Molecular Dynamics Simulations of Extracellular Enzyme-Mineral Interactions in Arctic Soils

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Jacob Tanner Perez

August 2024

Copyright © 2024 by Jacob Tanner Perez
All rights reserved.

ACKNOWLEDGEMENTS

The work that was put into this thesis would not have been possible without the amazing support and guidance of many people. My biggest thanks goes to my advisor, Dr. Andrew Steen, for bringing certainty, curiosity, and adventure into my graduate career. I would have never been able to imagine the places I'd go without him. I would also like to thank Dr. David Keffer for the wisdom, encouragement, and patience as I went on my computational science journey. Thank you to Dr. Karen Lloyd for not only your feedback and support on my research, but for also providing the bioinformatics wisdom to even start this thesis. And certainly thank you to Dr. Anna Szynkiewicz and Dr. Elizabeth Herndon for being incredibly supportive committee members from the very beginning of my time here at UT.

The field work involved with this research was funded by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program DE-SC0020369. Thank you to the AMP'D team and those that collected the samples in 2021: Drs. Karen Lloyd, Andrey Abramov, James Bradley, Julia Boike, Donato Giovannelli, Tatiana Vishnivetskaya, Katie Sipes, Matteo Selci, and Brianna Green, Margaret Cramm, and Sayali Mulay. The DEM of Svalbard provided by the Polar Geospatial Center under NSF-OPP awards 1043681, 1559691, 1542736, 1810976, and 2129685.

I would also like to thank other lab mates, both past and present: Brianna and Murray for both the lab work that fed into this thesis and for being great friends; Iyanu, Kami, and Adrianna for your support and insights; and a special thanks to Fumnanya and Raegan for making my first field expedition memorable. Another special thank you goes to the many people I have met during the International Geobiology Course 2022 and on the Nathaniel B. Palmer in the spring of 2023.

Thank you to my friends and family who have been a reliable support system through it all. Thank you to Carsen for being a great friend from day one. Thank you to Braden and Catie for the company and entertainment as I wrote this thesis. Thank you to Cam, Cat, Chance, Constance, Sara, and many others for the camaraderie and making Knoxville a fun and exciting place to be in. Thank you to my friends, Ashlee, Logan, Ezequiel, and Naila who knew how to keep my eyes on the prize even from afar. Thank you to my brothers and sisters-in-law for the tremendous amount of support, especially during breaks and holidays. Finally, thank you to my parents for the boundless amount of love and being a rock during this rollercoaster of graduate school.

ABSTRACT

Extracellular enzymes catalyze the hydrolysis of organic compounds, a critical initial step in the breakdown of complex soil organic matter. The ability of these enzymes to hydrolyze organic matter into simpler molecules is therefore an important kinetic control on microbial conversion of soil organic carbon into greenhouse gases. The Arctic environment is rapidly evolving – experiencing thawing of permafrost soils, increased thickening of the active layer into mineral soils, and more weathering. The purpose of this study is to investigate the association of extracellular enzymes with mineral surfaces and its implications on the enzymatic breakdown of complex organic carbon in Arctic active-layer environments. Using molecular dynamics simulations, we demonstrate that extracellular enzyme adsorption to mineral surfaces is dependent upon the charge characteristics of both the enzyme and the mineral surface. Additionally, the amount of structural deformation of the enzyme due to adsorption was dependent upon the structural characteristics of the enzyme itself. It is also shown that the potential for the enzyme to remain active upon adsorption can exist as the active site could remain accessible and flexible. The rest of the enzyme could decrease in flexibility, suggesting that adsorption may preserve these enzyme structures by making them more rigid. These simulations reveal the types of mechanisms behind enzyme-mineral adsorption, the importance of soil mineralogy for sorption mechanisms, and the influence of sorption on enzyme structure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1. INTRODUCTION

1.1. Background

The high Arctic hosts a unique and rapidly evolving environment. Low temperatures allow freezing of the upper layer of soils (i.e., the active layer) during cold seasons and thawing during warm seasons. Permafrost, or ground that has remained frozen for at least two years, underlays the active layer and stores roughly one third of the global surface soil carbon stock (~1,000 Pg of carbon) while only spanning 15% of the global soil area (Schuur et al., 2015). With warming in the Arctic being double the global average (Graversen et al., 2008), there comes landscape changes such as glacier retreat, where microbes and plants can begin to colonize on the newly exposed ground (Bradley et al., 2014). The rising temperatures also lead to permafrost thaw and active layer thickening (Strand et al., 2021). Permafrost thaw is resulting in increasing amounts of previously frozen soil organic matter becoming susceptible to degradation. This growing reservoir of carbon has the potential of being converted into greenhouse gases – such as carbon dioxide (CO_2) and methane (CH_4) – and becoming a positive feedback loop to climate change and furthering warming in the Arctic.

The breakdown of soil organic matter is mediated by the microbial communities within Arctic soils. Barbato et al. (2022) found that respiration rates during permafrost thaw were variable in samples within the same permafrost tunnel, potentially due to the heterogeneity of the microbial community assemblages. Additionally, the composition of the soil organic matter may control respiration rates. Recalcitrant forms of carbon, such as lignin and other polymeric carbohydrates, have rapidly degraded in other Arctic regions upon warming (Li et al., 2020). As the Arctic continues to change, its trajectory will be impacted by how both organic matter composition and the microbial communities change over time. How these factors change will depend on how well microbes can access the surrounding carbon pool.

Extracellular enzymes are enzymes that are secreted outside of the cell (and could either be tethered to the cell or free in the environment) and help breakdown large, complex carbon into smaller, more labile carbon that can be taken into the cell. The hydrolysis of larger carbon compounds via extracellular enzymes is the rate-limiting step of the overall breakdown of complex organic matter (Fox & Makam, 2011). How microbial communities produce extracellular enzymes and the efficiency of these enzymes are important controls in modeling the conversion of soil carbon to CO_2 during warming (Allison et al., 2010; Coolen et al., 2011). This

study aims to identify how interactions between extracellular enzymes in Arctic active layer soils and their environment – namely, adsorption to mineral surfaces – could impact carbon cycling in evolving high-latitude regions.

1.2. Organic-Microbe-Mineral Interactions

While the concentration of extracellular enzymes set the maximum activity, interactions with mineral surfaces are a primary control on the observed activity of extracellular enzymes in soils (Datta et al., 2017). Adsorption of various enzymes onto clay particles resulted mostly in a decrease in enzyme activity and protection from photodegradation (Tietjen & Wetzel, 2003). Variable amendments of different minerals showed that the specific activity of carbon-degrading enzymes could persist over many days (Olagoke et al., 2020). However, these experiments yielded variable results, and the mechanisms and sources of activity change were unclear. Therefore, studying interactions with numerous minerals within the active layer soils in the Arctic is needed to understand what geochemical influences mineral adsorption has on extracellular enzyme activity.

The extent and effect of surface adsorption on extracellular enzymes can be diverse in terms of known mechanisms for adsorption. For bovine serum albumin (BSA), adsorption to montmorillonite is interpreted to be mediated by electrostatic attraction when the pH is below the isoelectric point (pI) of the protein but hydrophobic attraction when the pH is close to the pI (Quiquampoix et al., 1993). Quiquampoix et al. also showed that surface area covered by BSA increased when electrostatic attraction was dominant and therefore further inferred denaturation from FTIR measurements. While not an enzyme, BSA serves as a model protein that provides insight into the potential interactions between enzymes and surfaces. Consideration for how different protein structures – especially ones that require specific properties within its structure, such as an enzyme active site – adsorb onto surfaces is important to decide which mechanisms are present and what effects it has on the enzyme structure.

1.3. Molecular Dynamics

Molecular dynamics (MD) simulations have been useful computational methods of studying the properties of various materials, including biological structures like proteins. MD uses Newton's equations of motion and calculations of forces to predict the trajectory of a system. The net force on a particle can be modeled by a set of intramolecular (bonded) interactions, and intermolecular (nonbonded) forces, such as Coulombic and van der Waals forces. The potential

energy for a given particle under these forces can be modeled by Equation 1 (MacKerell et al., 1998):

$$E_{\text{pot}} = \sum_{ij \text{ bonded}} K_{\text{r},ij}\left(r_{ij} - r_{0,ij}\right)^2 + \sum_{ijk \text{ bonded}} K_{\theta,ijk}\left(\theta_{ijk} - \theta_{0,ijk}\right)^2 + \sum_{ijkl \text{ bonded}} \frac{1}{2} V_{\varphi,ijkl}\left[1 + \cos\left(n\varphi_{ijkl} - \varphi_{0,ijkl}\right)\right] + \sum_{\substack{ijkl \text{ bonded} \\ \text{(in plane)}}} K_{\chi,ijkl}\left(\chi_{ijkl} - \chi_{0,ijkl}\right)^2 + \frac{1}{4\pi\epsilon_0\epsilon_r} \sum_{\substack{ij \text{ nonbonded} \\ \text{(1,2 and 1,3 excl)}}} \frac{q_i q_j}{r_{ij}} + \sum_{\substack{ij \text{ nonbonded} \\ \text{(1,2 and 1,3 excl)}}} \varepsilon_{0,ij}\left(\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{0,ij}}{r_{ij}}\right)^{6}\right) \quad (1)$$

The parameters for each potential energy term have been determined empirically for proteins and incorporated in forcefields such as CHARMM. This allows for the modeling of protein structures in a variety of solvents and thermodynamic conditions.

MD can also be used to study how biological structures interact with nanomaterials, given that the interaction terms between biomolecules and inorganic substrates have been parameterized. The INTERFACE forcefield has developed parameters to describe interactions between multiphase systems, including inorganic-biomolecular interfaces (Heinz et al., 2013). Many crystal structures have been incorporated into INTERFACE, such as silicates, clays, and various metals. INTERFACE has provided insight into CO_2 sequestration capacity and stability from industrial waste gas injection into model shale reservoir systems (Zhang et al., 2024). Having defined interaction terms between biological structures and solid surfaces allows MD to be used to investigate the relationship biology has with its complex environment.

1.4. Investigating Cold Adaptations of Extracellular Enzymes

Psychrophilic, or cold-adapted, enzymes are enzymes that have structural characteristics that optimize their activity for low temperatures. Cold-adapted enzymes have lower enthalpy and entropy for activation by decreasing stiffness throughout the structure (Bjelic et al., 2008). Becoming more flexible for optimal activity at lower temperatures could be at the cost of having structural stability. Mutations of residues near the catalytic center of a *Candida antarctica* lipase B showed an increased half-life accompanied by enhanced rigidity (Xie et al., 2014). Careful selection of mutations for thermostability while maintaining cavity flexibility can yield an enzyme structure that is both durable and active (Joo et al., 2010). MD revealed that shortening of protein loops in a thermophilic β-fructosidase helped maintain structural rigidity in higher

temperature simulations (Mazola et al., 2015). MD simulations are an effective way of demonstrating these molecular-level adaptations and could be useful in studying novel structures from the environment.

Predicting the structures of enzymes from natural cold environments is now easier with new bioinformatics tools. DNA reads can be annotated for different carbohydrate-active enzymes (CAZymes) by looking for signature domains of specific CAZyme families within the genes (Yin et al., 2012). Additionally, these sequences can further be annotated as secreted by detecting signal peptides structures (Petersen et al., 2011). This workflow can be used to distill sequences of extracellular enzymes within a metagenomic dataset. Three-dimensional structures can be predicted with AlphaFold, that utilizes machine learning in combination with multiple sequence alignments (Jumper et al., 2021). These tools are easier to use with the development of servers that implement these tools such as the dbCAN server (Zheng et al., 2023) and ColabFold (Mirdita et al., 2022).

This study uses computational tools to understand the potential mechanisms and effects of mineral adsorption for extracellular enzymes present in active layer soils in the Arctic. This work addresses the impact of structural characteristics of extracellular enzymes found in active layer soils on adsorption to mineral surfaces and the influence surface adsorption has on enzyme structure and dynamics. We hypothesized that the charged state (as determined by the isoelectric point of the enzyme) would predict adsorption – specifically, that enzymes with pI close to or above the porewater pH would exhibit adsorption onto negatively-charged surfaces. Additionally, a secondary hypothesis stated that enzymes that adsorbed would undergo structural deformation and reduce flexibility in a manner that is proportional to the amount of contact. By analyzing the resultant trajectories of select enzymes and surfaces, we found that electrostatic forces are the primary driving force for these systems and the amount of structural deformation and changes in flexibility to an enzyme are dependent on the relative amount of secondary structure (i.e., α-helices and β-sheets) within the enzyme.

SECTION 2. MATERIALS AND METHODS

2.1. Field Site Description and Sampling

In March 2021, Dr. Karen Lloyd and other team members (listed in the Acknowledgements) sampled a soil core from the Bayelva river catchment (78.92°N, 11.86°E) near Ny Ålesund on the Svalbard archipelago (Figure 1). This region has previously been described as a sandur, or an outwash plain formed by glacial meltwater, that is covered by patchy shrub moss (Wojcik et al., 2019).

The core was subsectioned by depth into 3-5 cm increments shortly after sampling. For DNA extractions, about 0.5g of soil were taken from the inside middle portion of the section and DNA was extracted using a QIAGEN PowerSoil DNA Kit within 24 hours of collection by scientists on the 2021 field expedition. DNA libraries were prepared using the SMARTer ThruPlex DNA-seq libraries and sequenced on Illumina NovaSeq 6000 S4. The DNA sequences from the 18 cm subsection were used for the metagenomic analysis. I chose this depth since it comes from the active layer and is not likely to be a recently thawed permafrost since it is closer to the surface.

2.2. Metagenomic Analysis

The DNA reads from the 18 cm depth subsection were imported into KBase where I used a standard metagenomic pipeline analysis to trim reads, assess quality, assemble reads, and bin contigs (Chivian et al., 2023). The reads were trimmed of adapters and bad quality reads using Trimmomatic (Bolger et al., 2014) where 97.6% of read pairs were maintained. The reads were assembled using metaSPAdes and produced 1,010 contigs (N50 = 4,400, L50 = 231). The metagenome assembly was annotated using Prokka and then exported from KBase for identification of secreted (extracellular) enzymes.

2.3. Determination of Extracellular Enzyme Structures

The metagenome assembly was annotated for CAZymes using a dbCAN3 server (Zheng et al., 2023). The list of CAZymes was subsetted for those that were secreted by the cell as annotated by the SignalP 4.0 (Petersen et al., 2011) implementation on the server. The pI was calculated from the amino acid sequence of each secreted CAZyme using pIR (Audain et al., 2016). Three extracellular enzymes were selected based on their pI values relative to the environmental pH, which is circumneutral (pH = 7.38 ± 0.19, Hodson et al., 2002). The mid-pI extracellular enzyme, here on referred to as MAN, was annotated with as EC 3.2.1.78, which is a mannan endo-1,4-beta-mannosidase. The high-pI enzyme, here on referred to as NAG, was

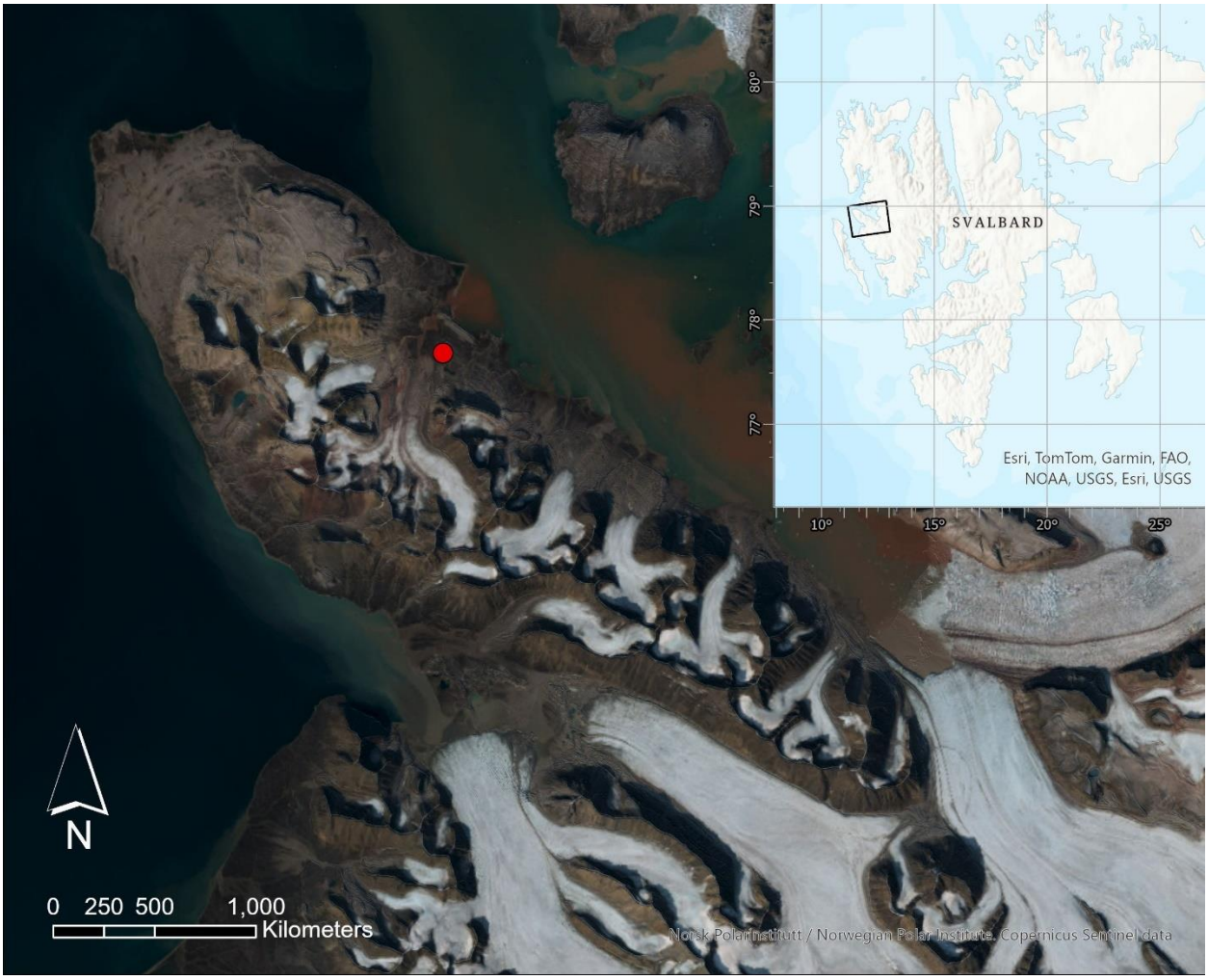

Figure 1: Borehole location (red dot) on the Brøgger peninsula where the DNA reads were extracted from. The sample location was within the Bayelva region near Ny Ålesund on the western side of Svalbard (inset). Imagery © Norwegian Polar Institute / USGS Landsat. DEM created by the Polar Geospatial Center from Maxar Imagery (Porter et al., 2023).

annotated within the Glycoside Hydrolase Family 84, which include activities such as beta-N-acetylglucosaminidase or hyaluronidase. The low-pI enzyme, here on referred to as CHI, was annotated within the Glycoside Hydrolase Family 23, which includes activities such as chitinase.

