Summer 2023

Using QM/MM methods to explore Sortase Enzyme Intermediates, Kinetics, and Stability

Kyle Whitham
Western Washington University, kylemwhitham@gmail.com

Using QM/MM methods to explore Sortase Enzyme Intermediates, Kinetics, and Stability

By

Kyle Maxwell Whitham

Accepted in Completion
Of the Requirements for the Degree
Master's of Science

ADVISORY COMMITTEE

Chair, Dr. Jay McCarty

Dr. John Antos

Dr. Tim Kowalczyk

GRADUATE SCHOOL

Dr. David L. Patrick, Dean

Master's Thesis

In presenting this thesis in fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Kyle Whitham

8/1/2023

Using QM/MM methods to explore Sortase Enzyme Intermediates, Kinetics, and Stability

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfilment
Of the Requirements for the Degree
Master's of Science

By
Kyle Maxwell Whitham
August 2023

Abstract

Biochemistry has seen advancements in methods and understanding of the inner workings of proteins, yet biochemists struggle to see real time reaction pathways of protein intermediates. This is where computational chemistry comes in and fills in the holes in knowledge through the use of Quantum Mechanical (QM) models. QM chemistry alone does not give results in a reasonable timescale to predict protein chemistry in a reasonable amount of time. Computational chemistry methods such as Quantum mechanical (QM)/ Molecular Mechanical (MM) (QM/MM), allow us to split the *in-silico* system into two regions that utilize a fast MM force field region and slow, but a chemically accurate, (QM) region. Our project will utilize docking programs and QM/MM methods to give accurate results in a reasonable timeframe to show the intermediate pathway in the Sortase A/B enzyme, meta-stable and stable intermediate information on the ligand and Sortase B enzyme structure. Several docking suites: AutoDock, FlexPepDock, a manual docking procedure, and a predicted folded structure from AlphaFold will be used to dock the *Bacillus anthracis* Sortase B (baSrtB) ligand to the active site of the enzyme. baSrtB simulation is supplemented by single point energy calculations of specific frames of structures to be used as a reference to compare to the Free Energy Surface (FES) results. This thesis will show that we have successfully modeled the Sortase B from *Bacillus anthracis* showing the accurate intermediate pathway.

Acknowledgements

I would like to give thanks to:

Dr. Jeanine Amacher, Sophie Jackson, and their Lab, for helping make this work come into its own and allowing me to learn wet lab biochemistry techniques that I otherwise would have missed out on.

Dr. John Antos and Dr. Tim Kowalczyk, for their contributions and help with formulating my thesis.

Dr. Jay McCarty for helping me properly learn and use GROMACS, CP2K, PLUMED, and his time mentoring me and giving guidance.

The NSF for funding my research over the Summer, Fall and winter of 2022 and 2023.

Table of Contents

List of Figures, Schemes, and Tables

1 Introduction:

In the field of biochemistry, a long-standing goal is to know at atomistic resolution a reaction pathway inside of a protein when multiple complex intermediates are present. High-energy transitions states that represent kinetic barriers are short-lived,[1] making conventional biochemistry methods too slow to see the intermediates happen in real time. Structural biochemistry methods that rely on creating a covalently bound intermediate with a resolved crystal structure, which is obtained either through X-ray crystallography or through Cryogenic Electron Microscopy (Cryo EM)[2], can take weeks or months to grow a crystal and a usable crystal can take longer to form. On top of growing a crystal, a protein crystal can fail in Cryo EM due to a crystal cracking that occurs when transferring a crystal to the Cryo solution rendering months of work useless. As an alternative, computational chemistry working in tandem with structural biochemistry can provide greater insight into the mechanisms of enzymes.

Computational chemistry allows one to investigate the dynamics of a molecule as the system evolves in time on the picosecond to up to microsecond time scales[3]. Sampling dynamic conformations at this time scale allows biochemists to quantify structural fluctuations, observe interactions of small molecules with amino acid side chains important for ligand specificity and recognition, and track reaction mechanisms[3].

In this work, several different, but complementary, computational chemistry methods will be employed. Molecular Dynamics (MD) uses classical Newtonian physics to propagate the system in time using Molecular Mechanics (MM) force fields [3]. Performing Classical MD on modern GPU hardware makes it possible to run trajectories in the microsecond timescale for medium-sized proteins in solution with an integration timestep of 1 fs. Classical MD is thus useful for studying protein conformational dynamics that evolve over 100s of nanoseconds where no covalent bonds are breaking or forming[3]. This makes MD simulations good for visualizing protein stability, hydrophobic/hydrophilic interactions, and dimerization interactions on the surface of the protein. For viewing bond breaking/forming events we need to use Quantum Mechanics (QM) to calculate the ground state electronic structure that changes during the formation or breaking of a chemical bond. QM simulations, while more accurate, are more computationally expensive than MD simulations[4]. In this work, I will describe my efforts

to use a hybrid approach, QM/MM that is capable of bridging the gap between these two methods and enable more efficient simulations that can show bond breaking and forming events within protein enzymes. My goal is to apply these methods to study sortase chemistry that has important application in human health and protein engineering.

The thesis is organized as follows. First, in the Introduction, I will describe the necessary background on sortase chemistry, motivation for this research, and the computational methods that will be used in this study. Next, I will describe different computational approaches I have employed to study sortase chemistry. I will discuss molecular docking approaches to predict the holoenzyme/ligand complex structure. This provides a starting structure from which to perform QM/MM simulations. I will then describe QM/MM simulations of three different systems: 1) sortase A in solution, 2) sortase B in solution, and 3) a smaller, minimal model of the sortase B active site in complex with the ligand, that is more computationally efficient than the full systems. After a discussion of the results of these simulations, I will discuss future directions and areas for continued research.

1.1 Sortase Chemistry:

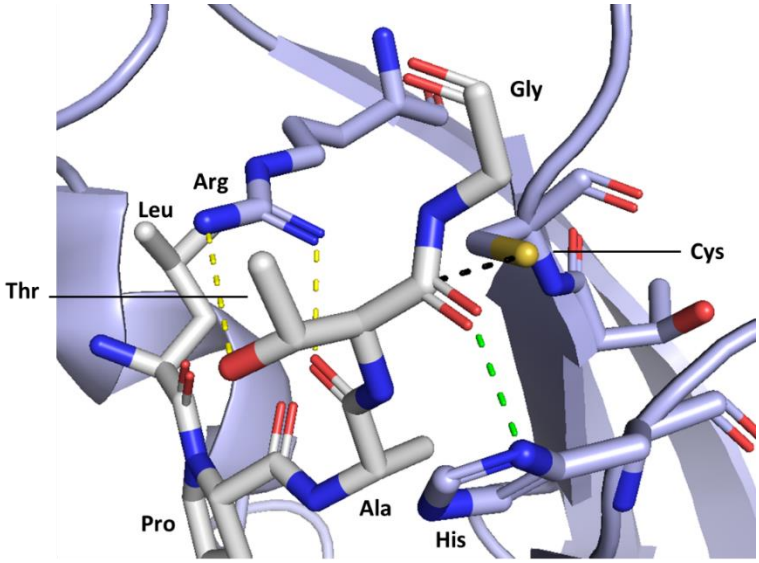
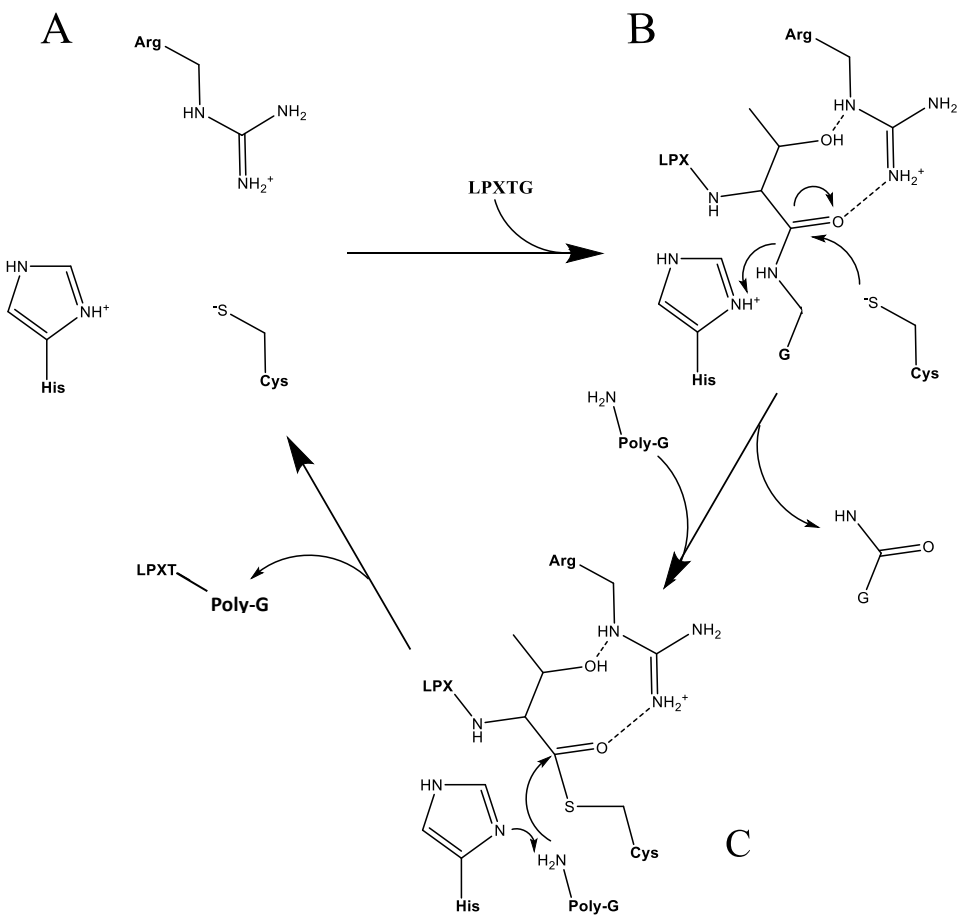


Figure 1) In white the LPATG substrate peptide is shown bound in the active site of *Staphylococcus aureus* Sortase A (Periwinkle). On the Sortase A protein the catalytic cysteine and histidine with the stabilizing arginine are labeled. The target for catalysis is shown with the black dotted line to form the thiol bond. The intended targets for the stabilization are shown in yellow and green for arginine and histidine respectively.

Sortase chemistry has been increasingly relevant to the field of protein engineering. Sortase enzymes are cysteine transpeptidases localized to the surface of Gram-positive bacteria. They play an important role in attaching proteins to the peptidoglycan cell wall. Sortases recognize and cleave specific target sequences, followed by ligation to a second fragment that contains an amine nucleophile [4-7]. The most studied sortase is Sortase A from *Staphylococcus aureus* (saSrtA). saSrtA recognizes the pentapeptide LPXTG motif, cleaving between the T and G residues, then ligating the LPXT- fragment to a lipid II molecule that is embedded in the cell wall[4-6]. The ligation of the LPXTG motif is made possible by a stable intermediate formed by the catalytic cysteine 208 and substrate Threonine (P1 position) peptide carbon shown in figure 1. The full catalytic pathway with the thioester intermediate formed by the cysteine is then broken when a poly-Glycine nucleophile comes in and ligates the thiol intermediate to the poly-Glycine tail and restores the cysteine active site shown in Scheme 1. While most of the literature concerns the saSrtA, other sortases recognize a different peptide motif and have different activity and selectivity. Despite these differences, every Sortase A and B is expected to follow the same basic catalysis method involving a catalytic triad of His, Cys, Arg that is highly conserved among class A and B sortases. However, the conservation of the catalytic triad does not explain the differences in enzyme kinetics between Sortase A and B. Differences in the recognition motif based on the class of sortase and organism it is found in[9] are shown in Table 1.

Table 1) Sortase Recognition Sites

Protein name	**Recognition Peptide Motif**
Staphylococcus Aureus Sortase A	LPXTG
Streptococcus Pyogenes Sortase A	LPXTG
Bacillus Anthracis Sortase B	NPKTG



Scheme 1) Shown is the full reaction pathway from Apo-enzyme (A) to the intermediate formed by the *Staphylococcus aureus* Sortase A of the LPXT-Cys thiol bond (B). Finally releasing the LPXT-thiol substrate via nucleophilic attack by the Poly-Glycine (C).

The transpeptidation reaction of the sortase enzyme makes sortase-mediated-ligation (SML) a powerful tool in protein engineering. A truncated version of the protein consisting of just the carboxy terminal catalytic domain is soluble in water, making the enzyme suitable for *in vitro* applications. SML utilizes [9] two different polypeptide chains with the five amino acid sortase recognition motif on one of the target proteins, and a poly-Glycine tail that has been added to the end of the other target protein. Following the same reaction scheme as Scheme 1, SML can be used to ligate drug payloads, polypeptides, synthetic proteins, and solid support particles to one another. The ability to ligate different targets, especially drug payloads, has allowed SML to come into the eye of biochemists. However, recent work suggests that saSrtA may not be broadly representative of other Sortase A enzymes and biochemists still have a very limited understanding of the catalytic mechanism of other classes of sortases including the class

B Sortases. Class B sortases have a different recognition motif, NPKTG, from Class A sortases. This change in recognition motif could guide us to make discoveries about binding specificity. With only two changes in amino acid sequence (from LPXTG to NPKTG) in saSrtA to baSrtB, we expect that a major charge-based interactions are occurring with the Asn and Lys of the baSrtB ligand to key residues on the protein structure to allow substrate recognition.

1.2 Molecular mechanics, Quantum Molecular Mechanics, and QM/MM:

Molecular dynamics (MD) is a computational method for investigating the motion of atoms given some initial value of the nuclear positions and velocities and the forces acting on each atom. The equation of motion is computed numerically on the computer at a finite (small) timestep. After a period of equilibration, the dynamic trajectory of the atomic positions as a function of time can be used to calculate observables of interest. These may be equilibrium (thermodynamic) quantities such as energies or dynamic quantities such as time correlation functions.

In general, MD simulations can be performed using any suitable method of describing the molecules and calculating the forces acting on the atomic nuclei. The starting point for any non-relativistic quantum system is the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(\{R_I\},\{r_i\},t) = \widehat{H}(\{R_I\},\{r_i\})\Psi(\{R_I\},\{r_i\},t) \qquad (1)$$

where $\Psi(\{R_I\},\{r_i\},t)$ is the wavefunction for the molecular system at time, t, that depends on the set of nuclear coordinates ($\{R_I\}$) and the set of electronic degrees of freedom ($\{r_i\}$), and $\widehat{H}(\{R_I\},\{r_i\})$ is the Hamiltonian operator that contains kinetic energy terms for both the nuclei and electrons and the electron-electron, electron-nuclear, and nuclear-nuclear Coulomb interactions. A next step is to introduce the famous "Born-Oppenheimer" approximation by considering the electronic part of the Hamiltonian for fixed nuclei. The ground state electronic energy is then computed from the *time-independent* Schrödinger equation for the electrons at a *fixed* nuclear configuration {**R**}:

$$\widehat{H_e}\psi_e = E_0\psi_e \qquad (2)$$

where it is noted that $\widehat{H_e}$ depends parametrically on the nuclear positions ({**R**}). In *ab initio* Molecular Dynamics (AIMD), the electronic structure problem is solved concurrently to propagating the atomic nuclei according to classical mechanics[10]. Because the electronic energy depends parametrically on the nuclear positions, the electronic energy must be computed at each iteration. From the perspective of the atomic nuclei, the nuclei move along an effective Born-Oppenheimer potential energy surface determined from the electronic degrees of freedom. The atomic nuclei are updated according to the equation of motion:

$$M_I \frac{d^2 R_I}{dt^2} = -\nabla_{R_I} \min_{\psi_0} \{\langle \psi_0 | \widehat{H_e} | \psi_0 \rangle\} \quad (3.a)$$

where

$$\min_{\psi_0} \{\langle \psi_0 | \widehat{H_e} | \psi_0 \rangle\} = V^{BO}(\{R\}) \quad (3.b)$$

defines an effective potential energy surface obtained by solving the time-independent Schrödinger equation (Equation 2). AIMD is not tied to any particular electronic structure method used to solve the time-independent Schrödinger equation, although very accurate methods are often prohibitively expensive. Several different methods have been used including density functional theory (DFT)[11], Hartree-Fock[12], generalized valence bond (GVB)[13], complete active space SCF (CASSCF)[13], second-order many-body perturbation theory (MP2)[14], coupled cluster[15], and semiempirical methods[12]. Because it is often the best compromise between accuracy and computational efficiency, Kohn-Sham (KS) DFT[16]is a popular QM method for AIMD applications. In KS DFT, the ground state energy of an interacting system of electrons is obtained by minimizing the Kohn-Sham energy functional of the electron density:

$$E_0 = \min_{\psi_0} \{\langle \psi_0 | \widehat{H_e} | \psi_0 \rangle\} = \min_{\{\phi_i\}} E^{KS}[\{\phi_i\}] \quad (4)$$

where the energy function $E^{KS}[\{\phi_i\}]$ is optimized with respect to a set of auxiliary functions $\{\phi_i\}$, the Kohn-Sham orbitals that satisfy the orthonormality relation:

$$\int d\boldsymbol{r}\, \phi_i^*(r)\phi_j(\boldsymbol{r}) = \delta_{ij} \quad (5)$$

The advantage of this approach is that the full many-body wavefunctions $\psi_0(r_1, \ldots, r_N)$ are replaced by solving a system of single-electron equations (the Kohn-Sham equations) for a

system of non-interacting particles with a local potential energy that generates the same density as the fully interacting system:

$$\rho(r) = \sum_i |\phi_i(r)|^2 \quad (6)$$

The KS energy functional has the form

$$E^{KS}[\{\phi_i\}] = E[\rho] = E_{KE}[\rho] + E_{ext}[\rho] + E_{Hartree}[\rho] + E_{ex}[\rho] \quad (7)$$

where the first term is the kinetic energy of a non-interacting reference system of electrons, the second term comes from the electron motion in a fixed external potential defined by nuclear positions and charges, the third term is the Hartree energy, and the last contribution is the exchange-correlation energy that results from the difference between the exact energy and the decomposition into the three previous Kohn-Sham terms. All the terms in the above expression are known apart from the exchange-correlation term. In practice, several different exchange-correlation functionals are widely used including the famous "Local Density Approximation" (LDA), "Generalized Gradient Approximation" (GGA)[17], or hybrid functionals that include to some extent the Hartree-Fock exchange in addition to the standard GGA expression[18]. In addition to the choice for the exchange-correlation functional, the basis set to use to represent the orbitals ϕ_i must also be specified. In general, the orbitals are represented as a linear combination of basis set functions:

$$\phi_i(r) = \sum_{k=1} a_k g_k(r) \quad (8)$$

where $\{g_k(r)\}$ are the basis functions that may be localized Gaussian-type basis functions (GTOs) or non-local plane waves (PW). The electronic structure in CP2K[19] is calculated using the Quickstep module that uses a plane wave auxiliary basis set within a Gaussian orbital scheme (Gaussian and Plane Waves scheme or GPW). Although DFT methods are reasonably fast for systems of up to 100 atoms, a number of semi-empirical methods have been developed for treating large system where the full DFT method would be too computationally expensive. Semi-empirical methods are based on the Hartree-Fock method, but make approximation, usually by introducing empirical parameter sets (a parameter based on experimental results) to speed up the calculation. Semiempirical methods implemented in the CP2K program include AM1[20,21], PM3, and PM6[22] semiempirical methods. An alternative approximation to DFT is the

DFTB (Density Functional based Tight-Binding) method[23]. A more recent tight-binding method is the GFN-xTB extended tight binding semiempirical method[24].

An even more drastic simplification is to approximate the Born Oppenheimer potential energy surface $V^{BO}(\{R\})$ with an analytical function, usually using only two-body terms:

$$V^{BO}(\{R\}) = \sum_{I<J} V(R_I, R_J) \; = \; V^{FF}(\{R_I\}) \qquad (9)$$

which is sometimes called the "force field." In this case the equation of motion (Equation 3) reduces to Newton's equation of motion for the atomic nuclei:

$$M_I \frac{d^2 R_I}{dt^2} = -\nabla_{R_I} V^{FF}(\{R_I\}) \qquad (10)$$

Equation 10 is the purely classical equation of motion used in "Standard" molecular dynamics (MD). In force-field based MD, the electronic degrees of freedom are no longer included explicitly and are replaced by a set of empirical interaction potentials. These molecular mechanics (MM) potentials give the potential energy of the molecular system in terms of the positions of the atomic nuclei and are generally divided into a covalent-bond-type component (including bond, angle, and dihedral terms) and a non-covalent component (including electrostatic and van der Waals terms).

