promoted through the proton affinity observed by the amine group of the Pro ring, where it is likely that one of the H_3O^+ hydrogen atoms coordinates with the amine during the electrophilic attack of H_3O^+ on the peptide bond of Pro-Ybb dipeptides.[547-549] Therefore, it is likely that the observed "Proline effect" would counteract the rigidity of the residue, and promote fragmentation under the environmental conditions associated with the Protein A chromatography stage of downstream manufacturing. Additionally, both Gly and Leu were found to promote fragmentation when covalently bound to C-terminus Pro, while Ser-Pro was found to be kinetically unfavorable. The unfavorability of the Ser-Pro fragmentation is likely representative of the hydroxyl group of the N-terminus Ser residue being more attractive to the free hydrogen compared to the amine of the Pro ring. As a result, the rigidness of the Pro residue would reduce its favorability towards non-enzymatic peptide bond fragmentation under acidic conditions.

7.5. Conclusions and Future Directions

In summary, a theoretical study on non-enzymatic peptide bond fragmentation under acidic conditions was completed to analyze the impact of the degradation mechanism on the stability of the mAb hinge region during the Protein A chromatography stage of mAb downstream manufacturing. Rate coefficients and Arrhenius parameters for the stepwise reaction mechanism were computed for 360 total reactions (90 reactions per modeled dielectric) using B3LYP and statistical thermodynamics. The forward rate coefficient of the proposed rate-determining step of the reaction mechanism was benchmarked against two experimental studies to validate our computational methodology and proposed chemical model.[142, 144] Furthermore, the impact of electrostatic interactions between the ionized residues and the attacking H_3O^+ was found to exhibit similar behavior as the study associated with the alkaline wash stage of downstream manufacturing, where the attraction of opposing charges resulted in a lower E_A while reactions containing like charges were found to kinetically unfavorable in comparison.[503, 513, 514, 517, 518]

Furthermore, evaluation of the correlation between the parameters associated with the developed model indicated that the Evans-Polanyi correlation was not an appropriate generalization for the reaction class associated with non-enzymatic peptide bond fragmentation under acidic conditions. Additionally, negative and weakly linear correlation was observed between k^F, the associated reaction net charge, propensity of the C-terminus residue to occur within the hinge region, and functionality of the N-terminus residue sidechain.[371, 372] As a result, we analyzed reaction propensity towards fragmentation as a function of three parameters. Evaluation of the parameters that were correlated to the forward rate coefficient indicated that the electrostatic interactions exhibited by the ionized residues and the pH-driven environment were conserved across pH extremes, where favorability was dependent on electrostatic attraction and repulsion.[503, 513, 514, 517, 518] Furthermore, the frequency of residues occurring in the hinge region was likely representative of its propensity towards fragmentation, where reactions containing small, flexible residues yielded a larger k^F than the reactions with bulky sidechains.[365, 503, 513-516, 523, 524] Reaction favorability was also found to resemble analytical results, such as the "Proline effect", where the affinity of the ring amine promoted fragmentation.[547-549] Future studies will evaluate the fragmentation reaction energetics of larger polypeptides to analyze the impact of the previously noted parameters on reaction favorability. Additionally, the thermochemical and kinetic parameters computed by our developed database will be used to generate a robust microkinetic model of protein degradation.

Chapter 8: Conclusions and Future Directions

8.1. Overall Summary of Projects

In summary, databases of thermochemical and kinetic parameters have been developed to represent the reaction energetics associated with two mAb CQAs along the different stages of the therapeutic manufacturing pipeline. The generated databases are also representative of how degradation impacts the generated formulation, where the chemical degradation-induced mutations associated with functional degradation reduce the efficacy of the final product, and structural degradation is observed via fragmentation of the peptide bond to cause instability within the protein structure.[82, 141, 147] The propensity of the protein to be impacted by these forms of degradation was found to be dependent on the stage of the manufacturing pipeline, where fragmentation was more likely to occur than deamidation during downstream manufacturing and both deamidation and fragmentation occurred under ambient conditions.[82, 141, 147] Furthermore, the propensity of the succinimide reaction pathway to occur under ambient environmental conditions indicated that the screening of cations to inhibit the reaction pathway was only applicable to the specified reaction pathway. Additionally, by using dipeptides with different sidechains and four different dielectrics to model the analyzed reactions the impact of residue functionalities on reaction energetics at the primary structure was quantified.[82, 141, 147] Via this analysis, we determined how the propensity of residues towards degradation via the specified reaction was correlated to its frequency within the protein. Additionally, the inclusion of the four modeled dielectrics gave insights towards reaction behavior at different sites within the protein structure, as well as for freely floating polypeptides in solution.[82, 141, 147]

The studies associated with the storage of the generated formulation under ambient conditions identified the most probable reaction pathway of Asn-X and Gln-X deamidation, quantified why Asn-X deamidation was more readily observed in mAbs relative to Gln-X deamidation, generalized the energetics of non-enzymatic peptide bond fragmentation along a linear trendline, and analyzed the reaction behavior of the sidechain functionalities at different portions of the protein. Structure-property relationships were observed between Asn-Gly and the modeled cation of interest, where the interaction site, oxidation state, spin multiplicity, and type of coordination influenced the binding free energies and effectiveness of the cation as an additive against the succinimide reaction pathway. Modeling the perturbations in pH observed during the downstream manufacturing stage of the mAb manufacturing pipeline indicated that the Evans-Polanyi correlation could be applied to generalize the energetics associated with non-enzymatic peptide bond fragmentation under alkaline conditions, but not acidic conditions. Additionally, the propensity of the hinge region towards fragmentation was correlated to the types of residues found to occur most frequently within the region.

8.2. Advancements and Limitations in Regards to mAb Formulations and Manufacturing

As we previously noted, the databases that were developed within this study were generated to complement prior analytical insights towards the non-enzymatic deamidation and peptide bond fragmentation as a function of different points within the pharmaceutical protein manufacturing pipeline. Specifically, to screen the impact of both intrinsic factors, such as the interactions between the N and C-terminus residue sidechains, and extrinsic factors, such as the modeled pH, dielectric, and cation-dipeptide complexes. Via this screening, we have provided the following advancements towards mAb formulations and manufacturing. First, by developing the thermodynamic and kinetic databases of mAb degradation as a function of the manufacturing pipeline, the forward rate coefficients of the specified degradation reaction can be predicted as a function of temperature. Table 8.2.1. depicts a two-sample t-test for the difference in means between the forward rate coefficients computed by *CalcK* relative to the forward rate coefficients that were computed using E_A^F and $\log(A)^F$ from the developed databases as a function of temperature. Evaluation of the two-sample t-test indicated that the difference in means was statistically insignificant at each specified temperature, suggesting that the E_A^F and $\log(A)^F$ could be used to accurately predict $\ln(k)^F$ without the use of the *CalcK* script.[275] Therefore, the forward rate coefficients associated with the different degradation pathways analyzed within this document can be evaluated at both the temperature indicated by the experimental studies used to validate the databases and the temperatures associated with long-term storage of mAbs.

Table 8.2.1. Two-sample t-test for the difference in means between $\ln(k)^F$ [CalcK] and $\ln(k)^F$ [Using dataset parameters] as a function of temperature (K) and over four modeled dielectrics (ε).

T = 233 K				
Database	**Reaction**	**t-value**	**DOF**	**p-value**
Asn-X Deamidation (Storage pH)	DH	0.31	158	0.75
	SF	-0.04	150	0.97
	SHD	1.01	150	0.32
	SHI	0.95	150	0.34
Gln-X Deamidation (Storage pH)	DH	1.10	158	0.27
	GF	0.04	150	0.97
	GHE	0.92	150	0.36
	GHI	0.47	150	0.64
Fragmentation (Storage pH)	Intermediate Formation	0.64	166	0.53
T = 273 K				
Database	**Reaction**	**t-value**	**DOF**	**p-value**
Asn-X Deamidation (Storage pH)	DH	0.18	158	0.86
	SF	-0.02	150	0.98
	SHD	0.60	150	0.55
	SHI	0.58	150	0.57
Gln-X Deamidation (Storage pH)	DH	0.64	158	0.53
	GF	0.01	150	0.99
	GHE	0.56	150	0.58
	GHI	0.28	150	0.78
Fragmentation (Storage pH)	Intermediate Formation	0.39	166	0.70
T = 281 K				

Database	Reaction	t-value	DOF	p-value
Asn-X Deamidation (Storage pH)	DH	0.16	158	0.87
	SF	-0.02	150	0.98
	SHD	0.52	150	0.60
	SHI	0.51	150	0.61
Gln-X Deamidation (Storage pH)	DH	0.56	158	0.57
	GF	0.00	150	1.00
	GHE	0.49	150	0.63
	GHI	0.22	150	0.82
Fragmentation (Storage pH)	Intermediate Formation	0.34	166	0.74
T = 298 K				
Database	**Reaction**	**t-value**	**DOF**	**p-value**
Asn-X Deamidation (Storage pH)	DH	0.13	158	0.90
	SF	0.00	150	1.00
	SHD	0.40	150	0.69
	SHI	0.37	150	0.71
Gln-X Deamidation (Storage pH)	DH	0.43	158	0.67
	GF	0.02	150	0.98
	GHE	0.37	150	0.72
	GHI	0.18	150	0.86
Fragmentation (Storage pH)	Intermediate Formation	0.26	166	0.80
Fragmentation (Basic pH)	Intermediate Formation	0.45	718	0.65
Fragmentation (Acidic pH)		0.29	718	0.77

By computing the instantaneous selectivity between the primary and secondary reaction pathways of Asn-X and Gln-X deamidation, we identified the most probable reaction pathway under ambient pH as a function of the C-terminus residue that Asn and Gln are covalently bound to, as well as the overall trends observed in regards to the modeled dielectric. Via this analysis, deamidation hotspots can be predicted, and additives and stabilizers can be developed to specifically target the more competitive degradation reaction pathway. Similarly, by generalizing the activation barrier associated with the proposed rate-determining step of non-enzymatic peptide bond fragmentation under ambient conditions, the energetics of potential hotspots within the hinge region can be predicted as a function of the enthalpy of reaction.

Additionally, as we previously specified, the patent filed by Rehder evaluated the effectiveness of cation-based additives in reducing the amount of iso-Asp generated within the formulation.[104] By evaluating the impact of the cations on reducing the spontaneity of the succinimide formation reaction, we were able to obtain novel insights regarding how the site of interaction, type of cation-peptide coordination, and cation classification impacted the binding free energy and spontaneity of the succinimide formation reaction. Via these learnings, we screened the cations in terms of their effectiveness in reducing the spontaneity of the succinimide formation reaction, which in turn would reduce the formation of iso-Asp within the formulation. In our downstream manufacturing studies, we combined chemical engineering insights with statistical learning methods and bioinformatics to identify the factors that most significantly impacted the kinetic favorability of non-enzymatic peptide bond fragmentation under acidic and basic conditions. Through this analysis, we were able to apply the Evans-Polanyi correlation to generalize the proposed rate-determining activation barriers of the dipeptide reactions most likely to occur within the hinge region, and correlate amino acid frequency within the hinge region to the hinge region propensity towards fragmentation. These insights provide fundamental insights towards

In addition to the insights that these works have provided regarding the manufacturing and formulation of mAbs, there are also limitations. Specifically, due to the tradeoff between model accuracy and computational cost, we chose to model the impact of residue and environmental interacts through the simplest chemical model, the model dipeptide. By selecting the model dipeptide as our model structure to evaluate the different protein degradation reactions, we reduced the computational complexity of our models. However, by reducing the computational complexity of our models, there is a possibility that the predicted trends identified in our studies are not representative of the degradation behavior exhibited by the mAb. Prior experimental findings can be used to validate the quantitative and qualitative trends observed by the developed models, as well as the applied methodology. However, the model validation is limited to the environments and polypeptides directly cited in literature. For the models that predicted the dipeptides most prone towards degradation, such as the non-enzymatic peptide bond fragmentation under acidic conditions, forced degradation at the predicted hotspots would be necessary to evaluate the accuracy of the predicted trends. Additionally, for our developed Evans-Polanyi correlations, further validation of the linear generalization is dependent on the experimental value of the enthalpy of reaction for the proposed rate-determining step, which we were unable to find within our literature review. For this purpose, further validation of the trends analyzed and the databases generated within this document is necessary.

8.3. Future Directions

A three-pronged approach is necessary to expand the findings of this work into the next stage. As we previously noted, all degradation reactions analyzed in this study were modeled at using the smallest structure, the dipeptide. As a result, the impact of the sidechain interactions on reaction energetics needs to be evaluated in the secondary and tertiary structures. By modeling these interactions, the trends reaction selectivity and site-specific propensity observed at the primary structure can be correlated to trends observed at the secondary and tertiary structures.

Experimental validation of the kinetic parameters predicted by the studies in this work is necessary to determine the associated error of building a microkinetic model based on these values, as well as if a scaling factor is necessary for any of the reactions modeled within this work. Nuclear Magnetic Resonance (NMR) and Isotope Enrichment NMR (^{1}H, ^{13}C, ^{15}N, ^{19}F etc) spectroscopy are two widely used techniques for evaluating the rates of protein degradation, where the shifts in the observed peaks associated with the labeled molecule of interest are evaluated to determine the rate of degradation.[142, 551] A study by Hu et al. indicated that quantification of the parameters of interest is tedious without the presence of the labeled isotopes due to the NMR signal intensities exhibited by other molecules within the solution, where the specificity of the spectroscopy method towards identifying degradation at a site of interest within the polypeptide sequence would be reduced.[551] Furthermore, the application of transverse relaxation optimization spectroscopy (TROSY) and replacing the hydrogen atom with deuterium at the residue of interest improved the sensitivity of the spectroscopy towards larger polypeptides.[551-553]

For this purpose, further experimental validation would be accomplished via Isotope Enrichment NMR. Specifically, for each of the reaction mechanisms modeled in the study, the dipeptides that were computed to yield the smallest and largest rate coefficients associated with the degradation reaction of interest would be tagged to analyze the accuracy of our developed model. By tagging the dipeptides associated with the endpoints of the computed rate coefficients, the need for a scaling factor can be evaluated through the margin of error associated with the experimental and computed rate coefficients.

Once the kinetic parameters predicted by the developed models associated with each reaction pathway have been validated, or have been flagged for the inclusion of a scaling factor, the microkinetic model can be tuned to generate the time-scale profiles necessary to quantify the degradation behavior of deamidation and fragmentation under different environments. Development of an appropriate microkinetic model for this work would be achieved via the "method of moments modeling approach".[554-559] The method of moments is a deterministic modeling approach that is popular for modeling the kinetics associated polymerization reactions. [554-559] Prior studies on the kinetic modeling of polymerization reactions have highlighted the method of moments approach due to its simplicity, where its popularity attributed to model simplicity in regards to the derivation of model parameters, the diverse array of reactions the methodology can be applied to, and its generation of average properties to produce chain-specific and full distributions.[554]

The sequence of the protein at its primary structure can be treated as a polypeptide chain, so it is likely that the method of moments approach would be an appropriate approach towards the generation of timescale profiles of protein degradation. Figure 8.2.1 depicts the proposed process flow diagram that would be used to generate the microkinetic model of protein degradation via the method of moments modeling approach. The results generated in our prior studies have indicated that multiple factors influence the propensity of polypeptides towards degradation, ranging from the environmental conditions associated with the different stages of the mAb manufacturing pipeline to the interactions of residue sidechains with the environment and with one another. Therefore, the conditions associated with the degradation reaction of interest should be specified at model initiation.

Conditions such as pH, the propensity of residues to occur within the modeled polypeptide chain, and the impact of the sidechain functionality on the energetics associated with the degradation reaction of interest influence the identification of the pertinent reactions to include within the developed model. For example, evaluation of deamidation as a function of time under ambient conditions via the cyclic intermediate pathways would exclude reactions containing Pro in the C-terminus from consideration, while the stability of the cyclic intermediate under acidic conditions would indicate that the direct hydrolysis pathway would be the only reaction considered.[82, 147, 153, 168, 560-562] Once the pertinent reactions have been identified and written as elementary reactions, the population mass balances associated the degradation mechanism can be derived.[554, 555] Generation of the population mass balances is the first step to account for the distributions that have not been predicted through our atomistic modeling, and takes both the length and type of the modeled polypeptide chains into consideration.[554, 555] Once the mass balances have been derived, they are transformed into moment balances to generate the differential equations associated with computing the distribution of the model parameter of interest.[554] Calculation of the average parameter of interest would then be accomplished via numerical or analytical methods, depending on how the derived differential equations are able to be solved.[554]

Once those properties have been computed via the method of moments approach, parameters such as the Arrhenius pre-exponential factor, activation barrier, and enthalpy of reaction can be implemented to scale the kinetic parameters associated with degradation to the length of the polypeptide chain, and increase the complexity of the generated model.[554, 555] Scaling of the kinetic parameters to accurately reflect the physical rates of protein degradation will be accomplished through the iterative process of model validation against experimental results, where the number of parameters included within the model will be increased first to determine if model complexity increases the accuracy of the generated results. If the accuracy of the model is not impacted by increasing its complexity, a correction factor will be implemented to scale the generated reaction energetics so they accurately reflect experimental results.

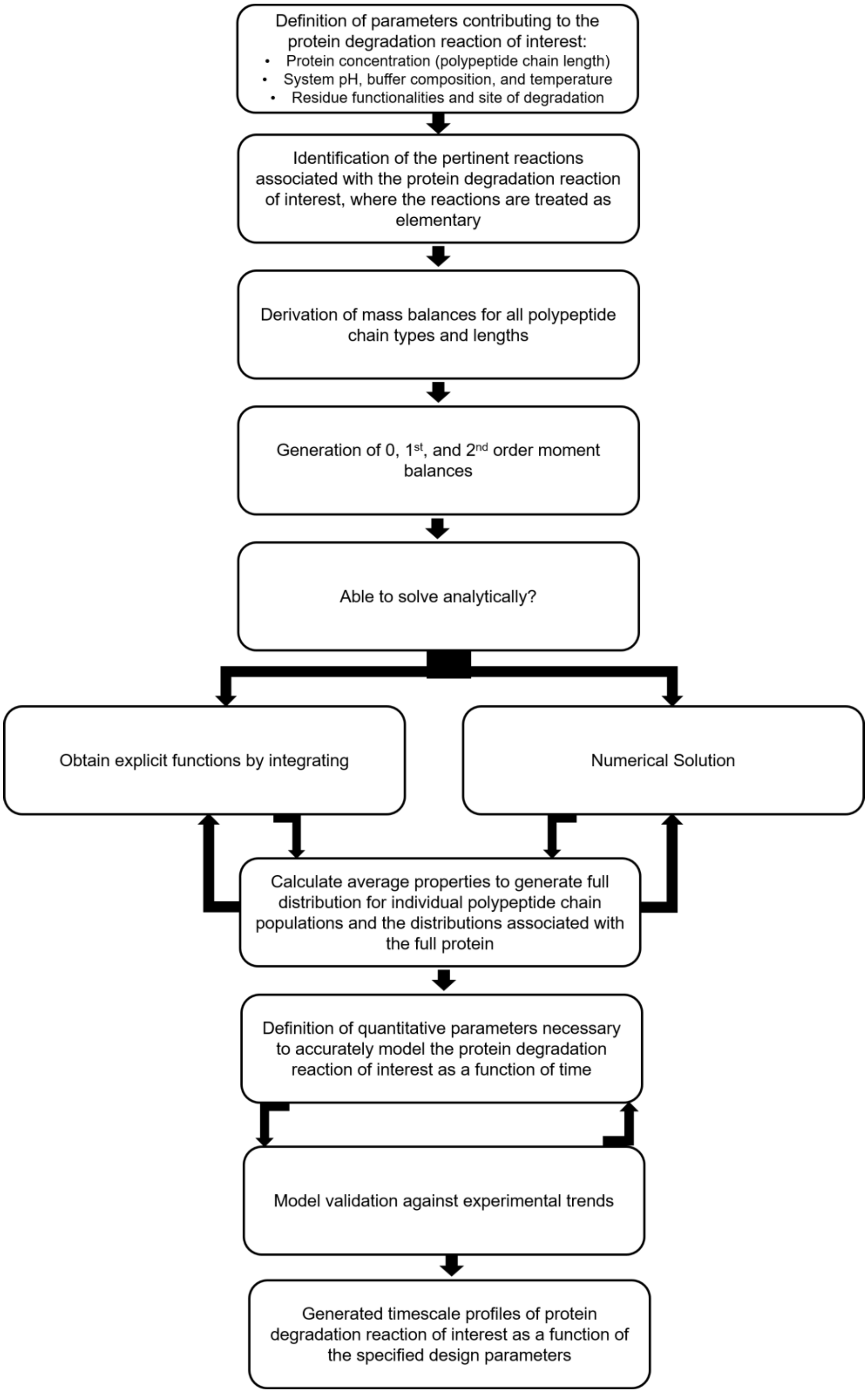

Figure 8.3.1. Proposed process flow diagram for the development of microkinetic models of protein degradation using the Methods of Moments Modeling Approach.[554, 555]

Appendix A: List of Dipeptides used to Map the Asn-X Deamidation Reaction Network during Storage (Chapter 2)

Table A.1. List of Dipeptides Used to Model Asn-X Direct Hydrolysis.

