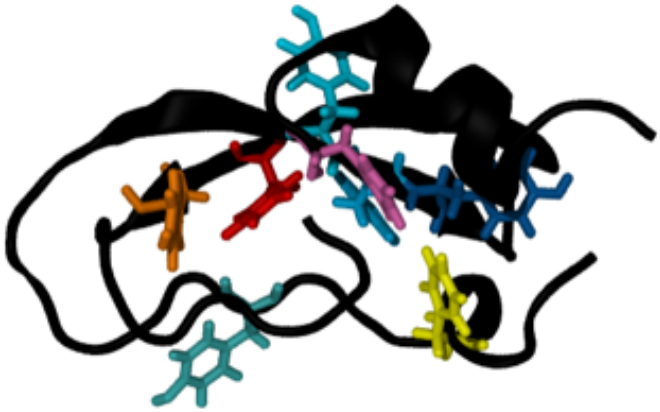
Aromatic Ring Flips and Local Volume Fluctuations in BPTI



By: Simon Enbom
Co-supervisor: Yulian Gavrilov
Supervisor: Pär Söderhjelm
Examiner: Mikael Akke

Division of Biophysical Chemistry
Department of Chemistry
Lund University

2nd January 2022

Acknowledgements

I want to begin by thanking my main supervisor Pär: There was really a lot that I had to learn in order to get started with the project, and you were generous with your time and made sure that I progressed. When I began to have my own ideas, you encouraged me and gave advice on relevant methods, and you valued my input not based on my experience but whether or not I could motivate ideas with reason.

I would also like to thank Yulian, who co-supervised me with Pär. You too were generous with your time and were always very quick in helping me, I unfortunately was not always as swift. I also want to thank you for our discussions and your advice on everything.

I had a great time collaborating with you and thank you both, Pär and Yulian, for this time.

I would also like to thank my examiner Mikael Akke for reviewing my work and I look forward to evaluate the report with you.

One of my teachers, Kristofer Modig, deserves thanks too. It was his courses in physical and later biophysical chemistry which convinced me that I should do my thesis in this field.

Thank you Sandra Arvidsson and Albin Olsson for choosing to oppose my work. I hope you will find it interesting.

Abstract

Aromatic ring flips are a classic example of a dynamic process in proteins. Discovered decades ago, methodological limitations hindered progress at the time. Ring flips have gained new interest and experimentalists can measure the flip rate and thermodynamic quantities associated with it. Molecular Dynamics (MD) simulations provide atomic resolution data and can complement the experimental work since it allows us to 1) compare theoretical and experimental values, 2) distinguish between local and global activation volumes and 3) elucidate the ring flip mechanism. Ring flips are slow processes and an enhanced sampling technique is necessary to sample them in reasonable computational times. Metadynamics, an enhanced sampling technique, works by applying a bias potential which "fills up" energy minima for a system coordinate, often called Collective Variable (CV) and while the dynamics are inherently non-physical, rare events can be sampled, thus creating extended ensembles. I have calculated the local activation volumes for buried residues in Bovine Pancreatic Trypsin Inhibitor (BPTI) in metadynamics trajectories and found qualitative agreement with experimental studies, showing that there is a positive activation volume associated with flips. I have also calculated the compressiblity of the transition state $^{\ddagger}\kappa$=5.40*10^{-4} bar $^{-1}$, developed an indirect CV that can force ring flips if biased and showed that solvation may be associated with flip kinetics.

Populärvetenskaplig Sammanfattning

Protein är inte bara “det man ska äta mycket av om man är fysiskt aktiv”. Det är en klass av molekyler som spelar en roll i alla biologiska funktioner som vi känner till. Allt från muskelaktivitet till igenkännandet av signalsubstanser i hjärnan. Proteiner är tillsammans med DNA och RNA livets hörnstenar och att förstå dem skulle drastiskt förbättra förutsättningarna för att utveckla säkra och effektiva läkemedel.

Proteiner är stora, långa molekyler bestående av hoppusslade kemiska enheter som kallas aminosyror. De får sin funktion i stor utsträckning från sin tredimensionella struktur och hur den ändrar sig på molekylär nivå. På detta sätt kan man likna dem vid olika typer av mänskligt skapade maskiner, men på molekylär skala är spelreglerna annorlunda. I allmänhet har vi god förståelse för de principer och lagar som gäller för mänskligt skapade maskiner, men så är inte fallet för protein.

Ett mål inom proteinforskningen är därför att förstå proteinernas rörelser, deras dynamik. Proteiners dynamik är dels associerad med dess funktion, exempelvis enzymatisk katalys, men det finns också rörelser som saknar direkt koppling till den utförda funktionen, vilka kan vara svåra att rationalisera.

I detta arbete har jag med hjälp av simuleringsverktyg försökt att förstå en känd dynamisk process som kallas aromatisk ringvändning, eller ”aromatic ring flip” på engelska. Dessa aromatiska ringar är ofta tätt packade i proteinernas mitt och för att de ska kunna vända sig krävs att utrymme skapas kring dem. Detta kräver energi och kan jämföras med hur en dörr blir tung att öppna om man ställer något i vägen för den. I projektet undersöks hur mycket volym som behöver skapas och hur denna är tryckberoende.

Contents

1 Introduction

1.1 Protein dynamics and its significance

Protein dynamics, structural motions in proteins, are essential to protein function. Protein functionality in turn spans every cellular function and includes drug-receptor interactions. In computational high-throughput screening of drug candidates, thousands of small molecules are screened against a drug target in different software. They calculate the binding free energy of a drug-receptor interaction by docking, putting a small molecule in the binding site of the target and then calculating the sum of the interactions.

This approach overlooks slow structural fluctuations in proteins, as the required simulation times for studying dynamic properties are too long for it to be a viable strategy [1]. A quantitative understanding of protein dynamics can play a role in improving computational drug discovery where the most important dynamic processes could be approximated without requiring long simulation times, as well as in fields such as protein engineering and enzyme technology.

There are many types of protein dynamics. There is global protein dynamics which concerns the rotational and translational dynamics of the protein. There is also conformational protein dynamics, which concerns protein motions that changes the distance between atoms. When it comes to the relation between protein dynamics and function, conformational dynamics is the relevant type of dynamics. Conformational dynamics are sometimes conceptualized with a two-step model in mind. There, initial and final states are defined for a motion along a reaction coordinate, which the system interchanges between at well-defined rates.

$$\mathrm{A} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_1}{\rightleftharpoons}} \mathrm{B} \tag{1.1}$$

Where k_1 and k_{-1} are the rate constants of the reaction in left and right direction and A and B are the generic states, defined based on the modeled state. [2] As an example, think of blinking; there, the eyelids interchange between being open and closed at certain rates. The rate for opening them is faster than the rate of closing them when we are awake, as our eyes are mostly open.

Protein dynamic properties are difficult to study. Experimental methods, such as NMR spectroscopy, give valuable information, but the experiments can be challenging to design and there is a limit to how many processes can be sampled at once. Computational methods on the other hand can sample all atom motions, but they can never prove something without experimental data. Therefore, combinations of these techniques are often employed, but even then, it can be challenging.

Understanding protein dynamics is a monumental task due to the methodological

challenges and its great complexity. If it can be done however, it would enable great technological advances, and understanding even a single process is worthwhile. The process that will be studied in this report is aromatic ring flips.