MM utilizes classical physics with point charges[4] to simulate the vibrations and attractions of atoms in the simulation. A typical molecular mechanics potential energy function for biological systems takes the form:

$$V^{FF}(\{R_I\}) = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,eq}\right)^2 + \sum_{angles} \frac{b_i}{2}\left(\theta_i - \theta_{i,eq}\right)^2 + \sum_{torsions} A_i[1 + cos(n\omega_i - \phi_i)] + \sum_{atom\,pairs} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{atom\,pairs} \frac{q_i q_j}{\varepsilon r_{ij}} \qquad (11)$$

The first three terms describe the covalent structure where the sum runs over all bonds, angles, and torsions. The last two terms account for the non-bonded van der Waals and electrostatic interactions, respectively, and the sum runs over all pairs of atoms separated by a distance $r_{ij} = |r_i - r_j|$, *excluding chemically bonded pairs*. Each step of the simulation calculates the force acting on each individual atom. Each atom's individual force is then applied to each atom over the time step in which your simulation is running. The position and velocities of the particles are updated in discrete time with a finite step that must be small enough to ensure stability of the numerical integration of the equation of motion (refer to eq 10). A typical

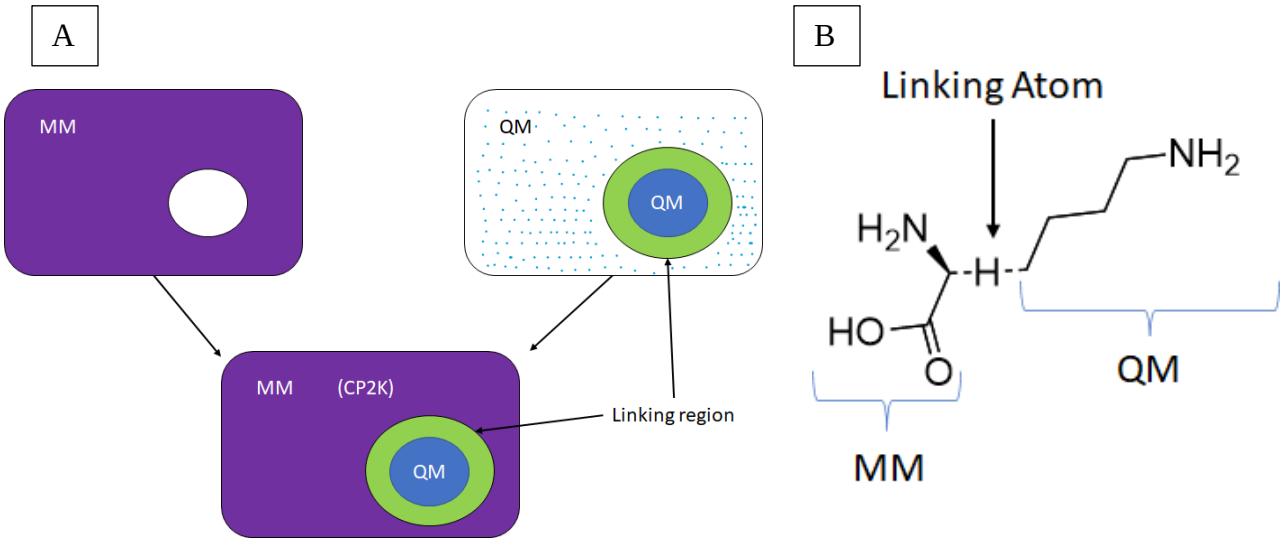
timestep used in classical MD is 2 fs. The accumulated data of positions and velocities as a function of time for all the atoms is called the trajectory, and provides information on both the structure and dynamics of the protein. The downside to classical MD is that through using classical physics (classical potential energy functions) we are unable to break and form bonds due to the neglect of the motion of electrons.

Any system over the size of roughly 1000 atoms is not feasible to simulate on a relevant time scale at the full quantum mechanical (QM) level with the current state of the art hardware and software technologies. However, typical biological systems contain much more than 1000 atoms. A typical MD simulation contains 100,000 atoms with simulations times approaching 1 microsecond. This leads us to a hybrid system that allows for faster simulation time but still allows specific bonds to be formed or broken, QM/MM.

Utilizing QM/MM allows computational chemists to speed up calculation time by treating the active site as a QM region and the rest of the system under MM parameters (figure 2A). The hybrid QM/MM potential energy contains three types of interactions in an additive scheme:

$$E^{Total} = E^{QM} + E^{MM} + E^{QM/MM} \qquad (12)$$

The first term is the interactions between particles within the QM region and is computed directly within the QM system. The second term is the interactions between atoms in the MM region and is calculated directly within the MM system. The final term is the coupling between the QM particles and MM (classical) particles. This coupling term contains both a non-bonded term to accounts for non-covalent interactions between the QM atoms and the MM atoms and a bonded term that is included when a QM atom is bonded to an MM atom. The long-range electrostatic contribution to the interaction energy can be categorized by the type of coupling scheme used. In “mechanical embedding”, there is no influence of the MM charge distribution on the QM atoms. In “electrostatic embedding” schemes, the electric field from the classical ions in the MM acts on the Gaussian basis functions through an additional term in the QM Hamiltonian. A third scheme is “polarized embedding” in which both the MM region and the QM region may polarize each other[25]. Throughout this work, we will use an electrostatic embedding scheme to couple the QM and MM regions.



$$\hat{\mathrm{H}}_{Tot} = \hat{\mathrm{H}}_{QM} + \hat{\mathrm{H}}_{MM} + \hat{\mathrm{H}}_{QM/MM}$$

Figure 2A/B) A cartoon image of the summation of the Hamiltonian wave functions in the simulation to give the total Hamiltonian of the system. In the white box each atom that is outside of the QM region is treated as a point charge affecting the QM specified atoms. In green the Linker region is shown as it is outside the MM and QM region. Shown in Figure 2B is the breakdown of a Linker Atom. This atom rests between the backbone carbon and the side chain carbon. This Linker Atom allows the atoms of the QM and MM region to interact with one another.

Sometimes partitioning the system into QM regions and MM regions leads to a situation where at least one covalent exists between a QM atom and an MM atom. A straightforward cut through the QM/MM boundary would lead to spurious unpaired electrons in the QM subsystem because in reality these electrons are paired in bonding orbitals with electrons belonging to atoms in the MM subsystem. In this case, to avoid these so-called dangling valences, the boundary of the QM and MM region requires a monovalent “linking atom” to bridge the gap (figure 2B). This linking atom can act both as a QM atom and a MM atom and is usually added as a hydrogen atom, as it has the smallest orbital. The linking atom also possesses both an electron able to covalently bond with the QM region and a point charge able to interact with the MM region. We use the “Integrated Molecular Orbital and Molecular Mechanics” (IMOMM) method as implemented in CP2K. In this method, the extra degrees of freedom due to the linking atom are removed by constraining the MM linking atom along the bond vector of the linking atom bond via the relation:

$$R_{MM} = \alpha R_{linking} + (1-\alpha) R_{QM} \qquad (13)$$

where R is the Cartesian coordinate of the subscripted atom, and α is a scaling factor. When a hydrogen linking atom is used to cap a single C-C bond, α is set to 1.38. Without the Linker atom the protein would either have to be simulated with full MM force fields or QM methods, sacrificing catalytic activity or extreme computational intensity respectively due to the protein being a continuous chain of atoms.

1.3 Free energy methods:

In the canonical ensemble (system of *N* particles in a fixed volume, *V*, and temperature, *T*), expectation values of any observable $\langle A \rangle$ are sampled by the Boltzmann configurational probability distribution:

$$\langle A \rangle = \frac{\int d\boldsymbol{R}\, A(R)\, e^{-U(R)/k_B T}}{Z} \qquad (14)$$

where the integral is overall all possible values of the atomic coordinates R, $U(R)$ is the potential energy, k_B is Boltzmann's constant, and *T* is the temperature. The denominator

$$Z = Z_0 \int d\boldsymbol{R}\; e^{-U(R)/k_B T} \qquad (15)$$

is the canonical partition function with Z_0 the partition function of an ideal gas (non-interacting particles). To compute averages according to Equation 14, we need to generate a sufficiently long trajectory that is consistent with the Boltzmann probability distribution under the ergodic hypothesis. The free energy is defined as the logarithm of the partition function:

$$F = -k_B T\; ln\, Z \; = F_0 \; - \; k_B T \int d\boldsymbol{R}\; e^{-U(R)/k_B T} \qquad (16)$$

where $F_0 = -k_B T\; \ln Z_0$ is an additive constant that will be subsequently ignored.

It is common in physics to reduce the complexity of a high-dimensional problem to a restricted number of variables called collective variables, CVs. Examples include so-called order parameters in Landau's theory of phase transitions or the solvent coordinate in Marcus theory of electron transfer. In enzyme catalysis we can think of a general reaction coordinate, $X(R)$, that defines the progress of the reaction from reactant to product along a low-dimensional coordinate. From transition state theory the rate constant for an elementary reaction is

$$k = \kappa \nu e^{-\Delta G^{\ddagger}/RT} \qquad (17)$$

where $\Delta G^{\ddagger}$ is the difference in free energy between the reactant and transition state, and $\kappa\nu$ is the pre-exponential factor. The transition state is defined as the high-energy, unstable complex corresponding to the maximum in free energy along the reaction coordinate. Once the free energy barrier to form the transition state is overcome, the reaction can progress and the product forms.

Having a set of the order parameters or CVs allow us to represent the progress of a reaction along the reaction coordinate in a concise and conceptually simple way. Mathematically, the CVs are defined as functions (usually nonlinear) of the atomic coordinates R: *S(R)*. The CVs should be able to distinguish between all relevant metastable states, and ideally would represent all the slow degrees of freedom. The choice of CVs in our work will be discussed shortly. Having defined the CVs, the free energy surface (FES) is defined as the logarithm of the equilibrium distribution of the CVs. Up to an arbitrary constant, the FES is then given as:

$$F(s) = -k_B T \ ln \int dR \ \delta(s - S(R)) e^{-U(R)/k_B T} \qquad (18)$$

The relative free energy difference between metastable states A and B along the reaction coordinate is determined from integrating the probabilities of the metastable states:

$$\Delta F_{A,B} = -k_B T \ ln \frac{\int_A ds \, e^{-F(s)/k_B T}}{\int_B ds \, e^{-F(s)/k_B T}} \qquad (19)$$

where the integrations are performed overall all configurations in CV space that define the two metastable states A and B, respectively.

From the above discussion, it would seem straightforward to evaluate the FES, and free energy changes simply from histograms of the sampled configurations. However, in practice this is often not feasible because the sampling is nonergodic on the timescales accessible to atomistic simulation. Often, the system will remain trapped in a local metastable state and will only fluctuate within this local free energy minimum. From this perspective, reactions that proceed from reactant to product by transitioning over a high-energy barrier through the transition state complex are seen as rare events that occur on time scales longer than that sampled from the simulation. We wish to enhance the probability of observing these transitions at a reasonable computational cost. One way to achieve this is to add an external bias potential

$V(s)$ that depends on the CVs, *s*, and thus also on the configuration *R* (through $S(R)$). The external bias enhances fluctuations along the CVs, leading to improved sampling along the reaction coordinate.

Metadynamics is an enhanced sampling method in which the bias is iteratively built through successive deposition of small repulsive biasing kernels in the form a localized Gaussian centered at the current CV value[24,25], *s'*:

$$G(s, s') = W\, e^{-\|s(t)-s'\|^2} \qquad (20)$$

where *W* is the height of the Gaussian and $\|s - s'\|^2$ is the distance between the value of *s* at time *t* and *s'*. When more than one CV is used, the distance metric is that of a multivariate Gaussian:

$$\|s(t) - s'\|^2 = \frac{1}{2}\sum_{i,j}(s_i - s'_i)\sum_{i,j}^{-1}(s_j - s'_j) \qquad (21)$$

where $\sum_{i,j}$ is the covariance matrix, which for simplicity is taken as diagonal with matrix elements, $\sum_{i,j} = \delta_{i,j}\sigma_i$, where σ_i is the width of the Gaussian for the *i*-th CV. The value of the Gaussian widths can be chosen from the unbiased fluctuations of the CV or estimated on the fly using an adaptive Gaussian approach. We use an open-source software known as PLUMED2[36-38] to perform metadynamics. As the simulation progresses, new Gaussian hills are continuously added, and these Gaussian peaks fill the FES with bias as the system runs. The addition of bias encourages the simulation to sample new regions of configuration space along the FES (Figure 3).

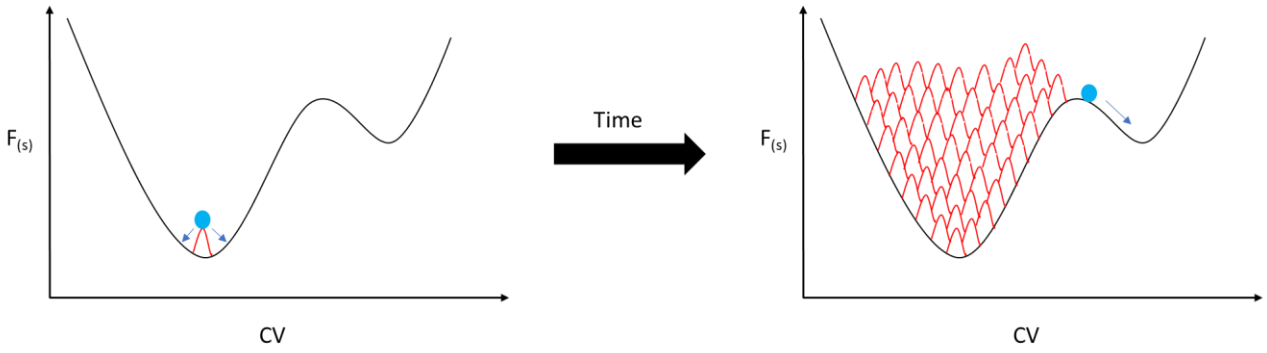


Figure 3) Shown is a FES ($F_{(s)}$) mapped over a CV, shown in black. Shown in red is the deposited bias that fills the FES until the simulation, shown as a blue ball, enters a new basin signaling a conformation change in the protein.

The bias, when built up high enough on the FES, can cause chemical reactions to take place as new FES basins are sampled[29]. Under the action of the bias potential, the instantaneous probability distribution will be that of a time-dependent, *modified* Boltzmann distribution

$$P(R,t) = \frac{e^{-(U(R)+V(s,t))/k_BT}}{Z_b} \qquad (21)$$

which can be recast in the form:

$$P(R,t) = e^{-(V(s,t)-c(t))/k_BT} \cdot P_0(R) \qquad (22)$$

where $P_0(R)$ is the unbiased Boltzmann distribution and the time-dependent function c(t) is

$$c(t) = k_BT \ ln\frac{Z}{Z_b} = k_BT \ ln\frac{\int ds\, e^{-F(s)/k_BT}}{\int ds\, e^{-(F(s)+V(s,t))/k_BT}} \qquad (23)$$

With these definitions it is possible to reweight any arbitrary position-dependent quantity to obtain an unbiased estimate of the ensemble average from averages over a biased trajectory according to[28]

$$\langle A \rangle_0 = \langle A \cdot e^{(V(s,t)-c(t))/k_BT} \rangle \qquad (24)$$

where the right-hand average is performed over a metadynamics trajectory. This allows us to reweight any equilibrium property from a biased simulation given enough sampling to obtain meaningful statistics on the ensemble averages. If we are interested in dynamic quantities such as rate constants, the acceleration factor due to the bias can be estimated within transition state theory assuming that no bias is deposited in the transition state region of the free energy. Under this assumption the ratio of the biased to unbiased reaction rate is given as[28]:

$$\frac{k_b}{k_0} = \frac{Z}{Z_b} = \langle e^{V(s,t)/k_B T} \rangle \qquad (25)$$

where the angular brackets denote an average over a metadynamics run confined to one of the metastable basins. This condition is met by increasing the lag time between Gaussian hill depositions to avoid completely filling the free energy basin during the simulation window.

PLUMED is also able to add constraints to our system to prevent the system from exploring irrelevant or nonphysical configurations[36-38]. These constraints add a harmonic restraint that puts an exponential force on two atoms either to keep them within or outside of a specified distance[36-38]. We also use a number of CVs to visualize our data from the simulation and track reaction progress. The distance CV is simply the distance from one atom to another[36-38]. This can be used to monitor the formation or breaking of covalent bonds. The coordination CV allows us to monitor the number of contacts formed between two groups of atoms (A and B) as the sum given by equation 26. The calculation consists of setting a cutoff distance for two groups of atoms. The cutoff distance adds a true (One) or false (Zero) value whenever an atom *(i)* comes within or leaves the cutoff distance of another atom *(j)* of a different group (equation 27).

$$\sum_{i \in A} \sum_{i \in B} S_{ij} \qquad (26)$$

The cutoff is implemented as a differentiable switching function of the form

$$S_{ij} = \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m} \qquad (27)$$

This CV will allow us to monitor the coordination number between two groups of atoms as they sample different conformations. As the simulation progresses, we look to see changes in how the interaction of groups of atoms progresses.

1.4 Project Goals:

In this thesis I am to produce a QM/MM model of the saSrtA and baSrtB systems with kinetic catalysis data and FES. This will be to understand the catalytic mechanism of the saSrtA and baSrtB enzymes at a quantum level, something not easily measured and viewed by

biochemists. Utilizing the FES will be integral to understanding the catalytic pathway in the baSrtB enzyme and provide understanding of the metastable states the are present in the catalytic pathway. Another goal is to understand the interactions between the ligand and key residue interactions with other amino acid residues of the protein. The second aim of my thesis will be to create a moving restraint to build a CV from that will drive our reaction. The moving restraint will be used to make the Harmonic Linear Discriminant Analysis (HLDA) CV which will be able to linearize our reaction and drive our reaction forward and backward. The third aim of my project is to build a QM/MM model of the saSrtA enzyme. The saSrtA model will be used to compare the catalytic pathway to baSrtB and help understand subtle differences in the catalytic mechanism. The fourth aim of my thesis is to run multiple docking simulations using AutoDock, FlexPepDock, a manual docking procedure, and AlphaFold. Running multiple docking simulations will allow us to tease out not only the best docked structure, but also allow us in the future to select the best docking software right away.

2 Discussion and Results :

2.1 Docking Experiments:

For modeling sortase catalysis, a starting point is to know the structure of the enzyme with its substrate in a bound confirmation. Since proteins and substrates are dynamic, there may be cooperative motion and many degrees of freedom involved. For the catalytic reaction to take place the enzyme and substrate must be arranged in near-attack conformation (NAC) for chemistry to occur. Because there are no high resolution structures available for Sortase B with a bound substrate, we turned to molecular modeling to construct a working model of sortase B with a peptide substrate.

For predicting binding poses of small molecule drugs in their protein targets, molecular docking software is a popular and computationally efficient method of evaluating potential binding sites and ranking ligand bound structures according to predicted binding affinity. Molecular docking methods use a search algorithm that generates a large number of possible bound structures. Typically, the receptor binding site is kept rigid, but different conformations

of the ligand are sampled. Typically, conformations are sampled using Monte Carlo moves, and binding poses are scored using an empirical scoring function. Traditional docking methods such as AutoDock use a scoring function that includes non-bonded van der Waals and Coulombic terms in addition to a phenomenological term to account for hydrogen bonds and/or hydrophobic interactions[30,31]. More recently, protein structure prediction methods such as AlphaFold use machine-learning, deep neural networks, and structural databases to predict folded protein structures with high accuracy[32]. When the ligand is a flexible peptide, the conformational search space is increased because of the large degrees of freedom of the peptide. Additionally, small peptides have many non-specific interactions with protein surfaces. These two features: many degrees of freedom and many non-specific interactions, makes the prediction of flexible docking of peptides particularly challenging. Recently, a new docking protocol for flexible peptides, FlexPepDock was introduced by the Furman Lab and included in the Rosetta suite of programs for protein structure prediction[31].

In this work, we evaluated three software packages on their ability to find a near-active conformation of ligand binding motif to the baSrtB binding cleft. The three software packages we used were,: AutoDock[30], which is a traditional docking program used for small molecule docking, FlexPepDock, developed for flexible peptide docking within Rosetta[31] , and AlphaFold[32], which uses a deep neural network to predict protein tertiary structure. AutoDock performed the worst out of the three and had the tendency to create an ouroboros type structure out of the ligand, probably because of the increased flexibility and degrees of freedom of the peptide not being properly accounted for in the model. FlexPepDock had a decent binding configuration, however it uses a two-step process that first involves breaking the protein and peptide into smaller fragments or chunks to explore different conformations and orientations. After favorable fragments are found, FlexPepDock proceeds to a rigid-body docking step where the best fragment combinations are assembled to form the complex. Unfortunately, an additional alignment in PyMol was needed due to the FlexPepDock procedure ruining the target binding protein PDB output file. Another strategy that was tried was a more manual approach. In the manual approach an iterative docking procedure was performed where a small fragment of the ligand was added to the active site. Then an energy

minimization was performed after manually placing in the amino acid string one by one until you have the completed the ligand motif. Lastly, AlphaFold was used by Dr. John Antos to create the best docked structure available to our group. In this procedure, Dr. Antos used the full baSrtB sequence with the peptide linked to the protein via a flexible linker (see Appendix).

2.1.1 AutoDock results:

Starting with AutoDock; an attempt to do the simplest docking procedure of putting the NPKTG motif into the binding site of baSrtB and running the program. After several attempts, AutoDock yielded one "good" structure. However, even AutoDock's best structure was curled into a "ouroboros" like stricture (see figure 4) where the C-terminus and N-terminus amino acid side chains (Glu and Asp) were interacting with each other and not lying flat in the binding cleft. Steps to curb this phenomenon were taken by increasing the box size to be more elongated and rectangular along the surface of the protein. Yet, these steps to fix the problem did not help due to the inability to rotate the box on an axis as it is fixed to Cartesian coordinates and would never lay parallel to the angle of the binding cleft. Quickly moving on from AutoDock to the next docking suite that had a better user interface (UI) and was browser based was FlexPepDock.

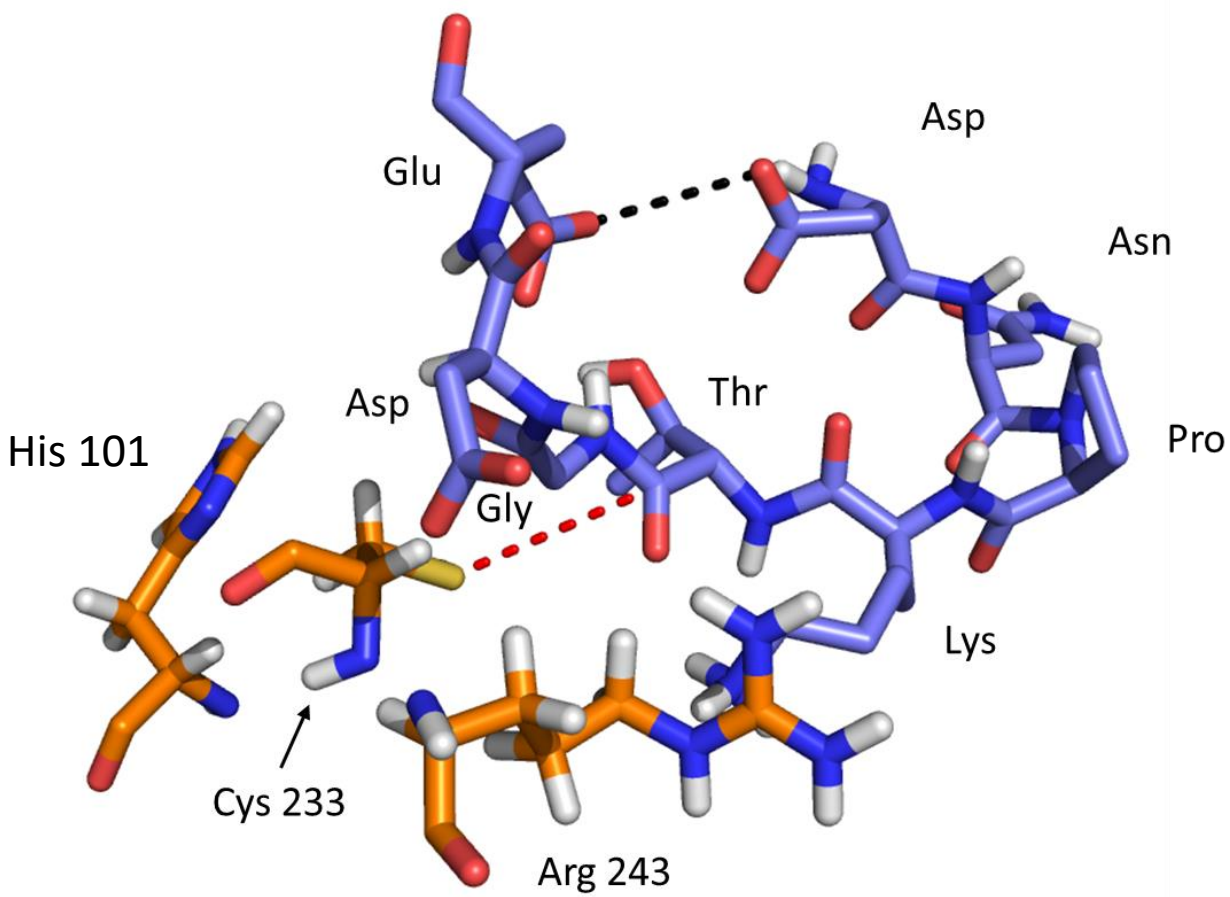


Figure 4) Shown in orange is the baSrtB catalytic triad (His 101, Cys 233, and Arg 243) and shown in slate is the binding ligand DNPKTGDE (Asp -> Glu). Shown by the black dotted line is the close contact the Autodock was making with two negatively charged sidechains at 3.8Å. The red dotted line (4.3Å) is showing the second problem with Autodock in that in never placed the ligand Thr in a close enough proximity to the active site.