The folded structures of each enzyme were predicted from their sequences using ColabFold, an implementation of AlphaFold2 (Mirdita et al., 2022). The signal peptide sequence was removed prior to predicting the structure and long chains at either terminus that were predicted to be disordered were also deleted from the structure since it is not likely they would preserve in the environment. The number of sequences used in the multiple sequence alignment (MSA) for each enzyme was sufficiently high (Figure 2A-C). For NAG and MAN, the predicted alignment error (PAE) was low (< 10 Å) for the bulk of the structure, with some high-valued (> 20 Å) regions correlated to disordered chains that are likely mobile (Figure 2D-F). NAG and MAN also had high scores (> 90) for the local difference distance test (plDDT, Figure 2G-I). For CHI, the PAE and plDDT values show lower quality predictions, suggesting that the accuracy of the predicted structure is lower than that of NAG and MAN. The highest ranked structure for each structure was used for the remainder of the analyses (Figure 3). Surface charge distributions were calculated using the Adaptive Poisson-Boltzmann Solver (Jurrus et al., 2018).

Homologous, well-characterized enzymes were identified by using UniProt BLAST tool (The UniProt Consortium et al., 2023). NAG matched to a hyaluronidase from *Thermobaculum terrenum* (E-value 9e-34). The structure of the reference protein was characterized at the protein level using X-ray diffraction and binding site residues have been inferred from experimental and computational evidence (Ostrowski et al., 2015). A BLAST search of the MAN protein sequence matched with endogluconase from *Martelella endophytica* (E-value 1e-8). This reference protein was isolated and characterized as an acidic, salt-tolerant enzyme that breaks down cellulose (Khan et al., 2021). Hypothetical active site residues within NAG and MAN were identified by aligning the protein sequences (Table 2.1). A well-characterized homologous protein was not found for CHI.

2.4. Molecular Dynamics Parameters

2.4.1. Enzyme-Only Systems

Input configurations for molecular dynamics simulations of each enzyme were generated using the Solution Builder tool in CHARMM-GUI (Jo et al., 2008; Lee et al., 2016, 2020). The simulation box size was -determined by simulating NAG, the largest enzyme, in different box

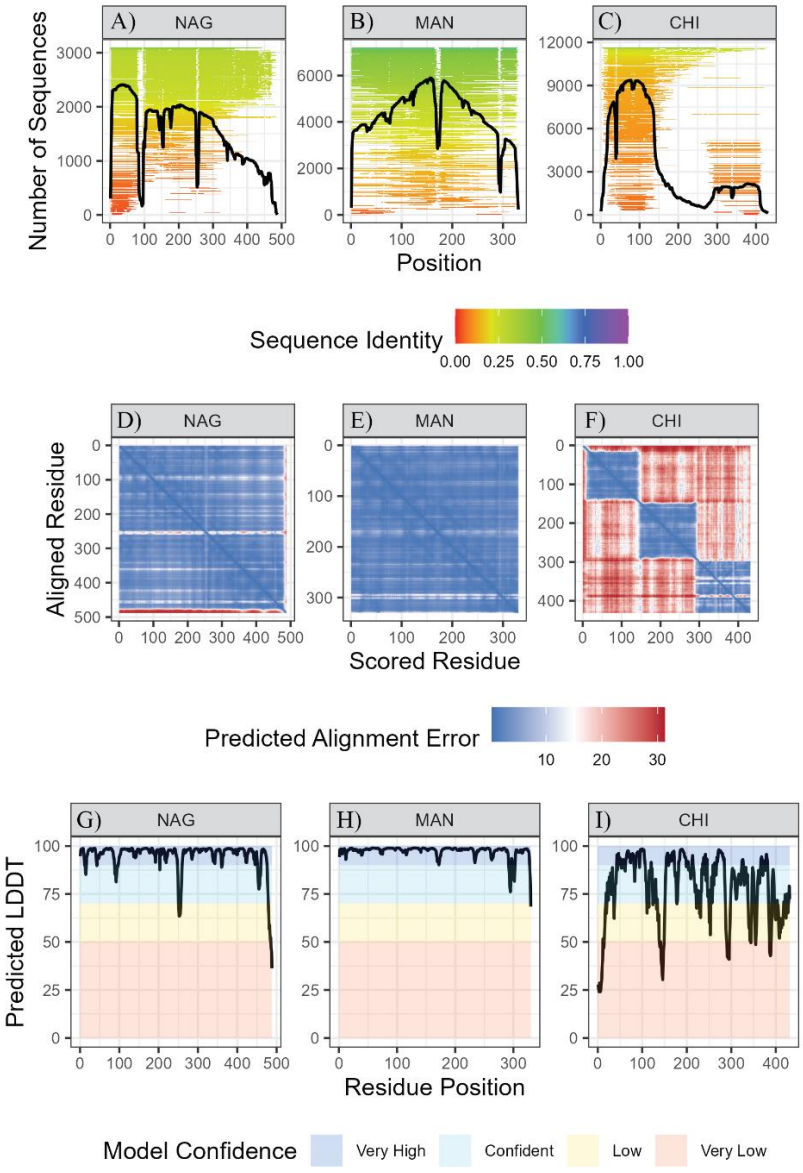

Figure 2: ColabFold prediction statistics for NAG (left column), MAN (middle column), and CHI (right column). A-C: Number of sequences at each position of the enzyme sequence used in the multiple sequence alignment (MSA). The colored regions represent the protein positions that match the queried sequence and are colored by overall sequence identity (red – low, blue – high). D-F: Predicted alignment error (PAE) for each position given the protein is aligned at a different position. Lower values (blue) show higher confidence in predicted position and only the highest-ranking structure is shown. G-I: Predicted Local Difference Distance Test (plDDT) for each structure produced. Higher values show higher confidence in the models.

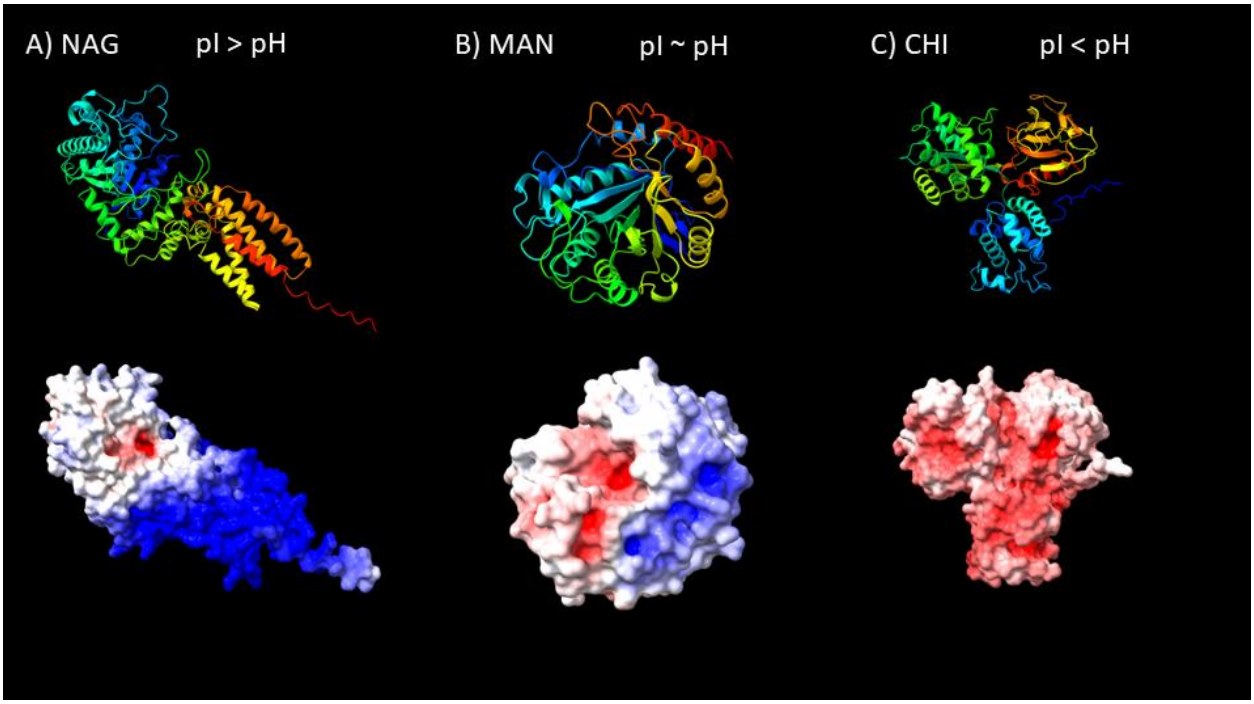

Figure 3: Predicted structures of hypothetical extracellular enzymes predicted by ColabFold (top row) and surface charge distributions calculated from Adaptive Poisson-Boltzmann Solver (bottom row). The enzymes are colored from N-terminus (blue) to C-terminus (red). Positively-charged regions are blue and negatively-charged regions are red. ChimeraX was used for visualization.

Table 2.1: Protein comparisons from UniProtKB BLAST and sequence alignment.

Enzyme	Reference UniProtKB Protein	Sequence Identity	Reference Binding Site Residues	Aligned Binding Site Resides
NAG	D1CDN2_THET1	27.40%	Gly12 Lys43 Asp119 Tyr168 Trp221 Asn223 Asp228 Asn257	Gly11 Lys42 Asp136 Tyr189 Trp242 Asn244 Asp249 Asn282
MAN	GUN_MAREN	26.96%	Glu158 Glu270	Glu150 Glu259

sizes. It was found that the 12 Å edge distance (140 Å x 140 Å x 140 Å) was a suitable minimal size to avoid system size effects on non-bonded energy interactions (see supplement in Appendix). Each enzyme was solvated in TIP3 water with Na^+ and Cl^- ions to balance charge and produce a 0.01 M NaCl solution to approximate the salinity of the sample porewater (Hodson et al., 2002). Each system was simulated in GROMACS (Abraham et al., 2015) using the CHARMM36 forcefield incorporated into CHARMM-GUI (Lee et al., 2016). An energy minimization step was performed using the steep method in GROMACS, followed by an equilibration stage (125 ps) in the NVT ensemble at 277.15 K – to simulate the colder soil temperatures without freezing – using the Nosé-Hoover thermostat with positional restraints on the enzyme backbone and sidechains. Afterwards, the MD simulation was in the NpT ensemble at 277.15 K and 1 bar using the Nosé-Hoover thermostat and the Parrinello-Rahman barostat, where pressure was maintained isotropically. The enzyme was simulated for 100 ns under these conditions.

2.4.2. Enzyme-Surface Systems

Mineral surfaces were generated using the Nanomaterial Modeler tool in CHARMM-GUI (Choi et al., 2022; Jo et al., 2008). The quartz surface was an α-quartz unit cell repeated to produce a rectangular box with 142.477 Å x 144.6632 Å x 17 Å dimensions with periodic image bonds in the x- and y-directions. The surface in the z-direction was cut along (001) direction of the crystal structure and 13.3% of surface -Si-OH were ionized (-Si-O^-) and charge was pacified with Na^+ ions. The muscovite surface was a rectangular box with 140.1786 Å x 144.2448 Å x 20 Å dimensions with the basal surface exposed in the z-direction. The exposed tetrahedral layer was half-saturated with K^+ ions. The kaolinite surface was a rectangular box with 144.312 Å x 143.071128 Å x 22 Å dimensions with the basal surfaces exposed in the z-direction. Preference for adsorption to either the tetrahedral or octahedral layers was investigated by running multiple starting orientations for each enzyme. Kaolinite systems with enzymes that interacted with the tetrahedral layer are referenced as “kaolinite(-)” and those with enzymes that interacted with the octahedral layer are referenced as “kaolinite(+)”, as it relates to the approximate surface charge each layer exhibits.

The mineral surfaces produced from the Nanomaterial Modeler and enzyme structures from the Solution Builder were used to create enzyme-surface systems using the Multicomponent Assembler tool in CHARMM-GUI (Jo et al., 2008; Kern et al., 2023). Each enzyme was placed

within proximity to the surface to minimize time of the enzyme approaching the surface and encourage adsorption if it was preferable. The system was solvated in TIP3 water with the solution thickness being at least 120 Å (most were 140 Å). Additional Na^+ and Cl^- ion were placed to produce 0.01 M NaCl solutions. Each system was simulated in GROMACS using the INTERFACE forcefield implemented within CHARMM-GUI (Emami et al., 2014; Heinz et al., 2013). An energy minimization step was performed using the steep method in GROMACS, followed by a step-wise equilibration stage for 250 ps in the NVT ensemble and 1,750 ps in the NpT ensemble at 277.15 K and 1 bar where pressure was being maintained anisotropically, where the x- and y-directions were incompressible to avoid artificial buckling of the mineral surface. Positional restraints were imposed on the backbone and side chain atoms with decreasing force constant values at each step during the equilibration. Afterwards, the MD simulation was in the NpT ensemble with no positional restraints and simulated for 100 ns.

2.4.3. Structural and Statistical Analysis

The root-mean-square deviation (RMSD) of the enzyme backbone was calculated at each saved trajectory frame using the gmx rms tool in GROMACS. The reference structure was taken as the initial enzyme configuration (i.e., the ColabFold predicted structure) and a least-squares fitting was performed prior to calculating structural deviations. Simulations were considered in steady-state if the enzyme became immobilized on the surface or if the enzyme exhibited consistent dynamics (such as remaining suspended in solution). Most systems reached steady-states by 75 ns of simulation time. Average values and standard errors were from the last 25 ns of each trajectory. RMSD was calculated using the gmx rms function in GROMACS that calculates RMSD as shown in Equation 2:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i - r_i^{ref}\right)^2} \quad (2)$$

where r_i is the atom position and r_i^{ref} is the atom position in the reference structure. Using a least squares fitting method, the enzyme was rotated and translated to best fit the reference structure prior to calculating RMSD at each trajectory frame.

The root-mean-square-fluctuation (RMSF) of a given residue within the enzyme structure was similarly calculated for each system (i.e., using the last 25 ns of the trajectory). The RMSF

of a residue represents the time-averaged deviation of the residue backbone from the average position. RMSF was calculated using the gmx rmsf function in GROMACS, using Equation 3:

$$RMSF_i = \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle} \quad (3)$$

where $\langle r_i \rangle$ is the time-averaged position of the backbone atom.

Enzyme dipole and radius of gyration was calculated using the gmx dipoles and gmx gyrate functions, respectively, in GROMACS. Energy interactions were calculated using the gmx energy function. Standard errors were calculated using a block-averaging method (Flyvbjerg & Petersen, 1989).

SECTION 3. RESULTS

3.1. Structural Characteristics of Extracellular Enzymes

The three enzymes selected for this study were chosen for the difference in isoelectric point (pI). The relative composition residues in each enzyme does not appear to vastly differ except for the number of basic residues, which is likely the source of the difference in pI (Figure 4). The relative compositions of residue types (Table 3.1) were significantly different ($p < 0.001$, chi-square), where CHI was different in composition from both NAG ($p < 0.001$) and MAN ($p < 0.01$) mostly due to its lack of basic residues. NAG and MAN were not found to be significantly different in relative composition but the difference in pI is likely sourced from NAG being larger in size (488 residues vs. 330 residues) and therefore possessing more basic residues in total.