Dipeptide	Net Charge
Asparginyl-Alanine (Asn-Ala)	0
Asparaginyl-Arginine (Asn-Arg)	1
Asparaginyl-Asparagine (Asn-Asn)	0
Asparaginyl-Aspartic Acid (Asn-Asp)	-1
Asparaginyl-Cysteine (Asn-Cys)	0
Asparaginyl-Glutamine (Asn-Gln)	0
Asparginyl-Glutamic Acid (Asn-Glu)	-1
Asparaginyl-Glycine (Asn-Gly)	0
Asparaginyl-Histidine (Asn-His)	0
Asparaginyl-Isoleucine (Asn-Ile)	0
Asparaginyl-Leucine (Asn-Leu)	0
Asparaginyl-Lysine (Asn-Lys)	1
Asparaginyl-Methionine (Asn-Met)	0
Asparaginyl-Phenylalanine (Asn-Phe)	0
Asparaginyl-Proline (Asn-Pro)	0
Asparaginyl-Serine (Asn-Ser)	0
Asparaginyl-Threonine (Asn-Thr)	0
Asparaginyl-Tryptophan (Asn-Trp)	0
Asparaginyl-Tyrosine (Asn-Tyr)	0
Asparaginyl-Valine (Asn-Val)	0

Table A.2. List of Dipeptides Used to Model Asn-X Succinimide Formation, Succinimide Hydrolysis into Asp-X, and Succinimide Hydrolysis into iso-Asp-X.

Dipeptide	Net Charge
Asparginyl-Alanine (Asn-Ala)	0
Asparaginyl-Arginine (Asn-Arg)	1
Asparaginyl-Asparagine (Asn-Asn)	0
Asparaginyl-Aspartic Acid (Asn-Asp)	-1
Asparaginyl-Cysteine (Asn-Cys)	0
Asparaginyl-Glutamine (Asn-Gln)	0
Asparginyl-Glutamic Acid (Asn-Glu)	-1
Asparaginyl-Glycine (Asn-Gly)	0
Asparaginyl-Histidine (Asn-His)	0
Asparaginyl-Isoleucine (Asn-Ile)	0
Asparaginyl-Leucine (Asn-Leu)	0
Asparaginyl-Lysine (Asn-Lys)	1
Asparaginyl-Methionine (Asn-Met)	0
Asparaginyl-Phenylalanine (Asn-Phe)	0
Asparaginyl-Serine (Asn-Ser)	0
Asparaginyl-Threonine (Asn-Thr)	0
Asparaginyl-Tryptophan (Asn-Trp)	0
Asparaginyl-Tyrosine (Asn-Tyr)	0
Asparaginyl-Valine (Asn-Val)	0

Appendix B: List of Dipeptides used to Map the Gln-X Deamidation Reaction Network during Storage (Chapter 3)

Table B.1. List of Dipeptides Used to Model Gln-X Direct Hydrolysis.

Dipeptide	Net Charge
Glutaminyl-Alanine (Gln-Ala)	0
Glutaminyl-Arginine (Gln-Arg)	1
Glutaminyl-Asparagine (Gln-Asn)	0
Glutaminyl-Aspartic Acid (Gln-Asp)	-1
Glutaminyl-Cysteine (Gln-Cys)	0
Glutaminyl-Glutamine (Gln-Gln)	0
Glutaminyl-Glutamic Acid (Gln-Glu)	-1
Glutaminyl-Glycine (Gln-Gly)	0
Glutaminyl-Histidine (Gln-His)	0
Glutaminyl-Isoleucine (Gln-Ile)	0
Glutaminyl-Leucine (Gln-Leu)	0
Glutaminyl-Lysine (Gln-Lys)	1
Glutaminyl-Methionine (Gln-Met)	0
Glutaminyl-Phenylalanine (Gln-Phe)	0
Glutaminyl-Proline (Gln-Pro)	0
Glutaminyl-Serine (Gln-Ser)	0
Glutaminyl-Threonine (Gln-Thr)	0
Glutaminyl-Tryptophan (Gln-Trp)	0
Glutaminyl-Tyrosine (Gln-Tyr)	0
Glutaminyl-Valine (Gln-Val)	0

Table B.2. List of Dipeptides Used to Model Gln-X Glutarimide Formation, Glutarimide Hydrolysis into Glu-X, and Glutarimide Hydrolysis into iso-Glu-X.

Dipeptide	Net Charge
Glutaminyl-Alanine (Gln-Ala)	0
Glutaminyl-Arginine (Gln-Arg)	1
Glutaminyl-Asparagine (Gln-Asn)	0
Glutaminyl-Aspartic Acid (Gln-Asp)	-1
Glutaminyl-Cysteine (Gln-Cys)	0
Glutaminyl-Glutamine (Gln-Gln)	0
Glutaminyl-Glutamic Acid (Gln-Glu)	-1
Glutaminyl-Glycine (Gln-Gly)	0
Glutaminyl-Histidine (Gln-His)	0
Glutaminyl-Isoleucine (Gln-Ile)	0
Glutaminyl-Leucine (Gln-Leu)	0
Glutaminyl-Lysine (Gln-Lys)	1
Glutaminyl-Methionine (Gln-Met)	0
Glutaminyl-Phenylalanine (Gln-Phe)	0
Glutaminyl-Serine (Gln-Ser)	0
Glutaminyl-Threonine (Gln-Thr)	0
Glutaminyl-Tryptophan (Gln-Trp)	0
Glutaminyl-Tyrosine (Gln-Tyr)	0
Glutaminyl-Valine (Gln-Val)	0

Appendix C: List of Dipeptides used to Model Non-Enzymatic Dipeptide Fragmentation during Storage (Chapter 4)

Table C.1. List of Dipeptides Used to Model Non-Enzymatic Peptide Bond Fragmentation**.**

Dipeptide	Net Charge
Alanyl-Lysine (Ala-Lys)	1
Alanyl-Serine (Ala-Ser)	0
Alanyl-Threonine (Ala-Thr)	0
Alanyl-Tyrosine (Ala-Tyr)	0
Arginyl-Aspartic Acid (Arg-Asp)	0
Arginyl-Lysine (Arg-Lys)	2
Aspartyl-Alanine (Asp-Ala)	-1
Aspartyl-Arginine (Asp-Arg)	0
Aspartyl-Glutamic Acid (Asp-Glu)	-2
Aspartyl-Serine (Asp-Ser)	-1
Aspartyl-Threonine (Asp-Thr)	-1
Glutamyl-Aspartic Acid (Glu-Asp)	-2
Glutamyl-Glycine (Glu-Gly)	-1
Glycyl-Glycine (Gly-Gly)	0
Glycyl-Proline (Gly-Pro)	0
Glycyl-Tyrosine (Gly-Tyr)	0
Prolyl-Arginine (Pro-Arg)	1
Prolyl-Glycine (Pro-Gly)	0
Tyrosyl-Glycine (Tyr-Gly)	0
Tyrosyl-Serine (Tyr-Ser)	0
Valyl-Alanine (Val-Ala)	0

Appendix D: List of Dipeptides used to Model Non-Enzymatic Dipeptide Fragmentation during Alkaline Wash Stage of Downstream Manufacturing (Chapter 6)

Table D.1. List of Dipeptides Used to Model Non-Enzymatic Peptide Bond Fragmentation.

Dipeptide	Net Charge
Alanyl-Alanine (Ala-Ala)	0
Alanyl-Aspartic Acid (Ala-Asp)	-1
Alanyl-Glycine (Ala-Gly)	0
Alanyl-Leucine (Ala-Leu)	0
Alanyl-Lysine (Ala-Lys)	1
Alanyl-Lysine (Ala-Lys)	0
Alanyl-Proline (Ala-Pro)	0
Alanyl-Serine (Ala-Ser)	0
Alanyl-Threonine (Ala-Thr)	0
Alanyl-Valine (Ala-Val)	0
Aspartyl-Alanine (Asp-Ala)	-1
Aspartyl-Aspartic Acid (Asp-Asp)	-2
Aspartyl-Glycine (Asp-Gly)	-1
Aspartyl-Leucine (Asp-Leu)	-1
Aspartyl-Lysine (Asp-Lys)	0
Aspartyl-Lysine (Asp-Lys)	-1
Aspartyl-Proline (Asp-Pro)	-1
Aspartyl-Serine (Asp-Ser)	-1
Aspartyl-Threonine (Asp-Thr)	-1
Aspartyl-Valine (Asp-Val)	-1
Glycyl-Alanine (Gly-Ala)	0
Glycyl-Aspartic Acid (Gly-Asp)	-1
Glycyl-Glycine (Gly-Gly)	0
Glycyl-Leucine (Gly-Leu)	0
Glycyl-Lysine (Gly-Lys)	1
Glycyl-Lysine (Gly-Lys)	0
Glycyl-Proline (Gly-Pro)	0
Glycyl-Serine (Gly-Ser)	0
Glycyl-Threonine (Gly-Thr)	0
Glycyl-Valine (Gly-Val)	0
Leucyl-Alanine (Leu-Ala)	0
Leucyl-Aspartic Acid (Leu-Asp)	-1
Leucyl-Glycine (Leu-Gly)	0
Leucyl-Leucine (Leu-Leu)	0
Leucyl-Lysine (Leu-Lys)	1
Leucyl-Lysine (Leu-Lys)	0
Leucyl-Proline (Leu-Pro)	0

Leucyl-Serine (Leu-Ser)	0
Leucyl-Threonine (Leu-Thr)	0
Leucyl-Valine (Leu-Val)	0
Lysyl-Alanine (Lys-Ala)	1
Lysyl-Aspartic Acid (Lys-Asp)	0
Lysyl-Aspartic Acid (Lys-Asp)	-1
Lysyl-Glycine (Lys-Gly)	1
Lysyl-Leucine (Lys-Leu)	1
Lysyl-Lysine (Lys-Lys)	1
Lysyl-Lysine (Lys-Lys)	0
Lysyl-Proline (Lys-Pro)	1
Lysyl-Proline (Lys-Pro)	0
Lysyl-Serine (Lys-Ser)	1
Lysyl-Threonine (Lys-Thr)	1
Lysyl-Threonine (Lys-Thr)	0
Lysyl-Valine (Lys-Val)	1
Prolyl-Alanine (Pro-Ala)	0
Prolyl-Aspartic Acid (Pro-Asp)	-1
Prolyl-Glycine (Pro-Gly)	0
Prolyl-Leucine (Pro-Leu)	0
Prolyl-Lysine (Pro-Lys)	1
Prolyl-Proline (Pro-Pro)	0
Prolyl-Serine (Pro-Ser)	0
Prolyl-Threonine (Pro-Thr)	0
Prolyl-Valine (Pro-Val)	0
Seryl-Alanine (Ser-Ala)	0
Seryl-Aspartic Acid (Ser-Asp)	-1
Seryl-Glycine (Ser-Gly)	0
Seryl-Leucine (Ser-Leu)	0
Seryl-Lysine (Ser-Lys)	1
Seryl-Lysine (Ser-Lys)	0
Seryl-Proline (Ser-Pro)	0
Seryl-Serine (Ser-Ser)	0
Seryl-Threonine (Ser-Thr)	0
Seryl-Valine (Ser-Val)	0
Threonyl-Alanine (Thr-Ala)	0
Threonyl-Aspartic Acid (Thr-Asp)	-1
Threonyl-Glycine (Thr-Gly)	0
Threonyl-Leucine (Thr-Leu)	0
Threonyl-Lysine (Thr-Lys)	1
Threonyl-Proline (Thr-Pro)	0
Threonyl-Serine (Thr-Ser)	0
Threonyl-Threonine (Thr-Thr)	0
Threonyl-Valine (Thr-Val)	0
Valyl-Alanine (Val-Ala)	0

Valyl-Aspartic Acid (Val-Asp)	-1
Valyl-Glycine (Val-Gly)	0
Valyl-Leucine (Val-Leu)	0
Valyl-Lysine (Val-Lys)	1
Valyl-Proline (Val-Pro)	0
Valyl-Serine (Val-Ser)	0
Valyl-Threonine (Val-Thr)	0
Valyl-Valine (Val-Val)	0

Appendix E: List of Dipeptides used to Model Non-Enzymatic Dipeptide Fragmentation during the Protein A Chromatography Stage of Downstream Manufacturing (Chapter 7)

Table E.1. List of Dipeptides Used to Model Non-Enzymatic Peptide Bond Fragmentation.

Dipeptide	Net Charge
Alanyl-Alanine (Ala-Ala)	0
Alanyl-Aspartic Acid (Ala-Asp)	0
Alanyl-Glycine (Ala-Gly)	0
Alanyl-Leucine (Ala-Leu)	0
Alanyl-Lysine (Ala-Lys)	1
Alanyl-Proline (Ala-Pro)	0
Alanyl-Serine (Ala-Ser)	0
Alanyl-Threonine (Ala-Thr)	0
Alanyl-Valine (Ala-Val)	0
Aspartyl-Alanine (Asp-Ala)	0
Aspartyl-Alanine (Asp-Ala)	-1
Aspartyl-Aspartic Acid (Asp-Asp)	0
Aspartyl-Aspartic Acid (Asp-Asp)	-2
Aspartyl-Glycine (Asp-Gly)	0
Aspartyl-Glycine (Asp-Gly)	-1
Aspartyl-Leucine (Asp-Leu)	0
Aspartyl-Leucine (Asp-Leu)	-1
Aspartyl-Lysine (Asp-Lys)	1
Aspartyl-Lysine (Asp-Lys)	0
Aspartyl-Proline (Asp-Pro)	0
Aspartyl-Serine (Asp-Ser)	0
Aspartyl-Serine (Asp-Ser)	-1
Aspartyl-Threonine (Asp-Thr)	0
Aspartyl-Valine (Asp-Val)	0
Glycyl-Alanine (Gly-Ala)	0
Glycyl-Aspartic Acid (Gly-Asp)	0
Glycyl-Aspartic Acid (Gly-Asp)	-1
Glycyl-Glycine (Gly-Gly)	0
Glycyl-Leucine (Gly-Leu)	0
Glycyl-Lysine (Gly-Lys)	1
Glycyl-Proline (Gly-Pro)	0
Glycyl-Serine (Gly-Ser)	0
Glycyl-Threonine (Gly-Thr)	0
Glycyl-Valine (Gly-Val)	0
Leucyl-Alanine (Leu-Ala)	0

Leucyl-Aspartic Acid (Leu-Asp)	0
Leucyl-Glycine (Leu-Gly)	0
Leucyl-Leucine (Leu-Leu)	0
Leucyl-Lysine (Leu-Lys)	1
Leucyl-Proline (Leu-Pro)	0
Leucyl-Serine (Leu-Ser)	0
Leucyl-Threonine (Leu-Thr)	0
Leucyl-Valine (Leu-Val)	0
Lysyl-Alanine (Lys-Ala)	1
Lysyl-Aspartic Acid (Lys-Asp)	1
Lysyl-Aspartic Acid (Lys-Asp)	0
Lysyl-Glycine (Lys-Gly)	1
Lysyl-Leucine (Lys-Leu)	1
Lysyl-Lysine (Lys-Lys)	1
Lysyl-Proline (Lys-Pro)	1
Lysyl-Serine (Lys-Ser)	1
Lysyl-Threonine (Lys-Thr)	1
Lysyl-Valine (Lys-Val)	1
Prolyl-Alanine (Pro-Ala)	0
Prolyl-Aspartic Acid (Pro-Asp)	0
Prolyl-Aspartic Acid (Pro-Asp)	-1
Prolyl-Glycine (Pro-Gly)	0
Prolyl-Leucine (Pro-Leu)	0
Prolyl-Lysine (Pro-Lys)	1
Prolyl-Proline (Pro-Pro)	0
Prolyl-Serine (Pro-Ser)	0
Prolyl-Threonine (Pro-Thr)	0
Prolyl-Valine (Pro-Val)	0
Seryl-Alanine (Ser-Ala)	0
Seryl-Aspartic Acid (Ser-Asp)	0
Seryl-Glycine (Ser-Gly)	0
Seryl-Leucine (Ser-Leu)	0
Seryl-Lysine (Ser-Lys)	1
Seryl-Proline (Ser-Pro)	0
Seryl-Serine (Ser-Ser)	0
Seryl-Threonine (Ser-Thr)	0
Seryl-Valine (Ser-Val)	0
Threonyl-Alanine (Thr-Ala)	0
Threonyl-Aspartic Acid (Thr-Asp)	0
Threonyl-Glycine (Thr-Gly)	0
Threonyl-Leucine (Thr-Leu)	0
Threonyl-Lysine (Thr-Lys)	1

Threonyl-Proline (Thr-Pro)	0
Threonyl-Serine (Thr-Ser)	0
Threonyl-Threonine (Thr-Thr)	0
Threonyl-Valine (Thr-Val)	0
Valyl-Alanine (Val-Ala)	0
Valyl-Aspartic Acid (Val-Asp)	0
Valyl-Glycine (Val-Gly)	0
Valyl-Leucine (Val-Leu)	0
Valyl-Lysine (Val-Lys)	1
Valyl-Proline (Val-Pro)	0
Valyl-Serine (Val-Ser)	0
Valyl-Threonine (Val-Thr)	0
Valyl-Valine (Val-Val)	0

References

[1] M. Akbarian, S.-H. Chen, Instability Challenges and Stabilization Strategies of Pharmaceutical Proteins, Pharmaceutics, 2022.
[2] B.G. de la Torre, F. Albericio, The Pharmaceutical Industry in 2022: An Analysis of FDA Drug Approvals from the Perspective of Molecules, Molecules, 2023.
[3] Novel Drug Approvals for 2022, US Food and Drug Administration, https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2022, 2022.
[4] A.L. Grilo, A. Mantalaris, The Increasingly Human and Profitable Monoclonal Antibody Market, Trends in Biotechnology 37 (2019) 9-16.
[5] M.A. Van Dijk, J.G.J. Van De Winkel, Human antibodies as next generation therapeutics, Current opinion in chemical biology 5 (2001) 368-374.
[6] G.P. Adams, L.M. Weiner, Monoclonal antibody therapy of cancer, Nature biotechnology 23 (2005) 1147-1157.
[7] K. Tsumoto, Y. Isozaki, H. Yagami, M. Tomita, Future perspectives of therapeutic monoclonal antibodies, Immunotherapy 11 (2019) 119-127.
[8] A. Pavlou, J. Reichert, Monoclonal antibodies market, Nature Reviews Drug Discovery 3 (2004) 383-384.
[9] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market, mAbs 7 (2015) 9-14.
[10] M. Vázquez-Rey, D.A. Lang, Aggregates in monoclonal antibody manufacturing processes, Biotechnology and bioengineering 108 (2011) 1494-1508.
[11] F.C. Breedveld, Therapeutic monoclonal antibodies, The Lancet 355 (2000) 735-740.
[12] Novel Drug Approvals for 2021, Food and Drug Administration, https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2021, 2021.
[13] Novel Drug Approvals for 2020, Food and Drug Administration, https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2020, 2020.
[14] Novel Drug Approvals for 2019, Food and Drug Administration, https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2019, 2019.
[15] Novel Drug Approvals for 2023, Food and Drug Administration (FDA), https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2023, 2023.
[16] History of the FDA, in: F.a.D.A. (FDA) (Ed.), 2018.
[17] History of the EMA, European Medicines Agency, 2019.
[18] How the committeees work, in: E.M.A. (EMA) (Ed.) Committees, 2021.
[19] Paediatric Committee (PDCO) Committees, European Medicines Agency (EMA), 2008.
[20] Committee for Advanced Therapies (CAT) Committees, European Medicines Agency (EMA), 2009.
[21] Committee on Herbal Medicinal Products (HMPC), Committees, European Medicines Agency (EMA), 2009.
[22] Committee for Orphan Medicinal Products (COMP) Committees, European Medicines Agency (EMA), 2007.