2.1.2 FlexPep Dock results:

FlexPepDock (FlexPep) Is a Web browser-based docking suite that specializes in flexible peptide docking. FlexPep also allows the user to see the docking box size and location before you submit the docking request, making the user experience very intuitive. When submitting the docking request, FlexPep asks several questions based on the chemistry of the ligand residues whether the residues are charged or uncharged, lacking or have additional hydrogens, and any bonds that should be non-rotatable. Another helpful feature was the ability to see exactly what the dimensions of the docking box looked like before submitting the job and FlexPep allows the user to see the changes made to the limits of the box made in real time when the user changes the X, Y, and Z dimensions of the box. Another plus for FlexPep is that

being web based is that you do not need a dedicated machine to dock a peptide on a computer that normally could not handle running a docking simulation such as a laptop. However, FlexPep was not the best docking suite we used. Though FlexPep had a better UI it suffered from some of the same problems as AutoDock, one of which was the self-interacting N and C terminus of the Ouroboros (see figure 4) on some of the lower scoring submissions. The higher scoring submissions had the problem of binding in the reverse orientation (see figure 5). It is unknown if the reverse orientation of the DNPKTGDE is active within the enzyme, but without published work stating the binding site of baSrtB, therefore models were selected that had similarities in both position and orientation with the inhouse *strep pyogenes* Sortase A structure[33]. Another failure of FlexPepDock was the baSrtB enzyme body was chopped up

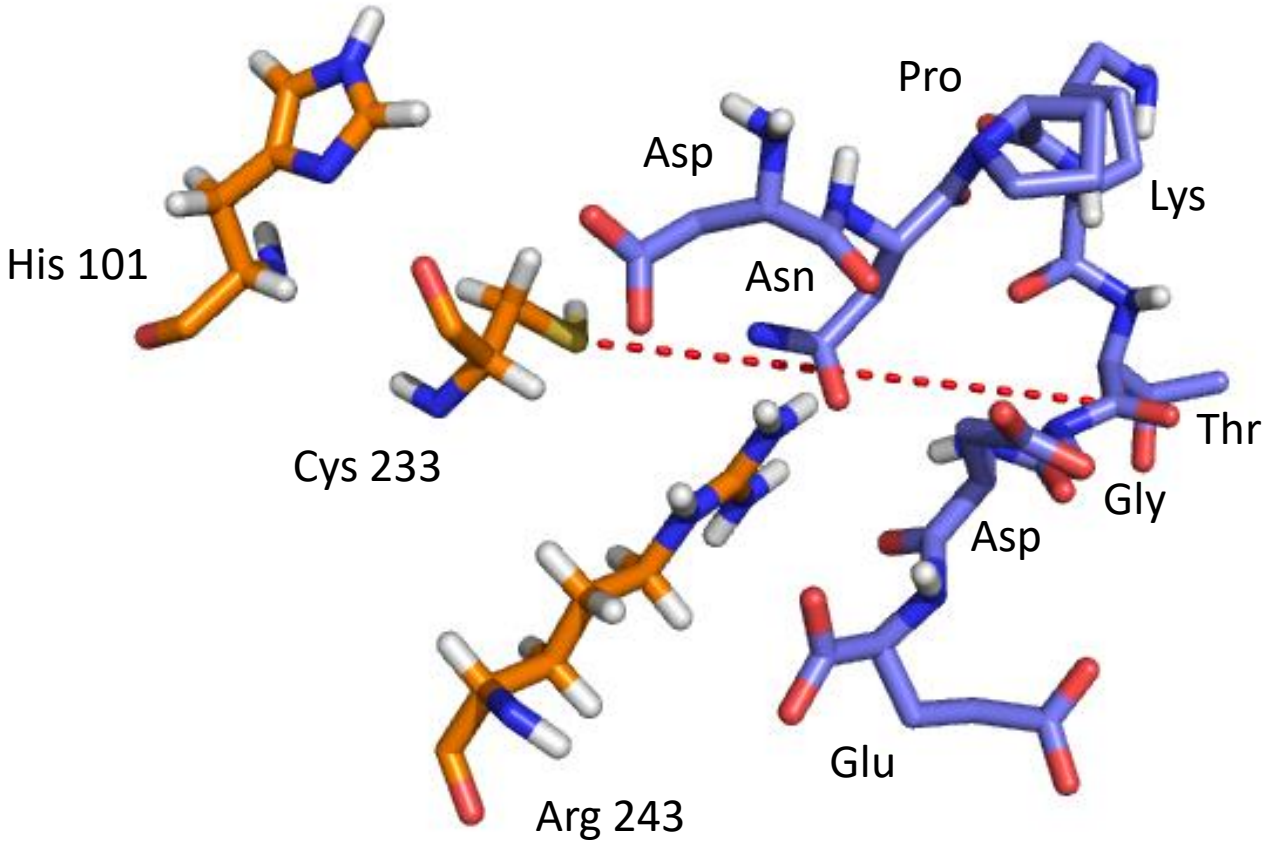


Figure 5) Shown in orange is the baSrtB catalytic triad (His 101, Cys 233, and Arg 243) and shown in slate is the binding ligand DNPKTGDE (Asp -> Glu). This structure was generated by FlexPep Dock and as you can see this is not a good binding due to the fact that not only is the structure in the reverse orientation and completely outside the binding pocket, the dotted red line connecting the catalytic Sulfur Cys 233 to the carbonyl carbon on the Thr is 11.3Å away from each other.

when the output file was returned. This failure is more of a minor concern and can easily be fixed with an alignment with the apo protein body and deletion of the output files enzyme body and saving the new combined structure in PyMol. Still a major problem existed with both FlexPep dock and AutoDock and that is neither docking suite gave a good Near Active Conformation (NAC) of the ligand bound in the enzyme. Therefore, it was decided to dock the ligand utilizing an iterative energy minimization process in CP2K and utilize PLUMED to keep the carbonyl carbon on the ligand Thr in place so we can keep the NAC of the ligand while adding amino acids to the ligand.

2.1.3 Iterative docking/refinement procedure:

Building our own docked ligand was done via an iterative energy minimization on several structures being built and modeled by hand in PyMol. The manual docking method requires that the user makes multiple files after each energy minimization step due to the fact that; adding atoms to the system after the energy minimization step ruins the topology file. This

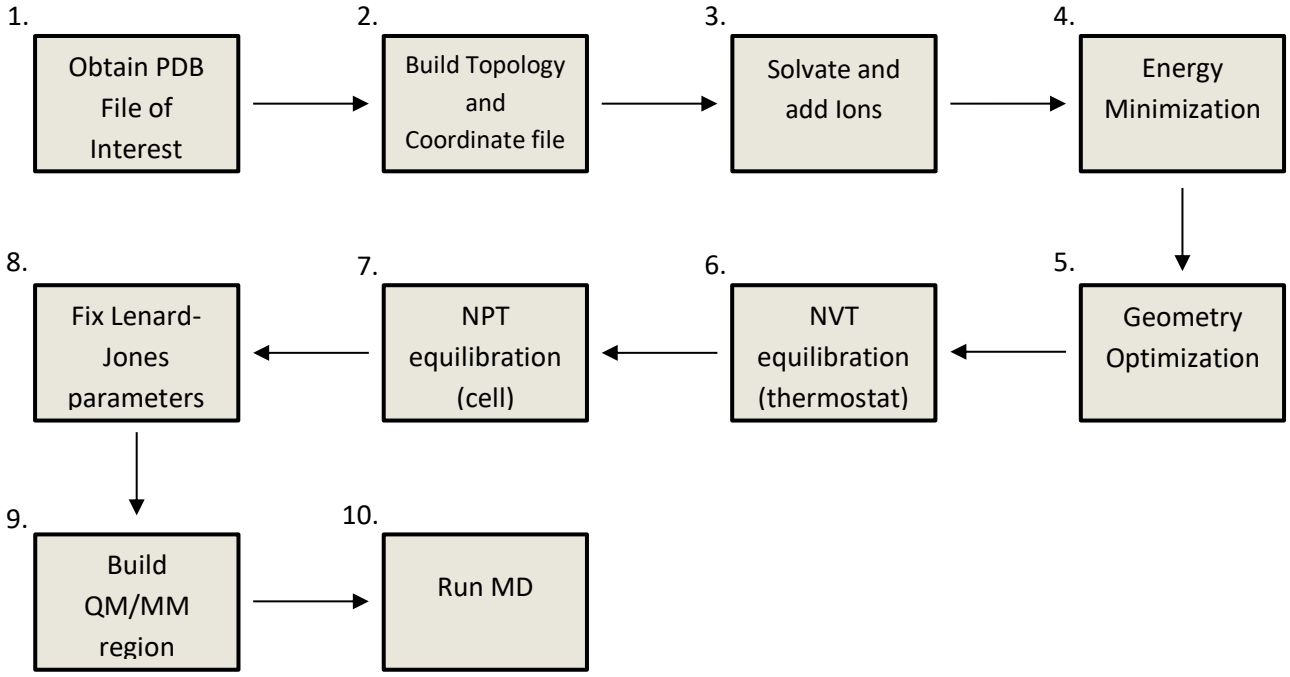


Figure 6): A flow chart depicting the major steps to building a fully equilibrated QM/MM system. Only building a system through steps 1-7 will still build you a fully functional MM system.

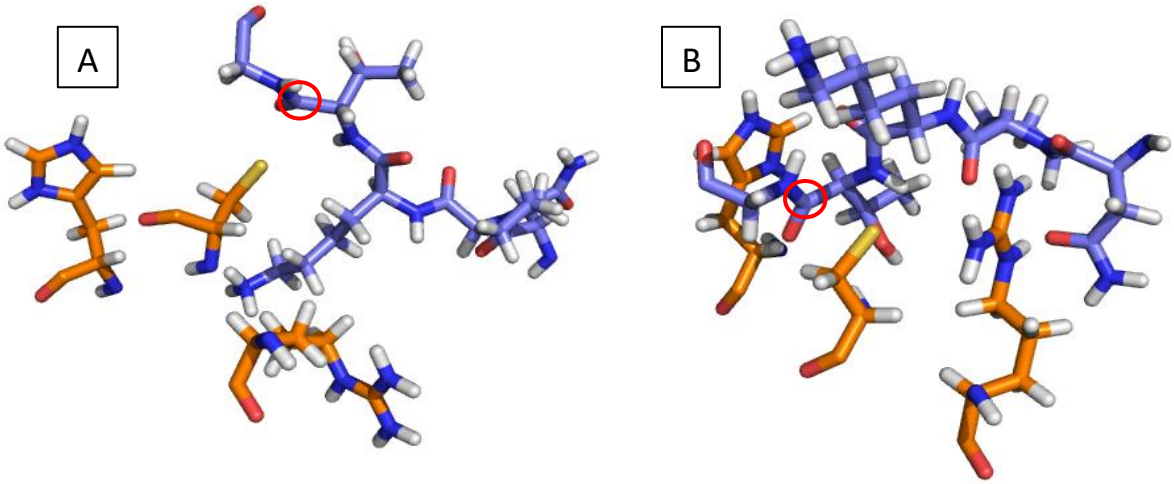


Figure 7) A) Shown is the baSrtB catalytic triad (orange) with the manually docked NPKTG binding motif. The reason for the higher confidence of this docking procedure is that the binding motif is laying flat across the binding region, the Lys was inside of a void in the protein binding domain, and the Thr carbonyl carbon (red circle) was 2.5Å away from the catalytic sulfur on Cys 208. B) However, as you can see the AlphaFold structure, by our collaborator Dr. John Antos, has the Thr carbonyl carbon (red circle) in close proximity to the Cys 208 sulfur, in addition to straddling the gap between His 101 and Cys 233. Arg 243 in the alpha fold model is also making hydrogen bond contacts with the Asn and Pro of the ligand binding motif.

requires the system to build a new topology file after step 5 (see figure 6) and figure 6 shows a flow chart that follows the method iteratively for steps 1-5 for each amino acid added to the ligand motif. However, even though this process is intensive and hands on, this method proved to be better than AutoDock or FlexPep. With PLUMED patched into CP2K the Thr was able to keep its carbonyl carbon within 2.5 Angstroms from the catalytic Cys while freely allowing the energy minimization and geometry optimization to minimize the energy of the other atoms. After finishing the ligand this method had a lot of important interactions with the acidic flankers, DNPKTGDE, on the ligand. Ultimately this method still was not perfect because the Lys was sticking into the binding pocket (see figure 7A) rather than out of the pocket (see figure 7B).

2.1.4 AlphaFold predicted structure:

The manual docking method was ultimately trumped by the Alpha Fold predicted structure (created by Dr. Antos) that had extreme similarities to the saSrtA binding where the Lys is facing out of the binding pocket. This similarity with the Sortase A structure is not

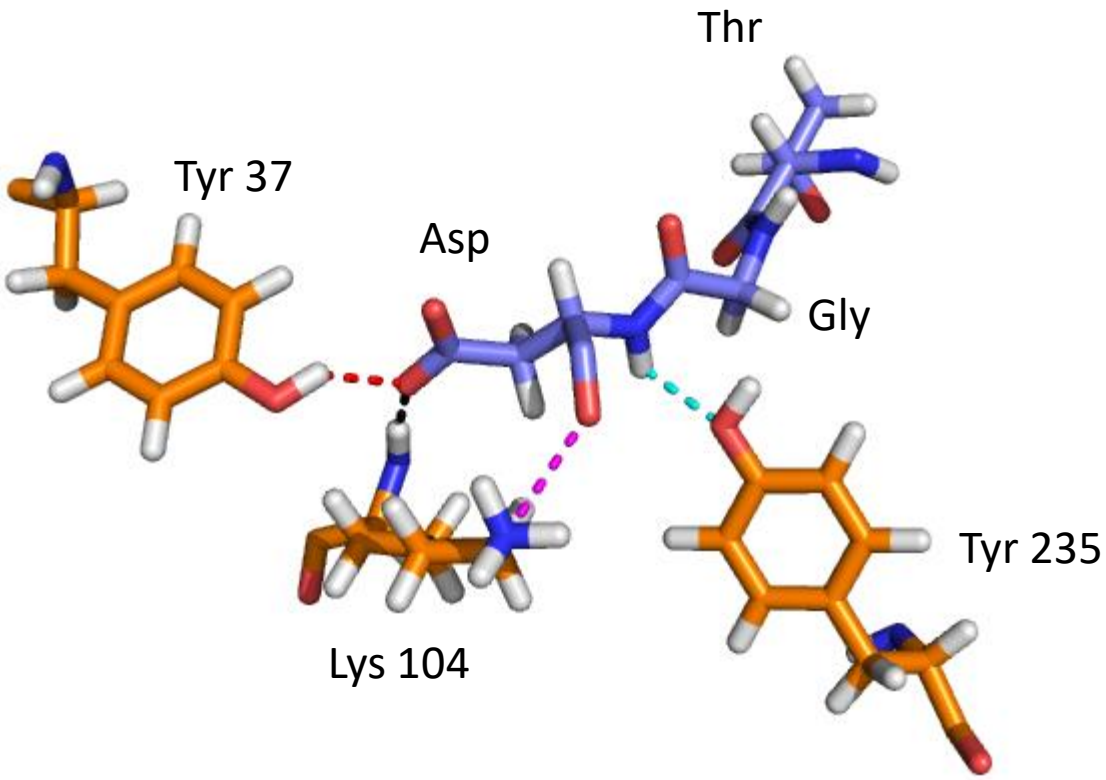


Figure 8) Shown in orange are three protein residues of Tyr 37, Lys 104, and Tyr 235 making hydrogen bond contacts with the N-terminal Asp on the binding ligand. The hydrogen bonds are shown by dotted lines with lengths of 2.2Å (Red), 1.6Å (Black), 3.1Å (Magenta), and 2.3Å (Cyan)

accidental since AlphaFold was trained on known folded structures in the PDB database. In the AlphaFold structure, the aspartic acid is coordinated to nothing on the protein and the side chain is solvent exposed. The asparagine on the N-terminus and the Proline are coordinated to arginine 243 of the catalytic triad (see figure 9A). The Lys is coordinated to the Glu 241 on the β7-β8 loop (see figure 9B). The Thr side chain is inside of a small pocket and is interacting with the backbone nitrogen on the catalytic Cys 233 (see figure 9C). The Gly residue on the ligand is making two contacts, one with the backbone oxygen to the backbone nitrogen on Arg 102 and the Gly backbone nitrogen to the ligand Thr backbone oxygen (see figure 9D). The flanking Glu interacts with three residues Tyr 37, Lys 104, and Tyr 235. Tyr 37 interacts with Glu by making a hydrogen bond to the carboxyl group on the side chain. Lys 104 makes two connections, one from the side chain nitrogen to the backbone oxygen and another from the backbone nitrogen to the carboxyl group on the side chain of the Glu (see figure 8). The final residue, Tyr 235

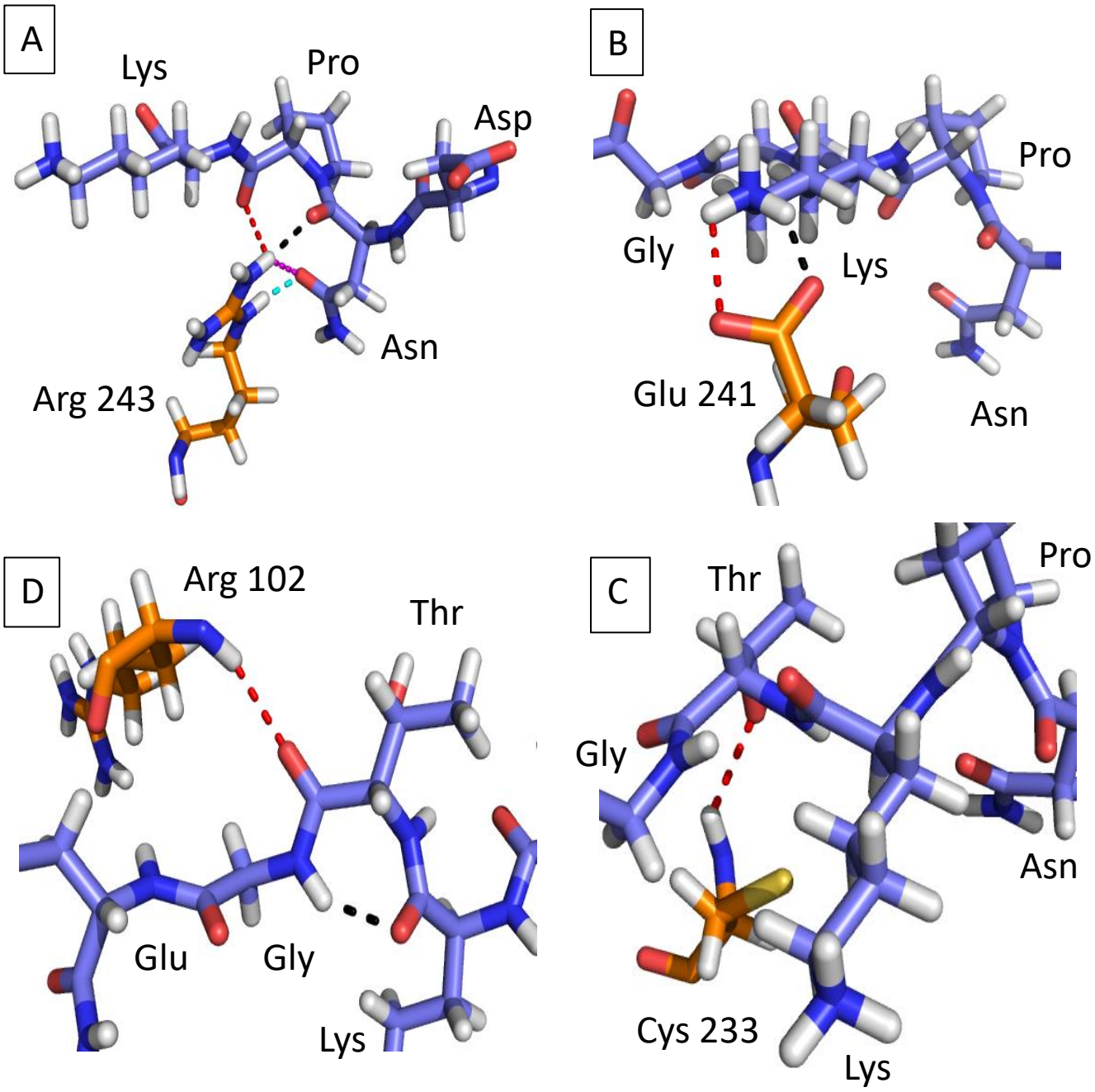


Figure 9) A) Show in orange is Arg 243 making close contacts with the binding motif ligand residues of Asn and Pro. Hydrogen bond contacts made by the residues are shown by the colored dotted lines at distances 2.7Å (Red), 2.0Å (Black), 3.2Å (Magenta), and 1.9Å (Cyan). B) Shown in orange is the protein Glu 241 making hydrogen bonds with the sidechain of the ligand binding motif Lys. The two hydrogen bonds are formed by Glu 241's sidechain carboxyl group and the nitrogen on the Lys at distances of 2.6Å (Red) and 1.8Å (Black). C) Shown in orange is Cys 233 forming a single hydrogen bond at 2.4Å (Red) between the backbone nitrogen of Cys 233 and the sidechain alcohol group of the ligand Thr. D) Shown in orange is Arg 102 making a hydrogen bond at a distance of 2.4Å (Red) with the backbone oxygen of the ligand Thr. The second hydrogen bond contact is between the ligand Gly nitrogen and the backbone oxygen of the ligand Lys at 2.1Å (Black).

makes a hydrogen bond with the backbone nitrogen on the Glu (see figure 8) The final Asp residue is once again solvent exposed.