1.2 Aromatic Ring Flips

Aromatic ring flips in proteins is a classic phenomenon in protein dynamics. In the 1970's, NMR studies on BPTI showed rings flipping, [3] which was supported by the first Molecular Dynamics (MD) simulation of a protein, also BPTI [4]. Methodological limitations hindered advances in aromatic ring flips, but researchers at Lund University have co-developed new NMR experiments which can determine the thermodynamics of the flip according to transition state theory [5] and have revised the kinetics for some well-studied rings [6].

Protein side-chains can be hydrophilic or hydrophobic. Side-chain polarity is important in protein folding, as the hydrophobic effect will concentrate apolar residues with other apolar residues in what is known as the protein hydrophobic core, whereas hydrophilic residues will be solvated [7]. Phenylalanine (Phe, F) and tyrosine (Tyr, Y) have aromatic rings and are often times buried in the hydrophobic core, but can also be solvated.

An aromatic ring flip is defined as the 180 degrees rotation along the χ_2 axis, from a main state (MS) into a different but structurally identical other main state. The two-state jump model can be an approriate tool to visualize this, with state A and B being identical but different. The transition state (TS) is defined as the 90 degree rotation along same axis, meaning that it is orthogonal to the main state. Figure 1.1 shows the χ_2 dihedral along with the χ_1, ϕ and ψ, which will also be referred to in the report.

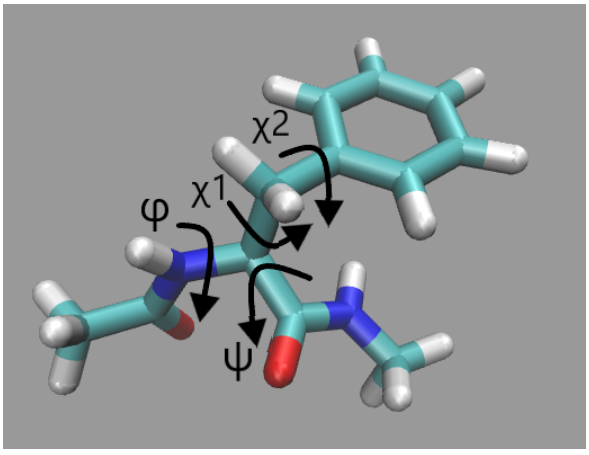


Figure 1.1: A phenylalanine dipeptide (two peptide/amide bonds), with the relevant dihedral angles marked. A 180 degrees rotation around χ_2 implies a ring flip.

Note that for many residues, there are also intermediate states, stable states where χ_1 rotates. What initially may seem like a trivial problem is shown to be a complex energy landscape, often with many intermediate and transition states [8]. Figure 1.2 shows the difference between rotating around the different axes. The reader can visualize that the phenylalanine of this example is buried inside a pocket, and that different conformations implicates different local environments for the ring.

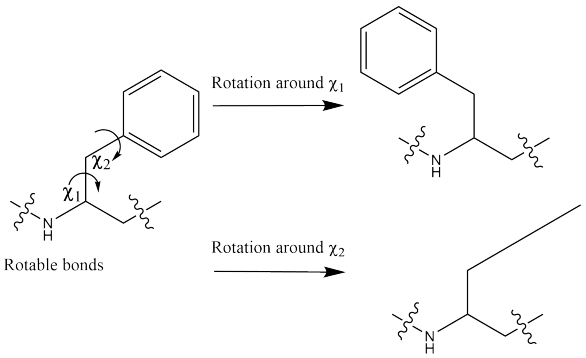


Figure 1.2: A schematic view of the rotations around χ_1 and χ_2. The rotation around χ_1 is 180 degrees and the rotation around χ_2 is 90 degrees, for illustrative purposes.

1.2.1 Aromatic ring flips and local volume fluctuations

In terms of other biophysical processes, ring flips are slow, occurring on millisecond-second timescales. This is explained by steric hindrances as a result of being packed in the protein interior. There is however some vacuum between the ring and its neighbouring atoms, called packing defects. In order for the flip to occur, the packing defects has to enlarge in order for the steric hindrances to be removed [9]. The correlation between the flip and the enlargement of packing defects is not fully understood; does the ring rotate, pushing the protein matrix aside as to increase packing defects, or does the packing defects themselves occasionally increase, allowing the flip to occur?

In thermodynamic terms, the quantity describing the packing defects is volume. Further, the flip can be understood in terms of transition state theory (formal description below), where a main state has a certain volume and the transition state is assumed to have a larger volume, due to the space having been created for the ring to rotate into a new symmetrical main state, which has the same volume as the first main state. The required increase in volume around the ring will be referred to as the local activation volume. This expansion is related to the free energy of activation in that a certain pressure-volume work is required to expand the volume around the ring. Note that it neither refers to the vacuum nor the ring alone but together, as vacuum is infinitely compressible and an atom incompressible.

There is experimental evidence for a global activation volume [5], where the entire protein expands in order to create the space required for the ring to flip. This is based on a measured free energy barrier increase with pressure, which is explained by

the theory in the next section. This is called the expansion mechanism. Under high pressure however, the contribution from an increase in global protein volume to the free energy barrier can be 0, but the flip still occurs. This indicates that the space can also be created around the ring by compacting internal protein voids, which is called the compensation mechanism.

1.3 The Thermodynamics of Aromatic Ring Flips

This section is dedicated to establishing the theoretical connection between the rate of aromatic ring flips and the activation volume. As previously mentioned, aromatic ring flips can be understood in the framework of transition state theory. The rate of the flip is related to the free energy of activation through the Eyring equation

$$k_{flip} = \frac{k_B T}{h} * e^{-\frac{\Delta^\ddagger G}{RT}} \tag{1.2}$$

where k_{flip} is the rate constant, k_B is Boltzmann's constant, T is temperature, h is Planck's constant, $\Delta^\ddagger G$ is the free energy of activation and R is the gas constant. In order to understand the relation between the flip rate and the volume, the free energy of activation is derived from a thermodynamic fundamental equation. The free energy of activation has a pressure and temperature dependency

$$d\Delta^\ddagger G = -\Delta^\ddagger S dT + \Delta^\ddagger V dp \tag{1.3}$$

Where p is pressure, $\Delta^\ddagger S$ is the entropy of activation and $\Delta^\ddagger V$ is the activation volume, the aforementioned required volumetric increase in packing defects necessary for the ring to flip. In this project, a thermostat was employed and as such we assign $dT = 0$ in order to simplify the expression and derivation. The derivation of the pressure and temperature dependence is given in [5]. The following equation which describes the pressure-dependence of a volume is integrated over pressure

$$dV = \alpha V dT - V \kappa dp \tag{1.4}$$

Where α is the expansivity coefficient and κ is the compressibility coefficient. Here, the volume compressibility is defined as $\kappa' = \kappa V$. We first reach the activation volume pressure dependence

$$\Delta^\ddagger V = \Delta^\ddagger V_0 - \Delta^\ddagger \kappa' (p - p_0) \tag{1.5}$$

Where V_0 is the volume at a reference pressure. By inserting the above equation into equation (1.3), the following expression of the Gibbs free energy of activation dependence on pressure is reached.

$$\Delta^\ddagger G = \Delta^\ddagger G_0 + \Delta^\ddagger V_0 (p - p_0) - \Delta^\ddagger \kappa' \frac{(p - p_0)^2}{2} \tag{1.6}$$

Equation (1.6) shows the volume- and pressure-dependence of the free energy of activation under isothermal conditions. It demonstrates that volume and compressibility are both related to the free energy of activation and as such to the flip rate.