[23] Committee for Veterinary Medicinal Products (CVMP), Committees, European Medicines Agency (EMA), 2022.
[24] Pharmacovigilance Risk Assessment Committee (PRAC), Committees, European Medicines Agency (EMA), 2013.
[25] Committee for Medicinal Products for Human Use (CHMP) Committees, European Medicines Agency (EMA), 2007.
[26] How the committeees work, Committees, European Medicines Agency (EMA), 2013.
[27] M.S. Epstein, E.D. Ehrenpreis, P.M. Kulkarni, F.D.A.R.M.C.o.t.A.C.o.G. the, Biosimilars: The Need, The Challenge, The Future: The FDA Perspective, Official journal of the American College of Gastroenterology | ACG 109 (2014).
[28] B.R. Hirsch, G.H. Lyman, Biosimilars: are they ready for primetime in the United States?, Journal of the National Comprehensive Cancer Network 9 (2011) 934-943.
[29] J.M. Hoffman, E. Li, J.G. Stevenson, Preparing for biosimilars: Scientific, regulatory, and practice management issues for pharmacists. Live webcast.
[30] M. McCamish, G. Woollett, The state of the art in the development of biosimilars, Clinical Pharmacology & Therapeutics 91 (2012) 405-417.
[31] X. Geng, X. Kong, H. Hu, J. Chen, F. Yang, H. Liang, X. Chen, Y. Hu, Research and development of therapeutic mAbs: An analysis based on pipeline projects, Hum Vaccin Immunother 11 (2015) 2769-2776.
[32] J.P. Blackstone Ea Fau - Fuhr, Jr., J.P. Fuhr, Jr., Innovation and Competition: Will Biosimilars Succeed?: The creation of an FDA approval pathway for biosimilars is complex and fraught with hazard. Yes, innovation and market competition are at stake. But so are efficacy and patient safety.
[33] B. Calo-Fernández, J.L. Martínez-Hurtado, Biosimilars: company strategies to capture value from the biologics market.
[34] L.A. Christl, J. Woodcock, S. Kozlowski, Biosimilars: the US regulatory framework, Annual review of medicine 68 (2017) 243-254.
[35] S. Barlas, President obama reopens debate on patented biologics: will the FDA quickly define an abbreviated pathway for biosimilars?
[36] R. Isbrucker, WHO Guideline for the safe production and quality control of monoclonal antibodies for use in humans, in: W.H. Organization (Ed.), Department of Health Products Policy and Standards, Geneva, Switzerland, 2021, pp. 55.
[37] Guideline on development, production, characterization, and specification for monoclonal antibodies and related products, in: C.f.m.p.f.h.u. (CHMP) (Ed.), European Medicines Agency, United Kingdom, 2016, pp. 13.
[38] FOR THE SUBMISSION OF CHEMISTRY, MANUFACTURING, AND CONTROLS INFORMATION FOR A THERAPEUTIC RECOMBINANT DNA-DERIVED PRODUCT OR A MONOCLONAL ANTIBODY PRODUCT FOR IN VIVO USE, in: C.f.B.E.a.R.C.C.f.D.E.a.R. (CDER) (Ed.), Food and Drug Administration, Rockville, Maryland, 1996, pp. 22.
[39] Z. Chen, A.-P. Zeng, Protein engineering approaches to chemical biotechnology, Current Opinion in Biotechnology 42 (2016) 198-205.
[40] P.A. Romero, F.H. Arnold, Exploring protein fitness landscapes by directed evolution, Nat Rev Mol Cell Biol 10 (2009) 866-876.
[41] D. Könning, H. Kolmar, Beyond antibody engineering: directed evolution of alternative binding scaffolds and enzymes using yeast surface display, Microbial Cell Factories 17 (2018) 32.
[42] W. Xiong, B. Liu, Y. Shen, K. Jing, T.R. Savage, Protein engineering design from directed evolution to de novo synthesis, Biochemical Engineering Journal 174 (2021) 108096.

[43] H.W. Hellinga, Rational protein design: Combining theory and experiment, Proceedings of the National Academy of Sciences 94 (1997) 10015-10017.
[44] I.V. Korendovych, Rational and Semirational Protein Design.
[45] W.M. Dawson, G.G. Rhys, D.N. Woolfson, Towards functional de novo designed proteins, Current Opinion in Chemical Biology 52 (2019) 102-111.
[46] P. Gronemeyer, R. Ditz, J. Strube, Trends in Upstream and Downstream Process Development for Antibody Manufacturing, Bioengineering, 2014, pp. 188-212.
[47] Y. Rajendra, D. Kiseljak, L. Baldi, D.L. Hacker, F.M. Wurm, A simple high-yielding process for transient gene expression in CHO cells, Journal of Biotechnology 153 (2011) 22-26.
[48] P. Girard, M. Derouazi, G. Baumgartner, M. Bourgeois, M. Jordan, B. Jacko, F.M. Wurm, 100-liter transient transfection, Cytotechnology 38 (2002) 15-21.
[49] B. Kelley, Industrialization of mAb production technology: The bioprocessing industry at a crossroads, mAbs 1 (2009) 443-452.
[50] A. Rita Costa, M. Elisa Rodrigues, M. Henriques, J. Azeredo, R. Oliveira, Guidelines to cell engineering for monoclonal antibody production, European Journal of Pharmaceutics and Biopharmaceutics 74 (2010) 127-138.
[51] W.-J. Xu, Y. Lin, C.-L. Mi, J.-Y. Pang, T.-Y. Wang, Progress in fed-batch culture for recombinant protein production in CHO cells, Applied Microbiology and Biotechnology 107 (2023) 1063-1075.
[52] H.F. Liu, J. Ma, C. Winter, R. Bayer, Recovery and purification process development for monoclonal antibody production, mAbs 2 (2010) 480-499.
[53] L. Blaas, M. Musteanu, R. Eferl, A. Bauer, E. Casanova, Bacterial artificial chromosomes improve recombinant protein production in mammalian cells, BMC Biotechnology 9 (2009) 3.
[54] F. Li, N. Vijayasankaran, A. Shen, R. Kiss, A. Amanullah, Cell culture processes for monoclonal antibody production, mAbs 2 (2010) 466-479.
[55] H. Narayanan, M. Sponchioni, M. Morbidelli, Integration and digitalization in the manufacturing of therapeutic proteins, Chemical Engineering Science 248 (2022) 117159.
[56] M. Johnson, B. Lanthier S Fau - Massie, G. Massie B Fau - Lefebvre, A.A. Lefebvre G Fau - Kamen, A.A. Kamen, Use of the Centritech Lab centrifuge for perfusion culture of hybridoma cells in protein-free medium.
[57] J.-M. Bielser, M. Wolf, J. Souquet, H. Broly, M. Morbidelli, Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review, Biotechnology Advances 36 (2018) 1328-1340.
[58] Z. Xing, B.M. Kenty, Z.J. Li, S.S. Lee, Scale-up analysis for a CHO cell culture process in large-scale bioreactors, Biotechnology and bioengineering 103 (2009) 733-746.
[59] A.S. Tait, M.L. Tarrant Rd Fau - Velez-Suberbie, D.I.R. Velez-Suberbie Ml Fau - Spencer, D.G. Spencer Di Fau - Bracewell, D.G. Bracewell, Differential response in downstream processing of CHO cells grown under mild hypothermic conditions.
[60] A.S. Tait, C.E.M. Hogwood, C.M. Smales, D.G. Bracewell, Host cell protein dynamics in the supernatant of a mAb producing CHO cell line, Biotechnology and Bioengineering 109 (2012) 971-982.
[61] K. Kaneyoshi, K. Kuroda, K. Uchiyama, M. Onitsuka, N. Yamano-Adachi, Y. Koga, T.A.-O. Omasa, Secretion analysis of intracellular "difficult-to-express" immunoglobulin G (IgG) in Chinese hamster ovary (CHO) cells.
[62] D.L. Hacker, F.M. Wurm, 2.29 - Recombinant Technology, in: M. Moo-Young (Ed.), Comprehensive Biotechnology (Second Edition), Academic Press, Burlington, 2011, pp. 401-406.

[63] K. Gupta, M. Parasnis, R. Jain, P. Dandekar, Vector-related stratagems for enhanced monoclonal antibody production in mammalian cells, Biotechnology Advances 37 (2019) 107415.
[64] M.A. Firer, Efficient elution of functional proteins in affinity chromatography, Journal of biochemical and biophysical methods 49 (2001) 433-442.
[65] V. Amritkar, S. Adat, V. Tejwani, A. Rathore, R. Bhambure, Engineering Staphylococcal Protein A for high-throughput affinity purification of monoclonal antibodies, Biotechnology Advances 44 (2020) 107632.
[66] A.A. Shukla, B. Hubbard, T. Tressel, S. Guhan, D. Low, Downstream processing of monoclonal antibodies—application of platform approaches, Journal of Chromatography B 848 (2007) 28-39.
[67] J. Wang, N.E. Jaffe, K. Patel, Use of alkaline washes during chromatography to remove impurities, in: W.I.P.O. (WIPO) (Ed.) World Intellectual Property Organization (WIPO), Bristol-Meyers Squib, Singapore, 2016.
[68] M. Yamanouchi, Imura, Y., Tagawa, T., Nonoyama, S. & Fujino, Y, Method for producing protein having Fc region by alkaline washing., in: W.I.P. Organization (Ed.) World Intellectual Property Organization, Genetic, Inc, France & Japan, 2016, pp. 47.
[69] Q5A Viral Safety Evaluation of Biotechnology Products Derived From Cell Lines of Human or Animal Origin, in: U.D.o.H.a.H. Services (Ed.), Food and Drug Administration, Center for Biologics Evaluation and Research (CBER), Rockville, MD, 1998, pp. 30.
[70] C. Gillespie, M. Holstein, L. Mullin, K. Cotoni, R. Tuccelli, J. Caulmare, P. Greenhalgh, Continuous In-Line Virus Inactivation for Next Generation Bioprocessing, Biotechnology Journal 14 (2019) 1700718.
[71] Y. Brodsky, C. Zhang, Y. Yigzaw, G. Vedantham, Caprylic acid precipitation method for impurity reduction: an alternative to conventional chromatography for monoclonal antibody purification, Biotechnology and Bioengineering 109 (2012) 2589-2598.
[72] A. Trapp, A. Faude, N. Hörold, S. Schubert, S. Faust, T. Grob, S. Schmidt, Multiple functions of caprylic acid-induced impurity precipitation for process intensification in monoclonal antibody purification, Journal of Biotechnology 279 (2018) 13-21.
[73] W. Lebing, K.M. Remington, C. Schreiner, H.I. Paul, Properties of a new intravenous immunoglobulin (IGIV-C, 10%) produced by virus inactivation with caprylate and column chromatography, Vox Sanguinis 84 (2003) 193-201.
[74] V. Morais, H. Massaldi, A model mechanism for protein precipitation by caprylic acid: Application to plasma purification, Biotechnology and Applied Biochemistry 59 (2012) 50-54.
[75] P.D. Boyer, F.G. Lum, G.A. Ballou, J.M. Luck, R.G. Rice, THE COMBINATION OF FATTY ACIDS AND RELATED COMPOUNDS WITH SERUM ALBUMIN: I. STABILIZATION AGAINST HEAT DENATURATION, Journal of Biological Chemistry 162 (1946) 181-198.
[76] T. Burnouf, E. Griffiths, A. Padilla, S. Seddik, M.A. Stephano, J.-M. Gutiérrez, Assessment of the viral safety of antivenoms fractionated from equine plasma, Biologicals 32 (2004) 115-128.
[77] P. Kheddo, M. Tracka, J. Armer, R.J. Dearman, S. Uddin, C.F. van der Walle, A.P. Golovanov, The effect of arginine glutamate on the stability of monoclonal antibodies in solution.
[78] S. Haindl, J. Stark, J. Dippel, S. Handt, A. Reiche, Scale-up of Microfiltration Processes, Chemie Ingenieur Technik 92 (2020) 746-758.
[79] N.K. Tripathi, A. Shrivastava, Recent developments in bioprocessing of recombinant proteins: expression hosts and process development, Frontiers in Bioengineering and Biotechnology 7 (2019) 420.

[80] N.E. Robinson, A.B. Robinson, Deamidation of human proteins, Proceedings of the National Academy of Sciences 98 (2001) 12409-12413.
[81] L. Wang, J. Dembecki, N.E. Jaffe, B.W. O'Mara, H. Cai, C.N. Sparks, J. Zhang, S.G. Laino, R.J. Russell, M. Wang, A safe, effective, and facility compatible cleaning in place procedure for affinity resin in large-scale monoclonal antibody purification, Journal of Chromatography A 1308 (2013) 86-95.
[82] K.E. Lawson, J.K. Dekle, M.N. Evans, A.J. Adamczyk, Deamidation reaction network mapping of pharmacologic and related proteins: impact of solvation dielectric on the degradation energetics of asparagine dipeptides, Reaction Chemistry & Engineering (2022).
[83] J.F. Carpenter, M.J. Pikal, B.S. Chang, T.W. Randolph, Rational Design of Stable Lyophilized Protein Formulations: Some Practical Advice, Pharmaceutical Research 14 (1997) 969-975.
[84] A.A.M. Morales, G. Núñez-Gandolff, N.P. Pérez, B.C. Véliz, I. Caballero-Torres, J. Ducongé, E. Fernández, F.Z. Crespo, A. Veloso, N. Iznaga-Escobar, Freeze-dried formulation for direct 99mTc-labeling ior-egf/r3 MAb: additives, biodistribution, and stability, Nuclear Medicine and Biology 26 (1999) 717-723.
[85] H. Svilenov, L. Gentiluomo, W. Friess, D. Roessner, G. Winter, A New Approach to Study the Physical Stability of Monoclonal Antibody Formulations—Dilution From a Denaturant, Journal of Pharmaceutical Sciences 107 (2018) 3007-3013.
[86] T.J. Zbacnik, R.E. Holcomb, D.S. Katayama, B.M. Murphy, R.W. Payne, R.C. Coccaro, G.J. Evans, J.E. Matsuura, C.S. Henry, M.C. Manning, Role of Buffers in Protein Formulations, Journal of Pharmaceutical Sciences 106 (2017) 713-733.
[87] J. Kang, X. Lin, J. Penera, Rapid formulation development for monoclonal antibodies, BioProcess Int 14 (2016) 40.
[88] S. Nema, R.J. Washkuhn, R.J. Brendel, Excipients and their use in injectable products, PDA Journal of Pharmaceutical Science and Technology 51 (1997) 166-171.
[89] S.H. Jeong, Analytical methods and formulation factors to enhance protein stability in solution, Archives of pharmacal research 35 (2012) 1871-1886.
[90] J.L. Cleland, M.F. Powell, S.J. Shire, The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation, Critical reviews in therapeutic drug carrier systems 10 (1993) 307-377.
[91] M. Luksic, R. Buchner, B. Hribar-Lee, V. Vlachy, Dielectric relaxation spectroscopy of aliphatic ionene bromides and fluorides in water: the role of the polyion's charge density and the nature of the counterions, Macromolecules 42 (2009) 4337-4342.
[92] S. Brudar, B.A.-O.X. Hribar-Lee, Effect of Buffer on Protein Stability in Aqueous Solutions: A Simple Protein Aggregation Model.
[93] H. Chen, L. Yuan, W. Song, Z. Wu, D. Li, Biocompatible polymer materials: Role of protein–surface interactions, Progress in Polymer Science 33 (2008) 1059-1087.
[94] Q. Wei, T. Becherer, S. Angioletti-Uberti, J. Dzubiella, C. Wischke, A.T. Neffe, A. Lendlein, M. Ballauff, R. Haag, Protein Interactions with Polymer Coatings and Biomaterials, Angewandte Chemie International Edition 53 (2014) 8004-8031.
[95] Y. Le Basle, P. Chennell, N. Tokhadze, A. Astier, V. Sautou, Physicochemical Stability of Monoclonal Antibodies: A Review, Journal of Pharmaceutical Sciences 109 (2020) 169-190.
[96] R. Rajan, S. Ahmed, N. Sharma, N. Kumar, A. Debas, K. Matsumura, Review of the current state of protein aggregation inhibition from a materials chemistry perspective: special focus on polymeric materials, Materials Advances 2 (2021) 1139-1176.