All structures contained secondary structures including α-helices and β-sheets (Table 3.2). NAG contained the most α-helices (22) and a parallel β-sheet. The tertiary structure of NAG contained α-β-α motifs that construct the parallel barrel β-sheet as well as many helix-helix interactions. This made NAG the largest enzyme with a large amount of secondary/tertiary structure. MAN was the smallest enzyme and contained a parallel barrel β-sheet with α-β-α motifs, like NAG. CHI was folded into three domains, two of which consisted primarily of α-helices and the third contained both an α-helix and an antiparallel β-sheet. This structure demonstrated substantial flexibility and deviation from the predicted structure in the simulations because there could be large displacements of each subdomain due to the mobile nature of the disordered chains connecting each subdomain.

3.2. Select Adsorption of Extracellular Enzymes on Different Mineral Surfaces

After 100 ns of simulation and testing different initial orientations for multiple enzyme-mineral pairs, only select pairs of enzymes and minerals demonstrated adsorption to the mineral surface. For NAG, the high-pI enzyme, adsorption was observed on quartz (a negatively-charged surface) and kaolinite(-) (i.e., the negatively-charged tetrahedral layer of kaolinite). MAN, the circumneutral pI, showed adsorption to quartz and kaolinite(+) (i.e., the positively-charged octahedral layer of kaolinite). Two additional orientations were tested for MAN and kaolinite to see whether a different enzyme face was more attracted to the mineral surface, but no sorption was observed in any of those simulations. For CHI, the low-pI enzyme, adsorption was only observed onto kaolinite(+). No adsorption was observed within the simulation time onto the muscovite surface. Overall, more contact with negatively-charged surfaces and less contact with

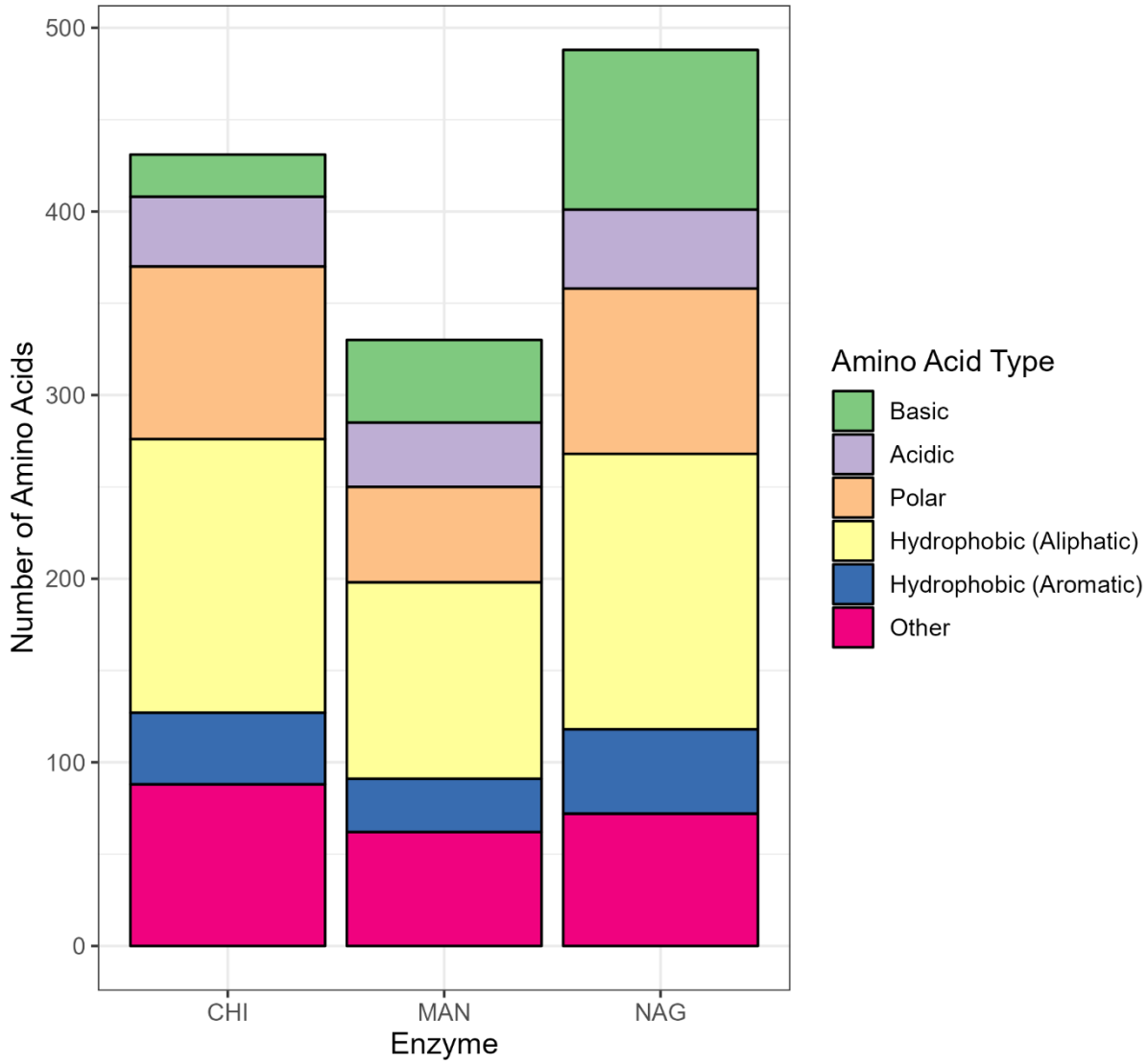

Figure 4: Composition of amino acid types for each enzyme. Basic residues include arginine, histidine, and lysine. Acidic residues include aspartic acid and glutamic acid. Polar residues include asparagine, cysteine, glutamine, serine, and threonine. Aliphatic hydrophobic residues include alanine, isoleucine, leucine, methionine, and valine. Aromatic hydrophobic residues include phenylalanine, tryptophan, and tyrosine. Other residues are glycine and proline.

Table 3.1: Primary structural characteristics of extracellular enzymes.

Enzyme	Total Residues	Types of Residues					pI	Weight (kDa)
		Basic	Acidic	Polar	Hydrophobic	Other		
NAG	488	87	43	90	196	72	11.4	50.8
MAN	330	45	35	52	136	62	6.41	33.4
CHI	431	23	38	94	188	88	4.64	41.7

Table 3.2: Secondary structural characteristics of extracellular enzymes. Regions of the enzyme were classified using PDBsum (Laskowski et al., 1997) that implements PROMOTIF (Hutchinson & Thornton, 1996) on submitted PDB files.

Enzyme	Alpha helix		Beta sheet		Chain
	Count	Num. Res (%)	Count	Num. Res (%)	Num. Res (%)
NAG	22	242 (49%)	2	43 (9%)	203 (42%)
MAN	13	105 (32%)	3	58 (17%)	167 (51%)
CHI	16	170 (39%)	3	61 (14%)	200 (46%)

kaolinite(+) was observed as pI increased (Figure 5).

For systems that showed adsorption onto negatively-charged surfaces, basic residues consisted of most of the residues that were in contact with the surface (Figure 6). For those that adsorbed onto kaolinite(+), polar residues were the highest in number. Hydrophobic residues were also prominent in the contact of CHI to the kaolinite surface.

3.3. Structural Changes in the Presence of Mineral Surfaces

The root-mean-square deviation (RMSD) of the enzyme structure from the structure predicted from ColabFold was measured at each saved trajectory frame using the GROMACS tool gmx rms (Table 3.3). From the three enzymes, NAG deviated least from the predicted structure across all simulations (Figure 7). The average RMSD values for each NAG simulation were all within 0.01 nm of each other. For NAG-Quartz and NAG-Kaolinite(-), the distribution of RMSD values within a simulation was narrower than NAG-No Surface and NAG-Muscovite. For MAN, quartz and muscovite simulations showed an increase in average values while kaolinite(+) showed a decrease in average values relative to the surface-free simulation. CHI showed a large deviation from the predicted structure in the surface-free simulation and the surface simulations showed higher deviations, comparable to MAN simulations. Smaller distributions of RMSD values were observed for the kaolinite(+) system for CHI.

The root-mean-square fluctuation (RMSF) of the enzyme backbone was measured during the equilibrated stage of the simulation (i.e., the last 25 ns). NAG fluctuated very little throughout the structure with the largest exception being the end of the chain (Figure 8). Differences between surface simulation and surface-free simulations for NAG were small with the end of the amino acid chain strongly decreasing in fluctuation in the quartz and kaolinite(-) simulations. Median values for RMSF of disordered chain regions and alpha helices were similar, whereas beta sheets were systematically lower across all NAG simulations (Figure 9). MAN possessed more regions of higher fluctuations relative to NAG and MAN (Figure 10). The MAN-Muscovite system showed the largest positive change in fluctuation. Median RMSF values were lower for both alpha helices and beta sheets relative to disordered chain regions (Figure 11). CHI had mostly higher values for RMSF compared to the other enzymes, with the beginning of the chain fluctuating the most within its structure (Figure 12). The median values were lowest for alpha helices in the CHI-simulations, except for the surface free simulation and disordered chains still had the highest median values (Figure 13).

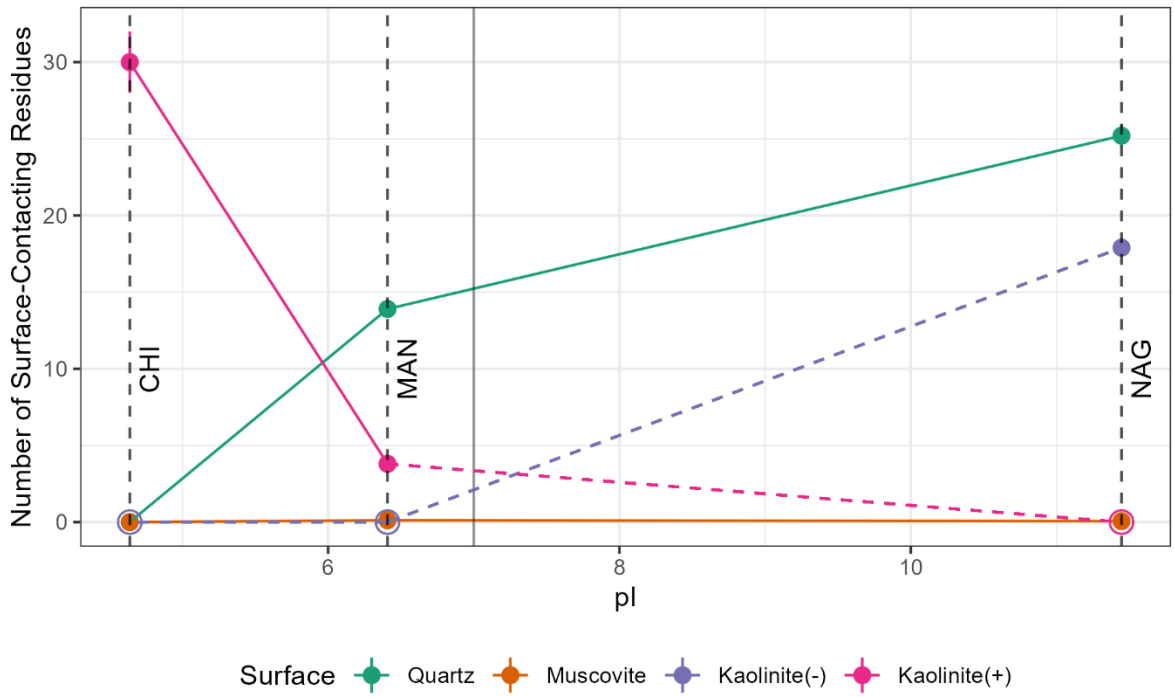

Figure 5: Average number of residues in contact with the mineral surface in the last 25 ns of the simulation as a function of the isoelectric point (pI) of the enzyme. Dotted lines and open circles represent interpreted zero values for kaolinite simulations, meaning that the enzyme preferred to adsorb to the oppositely-charged kaolinite face.

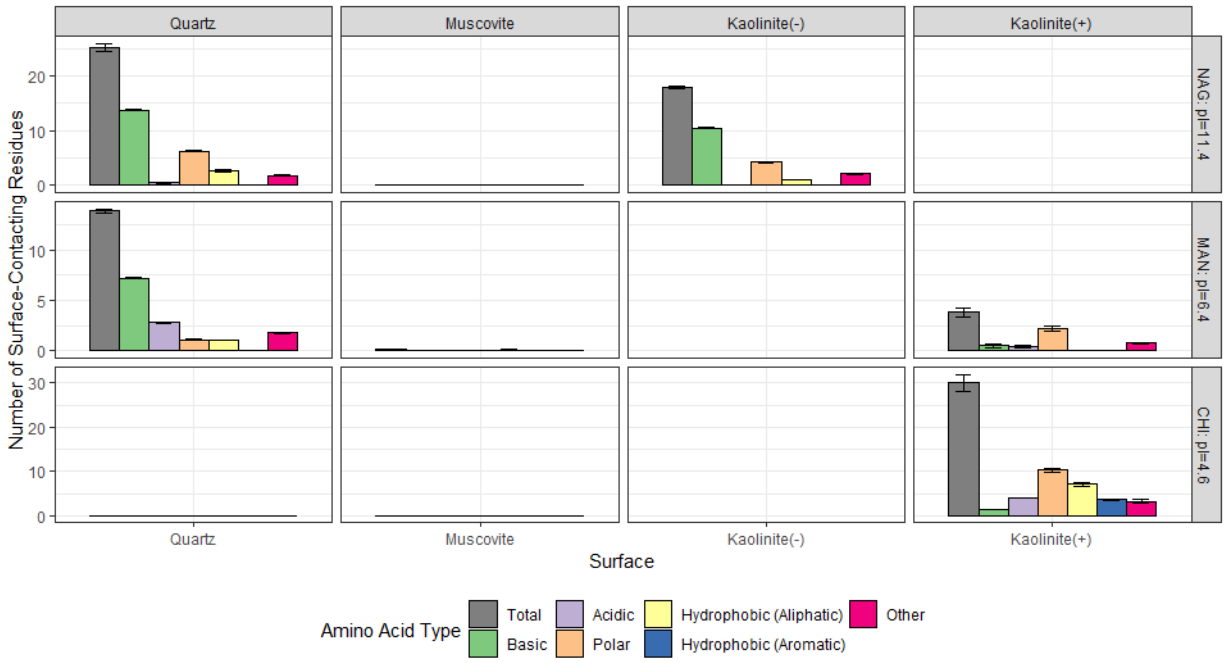

Figure 6: Categorization of contacting residues of each enzyme-mineral pair by amino acid type. Kaolinite is separated to clarify which basal face the enzyme adsorbed to. Values for all muscovite and CHI-Quartz are negligible.

Table 3.3: Average structural values for the last 25 ns of each simulation set. Root-mean-square-deviation (RMSD) was measured relative to the initial ColabFold structure.

Enzyme	Surface	RMSD (nm)	R_g (nm)	Dipole Moment (Debye)
NAG	No Surface	0.24 ± 0.01	2.706 ± 0.003	5,810 ± 20
	Quartz	0.250 ± 0.003	2.715 ± 0.004	6,011 ± 8
	Muscovite	0.23 ± 0.02	2.728 ± 0.003	5,880 ± 10
	Kaolinite(-)	0.227 ± 0.003	2.743 ± 0.007	5,920 ± 10
MAN	No Surface	0.42 ± 0.02	1.953 ± 0.002	1,420 ± 20
	Quartz	0.48 ± 0.01	1.978 ± 0.006	1,535 ± 6
	Muscovite	0.493 ± 0.003	1.981 ± 0.003	1,370 ± 30
	Kaolinite(+)	0.36 ± 0.02	1.953 ± 0.006	1,430 ± 20
CHI	No Surface	0.33 ± 0.01	2.427 ± 0.003	700 ± 10
	Quartz	0.40 ± 0.02	2.381 ± 0.001	570 ± 20
	Muscovite	0.42 ± 0.02	2.415 ± 0.005	780 ± 20
	Kaolinite(+)	0.509 ± 0.007	2.390 ± 0.002	918 ± 5

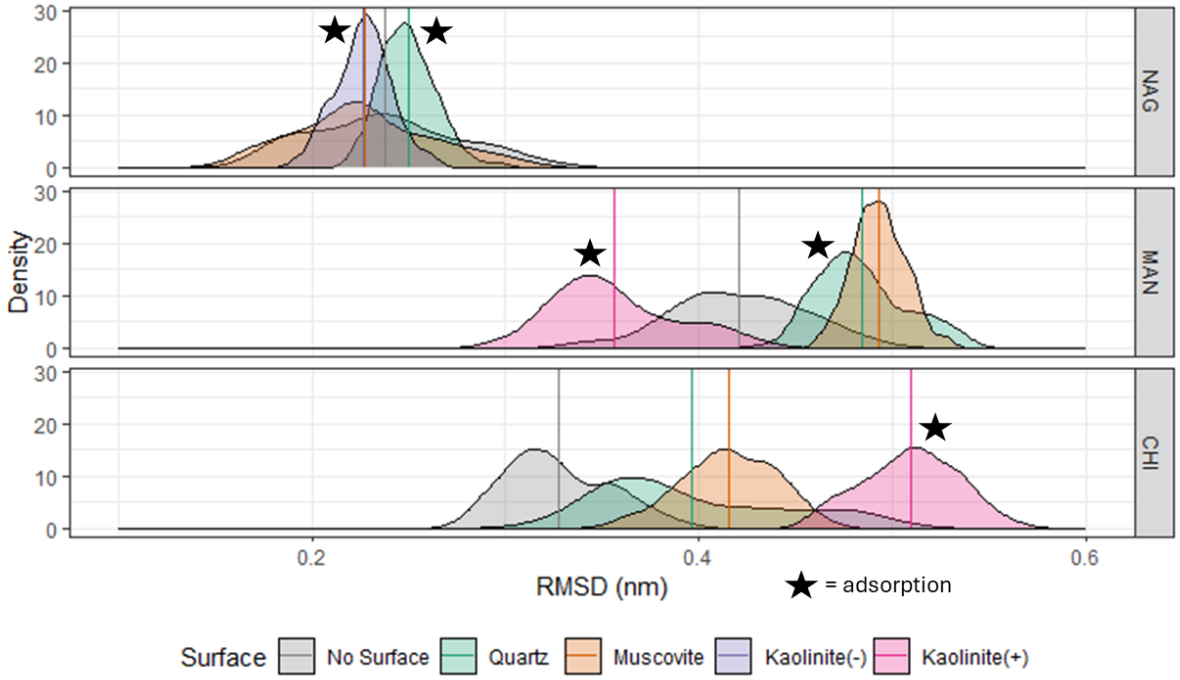

Figure 7: Distribution of root-mean-square deviation (RMSD) values in the last 25 ns of each simulation. Vertical lines represent the mean value of each system and stars indicate systems where the enzyme adsorbed to the corresponding surface.