For the discussion, the following expressions will be used, for the purpose of calculating the transition state compressibility, ${}^{\ddagger}\kappa$

$$\Delta^{\ddagger}\kappa' \simeq \Delta^{\ddagger}\kappa V_0 + \Delta^{\ddagger}V_0\kappa \tag{1.7}$$

$${}^{\ddagger}\kappa = \kappa + \Delta^{\ddagger}\kappa \tag{1.8}$$

1.4 Molecular Dynamics Simulations of Proteins

Molecular dynamics (MD) is a common simulation method for atomic resolution study of protein dynamics. Protein structures are determined experimentally, mainly through X-ray diffraction, NMR spectroscopy and cryo-EM and accessed through the Protein Data Bank (PDB). The structures are imported and the energy of bonded and non-bonded interactions (potential energy) for every atom is calculated. By differentiating over time and solving Newton's second law of motion on the acting forces on every atom, their motions over a time step can be calculated and a trajectory is obtained containing the changes in spatial coordinates over time. MD simulations can thus give dynamic data on every atom in the system, which makes it suited to complement experimental work, as the NMR experiments are more reliable but does not provide data on the dynamics of every atom.

Simulations performed need to show ergodicity which means that all accessible states must be visited. It is of course impossible to visit every state, but it is important to show that the simulations would have sampled every minima. Otherwise, there is the risk that the system is stuck in a local energy minimum, without ever visiting the global minimum. This puts a demand on computational times being long enough to visit all accessible states. MD simulations are computationally heavy, and running microsecond simulations can take weeks. Experimentally determined flip kinetics show that aromatic ring flips occur on many order of magnitudes greater time and thus that 1 μs simulations might not show any flips [6], making it unlikely that the simulations will be ergodic.

1.4.1 Metadynamics

A solution to the problem of not reaching ergodicity, of not sampling all states accessible to the system, is by employing an enhanced sampling technique. One of these is Metadynamics (MetaD), a powerful biasing algorithm. [10] Through MetaD, by defining Collective Variables (CVs) and applying a bias potential, the system is forced out of energy minima. This is analogous to filling a hole so to flatten a road. Where once a car tire could get stuck, it is now flat and the car drives over it with no problem. Metadynamics fills up holes in the energy landscape by depositing a well-defined

synthetic potential as a gaussian, the same mathematical structure that describes normal distributions. It is unphysical, but can be useful when there are problems with sampling. Sampling might seem like a strange word when referring to computer simulations, but it is impossible to calculate quantities for a structure which you do not have, and the simulation has to visit it for you to have it. MetaD has been shown to enable ring flips by biasing χ_1 and χ_2 rotations in [8].

1.5 BPTI

Bovine Pancreatic Trypsin Inhibitor, BPTI, is a common model protein due to its small size and stability. The aforementioned pioneering NMR [3] and MD [4] work was performed on BPTI, the latter enabled by early crystallographic studies [11]. It inhibits chymotrypsin, trypsin, plasmin and kallikrein and has been used as an anticoagulation drug [12]. BPTI has eight aromatic rings: F4, F22, F33, F45, Y10, Y21, Y23 and Y35. Four of them, F22, F45, Y23 and Y35 are buried and the others are partly solvated. All rings except Y21 were studied in this project.

1.6 Aim

The aim of the project was to study the volume and activation volume of aromatic ring flips in BPTI according to the following points:

- Calculate the distribution of local volumes in ensembles generated by previous metadynamics simulations.
- Based on these preliminary investigations, run more metadynamics simulations with various protocols.
- Run metadynamics simulations at different pressures and determine the activation volume and compressibility.
- Attempt to separate cavity size from ring flip kinetics by developing a CV that describes expansion around the ring.
- Develop a phenylalanine dipeptide model with the purpose of studying the impact of solvation on volume .

2 Methods

2.1 Simulation setup

Simulation setup for BPTI was performed and is described in a previous preprint [8] and the same files were used here. Briefly, the force field used was AMBER ff14sb [13] and the water model was TIP4P-EW [14]. The trajectory was then equilibrated under constant volume and constant pressure.

2.2 Metadynamics

Metadynamics trajectories at ambient pressure were produced in the same previous preprint. [8]. They were performed with bias on χ_1 and χ_2 with a bias factor of 12, initial gaussian height 1.2 kJ mole $^{-1}$ and a gaussian width of 0.07 radians. Biasing χ_2 forces the flip, the rotation from a main state via a transition state and back to a main state, but χ_1 was also required to get convergence for most residues. Residue Y23 required biasing of a different CV, based on a contact map called CMAP [8], which describes movements in the local environment of the ring and χ_2, with the same parameters as the other runs, but the details of CMAP are not explained here.

In this project, metadynamics simulations were performed under constant pressure at 1, 400, 800 and 1200 bar. They were equilibrated in the same way as the previous trajectories and every residue had the same bias as previously, but were performed under constant pressure.

All simulations were performed using Gromacs 2019.4 [15] and plumed 2.7.1 [16].

2.2.1 Well-tempered Metadynamics

The mode was well-tempered Metadynamics [17], where the bias decreases with time. The bias is deposited as gaussians, with a peak height W and width σ. In regular metadynamics, the function describing the bias is

$$V_G(S,t) = \int_0^t \mathrm{d}t\, \omega\, exp\left(-\sum_{i=1}^{n} \frac{S_i(R) - S_i(R(t))^2}{2\sigma_i^2}\right) \tag{2.1}$$

Where V_G is the bias, t is time, ω is the energy rate, S is a collection of n functions dependent on the microscopic coordinate R.

$$\omega = \frac{W}{\tau_G} \tag{2.2}$$

Where τ_G is the deposition stride. The integral is implemented as a sum in practice.

$$V_G(S,t) = \sum_{k\tau_G<t} W(k\tau_G)\, exp\left(-\sum_{i=1}^{n} \frac{S_i(R) - S_i(R(t))}{2\sigma_i^2}\right) \tag{2.3}$$

Well-tempered Metadynamics uses time-dependent function of the gaussian height

$$W(t) = W_0\, exp\left(-\frac{1}{\gamma - 1}\, \beta\, V_G(S,t)\right) \tag{2.4}$$

Where γ is the bias factor and $\beta = (k_B\, T)^{-1}$. This is a time-dependent function, and the gaussian height will decrease with time and as such, the bias deposited will decrease.

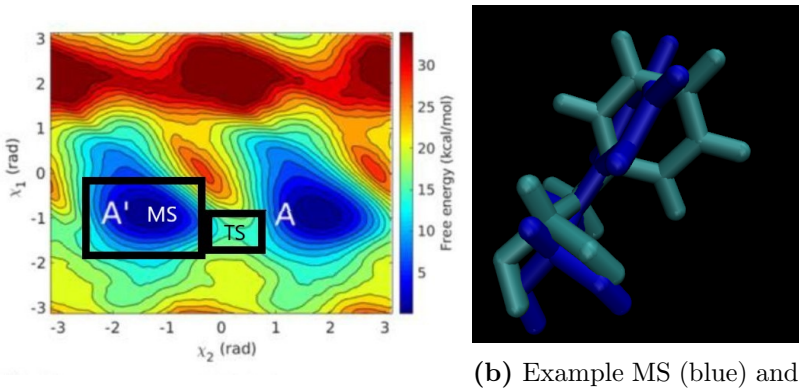
2.3 Local volume calculation

The raw trajectories were processed with the purpose of providing a stable spatial coordinate for the center of a residue of interest. The periodic boundary conditions of the simulation box were corrected and the distance between the center of the box and the most distant atom was minimized. The trajectory was then centered on the residue of interest and the center of the simulation box was set to 0,0,0 and the rotation and translation were fitted. This yields a simulation box containing BPTI and water centered on the residue of interest at coordinate 0,0,0.