[97] Y. Baek, N. Singh, A. Arunkumar, A.L. Zydney, Effects of Histidine and Sucrose on the Biophysical Properties of a Monoclonal Antibody, Pharmaceutical Research 34 (2017) 629-639.
[98] S. Fischer, J. Hoernschemeyer, H.-C. Mahler, Glycation during storage and administration of monoclonal antibody formulations, European Journal of Pharmaceutics and Biopharmaceutics 70 (2008) 42-50.
[99] C. Haeuser, P. Goldbach, J. Huwyler, W. Friess, A. Allmendinger, Excipients for Room Temperature Stable Freeze-Dried Monoclonal Antibody Formulations, Journal of Pharmaceutical Sciences 109 (2020) 807-817.
[100] S.J. Shire, Formulation and manufacturability of biologics, Current Opinion in Biotechnology 20 (2009) 708-714.
[101] M.E. Ressing, W. Jiskoot, H. Talsma, C.W. van Ingen, E.C. Beuvery, D.J.A. Crommelin, The Influence of Sucrose, Dextran, and Hydroxypropyl-β-cyclodextrin as Lyoprotectants for a Freeze-Dried Mouse IgG2a Monoclonal Antibody (MN12), Pharmaceutical Research 9 (1992) 266-270.
[102] S. Uchiyama, Liquid formulation for antibody drugs, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1844 (2014) 2041-2052.
[103] P.P. Bhojane, S. Joshi, S.J. Sahoo, A.A.-O. Rathore, Unexplored Excipients in Biotherapeutic Formulations: Natural Osmolytes as Potential Stabilizers Against Thermally Induced Aggregation of IgG1 Biotherapeutics.
[104] D. Rehder, P. Bondarenko, D. Chelius, A. McAuley, M. Matsumura, Stable polypeptide formulations, in: U.S.P.a.T. Office (Ed.) United States Patent and Trademark Office, Amgen Inc, United States, 2010.
[105] M. Pelliccia, P. Andreozzi, J. Paulose, M. D'Alicarnasso, V. Cagno, M. Donalisio, A. Civra, R.M. Broeckel, N. Haese, P. Jacob Silva, R.P. Carney, V. Marjomäki, D.N. Streblow, D. Lembo, F. Stellacci, V. Vitelli, S. Krol, Additives for vaccine storage to improve thermal stability of adenoviruses from hours to months, Nature Communications 7 (2016) 13520.
[106] F. Emami, A. Vatanara, E.J. Park, D.H. Na, Drying Technologies for the Stability and Bioavailability of Biopharmaceuticals, Pharmaceutics, 2018.
[107] X.M. Lam, J.Q. Oeswein, B. Ongpipattanakul, Z. Shahrokh, S.X. Wang, R.P. Weissburg, R.L. Wong, Antibody formulation, in: U.S.P.a.T. Office (Ed.) Patent Public Search Genetech, Inc, United States, 2005.
[108] F. Amini, E. Auma, Y. Hsia, S. Bilton, T. Hall, L. Ramkhelawon, P.T. Heath, K. Le Doare, Reliability of dried blood spot (DBS) cards in antibody measurement: A systematic review, PLOS ONE 16 (2021) e0248218.
[109] M. Johnson, Antibody shelf life/how to store antibodies, Mater Methods 2 (2012) 120.
[110] W. Wang, Instability, stabilization, and formulation of liquid protein pharmaceuticals, International Journal of Pharmaceutics 185 (1999) 129-188.
[111] S.V. Lale, A.K. Goyal M Fau - Bansal, A.K. Bansal, Development of lyophilization cycle and effect of excipients on the stability of catalase during lyophilization.
[112] L. Wang, L. Liu, X. Hong, D. Liu, Z. Cheng, A novel method for the storage and transport of biological samples of therapeutic proteins prior to the detection of analytes using ELISA, Scientific Reports 11 (2021) 8763.
[113] P. Espié, Y. He, P. Koo, D. Sickert, C. Dupuy, E. Chokoté, R. Schuler, H. Mergentaler, J. Ristov, J. Milojevic, First-in-human clinical trial to assess pharmacokinetics, pharmacodynamics, safety, and tolerability of iscalimab, an anti-CD40 monoclonal antibody, American Journal of Transplantation 20 (2020) 463-473.

[114] J.P. Amorij, J. Meulenaar, W.L.J. Hinrichs, T. Stegmann, A. Huckriede, F. Coenen, H.W. Frijlink, Rational design of an influenza subunit vaccine powder with sugar glass technology: Preventing conformational changes of haemagglutinin during freezing and freeze-drying, Vaccine 25 (2007) 6447-6457.
[115] J. Chesko, C. Fox, T. Dutill, T. Vedvick, S. Reed, Lyophilization and Stabilization of Vaccines, Development of Vaccines2011, pp. 385-397.
[116] R.J. Harris, S.J. Shire, C. Winter, Commercial manufacturing scale formulation and analytical characterization of therapeutic recombinant antibodies, Drug development research 61 (2004) 137-154.
[117] K.-T. Schneider, T. Kirmann, E.V. Wenzel, J.-H. Grosch, S. Polten, D. Meier, M. Becker, P. Matejtschuk, M. Hust, G. Russo, Shelf-Life Extension of Fc-Fused Single Chain Fragment Variable Antibodies by Lyophilization, Frontiers in Cellular and Infection Microbiology (2021) 1124.
[118] I. Roy, M.N. Gupta, Freeze-drying of proteins: some emerging concerns, Biotechnology and Applied Biochemistry 39 (2004) 165-177.
[119] Y.-F. Maa, S.J. Prestrelski, Biopharmaceutical powders particle formation and formulation considerations, Current pharmaceutical biotechnology 1 (2000) 283-302.
[120] J.F. Carpenter, J.H. Crowe, The mechanism of cryoprotection of proteins by solutes, Cryobiology 25 (1988) 244-255.
[121] T. Arakawa, S.J. Prestrelski, W.C. Kenney, J.F. Carpenter, Factors affecting short-term and long-term stabilities of proteins, Advanced drug delivery reviews 46 (2001) 307-326.
[122] P.L. Privalov, Cold denaturation of protein, Critical reviews in biochemistry and molecular biology 25 (1990) 281-306.
[123] J.A. Rupley, G. Careri, Protein hydration and function, Advances in protein chemistry 41 (1991) 37-172.
[124] S.J. Prestrelski, N. Tedeschi, T. Arakawa, J.F. Carpenter, Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers, Biophysical journal 65 (1993) 661-671.
[125] M.Z. Zhang, J. Wen, T. Arakawa, S.J. Prestrelski, A new strategy for enhancing the stability of lyophilized protein: the effect of the reconstitution medium on keratinocyte growth factor, Pharmaceutical research 12 (1995) 1447-1452.
[126] J.F. Carpenter, B.S. Chang, Lyophilization of protein pharmaceuticals, Biotechnology and biopharmaceutical manufacturing, processing, and preservation, CRC Press2020, pp. 199-264.
[127] A. Dong, S.J. Prestrelski, S.D. Allison, J.F. Carpenter, Infrared spectroscopic studies of lyophilization-and temperature-induced protein aggregation, Journal of pharmaceutical sciences 84 (1995) 415-424.
[128] C. Nowak, J. K. Cheung, S. M. Dellatore, A. Katiyar, R. Bhat, J. Sun, G. Ponniah, A. Neill, B. Mason, A. Beck, Forced degradation of recombinant monoclonal antibodies: A practical guide, MAbs, Taylor & Francis, 2017, pp. 1217-1230.
[129] S. Gupta, W. Jiskoot, C. Schoneich, A.S. Rathore, Oxidation and Deamidation of Monoclonal Antibody Products: Potential Impact on Stability, Biological Activity, and Efficacy, J Pharm Sci 111 (2022) 903-918.
[130] E.Y. Chi, S. Krishnan, T.W. Randolph, J.F. Carpenter, Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation, Pharmaceutical Research 20 (2003) 1325-1336.

[131] M.C. Manning, D.K. Chou, B.M. Murphy, R.W. Payne, D.S. Katayama, Stability of Protein Pharmaceuticals: An Update, Pharmaceutical Research 27 (2010) 544-575.
[132] F. Torkashvand, B. Vaziri, Main Quality Attributes of Monoclonal Antibodies and Effect of Cell Culture Components.
[133] N.B. Pham, W.S. Meng, Protein aggregation and immunogenicity of biotherapeutics, International Journal of Pharmaceutics 585 (2020) 119523.
[134] M.A.-O. Ajmal, Protein Misfolding and Aggregation in Proteinopathies: Causes, Mechanism and Cellular Response. LID - 10.3390/diseases11010030 [doi] LID - 30.
[135] L.N. Jessie Zhixin Miao, Stephan Altmaier, Katie Falls, Uma Sreenivasan, Workflow for Antibody Aggregates and Fragments Analysis, in: M. MilliporeSigma, Sepax (Ed.) Large Molecule HPLC, MilliporeSigma, 2023.
[136] B. Spanov, N. Govorukhina, N.C.v.d. Merbel, R. Bischoff, Analytical tools for the characterization of deamidation in monoclonal antibodies, Journal of Chromatography Open 2 (2022) 100025.
[137] Y.D. Liu, L. Cadang, K. Bol, X. Pan, K. Tschudi, M. Jazayri, J. Camperi, D. Michels, J. Stults, R.J. Harris, Challenges and strategies for a thorough characterization of antibody acidic charge variants, Bioengineering 9 (2022) 641.
[138] J. Vlasak, R. Ionescu, Fragmentation of monoclonal antibodies, mAbs 3 (2011) 253-263.
[139] M.E. Alam, T.R. Slaney, L. Wu, T.K. Das, S. Kar, G.V. Barnett, A. Leone, P.M. Tessier, Unique Impacts of Methionine Oxidation, Tryptophan Oxidation, and Asparagine Deamidation on Antibody Stability and Aggregation, Journal of Pharmaceutical Sciences 109 (2020) 656-669.
[140] N. Perico, J. Purtell, T.M. Dillon, M.S. Ricci, Conformational implications of an inversed pH-dependent antibody aggregation, Journal of pharmaceutical sciences 98 (2009) 3031-3042.
[141] K.E. Lawson, J.K. Dekle, A.J. Adamczyk, Towards pharmaceutical protein stabilization: DFT and statistical learning studies on non-enzymatic peptide hydrolysis degradation mechanisms, Computational and Theoretical Chemistry 1218 (2022) 113938.
[142] Y. Sun, M. Frenkel-Pinter, C.L. Liotta, M.A. Grover, The pH dependent mechanisms of non-enzymatic peptide bond cleavage reactions, Physical Chemistry Chemical Physics 22 (2020) 107-113.
[143] A. Radzicka, R. Wolfenden, Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases, Journal of the American Chemical Society 118 (1996) 6105-6109.
[144] R.M. Smith, D.E. Hansen, The pH-Rate Profile for the Hydrolysis of a Peptide Bond, Journal of the American Chemical Society 120 (1998) 8910-8913.
[145] T. Xiang, E. Lundell, Z. Sun, H. Liu, Structural effect of a recombinant monoclonal antibody on hinge region peptide bond hydrolysis, Journal of Chromatography B 858 (2007) 254-262.
[146] S. Gupta, K. Upadhyay, C. Schöneich, A.S. Rathore, Impact of various factors on the kinetics of non-enzymatic fragmentation of a monoclonal antibody, European Journal of Pharmaceutics and Biopharmaceutics 178 (2022) 131-139.
[147] K.E. Lawson, M.N. Evans, J.K. Dekle, A.J. Adamczyk, Computing the Differences between Asn-X and Gln-X Deamidation and Their Impact on Pharmaceutical and Physiological Proteins: A Theoretical Investigation Using Model Dipeptides, The Journal of Physical Chemistry A 127 (2023) 57-70.
[148] L. Huang, J. Lu, V.J. Wroblewski, J.M. Beals, R.M. Riggin, In Vivo Deamidation Characterization of Monoclonal Antibody by LC/MS/MS, Analytical Chemistry 77 (2005) 1432-1439.

[149] A.B. Robinson, C.J. Rudd, Deamidation of glutaminyl and asparaginyl residues in peptides and proteins, Current topics in cellular regulation, Elsevier1974, pp. 247-295.
[150] N.E. Robinson, Protein deamidation, Proceedings of the National Academy of Sciences 99 (2002) 5283-5288.
[151] N.E. Robinson, Z.W. Robinson, B.R. Robinson, A.L. Robinson, J.A. Robinson, M.L. Robinson, A.B. Robinson, Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides, The Journal of Peptide Research 63 (2004) 426-436.
[152] A.L. Pace, R.L. Wong, Y.T. Zhang, Y.-H. Kao, Y.J. Wang, Asparagine deamidation dependence on buffer type, pH, and temperature, Journal of pharmaceutical sciences 102 (2013) 1712-1723.
[153] K. Patel, R.T. Borchardt, Chemical Pathways of Peptide Degradation. II. Kinetics of Deamidation of an Asparaginyl Residue in a Model Hexapeptide, Pharmaceutical Research 7 (1990) 703-711.
[154] J. Vlasak, M.C. Bussat, S. Wang, E. Wagner-Rousset, M. Schaefer, C. Klinguer-Hamour, M. Kirchmeier, N. Corvaïa, R. Ionescu, A. Beck, Identification and characterization of asparagine deamidation in the light chain CDR1 of a humanized IgG1 antibody, Analytical Biochemistry 392 (2009) 145-154.
[155] X. Lu, R.P. Nobrega, H. Lynaugh, T. Jain, K. Barlow, T. Boland, A. Sivasubramanian, M. Vásquez, Y. Xu, Deamidation and isomerization liability analysis of 131 clinical-stage antibodies, mAbs 11 (2019) 45-57.
[156] W.F. Dall'Acqua, K.E. Cook, M.M. Damschroder, R.M. Woods, H. Wu, Modulation of the effector functions of a human IgG1 through engineering of its hinge region, The Journal of Immunology 177 (2006) 1129-1138.
[157] S.L. Cohen, C. Price, J. Vlasak, β-Elimination and Peptide Bond Hydrolysis: Two Distinct Mechanisms of Human IgG1 Hinge Fragmentation upon Storage, Journal of the American Chemical Society 129 (2007) 6976-6977.
[158] A.J. Cordoba, B.-J. Shyong, D. Breen, R.J. Harris, Non-enzymatic hinge region fragmentation of antibodies in solution, Journal of Chromatography B 818 (2005) 115-121.
[159] L.J. Harris, S.B. Larson, K.W. Hasel, A. McPherson, Refined structure of an intact IgG2a monoclonal antibody, Biochemistry 36 (1997) 1581-1597.
[160] M.M. Gromiha, 2.3.2 Amino acid composition, in: M.M. Gromiha (Ed.), Protein Bioinformatics, Academic Press, Singapore, 2010, pp. 29-62.
[161] L. Gentiluomo, D. Roessner, D. Augustijn, H. Svilenov, A. Kulakova, S. Mahapatra, G. Winter, W. Streicher, Å. Rinnan, G.H.J. Peters, P. Harris, W. Frieß, Application of interpretable artificial neural networks to early monoclonal antibodies development, European Journal of Pharmaceutics and Biopharmaceutics 141 (2019) 81-89.
[162] T. Ohta, M. Mori, H. Koshiba, J. Takada, T. Matsuyama, S. Ishii, Production of monoclonal antibodies specific for human bone γ-carboxyglutamic acid containing protein, Virchows Archiv A 418 (1991) 499-502.
[163] S.C. Flores, L.J. Lu, J. Yang, N. Carriero, M.B. Gerstein, Hinge Atlas: relating protein sequence to sites of structural flexibility, BMC Bioinformatics 8 (2007) 167.
[164] Amino Acids Reference Chart, Sigma Aldrich, https://www.sigmaaldrich.com/US/en/technical-documents/technical-article/protein-biology/protein-structural-analysis/amino-acid-reference-chart, 2023, pp. Properties of Common Amino Acids.

[165] D. Kahne, W.C. Still, Hydrolysis of a peptide bond in neutral water, Journal of the American Chemical Society 110 (1988) 7529-7534.
[166] D.L. Riggs, S.V. Gomez, R.R. Julian, Sequence and Solution Effects on the Prevalence of d-Isomers Produced by Deamidation, ACS Chemical Biology 12 (2017) 2875-2882.
[167] Q. Yan, M. Huang, M.J. Lewis, P. Hu, Structure Based Prediction of Asparagine Deamidation Propensity in Monoclonal Antibodies, mAbs 10 (2018) 901-912.
[168] R. Tyler-Cross, V. Schirch, Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides, Journal of Biological Chemistry 266 (1991) 22549-22556.
[169] J.F. Sydow, F. Lipsmeier, V. Larraillet, M. Hilger, B. Mautz, M. Molhoj, J. Kuentzer, S. Klostermann, J. Schoch, H.R. Voelger, J.T. Regula, P. Cramer, A. Papadimitriou, H. Kettenberger, Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions, PLoS One 9 (2014) e100736.
[170] C.L.D. Gamage, T.S. Hageman, D.D. Weis, Rapid Prediction of Deamidation Rates of Proteins to Assess Their Long-Term Stability Using Hydrogen Exchange–Mass Spectrometry, Journal of Pharmaceutical Sciences 108 (2019) 1964-1972.
[171] F. Hund, Early History of the Quantum Mechanical Treatment of the Chemical Bond, Angewandte Chemie International Edition in English 16 (1977) 87-91.
[172] F. Jensen, Introduction to computational chemistry, John wiley & sons2017.
[173] E. Nelson, Derivation of the Schrödinger equation from Newtonian mechanics, Physical review 150 (1966) 1079.
[174] N.H. March, W.H. Young, S. Sampanthar, The many-body problem in quantum mechanics, Courier Corporation1995.
[175] J.-M. Combes, P. Duclos, R. Seiler, The born-oppenheimer approximation, Rigorous atomic and molecular physics, Springer1981, pp. 185-213.
[176] M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln, Annalen der Physik 389 (1927) 457-484.
[177] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Physical review 136 (1964) B864.
[178] L.J. Sham, W. Kohn, One-particle properties of an inhomogeneous interacting electron gas, Physical Review 145 (1966) 561.
[179] N. Mardirossian, M. Head-Gordon, Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals, Molecular Physics 115 (2017) 2315-2372.
[180] R. Haunschild, A. Barth, B. French, A comprehensive analysis of the history of DFT based on the bibliometric method RPYS, Journal of Cheminformatics 11 (2019) 72.
[181] W. Kohn, W. kohn and lj sham, phys. rev. 140, a1133 (1965), phys. Rev 140 (1965) A1133.
[182] P.E.M. Siegbahn, M. Wirstam, M. Zimmer, Theoretical study of the mechanism of peptide ring formation in green fluorescent protein, International Journal of Quantum Chemistry 81 (2001) 169-186.
[183] M. Cutini, P. Ugliengo, Infrared harmonic features of collagen models at B3LYP-D3: From amide bands to the THz region, The Journal of Chemical Physics 155 (2021) 075102.
[184] B. Peters, B.L. Trout, Asparagine Deamidation: pH-Dependent Mechanism from Density Functional Theory, Biochemistry 45 (2006) 5384-5392.
[185] J.P. Perdew, K. Schmidt, Jacob's ladder of density functional approximations for the exchange-correlation energy, American Institute of Physics, pp. 1-20.