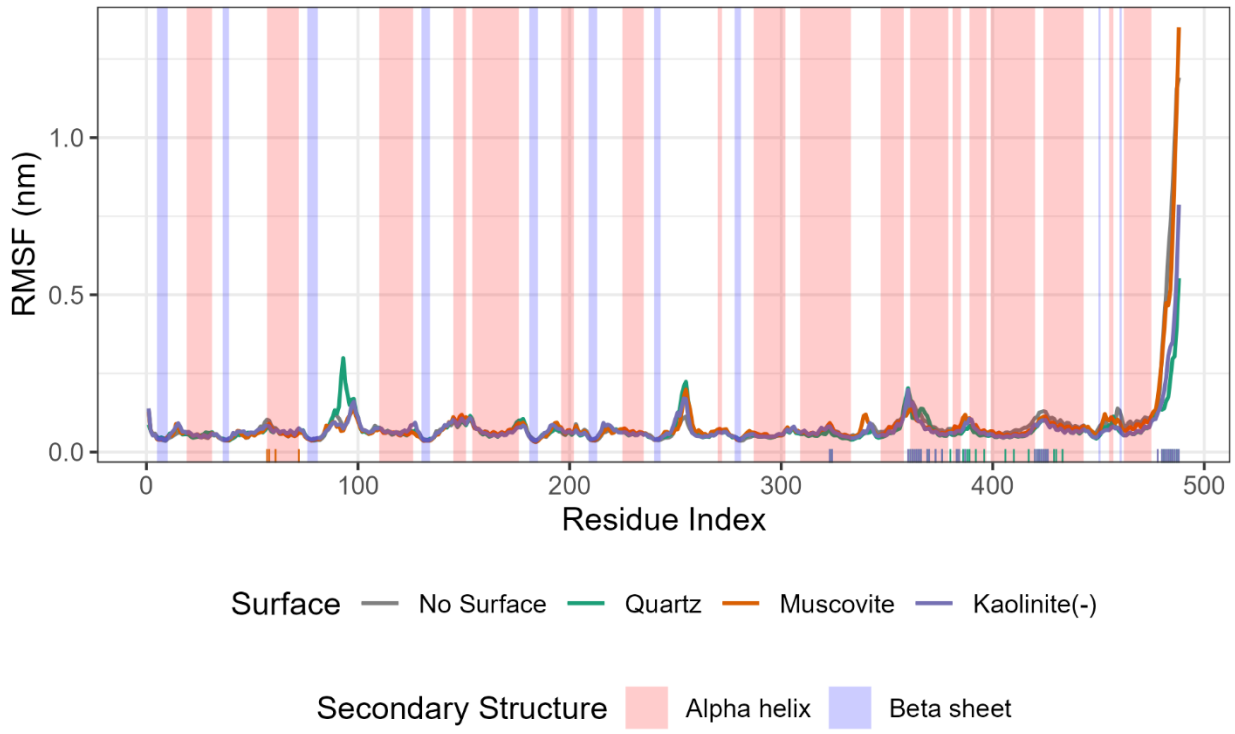

Figure 8: Root-mean-square-fluctuations (RMSF) for each residue in NAG during the last 25ns of each simulation. Blue and red shaded regions correspond to residues in beta sheets and alpha helices, respectively. Tick marks below the graph indicate residues that were in contact with the corresponding surface at any point in the last 25 ns.

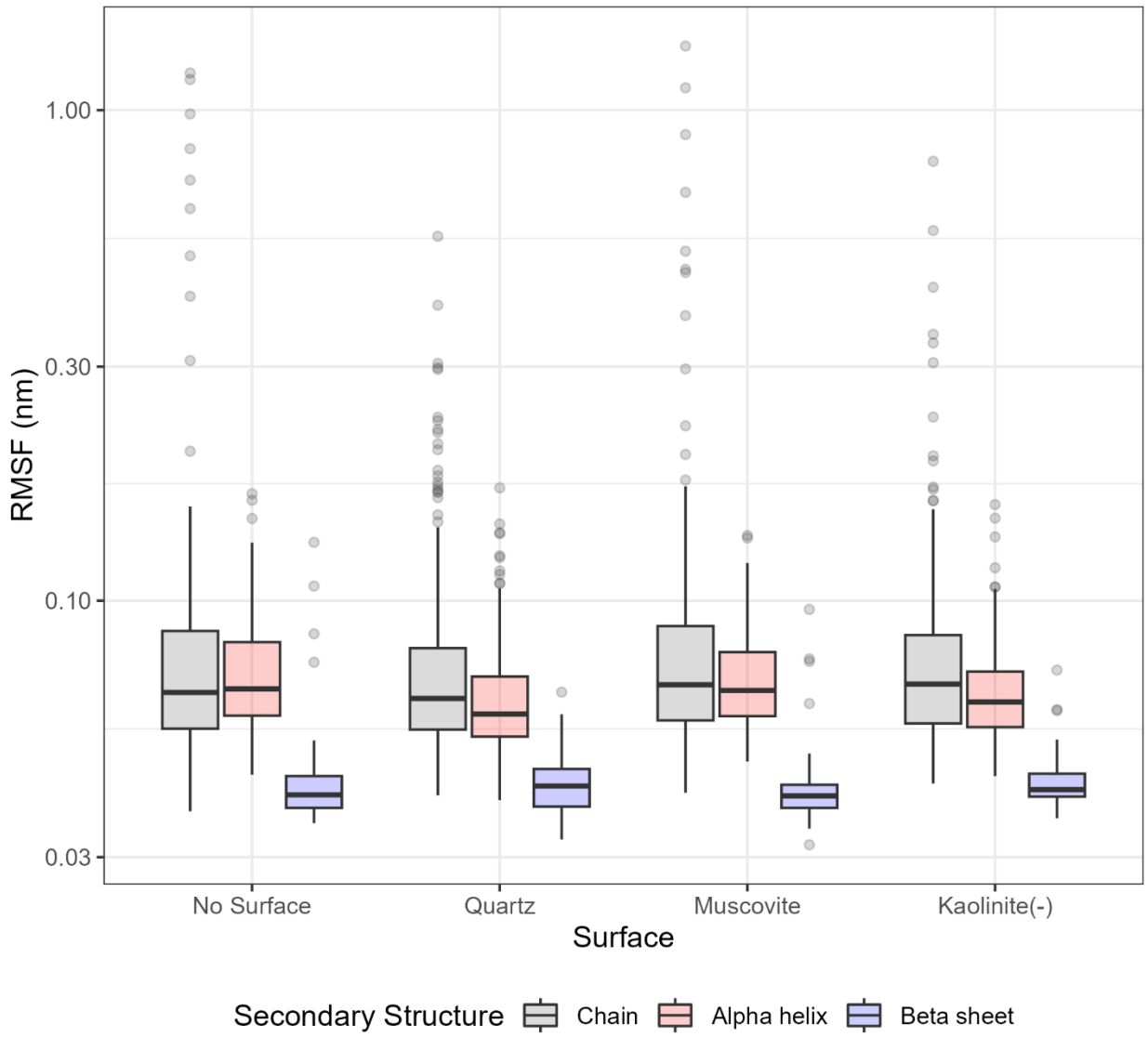

Figure 9: Boxplots of RMSF values in the last 25 ns of simulation for each NAG system simulations grouped by type of secondary structure. Residues in alpha helices or beta sheets are red and blue, respectively. Residues in neither type of those structures were considered part of the chain group and displayed in grey.

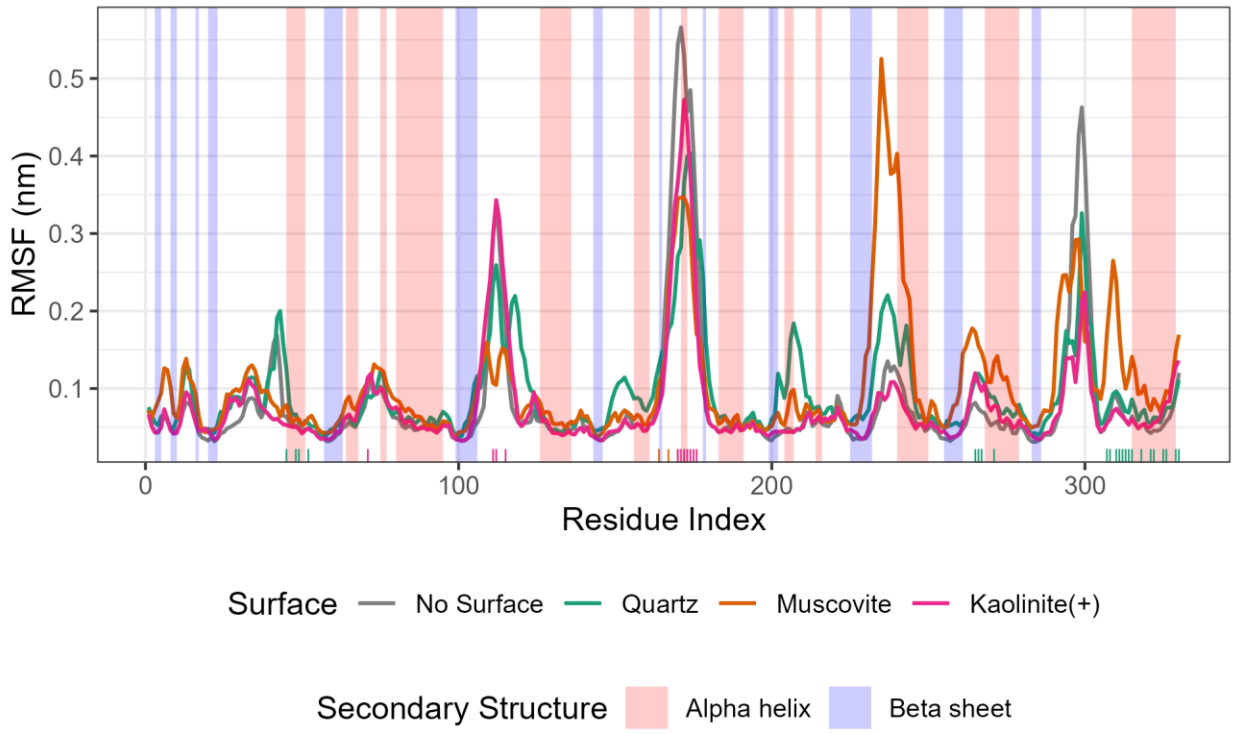

Figure 10: Root-mean-square-fluctuations (RMSF) for each residue in MAN during the last 25ns of each simulation. Blue and red shaded regions correspond to residues in beta sheets and alpha helices, respectively. Tick marks below the graph indicate residues that were in contact with the corresponding surface at any point in the last 25 ns.

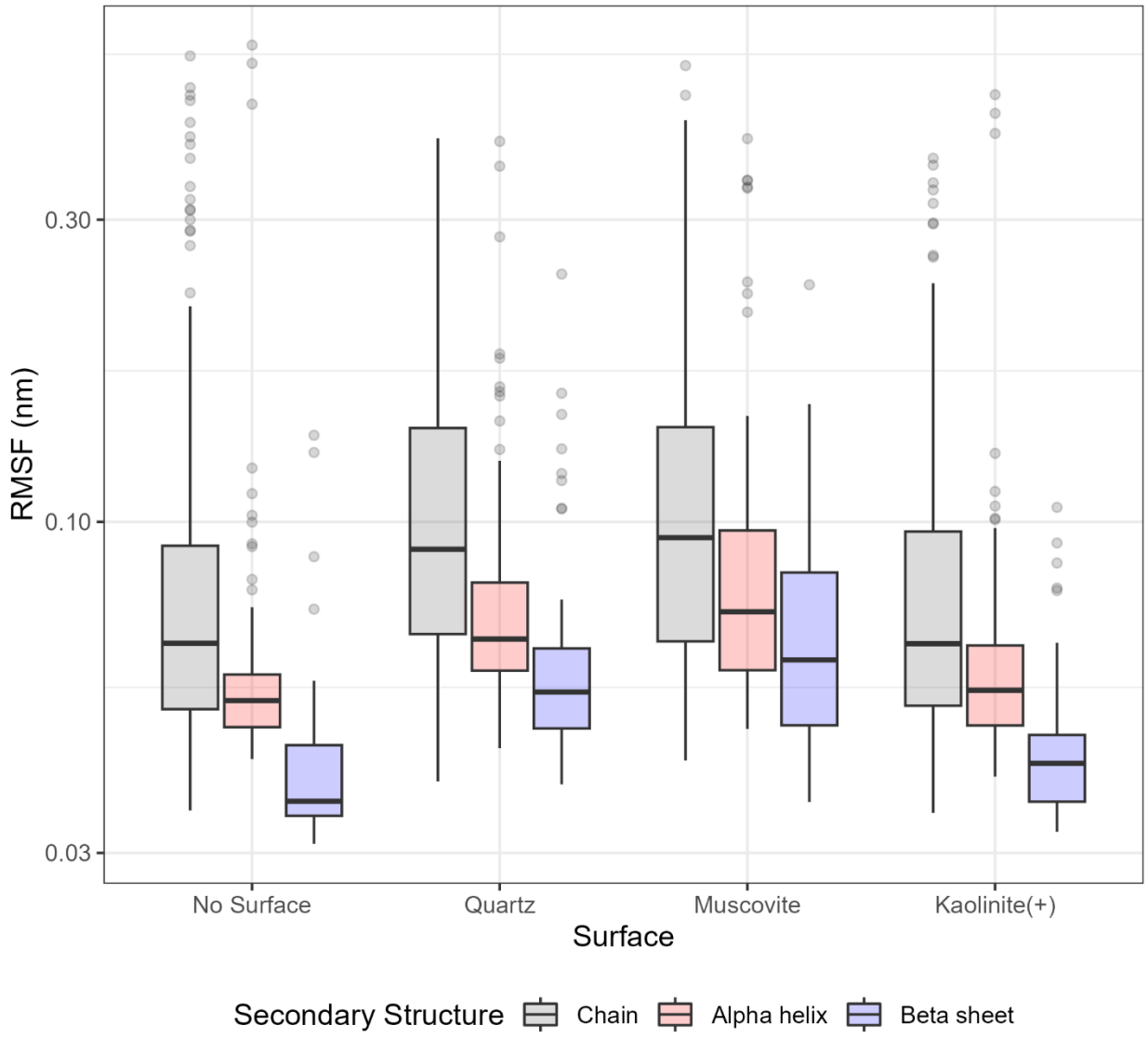

Figure 11: Boxplots of RMSF values for each MAN system simulations grouped by type of secondary structure. Residues in alpha helices or beta sheets are red and blue, respectively. Residues in neither type of those structures were considered part of the chain group and displayed in grey.