2.2 Sortase A/B Introduction

Sortase A from *staphylococcus aureus* (saSrtA) is the most widely studied sortase enzyme in the literature, and prior QM/MM simulations have been performed by Tian and Erikson[34] who found support for the reverse protonation mechanism where Cys208 and His141 exist as a thiolate anion and imidazolium cation before substrate binding. Tian and Erikson also argue that the Arg218 is not involved in stabilizing the tetrahedral intermediate as previously hypothesized[34]. Subsequent QM/MM simulations performed by Shrestha and Wereszczynski investigated the catalytic mechanism of saSrtA using QM/MM simulations with multi-dimensional metadynamics finding that the free energy barriers (FES) for the reaction are lower in the "Thr-In" binding motif where the threonine residue of the sorting signal is oriented into the binding site, suggesting the Thr-In configuration is the catalytically active state[35].

Initially, an attempt to build a working QM/MM simulation on the saSrtA system was done with the goal of using more advanced collective variables to try and sample the entirety of the reaction space using metadynamics. Work began by using the published NMR structure (PDB: 2KID). The 2KID structure has a substrate analog that is covalently linked to the enzyme to mimic the thioacyl intermediate. The first step was to replace the analog with the enzyme recognition LPATG motif and perform an energy minimization of the modified structure.

When building a QM/MM system several steps must be done first to ensure that the system is correctly built and equilibrated. The first 3 steps of this process use MM force fields and do not involve bond breaking or formation. In the MM force fields, covalent bonds are described with a harmonic potential energy, equivalent to the description that all the bonds and atoms are held together by ball and spring (see figure 10).

To set up the system we first perform a geometry optimization or energy minimization (EM). The EM step in CP2K uses the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) optimizer that is suitable for large systems. The LBFGS algorithm uses a variant of Newton's method based on the gradient of an objective function to find the minimum. For this step we update the positions of the atoms to find the configuration that minimizes the energy between all the bonded atoms by moving the system down the path of steepest descent on the energy funnel (see figure 11). This process of energy minimization will go on iteratively until the simulation cannot minimize the system past the energy value set to minimize past or until the

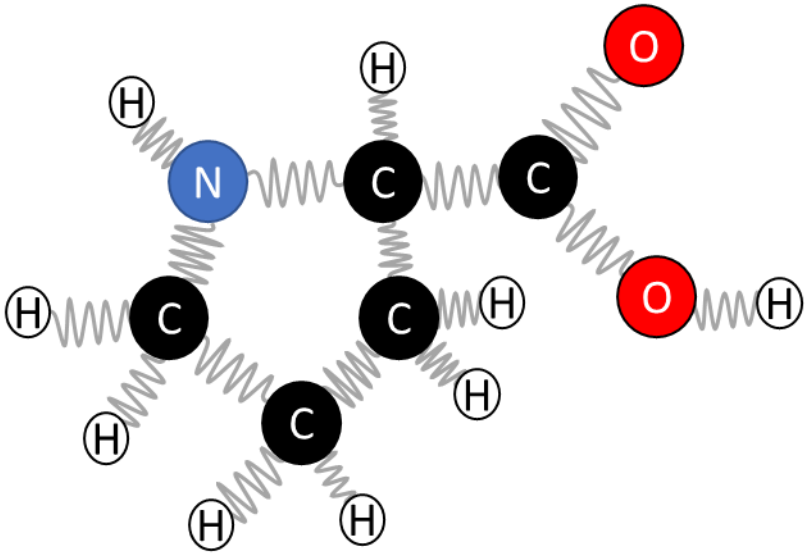


Figure 10): A representative "Ball and Spring" model of Proline showing the bonds as "Springs" and Atoms as "Balls". This assumption allows for a simplification during the MM calculation and use of a constant for determining the bond length based on what the atom is bonded to.

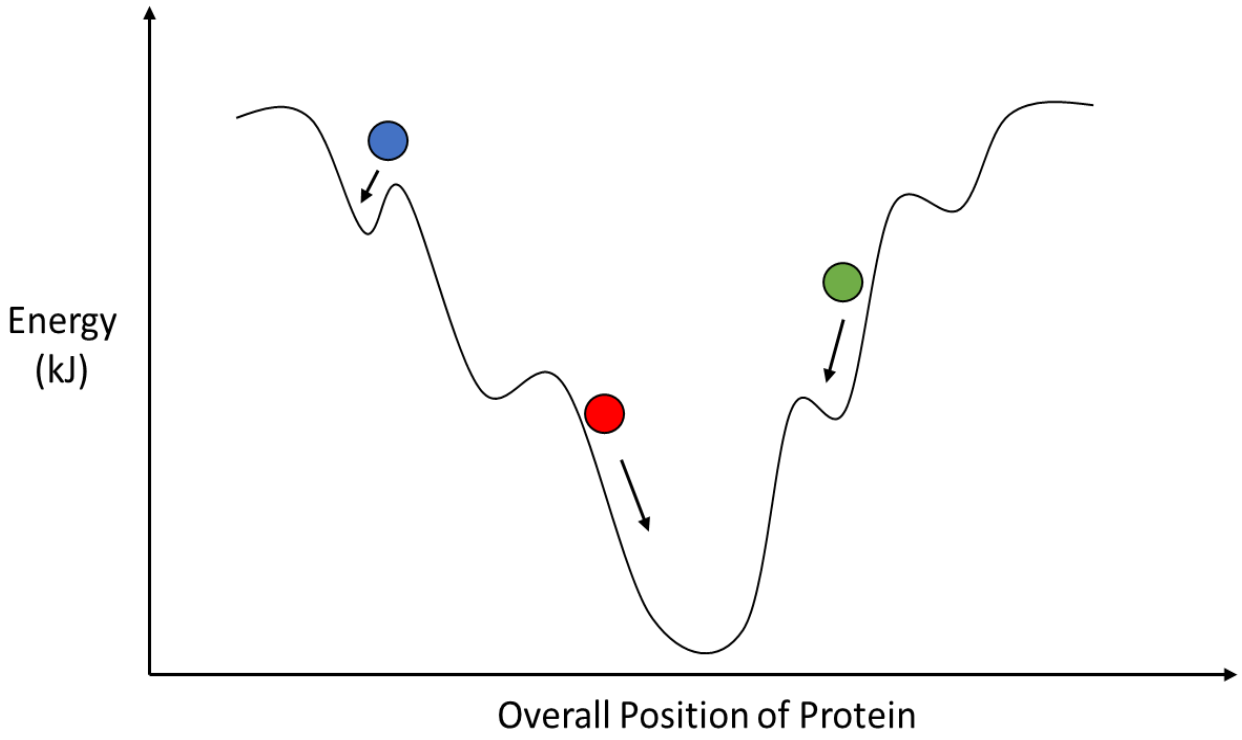


Figure 11) Shown is a simplistic 2D energy funnel. During the Energy Minimization process depending on your original configuration of the protein you could find the global minima (shown by red) or a local minimum with an energy barrier that is too high to cross over (shown by blue and green).

total step count has been reached. Once this step has finished you now have a system that is ready to be thermally equilibrated. The next step is to equilibrate the system using only MM force fields. The NVT step adds a thermal energy element to the simulation where the system is calibrated to a set temperature. Details of the thermostat used can be found in the Methods Section. NVE equilibration is important because the EM structure only considers structural energy minimization and not the thermal energy of the system. Now with a thermostat we can get Brownian motion to occur in our simulation through the addition or removal of kinetic energy from the system to maintain an equilibrium temperature in our system. Thermal fluctuations (Brownian motion) are important for overcoming kinetic barriers. The magnitude of random thermal motions is quantified by the thermal energy of the system, which is given by the value of k_BT (Boltzmann's constant times the temperature). Typically, kinetic barriers that are on the same order of magnitude as k_BT can be sampled during the simulation; however, metastable states that are separated by free energy barriers much larger than k_BT cannot be

sampled with conventional MD methods – hence the need for enhanced sampling methods such as metadynamics. Finally, the NPT equilibration implements a barostat and allows the system cell to fluctuate during the equilibration process to a set pressure of 1 atm. Details of the barostat can be found in the Methods section. This step is the most important because if the system cell structure was rigid it may be impossible to replicate benchtop experiments at 1atm. In other words, this allows us to perform simulations at a density that matches experimental conditions. The timeline for the EM, NVT, and NPT equilibration steps takes roughly 2 days (working on a single workstation described in the Methods Section) to fully complete (see figure 6). If you want to make any slight changes to the protein or ligand structure, such as a point mutation or change of protonation state, the system will need to start over completely from the energy minimization step. After NPT equilibration, the next step is to equilibrate the system at the QM/MM level. For the Sortase A QM/MM system, this is where all the problems began and did not stop. Building the QM/MM region is relatively simple in CP2K. All CP2K needs is a list of each atom that will be in the QM region and the two atoms (one in MM and one in QM) that you want to link (see figure 2B). The catalytic triad had the entire side chain in the QM region while the Alpha Carbon was connected to the Beta Carbon via a pseudo hydrogen atom. This set up should give you a system that is ready to run and collect data, this is what we thought too for the Sortase A system.

The first and most major problem that plagued the simulations was a “Cholesky decompose failed” error. Cholesky decomposition is a numerical procedure to factorize a positive definite matrix. This failure indicates that the matrix may be ill-conditioned or not positive definite. This failure occurs during the self-consistent field (SCF) step in the QM calculation. The likely cause of this error is that the structure is not stable causing the SCF to leave the loop without converging, but this is surprising since this step follows an energy minimization and equilibration of the system using the MM force fields. This error plagued every simulation that was ran on saSrtA work and some of the work on the Sortase B systems. Fortunately, we have solved this issue for the Sortase B system and have plans to use newly obtained knowledge of this solution on our Sortase A system. Understanding how the different

systems of MM and QM/MM interact with the close and tightly packed nature of the active sites in both Sortase A and B is key.

2.2.1 Sortase A:

The *staphylococcus aureus* Sortase A system proved to be a hurdle, as the system was not well defined with the given Protein Database files (PDB) available. The starting structure was taken from the PDB database entry 2KID, which was the truncated protein bound to a substrate mimetic. The initial thought was that it would be straightforward to build the peptide ligand into the catalytic binding site by converting the substrate mimetic to the peptide ligand LPXTG modified the structure in PyMol and performing the energy minimization, NVT, and NPT equilibration runs. When the saSrtA system was built with the QM/MM interface with the linker pseudo atoms, the Self Consistent Field (SCF) failed to converge. Every time the simulation was

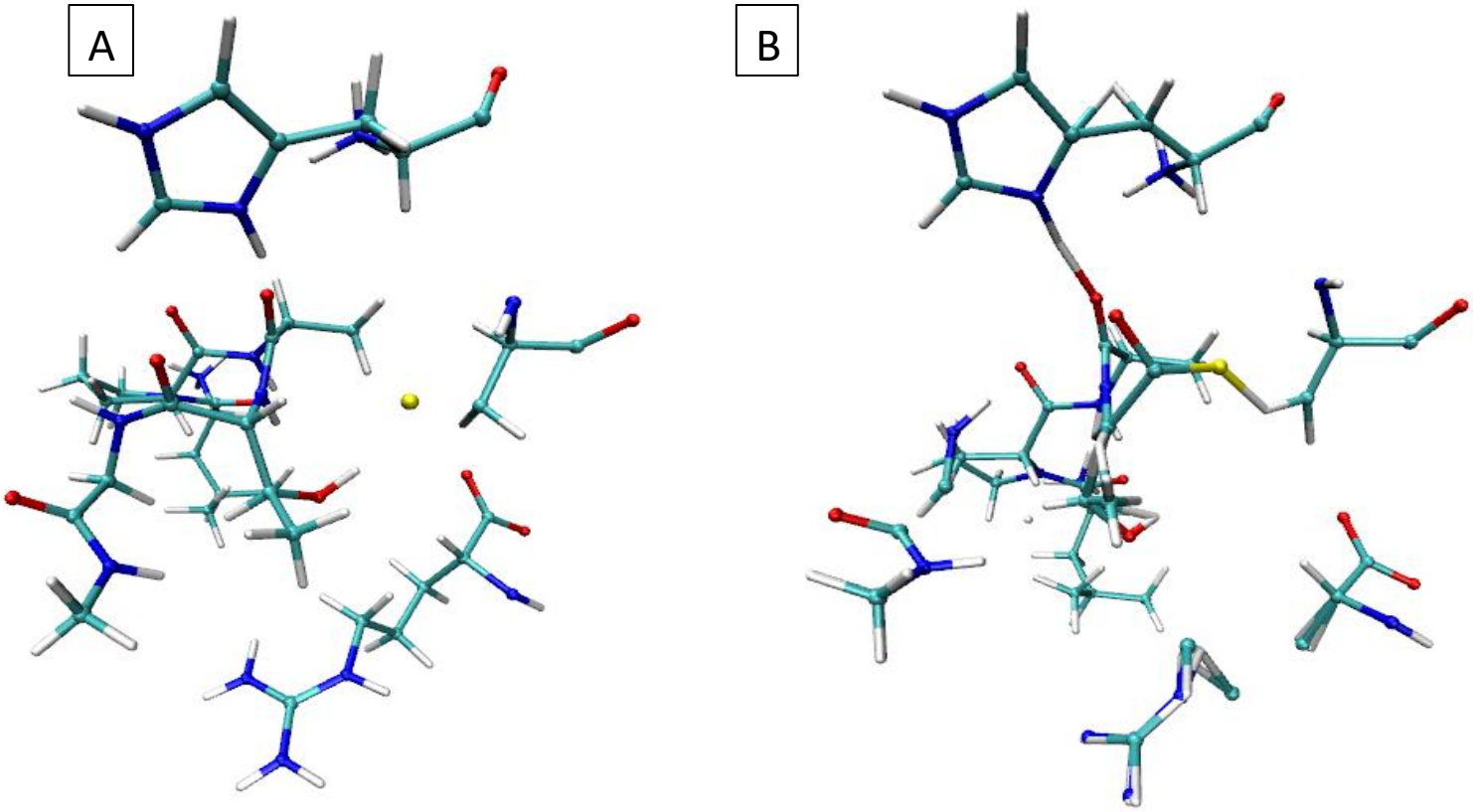


Figure 12) Shown on side A is the start of the QM/MM simulation on saSrtA that has been truncated down for simplicity. Shown on side B is the end of the 17.2 femtosecond simulation that is showing that the Arg 216's hydrogens are collapsing into the center atoms along the entire amino acid causing the simulation to crash.

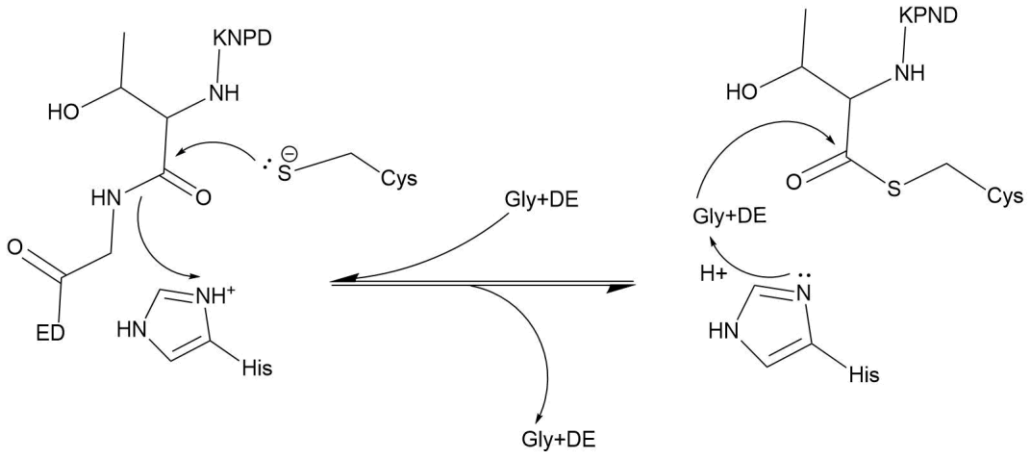
started the SCF would leave the loop after 50 steps. The first solution to fix this was to slow down the step time from 0.5fs to 0.1fs. The original hope was to allow the energy to converge to a stable state. This didn't work as the simulation still crashed this time just slower than before. However, the slower time step allowed a window to visualize in VMD what was happening to the simulation. In the simulation on saSrtA we noticed that the hydrogens on the side chains of the QM region of the catalytic triad were pulsating in and out at a rate that was too fast for a reasonable movement at a time step of 0.1fs (see figure 12). After a few steps of the hydrogen bonds pulsating, the system crashed, and the hydrogen atoms collapsed into the center atom (see figure 12B). This system was never finished due to this problem, but a fix has been found due to work on the baSrtB system and will be implemented in future runs.

2.2.2 Sortase B:

The start of the *bacillus anthracis* Sortase B system was not a straightforward endeavor. There were a lot of revisions done to the baSrtB system to get a structure suitable for QM/MM simulations, most of which were docking the NPKTG binding motif into the active site of the enzyme. After the docking simulations, however, a choice was made to use on of two viable structures to pursue; the manually docked binding motif and the Alpha fold structure provided by Dr. Antos. It was decided to pursue the Alpha fold structure as this ligand had binding similarities to the known *streptococcus pyogenes* Sortase A structure. The main binding similarities to the *pyogenes* structure include the position of the ligand Lysine tail and the Thr side chain pointing into the pocket. The workup began with the *bacillus anthracis* Sortase B PDB file that was created by Dr. Antos utilizing Alpha Fold to predict the binding motif of DNPKTGDE attached to a peptide chain off the N-terminus of the truncated protein. This structure was then edited until only the bench top truncated baSrtB protein was left and the ligand binding motif docked inside of the binding site. The N-terminal was capped with an Acyl group along with the N and C-terminus of the ligand binding motif with Acyl and N-Methyl respectively. The new Sortase B system was then put through the EM, NVT, NPT equilibration, and QM/MM system setup.

The “Cholesky decompose failed” error that was given was the same error that was happening with the Sortase A system, and the SCF left the loop and never converged. However, observations while viewing the output file showed that instead of converging to a single value, the SCF values alternated between two values. This can occur in numerical methods when the starting initial guess is poor. In such cases, convergence can be obtained by a better initial guess of the wavefunction. The starting guess were the atomic orbitals. It seemed like the system was having struggling finding the position that would best minimize the energy of the electrons in the QM region of the system, this was the same error that plagued the saSrtA system. This error can come about when the time independent Schrödinger equation takes the best guess and outputs an answer that gets put back in to equation only to give out the original input. To fix this we went back to the NPT equilibration step and increased the upper bound constraint on the carbonyl and sulfur distance from 2.5Å to 6.0Å and restarted the run for an additional 10ps. This allowed the carbonyl and sulfur distance to increase slightly. After this fix the SCF converged in 23 steps and has yet to become another problem for the simulation and allowed us to finally begin exploring the active site. This highlights the importance of having a well-equilibrated and reasonable starting structure for QM/MM simulations.

In order to explore the catalytic mechanism of Sortase B we need a configuration that is representative of the intermediate state of the enzyme. To do this a simulation was ran with a



Scheme 2) This scheme shows the forward and reverse reaction of the baSrtB enzyme with the DNPKTGDE ligand. On the left is the reactant state showing the nucleophilic attack from the sulfur on Cys 233 and proton transfer from His 101. Shown on the right is the ligand protein complex having a nucleophilic attack from the incoming ligand Gly. These two states represent the catalytic chemistry of baSrtB that we are exploring in our QM/MM simulations.

pulling restraint (steered MD), to pull the three bonds important to the catalytic mechanism to build the intermediate (product) state of the enzyme active site (see scheme 2). The Thr-Gly peptide bond needed to be broken, a proton transfer from the protonated histidine 101 to the Gly peptide nitrogen needed to be formed, and the thiol bond formation of the cysteine 233 Sulfur and the Ligand Thr Carbonyl Carbon. Details of the steered MD are described in the Methods Section. This intermediate configuration is necessary for the use of a diagonalized vector CVs that can represent the reaction coordinate and allow us to simulate the reaction with a single set of CVs. The steered MD was successfully able to make the tetrahedral intermediate, but unfortunately, the simulation has been unable to be restarted from the tetrahedral intermediate formed by the Thr carbonyl carbon, Gly nitrogen, catalytic Cys sulfur, and Thr backbone oxygen. This may be due to a problem with CP2K and initiating the restart file, where the initial guess for the wavefunction should be the molecular wavefunction of the previous step, and not the atomic orbitals. This is due to each time we try restarting the run from the tetrahedral formation the SCF fails to converge just like previous errors. There are two possible causes to this failure; one was the QM region was too small to allow the electron cloud to move and extend the range of the backbone oxygen to single bond distance. The solution to give more electron cloud mobility to the backbone oxygen was to increase the QM region on either side of the Thr carbonyl carbon and Gly peptide nitrogen as to make the electron cloud more mobile allowing the double bonded oxygen to pull the electron density toward it, lengthening the double bond to a single bond distance (see figure 14). The second problem could be that CP2K is unable to properly read the wave function file from the wave function restart file. Since the latter option involves dealing with the source code of CP2K and likely wouldn't finish in time for writing this thesis, it was decided to go with the first option and rebuild the system from scratch.