[186] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Physical Review A 38 (1988) 3098-3100.
[187] M.A. Krestyaninov, E.G. Odintsova, A.M. Kolker, M.G. Kiselev, The structure of water – Acetamide hydrogen bonded complexes. Quantum chemical analysis, Journal of Molecular Liquids 264 (2018) 343-351.
[188] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, The Journal of Chemical Physics 132 (2010) 154104.
[189] S. Grimme, Density functional theory with London dispersion corrections, WIREs Computational Molecular Science 1 (2011) 211-228.
[190] T. Ichibha, O. Srihakulung, G. Chao, A.T. Hanindriyo, L. Lawtrakul, K. Hongo, R. Maezono, Importance of vdW and long-range exchange interactions to DFT-predicted docking energies between plumbagin and cyclodextrins, arXiv preprint arXiv:1904.02503 (2019).
[191] D. McQuarrie, Statistical mechanics. 2000, Sausalito, Calif.: University Science Books 12 (2004) 641.
[192] D.A. McQuarrie, Statistical thermodynamics, (1973).
[193] W. Ebeling, I.M. Sokolov, Statistical thermodynamics and stochastic theory of nonequilibrium systems, World Scientific2005.
[194] T.L. Hill, An introduction to statistical thermodynamics, Courier Corporation1986.
[195] L.D. Landau, E.M. Lifshits, L.P. Pitaevskiĭ, Course of theoretical physics: statistical physics, Elsevier Butterworth Heinemann1980.
[196] A. Fernández-Ramos, B.A. Ellingson, R. Meana-Pañeda, J.M.C. Marques, D.G. Truhlar, Symmetry numbers and chemical reaction rates, Theor. Chem. Acc. 118 (2007) 813-826.
[197] R. Chang, Physical chemistry for the biosciences, University Science Books2005.
[198] K.J. Laidler, The development of the Arrhenius equation, Journal of Chemical Education 61 (1984) 494.
[199] V. Jagannadham, How Do We Introduce the Arrhenius Pre-Exponential Factor (A) to Graduate Students?, Creative Education 1 (2010) 128.
[200] S.S. Zumdahl, D.J. DeCoste, Introductory chemistry: A foundation, Cengage learning2018.
[201] K.J. Laidler, Chemical Kinetics.(3rdedn), Pearson Education2008.
[202] Z. Qin, S.K. Balasubramanian, W.F. Wolkers, J.A. Pearce, J.C. Bischof, Correlated parameter fit of arrhenius model for thermal denaturation of proteins and cells, Ann Biomed Eng 42 (2014) 2392-2404.
[203] M. Schwaab, J.C. Pinto, Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant, Chemical Engineering Science 62 (2007) 2750-2764.
[204] C.J. Cramer, D.G. Truhlar, Implicit solvation models: equilibria, structure, spectra, and dynamics, Chemical Reviews 99 (1999) 2161-2200.
[205] A. Okur, L. Wickstrom, M. Layten, R. Geney, K. Song, V. Hornak, C. Simmerling, Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model, Journal of Chemical Theory and Computation 2 (2006) 420-433.
[206] S.N. Steinmann, P. Sautet, C. Michel, Solvation free energies for periodic surfaces: comparison of implicit and explicit solvation models, Physical Chemistry Chemical Physics 18 (2016) 31850-31861.
[207] R.C. DeMille, T.E. Cheatham, V. Molinero, A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions, The Journal of Physical Chemistry B 115 (2011) 132-142.

[208] V. Barone, M. Cossi, Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model, The Journal of Physical Chemistry A 102 (1998) 1995-2001.
[209] B. Mennucci, Polarizable continuum model, Wiley Interdisciplinary Reviews: Computational Molecular Science 2 (2012) 386-404.
[210] C.N. Schutz, A. Warshel, What are the dielectric “constants” of proteins and how to validate electrostatic models?, Proteins: Structure, Function, and Bioinformatics 44 (2001) 400-417.
[211] A. Warshel, S.T. Russell, Calculations of electrostatic interactions in biological systems and in solutions, Quarterly Reviews of Biophysics 17 (1984) 283-422.
[212] L. Li, C. Li, Z. Zhang, E. Alexov, On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi, Journal of Chemical Theory and Computation 9 (2013) 2126-2136.
[213] L. Wang, Z. Zhang, W. Rocchia, E. Alexov, Using DelPhi Capabilities to Mimic Protein's Conformational Reorganization with Amino Acid Specific Dielectric Constants, Commun Comput Phys 13 (2013) 13-30.
[214] A. Warshel, A. Papazyan, Electrostatic effects in macromolecules: fundamental concepts and practical modeling, Current Opinion in Structural Biology 8 (1998) 211-217.
[215] M.K. Gilson, B.H. Honig, The dielectric constant of a folded protein, Biopolymers 25 (1986) 2097-2119.
[216] T. Simonson, D. Perahia, Polar fluctuations in proteins: molecular-dynamic studies of cytochrome c in aqueous solution, Faraday Discussions 103 (1996) 71-90.
[217] D.L. Mobley, K.A. Dill, J.D. Chodera, Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules, The Journal of Physical Chemistry B 112 (2008) 938-946.
[218] H. Gouda, I.D. Kuntz, D.A. Case, P.A. Kollman, Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods, Biopolymers 68 (2003) 16-34.
[219] M. Kato, A.V. Pisliakov, A. Warshel, The barrier for proton transport in aquaporins as a challenge for electrostatic models: The role of protein relaxation in mutational calculations, Proteins: Structure, Function, and Bioinformatics 64 (2006) 829-844.
[220] M. Amin, J. Küpper, Variations in Proteins Dielectric Constants, ChemistryOpen 9 (2020) 691-694.
[221] A. Acharya, D. Yi, A. Pavlova, V. Agarwal, J.C. Gumbart, Resolving the hydride transfer pathway in oxidative conversion of proline to pyrrole, Biochemistry 61 (2022) 206-215.
[222] C.M. Lemon, M.A. Marletta, Corrole-Substituted Fluorescent Heme Proteins, Inorganic chemistry 60 (2021) 2716-2729.
[223] S. Vicatos, M. Roca, A. Warshel, Effective approach for calculations of absolute stability of proteins using focused dielectric constants, Proteins: Structure, Function, and Bioinformatics 77 (2009) 670-684.
[224] S. Roy, S. Goedecker, V. Hellmann, Bell-Evans-Polanyi principle for molecular dynamics trajectories and its implications for global optimization, Physical Review E 77 (2008) 056707.
[225] M.R. Chernick, R.A. LaBudde, An introduction to bootstrap methods with applications to R, John Wiley & Sons2014.
[226] B. Efron, R. Tibshirani, The Bootstrap Method for Assessing Statistical Accuracy, Behaviormetrika 12 (1985) 1-35.

[227] J.S. Haukoos, R.J. Lewis, Advanced Statistics: Bootstrapping Confidence Intervals for Statistics with "Difficult" Distributions, Academic Emergency Medicine 12 (2005) 360-365.
[228] A.R. Henderson, The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data, Clinica Chimica Acta 359 (2005) 1-26.
[229] W. Jiang, R. Simon, A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification, Statistics in Medicine 26 (2007) 5320-5334.
[230] D. Kushary, Bootstrap methods and their application, Taylor & Francis, 2000.
[231] R.J. Tibshirani, B. Efron, An introduction to the bootstrap, Monographs on statistics and applied probability 57 (1993) 1-436.
[232] R. Wehrens, H. Putter, L.M.C. Buydens, The bootstrap: a tutorial, Chemometrics and Intelligent Laboratory Systems 54 (2000) 35-52.
[233] A.A.-O. Camargo, PCAtest: testing the statistical significance of Principal Component Analysis in R.
[234] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2016) 20150202.
[235] J. Lever, M. Krzywinski, N. Altman, Points of significance: Principal component analysis, Nature methods 14 (2017) 641-643.
[236] M. Ringnér, What is principal component analysis?, Nature Biotechnology 26 (2008) 303-304.
[237] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and Intelligent Laboratory Systems 2 (1987) 37-52.
[238] B. Godugu, P. Neta, Y. Simón-Manso, S.E. Stein, Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions, Journal of the American Society for Mass Spectrometry 21 (2010) 1169-1176.
[239] J. Kuriyan, B. Konforti, D. Wemmer, The molecules of life: Physical and chemical principles, Garland Science2012.
[240] D.L. Riggs, J.W. Silzel, Y.A. Lyon, A.S. Kang, R.R. Julian, Analysis of Glutamine Deamidation: Products, Pathways, and Kinetics, Analytical Chemistry 91 (2019) 13032-13038.
[241] D.-B. Huang, C.-H. Chang, C. Ainsworth, G. Johnson, A. Solomon, F.J. Stevens, M. Schiffer, Variable domain structure of κIV human light chain Len: High homology to the murine light chain McPC603, Molecular Immunology 34 (1997) 1291-1301.
[242] L. Jia, Y. Sun, Protein asparagine deamidation prediction based on structures with machine learning methods, PloS one 12 (2017) e0181347-e0181347.
[243] S. Catak, G. Monard, V. Aviyente, M.F. Ruiz-López, Deamidation of Asparagine Residues: Direct Hydrolysis versus Succinimide-Mediated Deamidation Mechanisms, The Journal of Physical Chemistry A 113 (2009) 1111-1120.
[244] M.A. Halim, M.H. Almatarneh, R.A. Poirier, Mechanistic Study of the Deamidation Reaction of Glutamine: A Computational Approach, The Journal of Physical Chemistry B 118 (2014) 2316-2330.
[245] L.J.M. Kempkes, G.C. Boles, J. Martens, G. Berden, P.B. Armentrout, J. Oomens, Deamidation of Protonated Asparagine-Valine Investigated by a Combined Spectroscopic, Guided Ion Beam, and Theoretical Study, J Phys Chem A 122 (2018) 2424-2436.
[246] A.L. Heaton, P.B. Armentrout, Thermodynamics and mechanism of the deamidation of sodium-bound asparagine, Journal of the American Chemical Society 130 (2008) 10227-10232.

[247] A.L. Heaton, P.B. Armentrout, Thermodynamics and mechanism of protonated asparagine decomposition, Journal of the American Society for Mass Spectrometry 20 (2011) 852-866.
[248] A.A. Wakankar, R.T. Borchardt, Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization, Journal of Pharmaceutical Sciences 95 (2006) 2321-2336.
[249] Y. Kori, R. Patel, A. Neill, H. Liu, A conventional procedure to reduce Asn deamidation artifacts during trypsin peptide mapping, Journal of Chromatography B 1009 (2016) 107-113.
[250] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 16 Rev. C.01, Wallingford, CT, 2016.
[251] Z. Wang, A. Nautiyal, C. Alexopoulos, R. Aqrawi, X. Huang, A. Ali, K.E. Lawson, K. Riley, A.J. Adamczyk, P. Dong, X. Zhang, Fentanyl Assay Derived from Intermolecular Interaction-Enabled Small Molecule Recognition (iMSR) with Differential Impedance Analysis for Point-of-Care Testing, Analytical Chemistry 94 (2022) 9242-9251.
[252] A.G. Baboul, L.A. Curtiss, P.C. Redfern, K. Raghavachari, Gaussian-3 theory using density functional geometries and zero-point energies, J Chem Phys 110 (1999) 7650-7657.
[253] M.A. Krestyaninov, E.G. Odintsova, A.M. Kolker, M.G. Kiselev, The structure of water - Acetamide hydrogen bonded complexes. Quantum chemical analysis, Journal of Molecular Liquids 264 (2018) 343-351.
[254] A.P. Scott, L. Radom, Harmonic vibrational frequencies: an evaluation of Hartree− Fock, Møller− Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors, The Journal of Physical Chemistry 100 (1996) 16502-16513.
[255] A.J. Adamczyk, L.J. Broadbelt, The role of multifunctional kinetics during early-stage silicon hydride pyrolysis: reactivity of Si2H2 isomers with SiH4 and Si2H6, J Phys Chem A 115 (2011) 2409-2422.
[256] A.J. Adamczyk, M.F. Reyniers, G.B. Marin, L.J. Broadbelt, Kinetics of substituted silylene addition and elimination in silicon nanocluster growth captured by group additivity, Chemphyschem 11 (2010) 1978-1994.
[257] Y. Choi, T.J. Preston, A.J. Adamczyk, Data-Driven Investigation of Monosilane and Ammonia Co-Pyrolysis to Silicon-Nitride-Based Ceramic Nanomaterials, Chemphyschem 21 (2020) 2627-2642.
[258] Y. Choi, A.J. Adamczyk, Tuning Hydrogenated Silicon, Germanium, and SiGe Nanocluster Properties Using Theoretical Calculations and a Machine Learning Approach, Journal of Physical Chemistry A 122 (2018) 9851-9868.
[259] A.J. Adamczyk, L.J. Broadbelt, Thermochemical property estimation of hydrogenated silicon clusters, J Phys Chem A 115 (2011) 8969-8982.

[260] A.J. Adamczyk, M.F. Reyniers, G.B. Marin, L.J. Broadbelt, Hydrogenated amorphous silicon nanostructures: novel structure-reactivity relationships for cyclization and ring opening in the gas phase, Theor. Chem. Acc. 128 (2011) 91-113.
[261] A.J. Adamczyk, M.F. Reyniers, G.B. Marin, L.J. Broadbelt, Kinetic correlations for H2 addition and elimination reaction mechanisms during silicon hydride pyrolysis, Phys Chem Chem Phys 12 (2010) 12676-12696.
[262] Y. Choi, A.J. Adamczyk, Competitive Hydrogen Migration in Silicon Nitride Nanoclusters: Reaction Kinetics Generalized from Supervised Machine Learning, J Phys Chem A 126 (2022) 2677-2689.
[263] A.J. Adamczyk, M.-F. Reyniers, G.B. Marin, L.J. Broadbelt, Exploring 1, 2-hydrogen shift in silicon nanoparticles: reaction kinetics from quantum chemical calculations and derivation of transition state group additivity database, The Journal of Physical Chemistry A 113 (2009) 10933-10946.
[264] H.B. Schlegel, Optimization of equilibrium geometries and transition structures, Journal of computational chemistry 3 (1982) 214-218.
[265] C. Peng, P.Y. Ayala, H.B. Schlegel, M.J. Frisch, Using redundant internal coordinates to optimize equilibrium geometries and transition states, Journal of Computational Chemistry 17 (1996) 49-56.
[266] H. Bernhard Schlegel, Estimating the hessian for gradient-type geometry optimizations, Theoretica chimica acta 66 (1984) 333-340.
[267] P.Y. Ayala, H.B. Schlegel, A combined method for determining reaction paths, minima, and transition state geometries, The Journal of chemical physics 107 (1997) 375-384.
[268] A. Jalan, I.M. Alecu, R. Meana-Pañeda, J. Aguilera-Iparraguirre, K.R. Yang, S.S. Merchant, D.G. Truhlar, W.H. Green, New pathways for formation of acids and carbonyl products in low-temperature oxidation: The Korcek decomposition of γ-ketohydroperoxides, Journal of the American Chemical Society 135 (2013) 11100-11114.
[269] H.S. Fogler, Essentials of Chemical Reaction Engineering: Essenti Chemica Reactio Engi, Pearson Education2010.
[270] L.P. Stratton, R.M. Kelly, J. Rowe, J.E. Shively, D.D. Smith, J.F. Carpenter, M.C. Manning, Controlling deamidation rates in a model peptide: effects of temperature, peptide concentration, and additives, Journal of pharmaceutical sciences 90 (2001) 2141-2148.
[271] D. Ouellette, C. Chumsae, A. Clabbers, C. Radziejewski, I. Correia, Comparison of the in vitro and in vivo stability of a succinimide intermediate observed on a therapeutic IgG1 molecule, mAbs 5 (2013) 432-444.
[272] J. Zhang, H. Yip, V. Katta, Identification of isomerization and racemization of aspartate in the Asp-Asp motifs of a therapeutic protein, Anal Biochem 410 (2011) 234-243.
[273] S. Clarke, Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins, International journal of peptide and protein research 30 (1987) 808-821.
[274] R.C. Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2021.
[275] N.A.C. Cressie, H.J. Whitford, How to use the two sample t-test, Biometrical Journal 28 (1986) 131-148.
[276] D. Liu, T. Wyttenbach, C.J. Carpenter, M.T. Bowers, Investigation of Noncovalent Interactions in Deprotonated Peptides: Structural and Energetic Competition between Aggregation and Hydration, Journal of the American Chemical Society 126 (2004) 3261-3270.

[277] S. Catak, V. Monard G Fau - Aviyente, M.F. Aviyente V Fau - Ruiz-López, M.F. Ruiz-López, Reaction mechanism of deamidation of asparaginyl residues in peptides: effect of solvent molecules.
[278] I. Kaliman, A. Nemukhin, S. Varfolomeev, Free energy barriers for the N-terminal asparagine to succinimide conversion: quantum molecular dynamics simulations for the fully solvated model, Journal of Chemical Theory and Computation 6 (2010) 184-189.
[279] T. Geiger, S. Clarke, Deamidation, Isomerization, and Racemization at Asparaginyl and Aspartyl Residues in Peptides - Succinimide-Linked Reactions That Contribute to Protein-Degradation, Journal of Biological Chemistry 262 (1987) 785-794.
[280] S. Catak, G. Monard, V. Aviyente, M.F. Ruiz-López, Computational study on nonenzymatic peptide bond cleavage at asparagine and aspartic acid, The Journal of Physical Chemistry A 112 (2008) 8752-8761.
[281] L. Gorb, A. Asensio, I. Tuñón, M.F. Ruiz-López, The mechanism of formamide hydrolysis in water from ab initio calculations and simulations, Chemistry–A European Journal 11 (2005) 6743-6753.
[282] A.A. Morgan, E. Rubenstein, Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome, PLOS ONE 8 (2013) e53785.
[283] R.C. Stephenson, S. Clarke, Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins, Journal of Biological Chemistry 264 (1989) 6164-6170.
[284] H. Yang, R.A. Zubarev, Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides, Electrophoresis 31 (2010) 1764-1772.
[285] J. Pitarch, M.F. Ruiz-López, E. Silla, J.-L. Pascual-Ahuir, I. Tuñón, Neutral and alkaline hydrolyses of model β-lactam antibiotics. An ab initio study of water catalysis, Journal of the American Chemical Society 120 (1998) 2146-2155.
[286] J. Pitarch, M.F. Ruiz-López, J.-L. Pascual-Ahuir, E. Silla, I. Tuñón, Ab initio calculations on neutral and alkaline hydrolyses of β-lactam antibiotics. A theoretical study including solvent effects, The Journal of Physical Chemistry B 101 (1997) 3581-3588.
[287] F.A.S. Konuklar, V. Aviyente, Modelling the hydrolysis of succinimide: formation of aspartate and reversible isomerization of aspartic acid via succinimide, Organic & biomolecular chemistry 1 (2003) 2290-2297.
[288] S. Dagan, T. Hagai, Y. Gavrilov, R. Kapon, Y. Levy, Z. Reich, Stabilization of a protein conferred by an increase in folded state entropy, Proceedings of the National Academy of Sciences 110 (2013) 10628.
[289] F.H. Schumann, R. Varadan, P.P. Tayakuniyil, J.H. Grossman, J.A. Camarero, D. Fushman, Changing the topology of protein backbone: the effect of backbone cyclization on the structure and dynamics of a SH3 domain, Frontiers in Chemistry 3 (2015) 26.
[290] F.C. Lightstone, T.C. Bruice, Enthalpy and Entropy in Ring Closure Reactions, Bioorganic Chemistry 26 (1998) 193-199.
[291] A.K. Pearce, J.C. Foster, R.K. O'Reilly, Recent developments in entropy-driven ring-opening metathesis polymerization: Mechanistic considerations, unique functionality, and sequence control, Journal of Polymer Science Part A: Polymer Chemistry 57 (2019) 1621-1634.
[292] L. Grassi, C. Cabrele, Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions, Amino Acids 51 (2019) 1409-1431.