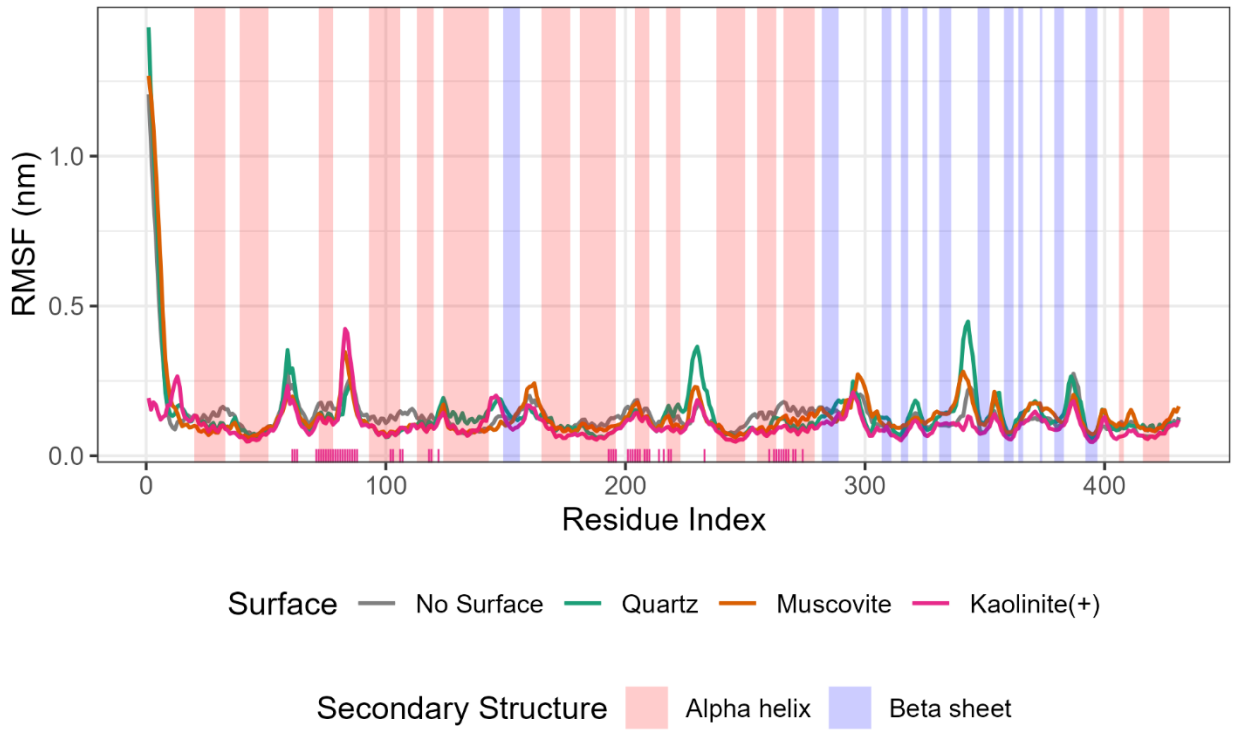

Figure 12: Root-mean-square-fluctuations (RMSF) for each residue in CHI during the last 25ns of each simulation. Blue and red shaded regions correspond to residues in beta sheets and alpha helices, respectively. Tick marks below the graph indicate residues that were in contact with the corresponding surface at any point in the last 25 ns.

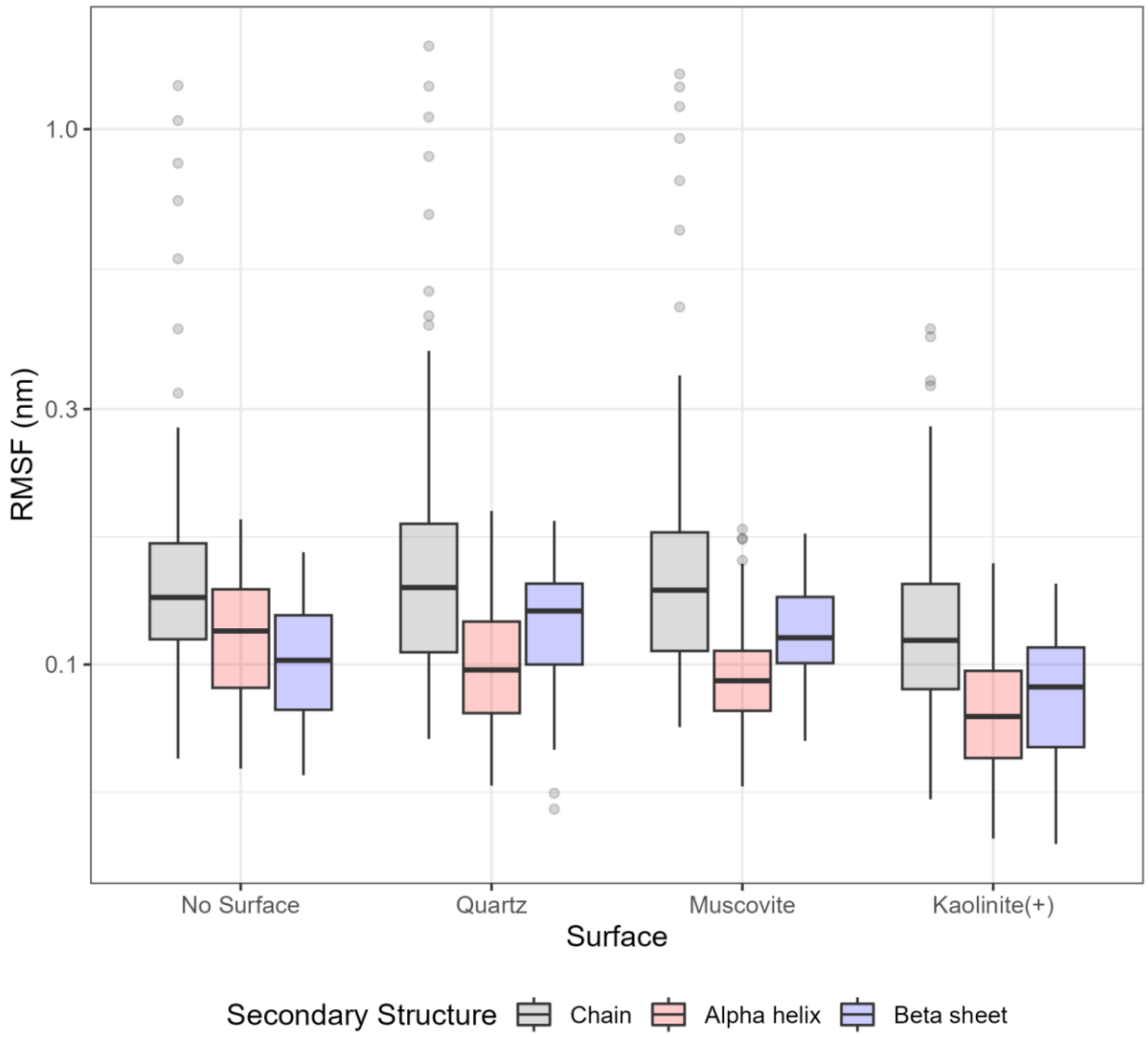

Figure 13: Boxplots of RMSF values for each CHI system simulations grouped by type of secondary structure. Residues in alpha helices or beta sheets are red and blue, respectively. Residues in neither type of those structures were considered part of the chain group and displayed in grey.

3.4. Energetic Changes in the Presence of Mineral Surfaces

Non-bonded energy interactions were averaged over the last 25 ns of each simulation, including Coulombic and van der Waals (modeled by Lennard-Jones potential) potential energies sourced from both short-range interactions and 1-4 interactions (i.e., interactions between two atoms that are separated by two bonded atoms). The total Coulombic and Lennard Jones (LJ) interaction energies were estimated by summing the short-range and 1-4 energies (Table 3.4 and Table 3.5). Changes in energies for each surface simulation were calculated by subtracting the corresponding average energy from the surface-free simulation, such that a negative energy change indicates a preferable shift in energy interactions (Figure 14 and Figure 15). The enzyme-surface energy was defined as zero for the surface-free simulation. The solvent group included surface ions on the quartz and muscovite surface since ions were free to move into the solution.

For NAG, there was a large negative Coulombic energy interaction between the enzyme and the quartz (1,420 ± 20 kJ/mol) and kaolinite(-) (-250 ± 20 kJ/mol) surface, with quartz having the largest decrease in Coulombic energy. There was a similar decrease in LJ non-bonded energy interactions, -99 ± 2 and -151 ± 1 kJ/mol for quartz and kaolinite(-), respectively, where kaolinite(-) had the largest decrease in energy. The energy change for the NAG-Muscovite interaction was nearly zero (2 ± 1 kJ/mol). There was a small increase in non-bonded Coulombic energies within the enzyme structure for quartz (300 ± 70 kJ/mol) and kaolinite(-) (480 ± 210 kJ/mol) and muscovite did not show an appreciable change in energy (80 ± 120 kJ/mol). The non-bonded LJ energies for enzyme-enzyme interactions increased across all surface simulations relative to the surface-free simulation, where muscovite showed the least increase (40 ± 20 kJ/mol) and kaolinite(-) the most (200 ± 20 kJ/mol). The Coulombic interactions with the solution minorly increased for the quartz simulation (320 ± 130 kJ/mol) and decreased for kaolinite(-) (-1,400 ± 400 kJ/mol). The LJ interactions with the solvent changed very little for all surface simulations.

For MAN, there was only a large decrease in Coulombic energy between the enzyme and the surface for the quartz simulation (-620 ± 20 kJ/mol). There were decreases in LJ energies for quartz (-60 ± 4 kJ/mol) and kaolinite(+) (-15 ± 1 kJ/mol). The non-bonded interactions, both Coulombic and LJ, between the enzyme and itself increased across all surface simulations with quartz showing the highest increases in energy (2,390 ± 110 kJ/mol for Coulombic and 280 ± 30 kJ/mol). The non-bonded interactions between the enzyme and the solvent decreased across all

Table 3.4: Summary of interaction energies between the enzyme and the mineral surfaces. All values are reported in kJ/mol.

Enzyme	Surface	U_{Coul}	U_{vdW}	U_{total}
NAG	Quartz	-1,420 ± 20	-99 ± 2	-1,520 ± 20
	Muscovite	-1.8 ± 0.9	-0.3 ± 0.4	-2 ± 1
	Kaolinite(-)	-250 ± 20	-151 ± 1	-400 ± 20
MAN	Quartz	-620 ± 20	-60 ± 4	-680 ± 20
	Muscovite	-3.0 ± 2.0	-0.6 ± 0.4	-4 ± 2
	Kaolinite(+)	-3.5 ± 0.7	-15 ± 1	-12 ± 1
CHI	Quartz	0	0	0
	Muscovite	-80 ± 20	-0.4 ± 0.8	-80 ± 20
	Kaolinite(+)	-19 ± 1	-125 ± 5	-144 ± 5

Table 3.5: Summary of non-bonded interaction energies between the enzyme with itself. All values are reported in kJ/mol.

Enzyme	Surface	U_{Coul}	U_{vdW}	U_{total}
NAG	No Surface	-86,580 ± 60	-7,780 ± 20	-94,360 ± 60
	Quartz	-86,280 ± 40	-7,660 ± 20	-93,940 ± 40
	Muscovite	-86,500 ± 100	-7,740 ± 10	-94,200 ± 100
	Kaolinite(-)	-86,100 ± 200	-7,580 ± 10	-93,700 ± 200
MAN	No Surface	-37,590 ± 50	-5,260 ± 20	-42,850 ± 50
	Quartz	-35,200 ± 100	-4,980 ± 20	-40,200 ± 100
	Muscovite	-35,930 ± 90	-5,060 ± 30	-40,990 ± 90
	Kaolinite(+)	-36,500 ± 100	-5,080 ± 10	-41,600 ± 100
CHI	No Surface	-29,510 ± 100	-5,890 ± 10	-35,400 ± 100
	Quartz	-29,130 ± 70	-5,730 ± 10	-34,860 ± 70
	Muscovite	-29,550 ± 90	-5,770 ± 30	-35,320 ± 90
	Kaolinite(+)	-30,400 ± 100	-5,990 ± 10	-36,400 ± 100

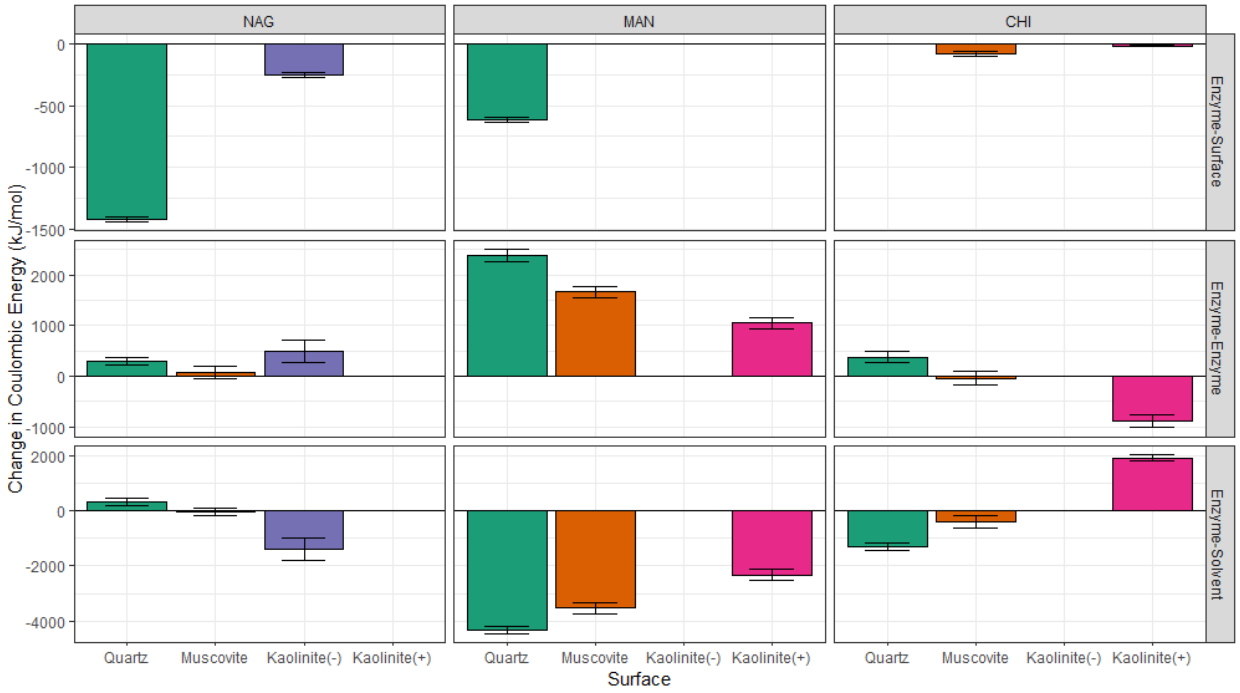

Figure 14: The changes in the Coulombic interaction energies between various groups in the simulation. Kaolinite(+) for NAG and Kaolinite(-) for MAN and CHI are absent since those simulations did not show interactions towards those surfaces. Standard errors were propagated from energy measurements and errors obtained using a block-averaging method. The total energies include both short-range and 1-4 Coulombic interactions.

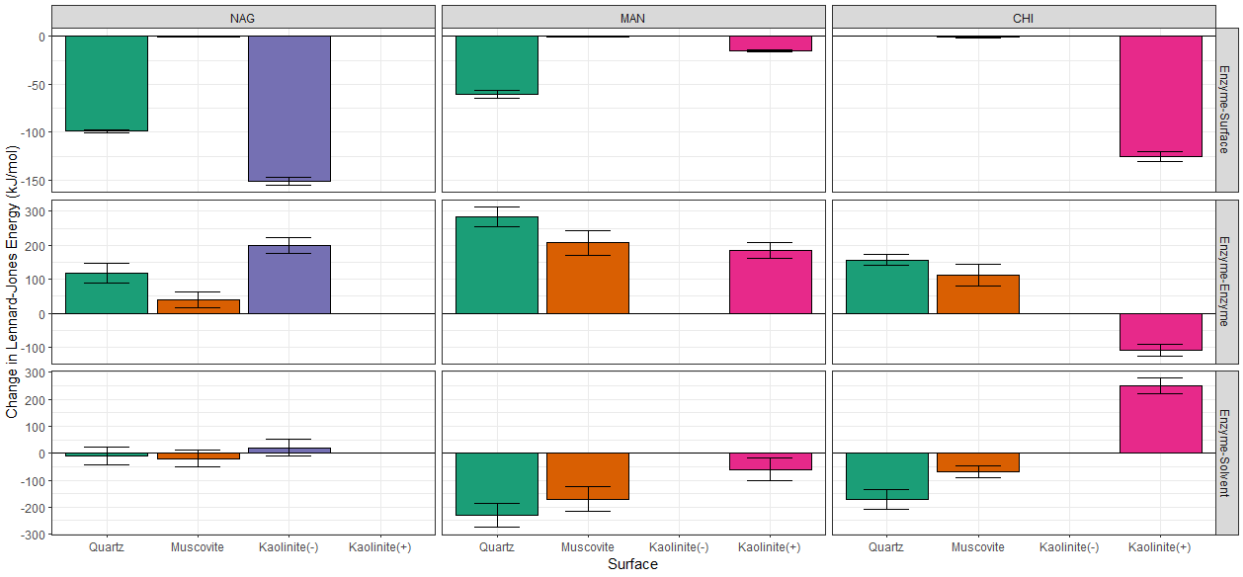

Figure 15: The changes in the Lennard-Jones interaction energies between various groups in the simulation. Kaolinite(+) for NAG and Kaolinite(-) for MAN and CHI are absent since those simulations did not show interactions towards those surfaces. Standard errors were propagated from energy measurements and errors obtained using a block-averaging method. The total energies include both short-range and 1-4 Lennard-Jones interactions.

surfaces, from -2,320 ± 210 to -4,320 ± 120 kJ/mol for Coulombic and -60 ± 40 to -230 ± 40 kJ/mol for LJ energies.