The trajectory was partitioned into different ensembles, using $\chi_1\chi_2$ regions in the free energy surface (FES) to define main, intermediate and transition states, exemplified in Figure 2.1. This was accomplished by visual inspection and using an in-house developed script. It was assumed that χ_1 and χ_2 were sufficient descriptors of the transition state.

Finally, the volume was calculated using Epock [18], a software developed for calculating binding pockets in proteins. First, one defines a geometric inclusion volume, by defining center, shape and size. The inclusion volume obstructed by water or protein van der Waals volume is removed. The remaining volume is filled with a grid of spheres with a fixed sphere radius and distance between sphere centers, which allow for total volume calculation. Epock also has an algorithm for excluding cavities within the inclusion volume not connected to the main cavity. This avoids the potential problem of two discrete cavities computed as one.

Epock was used on the partitioned trajectories, where the residue of interest had been removed. Thus, the volume measured was the packing defects plus the ring, which is referred to as either the Epock volume or the ring cavity volume. The inclusion volume was a sphere with an 8 Å radius, 4 Å contiguous seed volume centered on 0,0,0 and the grid spheres had 1.4 Å radius and were set to a 0.3 Å distance. A sphere with an 8 Å radius has a volume of 2144 Å^3 volume and one with a 4 Å radius has a 268 Å^3 volume. The radius of the grid spheres was set to 1.4 Å as it is the van der



(a) FES of residue F45.

(b) Example MS (blue) and TS (cyan) of residue F45.

Figure 2.1: The free energy surface of Residue F45 (a) shows the energy minima and the lowest energy path between them, across the saddle point. The colored area in the $\chi_1\chi_2$ plane contains free energy of among others the main state and the transition state. The rectangle enclosing the dark blue free energy (low) is the main state, and rectangle enclosing the turquoise free energy is the transition state. Structures from the different ensembles have been extracted and is shown in (b) color coded correspondingly to their free energy in (a).

Waals radius of water, and the distance is set as short as possible within reasonable computation times, as lower sphere distances will result in closer approximations of the given topology, and 0.3 Å satisfied these conditions the best.

By subtracting the MS ensemble average volume from the TS ensemble average volume, a volume difference is acquired. This is an attempt to approximate the activation volume but is not formally it. This does correct for the van der Waals volume of the aromatic ring and yields the difference in packing defects, but the numerically computed volumes still have limitations. The problem with Epocks algorithm can be illustrated by a 2-dimensional representation, although it is equally valid in three dimensions. Figure 2.2 is a schematic image which shows that two rectangles with different sizes are both filled by two circles of equal size, and if the circles were used to describe the areas of the rectangles, they would seem equal, but are obviously not. The problem of Epock is the same, but in three dimensions.

2.4 Phe dipeptide

The smallest protein-like molecule containing an amino acid is the dipeptide, which has two peptide bonds, but only one amino acid. The phenylalanine (Phe, F) dipeptide is the molecule exemplifying the dihedral angles in Figure 1.1, and the peptide bonds are between Phe and N- and C-terminals, and not other amino acids. There is some inconsistencies in the literature as to what a dipeptide is and an alernative use is "di" to refer to the number of amino acids. Then, a phenylalanine dipeptide would be a phenylalanyl phenylalanine, containing two phenylalanines, but that is not the definition used here.

A Phenylalanine dipeptide was prepared using AMBER ff14SB, put into a cubic box

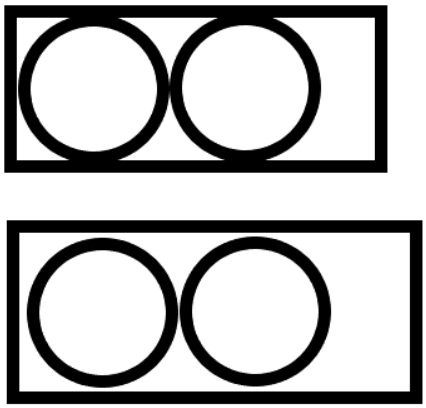


Figure 2.2: Two rectangles with different sizes. If their area is approximated by filling them with circles with the size of the example, their volumes seem identical.

with 10 Å side and solvated with the TIP4P water model [19]. The simulations were performed at 1 bar, 300 K and with a time step of 2 fs. The structure was energy minimized and then equilibrated in the NVT and then NPT ensembles (constant amount of substance, temperature, and volume or pressure. The production runs were performed in the NPT ensemble. Metadynamics simulations with bias on the dihedrals ϕ and ψ were performed using the same CV parameters as the metadynamics trajectories biasing χ_1 and χ_2, a bias factor of 12, initial gaussian height 1.2 kJ mole $^{-1}$ and a gaussian width of 0.07 radians. The χ_1 and χ_2 dihedrals are more flexible than ϕ and ψ in the dipeptide, explaining the choice of CVs for biasing.

The dipeptide results were analyzed by studying its autocorrelation function. The autocorrelation function is a statistical function determining the correlation between consecutive values in a data set at different distances, reporting on to which degree consecutive values can be predicted from current values. This can be used as a rough estimation of flip kinetics, as it will be related to the probability that the ring has flipped after a certain time. The volume can also be analyzed in this way, but the conclusions will be different. It can report on how fast the local environment of the ring changes.

The autocorrelation function was calculated using the gromacs tool gmx analyze. It returns a data set containing the correlation between a value at a certain time with values at consecutive for $\frac{n}{2}$ values where n is the size of the data set. It returns 1 if the values are identical, 0 if there is no correlation, and in between if there is some correlation. An exponential function was fitted to the autocorrelation of the χ_2 values and a double exponential function was fitted to the autocorrelation of the V values.

$$y = exp\left(-\frac{x}{a_0}\right) \tag{2.5}$$

$$y = a_0\, exp\left(-\frac{x}{a_1}\right) + (1 - a_0)\, exp\left(-\frac{x}{a_2}\right) \tag{2.6}$$

Where a_n is fitted parameters. Which function was used was based on which used the least number of variables to capture the initial decay. Residue F4 and F45 were

analysed in the same way.

2.5 Biasing cavity expansion

A CV based on the coordination number (CN) of the ring atoms and carbon atoms in the protein was developed using the plumed keyword COORDINATION. It is based on a switching function:

$$s_{ij} = \frac{1 - (\frac{(r_{ij} - d_0)^n}{r_0})}{1 - (\frac{(r_{ij} - d_0)^m}{r_0})} \tag{2.7}$$

where s_{ij} is the contribution to the coordination number from the interaction between atoms i and j, r_{ij} is the distance between atoms i and j, and d_0, r_0, n and m are design parameters, manipulated to get functions of interest. The atom groups for which the contacts were calculated were ring carbons 2-6 and the phenolic oxygen if the residue was tyrosine, and all carbon atoms in the protein. By looking at the contacts between the ring and protein carbon atoms, most side chain and backbone atoms are covered. The coordination number CV was implemented with equation (2.7) with the parameters $n = 3$, $m = 6$, $d_0 = 2$ Å and $r_0 = 3$ Å.

A metadynamics simulation with bias on CN was performed on residue Y23 which had required biasing the CMAP CV to reach χ_2 convergence [8] It had a bias factor of 12, initial gaussian height 1.2 kJ mole^{-1} and a gaussian width of 1.65.