[293] T.V. Brennan, S. Clarke, Spontaneous Degradation of Polypeptides at Aspartyl and Asparaginyl Residues - Effects of the Solvent Dielectric, Protein Sci 2 (1993) 331-338.
[294] L. Serrano, M. Bycroft, A.R. Fersht, Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles, Journal of molecular biology 218 (1991) 465-475.
[295] J.R. Murdoch, Rate-equilibriums relations and proton-transfer reactions, Journal of the American Chemical Society 94 (1972) 4410-4418.
[296] S. Khemaissa, S. Sagan, A. Walrant, Tryptophan, an Amino-Acid Endowed with Unique Properties and Its Many Roles in Membrane Proteins, Crystals, 2021.
[297] N.P. Sargaeva, C. Lin, P.B. O'Connor, Differentiating N-terminal aspartic and isoaspartic acid residues in peptides, Analytical chemistry 83 (2011) 6675-6682.
[298] N.P. Sargaeva, C. Lin, P.B. O'Connor, Identification of aspartic and isoaspartic acid residues in amyloid beta peptides, including Abeta1-42, using electron-ion reactions, Anal Chem 81 (2009) 9778-9786.
[299] S. Capasso, Thermodynamic parameters of the reversible isomerization of aspartic residues via a succinimide derivative, Thermochimica Acta 286 (1996) 41-50.
[300] N.P. Sargaeva, C. Lin, P.B. O’Connor, Identification of aspartic and isoaspartic acid residues in amyloid β peptides, including Aβ1− 42, using electron− ion reactions, Analytical chemistry 81 (2009) 9778-9786.
[301] N. DeGraan-Weber, J. Zhang, J.P. Reilly, Distinguishing Aspartic and Isoaspartic Acids in Peptides by Several Mass Spectrometric Fragmentation Methods, Journal of the American Society for Mass Spectrometry 27 (2016) 2041-2053.
[302] K. Bhanuramanand, S. Ahmad, N.M. Rao, Engineering deamidation-susceptible asparagines leads to improved stability to thermal cycling in a lipase, Protein science : a publication of the Protein Society 23 (2014) 1479-1490.
[303] R. Jaenicke, Protein structure and function at low temperatures, Philosophical Transactions of the Royal Society of London. B, Biological Sciences 326 (1990) 535-553.
[304] P. Tarakeshwar, S. Manogaran, Vibrational frequencies of cysteine and serine zwitterions—an ab initio assignment, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 51 (1995) 925-928.
[305] C.K. Wang, J.E. Swedberg, S.E. Northfield, D.J. Craik, Effects of Cyclization on Peptide Backbone Dynamics, The Journal of Physical Chemistry B 119 (2015) 15821-15830.
[306] S. Sinha, S. Zhang L Fau - Duan, T.D. Duan S Fau - Williams, J. Williams Td Fau - Vlasak, R. Vlasak J Fau - Ionescu, E.M. Ionescu R Fau - Topp, E.M. Topp, Effect of protein structure on deamidation rate in the Fc fragment of an IgG1 monoclonal antibody.
[307] M. Xie, R.L. Schowen, Secondary structure and protein deamidation, Journal of pharmaceutical sciences 88 (1999) 8-13.
[308] F.J. Irudayanathan, J. Zarzar, J. Lin, S. Izadi, Deciphering deamidation and isomerization in therapeutic proteins: Effect of neighboring residue, mAbs 14 (2022) 2143006.
[309] J. Hanna, A. Guerra-Moreno, J. Ang, Y. Micoogullari, Protein Degradation and the Pathologic Basis of Disease, Am J Pathol 189 (2019) 94-103.
[310] K.R. Kalli, P. Hsu, D.T. Fearon, Therapeutic uses of recombinant complement protein inhibitors, Springer, pp. 417-431.
[311] X. Li, C. Lin, P.B. O'Connor, Glutamine deamidation: differentiation of glutamic acid and gamma-glutamic acid in peptides by electron capture dissociation, Anal Chem 82 (2010) 3606-3615.

[312] L.J. Kempkes, J. Martens, J. Grzetic, G. Berden, J. Oomens, Deamidation Reactions of Asparagine- and Glutamine-Containing Dipeptides Investigated by Ion Spectroscopy, J Am Soc Mass Spectrom 27 (2016) 1855-1869.
[313] E.C. Peterson, R. Celikel, K. Gokulan, K.I. Varughese, Structural Characterization of a Therapeutic Anti-Methamphetamine Antibody Fragment: Oligomerization and Binding of Active Metabolites, PLOS ONE 8 (2013) e82690.
[314] S. Lee, Y. Xue, J. Hu, Y. Wang, X. Liu, B. Demeler, Y. Ha, The E2 Domains of APP and APLP1 Share a Conserved Mode of Dimerization, Biochemistry 50 (2011) 5453-5464.
[315] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: an indicator of antiquity, or preservational quality?, Rapid Commun Mass Spectrom 30 (2016) 251-255.
[316] L. Steinman, S.S. Ousman, W.H. Robinson, Alpha B-crystallin as a therapy for multiple sclerosis, in: W.I.P.O. (WIPO) (Ed.) World Intellectual Property Organization (WIPO)United States, 2011.
[317] L.V.S. Kumar, C.M. Rao, Chimeric protein α BNAC crystallin with extraordinarily high chaperone-like activity and a method related to the use thereof, in: W.I.P.O. (WIPO) (Ed.) World Intellectual Property Organization (WIPO), Council of Scientific and Industrial Research, United States, 2005.
[318] B.H. Toyama, J.N. Savas, S.K. Park, M.S. Harris, N.T. Ingolia, J.R. Yates, 3rd, M.W. Hetzer, Identification of long-lived proteins reveals exceptional stability of essential cellular structures, Cell 154 (2013) 971-982.
[319] N.E. Robinson, A. Robinson, Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins, Althouse press2004.
[320] S. Capasso, L. Mazzarella, F. Sica, A. Zagari, First evidence of spontaneous deamidation of glutamine residue via cyclic imide to α-and γ-glutamic residue under physiological conditions, Journal of the Chemical Society, Chemical Communications (1991) 1667-1668.
[321] K. Kato, T. Nakayoshi, E. Kurimoto, A. Oda, Computational Studies on the Nonenzymatic Deamidation Mechanisms of Glutamine Residues, ACS Omega 4 (2019) 3508-3513.
[322] K. Nguyen, M. Iskandar, D.L. Rabenstein, Kinetics and equilibria of cis/trans isomerization of secondary amide peptide bonds in linear and cyclic peptides, J Phys Chem B 114 (2010) 3387-3392.
[323] R. Bischoff, H.V. Kolbe, Deamidation of asparagine and glutamine residues in proteins and peptides: structural determinants and analytical methodology, J Chromatogr B Biomed Appl 662 (1994) 261-278.
[324] L. Sawyer, M.N. James, Carboxyl-carboxylate interactions in proteins, Nature 295 (1982) 79-80.
[325] H. Liu, G. Gaza-Bulseco, C. Chumsae, Glutamine deamidation of a recombinant monoclonal antibody, Rapid Commun Mass Spectrom 22 (2008) 4081-4088.
[326] K.B. Wiberg, The Concept of Strain in Organic-Chemistry, Angew Chem Int Edit 25 (1986) 312-322.
[327] S.W. Benson, Thermochemical kinetics, Wiley1976.
[328] O. Herbinet, B. Sirjean, R. Bounaceur, R. Fournet, F. Battin-Leclerc, G. Scacchi, P.M. Marquaire, Primary mechanism of the thermal decomposition of tricyclodecane, J Phys Chem A 110 (2006) 11298-11314.
[329] L.J.M. Kempkes, G.C. Boles, J. Martens, G. Berden, P.B. Armentrout, J. Oomens, Deamidation of Protonated Asparagine-Valine Investigated by a Combined Spectroscopic, Guided Ion Beam, and Theoretical Study, J Phys Chem A 122 (2018) 2424-2436.

[330] S.P. de Visser, G. Mukherjee, H.S. Ali, C.V. Sastri, Local Charge Distributions, Electric Dipole Moments, and Local Electric Fields Influence Reactivity Patterns and Guide Regioselectivities in alpha-Ketoglutarate-Dependent Non-heme Iron Dioxygenases, Acc Chem Res 55 (2022) 65-74.
[331] S.J. Holland, A.M. Jolly, M. Yasin, B.J. Tighe, Polymers for biodegradable medical devices: II. Hydroxybutyrate-hydroxyvalerate copolymers: Hydrolytic degradation studies, Biomaterials 8 (1987) 289-295.
[332] J.K. Scott, Discovering peptide ligands using epitope libraries, Trends in biochemical sciences 17 (1992) 241-245.
[333] A.L. Daugherty, R.J. Mrsny, Formulation and delivery issues for monoclonal antibody therapeutics, Advanced drug delivery reviews 58 (2006) 686-706.
[334] B. Pan, M.S. Ricci, B.L. Trout, A molecular mechanism of hydrolysis of peptide bonds at neutral pH using a model compound, The Journal of Physical Chemistry B 115 (2011) 5958-5970.
[335] Y. Qian, M.H. Engel, S.A. Macko, S. Carpenter, J.W. Deming, Kinetics of peptide hydrolysis and amino acid decomposition at high temperature, Geochimica et Cosmochimica Acta 57 (1993) 3281-3293.
[336] A. Rimola, P. Ugliengo, M. Sodupe, Formation versus hydrolysis of the peptide bond from a quantum-mechanical viewpoint: the role of mineral surfaces and implications for the origin of life, International journal of molecular sciences 10 (2009) 746-760.
[337] T.M. Lauer, G.P.F. Wood, D. Farkas, H.A. Sathish, H.S. Samra, B.L. Trout, Molecular Investigation of the Mechanism of Non-Enzymatic Hydrolysis of Proteins and the Predictive Algorithm for Susceptibility, Biochemistry 55 (2016) 3315-3328.
[338] A.L.L. East, On the hydrolysis mechanisms of amides and peptides, International Journal of Chemical Kinetics 50 (2018) 705-709.
[339] S. Antonczak, M.F. Ruiz-Lopez, J.L. Rivail, Ab initio analysis of water-assisted reaction mechanisms in amide hydrolysis, Journal of the American Chemical Society 116 (1994) 3912-3921.
[340] S. Ravuluri, R. Bansal, N. Chhabra, A.S. Rathore, Kinetics and Characterization of Non-enzymatic Fragmentation of Monoclonal Antibody Therapeutics, Pharmaceutical Research 35 (2018) 142.
[341] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, Journal of molecular graphics 14 (1996) 33-38.
[342] BIOVIA, Discovery Studio, Dassault Systemes, San Diego, 2022.
[343] B. Brumshtein, S.A.-O. Esswein, M.R. Sawaya, G. Rosenberg, A.T. Ly, M. Landau, D.A.-O. Eisenberg, Identification of two principal amyloid-driving segments in variable domains of Ig light chains in systemic light-chain amyloidosis.
[344] K.L. Zapadka, F.J. Becher, A.L. Gomes dos Santos, S.E. Jackson, Factors affecting the physical stability (aggregation) of peptide therapeutics, Interface Focus 7 (2017) 20170030.
[345] Q. Ma, P.J. Shuler, C.W. Aften, Y. Tang, Theoretical studies of hydrolysis and stability of polyacrylamide polymers, Polymer Degradation and Stability 121 (2015) 69-77.
[346] J.I. Mujika, J.M. Mercero, X. Lopez, Water-promoted hydrolysis of a highly twisted amide: Rate acceleration caused by the twist of the amide bond, Journal of the American Chemical Society 127 (2005) 4445-4453.
[347] R. Bischoff, H. Schlüter, Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications, Journal of Proteomics 75 (2012) 2275-2296.
[348] A. Brunelli, A synopsis of resampling techniques.

[349] J. Gruszczyk, R.K. Huang, L.-J. Chan, S. Menant, C. Hong, J.M. Murphy, Y.-F. Mok, M.D.W. Griffin, R.D. Pearson, W. Wong, A.F. Cowman, Z. Yu, W.-H. Tham, Cryo-EM structure of an essential Plasmodium vivax invasion complex, Nature 559 (2018) 135-139.
[350] S. Bangaru, H. Zhang, I.M. Gilchuk, T.G. Voss, R.P. Irving, P. Gilchuk, P. Matta, X. Zhu, S. Lang, T. Nieusma, J.A. Richt, R.A. Albrecht, H.A. Vanderven, R. Bombardi, S.J. Kent, A.B. Ward, I.A. Wilson, J.E. Crowe, A multifunctional human monoclonal neutralizing antibody that targets a unique conserved epitope on influenza HA, Nature Communications 9 (2018) 2669.
[351] S. Tan, K. Liu, Y. Chai, C.W.H. Zhang, S. Gao, G.F. Gao, J. Qi, Distinct PD-L1 binding characteristics of therapeutic monoclonal antibody durvalumab, Protein & Cell 9 (2018) 135-139.
[352] G. Scapin, X. Yang, W.W. Prosise, M. McCoy, P. Reichert, J.M. Johnston, R.S. Kashi, C. Strickland, Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab, Nature Structural & Molecular Biology 22 (2015) 953-958.
[353] T. Oie, G.H. Loew, S.K. Burt, J.S. Binkley, R.D. MacElroy, Quantum chemical studies of a model for peptide bond formation: formation of formamide and water from ammonia and formic acid, Journal of the American Chemical Society 104 (1982) 6169-6174.
[354] D. Chandler, Interfaces and the driving force of hydrophobic assembly, Nature 437 (2005) 640-647.
[355] S. Garde, A.J. Patel, Unraveling the hydrophobic effect, one molecule at a time, Proceedings of the National Academy of Sciences 108 (2011) 16491-16492.
[356] F.H. Stillinger, Water revisited, science 209 (1980) 451-457.
[357] A. Godec, F. Merzel, Physical Origin Underlying the Entropy Loss upon Hydrophobic Hydration, Journal of the American Chemical Society 134 (2012) 17574-17581.
[358] H. Ishikita, K. Saito, Proton transfer reactions and hydrogen-bond networks in protein environments, Journal of The Royal Society Interface 11 (2014) 20130518.
[359] C.J. Pace, J. Gao, Exploring and Exploiting Polar−π Interactions with Fluorinated Aromatic Amino Acids, Accounts of Chemical Research 46 (2013) 907-915.
[360] S. Mecozzi, A.P. West Jr, D.A. Dougherty, Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide, Proceedings of the National Academy of Sciences 93 (1996) 10566-10571.
[361] N.-Z. Xie, Q.-S. Du, J.-X. Li, R.-B. Huang, Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design, PloS one 10 (2015) e0137113-e0137113.
[362] P. Horx, A. Geyer, Comparing the Hinge-Type Mobility of Natural and Designed Intermolecular Bi-disulfide Domains, Frontiers in Chemistry 8 (2020).
[363] R. Nezlin, CHAPTER 1 - General Characteristics of Immunoglobulin Molecules, in: R. Nezlin (Ed.), The Immunoglobulins, Academic Press, New York, 1998, pp. 3-73.
[364] R. Najmanovich, J. Kuttner, V. Sobolev, M. Edelman, Side-chain flexibility in proteins upon ligand binding, Proteins: Structure, Function, and Bioinformatics 39 (2000) 261-268.
[365] F. Huang, W.M. Nau, A conformational flexibility scale for amino acids in peptides, Angewandte Chemie International Edition 42 (2003) 2269-2272.
[366] R. Vinu, L.J. Broadbelt, Unraveling Reaction Pathways and Specifying Reaction Kinetics for Complex Systems, Annual Review of Chemical and Biomolecular Engineering 3 (2012) 29-54.
[367] M.G. Evans, M. Polanyi, Inertia and driving force of chemical reactions, Transactions of the Faraday Society 34 (1938) 11-24.

[368] J.P. Stevens, Outliers and influential data points in regression analysis, Psychological bulletin 95 (1984) 334.
[369] C.P. Dhakal, Dealing with outliers and influential points while fitting regression, Journal of Institute of Science and Technology 22 (2017) 61-65.
[370] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, Journal of the Chemical Society, Perkin Transactions 2 (1993) 799-805.
[371] P. Schober, C. Boer, L.A. Schwarte, Correlation Coefficients: Appropriate Use and Interpretation, Anesthesia & Analgesia 126 (2018).
[372] R. Taylor, Interpretation of the Correlation Coefficient: A Basic Review, Journal of Diagnostic Medical Sonography 6 (1990) 35-39.
[373] J. Alcantara, R. Stix, K. Huang, A. Connor, R. East, V. Jaramillo-Martinez, E.J. Stollar, K.A. Ball, An Unbound Proline-Rich Signaling Peptide Frequently Samples Cis Conformations in Gaussian Accelerated Molecular Dynamics Simulations, Frontiers in Molecular Biosciences 8 (2021).
[374] D.E. Stewart, A. Sarkar, J.E. Wampler, Occurrence and role ofcis peptide bonds in protein structures, Journal of Molecular Biology 214 (1990) 253-260.
[375] M.S. Weiss, H.J. Metzner, R. Hilgenfeld, Two non-proline cis peptide bonds may be important for factor XIII function, FEBS Letters 423 (1998) 291-296.
[376] P. Craveur, A.P. Joseph, P. Poulain, A.G. de Brevern, J. Rebehmed, Cis–trans isomerization of omega dihedrals in proteins, Amino acids 45 (2013) 279-289.
[377] B.J. Finney Jl Fau - Gellatly, I.C. Gellatly Bj Fau - Golton, J. Golton Ic Fau - Goodfellow, J. Goodfellow, Solvent effects and polar interactions in the structural stability and dynamics of globular proteins.
[378] K. Son, C. Kwon, Stabilization of Human Epidermal Growth Factor (hEGF) in Aqueous Formulation, Pharmaceutical Research 12 (1995) 451-454.
[379] Z. Shahrokh, G. Eberlein, D. Buckley, M.V. Paranandi, D.W. Aswad, P. Stratton, R. Mischak, Y.J. Wang, Major Degradation Products of Basic Fibroblast Growth Factor: Detection of Succinimide and Iso-aspartate in Place of Aspartate15, Pharmaceutical Research 11 (1994) 936-944.
[380] A.A. Wakankar, R.T. Borchardt, C. Eigenbrot, S. Shia, Y.J. Wang, S.J. Shire, J.L. Liu, Aspartate Isomerization in the Complementarity-Determining Regions of Two Closely Related Monoclonal Antibodies, Biochemistry 46 (2007) 1534-1544.
[381] G. Teshima, J. Porter, K. Yim, V. Ling, A. Guzzetta, Deamidation of soluble CD4 at asparagine-52 results in reduced binding capacity for the HIV-1 envelope glycoprotein gp120, Biochemistry 30 (1991) 3916-3922.
[382] Y.D. Liu, J.Z. van Enk, G.C. Flynn, Human antibody Fc deamidation in vivo, Biologicals 37 (2009) 313-322.
[383] C.M. Eakin, A. Miller, J. Kerr, J. Kung, A. Wallace, Assessing analytical methods to monitor isoAsp formation in monoclonal antibodies, Frontiers in Pharmacology 5 (2014).
[384] J. Kuang, Y. Tao, Y. Song, L. Chemmalil, N. Mussa, J. Ding, Z.J. Li, Understanding the pathway and kinetics of aspartic acid isomerization in peptide mapping methods for monoclonal antibodies, Analytical and Bioanalytical Chemistry 413 (2021) 2113+.
[385] J.O. Hui, T. Flick, J.A. Loo, I.D.G. Campuzano, Unequivocal Identification of Aspartic Acid and isoAspartic Acid by MALDI-TOF/TOF: From Peptide Standards to a Therapeutic Antibody, Journal of the American Society for Mass Spectrometry 32 (2021) 1901-1909.