2.2.3 Stripped Sortase B System:

The first step was to construct a minimal system that could be simulated efficiently in CP2K. Using a reduced system gave two benefits: It allowed us to speed up the computational time for each timestep. And second, it allowed for the use of an expanded QM region to encompass more of the ligand inside of the QM region (see figure 13 and 14). Encompassing more atoms inside of the QM region allows more of the electron cloud to be moved and manipulated by atoms in proximity within the QM region. Also, by stripping the system of most of the non-essential atoms the simulation has decreased the time per step by a factor of ten allowing for simulations to only take overnight to produce results rather than 3 days for useful results. However, this new structure has its own set of challenges that needed to be solved before it was running smoothly. When we expanded the QM region, more constraints had to be put on atoms to prevent bonds breaking between atom

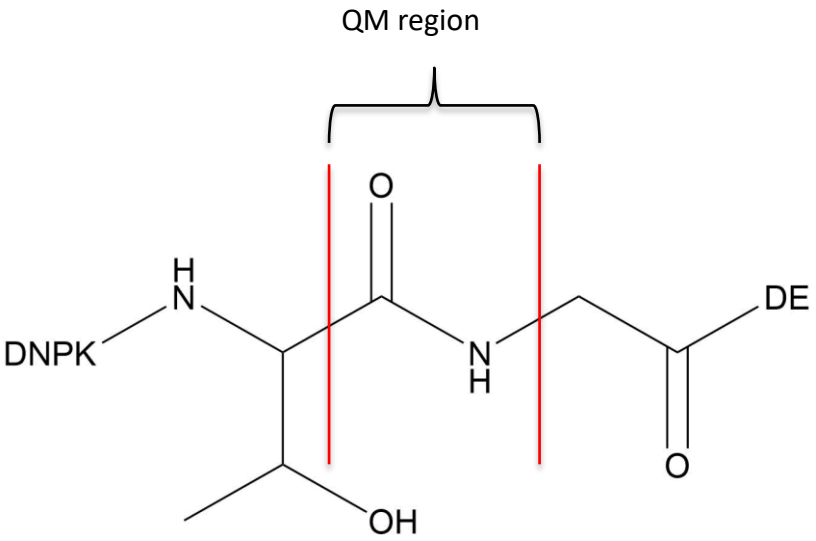


Figure 13) The representative atoms inside the QM region of the baSrtB system are shown between the two red lines consisting of the carbonyl carbon, backbone oxygen, peptide nitrogen, and amine hydrogen.

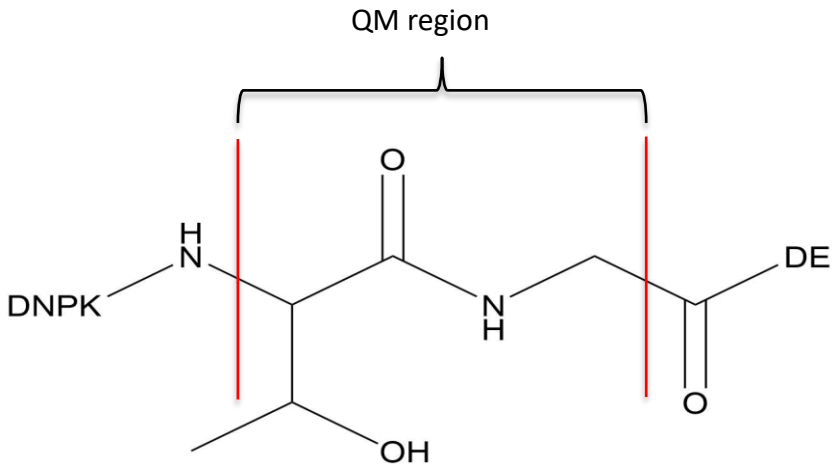


Figure 14) The representative atoms inside the QM region of the stripped baSrtB system are shown between the two red lines consisting of the carbonyl carbon, backbone oxygen, peptide nitrogen, amine hydrogen, in addition to the entire Thr side chain and both Thr and Gly α-carbons.

pairs that are not important to the overall chemistry of the Sortase B enzyme. Some atoms that needed to be restrained to their neighboring atoms were the second nitrogen on the catalytic histidine. The hydrogen atoms ended up swapping with the Gly peptide nitrogen's hydrogen and the first hydrogen on the histidine so that in the end each hydrogen had swapped positions from their original partner. The second problem was that the alcohol group on the side chain of the ligand Threonine needed to be restrained as well as when biasing the proton transfer because this led to the creation of a spurious water molecule that flew off in the simulation. During this time, some CVs were decoupled as it was discovered that the simulation may be missing the tetrahedral, state due to the coupled CVs being anti-correlated. The HLDA method, discussed above in section 1.3 and in the methods section to generate a single diagonalized

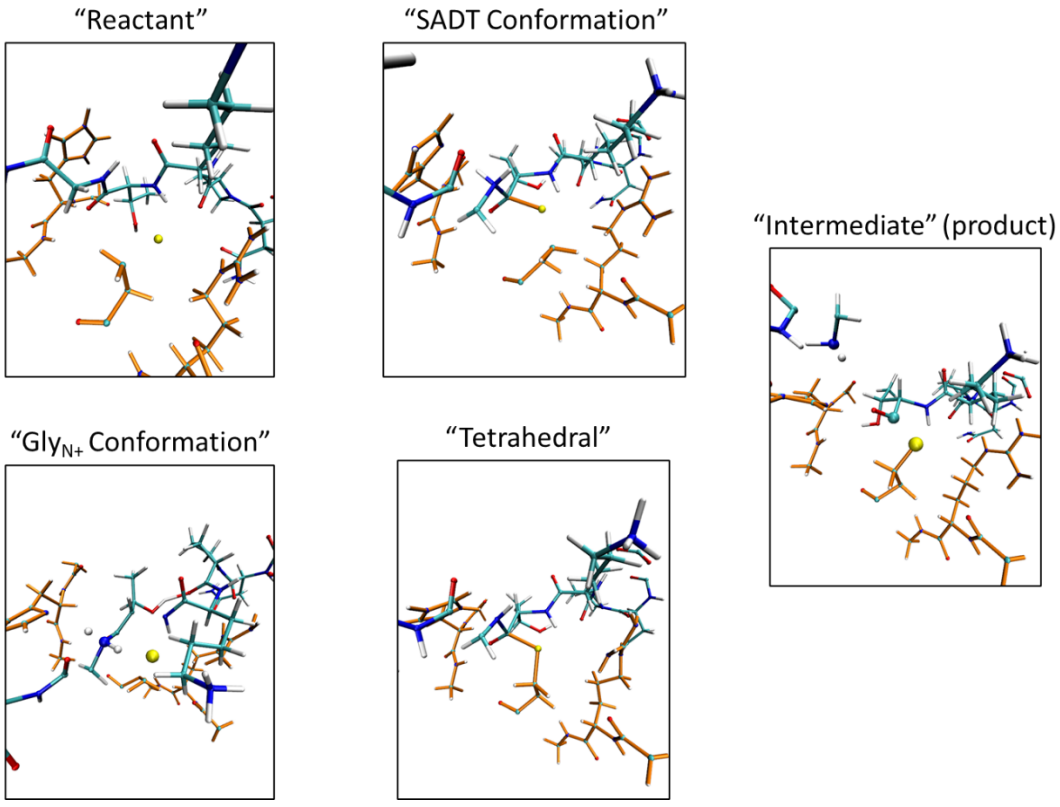


Figure 15) Five of the most important snapshots of frames from the baSrtB trajectory showing the catalytic triad (orange) and the ligand (teal) in different conformations. The representative snapshots have corresponding single point energy calculations in Table 2.

vector from reactant to product, from the steered MD, was too low dimensional for this complex reaction. The one-dimensional CV coupled the anti-correlated peptide bond breaking and threonine carbonyl carbon and catalytic cysteine sulfur thiol bond formation; the

correlation of the two was concerning due to wanting to find the tetrahedral state. So, the CVs were split into three separate CVs. The first CV is the threonine carbonyl carbon and catalytic cysteine sulfur distance (Thr_C-Cys_S), the second CV is the ligand peptide bond of the Threonine carbon and Glycine nitrogen (Thr_C-Gly_N), the third CV is a combined CV encompassing the proton transfer from the catalytic histidine to the peptide nitrogen on the Glycine (His_N-His_H minus the distance from Gly_N-His_H). The combined CV is the anti-correlated bond distances from the catalytic His nitrogen and hydrogen and the Gly peptide nitrogen and catalytic His hydrogen. As this reaction is a proton transfer the reaction mechanism is a quick and simple proton transfer and coupling the two distances to one CV should not interfere with the formation of the tetrahedral intermediate. With the three CVs working in unison, the simulation was successfully able to model the active site of the baSrtB enzyme and started to characterize the active site by taking single point energies of key frames along the trajectory. Details of the metadynamics procedure are included in the Methods Section.

From the trajectory, snapshots were taken of representative structures to perform a single energy point calculation using DFT with the DZVP basis set and both the PBE or BLYP exchange correlation functional were used. The trajectory FES yielded results, while not what was initially expected, that agreed with our single point DFT energy calculations with both BLYP and PBE. The results show a deviation from the common school of thought that the baSrtB has

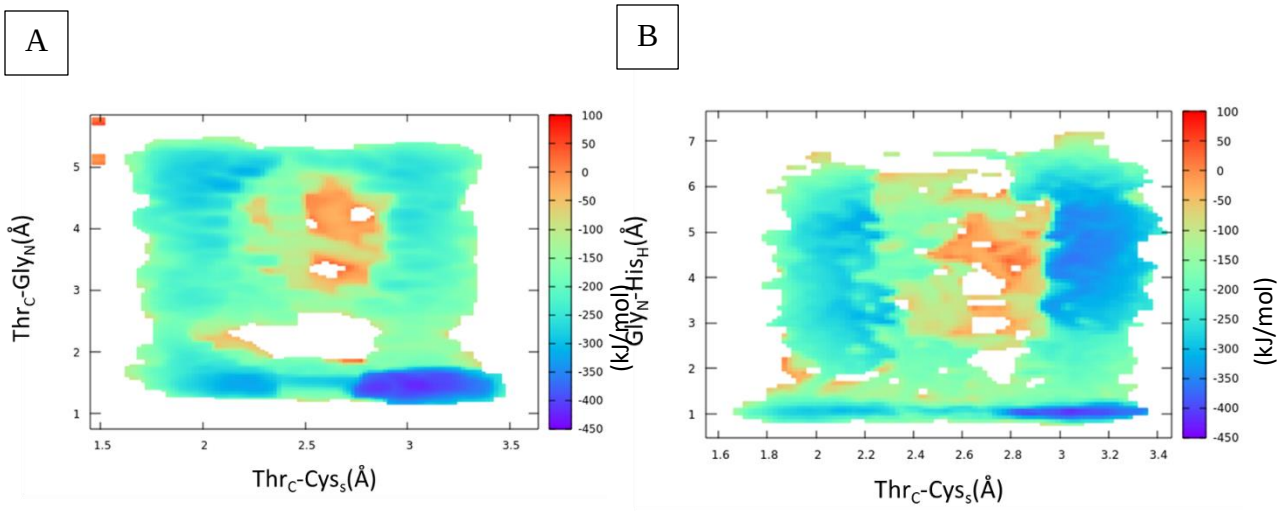


Figure 16) Shown are two FES plot with the distances from Thr_C-Cys_S, Thr_C-Gly_N, and Gly_N-His_H. FES A shows the initial starting position [3,1.5], the intermediate step as the area in the bottom left [2,1.5], and the product state [1.9,4.5]. For FES B the surface isn't as clearly defined but can be thought of as a "U" shape in its forward reaction going from top right, bottom right, bottom left, and finally top left.

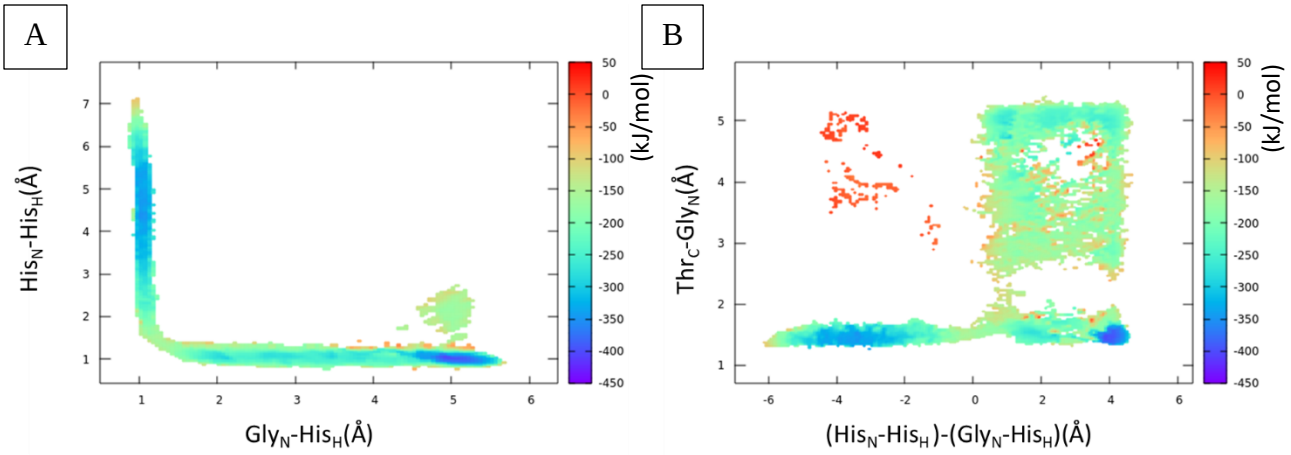


Figure 17) Shown are two FES plot with the distances from His_N-His_H, and Gly_N-His_H for part A and Thr_C-Gly_N the combined distances from His_N-His_H, minus the distance from Gly_N-His_H (This represents the anti-correlated? CV). FES A Shows the proton transfer from the His 142 to the Gly_N. FES B shows the correlation between the proton transfer CV and the from Thr_C-Gly_N, CV.

only a tetrahedral intermediate with the Thr carbonyl carbon that gets cleaved via the proton transfer. Instead, the results show that baSrtB undergoes a secondary amide dual tetrahedral (SADT) meta-stable state. When the 3 CVs are plotted against one another they show that the Cys 233 sulfur and Gly peptide nitrogen must form a tetrahedral complex in order to form the product state (see figure 15). Another requirement to form the product state is that both the proton transfer to the Gly peptide nitrogen and Cys 233 tetrahedral intermediate must be formed to have the shortest distance to get to the next basin (see figure 16). When combining the FES this with the single point energy calculations on representative frames (see table 2), it can be seen that the intermediate (Thiol bound enzyme intermediate) has the lowest energy of the whole system and is hence given the baseline of zero kJ/mol, the reactant (starting position of the baSrtB trajectory) is the next lowest at 195.9 kJ/mol for DFT-BLYP and 183.7 kJ/mol DFT-PBE, and the meta stable state of the SADT had the third lowest energy of 232.0 kJ/mol for DFT-BLYP and 203.9 for DFT-PBE. The Gly_{N+} state was over 200 kJ/mol away from that state. Even though The PBE and BLYP differed by 72.9 kJ/mol the DFT-PBE calculation puts it at over 130kJ/mol higher than the SADT configuration (see table 2). The SADT configuration also alludes to the fact how the tetrahedral could never be seen due to the speed of the chemistry without the proton transfer happening first and forming the Gly_{N+} state. The stability of the Gly_{N+} state could be due to the relative simplicity of

the proton transfer FES when compared to the other FES graphs and the amount of space the other FES must sample.

Table 2: Single Point Energy Calculations of Key Frames Along the Trajectory

DFT Basis Set	Reactant (kJ/mol)	Protonated Cys 233 (kJ/mol)	SADT Conformation (kJ/mol)	Gly_{N+} (kJ/mol)	Tetrahedral (kJ/mol)	Intermediate (product) (kJ/mol)	Broken intermediate (kJ/mol)
BLYP	195.9	339.3	232.0	422.2	410.1	0	567.0
PBE	183.7	337.0	203.9	413.7	337.2	0	607.1

Table 2) Note: the "Protonated Cys 208" column is shown to be higher in energy that the "Reactant" state. And this is possibly due to an upper wall restraint that is causing an incomplete covalent bond to the hydrogen to occur between the Gly nitrogen and Cys sulfur that exceeds distances of 1Å away from either atom in the frame of the trajectory but is shown for in the table for completeness.

3 Conclusion:

Having explored using QM/MM simulations to study the catalytic mechanisms of sortase enzymes. For Sortase A, simulations were unable to sufficient to study the system and equilibrate it with QM/MM methods. The prevailing issue was the SCF convergence error that was never resolved. A new system is needed for saSrtA to study the active site before any definitive conclusions can be made about the chemistry of saSrtA's catalysis. For Sortase B, the system was able to overcome this problem by relaxing the constraint on the carbonyl and sulfur distance. This approach of relaxing the restraint will likely work for saSrtA as well, and future work should be devoted to getting a working model of saSrtA.

For the Sortase B system, the current results validate the AlphaFold predicted binding motif, as the system was able to form the thiol tetrahedral intermediate and cleave the peptide bond using MD simulations. Unfortunately, the baSrtB system was unable to drive the reaction using a single low-dimensional reaction coordinate from reactant to the intermediate (product) state. Instead, we were able to observe the product formation using a combination of distance CVs similar to the approach used by Shrestha and Wereszczynski[35]. The QM/MM simulations support the hypothesis that the baSrtB catalysis proceeds via a protonation of the Gly peptide nitrogen first or alongside of the nucleophilic attack from the Cys 233 sulfur, exactly the same

method of activation on saSrtA discussed by Shrestha and Wereszcynski[35]. This is supported by the single point energy calculations and trajectory showing that the proton transfer finds a metastable state within the FES in the Gly_{N+} state and SADT state. It would be interesting in future work to try using machine learning based CVs to find an optimal reaction coordinate for biased simulations.

This finding is against the conventional wisdom that the protonation of the amine would be unfavorable because of the overall stability of the peptide bond. Going forward with exploring the baSrtB enzyme and other classes of sortases; we must be mindful of how enzymes subvert the common line on logic through the lens of organic chemistry. Though the research on baSrtB is fledgling at best, due to the poor enzymatic activity of the enzyme, the results show an interesting metastable state appearing where conventional organic chemistry logic would dictate the tetrahedral state would be the most stable intermediate state. More rigorous testing must be done on the baSrtB system rerunning the system with lower bias overall to give the system a better chance of sampling the SADT conformation before the enzyme catalysis. In future simulations, the bias could be lowered by reducing the Gaussian hill deposition stride on the proton transfer CV to slow down the rate of transfer in relation to the other CVs. This would be a step in the right direction to hopefully create a more realistic and stable baSrtB system in the future. Taking the lessons from the baSrtB system we should be able to quickly and accurately make a new saSrtA system. The theory to why the saSrtA failed every time the simulation was put it into a QM/MM system is thought to be because the distance between the ligand Thr carbonyl carbon and Cys 208's sulfur was too close and lead to the failure of the SCF being unable to converge and crashing the simulation. The fix for the saSrtA system would simply be allowing the Thr carbonyl carbon to drift father away during the NPT and NVT calibration runs. The same problem that was occurring with the baSrtB system and allowing the Thr carbonyl carbon to relax fixed the problem. Hopefully this work on enzyme catalysis using QM/MM computational analysis can give us greater detail about enzymatic processes and that experiences and lessons learned here may translate to other enzymes for future simulations.

4 Materials and Methods:

4.1 Ligand-Receptor Docking:

The preparation of ligand receptor structures for AutoDock consists in preparing a PDBQT file that is an extension of the PDB format with an additional field for partial charges and atom type. The PDB files of the ligand and protein were converted into PDBQT format using scripts provided with the AutoDockTools suite. In scoring and docking with AutoDock Vina and FlexPep Dock a grid was prepared for the receptor with the grid cent chosen as the center of the binding site. The grid dimensions were extend by 10 Å in the x, y, and z directions with the default grid spacing of 0.375 Å, In the case of AutoDock Vina version 1.1.2 an exhaustiveness of 100 was used in the single point mode. FlexPepDock docking was performed using the online Server at flexpepdock.furmanlab.cs.huji.ac.il/index.php. Default parameters were used in the docking procedure.

4.1 Docking Refinement for DNPKTGDE Ligand:

The docking refinement for the recognition motif of baSrtB was done by using PYMOL's builder. The Thr and Gly residues of the ligand were inserted into the active site of baSrtB such that the carbonyl carbon on the Thr was 2Å away from the catalytic Cys. PLUMED was then used to create an upper wall boundary at 2.5Å with a kappa of 600. The system was then treated with the ff99SB-ILDN[37] and TIP3P water molecules. A Short Energy Minimization (EM) simulation was run for 1000 steps and using the path of steepest decent patched with PLUMED. The lowest energy structure was extracted, and the procedure was repeated except adding a single amino acid of the recognition motif. This was done iteratively alternating C-terminal and N-terminal amino acids of the ligand PyMol[38] until the baSrtB ligand was complete.

4.2 Creating the Product and Reactant States via Moving Restaint:

All QM/MM simulations on baSrtB used CP2K version 8.1[19] using the semiempirical PM6 method and Coulomb embedding with the MM region described by the Amber force field ff99SB-ILDN[37] and patched with PLUMED 2.8[36-38]. We used a timestep of 0.5 fs to update the atomic positions. Utilizing the moving restraint function[37] of PLUMED[35-37] and QMMM linking atom[44-48], we ran the reactant state unbiased for 50,000 steps. Over the next period of 50,000 steps, we utilized a 3-stage moving restraint that covered the entire catalytic process of the baSrtB enzyme. The moving restraint is a harmonic restraint given by

$$V(s,t) = \frac{1}{2}\kappa(t)(s - s_0(t))^2$$

where κ is a time-dependent force constant and $s_0$is a moving target value of the CV. Our moving restraint had three phases; the first 20,000 steps applied 3 moving restraints to the distances between the peptide bond of the Thr and Gly of the ligand, the catalytic Cys and the Thr of the ligand, and the proton and the Gly nitrogen. The peptide bond was set at zero kappa and increased to 300 kappa at 5.0Å. The catalytic Cys and Thr bond was set at zero kappa and increased to 600 kappa and 1.5Å. The final distance of the Hydrogen transfer to the Nitrogen on the Gly was set at zero kappa and increased to 600 kappa and 1.5Å. All the starting distances were chosen to be slightly longer than the starting bond distance of the starting structure. These bonds were then held for 10,000 steps before being slowly released over 20,000 steps where the kappa was reduced to zero. The system was then allowed to equilibrate with no bias for 50,000 steps.