Other simulations were performed with the plumed keyword UPPERWALLS to confine the system to low coordination numbers, 25 and 40. It applies a constant bias whenever the CV is higher than the limit, and is defined as:

$$V = k(x - a)^n \tag{2.8}$$

Where V is the constant bias, k is an energy constant, x is the value of the CV, a is the limit and n is an exponent. The simulation with the limit $a = 40$ had $k = 10$, and $n = 4$ and the simulation with $a = 25$ had $k = 20$, and $n = 4$. If a particular range of values of a CV is of interest, this method can be used to ensure that only that range is sampled.

Metadynamics simulations on the buried residues (F22, F45, Y23, Y35) were performed with bias on χ_2 and and CN, with the biases as previously described. Converging χ_2 has required biasing an additional CV, and by comparing these results with those of the simulations biasing χ_1 or CMAP and χ_2, which are known to converge χ_2, the impact of the CN can be further evaluated.

3 Results and discussion

The volume distributions of every defined state in every residue are calculated and analysed in this section. The volume distributions for F45 at different pressures are also calculated, and they are used as a basis for calculating its relation to the flip rate. The phenylalanine dipeptide was compared to a buried and a solvated residue. Finally, the proposed CV for biasing cavity expansion, the coordination number is evaluated in terms of its potential to bias ring flips together with χ_2.

3.1 Calculating the local volume

The following section will illustrate volume calculation qualitatively. Residue F45 will be discussed, as its free energy surface has already been used as an example in Methods. The volume calculation method is based on removing the sidechain of the residue of interest from the trajectory, meaning that the volume calculated is that of the ring and of the packing defects, shown in Figure 3.1. The ring was not there when the volume was calculated, and the cavity that was created by removing it was filled with the spheres. This quantity is referred to as the ring cavity volume, and sometimes local volume. When it concerns the specific local volume of a specific ensemble, they might be referred to as main state volumes.

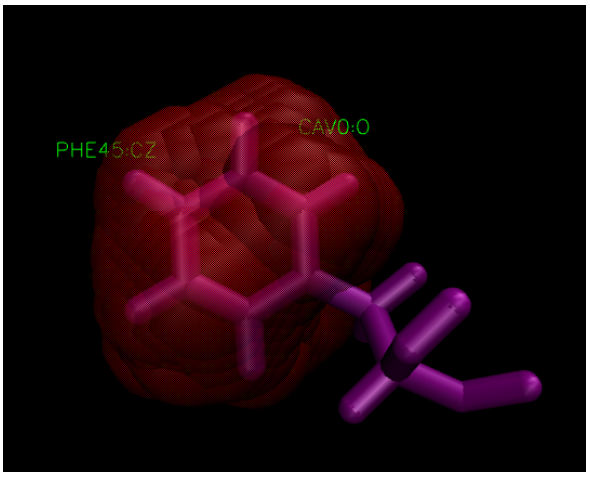


Figure 3.1: Residue F45 (magenta ring) and the associated volume (red cloud of spheres) at a time step in the MS ensemble. The red volume is the cavity generated by Epock (ring cavity volume) and the aromatic ring is residue F45. Due to the packing defects, the ring cavity volume is not homogeneous.

The cavity seems to encapsulate the ring. It is also clear that the ring cavity volume, unlike the ring is, not symmetric. This is because some of the packing defects are captured by Epock. As mentioned in the Methods section, ensemble average of the ring cavity volume can be calculated from the red spheres. The ring cavity volume

versus time is shown in Figure 3.2. There are some outliers but overall, the volumes are stable in intervals. It is clear that the MS has lower volumes than the TS in the plot where the TS volumes are overlaid on the MS.

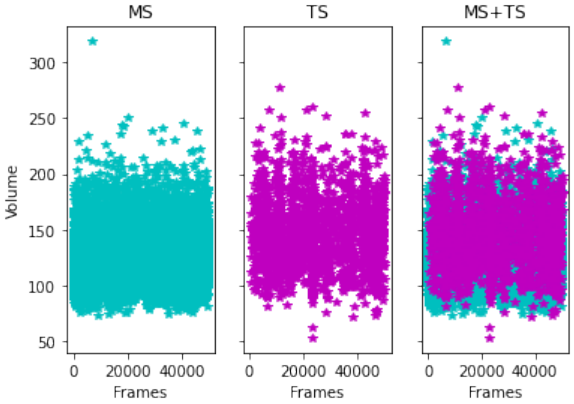


Figure 3.2: The ring cavity volume plotted against time for the main state, transition state, and main state and transition state. It implies that the transition state volumes have a higher average than the main state volumes and thusly that the transition state cavity is larger than the main state cavity.

The volume distribution is shown in Figure 3.3. It is clear that the average volume is greater in the TS, but also that the distribution is wider.

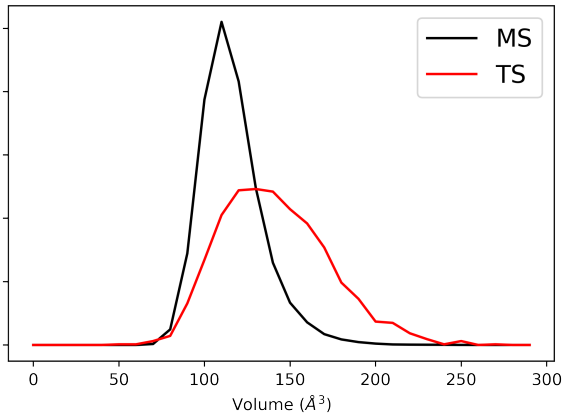


Figure 3.3: The MS and TS volume distribution of residue F45. TS has a higher average, greater variance and has 200+ Å ^{3} states in the ensemble.

The ensemble averages with standard errors for MS and TS and the differences (presented in Table 3.3 were determined by statistical bootstrapping. The distributions were compared using the Kolmogorov-Smirnov statistical test, testing for statistically significant differences between the distributions. There was extremely low probability of identity ($p = 2.8 * 10^{-290}$). Taken together, the evidence presented here suggest that the proposed method can distinguish between main and transition states and implementation shows that the ring cavity volume is significantly larger in the TS than in the MS for residue F45.

3.2 Volume and local volume fluctuations

This section takes a quantitative approach to the calculating volume of every state in Residue F4, F22, F33, F45, Y10, Y23 and Y35. The states of all residues except F45 are shown and marked in Figure 3.4. These mark the main state (MS), transition states (TS 1-3) and intermediate states (IS 1-2). For every marked state, there is a periodic state which was also used, shifted π radians along χ_2. Occasionally, there are states at the χ_1 limits at π or $-\pi$, in which case both parts are marked with a rectangle but only one is labelled.