[386] M.V. Paranandi, A.W. Guzzetta, W.S. Hancock, D.W. Aswad, Deamidation and isoaspartate formation during in vitro aging of recombinant tissue plasminogen activator, Journal of Biological Chemistry 269 (1994) 243-253.
[387] K. Kato, T. Nakayoshi, E. Kurimoto, A. Oda, Mechanisms of Deamidation of Asparagine Residues and Effects of Main-Chain Conformation on Activation Energy. LID - 10.3390/ijms21197035 [doi] LID - 7035.
[388] M. Harding, The architecture of metal coordination groups in proteins, Acta Crystallographica Section D 60 (2004) 849-859.
[389] H.I. Okur, J. Kherb, P.S. Cremer, Cations Bind Only Weakly to Amides in Aqueous Solutions, Journal of the American Chemical Society 135 (2013) 5062-5067.
[390] K. Lewinski, L. Lebioda, Preferred geometry of cation-amide bonding. Crystal structures of [Mg(NMA)2(H2O)4](NO3)2 and Ca(NO3)2.4NMA (NMA = N-methylacetamide), Journal of the American Chemical Society 108 (1986) 3693-3696.
[391] J. Zhao, T. Dong, P. Yu, J. Wang, Conformation and Metal Cation Binding of Zwitterionic Alanine Tripeptide in Saline Solutions by Infrared Vibrational Spectroscopy and Molecular Dynamics Simulations, The Journal of Physical Chemistry B 126 (2022) 161-173.
[392] S. Veeralakshmi, S. Nehru, S. Arunachalam, P. Kumar, M. Govindaraju, Study of single and double chain surfactant–cobalt (III) complexes and their hydrophobicity, micelle formation, interaction with serum albumins and antibacterial activities, Inorganic Chemistry Frontiers 1 (2014) 393-404.
[393] S.V. Wegner, J.P. Spatz, Cobalt (III) as a stable and inert mediator ion between NTA and His6-tagged proteins, Angewandte Chemie International Edition 52 (2013) 7593-7596.
[394] Y. Zhang, C. Schöneich, Visible Light Induces Site-Specific Oxidative Heavy Chain Fragmentation of a Monoclonal Antibody (IgG1) Mediated by an Iron (III)-Containing Histidine Buffer, Molecular Pharmaceutics (2022).
[395] M. Engelhard, B. Pevec, B. Hess, Modification of two peptides of bacteriorhodopsin with a (pentaammine) cobalt (III) complex, Biochemistry 28 (1989) 5432-5438.
[396] S. Rafique, M. Idrees, A. Nasim, H. Akbar, A. Athar, Transition metal complexes as potential therapeutic agents, Biotechnology and Molecular Biology Reviews 5 (2010) 38-45.
[397] M.A. Zoroddu, H. Kowalik-Jankowska T Fau - Kozlowski, K. Kozlowski H Fau - Salnikow, M. Salnikow K Fau - Costa, M. Costa, Ni(II) and Cu(II) binding with a 14-aminoacid sequence of Cap43 protein, TRSRSHTSEGTRSR.
[398] T. Shutova, G. Nikitina J Fau - Deikus, B. Deikus G Fau - Andersson, V. Andersson B Fau - Klimov, G. Klimov V Fau - Samuelsson, G. Samuelsson, Structural dynamics of the manganese-stabilizing protein-effect of pH, calcium, and manganese.
[399] A. Simeonov, Recent developments in the use of differential scanning fluorometry in protein and small molecule discovery and characterization, Expert opinion on drug discovery 8 (2013) 1071-1082.
[400] S.I. Mann, A. Nayak, G.T. Gassner, M.J. Therien, W.F. DeGrado, De Novo Design, Solution Characterization, and Crystallographic Structure of an Abiological Mn–Porphyrin-Binding Protein Capable of Stabilizing a Mn (V) Species, Journal of the American Chemical Society 143 (2020) 252-259.
[401] M.A. Halcrow, The Effect of Ligand Design on Metal Ion Spin State—Lessons from Spin Crossover Complexes, Crystals, 2016.
[402] K.-J. Appenroth, Definition of “heavy metals” and their role in biological systems, Soil heavy metals (2010) 19-29.

[403] E. Nieboer, D.H.S. Richardson, The replacement of the nondescript term 'heavy metals' by a biologically and chemically significant classification of metal ions, Environmental Pollution Series B, Chemical and Physical 1 (1980) 3-26.
[404] M.E. Maguire, J.A. Cowan, Magnesium chemistry and biochemistry, Biometals 15 (2002) 203-210.
[405] A.C. Warden, L. Hearn Mt Fau - Spiccia, L. Spiccia, Novel acetate binding modes in [Na2Cu(CH3COO)4(H2O)].H2O.
[406] W. Buijs, I.A. Hussein, M. Mahmoud, A.T. Onawole, M.A. Saad, G.R. Berdiyorov, Molecular Modeling Study toward Development of H2S-Free Removal of Iron Sulfide Scale from Oil and Gas Wells, Industrial & Engineering Chemistry Research 57 (2018) 10095-10104.
[407] M.-S. Eum, C.S. Chin, S.Y. Kim, C. Kim, S.K. Kang, N.H. Hur, J.H. Seo, G.Y. Kim, Y.K. Kim, Ligand effects on luminescence of new type blue light-emitting mono (2-phenylpyridinato) iridium (III) complexes, Inorganic chemistry 47 (2008) 6289-6295.
[408] S. Hein, M. Jähnig, N. Kannengießer, J. Pape, T. Laporte, G. Schnakenburg, R.K. Kremer, W. Urland, R. Glaum, The weak ligand field in lanthanoid(III) hydrogensulfate-sulfates, Zeitschrift für anorganische und allgemeine Chemie 649 (2023) e202200342.
[409] L. Rulíšek, Z. Havlas, Ab Initio Calculations of Monosubstituted (CH3OH, CH3SH, NH3) Hydrated Ions of Zn2+ and Ni2+, The Journal of Physical Chemistry A 103 (1999) 1634-1639.
[410] S. Bestgen, C. Schoo, B.L. Neumeier, T.J. Feuerstein, C. Zovko, R. Köppe, C. Feldmann, P.W. Roesky, Intensely Photoluminescent Diamidophosphines of the Alkaline-Earth Metals, Aluminum, and Zinc, Angewandte Chemie International Edition 57 (2018) 14265-14269.
[411] E. Pluhařová, M.D. Baer, C.J. Mundy, B. Schmidt, P. Jungwirth, Aqueous Cation-Amide Binding: Free Energies and IR Spectral Signatures by Ab Initio Molecular Dynamics, The Journal of Physical Chemistry Letters 5 (2014) 2235-2240.
[412] C. Bălăiu, A. Lupan, R. Bruce King, Combining a weak-field rigid chelating bidentate dicarbene ligand with a strong-field carbonyl ligand in binuclear cyclopentadienyliron carbonyl derivatives, Polyhedron 219 (2022) 115778.
[413] G.H. Cohen, J.L. Hoard, The Structure of the Seven-Coordinate trans-1, 2-Diaminocyclohexane-N, N'-tetraacetatoaquoferrate (III) Ion in Crystals of the Calcium Salt1, 2, Journal of the American Chemical Society 88 (1966) 3228-3234.
[414] N.E. Kime, J.A. Ibers, The Co (III)–N bond length in relation to the Co (II)–N bond length. The crystal structure of hexaamminecobalt (III) iodide [Co (NH3) 6] I3, Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry 25 (1969) 168-169.
[415] J.P. Collman, R.R. Gagne, C.A. Reed, W.T. Robinson, G.A. Rodley, Structure of an Iron(II) Dioxygen Complex; A Model for Oxygen Carrying Hemeproteins, Proceedings of the National Academy of Sciences 71 (1974) 1326-1329.
[416] M. Nara, H. Torii, M. Tasumi, Correlation between the Vibrational Frequencies of the Carboxylate Group and the Types of Its Coordination to a Metal Ion: An ab Initio Molecular Orbital Study, The Journal of Physical Chemistry 100 (1996) 19812-19817.
[417] R. Gupta, T. Taguchi, B. Lassalle-Kaiser, E.L. Bominaar, J. Yano, M.P. Hendrich, A.S. Borovik, High-spin Mn-oxo complexes and their relevance to the oxygen-evolving complex within photosystem II.
[418] R. Gupta, C.E. MacBeth, V.G. Young, A.S. Borovik, Isolation of Monomeric MnIII/II−OH and MnIII−O Complexes from Water: Evaluation of O−H Bond Dissociation Energies, Journal of the American Chemical Society 124 (2002) 1136-1137.

[419] B. Pandey, A. Ansari, N. Vyas, G. Rajaraman, Structures, bonding and reactivity of iron and manganese high-valent metal-oxo complexes: A computational investigation, Journal of Chemical Sciences 127 (2015) 343-352.
[420] N. Nabil, O.M.I. Adly, M. Shebl, A. Taha, F. Samy, Ni II and Co II binary and ternary complexes of 3-formylchromone: spectroscopic characterization, antimicrobial activities, docking and modeling studies, RSC advances 12 (2022) 29939-29958.
[421] D.K. Garner, R.A. Allred, K.J. Tubbs, A.M. Arif, L.M. Berreau, Synthesis and Characterization of Mononuclear Zinc Aryloxide Complexes Supported by Nitrogen/Sulfur Ligands Possessing an Internal Hydrogen Bond Donor, Inorganic Chemistry 41 (2002) 3533-3541.
[422] M. Nara, M. Morii H Fau - Tanokura, M. Tanokura, Coordination to divalent cations by calcium-binding proteins studied by FTIR spectroscopy.
[423] J. Zhao, T. Dong, P. Yu, J. Wang, Conformation and metal cation binding of zwitterionic alanine tripeptide in saline solutions by infrared vibrational spectroscopy and molecular dynamics simulations, The Journal of Physical Chemistry B 126 (2021) 161-173.
[424] G.B. Deacon, R.J. Phillips, Relationships between the carbon-oxygen stretching frequencies of carboxylato complexes and the type of carboxylate coordination, Coordination chemistry reviews 33 (1980) 227-250.
[425] K. Nakamoto, Infrared and Raman Spectra of Inorganic and Coordination Compounds: Part A: Theory and Applications in Inorganic Chemistry.
[426] A.D. McNaught, A. Wilkinson, Compendium of chemical terminology, Blackwell Science Oxford1997.
[427] R.S. Devan, R.A. Patil, J.H. Lin, Y.R. Ma, One-dimensional metal-oxide nanostructures: recent developments in synthesis, characterization, and applications, Advanced Functional Materials 22 (2012) 3326-3370.
[428] M.J. Limo, A. Sola-Rabada, E. Boix, V. Thota, Z.C. Westcott, V. Puddu, C.C. Perry, Interactions between Metal Oxides and Biomolecules: from Fundamental Understanding to Applications, Chemical Reviews 118 (2018) 11118-11193.
[429] J. Lee, M. Christopher Orilall, S.C. Warren, M. Kamperman, F.J. Disalvo, U. Wiesner, Direct access to thermally stable and highly crystalline mesoporous transition-metal oxides with uniform pores, Nature materials 7 (2008) 222-228.
[430] V.M. Dadarlat, R.D. Skeel, Dual Role of Protein Phosphorylation in DNA Activator/Coactivator Binding, Biophysical Journal 100 (2011) 469-477.
[431] A. Del Rio, M. Sgobba, M.D. Parenti, G. Degliesposti, R. Forestiero, C. Percivalle, P.F. Conte, M. Freccero, G. Rastelli, A computational workflow for the design of irreversible inhibitors of protein kinases, Journal of Computer-Aided Molecular Design 24 (2010) 183-194.
[432] A.M. Dos Santos, A.R.S. Oliveira, C.H.S. da Costa, P.W. Kenny, C.A. Montanari, J.d.J.G.J. Varela, J. Lameira, Assessment of Reversibility for Covalent Cysteine Protease Inhibitors Using Quantum Mechanics/Molecular Mechanics Free Energy Surfaces, Journal of Chemical Information and Modeling 62 (2022) 4083-4094.
[433] T. Hansson, J. Åqvist, Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations, Protein Engineering, Design and Selection 8 (1995) 1137-1144.
[434] H. Hata, D. Phuoc Tran, M. Marzouk Sobeh, A. Kitao, Binding free energy of protein/ligand complexes calculated using dissociation Parallel Cascade Selection Molecular Dynamics and Markov state model.

[435] J.S. Patel, F.M. Ytreberg, Fast Calculation of Protein–Protein Binding Free Energies Using Umbrella Sampling with a Coarse-Grained Model, Journal of Chemical Theory and Computation 14 (2018) 991-997.
[436] J.W. Perthold, C. Oostenbrink, Simulation of Reversible Protein–Protein Binding and Calculation of Binding Free Energies Using Perturbed Distance Restraints, Journal of Chemical Theory and Computation 13 (2017) 5697-5708.
[437] S. Wan, A.P. Bhati, S.J. Zasada, P.V. Coveney, Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction, Interface Focus 10 (2020) 20200007.
[438] S. Wan, B. Knapp, D.W. Wright, C.M. Deane, P.V. Coveney, Rapid, Precise, and Reproducible Prediction of Peptide-MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment.
[439] H.-J. Woo, B. Roux, Calculation of absolute protein–ligand binding free energy from computer simulations, Proceedings of the National Academy of Sciences 102 (2005) 6825-6830.
[440] P.L. Kastritis, A.M.J.J. Bonvin, Are scoring functions in protein− protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark, Journal of proteome research 9 (2010) 2216-2225.
[441] P.A. Jackson, J.C. Widen, D.A. Harki, K.M. Brummond, Covalent modifiers: A chemical perspective on the reactivity of α, β-unsaturated carbonyls with thiols via hetero-Michael addition reactions, Journal of medicinal chemistry 60 (2017) 839-885.
[442] E. Awoonor-Williams, A.G. Walsh, C.N. Rowley, Modeling covalent-modifier drugs, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1865 (2017) 1664-1675.
[443] D.S. Johnson, B.F. Weerapana E Fau - Cravatt, B.F. Cravatt, Strategies for discovering and derisking covalent, irreversible enzyme inhibitors.
[444] R.A.-O. Fossmark, T.C. Martinsen, H.A.-O. Waldum, Adverse Effects of Proton Pump Inhibitors-Evidence and Plausibility. LID - 10.3390/ijms20205203 [doi] LID - 5203.
[445] X. Du, Y. Li, Y.L. Xia, S.M. Ai, J. Liang, P. Sang, X.L. Ji, S.Q. Liu, Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. LID - 10.3390/ijms17020144 [doi] LID - 144.
[446] Q. Tan, Y. Ding, Z. Qiu, J. Huang, Binding Energy and Free Energy of Calcium Ion to Calmodulin EF-Hands with the Drude Polarizable Force Field, ACS Physical Chemistry Au 2 (2022) 143-155.
[447] W.J. Song, P.A. Sontz, X.I. Ambroggio, F.A. Tezcan, Metals in Protein–Protein Interfaces, Annual Review of Biophysics 43 (2014) 409-431.
[448] A.T. Onawole, I.A. Hussein, A. Sultan, S. Abdel-Azeim, M. Mahmoud, M.A. Saad, Molecular and electronic structure elucidation of Fe^{2+}/Fe^{3+} complexed chelators used in iron sulphide scale removal in oil and gas wells, The Canadian Journal of Chemical Engineering 97 (2019) 2021-2027.
[449] T. Dudev, C. Lim, Effect of Carboxylate-Binding Mode on Metal Binding/Selectivity and Function in Proteins, Accounts of Chemical Research 40 (2007) 85-93.
[450] A.K. Katz, J.P. Glusker, S.A. Beebe, C.W. Bock, Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc, Journal of the American Chemical Society 118 (1996) 5752-5763.
[451] N. Gangwar, R. Mishra, N. Budholiya, A.S. Rathore, Effect of vitamins and metal ions on productivity and charge heterogeneity of IgG1 expressed in CHO cells, Biotechnology Journal 16 (2021) 2000464.

[452] Q.-Y. He, A.B. Mason, R.C. Woodworth, B.M. Tam, T. Wadsworth, R.T.A. MacGillivray, Effects of Mutations of Aspartic Acid 63 on the Metal-Binding Properties of the Recombinant N-Lobe of Human Serum Transferrin, Biochemistry 36 (1997) 5522-5528.
[453] A. Mason, B. He Qy Fau - Tam, R.A. Tam B Fau - MacGillivray, R. MacGillivray Ra Fau - Woodworth, R. Woodworth, Mutagenesis of the aspartic acid ligands in human serum transferrin: lobe-lobe interaction and conformation as revealed by antibody, receptor-binding and iron-release studies.
[454] H.M. Baker, B.F. Anderson, E.N. Baker, Dealing with iron: Common structural principles in proteins that transport iron and heme, Proceedings of the National Academy of Sciences 100 (2003) 3579-3583.
[455] J.D. Carballeira, J. Fernandez-Lucas, M.A. Quezada, M.J. Hernaiz, A.R. Alcantara, Y. Simeó, J.V. Sinisterra, Biotransformations, in: M. Schaechter (Ed.), Encyclopedia of Microbiology (Third Edition), Academic Press, Oxford, 2009, pp. 212-251.
[456] A. Marx, M. Galilee, A. Alian, Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes.
[457] Z. Liu, S. Frutos, M.J. Bick, M. Vila-Perelló, G.T. Debelouchina, S.A. Darst, T.W. Muir, Structure of the branched intermediate in protein splicing, Proceedings of the National Academy of Sciences 111 (2014) 8422-8427.
[458] M. Mandado, M.N.D.S. Cordeiro, On the stability of metal–aminoacid complexes in water based on water–ligand exchange reactions and electronic properties: Detailed study on iron–glycine hexacoordinated complexes, Journal of Computational Chemistry 31 (2010) 2735-2745.
[459] T. Scherpf, C.R. Carr, L.J. Donnelly, Z.S. Dubrawski, B.S. Gelfand, W.E. Piers, A Mesoionic Carbene–Pyridine Bidentate Ligand That Improves Stability in Electrocatalytic CO2 Reduction by a Molecular Manganese Catalyst, Inorganic Chemistry 61 (2022) 13644-13656.
[460] S. Sommerfeld, J. Strube, Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies, Chemical Engineering and Processing: Process Intensification 44 (2005) 1123-1137.
[461] M. Vázquez-Rey, D.A. Lang, Aggregates in monoclonal antibody manufacturing processes, Biotechnology and Bioengineering 108 (2011) 1494-1508.
[462] B. Somasundaram, K. Pleitt, E. Shave, K. Baker, L.H.L. Lua, Progression of continuous downstream processing of monoclonal antibodies: Current trends and challenges, Biotechnology and Bioengineering 115 (2018) 2893-2907.
[463] R.L. Fahrner, H.L. Knudsen, C.D. Basey, W. Galan, D. Feuerhelm, M. Vanderlaan, G.S. Blank, Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes, Biotechnology and Genetic Engineering Reviews 18 (2001) 301-327.
[464] E. Mahajan, A. George, B. Wolk, Improving affinity chromatography resin efficiency using semi-continuous chromatography, Journal of chromatography A 1227 (2012) 154-162.
[465] T. Breece, R. Fahrner, J. Gorrell, K. Lazzareschi, P. Lester, D. Peng, Protein purification, in: E.P. Office (Ed.) World Intellectural Property Organization, Genetec, Inc, 2003.
[466] K. Brorson, S. Krejci, K. Lee, E. Hamilton, K. Stein, Y. Xu, Bracketed generic inactivation of rodent retroviruses by low pH treatment for monoclonal antibodies and recombinant proteins, Biotechnology and Bioengineering 82 (2003) 321-329.
[467] Q. Chen, Viral clearance using traditional, well-understood unit operations (session I): low-pH inactivation, PDA Journal of Pharmaceutical Science and Technology 68 (2014) 17-22.