4.3 HLDA Calculations:

Harmonic Linear Discriminant Analysis (HLDA) is a new method to construct CVs for studying complex reactions in *ab initio* MD with metadynamics[43]. Briefly, we used 25 ps of an unbiased QM/MM trajectory in both the reactant state and product state. The product state was created from the moving-restrain simulations described immediately above. We then define a vector as a set of descriptors, $\boldsymbol{d} = \{d_i\}$, that are able to distinguish between the reactant and product states. From the unbiased trajectory data we also compute both the mean, $\mu = \{\overline{d_i}\}$, and the

covariance matrix, Σ, of the set of descriptors for each state. Equation 2 (shown below) is the HLDA equation that gives a low-dimensional CV that can be biased via metadynamics:

$$\tilde{s}_H(\boldsymbol{R}) = (\mu_R - \mu_P)^T \left(\frac{\mathbf{1}}{\tilde{\Sigma}_{\boldsymbol{R}}} + \frac{\mathbf{1}}{\tilde{\Sigma}_{\boldsymbol{P}}}\right) \widehat{\boldsymbol{d}}(\boldsymbol{R}) \qquad 2)$$

Where $\tilde{s}_H$ is our HLDA CV which is a linear combination of the descriptors and is capable of describing the transition from one local minima to our desired product. $(\mu_R - \mu_P)^T$ are the mean of the reactants and products respectively and $\left(\frac{\mathbf{1}}{\tilde{\Sigma}_{\boldsymbol{R}}} + \frac{\mathbf{1}}{\tilde{\Sigma}_{\boldsymbol{P}}}\right)$ are the inverse covariance matrices, and $\widehat{\boldsymbol{d}}$ is the descriptors in any configuration $(\boldsymbol{R})$ that you are using for the simulation. In our simulation, our descriptors are three distance CVs to monitor the reaction. The distances used were the distance from the carbonyl carbon to the catalytic cysteine, the peptide bond between the ligand threonine and glycine, and the final distance monitored was the distance from the proton and the nitrogen on the glycine. The final HLDA CV is a linear combination of the three distances with coefficients of -0.018, 0.989, and 0.144 respectively.

4.4 QM/MM Simulations:

All QM/MM simulations were done using CP2K version 8.1[19] patched with PLUMED 2.8.[36-38] with the linker atom function built into CP2K. A timestep of 0.5 fs was used and a constant temperature of 300 K was maintained using the velocity rescaling thermostat. The system was built using the Ambertools Software package. The protein was described by the Amber99SB-ILDN force fields with the TIP3P water model. Long range electrostatics are computed using the Smoothed Particle Mesh Ewald (SPME) method. The MM step in CP2K uses the FIST molecular mechanics module. The QM force evaluation uses the Quickstep algorithm based on the Gaussian and plane waves method (GPW) and its augmented extension (GAPW). The QM system is described at the PM6 semiempirical level. . Metadynamics simulations were run with a bias factor of 80, a hill deposition stride of 100, and initial hill height of W=5.00 . We used an

adaptive Gaussian scheme for the hill widths. All simulations were performed on an 8-core Linux workstation with Intel i7-9700 CPUs.

4.4 DFT Energy Calculations:

Single point energy calculations were performed using CP2K version 8.1 using the DZVP-MOLOPT-GTH basis set[19] and GTH-PBE or GTH-BLYP pseudopotential. We tested two different exchange correlation functionals, the PBE (Perdew-Burke-Ernzerhof GGA functional)[20] and the BLYP (Becke-Lee-Yang-Parr hybrid functional)[21,22]. For electrostatic embedding, we used the efficient Gaussian Expansion of the Electrostatic Potential (GEEP) method as implemented in CP2K[19].

Bibliography:

1. Alberts, B. *et al.* Protein Function. (2002).

2. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85 (2016).

3. Lin, Q. *et al.* The seahorse genome and the evolution of its specialized morphology. *Nature* **540**, 395–399 (2016).

4. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed* **48**, 1198–1229 (2009).

5. Antos, J. M., Truttmann, M. C. & Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118 (2016).

6. David Row, R., Roark, T. J., Philip, M. C., Perkins, L. L. & Antos, J. M. Enhancing the efficiency of sortase-mediated ligations through nickel-peptide complex formation. *Chem. Commun.* **51**, 12548–12551 (2015).

7. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. Sortase transpeptidases: structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264 (2017).

8. Spirig, T., Weiner, E. M. & Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059 (2011).

9. Antos, J. M., Truttmann, M. C. & Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118 (2016).

10. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).

11. Barnett, R. N. & Landman, U. Born-Oppenheimer molecular-dynamics simulations of finite systems: Structure and dynamics of (O. *Phys. Rev. B* **48**, 2081–2097 (1993).

12. Field, M. J. Simulated annealing, classical molecule dynamics and the Hartree—Fock method: the NDDO approximation. *Chem. Phys. Lett.* **172**, 83–88 (1990).

13. Gibson, D. A. & Carter, E. A. Time-reversible multiple time scale ab initio molecular dynamics. *J. Phys. Chem.* **97**, 13429–13434 (1993).

14. Schlegel, H. B. *et al. Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **114**, 9758–9763 (2001).

15. Kvaal, S. Ab initio quantum dynamics using coupled-cluster. *J. Chem. Phys.* **136**, 194109 (2012).

16. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

17. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.* **38**, 3098–3100 (1988).

18. Görling, A. Exact treatment of exchange in Kohn-Sham band-structure schemes. *Phys Rev, B Condens Matter* **53**, 7024–7029 (1996).

19. Kühne, T. D. *et al.* CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).

20. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).

21. Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **10**, 221–264 (1989).

22. Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **10**, 221–264 (1989).

23. Porezag, Dirk, et al. "Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon." *Physical Review B* 51.19 (1995): 12947.

24. Bannwarth, Christoph, et al. "Extended tight-binding quantum chemistry methods." Wiley Interdisciplinary Reviews: Computational Molecular Science 11.2 (2021): e1493.

25.Kowalczyk, T., Wang, L.-P. & Van Voorhis, T. Simulation of solution phase electron transfer in a compact donor-acceptor dyad. *J. Phys. Chem. B* **115**, 12135–12144 (2011).

26. Laio, Alessandro, and Michele Parrinello. "Escaping free-energy minima." Proceedings of the national academy of sciences 99.20 (2002): 12562-12566.

27. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).

28. Tiwary, P. & Parrinello, M. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **119**, 736–742 (2015).

29. Tiwary, P. & Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **111**, 230602 (2013).

30. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: Applications of autodock. *J. Mol. Recognit.* **9**, 1–5 (1996).

31. London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* **39**, W249-53 (2011).

32. Jumper, J. & Hassabis, D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* **19**, 11–12 (2022).

33. Johnson, D. A. *et al.* Structures of Streptococcus pyogenes class A sortase in complex with substrate and product mimics provide key details of target recognition. *J. Biol. Chem.* **298**, 102446 (2022).

34. Tian, B.-X. & Eriksson, L. A. Catalytic mechanism and roles of Arg197 and Thr183 in the Staphylococcus aureus sortase A enzyme. *J. Phys. Chem. B* **115**, 13003–13011 (2011).

35. Shrestha, P. & Wereszczynski, J. Discerning the catalytic mechanism of Staphylococcus aureus sortase A with QM/MM free energy calculations. *J. Mol. Graph. Model.* **67**, 33–43 (2016).

36. M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello. *PLUMED: a portable plugin for free energy calculations with molecular dynamics*, Comp. Phys. Comm. 180, 1961 (2009), preprint available as arXiv:0902.0874

37. The PLUMED consortium. *Promoting transparency and reproducibility in enhanced molecular simulations*, Nat. Methods 16, 670 (2019)

38. G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. *PLUMED2: New feathers for an old bird*, Comp. Phys. Comm. 185, 604 (2014), preprint available as arXiv:1310.0980

39. Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 (2009).

40. The PyMOL Molecular Graphics System, Version 1.8.4.0 Schrödinger, LLC.

41. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).

42. Xiao, C. & Zhang, Y. Design-atom approach for the quantum mechanical/molecular mechanical covalent boundary: a design-carbon atom with five valence electrons. *J. Chem. Phys.* **127**, 124102 (2007).

43. Mendels, D., Piccini, G. & Parrinello, M. Collective Variables from Local Fluctuations. *J. Phys. Chem. Lett.* **9**, 2776–2781 (2018).

44. Li, W.-L., Chen, K., Rossomme, E., Head-Gordon, M. & Head-Gordon, T. Optimized pseudopotentials and basis sets for semiempirical density functional theory for electrocatalysis applications. *J. Phys. Chem. Lett.* **12**, 10304–10309 (2021).

45. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

46. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.* **38**, 3098–3100 (1988).

47. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev, B Condens Matter* **37**, 785–789 (1988).

48. Laino, T., Mohamed, F., Laio, A. & Parrinello, M. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **1**, 1176–1184 (2005).

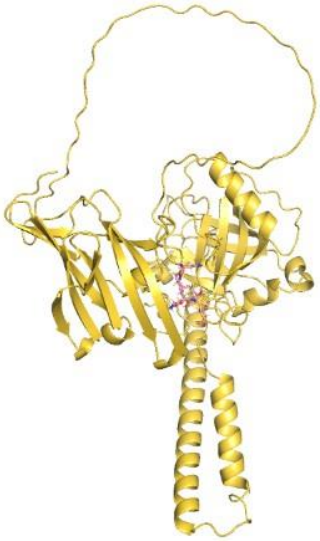
5 Appendix:

SrtB B. Anthracis

Full length precursors (from NCBI):

>APT28041.1 sortase B [Bacillus anthracis]
MSSEKERKKKIFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAEAQNIYEKSPMEEQSQDGEVR
KQFKALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYKGEDMRAGSIFMDYRNDVKSQNRNTI
LYGHRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTS
FLEKIQEKSLYKTDTTVTAGDQIVTLSTCDYALDPEAGRLVVHAKLVKRQ

>WP_154136074.1 heme uptake protein IsdC [Bacillus anthracis]
MKKVSVLPAFIITFVCMLAFLVMPYGNASAKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKN
GEMRVQVPMNHSAWITEFKAPENGNFVDAKVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYD
HHYTIRFAFDANVKAVGGDNGVAATTKNNDQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKT
DNPKTGDEARIGLFAALILISGVFLI RKVKLSK



Galaxy36-[Alphafold_2_on__Model_5]_baSrtB.pbd

Sequence of chimera modeled withAlphaFold:

MKLADGTYDINYVIQKAENDSASMANDYFEKPAKLIVKNGEMRVQVPMNHSAWITEFKAPENGNFVDA
KVVSKDESADKRTVEFKVDDLSKPEAVKIHVVVPNANYDHHYTIRFAFDANVKAVGGDNGVAATTKNN
DQAKTDTQVKEEKTKVESKETAKEVNKETKNENGKAEKTDNPKTGDEARIGLFAALILISGVFLI GGG
GSGGGGSIFFQRILTVVFLGTFFYSVYELGDIFMDYYENRKVMAEAQNIYEKSPMEEQSQDGEVRKQF
KALQQINQEIVGWITMDDTQINYPIVQAKDNDYYLFRNYKGEDMRAGSIFMDYRNDVKSQNRNTILYG
HRMKDGSMFGSLKKMLDEEFFMSHRKLYYDTLFEGYDLEVFSVYTTTTDFYYIETDFSSDTEYTSFLE
KIQEKSLYKTDTTVTAGDQIVTLSTCDYALDPEAGRLVVHAKLVKRQ

baSrtB Alphafold sequence courtesy of Dr. John Antos of Western Washington University.

Starting coordinates to the Stripped baSrtB PDB file:

CRYST1 30.000 30.000 30.000 90.00 90.00 90.00 P 1 1
ATOM 1 H1 ACE 1 -1.079 3.099 16.241 1.00 0.00
ATOM 2 CH3 ACE 1 -2.165 3.035 16.306 1.00 0.00
ATOM 3 H2 ACE 1 -2.559 3.961 16.723 1.00 0.00
ATOM 4 H3 ACE 1 -2.442 2.200 16.949 1.00 0.00
ATOM 5 C ACE 1 -2.781 2.819 14.941 1.00 0.00
ATOM 6 O ACE 1 -2.074 2.736 13.928 1.00 0.00
ATOM 7 N ASN 2 -4.062 2.717 14.831 1.00 0.00
ATOM 8 H ASN 2 -4.660 2.858 15.632 1.00 0.00
ATOM 9 CA ASN 2 -4.688 2.507 13.528 1.00 0.00
ATOM 10 HA ASN 2 -4.039 2.854 12.724 1.00 0.00
ATOM 11 CB ASN 2 -4.931 0.999 13.348 1.00 0.00
ATOM 12 HB2 ASN 2 -5.426 0.685 14.267 1.00 0.00

```
ATOM     13  HB3 ASN     2      -5.598   0.835  12.502  1.00  0.00
ATOM     14  CG  ASN     2      -3.688   0.155  13.155  1.00  0.00
ATOM     15  OD1 ASN     2      -2.602   0.624  12.865  1.00  0.00
ATOM     16  ND2 ASN     2      -3.826  -1.149  13.252  1.00  0.00
ATOM     17 HD21 ASN     2      -3.023  -1.749  13.130  1.00  0.00
ATOM     18 HD22 ASN     2      -4.735  -1.544  13.448  1.00  0.00
ATOM     19  C   ASN     2      -6.007   3.251  13.836  1.00  0.00
ATOM     20  O   ASN     2      -6.570   3.075  14.915  1.00  0.00
ATOM     21  N   NME     3      -6.585   4.169  12.843  1.00  0.00
ATOM     22  H   NME     3      -6.323   4.474  11.917  1.00  0.00
ATOM     23  CH3 NME     3      -7.804   4.703  13.425  1.00  0.00
ATOM     24 HH31 NME     3      -7.948   4.281  14.420  1.00  0.00
ATOM     25 HH32 NME     3      -7.727   5.788  13.499  1.00  0.00
ATOM     26 HH33 NME     3      -8.653   4.442  12.794  1.00  0.00
TER
ATOM     27  H1  ACE     4      -4.687   6.702  10.208  1.00  0.00
ATOM     28  CH3 ACE     4      -5.553   7.265  10.556  1.00  0.00
ATOM     29  H2  ACE     4      -5.496   8.288  10.183  1.00  0.00
ATOM     30  H3  ACE     4      -5.566   7.276  11.646  1.00  0.00
ATOM     31  C   ACE     4      -6.842   6.642  10.063  1.00  0.00
ATOM     32  O   ACE     4      -6.828   5.623   9.359  1.00  0.00
ATOM     33  N   LEU     5      -7.969   7.181  10.380  1.00  0.00
ATOM     34  H   LEU     5      -8.005   7.976  11.002  1.00  0.00
ATOM     35  CA  LEU     5      -9.227   6.606   9.921  1.00  0.00
ATOM     36  HA  LEU     5      -9.070   5.990   9.036  1.00  0.00
ATOM     37  CB  LEU     5      -9.799   5.720  11.041  1.00  0.00
ATOM     38  HB2 LEU     5      -9.021   5.012  11.326  1.00  0.00
ATOM     39  HB3 LEU     5      -9.995   6.398  11.872  1.00  0.00
ATOM     40  CG  LEU     5     -11.080   4.957  10.693  1.00  0.00
ATOM     41  HG  LEU     5     -11.829   5.680  10.370  1.00  0.00
ATOM     42  CD1 LEU     5     -10.853   3.928   9.584  1.00  0.00
ATOM     43 HD11 LEU     5     -10.104   3.205   9.907  1.00  0.00
ATOM     44 HD12 LEU     5     -11.788   3.411   9.370  1.00  0.00
ATOM     45 HD13 LEU     5     -10.504   4.434   8.684  1.00  0.00
ATOM     46  CD2 LEU     5     -11.568   4.188  11.925  1.00  0.00
ATOM     47 HD21 LEU     5     -11.772   4.889  12.734  1.00  0.00
ATOM     48 HD22 LEU     5     -12.480   3.645  11.677  1.00  0.00
ATOM     49 HD23 LEU     5     -10.800   3.482  12.241  1.00  0.00
ATOM     50  C   LEU     5     -10.229   7.693   9.480  1.00  0.00
ATOM     51  O   LEU     5     -10.405   8.676  10.145  1.00  0.00
ATOM     52  N   NME     6     -10.978   7.522   8.228  1.00  0.00
ATOM     53  H   NME     6     -11.014   6.812   7.510  1.00  0.00
ATOM     54  CH3 NME     6     -11.828   8.690   8.084  1.00  0.00
ATOM     55 HH31 NME     6     -11.676   9.358   8.932  1.00  0.00
```

```
ATOM     56 HH32 NME     6     -11.576   9.212   7.161  1.00  0.00
ATOM     57 HH33 NME     6     -12.872   8.378   8.051  1.00  0.00
TER
ATOM     58  H1  ACE     7     -13.037  14.626   8.049  1.00  0.00
ATOM     59  CH3 ACE     7     -13.113  14.164   7.065  1.00  0.00
ATOM     60  H2  ACE     7     -12.114  14.000   6.662  1.00  0.00
ATOM     61  H3  ACE     7     -13.670  14.820   6.397  1.00  0.00
ATOM     62  C   ACE     7     -13.828  12.833   7.142  1.00  0.00
ATOM     63  O   ACE     7     -14.244  12.394   8.223  1.00  0.00
ATOM     64  N   ARG     8     -14.009  12.145   6.066  1.00  0.00
ATOM     65  H   ARG     8     -14.216  12.775   5.304  1.00  0.00
ATOM     66  CA  ARG     8     -14.701  10.848   6.106  1.00  0.00
ATOM     67  HA  ARG     8     -14.694  10.419   7.108  1.00  0.00
ATOM     68  CB  ARG     8     -16.154  11.110   5.696  1.00  0.00
ATOM     69  HB2 ARG     8     -16.515  11.936   6.309  1.00  0.00
ATOM     70  HB3 ARG     8     -16.141  11.413   4.649  1.00  0.00
ATOM     71  CG  ARG     8     -17.086   9.901   5.867  1.00  0.00
ATOM     72  HG2 ARG     8     -16.660   9.054   5.328  1.00  0.00
ATOM     73  HG3 ARG     8     -17.153   9.660   6.928  1.00  0.00
ATOM     74  CD  ARG     8     -18.481  10.199   5.325  1.00  0.00
ATOM     75  HD2 ARG     8     -19.119   9.327   5.465  1.00  0.00
ATOM     76  HD3 ARG     8     -18.905  11.048   5.860  1.00  0.00
ATOM     77  NE  ARG     8     -18.432  10.524   3.885  1.00  0.00
ATOM     78  HE  ARG     8     -17.636  10.141   3.396  1.00  0.00
ATOM     79  CZ  ARG     8     -19.302  11.240   3.205  1.00  0.00
ATOM     80  NH1 ARG     8     -20.401  11.690   3.742  1.00  0.00
ATOM     81 HH11 ARG     8     -20.605  11.490   4.711  1.00  0.00
ATOM     82 HH12 ARG     8     -21.044  12.236   3.187  1.00  0.00
ATOM     83  NH2 ARG     8     -19.053  11.508   1.956  1.00  0.00
ATOM     84 HH21 ARG     8     -18.204  11.166   1.529  1.00  0.00
ATOM     85 HH22 ARG     8     -19.710  12.056   1.420  1.00  0.00
ATOM     86  C   ARG     8     -13.589   9.862   6.062  1.00  0.00
ATOM     87  O   ARG     8     -12.389  10.074   6.158  1.00  0.00
ATOM     88  N   NME     9     -13.886   8.403   5.757  1.00  0.00
ATOM     89  H   NME     9     -13.738   7.725   6.491  1.00  0.00
ATOM     90  CH3 NME     9     -14.370   7.931   4.471  1.00  0.00
ATOM     91 HH31 NME     9     -14.478   8.776   3.791  1.00  0.00
ATOM     92 HH32 NME     9     -13.659   7.218   4.054  1.00  0.00
ATOM     93 HH33 NME     9     -15.337   7.445   4.602  1.00  0.00
TER
ATOM     94  H1  ACE    10      -7.329   2.625   4.659  1.00  0.00
ATOM     95  CH3 ACE    10      -7.824   3.390   4.060  1.00  0.00
ATOM     96  H2  ACE    10      -7.386   3.410   3.062  1.00  0.00
ATOM     97  H3  ACE    10      -7.693   4.363   4.533  1.00  0.00
```

```
ATOM     98  C   ACE    10     -9.306   3.109   3.933  1.00  0.00
ATOM     99  O   ACE    10     -9.258   2.075   4.613  1.00  0.00
ATOM    100  N   HIP    11     -9.918   3.413   2.839  1.00  0.00
ATOM    101  H   HIP    11     -9.433   3.890   2.092  1.00  0.00
ATOM    102  CA  HIP    11    -11.355   3.154   2.682  1.00  0.00
ATOM    103  HA  HIP    11    -11.689   2.381   3.374  1.00  0.00
ATOM    104  CB  HIP    11    -12.129   4.481   3.001  1.00  0.00
ATOM    105  HB2 HIP    11    -11.683   5.323   2.471  1.00  0.00
ATOM    106  HB3 HIP    11    -13.159   4.348   2.670  1.00  0.00
ATOM    107  CG  HIP    11    -12.192   4.878   4.484  1.00  0.00
ATOM    108  ND1 HIP    11    -11.067   5.011   5.304  1.00  0.00
ATOM    109  HD1 HIP    11    -10.072   5.011   5.133  1.00  0.00
ATOM    110  CE1 HIP    11    -11.675   5.157   6.523  1.00  0.00
ATOM    111  HE1 HIP    11    -10.932   5.139   7.320  1.00  0.00
ATOM    112  NE2 HIP    11    -13.035   5.295   6.567  1.00  0.00
ATOM    113  HE2 HIP    11    -13.704   5.473   7.303  1.00  0.00
ATOM    114  CD2 HIP    11    -13.355   5.118   5.219  1.00  0.00
ATOM    115  HD2 HIP    11    -14.393   5.178   4.893  1.00  0.00
ATOM    116  C   HIP    11    -11.673   2.626   1.315  1.00  0.00
ATOM    117  O   HIP    11    -11.258   3.037   0.234  1.00  0.00
ATOM    118  N   NME    12    -12.590   1.479   1.262  1.00  0.00
ATOM    119  H   NME    12    -13.085   0.930   1.950  1.00  0.00
ATOM    120  CH3 NME    12    -12.755   1.137  -0.140  1.00  0.00
ATOM    121 HH31 NME    12    -12.162   1.816  -0.752  1.00  0.00
ATOM    122 HH32 NME    12    -12.421   0.113  -0.306  1.00  0.00
ATOM    123 HH33 NME    12    -13.806   1.225  -0.415  1.00  0.00
TER
ATOM    124  H1  ACE    13    -15.541   3.227   0.463  1.00  0.00
ATOM    125  CH3 ACE    13    -15.804   4.087  -0.153  1.00  0.00
ATOM    126  H2  ACE    13    -14.912   4.461  -0.655  1.00  0.00
ATOM    127  H3  ACE    13    -16.221   4.872   0.477  1.00  0.00
ATOM    128  C   ACE    13    -16.829   3.713  -1.202  1.00  0.00
ATOM    129  O   ACE    13    -17.264   2.557  -1.285  1.00  0.00
ATOM    130  N   LYS    14    -17.257   4.613  -2.021  1.00  0.00
ATOM    131  H   LYS    14    -16.956   5.573  -1.935  1.00  0.00
ATOM    132  CA  LYS    14    -18.257   4.282  -3.056  1.00  0.00
ATOM    133  HA  LYS    14    -18.241   3.211  -3.259  1.00  0.00
ATOM    134  CB  LYS    14    -19.666   4.670  -2.579  1.00  0.00
ATOM    135  HB2 LYS    14    -19.644   5.680  -2.169  1.00  0.00
ATOM    136  HB3 LYS    14    -20.356   4.636  -3.422  1.00  0.00
ATOM    137  CG  LYS    14    -20.132   3.693  -1.500  1.00  0.00
ATOM    138  HG2 LYS    14    -20.275   2.723  -1.977  1.00  0.00
ATOM    139  HG3 LYS    14    -19.335   3.619  -0.760  1.00  0.00
ATOM    140  CD  LYS    14    -21.430   4.119  -0.811  1.00  0.00
```

```
ATOM    141  HD2 LYS    14     -21.359   5.121  -0.388  1.00  0.00
ATOM    142  HD3 LYS    14     -22.286   4.069  -1.484  1.00  0.00
ATOM    143  CE  LYS    14     -21.609   3.101   0.318  1.00  0.00
ATOM    144  HE2 LYS    14     -21.914   2.160  -0.139  1.00  0.00
ATOM    145  HE3 LYS    14     -20.639   2.974   0.798  1.00  0.00
ATOM    146  NZ  LYS    14     -22.603   3.502   1.323  1.00  0.00
ATOM    147  HZ1 LYS    14     -23.502   3.619   0.879  1.00  0.00
ATOM    148  HZ2 LYS    14     -22.669   2.788   2.034  1.00  0.00
ATOM    149  HZ3 LYS    14     -22.320   4.374   1.747  1.00  0.00
ATOM    150  C   LYS    14     -17.960   4.941  -4.397  1.00  0.00
ATOM    151  O   LYS    14     -17.634   6.126  -4.426  1.00  0.00
ATOM    152  N   NME    15     -18.059   4.168  -5.644  1.00  0.00
ATOM    153  H   NME    15     -18.297   3.214  -5.873  1.00  0.00
ATOM    154  CH3 NME    15     -17.718   5.072  -6.728  1.00  0.00
ATOM    155 HH31 NME    15     -17.487   6.057  -6.322  1.00  0.00
ATOM    156 HH32 NME    15     -16.850   4.688  -7.263  1.00  0.00
ATOM    157 HH33 NME    15     -18.561   5.151  -7.414  1.00  0.00
TER
ATOM    158  H1  ACE    16      -6.532 -14.271  11.346  1.00  0.00
ATOM    159  CH3 ACE    16      -6.424 -13.222  11.072  1.00  0.00
ATOM    160  H2  ACE    16      -6.062 -12.657  11.931  1.00  0.00
ATOM    161  H3  ACE    16      -7.390 -12.826  10.759  1.00  0.00
ATOM    162  C   ACE    16      -5.441 -13.055   9.934  1.00  0.00
ATOM    163  O   ACE    16      -4.873 -14.034   9.432  1.00  0.00
ATOM    164  N   TYR    17      -5.191 -11.873   9.480  1.00  0.00
ATOM    165  H   TYR    17      -5.704 -11.072   9.820  1.00  0.00
ATOM    166  CA  TYR    17      -4.240 -11.675   8.373  1.00  0.00
ATOM    167  HA  TYR    17      -3.537 -12.506   8.322  1.00  0.00
ATOM    168  CB  TYR    17      -5.012 -11.606   7.053  1.00  0.00
ATOM    169  HB2 TYR    17      -4.280 -11.533   6.249  1.00  0.00
ATOM    170  HB3 TYR    17      -5.560 -12.544   6.960  1.00  0.00
ATOM    171  CG  TYR    17      -5.979 -10.446   6.962  1.00  0.00
ATOM    172  CD1 TYR    17      -7.343 -10.627   7.249  1.00  0.00
ATOM    173  HD1 TYR    17      -7.709 -11.614   7.532  1.00  0.00
ATOM    174  CE1 TYR    17      -8.234  -9.543   7.172  1.00  0.00
ATOM    175  HE1 TYR    17      -9.276  -9.662   7.468  1.00  0.00
ATOM    176  CZ  TYR    17      -7.758  -8.280   6.755  1.00  0.00
ATOM    177  OH  TYR    17      -8.621  -7.255   6.566  1.00  0.00
ATOM    178  HH  TYR    17      -9.533  -7.495   6.744  1.00  0.00
ATOM    179  CE2 TYR    17      -6.393  -8.095   6.486  1.00  0.00
ATOM    180  HE2 TYR    17      -6.023  -7.113   6.191  1.00  0.00
ATOM    181  CD2 TYR    17      -5.504  -9.176   6.597  1.00  0.00
ATOM    182  HD2 TYR    17      -4.442  -9.031   6.400  1.00  0.00
ATOM    183  C   TYR    17      -3.314 -10.458   8.542  1.00  0.00
```

```
ATOM    184  O   TYR    17     -3.775  -9.384   8.913  1.00  0.00
ATOM    185  N   NME    18     -1.877 -10.584   8.261  1.00  0.00
ATOM    186  H   NME    18     -1.273 -11.334   7.955  1.00  0.00
ATOM    187  CH3 NME    18     -1.285  -9.281   8.511  1.00  0.00
ATOM    188 HH31 NME    18     -2.060  -8.585   8.832  1.00  0.00
ATOM    189 HH32 NME    18     -0.530  -9.368   9.292  1.00  0.00
ATOM    190 HH33 NME    18     -0.821  -8.911   7.597  1.00  0.00
TER
ATOM    191  H1  ACE    19     -3.244  -5.033   1.057  1.00  0.00
ATOM    192  CH3 ACE    19     -3.675  -4.193   1.601  1.00  0.00
ATOM    193  H2  ACE    19     -3.926  -3.396   0.901  1.00  0.00
ATOM    194  H3  ACE    19     -2.953  -3.823   2.329  1.00  0.00
ATOM    195  C   ACE    19     -4.933  -4.606   2.334  1.00  0.00
ATOM    196  O   ACE    19     -5.354  -5.770   2.273  1.00  0.00
ATOM    197  N   SER    20     -5.578  -3.735   3.033  1.00  0.00
ATOM    198  H   SER    20     -5.228  -2.793   3.132  1.00  0.00
ATOM    199  CA  SER    20     -6.795  -4.101   3.752  1.00  0.00
ATOM    200  HA  SER    20     -7.237  -5.001   3.324  1.00  0.00
ATOM    201  CB  SER    20     -6.443  -4.378   5.214  1.00  0.00
ATOM    202  HB2 SER    20     -5.680  -5.156   5.248  1.00  0.00
ATOM    203  HB3 SER    20     -6.048  -3.464   5.657  1.00  0.00
ATOM    204  OG  SER    20     -7.585  -4.803   5.944  1.00  0.00
ATOM    205  HG  SER    20     -7.338  -4.970   6.856  1.00  0.00
ATOM    206  C   SER    20     -7.851  -3.010   3.636  1.00  0.00
ATOM    207  O   SER    20     -7.602  -1.846   3.949  1.00  0.00
ATOM    208  N   THR    21     -9.058  -3.396   3.224  1.00  0.00
ATOM    209  H   THR    21     -9.150  -4.317   2.820  1.00  0.00
ATOM    210  CA  THR    21    -10.224  -2.507   3.200  1.00  0.00
ATOM    211  HA  THR    21    -10.171  -1.744   3.977  1.00  0.00
ATOM    212  CB  THR    21    -10.327  -1.813   1.838  1.00  0.00
ATOM    213  HB  THR    21     -9.326  -1.459   1.593  1.00  0.00
ATOM    214  CG2 THR    21    -10.850  -2.687   0.701  1.00  0.00
ATOM    215 HG21 THR    21    -11.852  -3.041   0.945  1.00  0.00
ATOM    216 HG22 THR    21    -10.885  -2.104  -0.219  1.00  0.00
ATOM    217 HG23 THR    21    -10.187  -3.541   0.564  1.00  0.00
ATOM    218  OG1 THR    21    -11.156  -0.694   1.961  1.00  0.00
ATOM    219  HG1 THR    21    -11.226  -0.253   1.111  1.00  0.00
ATOM    220  C   THR    21    -11.498  -3.255   3.600  1.00  0.00
ATOM    221  O   THR    21    -11.543  -4.484   3.528  1.00  0.00
ATOM    222  N   CYM    22    -12.547  -2.554   4.037  1.00  0.00
ATOM    223  H   CYM    22    -12.442  -1.569   4.236  1.00  0.00
ATOM    224  CA  CYM    22    -13.840  -3.185   4.314  1.00  0.00
ATOM    225  HA  CYM    22    -13.701  -4.025   4.994  1.00  0.00
ATOM    226  CB  CYM    22    -14.802  -2.184   4.962  1.00  0.00
```

```
ATOM    227  HB3 CYM    22     -15.826  -2.541   4.853  1.00  0.00
ATOM    228  HB2 CYM    22     -14.702  -1.215   4.474  1.00  0.00
ATOM    229  SG  CYM    22     -14.441  -2.066   6.733  1.00  0.00
ATOM    230  C   CYM    22     -14.455  -3.780   3.046  1.00  0.00
ATOM    231  O   CYM    22     -14.464  -3.154   1.993  1.00  0.00
TER
ATOM    232  H1  ACE    23     -16.783  -6.081   4.776  1.00  0.00
ATOM    233  CH3 ACE    23     -17.357  -6.407   3.908  1.00  0.00
ATOM    234  H2  ACE    23     -18.366  -6.676   4.219  1.00  0.00
ATOM    235  H3  ACE    23     -16.874  -7.274   3.457  1.00  0.00
ATOM    236  C   ACE    23     -17.447  -5.309   2.870  1.00  0.00
ATOM    237  O   ACE    23     -17.029  -4.147   2.777  1.00  0.00
ATOM    238  N   NME    24     -15.042  -5.126   3.091  1.00  0.00
ATOM    239  H   NME    24     -14.044  -5.042   3.225  1.00  0.00
ATOM    240  CH3 NME    24     -15.536  -5.415   1.755  1.00  0.00
ATOM    241 HH31 NME    24     -16.624  -5.475   1.774  1.00  0.00
ATOM    242 HH32 NME    24     -15.228  -4.622   1.074  1.00  0.00
ATOM    243 HH33 NME    24     -15.127  -6.366   1.413  1.00  0.00
TER
ATOM    244  N   TYR    25     -18.436  -5.723   2.152  1.00  0.00
ATOM    245  H   TYR    25     -18.129  -6.384   1.453  1.00  0.00
ATOM    246  CA  TYR    25     -19.767  -5.900   2.738  1.00  0.00
ATOM    247  HA  TYR    25     -19.693  -6.032   3.817  1.00  0.00
ATOM    248  CB  TYR    25     -20.615  -4.652   2.454  1.00  0.00
ATOM    249  HB2 TYR    25     -20.830  -4.688   1.386  1.00  0.00
ATOM    250  HB3 TYR    25     -21.545  -4.749   3.015  1.00  0.00
ATOM    251  CG  TYR    25     -19.957  -3.328   2.792  1.00  0.00
ATOM    252  CD1 TYR    25     -19.581  -3.050   4.120  1.00  0.00
ATOM    253  HD1 TYR    25     -19.715  -3.807   4.893  1.00  0.00
ATOM    254  CE1 TYR    25     -19.034  -1.798   4.450  1.00  0.00
ATOM    255  HE1 TYR    25     -18.683  -1.604   5.464  1.00  0.00
ATOM    256  CZ  TYR    25     -18.871  -0.822   3.450  1.00  0.00
ATOM    257  OH  TYR    25     -18.462   0.418   3.788  1.00  0.00
ATOM    258  HH  TYR    25     -18.313   0.520   4.731  1.00  0.00
ATOM    259  CE2 TYR    25     -19.203  -1.108   2.110  1.00  0.00
ATOM    260  HE2 TYR    25     -19.037  -0.365   1.329  1.00  0.00
ATOM    261  CD2 TYR    25     -19.752  -2.363   1.783  1.00  0.00
ATOM    262  HD2 TYR    25     -20.019  -2.589   0.751  1.00  0.00
ATOM    263  C   TYR    25     -20.462  -7.171   2.242  1.00  0.00
ATOM    264  O   TYR    25     -20.435  -7.441   1.043  1.00  0.00
ATOM    265  N   NME    26     -21.154  -8.058   3.189  1.00  0.00
ATOM    266  H   NME    26     -21.311  -8.053   4.187  1.00  0.00
ATOM    267  CH3 NME    26     -21.699  -9.158   2.412  1.00  0.00
ATOM    268 HH31 NME    26     -21.447  -9.021   1.361  1.00  0.00
```

```
ATOM    269 HH32 NME    26     -21.277 -10.098   2.767  1.00  0.00
ATOM    270 HH33 NME    26     -22.783  -9.181   2.525  1.00  0.00
TER
ATOM    271  H1  ACE    27     -22.562  -4.747   6.538  1.00  0.00
ATOM    272  CH3 ACE    27     -21.932  -5.486   7.034  1.00  0.00
ATOM    273  H2  ACE    27     -21.278  -5.958   6.300  1.00  0.00
ATOM    274  H3  ACE    27     -22.560  -6.244   7.501  1.00  0.00
ATOM    275  C   ACE    27     -21.075  -4.843   8.102  1.00  0.00
ATOM    276  O   ACE    27     -21.138  -3.626   8.325  1.00  0.00
ATOM    277  N   GLU    28     -20.266  -5.573   8.791  1.00  0.00
ATOM    278  H   GLU    28     -20.228  -6.573   8.655  1.00  0.00
ATOM    279  CA  GLU    28     -19.422  -4.981   9.832  1.00  0.00
ATOM    280  HA  GLU    28     -19.306  -3.909   9.672  1.00  0.00
ATOM    281  CB  GLU    28     -20.099  -5.205  11.201  1.00  0.00
ATOM    282  HB2 GLU    28     -20.241  -6.279  11.319  1.00  0.00
ATOM    283  HB3 GLU    28     -19.410  -4.845  11.965  1.00  0.00
ATOM    284  CG  GLU    28     -21.454  -4.488  11.353  1.00  0.00
ATOM    285  HG2 GLU    28     -22.088  -4.871  10.553  1.00  0.00
ATOM    286  HG3 GLU    28     -21.905  -4.722  12.317  1.00  0.00
ATOM    287  CD  GLU    28     -21.315  -2.966  11.210  1.00  0.00
ATOM    288  OE1 GLU    28     -22.077  -2.343  10.436  1.00  0.00
ATOM    289  OE2 GLU    28     -20.396  -2.392  11.839  1.00  0.00
ATOM    290  C   GLU    28     -18.001  -5.575   9.816  1.00  0.00
ATOM    291  O   GLU    28     -17.799  -6.793   9.712  1.00  0.00
ATOM    292  N   NME    29     -16.849  -4.668   9.926  1.00  0.00
ATOM    293  H   NME    29     -16.490  -3.911   9.362  1.00  0.00
ATOM    294  CH3 NME    29     -16.137  -5.017  11.142  1.00  0.00
ATOM    295 HH31 NME    29     -16.644  -5.846  11.636  1.00  0.00
ATOM    296 HH32 NME    29     -15.117  -5.311  10.895  1.00  0.00
ATOM    297 HH33 NME    29     -16.114  -4.156  11.810  1.00  0.00
TER
ATOM    298  H1  ACE    30     -16.843  -7.162   6.596  1.00  0.00
ATOM    299  CH3 ACE    30     -16.208  -7.888   6.089  1.00  0.00
ATOM    300  H2  ACE    30     -16.733  -8.288   5.221  1.00  0.00
ATOM    301  H3  ACE    30     -15.967  -8.701   6.774  1.00  0.00
ATOM    302  C   ACE    30     -14.918  -7.251   5.622  1.00  0.00
ATOM    303  O   ACE    30     -14.688  -6.051   5.828  1.00  0.00
ATOM    304  N   ARG    31     -14.043  -7.966   5.001  1.00  0.00
ATOM    305  H   ARG    31     -14.196  -8.953   4.852  1.00  0.00
ATOM    306  CA  ARG    31     -12.784  -7.380   4.535  1.00  0.00
ATOM    307  HA  ARG    31     -12.887  -6.298   4.454  1.00  0.00
ATOM    308  CB  ARG    31     -11.648  -7.694   5.525  1.00  0.00
ATOM    309  HB2 ARG    31     -11.596  -8.780   5.602  1.00  0.00
ATOM    310  HB3 ARG    31     -10.730  -7.318   5.073  1.00  0.00
```

```
ATOM    311  CG  ARG    31     -11.813  -7.087   6.933  1.00  0.00
ATOM    312  HG2 ARG    31     -12.778  -7.406   7.327  1.00  0.00
ATOM    313  HG3 ARG    31     -11.016  -7.479   7.565  1.00  0.00
ATOM    314  CD  ARG    31     -11.748  -5.552   6.920  1.00  0.00
ATOM    315  HD2 ARG    31     -10.784  -5.251   6.510  1.00  0.00
ATOM    316  HD3 ARG    31     -12.548  -5.178   6.281  1.00  0.00
ATOM    317  NE  ARG    31     -11.898  -4.956   8.263  1.00  0.00
ATOM    318  HE  ARG    31     -11.022  -4.679   8.683  1.00  0.00
ATOM    319  CZ  ARG    31     -13.014  -4.762   8.946  1.00  0.00
ATOM    320  NH1 ARG    31     -14.197  -5.104   8.501  1.00  0.00
ATOM    321 HH11 ARG    31     -14.286  -5.539   7.594  1.00  0.00
ATOM    322 HH12 ARG    31     -15.015  -4.931   9.067  1.00  0.00
ATOM    323  NH2 ARG    31     -12.974  -4.200  10.119  1.00  0.00
ATOM    324 HH21 ARG    31     -12.087  -3.912  10.507  1.00  0.00
ATOM    325 HH22 ARG    31     -13.830  -4.055  10.635  1.00  0.00
ATOM    326  C   ARG    31     -12.431  -7.903   3.151  1.00  0.00
ATOM    327  O   ARG    31     -12.534  -9.101   2.901  1.00  0.00
ATOM    328  N   NME    32     -11.968  -6.973   2.110  1.00  0.00
ATOM    329  H   NME    32     -11.805  -5.977   2.065  1.00  0.00
ATOM    330  CH3 NME    32     -11.717  -7.766   0.920  1.00  0.00
ATOM    331 HH31 NME    32     -11.938  -8.813   1.127  1.00  0.00
ATOM    332 HH32 NME    32     -12.353  -7.414   0.108  1.00  0.00
ATOM    333 HH33 NME    32     -10.671  -7.667   0.631  1.00  0.00
TER
ATOM    334  H1  ACE    33      -5.588  -6.730  18.954  1.00  0.00
ATOM    335  CH3 ACE    33      -5.643  -5.839  18.329  1.00  0.00
ATOM    336  H2  ACE    33      -4.790  -5.817  17.650  1.00  0.00
ATOM    337  H3  ACE    33      -5.627  -4.950  18.960  1.00  0.00
ATOM    338  C   ACE    33      -6.915  -5.827  17.508  1.00  0.00
ATOM    339  O   ACE    33      -7.737  -6.748  17.592  1.00  0.00
ATOM    340  N   ASP    34      -7.145  -4.840  16.710  1.00  0.00
ATOM    341  H   ASP    34      -6.524  -4.045  16.659  1.00  0.00
ATOM    342  CA  ASP    34      -8.369  -4.816  15.913  1.00  0.00
ATOM    343  HA  ASP    34      -8.756  -5.829  15.801  1.00  0.00
ATOM    344  CB  ASP    34      -9.454  -3.943  16.604  1.00  0.00
ATOM    345  HB2 ASP    34      -8.952  -2.999  16.817  1.00  0.00
ATOM    346  HB3 ASP    34     -10.259  -3.773  15.889  1.00  0.00
ATOM    347  CG  ASP    34     -10.036  -4.517  17.903  1.00  0.00
ATOM    348  OD1 ASP    34      -9.292  -4.651  18.896  1.00  0.00
ATOM    349  OD2 ASP    34     -11.232  -4.876  17.915  1.00  0.00
ATOM    350  C   ASP    34      -8.093  -4.327  14.511  1.00  0.00
ATOM    351  O   ASP    34      -7.180  -3.533  14.237  1.00  0.00
ATOM    352  N   ASN    35      -8.874  -4.835  13.548  1.00  0.00
ATOM    353  H   ASN    35      -9.523  -5.567  13.800  1.00  0.00
```

```
ATOM    354  CA  ASN    35     -8.797  -4.432  12.153  1.00  0.00
ATOM    355  HA  ASN    35     -7.782  -4.137  11.886  1.00  0.00
ATOM    356  CB  ASN    35     -9.198  -5.629  11.297  1.00  0.00
ATOM    357  HB2 ASN    35     -8.595  -6.464  11.652  1.00  0.00
ATOM    358  HB3 ASN    35    -10.255  -5.871  11.410  1.00  0.00
ATOM    359  CG  ASN    35     -8.906  -5.412   9.830  1.00  0.00
ATOM    360  OD1 ASN    35     -9.224  -4.409   9.213  1.00  0.00
ATOM    361  ND2 ASN    35     -8.344  -6.408   9.211  1.00  0.00
ATOM    362 HD21 ASN    35     -8.122  -6.328   8.229  1.00  0.00
ATOM    363 HD22 ASN    35     -8.132  -7.257   9.715  1.00  0.00
ATOM    364  C   ASN    35     -9.702  -3.210  11.913  1.00  0.00
ATOM    365  O   ASN    35    -10.925  -3.356  12.041  1.00  0.00
ATOM    366  N   PRO    36     -9.161  -2.048  11.512  1.00  0.00
ATOM    367  CD  PRO    36     -7.751  -1.774  11.279  1.00  0.00
ATOM    368  HD2 PRO    36     -7.482  -2.157  10.295  1.00  0.00
ATOM    369  HD3 PRO    36     -7.167  -2.290  12.041  1.00  0.00
ATOM    370  CG  PRO    36     -7.620  -0.264  11.373  1.00  0.00
ATOM    371  HG2 PRO    36     -6.799   0.091  10.750  1.00  0.00
ATOM    372  HG3 PRO    36     -7.453   0.045  12.405  1.00  0.00
ATOM    373  CB  PRO    36     -8.960   0.242  10.866  1.00  0.00
ATOM    374  HB2 PRO    36     -9.001   0.306   9.779  1.00  0.00
ATOM    375  HB3 PRO    36     -9.213   1.211  11.296  1.00  0.00
ATOM    376  CA  PRO    36     -9.941  -0.821  11.362  1.00  0.00
ATOM    377  HA  PRO    36    -10.299  -0.538  12.352  1.00  0.00
ATOM    378  C   PRO    36    -11.134  -0.968  10.411  1.00  0.00
ATOM    379  O   PRO    36    -11.055  -1.639   9.373  1.00  0.00
ATOM    380  N   LYS    37    -12.271  -0.369  10.788  1.00  0.00
ATOM    381  H   LYS    37    -12.303   0.027  11.717  1.00  0.00
ATOM    382  CA  LYS    37    -13.506  -0.376   9.991  1.00  0.00
ATOM    383  HA  LYS    37    -13.561  -1.306   9.426  1.00  0.00
ATOM    384  CB  LYS    37    -14.759  -0.264  10.875  1.00  0.00
ATOM    385  HB2 LYS    37    -14.742  -1.062  11.617  1.00  0.00
ATOM    386  HB3 LYS    37    -14.751   0.702  11.381  1.00  0.00
ATOM    387  CG  LYS    37    -16.037  -0.385  10.020  1.00  0.00
ATOM    388  HG2 LYS    37    -15.989   0.367   9.233  1.00  0.00
ATOM    389  HG3 LYS    37    -16.051  -1.379   9.572  1.00  0.00
ATOM    390  CD  LYS    37    -17.312  -0.179  10.846  1.00  0.00
ATOM    391  HD2 LYS    37    -17.349  -0.926  11.638  1.00  0.00
ATOM    392  HD3 LYS    37    -17.294   0.817  11.288  1.00  0.00
ATOM    393  CE  LYS    37    -18.551  -0.320   9.953  1.00  0.00
ATOM    394  HE2 LYS    37    -18.473   0.405   9.142  1.00  0.00
ATOM    395  HE3 LYS    37    -18.562  -1.328   9.539  1.00  0.00
ATOM    396  NZ  LYS    37    -19.801  -0.084  10.706  1.00  0.00
ATOM    397  HZ1 LYS    37    -19.792   0.850  11.090  1.00  0.00
```