For CHI, there was a small change in Coulombic energies between the enzyme and surface for muscovite (-80 ± 20 kJ/mol) and a large change in LJ energies for kaolinite(+) (-125 ± 5 kJ/mol). The Coulombic energies increased for the enzyme with itself for quartz (380 ± 120 kJ/mol) and decreased for kaolinite(+) (-890 ± 140 kJ/mol). No significant relative change was observed for muscovite (-40 ± 130 kJ/mol). The LJ interactions between the enzyme and itself increased for quartz (160 ± 10 kJ/mol) and muscovite (120 ± 30 kJ/mol) and decreased for kaolinite(+) (-100 ± 10 kJ/mol). The Coulombic and LJ interactions between the enzyme and the solvent decreased for quartz (-1,300 ± 100 and -170 ± 40 kJ/mol) and increased for kaolinite(+) (1,900 ± 100 and 250 ± 30 kJ/mol), with no large change for muscovite (-400 ± 200 and -70 ± 20 kJ/mol).

SECTION 4. DISCUSSION

4.1. Structural Drivers of Adsorption to Mineral Surfaces

All three simulated extracellular enzymes exhibited adsorption to at least one mineral surface in the set of MD simulations, implying that adsorption is dictated by factors in addition to the isoelectric point of the enzyme. More broadly, adsorption occurred when there was a charge compatibility between the enzyme and the surface. Positively-charged enzymes and positively-charged regions of enzymes (as in the case for MAN) adsorbed onto negatively-charged surfaces like quartz and the tetrahedral face of kaolinite while the negatively-charged enzyme and the negatively-charged region of enzymes adsorbed onto positively-charged surfaces like the octahedral face of kaolinite (Figure 5). The lack of adsorption to muscovite is likely due to the negatively-charged tetrahedral layer being pacified by potassium cations, creating an effectively charge-neutral surface and minimizing electrostatic attraction.

The mechanisms for adsorption appeared to differ across enzyme and mineral pairings. The types of residues in contact with the surface depended on the type of surface (Figure 6). For negatively-charged surfaces (i.e., quartz and the tetrahedral face of kaolinite)), most residues that were in contact with the surface were basic. This electrostatic attraction to the surface is further supported by the large decrease in non-bonded Coulombic energies between the enzyme and the surface (Table 3.4). For the positively-charged kaolinite surface, a plurality of residues in contact with the mineral surfaces were polar residues within the circumneutral (MAN) and negatively-charged (CHI) enzymes. The negatively-charged enzyme (CHI) also had a high number of hydrophobic residues in contact with the surface, which is consistent with the higher van der Waals potential energies between the enzyme and mineral surface relative the Coulombic interactions. These weaker interactions may imply that adsorption is not as strong and may be overcome by stronger forces, but they may also be less resistant to changes in geochemistry since it is not based on electrostatic charge.

These molecular dynamics simulations suggest that enzyme-mineral interactions are diverse and can depend on both the structure of the enzyme and identity of the mineral surface. Further characterizing the extracellular enzyme pool present within the active-layer community by describing their primary structure and folded state could help predict the sensitivity of adsorption of different CAZymes to the presence of different minerals and changes to the geochemistry of the soil.

4.2. Influence of Adsorption to Enzyme Structures

From the set of enzymes and surfaces simulated in this study, there was no general trend of enzyme distortion (as measured from RMSD) as a function of the amount of contact the enzyme had during adsorption to a mineral surface. Therefore, these simulations do not support the secondary hypothesis that more deformation would occur as more contact is made between an enzyme and a mineral surface. The changes in enzyme structure appear to be specific to each system. NAG simulations exhibited the least amount of deviation from the ColabFold prediction and each system's average RMSD were grouped together, most within error of each other (Table 3.3). The smaller distribution of RMSD values for NAG on quartz and kaolinite suggests that adsorption stabilized the structure by decreasing the fluctuations of the backbone. This low degree of deformation implies that this enzyme did not dramatically unfold, but rather had subtle changes in structure.

For MAN and CHI, there were consistent increases in mean RMSD values relative to the surface-free simulations for the quartz and muscovite systems (Figure 7). This relationship does not appear to be related to adsorption as only one pair resulted in adsorption (MAN-Qtz). I hypothesize that this consistent increase in enzyme backbone distortion for quartz and muscovite systems is related to the increase of ions in the solution phase. During the simulation, surface sodium ions for quartz and potassium ions for muscovite dissociated from the surface and were free in solution. These ions likely become associated with the enzyme structure and cause it to "bloat". This is supported by the large decrease in electrostatic interaction energies between the enzyme and the solution (Figure 14), likely caused by the close proximity of solution ions to the enzyme structure. There was also an increase in non-bonded interactions between the enzyme itself, suggesting an unfavorable structural conformation. There was not a comparable increase in deformation for NAG with quartz and muscovite systems.

4.3. Potential Consequences for Enzyme Activity and Lifetimes

The comparison of differently temperature-adapted enzymes shows that different structural characteristics impact the enzyme's activity and lifetimes (Feller & Gerday, 2003). Specifically, adding flexibility to the enzyme structure has been shown to lower the energy transition states of psychrophilic enzymes, thus promoting their activity, while increasing structural rigidity and allowing thermophilic enzymes to persist at higher temperatures (Bjelic et al., 2008). This implies that enzymes in cold environments may be more active with shorter lifetimes or be less

active with longer lifetimes. Where an enzyme lands on this spectrum can be inferred from the fluctuations of the structure in the MD simulations.

NAG demonstrated the lowest overall backbone fluctuations at equilibrium, excluding some of the terminal residues that had some of the highest RMSF values (Figure 8). Regions that did have higher RMSF values were typically correlated to sections of residues that were not in secondary structures such as helices or sheets. Thus, the overall lower RMSF values for the NAG enzyme is likely due to having the most residues in secondary structures relative to the other enzymes in the study (Table 3.2). The beta sheets were also lower in RMSF values than the other residues and the highest values were associated with disordered chain regions (Figure 9). Mean RMSF values of the bulk structure for NAG decreased for quartz and kaolinite systems relative to the surface-free simulation, but variations in values were high (Table 4.1). Separately, binding site residues seemed to maintain flexibility if not slightly increase it (Figure 16). This may indicate that adsorption onto mineral surfaces may decrease the flexibility of the enzyme structure, therefore making it more resistant to degradation. Additionally, the binding site is separate from the adsorbing region of the enzyme and site residues maintain flexibility. Therefore, NAG may be an enzyme that maintains enzymatic activity while preserving itself on mineral surfaces.

For MAN, there were more regions with higher fluctuations (>0.3 nm) than in the other enzymes (Figure 10). This is likely a result of MAN having more residues in disordered chains than NAG or CHI (Table 3.2). The lower median RMSF values between the alpha helix and beta sheet residues relative to the disordered chain regions is more pronounced. The bulk structure fluctuations were slightly higher for quartz and muscovite simulations. The active-site residues also increased for these systems (Figure 17). The increased flexibility of MAN in systems that involve release of salt ions into solution may be related to the halotolerance of similar structures. This has been previously observed in MD simulations of a halophilic *Vibrio salmonicida* endonuclease that demonstrated higher flexibility at higher salt concentrations (Benrezkallah et al., 2015). Structures like MAN may be capable of increased activity during cold seasons where freezing of porewater concentrates solutes into brine veins. While the overall structure flexibility allows for high activity in cold temperatures, higher structural flexibility is also typically associated with lower thermodynamic stability (Feller et al., 1999). Microbes would therefore need to invest in producing more of these enzymes to keep degrading complex organic carbon

Table 4.1: Average RMSF for active site and non-active site residues. Bulk values represent averages that include both types of residues.

Enzyme	Surface	RMSF (nm)		
		Bulk	Non-Active Site	Active Site
NAG	No Surface	0.1 ± 0.1	0.1 ± 0.1	0.046 ± 0.007
	Quartz	0.07 ± 0.04	0.07 ± 0.04	0.052 ± 0.007
	Muscovite	0.1 ± 0.1	0.1 ± 0.1	0.051 ± 0.005
	Kaolinite(-)	0.07 ± 0.05	0.07 ± 0.05	0.05 ± 0.01
MAN	No Surface	0.08 ± 0.09	0.08 ± 0.09	0.041 ± 0.004
	Quartz	0.09 ± 0.06	0.09 ± 0.06	0.08 ± 0.03
	Muscovite	0.10 ± 0.08	0.10 ± 0.08	0.074 ± 0.008
	Kaolinite(+)	0.08 ± 0.06	0.08 ± 0.06	0.042 ± 0.005

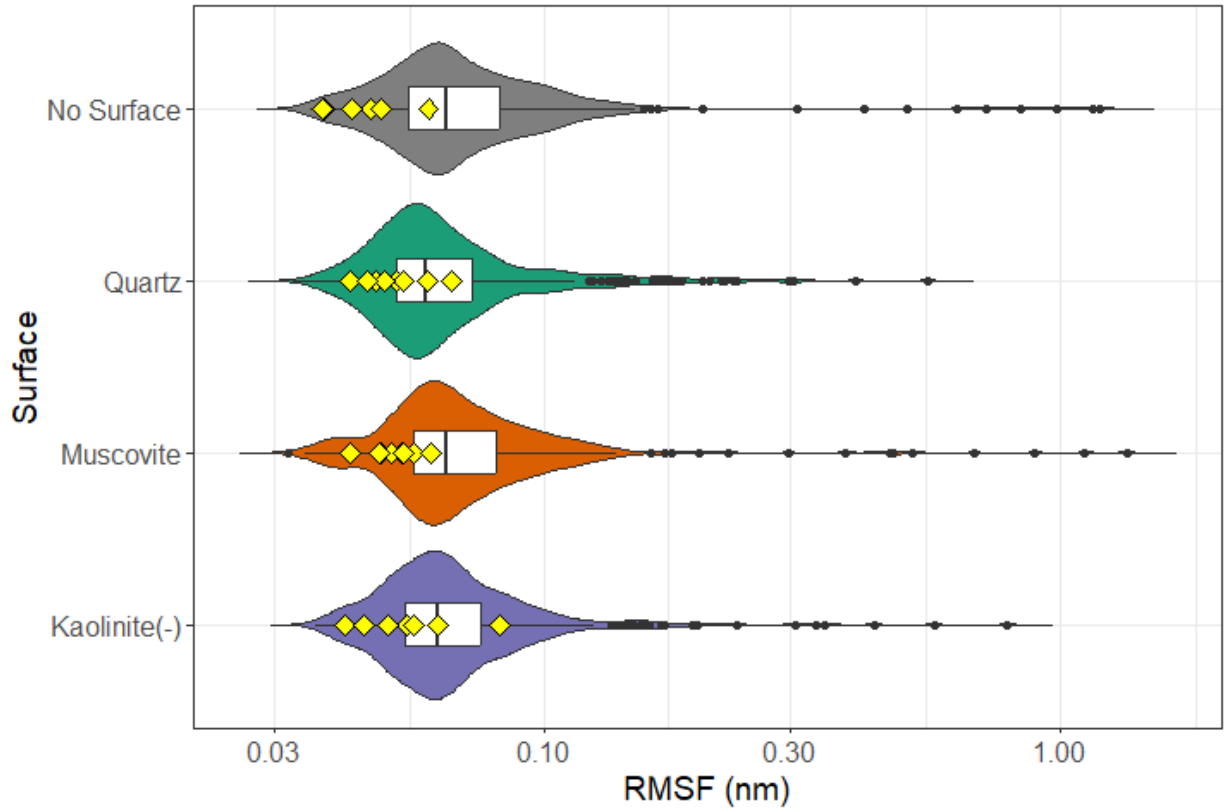

Figure 16: Distribution of RMSF values in the last 25ns of NAG simulations. Yellow diamonds indicate RMSF of residues hypothesized to be involved in the binding site of the enzyme. Values are plotted on a logarithmic scale for better visualization of the distribution of values.

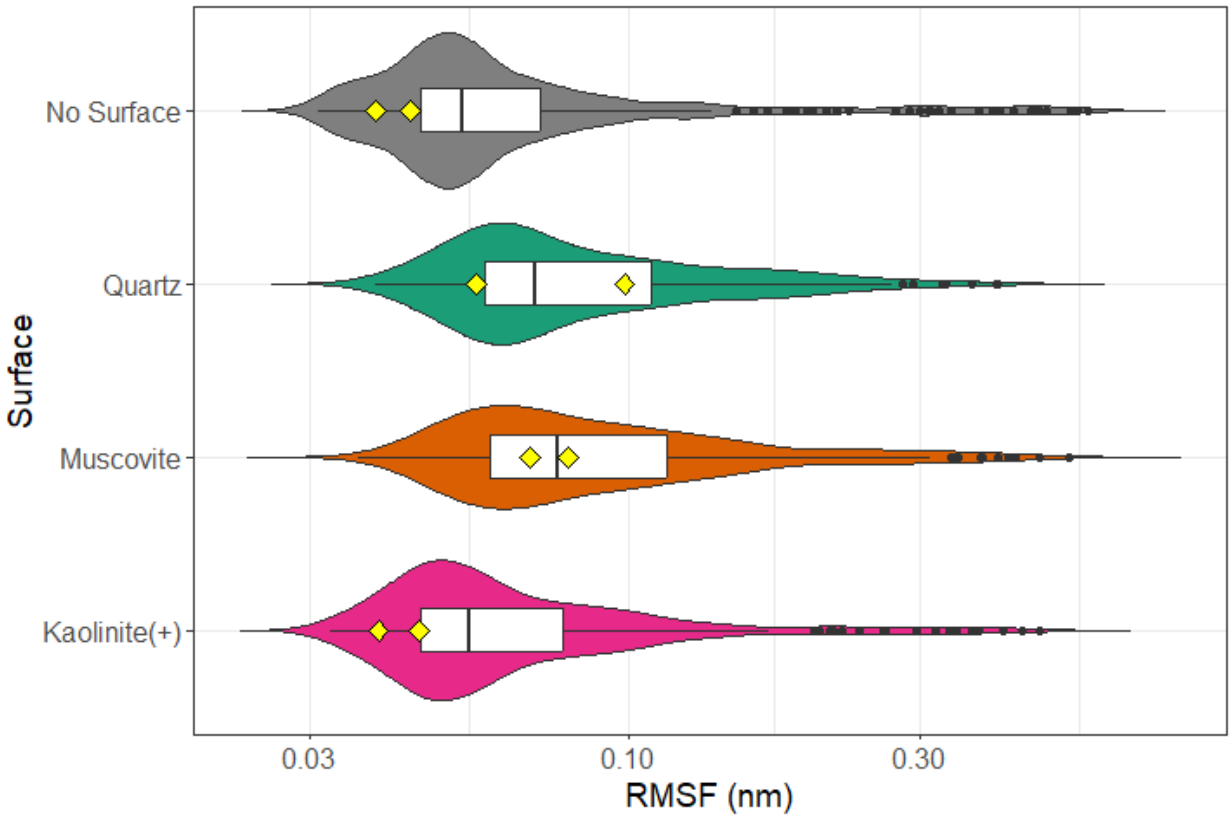

Figure 17: Distribution of RMSF values in the last 25 ns of MAN simulations. Yellow diamonds indicate residues hypothesized to be involved in the binding site of the enzyme.

like cellulose, presuming that the primary secreted enzymes to do this is similar to MAN.

For CHI, fluctuations were highest for regions that were within disordered chains (Figure 11). Residues within both alpha helices and beta sheets were systematically lower in fluctuations relative to the disordered chains (Figure 12). The fluctuations were also lowest for the kaolinite simulation, where CHI demonstrated adsorption. This supports that adsorption likely reduces enzyme flexibility and therefore increases lifetime and decreases activity.

4.4. Environmental Implications

As the Arctic becomes warmer, various processes may change the geochemistry of the soil and thus change the electrostatic interactions between enzymes and minerals. Change in porewater pH impacts the protonation state of variably-charged surface features of minerals as well as some amino acids. Weathering of calcite due to glacial meltwater runoff within Bayelva would result in increasing pH in the developing active layer soil (Hodson et al., 2002). This would result in more negatively-charged surfaces in the case of quartz. Permanent charged features of minerals may also form due to silicate weathering releasing divalent cations that can substitute aluminum within the structure and create more negative charge. Since the side chains of basic residues are not likely to change protonation state as long as pH is below 10, enzymes are likely to retain their positive charge. This would result in stronger electrostatic interactions and likely more adsorption. The increase of water in the liquid phase due to permafrost thaw increases moisture availability, which could increase chemical weathering (Hall et al., 2002). The acceleration of soil mineral weathering is likely to increase interactions between extracellular enzymes and minerals by increased structural substitutions or mineral weathering products. It is also possible that changing weather conditions will change the character of soil organic matter, leading to different concentrations of humic, fulvic, or tannic acids, thus changing soil pH, but it is not obvious in what direction that is likely to change (Adeleke et al., 2017).

Seeing deformation of enzyme structures in relation to salinity shows that active-layer soil extracellular enzymes may be sensitive to periods of porewater freezing, as ice formation excludes solutes and forms brine veins (Gilichinsky et al., 2003). The extent of this effect appears to be structure-dependent. Characterizing the enzyme pool as having small salinity-caused deformation or large salinity-caused deformation (e.g., NAG-like or MAN-like) in other depths of the active-layer and even permafrost soils that will become active-layer soils will be informative of how effective the enzyme pool is at persisting through freezing cycles and

continuing to breakdown soil organic carbon. Extracellular enzymes have persistent activity in soils by being stabilized on mineral surfaces (Allison, 2006; Schimel et al., 2017), so further studying how adsorption and salinity impact structure together would be important.