[468] A. Frauenschuh, K. Bill, Wash solution and method for affinity chromatography, in: W.I.P.O. (WIPO) (Ed.) World Intellectual Property Organization (WIPO), Novartis AG, 2011.
[469] I.C.H.H.T. Guideline, viral safety evaluation of biotechnology products derived from cell lines of human or animal origin Q5A (R1), Current Step 4 (1997).
[470] Y. Imura, T. Tagawa, Y. Miyamoto, S. Nonoyama, H. Sumichika, Y. Fujino, M. Yamanouchi, H. Miki, Washing with alkaline solutions in protein A purification improves physicochemical properties of monoclonal antibodies, Scientific Reports 11 (2021) 1-11.
[471] J. Mattila, M. Clark, S. Liu, J. Pieracci, T.R. Gervais, E. Wilson, O. Galperina, X. Li, D. Roush, K. Zoeller, Retrospective evaluation of low-pH viral inactivation and viral filtration data from a multiple company collaboration, PDA journal of pharmaceutical science and technology 70 (2016) 293-299.
[472] G. Miesegaes, S. Lute, K. Brorson, Analysis of viral clearance unit operations for monoclonal antibodies, Biotechnology and Bioengineering 106 (2010) 238-246.
[473] A.A. Shukla, P. Hinckley, Host cell protein clearance during protein A chromatography: development of an improved column wash step.
[474] A.A. Shukla, P. Hinckley, Host cell protein clearance during protein a chromatography: Development of an improved column wash step, Biotechnology Progress 24 (2008) 1115-1121.
[475] O. World Health, Guidelines on viral inactivation and removal procedures intended to assure the viral safety of human blood plasma products, WHO Technical Report, Series 924 (2004) 150-224.
[476] S.X. Gao, Y. Zhang, K. Stansberry-Perkins, A. Buko, S. Bai, V. Nguyen, M.L. Brader, Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity, Biotechnology and bioengineering 108 (2011) 977-982.
[477] M.F. García-Mayoral, R. del Pozo Am Fau - Campos-Olivas, J.G. Campos-Olivas R Fau - Gavilanes, J. Gavilanes Jg Fau - Santoro, M. Santoro J Fau - Rico, D.V. Rico M Fau - Laurents, M. Laurents Dv Fau - Bruix, M. Bruix, pH-Dependent conformational stability of the ribotoxin alpha-sarcin and four active site charge substitution variants.
[478] S. Sato, D.P. Raleigh, pH-dependent Stability and Folding Kinetics of a Protein with an Unusual α–β Topology: The C-terminal Domain of the Ribosomal Protein L9, Journal of molecular biology 318 (2002) 571-582.
[479] H.A.-O. Sjuts, H. Schreuder, C.K. Engel, T. Bussemer, Y. Gokarn, Matching pH values for antibody stabilization and crystallization suggest rationale for accelerated development of biotherapeutic drugs.
[480] G. Gaza-Bulseco, H. Liu, Fragmentation of a Recombinant Monoclonal Antibody at Various pH, Pharmaceutical Research 25 (2008) 1881-1890.
[481] P.P. Nugrahadi, W.L.J. Hinrichs, H.W. Frijlink, C. Schöneich, C. Avanti, Designing Formulation Strategies for Enhanced Stability of Therapeutic Peptides in Aqueous Solutions: A Review, Pharmaceutics, 2023.
[482] S. Kumar, R. Nussinov, Close-range electrostatic interactions in proteins.
[483] S. Sokalingam, B. Madan, G. Raghunathan, S.-G. Lee, In silico study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability, Biotechnology and bioprocess engineering 18 (2013) 18-26.
[484] L. Serrano, A. Horovitz, B. Avron, M. Bycroft, A.R. Fersht, Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles, Biochemistry 29 (1990) 9343-9352.

[485] H.A.-O. Zhou, X. Pang, Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation.
[486] C. Nick Pace, R.W. Alston, K.L. Shaw, Charge–charge interactions influence the denatured state ensemble and contribute to protein stability, Protein Sci 9 (2000) 1395-1398.
[487] I. André, S. Linse, F.A.A. Mulder, Residue-Specific pKa Determination of Lysine and Arginine Side Chains by Indirect 15N and 13C NMR Spectroscopy: Application to apo Calmodulin, Journal of the American Chemical Society 131 (2009) 6640-6640.
[488] D.G. Isom, C.A. Castañeda, B.R. Cannon, B. García-Moreno E, Large shifts in pKa values of lysine residues buried inside a protein, Proceedings of the National Academy of Sciences 108 (2011) 5260-5265.
[489] P.E. Bowling, D.R. Broderick, J.M. Herbert, Fragment-Based Calculations of Enzymatic Thermochemistry Require Dielectric Boundary Conditions, The Journal of Physical Chemistry Letters 14 (2023) 3826-3834.
[490] J. Shlens, A tutorial on principal component analysis, arXiv preprint arXiv:1404.1100 (2014).
[491] Zach, PCA.R, in: Statology (Ed.) R-Guides, Statology, Github.com, 2020.
[492] N.R. Draper, H. Smith, Applied regression analysis, John Wiley & Sons1998.
[493] D. Brzyski, A. Gossmann, W. Su, M. Bogdan, Group SLOPE - adaptive selection of groups of predictors, J Am Stat Assoc 114 (2019) 419-433.
[494] D.B. Suits, Use of Dummy Variables in Regression Equations, J Am Stat Assoc 52 (1957) 548-551.
[495] Zach, How to Create Dummy Variables in R (Step-by-Step), Statology, 2021.
[496] I. André, S. Linse, F.A.A. Mulder, Residue-Specific pKa Determination of Lysine and Arginine Side Chains by Indirect 15N and 13C NMR Spectroscopy: Application to apo Calmodulin, Journal of the American Chemical Society 129 (2007) 15805-15813.
[497] M. Mozurkewich, S.W. Benson, Negative activation energies and curved Arrhenius plots. 1. Theory of reactions over potential wells, The Journal of Physical Chemistry 88 (1984) 6429-6435.
[498] Y. Takayama, C.A. Castañeda, M. Chimenti, B. García-Moreno, J. Iwahara, Direct Evidence for Deprotonation of a Lysine Side Chain Buried in the Hydrophobic Core of a Protein, Journal of the American Chemical Society 130 (2008) 6714-6715.
[499] F. Persson, B. Halle, How amide hydrogens exchange in native proteins, Proc Natl Acad Sci U S A 112 (2015) 10383-10388.
[500] Y. Bai, J.S. Milne, L. Mayne, S.W. Englander, Primary structure effects on peptide group hydrogen exchange, Proteins 17 (1993) 75-86.
[501] R.S. Molday, S.W. Englander, R.G. Kallen, Primary structure effects on peptide group hydrogen exchange, Biochemistry 11 (1972) 150-158.
[502] D.G. Isom, C.A. Castañeda, B.R. Cannon, P.D. Velu, B. García-Moreno E, Charges in the hydrophobic interior of proteins, Proceedings of the National Academy of Sciences 107 (2010) 16096-16100.
[503] H.-X. Zhou, X. Pang, Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation, Chemical reviews 118 (2018) 1691-1741.
[504] K.B. Schowen, H.H. Limbach, G.S. Denisov, R.L. Schowen, Hydrogen bonds and proton transfer in general-catalytic transition-state stabilization in enzyme catalysis, Biochimica et Biophysica Acta (BBA) - Bioenergetics 1458 (2000) 43-62.
[505] D. Zahn, Car–Parrinello molecular dynamics simulation of base-catalyzed amide hydrolysis in aqueous solution, Chemical Physics Letters 383 (2004) 134-137.

[506] D. Zahn, On the role of water in amide hydrolysis, European Journal of Organic Chemistry 2004 (2004) 4020-4023.
[507] L.N. Krasnoperov, J. Peng, P. Marshall, Modified Transition State Theory and Negative Apparent Activation Energies of Simple Metathesis Reactions: Application to the Reaction CH3 + HBr → CH4 + Br, The Journal of Physical Chemistry A 110 (2006) 3110-3120.
[508] J.M. Nicovich, C.A. Van Dijk, K.D. Kreutter, P.H. Wine, Kinetics of the reactions of alkyl radicals with hydrogen bromide and deuterium bromide, The Journal of Physical Chemistry 95 (1991) 9890-9896.
[509] P.W. Seakins, M.J. Pilling, J.T. Niiranen, D. Gutman, L.N. Krasnoperov, Kinetics and thermochemistry of R+ hydrogen bromide. dblarw. RH+ bromine atom reactions: determinations of the heat of formation of ethyl, isopropyl, sec-butyl and tert-butyl radicals, The Journal of Physical Chemistry 96 (1992) 9847-9855.
[510] D.M. Hawkins, The Problem of Overfitting, Journal of Chemical Information and Computer Sciences 44 (2004) 1-12.
[511] M.A. Babyak, What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models, Psychosomatic medicine 66 (2004) 411-421.
[512] Z. Zhang, Too much covariates in a multivariable model may cause the problem of overfitting.
[513] S. Lindman, W.-F. Xue, O. Szczepankiewicz, M.C. Bauer, H. Nilsson, S. Linse, Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability, Biophysical Journal 90 (2006) 2911-2921.
[514] E.C. Cook, T.P. Creamer, Influence of electrostatic forces on the association kinetics and conformational ensemble of an intrinsically disordered protein, Proteins: Structure, Function, and Bioinformatics 88 (2020) 1607-1619.
[515] C.N. Pace, G.R. Grimsley, J.M. Scholtz, Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility*, Journal of Biological Chemistry 284 (2009) 13285-13289.
[516] D. Dell'Orco, W.-F. Xue, E. Thulin, S. Linse, Electrostatic Contributions to the Kinetics and Thermodynamics of Protein Assembly, Biophysical Journal 88 (2005) 1991-2002.
[517] X. Pang, H.-X. Zhou, Rate Constants and Mechanisms of Protein–Ligand Binding, Annual Review of Biophysics 46 (2017) 105-130.
[518] G. Schreiber, G. Haran, H.X. Zhou, Fundamental Aspects of Protein−Protein Association Kinetics, Chemical Reviews 109 (2009) 839-860.
[519] K.H. Roux, L. Strelets, T.E. Michaelsen, Flexibility of human IgG subclasses, Journal of immunology (Baltimore, Md.: 1950) 159 (1997) 3372-3382.
[520] R. Jefferis, J. Lund, Interaction sites on human IgG-Fc for FcγR: current models, Immunology Letters 82 (2002) 57-65.
[521] U.J. Harris, S.B. Larson, A. McPherson, Comparison of intact antibody structures and the implications for effector function, Advances in immunology 72 (1999) 191-208.
[522] L.M. Amzel, R.J. Poljak, Three-Dimensional Structure of Immunoglobulins, Annual Review of Biochemistry 48 (1979) 961-997.
[523] R. Senthil, S. Usha, K.M. Saravanan, Importance of Fluctuating Amino Acid Residues in Folding and Binding of Proteins.
[524] A.M. Ruvinsky, I.A. Vakser, Sequence composition and environment effects on residue fluctuations in protein structures.

[525] D.F. Hunt, J.R. Yates 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proceedings of the National Academy of Sciences 83 (1986) 6233-6237.
[526] L.A. Breci, D.L. Tabb, J.R. Yates, V.H. Wysocki, Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra, Analytical Chemistry 75 (2003) 1963-1971.
[527] J.A. Loo, C.G. Edmonds, R.D. Smith, Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization, Analytical Chemistry 65 (1993) 425-438.
[528] D.J. Foreman, E.T. Dziekonski, S.A. McLuckey, Maximizing Selective Cleavages at Aspartic Acid and Proline Residues for the Identification of Intact Proteins, Journal of The American Society for Mass Spectrometry 30 (2019) 34-44.
[529] E.A. Kapp, F. Schütz, G.E. Reid, J.S. Eddes, R.L. Moritz, R.A.J. O'Hair, T.P. Speed, R.J. Simpson, Mining a Tandem Mass Spectrometry Database To Determine the Trends and Global Factors Influencing Peptide Fragmentation, Analytical Chemistry 75 (2003) 6251-6264.
[530] N. Spitzner, F. Löhr, S. Pfeiffer, A. Koumanov, A. Karshikoff, H. Rüterjans, Ionization properties of titratable groups in ribonuclease T1, European Biophysics Journal 30 (2001) 186-197.
[531] M.A.S. Hass, F.A.A. Mulder, Contemporary NMR Studies of Protein Electrostatics, Annual Review of Biophysics 44 (2015) 53-75.
[532] P. Linus, Nature of the Chemical Bond and the Structure of Molecules and Crystals: an Introduction to Modern Structural Chemistry, Oxford & IBH Publishing Company, 1960.
[533] C.A.-O. Tantardini, A.A.-O. Oganov, Thermochemical electronegativities of the elements.
[534] Z. Keita, PCA_Protein_Data, in: Datacamp (Ed.) Learn R: Principal Component Analysis in R Tutorial, Datacamp, https://app.datacamp.com/workspace/w/6422b279-e9b8-41e5-9491-8f64e35f9250, 2023, pp. Training code to teach individuals how to apply PCA through R coding.
[535] S.W. Englander, Hydrogen exchange and mass spectrometry: A historical perspective.
[536] E. Liepinsh, G. Otting, Proton exchange rates from amino acid side chains—implications for image contrast, Magnetic resonance in medicine 35 (1996) 30-42.
[537] Y. Bai, J.J. Englander, L. Mayne, J.S. Milne, S.W. Englander, [15] Thermodynamic parameters from hydrogen exchange measurements, Methods in Enzymology, Academic Press1995, pp. 344-356.
[538] D. Nguyen, L. Mayne, M.C. Phillips, S. Walter Englander, Reference Parameters for Protein Hydrogen Exchange Rates, Journal of the American Society for Mass Spectrometry 29 (2018) 1936-1939.
[539] J.P. Schessner, E. Voytik, I. Bludau, A practical guide to interpreting and generating bottom-up proteomics data visualizations, PROTEOMICS 22 (2022) 2100103.
[540] C.N. Pace, K. Fu H Fau - Lee Fryar, J. Lee Fryar K Fau - Landua, S.R. Landua J Fau - Trevino, D. Trevino Sr Fau - Schell, R.L. Schell D Fau - Thurlkill, S. Thurlkill Rl Fau - Imura, J.M. Imura S Fau - Scholtz, K. Scholtz Jm Fau - Gajiwala, J. Gajiwala K Fau - Sevcik, L. Sevcik J Fau - Urbanikova, J.K. Urbanikova L Fau - Myers, K. Myers Jk Fau - Takano, E.J. Takano K Fau - Hebert, B.A. Hebert Ej Fau - Shirley, G.R. Shirley Ba Fau - Grimsley, G.R. Grimsley, Contribution of hydrogen bonds to protein stability.
[541] S. Kumar, R. Nussinov, Salt bridge stability in monomeric proteins11Edited by J. M. Thornton, Journal of Molecular Biology 293 (1999) 1241-1255.
[542] A. Nayek, P.S. Sen Gupta, S. Banerjee, B. Mondal, A.K. Bandyopadhyay, Salt-Bridge Energetics in Halophilic Proteins, PLOS ONE 9 (2014) e93862.

[543] A.K. Bandyopadhyay, R.N.U. Islam, D. Mitra, S. Banerjee, A. Goswami, Stability of buried and networked salt-bridges (BNSB)in thermophilic proteins.
[544] A.K. Bandyopadhyay, R.N. Ul Islam, N. Hazra, Salt-bridges in the microenvironment of stable protein structures.
[545] M. Baumgart, M. Röpke, M.E. Mühlbauer, S. Asami, S.L. Mader, K. Fredriksson, M. Groll, A.P. Gamiz-Hernandez, V.R.I. Kaila, Design of buried charged networks in artificial proteins, Nature Communications 12 (2021) 1895.
[546] R.N. Grewal, H. El Aribi, A.G. Harrison, K.W.M. Siu, A.C. Hopkinson, Fragmentation of Protonated Tripeptides: The Proline Effect Revisited, The Journal of Physical Chemistry B 108 (2004) 4899-4908.
[547] A.G. Harrison, The gas-phase basicities and proton affinities of amino acids and peptides, Mass Spectrometry Reviews 16 (1997) 201-217.
[548] C. Bleiholder, S. Suhai, B. Paizs, Revising the proton affinity scale of the naturally occurring α-amino acids, Journal of the American Society for Mass Spectrometry 17 (2006) 1275-1281.
[549] C. Bleiholder, S. Suhai, A.G. Harrison, B. Paizs, Towards Understanding the Tandem Mass Spectra of Protonated Oligopeptides. 2: The Proline Effect in Collision-Induced Dissociation of Protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp), Journal of the American Society for Mass Spectrometry 22 (2011) 1032-1039.
[550] D. Huo, T. Qin, L. Zu, Energetic switch of the proline effect in collision-induced dissociation of singly and doubly protonated peptide Ala-Ala-Arg-Pro-Ala-Ala, Journal of Mass Spectrometry 54 (2019) 55-65.
[551] Y. Hu, K. Cheng, L. He, X. Zhang, B. Jiang, L. Jiang, C. Li, G. Wang, Y. Yang, M. Liu, NMR-Based Methods for Protein Analysis, Analytical Chemistry 93 (2021) 1866-1879.
[552] J.J. Englander, C. Del Mar, W. Li, S.W. Englander, J.S. Kim, D.D. Stranz, Y. Hamuro, V.L. Woods, Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry, Proceedings of the National Academy of Sciences 100 (2003) 7057-7062.
[553] T.L. Religa, L.E. Kay, Optimal methyl labeling for studies of supra-molecular systems, Journal of Biomolecular NMR 47 (2010) 163-169.
[554] E. Mastan, S. Zhu, Method of moments: A versatile tool for deterministic modeling of polymerization kinetics, European Polymer Journal 68 (2015) 139-160.
[555] T. M. Kruse, O. Sang Woo, L. J. Broadbelt, Detailed mechanistic modeling of polymer degradation: application to polystyrene, Chemical Engineering Science 56 (2001) 971-979.
[556] C.H. Bamford, H. Tompa, The calculation of molecular weight distributions from kinetic schemes, Transactions of the Faraday Society 50 (1954) 1097-1115.
[557] D.L. Marchisio, J.T. Pikturna, R.O. Fox, R.D. Vigil, A.A. Barresi, Quadrature method of moments for population-balance equations, AIChE Journal 49 (2003) 1266-1276.
[558] E. Mastan, X. Li, S. Zhu, Modeling and theoretical development in controlled radical polymerization, Progress in Polymer Science 45 (2015) 71-101.
[559] Y.-N. Zhou, Z.-H. Luo, State-of-the-Art and Progress in Method of Moments for the Model-Based Reversible-Deactivation Radical Polymerization, Macromolecular Reaction Engineering 10 (2016) 516-534.
[560] S. Capasso, L. Mazzarella, F. Sica, A. Zagari, S. Salvadori, Kinetics and mechanism of succinimide ring formation in the deamidation process of asparagine residues, Journal of the Chemical Society, Perkin Transactions 2 (1993) 679-682.

[561] Y. Song, R.L. Schowen, R.T. Borchardt, E.M. Topp, Effect of ‘pH’on the rate of asparagine deamidation in polymeric formulations:‘pH’–rate profile, Journal of pharmaceutical sciences 90 (2001) 141-156.
[562] M. Xie, D. Vander Velde, M. Morton, R.T. Borchardt, R.L. Schowen, pH-Induced Change in the Rate-Determining Step for the Hydrolysis of the Asp/Asn-Derived Cyclic-Imide Intermediate in Protein Degradation, Journal of the American Chemical Society 118 (1996) 8955-8956.