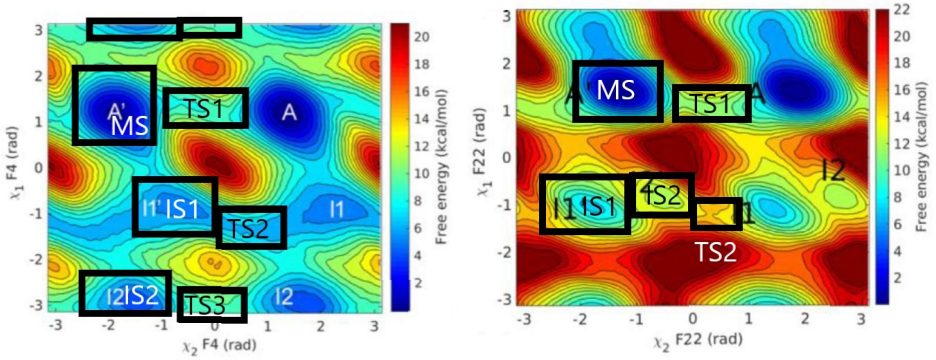
The ensemble average volume of all states based on their definition in the free energy surface is presented in Table 3.1. For the distributions, see Appendix A. They show the bootstrapped average volumes with the bootstrapped standard error for all stable and transition states for the seven studied rings. The main state volume is almost always significantly lower than the transition states. Residue Y23 and Y35, where this is not true, had problems with convergence, which may partly explain the variation. This method assumes that χ_1 or CMAP and χ_2 describe the important slow system coordinates for the flip, but this may not be the case for these residues. The assumption that the biased CVs sufficiently describe the flip is in fact critical and necessary for the validity of the state definitions.

The values in Table 3.1 are color coded based on which stable and transition states are connected. Take (a) in Figure 3.4 as an example; in residue F4, MS and TS1, IS1 and TS2, and IS2 and TS3 are connected. These color codes have nothing to do with the heat map in the free energy surfaces, but how states are connected in it.

Table 3.1: Ring cavity volumes of all main-, intermediate- and transition states. If there is no value, the particular ring did not have any such state. The values are color coded, showing which states are connected. MS and TS1 are always connected, as can be deduced from Figure 3.4

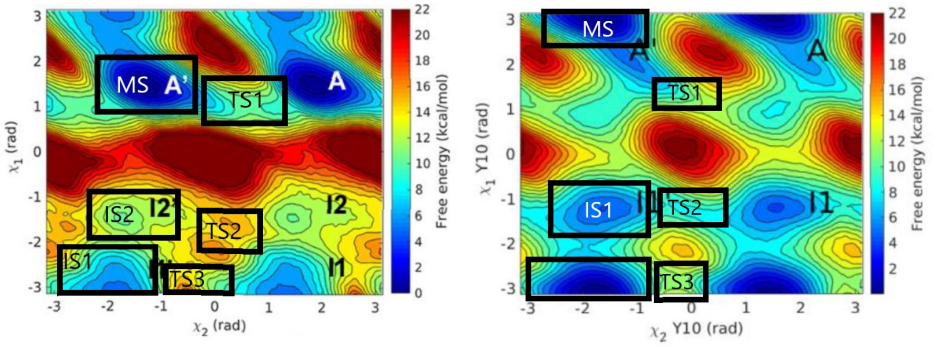
Residue	MS (Å^3)	TS1 (Å^3)	TS2 (Å^3)	TS3 (Å^3)	IS1 (Å^3)	IS2 (Å^3)
F4	137.5 ± 0.2	147.8 ± 0.5	163.6 ± 0.5	157.2 ± 0.6	168 ± 0.5	150.7 ± 0.4
F22	122.6 ± 0.1	132.1 ± 0.4	122.7 ± 0.5	-	121.4 ± 0.2	126.7 ± 0.4
F33	130.7 ± 0.2	132.4 ± 0.9	146.9± 1.3	143.7 ± 1.8	157.4 ± 1.3	173.4± 0.5
F45	121.1 ± 0.1	146.2 ± 0.7	-	-	-	-
Y10	132.0 ± 0.2	137.6 ± 0.3	138.9 ± 0.4	135.2 ± 0.6	134.1± 0.3	-
Y23	135.8 ± 0.2	136.2 ± 0.8	141.6± 3.7	131.0 ± 3.1	146.4 ± 1.5	-
Y35	129.7 ± 0.1	126.6 ± 0.4	132.3± 0.4	140.1 ± 0.5	129.8 ± 0.2	130.6 ± 0.2

The transition states are connected either to a MS or an IS. This is used to determine which path across χ_2 in the FES that is associated with the lowest energy, and if that correlates with the lowest volume difference. In many cases, the transition with the lowest energy TS has the highest volume difference. This is true for F22, F33 and Y10, which all have multiple transition states. There are two values for TS2 in residue F22, IS1 and IS2 as both are connected through it. Based on the values in Table 3.1, Table 3.2 was constructed, and it contains the volume differences of the states, based on the color coding.



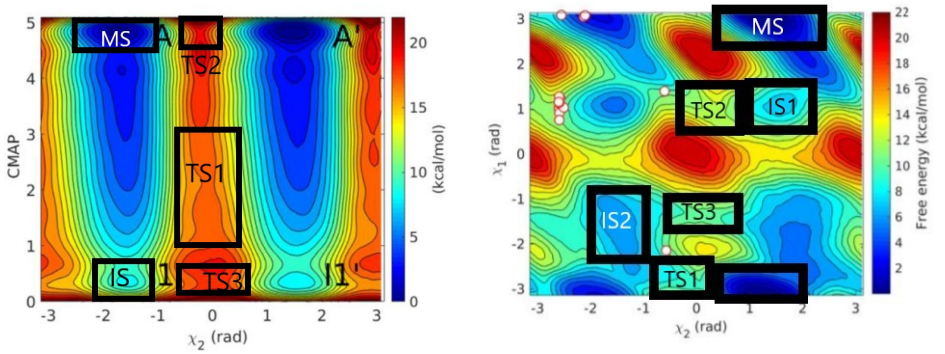
(a) FES of residue F4 with states marked.

(b) FES of residue F22 with states marked.



(c) FES of residue F33 with states marked.

(d) FES of residue Y10 with states marked.



(e) FES of residue Y23 with states marked.

(f) FES of residue Y35 with states marked.

Figure 3.4: The free energy surfaces of all residues except F45, which has already been shown in the methods section. The area contained by the black rectangles represent the different studied states extracted from the simulations. Figures are adapted from [8] and contain their notation for the states.

The volume differences are not always positive, which is sometimes difficult to explain. The most drastic difference, between IS2 and TS2 in residue F33, had a "volume leakage" in IS2. This means that Epock found a contiguous path from the ring cavity to the protein interior. That volume was obviously not part of the ring cavity, but was still included in the volume calculation, which explains the increase. For residue Y23, the number of frames in IS, TS2 and TS3 were lower than 100, as compared to every other ensemble which had at least 1000 frames. When biasing, the Boltzmann dis-

Table 3.2: The volume difference between a stable state across the different transition states for every residue. The subscripts here refer to which transition state was traveled through and the corresponding main state can be deduced from 3.1.

Residue	ΔV_{TS1} (Å^3)	ΔV_{TS2} (Å^3)	ΔV_{TS3} (Å^3)
F4	10.3 ± 0.3	-4.6 ± 0.6	6.5 ± 0.5
F22	9.6 ± 0.2	-4.0 ± 0.6, 1.2 ± 0.3	-
F33	1.6 ± 0.3	-26.5 ± 0.7	-13.7 ± 1.6
F45	25.0 ± 0.2	-	-
Y10	5.6 ± 0.2	4.9 ± 0.4	3.2± 0.2
Y23	0.4 ± 0.3	5.8 ±0 .3	-15.4 ± 1.8
Y35	-3.1 ± 02	10.4 ± 0.3	-4.0 ± 0.3

tribution is skewed towards higher energy states along the biased CVs. Intermediate states and transition states should be visited when biasing, as the intermediate state is metastable and the transition state includes the saddle point of the free energy barrier. There is a purely statistical argument that the sample size would barely be acceptable for the central limit theorem. There is also a statistical mechanical argument that as the energy landscape was sampled (ergodicity) and the ensembles as defined are so small, they can be assumed to be irrelevant.