SECTION 5. CONCLUSIONS AND FUTURE WORKS

Three hypothetical extracellular enzymes were identified from metagenomic reads from an active-layer sample site from Svalbard. They were selected based on their identification as an enzyme that degrades carbon substrates, the presence of a signal peptide, and their isoelectric point. Folded structures were predicted using ColabFold and simulated in solution using GROMACS. Simulations were also run with various mineral surfaces that are representative of the environment's mineralogy: quartz, muscovite, and kaolinite.

Within 100 ns of simulation time, I observed adsorption in some, but not all, systems. Positively-charged regions of enzymes adsorbed onto negatively-charged surfaces, such as quartz and the tetrahedral layer of kaolinite. Negatively-charged regions of enzymes adsorbed onto positively-charged surfaces, like the octahedral layer of kaolinite. These results demonstrated that electrostatics is a large driver for enzyme attraction to mineral surfaces and that Coulombic interactions of charged residues with surfaces help decrease the energy of the system. Attraction of polar and hydrophobic residues to mineral surfaces were present in some simulations and had a smaller energetic change to the systems. This suggests that enzyme-mineral interactions will strongly be affected by processes that impact both charge of the enzymes and charge of the mineral surfaces. This includes changes in pH, where residues on enzymes and silanol groups on quartz can change protonation state. Weathering of minerals will also impact electrostatic interactions by substitution of cations within the crystal structure or weathering existing minerals into new ones (e.g., micas into kaolinite).

I observed changes to the overall structure of the enzymes during simulation near mineral surfaces. The influence of adsorption to the structure of an enzyme is dependent upon the nature of the structure itself. Some structures contain many secondary structural features that make it resistant to deformation upon adsorption, whereas others contain disordered chains that are more free to deviate from initial predicted positions. Differences in structural deformation upon adsorption could mean different changes in enzyme activity, as some structures may deform enough to inactivate the binding site. Two structures (NAG and MAN) with well-characterized homologues demonstrate that adsorption results in less flexibility to the broader structure while the residues associated with enzymatic activity may remain flexible. This proposes the potential that enzymes may be preserved on mineral surfaces and remain active.

This work provides a foundational set of simulations that serves the understanding of enzyme-mineral interactions of enzymes derived from metagenomics. The workflow to generate and run the MD simulations was accessible, meaning most lab groups could extend this basis set with resources most universities possess. This work has demonstrated that certain surface interactions and enzyme structural changes are influenced by its overall tertiary structure. With the accessibility of metagenomic datasets sampled all over the Arctic, those with common computational resources could run similar simulations on other enzyme structures found in a variety of geochemically-diverse environments, providing an opportunity for accessible research in the geomicrobiology of polar regions. With more extensive sampling of extracellular enzyme structures across geochemical gradients of cold regions, we could gain further insights into the functional capacity of cold-adapted enzymes in a heterogeneous system (i.e., surface and solution phases). This has the potential for applications in biotechnical fields that may be searching for lower temperature enzymes to use.

Molecular dynamics simulations illuminate the detailed changes in structure that were previously interpolated from bulk geochemical analyses. From this detail, I propose that enzyme structure may still be preserved when interacting with mineral surfaces. Further investigation into the types of structures that exist in the pool of extracellular enzymes in these soils can help determine if the present microbial community construct enzymes are resilient to their environmental conditions or if they may be more short-lived. Further simulations that vary salinity may also give insight into how these structures behave during winter seasons when porewater begins to form ice that excludes solutes and creates brine veins. This study provides preliminary evidence that increased ion concentrations may make enzyme structure more flexible. Constraining the potential activity extracellular enzymes possess under different geochemical stresses will help improve our understanding of kinetic controls on carbon cycling in the Arctic. This could be later used to improve models of carbon fluxes to the atmosphere and the positive feedback on climate change.

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Adeleke, R., Nwangburuka, C., & Oboirien, B. (2017). Origins, roles and fate of organic acids in soils: A review. *South African Journal of Botany*, *108*, 393–406. https://doi.org/10.1016/j.sajb.2016.09.002

Allison, S. D. (2006). Soil minerals and humic acids alter enzyme stability: implications for ecosystem processes. *Biogeochemistry*, *81*(3), 361–373. https://doi.org/10.1007/s10533-006-9046-2

Allison, S. D., Wallenstein, M. D., & Bradford, M. A. (2010). Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, *3*(5), 336–340. https://doi.org/10.1038/ngeo846

Audain, E., Ramos, Y., Hermjakob, H., Flower, D. R., & Perez-Riverol, Y. (2016). Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. *Bioinformatics*, *32*(6), 821–827. https://doi.org/10.1093/bioinformatics/btv674

Barbato, R. A., Jones, R. M., Douglas, T. A., Doherty, S. J., Messan, K., Foley, K. L., et al. (2022). Not all permafrost microbiomes are created equal: Influence of permafrost thaw on the soil microbiome in a laboratory incubation study. *Soil Biology and Biochemistry*, *167*, 108605. https://doi.org/10.1016/j.soilbio.2022.108605

Benrezkallah, D., Dauchez, M., & Krallafa, A. M. (2015). Molecular dynamics of the salt dependence of a cold-adapted enzyme: endonuclease I. *Journal of Biomolecular Structure and Dynamics*, *33*(11), 2511–2521. https://doi.org/10.1080/07391102.2014.1002007

Bjelic, S., Brandsdal, B. O., & Åqvist, J. (2008). Cold Adaptation of Enzyme Reaction Rates. *Biochemistry*, *47*(38), 10049–10057. https://doi.org/10.1021/bi801177k

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bradley, J. A., Singarayer, J. S., & Anesio, A. M. (2014). Microbial community dynamics in the forefield of glaciers. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1795), 20140882. https://doi.org/10.1098/rspb.2014.0882

Chivian, D., Jungbluth, S. P., Dehal, P. S., Wood-Charlson, E. M., Canon, R. S., Allen, B. H., et al. (2023). Metagenome-assembled genome extraction and analysis from microbiomes using KBase. *Nature Protocols*, *18*(1), 208–238. https://doi.org/10.1038/s41596-022-00747-x

Choi, Y. K., Kern, N. R., Kim, S., Kanhaiya, K., Afshar, Y., Jeon, S. H., et al. (2022). CHARMM-GUI Nanomaterial Modeler for Modeling and Simulation of Nanomaterial Systems. *Journal of Chemical Theory and Computation*, *18*(1), 479–493. https://doi.org/10.1021/acs.jctc.1c00996

Coolen, M. J. L., Van De Giessen, J., Zhu, E. Y., & Wuchter, C. (2011). Bioavailability of soil organic matter and microbial community dynamics upon permafrost thaw. *Environmental Microbiology*, *13*(8), 2299–2314. https://doi.org/10.1111/j.1462-2920.2011.02489.x

Datta, R., Anand, S., Moulick, A., Baraniya, D., Pathan, S. I., Rejsek, K., et al. (2017). How enzymes are adsorbed on soil solid phase and factors limiting its activity: A Review. *International Agrophysics*, *31*(2), 287–302. https://doi.org/10.1515/intag-2016-0049

Emami, F. S., Puddu, V., Berry, R. J., Varshney, V., Patwardhan, S. V., Perry, C. C., & Heinz, H. (2014). Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials*, *26*(8), 2647–2658. https://doi.org/10.1021/cm500365c

Feller, G., & Gerday, C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, *1*(3), 200–208. https://doi.org/10.1038/nrmicro773

Feller, G., d'Amic, D., & Gerday, C. (1999). Thermodynamic Stability of a Cold-Active α-Amylase from the Antarctic Bacterium *Alteromonas haloplanctis*. *Biochemistry*, *38*(14), 4613–4619. https://doi.org/10.1021/bi982650+

Flyvbjerg, H., & Petersen, H. G. (1989). Error estimates on averages of correlated data. *The Journal of Chemical Physics*, *91*(1), 461–466. https://doi.org/10.1063/1.457480

Fox, P., & Makam, R. (2011). Kinetics of model high molecular weight organic compounds biodegradation in soil aquifer treatment. *Water Research*, *45*(15), 4419–4427. https://doi.org/10.1016/j.watres.2011.05.023

Gilichinsky, D., Rivkina, E., Shcherbakova, V., Laurinavichuis, K., & Tiedje, J. (2003). Supercooled Water Brines Within Permafrost—An Unknown Ecological Niche for Microorganisms: A Model for Astrobiology. *Astrobiology*, *3*(2), 331–341. https://doi.org/10.1089/153110703769016424

Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., & Svensson, G. (2008). Vertical structure of recent Arctic warming. *Nature*, *451*(7174), 53–56. https://doi.org/10.1038/nature06502

Hall, K., Thorn, C. E., Matsuoka, N., & Prick, A. (2002). Weathering in cold regions: some thoughts and perspectives. *Progress in Physical Geography: Earth and Environment*, *26*(4), 577–603. https://doi.org/10.1191/0309133302pp353ra

Heinz, H., Lin, T.-J., Kishore Mishra, R., & Emami, F. S. (2013). Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir*, *29*(6), 1754–1765. https://doi.org/10.1021/la3038846

Hodson, A., Tranter, M., Gurnell, A., Clark, M., & Hagen, J. O. (2002). The hydrochemistry of Bayelva, a high Arctic proglacial stream in Svalbard. *Journal of Hydrology*, *257*(1–4), 91–114. https://doi.org/10.1016/S0022-1694(01)00543-1

Hutchinson, E. G., & Thornton, J. M. (1996). PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Science*, *5*(2), 212–220. https://doi.org/10.1002/pro.5560050204

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*(11), 1859–1865. https://doi.org/10.1002/jcc.20945

Joo, J. C., Pohkrel, S., Pack, S. P., & Yoo, Y. J. (2010). Thermostabilization of Bacillus circulans xylanase via computational design of a flexible surface cavity. *Journal of Biotechnology*, *146*(1–2), 31–39. https://doi.org/10.1016/j.jbiotec.2009.12.021

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, *27*(1), 112–128. https://doi.org/10.1002/pro.3280

Kern, N. R., Lee, J., Choi, Y. K., & Im, W. (2023). *CHARMM-GUI Multicomponent Assembler for Modeling and Simulation of Complex Multicomponent Systems* (preprint). Biophysics. https://doi.org/10.1101/2023.08.30.555590

Khan, A., Khan, H., Faheem, M., Zeb, A., Badshah, M., & Chung, Y. R. (2021). Isolation and Characterization of an Acidic, Salt-Tolerant Endoglucanase Cel5A from a Bacterial Strain Martelella endophytica YC6887 Genome. *Molecular Biotechnology*, *63*(4), 305–315. https://doi.org/10.1007/s12033-020-00295-3

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., & Thornton, J. M. (1997). PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences*, *22*(12), 488–490. https://doi.org/10.1016/S0968-0004(97)01140-7

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., et al. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935

Lee, J., Hitzenberger, M., Rieger, M., Kern, N. R., Zacharias, M., & Im, W. (2020). CHARMM-GUI supports the Amber force fields. *The Journal of Chemical Physics*, *153*(3), 035103. https://doi.org/10.1063/5.0012280

Li, F., Peng, Y., Chen, L., Yang, G., Abbott, B. W., Zhang, D., et al. (2020). Warming alters surface soil organic matter composition despite unchanged carbon stocks in a Tibetan permafrost ecosystem. *Functional Ecology*, *34*(4), 911–922. https://doi.org/10.1111/1365-2435.13489

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, *102*(18), 3586–3616. https://doi.org/10.1021/jp973084f

Mazola, Y., Guirola, O., Palomares, S., Chinea, G., Menéndez, C., Hernández, L., & Musacchio, A. (2015). A comparative molecular dynamics study of thermophilic and mesophilic β-fructosidase enzymes. *Journal of Molecular Modeling*, *21*(9), 228. https://doi.org/10.1007/s00894-015-2772-4

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, *19*(6), 679–682. https://doi.org/10.1038/s41592-022-01488-1

Olagoke, F. K., Kaiser, K., Mikutta, R., Kalbitz, K., & Vogel, C. (2020). Persistent Activities of Extracellular Enzymes Adsorbed to Soil Minerals. *Microorganisms*, *8*(11), 1796. https://doi.org/10.3390/microorganisms8111796

Ostrowski, A., Gundogdu, M., Ferenbach, A. T., Lebedev, A. A., & Van Aalten, D. M. F. (2015). Evidence for a Functional O-Linked N-Acetylglucosamine (O-GlcNAc) System in the Thermophilic Bacterium Thermobaculum terrenum. *Journal of Biological Chemistry*, *290*(51), 30291–30305. https://doi.org/10.1074/jbc.M115.689596

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Porter, C., Howat, I., Noh, M.-J., Husby, E., Khuvis, S., Danish, E., et al. (2023). ArcticDEM - Mosaics, Version 4.1 [Data set]. [object Object]. https://doi.org/10.7910/DVN/3VDC4W

Quiquampoix, H., Staunton, S., Baron, M.-H., & Ratcliffe, R. G. (1993). Interpretation of the pH dependence of protein adsorption on clay mineral surfaces and its relevance to the understanding of extracellular enzyme activity in soil. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *75*, 85–93. https://doi.org/10.1016/0927-7757(93)80419-F

Schimel, J., Becerra, C. A., & Blankinship, J. (2017). Estimating decay dynamics for enzyme activities in soils from different ecosystems. *Soil Biology and Biochemistry*, *114*, 5–11. https://doi.org/10.1016/j.soilbio.2017.06.023

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. https://doi.org/10.1038/nature14338

Strand, S. M., Christiansen, H. H., Johansson, M., Åkerman, J., & Humlum, O. (2021). Active layer thickening and controls on interannual variability in the Nordic Arctic compared to the circum-Arctic. *Permafrost and Periglacial Processes*, *32*(1), 47–58. https://doi.org/10.1002/ppp.2088

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., et al. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Tietjen, T., & Wetzel, R. (2003). Extracellular enzyme-clay mineral complexes: Enzyme adsorption, alteration of enzyme activity, and protection from photodegradation. *Aquatic Ecology*, *37*(4), 331–339. https://doi.org/10.1023/B:AECO.0000007044.52801.6b

Wojcik, R., Palmtag, J., Hugelius, G., Weiss, N., & Kuhry, P. (2019). Land cover and landform-based upscaling of soil organic carbon stocks on the Brøgger Peninsula, Svalbard. *Arctic, Antarctic, and Alpine Research*, *51*(1), 40–57. https://doi.org/10.1080/15230430.2019.1570784

Xie, Y., An, J., Yang, G., Wu, G., Zhang, Y., Cui, L., & Feng, Y. (2014). Enhanced Enzyme Kinetic Stability by Increasing Rigidity within the Active Site. *Journal of Biological Chemistry*, *289*(11), 7994–8006. https://doi.org/10.1074/jbc.M113.536045

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), W445–W451. https://doi.org/10.1093/nar/gks479

Zhang, Y., Li, D., Xin, G., Jiu, H., & Ren, S. (2024). Molecular Insights into CO_2 -Rich Industrial Waste Gas Enhanced Shale Gas Recovery and Sequestration in Real Shale Gas Reservoirs. *Energy & Fuels*, *38*(1), 285–313. https://doi.org/10.1021/acs.energyfuels.3c03821

Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., & Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, *51*(W1), W115–W121. https://doi.org/10.1093/nar/gkad328

APPENDIX

System Size Suitability

To assure that there are no artifacts within the dynamics and energy interactions due to poor system size, multiple simulations were run of different dimensions to find the smallest sized box that converges upon energy values of larger sized boxes but minimizes computational costs. Each simulation used NAG as the example enzyme since it was the largest enzyme. It is expected that the smallest suitable simulation box for NAG is also suitable for the smaller enzymes. The simulations are summarized in Table A1.

The nonbonded interaction terms (Coulombic and Lennard-Jones) were calculated during the trajectory (Figure 18). It appears that each system achieved an equilibrium with respect to energy interactions by 5ns. Mean values for each energy term were calculated from 5 to 10ns and standard errors were estimated using a block averaging method (Table A2, Flyvbjerg & Petersen, 1989).

To best evaluate the system size dependencies of these different energy terms, we created linear models and plots of the energy values as a function of the number of water molecules in the simulation (Figure 19).

All slope values were also reported in terms of the percent of the energy value of the 200 x 200 x 200 simulation (Table A3). Since this system is the largest, it is expected to have the least system size effect and thus would be closest to the “true” energy value. The largest magnitude slope (0.28%) was for the short-range Coulombic forces between the enzyme and solvent. This slope implies that our smallest simulation size may be up to 5% lower just from having less water molecules. However, the actual measured value for our smallest simulation was higher than our largest simulation. This may be due to variability in values from the other systems affecting the linear model. I interpret this to mean that the smallest simulation box was sufficiently large to have minimal system size effects.

Table A1: System details of simulation set used for system size investigations. The smallest simulation was run for 100 ns since it was used for analysis in the remainder of the thesis.