ATOM 398 HZ2 LYS 37 -20.592 -0.185 10.086 1.00 0.00
ATOM 399 HZ3 LYS 37 -19.874 -0.756 11.457 1.00 0.00
ATOM 400 C LYS 37 -13.417 0.709 8.916 1.00 0.00
ATOM 401 O LYS 37 -14.002 1.781 8.993 1.00 0.00
ATOM 402 N THR 38 -12.653 0.403 7.884 1.00 0.00
ATOM 403 H THR 38 -12.199 -0.498 7.933 1.00 0.00
ATOM 404 CA THR 38 -12.311 1.297 6.767 1.00 0.00
ATOM 405 HA THR 38 -12.274 2.333 7.104 1.00 0.00
ATOM 406 CB THR 38 -10.931 0.904 6.246 1.00 0.00
ATOM 407 HB THR 38 -10.673 1.591 5.440 1.00 0.00
ATOM 408 CG2 THR 38 -9.854 0.953 7.317 1.00 0.00
ATOM 409 HG21 THR 38 -10.111 0.265 8.123 1.00 0.00
ATOM 410 HG22 THR 38 -8.897 0.663 6.884 1.00 0.00
ATOM 411 HG23 THR 38 -9.781 1.966 7.714 1.00 0.00
ATOM 412 OG1 THR 38 -10.946 -0.417 5.747 1.00 0.00
ATOM 413 HG1 THR 38 -10.072 -0.647 5.424 1.00 0.00
ATOM 414 C THR 38 -13.415 1.308 5.695 1.00 0.00
ATOM 415 O THR 38 -13.187 1.220 4.493 1.00 0.00
ATOM 416 N GLY 39 -14.675 1.417 6.115 1.00 0.00
ATOM 417 H GLY 39 -14.826 1.605 7.096 1.00 0.00
ATOM 418 CA GLY 39 -15.842 1.282 5.249 1.00 0.00
ATOM 419 HA2 GLY 39 -15.521 0.830 4.311 1.00 0.00
ATOM 420 HA3 GLY 39 -16.559 0.625 5.742 1.00 0.00
ATOM 421 C GLY 39 -16.518 2.622 4.950 1.00 0.00
ATOM 422 O GLY 39 -16.502 3.523 5.775 1.00 0.00
ATOM 423 N ASP 40 -17.118 2.776 3.769 1.00 0.00
ATOM 424 H ASP 40 -17.008 2.062 3.063 1.00 0.00
ATOM 425 CA ASP 40 -17.840 4.006 3.411 1.00 0.00
ATOM 426 HA ASP 40 -17.480 4.848 4.002 1.00 0.00
ATOM 427 CB ASP 40 -17.599 4.307 1.939 1.00 0.00
ATOM 428 HB2 ASP 40 -16.519 4.275 1.792 1.00 0.00
ATOM 429 HB3 ASP 40 -18.064 3.487 1.392 1.00 0.00
ATOM 430 CG ASP 40 -18.144 5.638 1.421 1.00 0.00
ATOM 431 OD1 ASP 40 -18.589 6.507 2.201 1.00 0.00
ATOM 432 OD2 ASP 40 -18.102 5.790 0.182 1.00 0.00
ATOM 433 C ASP 40 -19.332 3.907 3.772 1.00 0.00
ATOM 434 O ASP 40 -20.114 3.180 3.140 1.00 0.00
ATOM 435 N GLU 41 -19.736 4.655 4.813 1.00 0.00
ATOM 436 H GLU 41 -19.088 5.311 5.225 1.00 0.00
ATOM 437 CA GLU 41 -21.063 4.524 5.410 1.00 0.00
ATOM 438 HA GLU 41 -21.452 3.519 5.244 1.00 0.00
ATOM 439 CB GLU 41 -20.978 4.778 6.943 1.00 0.00
ATOM 440 HB2 GLU 41 -20.135 4.189 7.304 1.00 0.00
ATOM 441 HB3 GLU 41 -20.758 5.839 7.066 1.00 0.00

```
ATOM   442  CG  GLU   41    -22.258   4.413   7.763  1.00  0.00
ATOM   443  HG2 GLU   41    -22.661   3.519   7.288  1.00  0.00
ATOM   444  HG3 GLU   41    -21.881   4.159   8.754  1.00  0.00
ATOM   445  CD  GLU   41    -23.353   5.455   7.890  1.00  0.00
ATOM   446  OE1 GLU   41    -23.141   6.647   7.597  1.00  0.00
ATOM   447  OE2 GLU   41    -24.467   5.049   8.287  1.00  0.00
ATOM   448  C   GLU   41    -22.040   5.481   4.771  1.00  0.00
ATOM   449  O   GLU   41    -21.712   6.692   4.559  1.00  0.00
ATOM   450  N   NME   42    -23.216   5.069   4.436  1.00  0.00
ATOM   451  H   NME   42    -23.507   4.114   4.590  1.00  0.00
ATOM   452  CH3 NME   42    -24.160   5.987   3.820  1.00  0.00
ATOM   453 HH31 NME   42    -23.700   6.970   3.722  1.00  0.00
ATOM   454 HH32 NME   42    -24.437   5.616   2.833  1.00  0.00
ATOM   455 HH33 NME   42    -25.052   6.064   4.442  1.00  0.00
TER
END
```

Below is the input for CP2K that we used:

```
&GLOBAL
  PROJECT MONITOR              ! Name of the calculation
  PRINT_LEVEL LOW              ! Verbosity of the output
  RUN_TYPE MD                  ! Calculation type: MD
&END GLOBAL

&FORCE_EVAL                    ! parameters needed to calculate energy and forces
  METHOD QMMM                  ! Hybrid quantum classical
  STRESS_TENSOR ANALYTICAL     ! Compute the stress tensor analytically (if available).
   &DFT
  CHARGE  1 #Net charge
  MULTIPLICITY    1 #Spin multiplicity
    &QS
     METHOD PM6
     &SE
       &COULOMB
         CUTOFF [angstrom] 10.0
       &END
       &EXCHANGE
         CUTOFF [angstrom] 10.0
       &END
     &END
     &END QS
    &POISSON
```

```
    &EWALD
      EWALD_TYPE SPME
      GMAX 80
      O_SPLINE 6
    &END EWALD
   &END POISSON
   &SCF
    SCF_GUESS ATOMIC
    EPS_SCF 1.0E-5
    MAX_SCF 50
    &OUTER_SCF
       EPS_SCF 1.0E-7
       MAX_SCF 500
    &END
   &END SCF
  &END DFT


  &MM                     ! Parameters to run a MM calculation
   &FORCEFIELD              ! Set up a force_field for the classical calculations
    PARMTYPE AMBER          ! Kind of torsion potential
    ! Filename that contains the parameters of the FF
    PARM_FILE_NAME system_qm-charge.parm7
    &SPLINE               ! Parameters to set up the splines used in the nonboned interactions
     EMAX_SPLINE 1.0E8       ! Maximum value of the potential up to which splines will be
constructed
     RCUT_NB [angstrom] 10    ! Cutoff radius for nonbonded interactions
    &END SPLINE
   &END FORCEFIELD
   &POISSON
    &EWALD
    ! Ewald parameters controlling electrostatic
     EWALD_TYPE SPME        ! Type of ewald
     ALPHA .40            ! Alpha parameter associated with Ewald (EWALD|PME|SPME)
     GMAX 80              ! Number of grid points (SPME and EWALD)
    &END EWALD
   &END POISSON
  &END MM

  &SUBSYS                  ! a subsystem: coordinates, topology, molecules and cell
   &CELL                  !Set box dimensions here
    ABC [angstrom] 300.0000000  300.0000000  300.0000000
    ALPHA_BETA_GAMMA 90.0 90.0 90.0
   &END CELL
```

```
   &TOPOLOGY                ! Topology for classical runs
     CONN_FILE_FORMAT AMBER
     CONN_FILE_NAME system_qm-charge.parm7
   &END TOPOLOGY
   !NA+ is not recognized by CP2K, so it is necessary to define it here using KIND
   &KIND NA+
    ELEMENT Na
   &END KIND
   &KIND N
    ELEMENT N
   &END KIND
   &KIND CA2+
    ELEMENT Ca
   &END KIND
   &KIND CL-
    ELEMENT Cl
   &END KIND
   &KIND NS4
    ELEMENT N
   &END KIND
   &KIND NS1
    ELEMENT N
   &END KIND
   &KIND NS2
    ELEMENT N
   &END KIND
  &END SUBSYS
  &QMMM                     ! Input for QM/MM calculations
   ECOUPL COULOMB              ! type of the QM - MM electrostatic coupling
   &CELL                   ! Set box dimensions here
     ABC 90 90 90
     ALPHA_BETA_GAMMA 90 90 90
   &END CELL
   @INCLUDE './QMresidue-index.inc'
 &END QMMM
&END FORCE_EVAL

&MOTION                     ! Parameter for the motion of the nuclei
  &MD                     ! set of parameters needed perform an MD run
  ENSEMBLE NVT              ! Ensemble/integrator that you want to use for MD
  TIMESTEP [fs] 0.5         ! Time step
  STEPS    5000000          ! Number of MD steps to perform
  TEMPERATURE 298           ! Temperature in K
  &THERMOSTAT               ! Parameters of Thermostat.
```

```
   REGION GLOBAL              ! region each thermostat is attached to.
   TYPE CSVR                  ! canonical sampling through velocity rescaling
   &CSVR
    TIMECON [fs] 100.
   &END CSVR
  &END THERMOSTAT
  COMVEL_TOL 1e-8             ! remove COM motion
  &END MD
  &CONSTRAINT
   &FIXED_ATOMS
   LIST 2 5 7 9 19 21 23 28 31 33 35 50 52 54 59 62 64 66 86 88 90 95 98 100 102 116 118 120
125 128 130 132 150 152 154 159 162 164 166 183 185 187 192 195 197 199 206 208 210 220
222 224 230 238 240 233 236  244 246 263 265 272 275 277 279 290 292 294 299 302 304 306
326 328 330 335 338 450 452        !Tells cp2k what atoms to fix
   COMPONENTS_TO_FIX XYZ                  !tells cp2k what cordinates to fix i.e. (X,Y,Z)
  &END FIXED_ATOMS
  &END CONSTRAINT
  &FREE_ENERGY
    &METADYN
     USE_PLUMED .TRUE.
     PLUMED_INPUT_FILE ./plumed.dat
    &END METADYN
  &END FREE_ENERGY
  &PRINT                     ! Printing properties during an MD run
   &RESTART                  ! Printing of restart files
    &EACH                    ! A restart file will be printed every 25000 md steps
      MD 10000
    &END
   &END
   &TRAJECTORY               ! Controls the output of the trajectory
    FORMAT XYZ               ! Format of the output trajectory is DCD
    &EACH                    ! New trajectory frame will be printed each 5000 md steps
      MD 100
    &END
   &END
   &RESTART_HISTORY          ! Controls printing of unique restart files during the run keeping
all of them.
    &EACH                    ! A new restart file will be printed every 5000 md steps
      MD 10000
    &END
   &END
  &END PRINT
&END MOTION
```

```
&EXT_RESTART
  RESTART_FILE_NAME NPT-1.restart
  RESTART_COUNTERS .FALSE.
  RESTART_THERMOSTAT .FALSE.

&END
```

Below is the PLUMED file for upper and lower bounds and metadynamic parameters:

```
#RESTART
UNITS LENGTH=A TIME=0.1  #Amstroeng, hartree, fs

energy: ENERGY

        # (***)=numbers that have been changed to reflect the new structure.

d1: DISTANCE ATOMS=414,229 NOPBC #Thr C to Cys S ***
d2: DISTANCE ATOMS=414,416 NOPBC #Thr C to Gly N peptide bond ***
d3: DISTANCE ATOMS=421,418 NOPBC #Gly Alpha C and C restraint ***
d4: DISTANCE ATOMS=108,109 NOPBC #His nitrogen to His H ***
d5: DISTANCE ATOMS=104,105 NOPBC #His beta C to H1 ***
d6: DISTANCE ATOMS=104,106 NOPBC #His beta C to H2 ***
d7: DISTANCE ATOMS=108,229 NOPBC  #His to Cys distance ***
d8: DISTANCE ATOMS=416,417 NOPBC #Gly N to Gly H distance ***
d9: DISTANCE ATOMS=112,113 NOPBC #His N2 to His H2 ***
d10: DISTANCE ATOMS=412,406 NOPBC #Thr sidechain Oxy to Beta C ***
d11: DISTANCE ATOMS=108,109 NOPBC  #His to H+ distance ***
d12: DISTANCE ATOMS=229,109 NOPBC  #Cys and H+ distance ***
d13: DISTANCE ATOMS=416,109 NOPBC  #Gly peptide nitrogen to His H distance ***
d14: DISTANCE ATOMS=414,415 NOPBC #Thr C to Thr O bond (used for moving restraint)***
d15: DISTANCE ATOMS=414,404 NOPBC #Thr C to Thr CA bond (used for moving restraint)***
d16: DISTANCE ATOMS=415,229 NOPBC  #Thr backbone Oxy to Cys distance ***
d17: DISTANCE ATOMS=108,414 NOPBC  #His to Thr C distance ***
d18: DISTANCE ATOMS=109,415 NOPBC #His H (catalysis) to Thr O bond (used for moving
restraint)***
d19: DISTANCE ATOMS=418,420 NOPBC #Gly alpha carbon to H1 ***
d20: DISTANCE ATOMS=418,419 NOPBC #Gly alpha carbon to H1 ***
d21: DISTANCE ATOMS=222,223 NOPBC #Ser 232 backbone nitrogen and H1 restraint ***

#comb1:  COMBINE ARG=d1,d2,d13,d4 PERIODIC=NO COEFFICIENTS=-1,1,-1,1              ###
HDLA COEFFICIENTS are d1=-0.018, d2=0.989, d13=-0.144
comb1:  COMBINE ARG=d4,d13 PERIODIC=NO COEFFICIENTS=1,-1
comb2:  COMBINE ARG=d1,d2 PERIODIC=NO COEFFICIENTS=1,-1
```

```
c1: COORDINATION GROUPA=414 GROUPB=415,229,416,404 R_0=2.0

#MOVINGRESTRAINT ...
#  ARG=d13
#  STEP0=0 AT0=2.5 KAPPA0=0
#  STEP1=20000 AT1=1.5 KAPPA1=600.0
#  STEP2=30000 AT2=1.5 KAPPA2=600.0
#  STEP3=50000 AT3=1.5 KAPPA3=0.0
#  LABEL=restraint
#...
#MOVINGRESTRAINT ...
#  ARG=d1
#  STEP0=0 AT0=4.3 KAPPA0=0.0
#  STEP1=20000 AT1=1.5 KAPPA1=600.0
#  STEP2=30000 AT2=1.5 KAPPA2=600.0
#  STEP3=50000 AT3=1.5 KAPPA3=0.0
#  LABEL=restraint2
#...
#MOVINGRESTRAINT ...
#  ARG=d2
#  STEP0=0 AT0=2.0 KAPPA0=300.0
#  STEP1=20000 AT1=5.0 KAPPA1=300.0
#  STEP2=30000 AT2=10.0 KAPPA2=300.0
#  STEP3=50000 AT3=10.0 KAPPA3=0.0
#  LABEL=restraint3
#...

UPPER_WALLS ARG=d1 AT=3.0 KAPPA=600.0 EXP=2 LABEL=uwall_5
UPPER_WALLS ARG=d14 AT=2.5 KAPPA=300.0 EXP=2 LABEL=uwall_6
UPPER_WALLS ARG=d11 AT=5.0 KAPPA=300.0 EXP=2 LABEL=uwall_7
UPPER_WALLS ARG=d2 AT=5.0 KAPPA=600.0 EXP=2 LABEL=uwall_8
UPPER_WALLS ARG=d15 AT=2.5 KAPPA=300.0 EXP=2 LABEL=uwall_9
UPPER_WALLS ARG=d13 AT=5.0 KAPPA=50.0 EXP=2 LABEL=uwall_13
UPPER_WALLS ARG=d3 AT=1.8 KAPPA=600.0 EXP=2 LABEL=uwall_11
UPPER_WALLS ARG=d8 AT=1.7 KAPPA=600.0 EXP=2 LABEL=uwall_12
UPPER_WALLS ARG=d9 AT=1.7 KAPPA=600.0 EXP=2 LABEL=uwall_14
UPPER_WALLS ARG=d5 AT=1.7 KAPPA=600.0 EXP=2 LABEL=uwall_15
UPPER_WALLS ARG=d6 AT=1.7 KAPPA=600.0 EXP=2 LABEL=uwall_16
UPPER_WALLS ARG=d10 AT=2.0 KAPPA=150.0 EXP=2 LABEL=uwall_17
UPPER_WALLS ARG=d19 AT=1.5 KAPPA=600.0 EXP=2 LABEL=uwall_19
UPPER_WALLS ARG=d20 AT=1.5 KAPPA=600.0 EXP=2 LABEL=uwall_20
UPPER_WALLS ARG=d21 AT=1.5 KAPPA=600.0 EXP=2 LABEL=uwall_21
```

```
LOWER_WALLS ARG=d16 AT=3.00 KAPPA=600.0 EXP=2 LABEL=lwall_1
LOWER_WALLS ARG=d17 AT=2.5 KAPPA=600.0 EXP=2 LABEL=lwall_2
LOWER_WALLS ARG=d18 AT=1.75 KAPPA=600.0 EXP=2 LABEL=lwall_3

#UPPER_WALLS ARG=d1 AT=+5.0 KAPPA=150.0  EXP=2 LABEL=uwall_1
#UPPER_WALLS ARG=d2 AT=+5.0 KAPPA=150.0  EXP=2 LABEL=uwall_2

#metad: METAD ARG=d1 PACE=100 HEIGHT=5.0 SIGMA=0.5 GRID_MIN=-100 GRID_MAX=100
GRID_BIN=200 FILE=HILLS BIASFACTOR=80 TEMP=300.0 ADAPTIVE=GEOM SIGMA_MIN=0.02
SIGMA_MAX=2.0
#metad: METAD ARG=comb1,d1 PACE=100 HEIGHT=5.0 SIGMA=0.5 GRID_MIN=-100,-100
GRID_MAX=100,100 GRID_BIN=200,200 FILE=HILLS BIASFACTOR=80 TEMP=300.0
ADAPTIVE=GEOM SIGMA_MIN=0.02,0.02 SIGMA_MAX=2.0,2.0
metad: METAD ARG=comb1,d1,d2 PACE=100 HEIGHT=5.0 SIGMA=0.5 GRID_MIN=-100,-100,-
100 GRID_MAX=100,100,100 GRID_BIN=200,200,200 FILE=HILLS_3CV BIASFACTOR=80
TEMP=300.0 ADAPTIVE=GEOM SIGMA_MIN=0.02,0.02,0.02 SIGMA_MAX=2.0,2.0,2.0

FLUSH STRIDE=1
PRINT ...

ARG=d1,d2,d3,d4,d5,d6,d7,d8,d9,d10,d11,d12,d13,d14,d15,d16,d17,c1,comb1,comb2,metad.*,
uwall_5.*,lwall_1.*,uwall_13.*,lwall_2.*#restraint.bias,#restraint2.bias,#restraint3.bias,#metad
.*
 STRIDE=1
 FILE=COLVAR_3CV
... PRINT
```