As for the negative activation volumes not discussed, it is not clear why they were negative and it may indeed be physical. If they are taken to be physical, one interpretation might be that depending on the topology of the cavity, different volumes are necessary for rotation. It may be that the mechanism allowing the topology change of the ring cavity (rotation), inherently shrinks it for residue Y35.

There was also a challenge in defining appropriate areas in the free energy surface with rectangles. With residue F33 for instance, TS2 and TS3 contains saddle points, but also local maxima, which could mean that some high energy states are included where they should not be and thus affect the ensemble average in a non-representative way. There are some cases where negative volume differences are challenging to explain and they may indeed be real. Overall though, most flips are associated with a positive volume increase, especially for the main states.

The connection between the volume differences and the previously determined ring flip kinetics is not obvious. For instance, residue F4 was shown to have a flip rate three orders of magnitude greater than that of F22 [8], but they have quite similar V_{TS1}. This might be due to some other thermodynamic quantity which is not examined here, such as the entropy of activation $\Delta^{\ddagger}S$, contributing more to the free energy barrier, which could be examined with experimental methods [5].

In almost all cases, the MS-TS1 transition is associated with the highest volume difference. On the one hand, this is in agreement with the theory that the packing defects expand during flips, but on the other hand, this means that this path has the greatest positive contribution to the free energy barrier. The volume of MS and TS1, and the difference between MS and TS1 are shown in Table 3.3 for further discussion. Values for the global activation volume are included for some residues. It describes

the volumetric increase of the entire protein, and is needed to evaluate whether or not the data supports the expansion mechanism, which was mentioned in the introduction.

Table 3.3: Main state and the closest transition state and the difference between them, including a global volume difference calculation for comparison.

Residue	V_{MS} (Å^3)	V_{TS} (Å^3)	ΔV_{local} (Å^3)	ΔV_{global} (Å^3) [20]
F4	137.5 ± 0.2	147.8 ± 0.4	10.3 ± 0.3	-
F22	122.6 ± 0.1	132.1 ± 0.5	9.6± 0.2	15.8 ± 1.1
F33	130.7 ± 0.2	132.4 ± 0.9	1.6± 0.3	-
F45	121.1 ± 0.1	146.2 ± 0.7	25.0± 0.2	25.6 ± 1.3
Y10	132.1 ± 0.2	137.6± 0.3, 135.2 ± 0.6	5.6± 0.2, 3.2± 0.2	-
Y23	135.8 ± 0.2	136.2 ± 0.8	0.4± 0.3	16.6 ± 1.2
Y35	129.7 ± 0.1	126.6 ± 0.5	-3.1± 0.2	3.7 ± 1.5

The results for the local volume are positive volume differences for all residues except Y35, indicating that the transition state is characterized by an increase in protein volume, consistent with theory presented in among other places here [5]. The negative volume difference of residue Y35 means that the ring cavity is actually shrinking which is inconsistent with the theory and the other results. There was as previously mentioned convergence problems for it, which may partly explain this anomaly.

The local volume differences and the global volume differences determined by Gavrilov [20] is always significantly higher than the local volume differences. For the global volume, a different software was used, Protein Volume 1.3 (PV 1.3) [21] which is not explained here. Although PV 1.3 uses a different algorithm and calculates global volume, comparing the MS and TS differences is still relevant, as systematic differences are eliminated when subtracting them. In the cases where there is comparable data, the global volumes are significantly larger than the local volumes except for residue F45. This indicates that the entire increase in global volume is towards the local volume for F45 and that the global volume increase in the other residue pockets may enlarge other cavities in the protein as well as the local volume of those rings. This supports the expansion mechanism mentioned in the introduction, with positive local and global volume differences.

3.3 Volume and local volume fluctuations at high pressure

As mentioned in the introduction, BPTI is a stable protein and previous simulations have been performed at up to 10000 bar [22]. Studying volume differences associated with flips under varying pressure can give information on the compressibility, free energy and ultimately the kinetics, which was shown in the derivation of equation 1.3.

The volume distributions of residue F45 is shown as histograms in Figure 3.5. The MS and TS volumes are shifted to lower volumes when pressure increases, which is

exemplified by the TS peak moving from right of the intersection of the MS and TS volume distributions to left of it, when comparing the histograms at 1 and 1200 bar. The MS also decreases, shown in Table 3.4, but it decreases less than the TS, resulting in a net decrease in the difference. The volume decreases with increasing pressure, which shows that both states are compressible. The net decrease in the difference can be interpreted as an activation volume decrease driven by the activation compressibility, according to equation (1.5). This means that the increased pressure-volume work required as a result of increased pressure leads to the volume expanding less. In a sense, the expansion becomes more economic in terms of the volume required for the flip to occur. This seems like a reasonable proposition, as the transition state has been shown to be larger than the main state, and that the increase is as a result of packing defect expansion and as such that more vacuum is created around the ring.

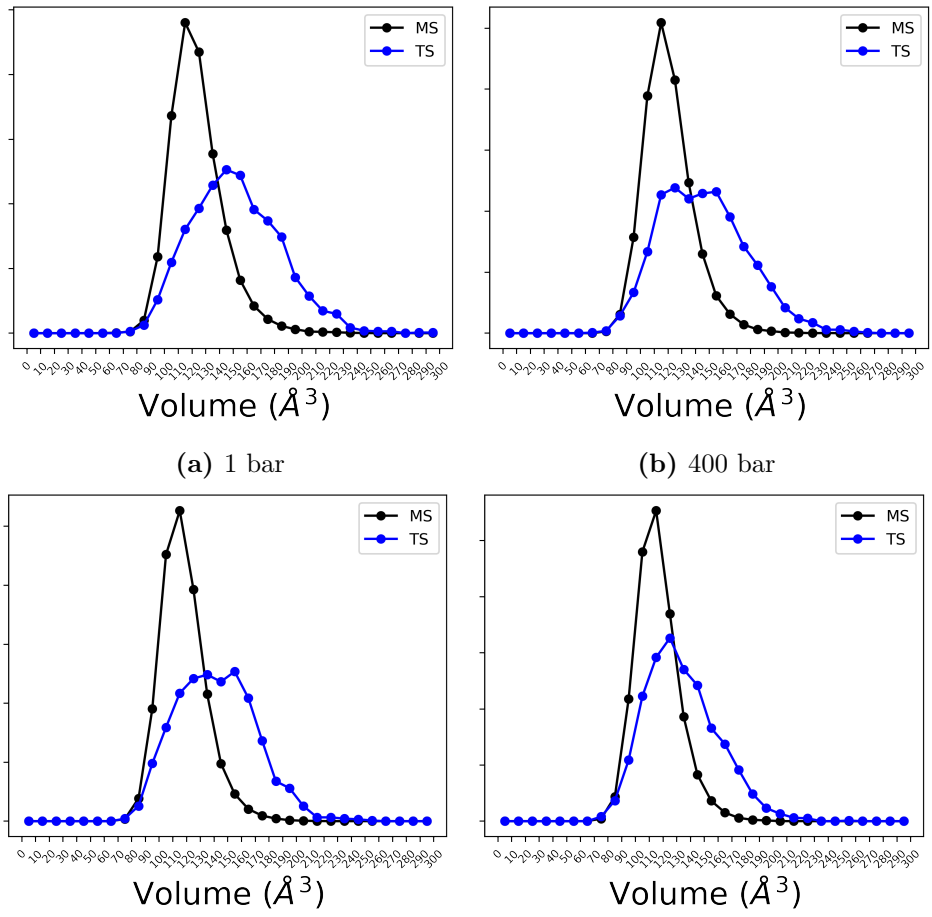
Table 3.4: ΔV of residue F45 at 1, 400 800 and 1200 bar.