Initial Box Dimensions	No. water molecules	No. ions (Na^+/Cl^-)	Simulation Time (ns)
140 x 140 x 140	84,895	15/43	100
150 x 150 x 150	105,079	19/47	10
175 x 175 x 175	168,987	31/59	10
200 x 200 x 200	253,289	47/75	10

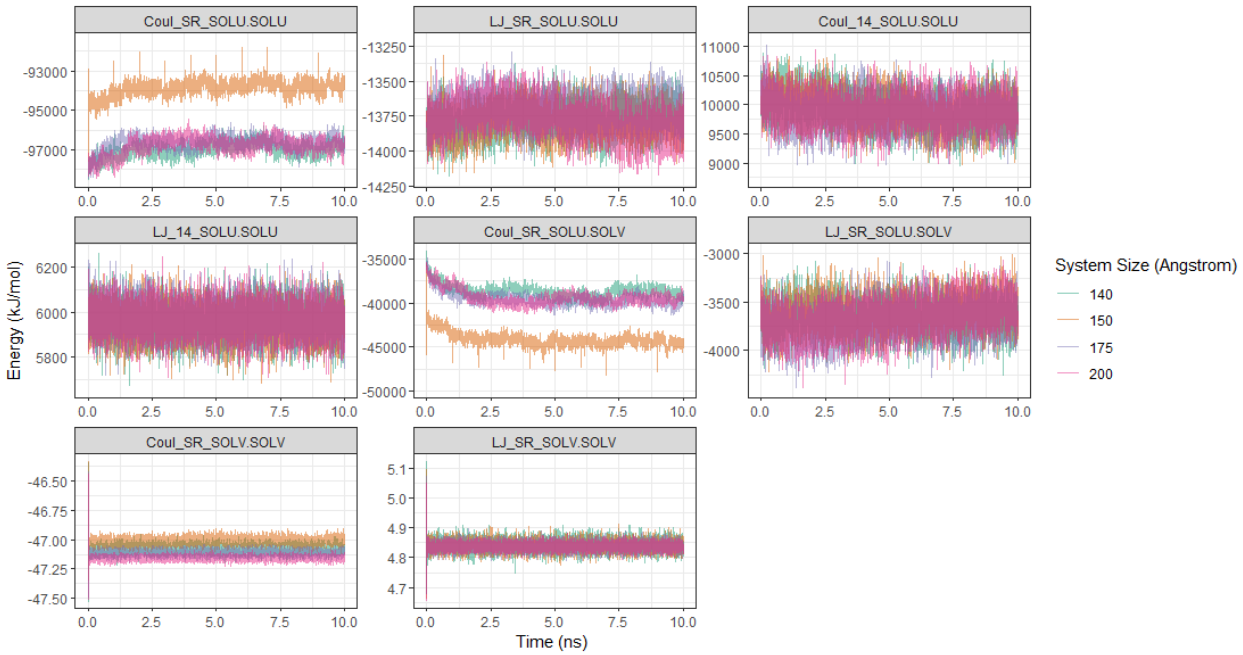

Figure 18: Energy values for different sized simulation boxes in from 5 to 10 ns for Coulombic (Coul) and Lennard-Jones (LJ) forces for short-ranged (SR) and 1-4 (14) interactions. Each facet represents a different pairing of enzyme (SOLU) and solvent (SOLV). Solvent-solvent interactions were normalized by the number of water molecules.

Table A2: Interaction potential energies of the enzyme and solvent groups for each system size. System size values are in Angstroms and represent the cubic cell side length. Solvent-solvent values are normalized by the number of water molecules.

Group Interaction	Interaction Energy (kJ/mol)	System Size			
		140	150	175	200
Enzyme-Enzyme	Coul (SR)	-96,900 ± 100	-93,750 ± 40	-96,580 ± 40	-96,760 ± 40
	Coul (1-4)	9,840 ± 20	9,820 ± 30	9,860 ± 20	9,900 ± 20
	LJ (SR)	-13,770 ± 10	-13,770 ± 6	-13,695 ± 6	-13,850 ± 20
	LJ (1-4)	5,960 ± 3	5,953 ± 2	5,980 ± 3	5,958 ± 2
Enzyme-Solvent	Coul (SR)	-38,900 ± 200	-44,420 ± 50	-39,820 ± 80	-39,490 ± 70
	LJ (SR)	-3,669 ± 8	-3,590 ± 10	-3,670 ± 10	-3,600 ± 20
Solvent-Solvent	Coul (SR)	-47.0777 ± 0.0007	-47.006 ± 0.003	-47.1139 ± 0.0007	-47.1575 ± 0.0004
	LJ (SR)	4.8430 ± 0.0008	4.8412 ± 0.0004	4.8386 ± 0.0002	4.8381 ± 0.0004

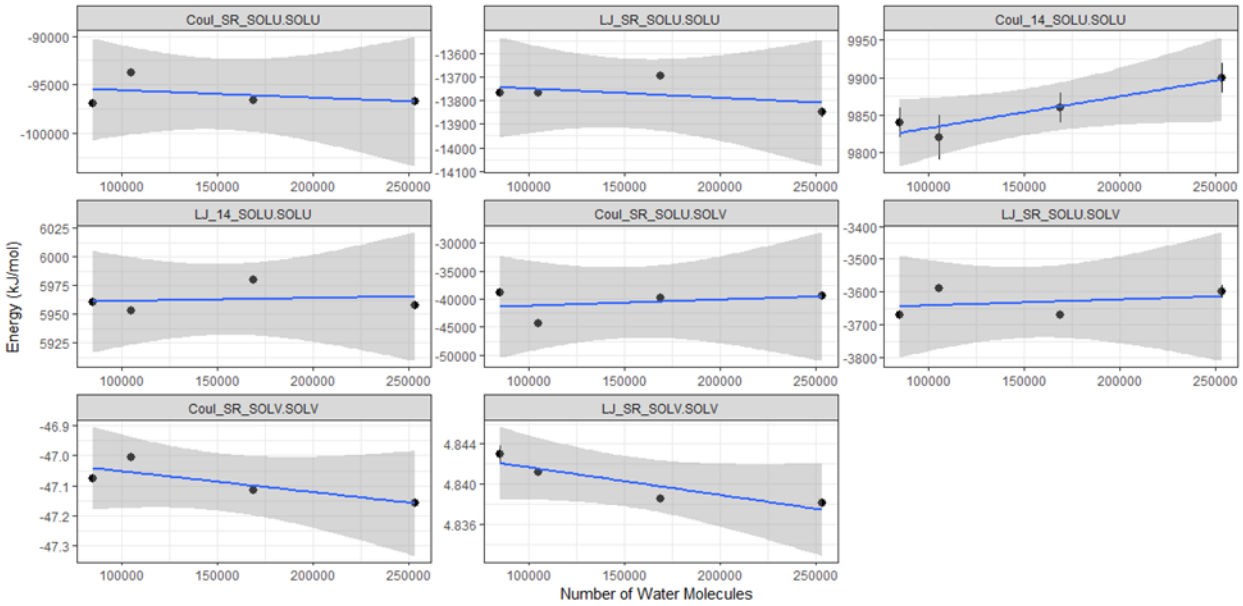

Figure 19: Estimated equilibrium energy values for each simulation system with linear models calculated in R.

Table A3: Linear model slopes as a function of the number of water molecules. Percentages are calculated relative to the largest system energy values.

Group Interaction	Interaction Energy	Model Slope (kJ/mol per 10,000 molecules)
Enzyme-Enzyme	Coul (SR)	-76.54 (0.08%)
	Coul (1-4)	4.228 (0.04%)
	LJ (SR)	-3.956 (0.03%)
	LJ (1-4)	4.228 (0.07%)
Enzyme-Solvent	Coul (SR)	110.8 (0.28%)
	LJ (SR)	1.803 (0.05%)
Solvent-Solvent	Coul (SR)	-0.006966 (0.01%)
	LJ (SR)	-0.0002753 (0.006%)

Table A4: Simulation system descriptions.

System	Simulation Box Dimensions	Number of ions (Na^+/Cl^-)	Surface Ions	Number of Water Molecules
NAG-Solvated	140 x 140 x 140	15/43	NA	84,895
NAG-Qtz	142.477 x 144.6632 x 141	13/41	524 Na^+	76,288
NAG-Musc	140.1786 x 144.2448 x 160	15/43	864 K^+	86,307
NAG-Kaol	144.312 x 143.071128 x 172	17/45	NA	95,361
MAN-Solvated	140 x 140 x 140	20/16	NA	85,969
MAN-Qtz	142.477 x 144.6632 x 157	20/16	524 Na^+	88,139
MAN-Musc	140.1786 x 144.2448 x 160	20/16	864 K^+	87,294
MAN-Kaol	144.312 x 143.071128 x 172	20/16	NA	89,286
CHI-Solvated	140 x 140 x 140	35/16	NA	85,429
CHI-Qtz	142.477 x 144.6632 x 157	37/16	524 Na^+	87,616
CHI-Musc	140.1786 x 144.2448 x 160	36/15	864 K^+	86,798
CHI-Kaol	144.312 x 143.071128 x 162	35/16	NA	88,892

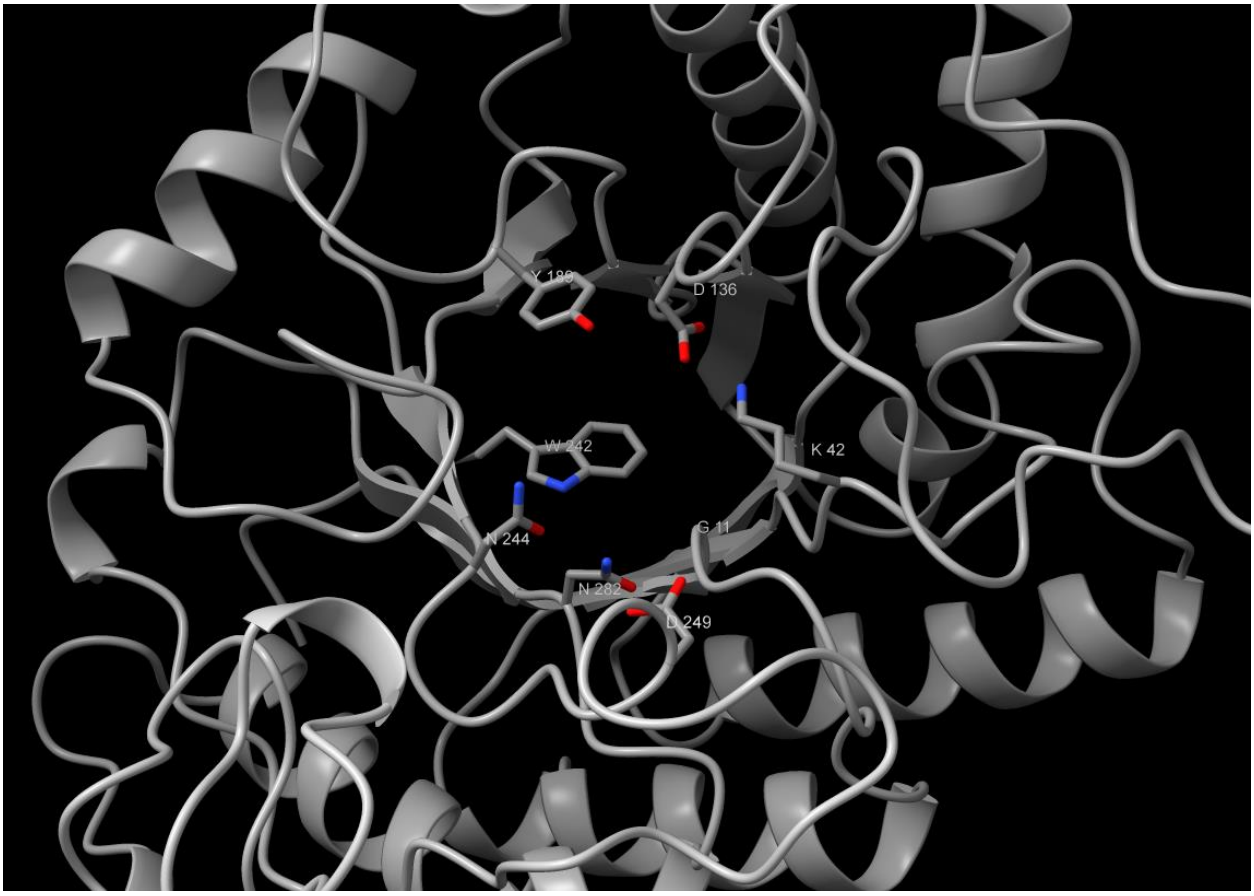

Figure 20: Hypothetical binding residues within the NAG structure. Binding site atoms are explicitly shown and labelled with residue code and number.

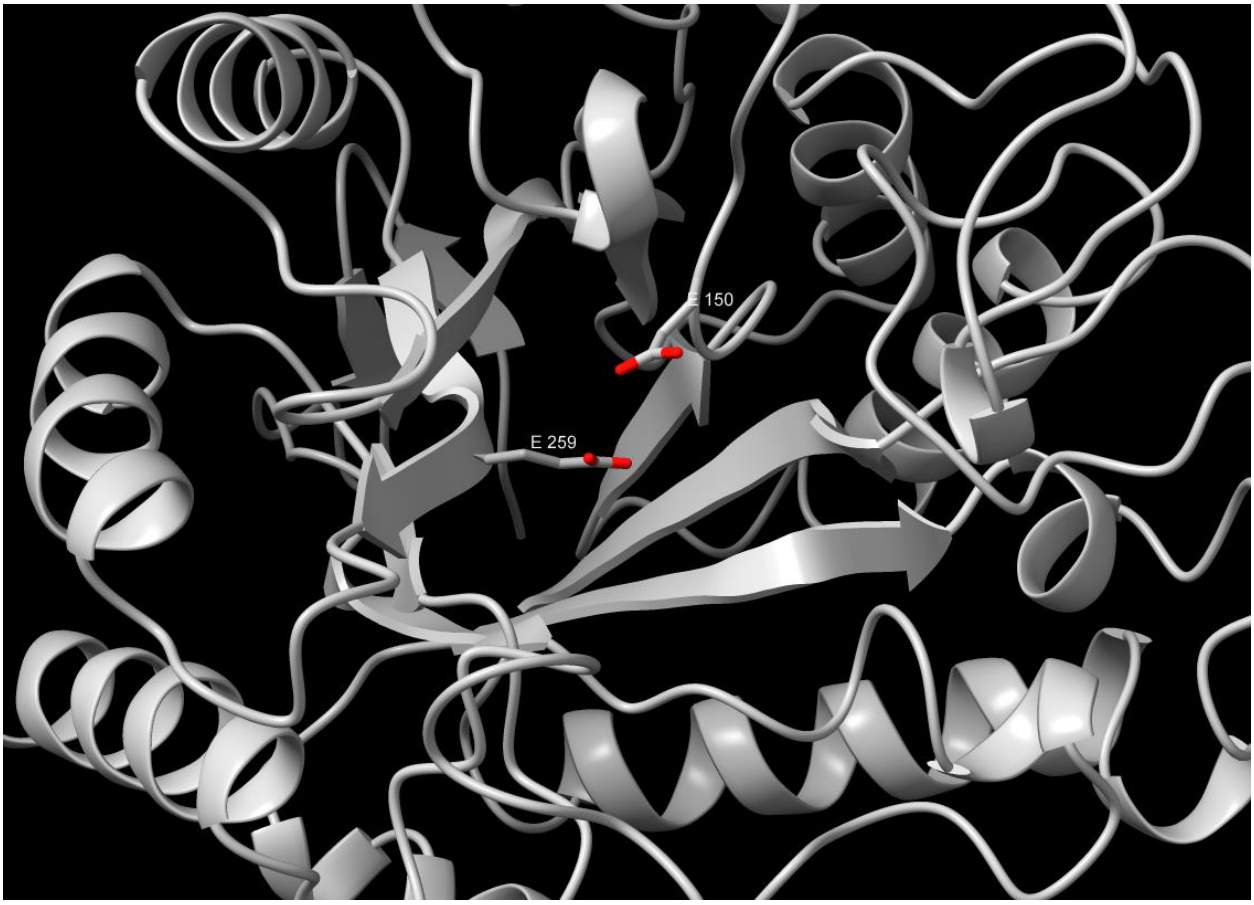

Figure 21: Hypothetical active site residues for the MAN structure. Active site atoms are explicitly shown and labelled with residue code and number.

VITA

Jacob Perez was born and raised in Phoenix, Arizona. Jacob attended Arizona State University where he received two Bachelor of Science degrees in Chemistry and Physics in 2018. He also completed his honors thesis titled “Practices and Perceptions of Supplements among College Students” with Barrett, the Honors College. After completing his undergraduate degrees, Jacob worked as a quality control chemist where he expanded his laboratory skills operating an ICP-MS and UPLC. With these new analytical tools under his belt, he searched to further his interest in exploring many of the world’s phenomena by beginning graduate school. That is when Jacob started his studies in geology in the Earth and Planetary Sciences department at the University of Tennessee, Knoxville in 2020. After gaining a foundation in the field of geochemistry, Jacob joined Dr. Andrew Steen’s lab in the summer of 2021 to study the geomicrobiology of Arctic soils. In March 2022, Jacob went on his first field expedition in Ny Ålesund, Svalbard where he assisted in taking gas flux measurements over the area’s frozen soils. After his adventure in the northernmost active settlement, Jacob solidified his love for studying the microbiology in the world’s poles. In the summer of 2022, Jacob participated in the International Geobiology Course hosted by Caltech where he studied geomicrobiology of Mono Lake and ancient methane seeps in California. He later presented his groups work at the American Geophysical Union Fall Meeting in Chicago, IL in December 2022. In the spring of 2023, Jacob then embarked on a one-month journey on the RVIB Nathaniel B. Palmer to sample marine sediments from the continental shelf across East Antarctica. Outside of his research and collaborations, he also worked as a graduate teaching assistant for introductory environmental geology labs.