p (bar)	V_{MS} (Å^3)	V_{TS} (Å^3)	ΔV (Å^3)
1	123.4	151.6	28.2
400	120.3	145.6	25.4
800	117.7	141.3	23.6
1200	116.0	133.5	17.2

Two local maxima can be distinguished in the volume distributions at 400 and 800 bar in Figure 3.5. This may be explained by the fact that residue F45 can be either buried or solvated. Solvated residues flip faster and as a result, the two local maxima found in the histograms indicate two different clusters in the transition state ensemble, the large volume cluster being solvated and the low volume cluster being buried. They seem however to occur at similar frequency. The higher volume peak was not visible under constant volume in Figure 3.3. The reason for this may be that the buried and solvated ensembles are affected differently by pressure and that changing the pressure could be what exposes them here.

The bootstrapped volume averages are different in Table 3.4 and Table 3.3. This is probably explained by the fact that the values in Table 3.3 were produced under constant volume, but the values in Table 3.4 were produced under constant pressure. If something is to expand under constant volume, the space may be created only by compacting something else. If it expands under constant pressure, the expansion may be facilitated through enlarging the box which may be the explanation for the increase in ensemble average volumes for residue F45 at constant pressure compared to constant volume. The decrease in ΔV does not necessarily show support for the compensation mechanism, as it only reports on the local volume. However, it shows that if the local expansion is facilitated by global protein expansion, the global expansion required will be less.

The activation volume is a formal thermodynamic quantity and is related to the free energy under isothermal conditions according to (1.3). It describes the difference between a main and a transition state and is mainly determined through experiments. As such calculating the TS and MS ensemble averages and the average differences between them does not formally result in the activation volume. The difference can



(a) 1 bar **(b)** 400 bar

(c) 800 bar **(d)** 1200 bar

Figure 3.5: The volume distribution of the main and transition states of residue F45 at a) 1 bar b) 400 bar c) 800 bar d) 1200 bar. With increasing pressure, the distributions are shifted towards lower volumes, but there is a greater effect on the TS.

however be argued to approximate it. Then, a geometric volume is assumed to have a fundamental connection to the rate constant through the activation volume via equations (1.2) and (1.3). In the following discussion, this assumption will be made, for the purpose of assessing the quality of the approximation, as there is experimental data to compare results to [5].

The volume data at different pressures allow for the usage of equation (1.5), to determine what is called the volume compressibility κ' by linear regression. The data and the regression line is shown in Figure 3.6, with $\Delta^{\ddagger}\kappa' = 174$ Å^3 GPa^{-1} and $\Delta^{\ddagger}V_0 = 28.8$ Å^3). This is in qualitative agreement with the results of Dreydoppel et al as the volume compressibility has a positive sign, with the value being off with a factor of 2. The volume compressibility is not trivial to grasp, but a positive activation volume compressibility can be interpreted as the TS being more porous than the MS, consistent with the view that the packing defects are enlarged in the TS.

Assuming that the volume difference here computed is the activation volume, the activation compressibility coefficient can be calculated. Recall equation (1.7), $\Delta^{\ddagger}\kappa' = \Delta^{\ddagger}\kappa V_0 + \Delta^{\ddagger}V_0\kappa$. By using a previously determined compressibility $\kappa = 0.23$ GPa^{-1} [23]

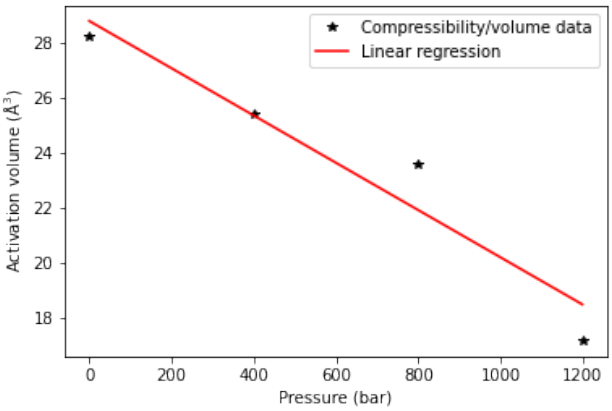


Figure 3.6: The different values of the volume difference of residue F45 at 1, 400, 800 and 1200 bar, and the curve fitted with linear regression.

(under the same force field and water model), the previously calculated $\Delta^{\ddagger}V_0$ and $\Delta^{\ddagger}\kappa'$ and setting V_0 equal to the main state volume at 1 bar, 123.4, the activation compressibility coefficient $\Delta^{\ddagger}\kappa$ is calculated to 5.06 GPa^{-1} and equation (1.8), ${}^{\ddagger}\kappa = \kappa + \Delta^{\ddagger}\kappa$, is used to calculate the transition state compressibility coefficient ${}^{\ddagger}\kappa$ to 5.29 GPa^{-9}, similar to compressible gravels and half an order of magnitude larger than ethanol ($\kappa = 1.15$ GPa^{-1}) [24] and Dreydoppel et als (${}^{\ddagger}\kappa = 1.6$ GPa^{-1}) [5].

There is some ambiguity in which V_0 to use. Here, it is assigned to the local volume estimated by Epock. Perhaps the volume difference should not be assigned to merely the cavity, but include the protein matrix. For reference, a volume identical to that of the contiguous seed sphere gives only a 20% decrease in the calculated compressibility. The order of magnitude of the value thus seems stable, but the understanding of the process is quite different as the method used implies that the cavity expansion is completely unrelated to eventual changes in the protein matrix which may be a too simplistic view of the phenomenon.

The compressibility can be calculated from the variance in the volume distributions. It is however not clear to which extent the variance is affected by the methodology, but the averages appeared to be stable in testing.

With the activation volume compressibility determined, the change in the free energy of activation as a result of increased pressure can be computed by performing some slight algebra on equation (1.6) into

$$\Delta^{\ddagger}\Delta G = \Delta^{\ddagger}G - \Delta^{\ddagger}G_0 = \Delta^{\ddagger}V_0(p - p_0) - \Delta^{\ddagger}\kappa'\frac{(p - p_0)^2}{2}. \tag{3.1}$$

The change was 0.5, 0.7 and 0.6 kJ mol^{-1} for 400, 800 and 1200 bar respectively, with 1 bar as reference. It seems counterintuitive that the pressure contribution should decrease with pressure, but the equation has a global minimum and it is not obvious that the pressure-volume work spent on enlarging the cavity should increase with pressure, even though the work per expanded volume unit does.

Reservations are necessary due to the speculative nature of the assumption that $\Delta V = \Delta^{\ddagger} V$ and since Dreydoppel et al studied a ring in a different protein, but its noteworthy that the order of magnitude ends up reasonable.

3.4 Phenylalanine Dipeptide

Some rings are solvated, and it is important to understand how this affects ring flips. The phenylalanine dipeptide is the solvated F residue in its extreme and serves as a reference in terms of degree of solvation. By studying the autocorrelation of χ_2 values, a rough estimate on the flip rate can be obtained. The autocorrelation of V can give insights into the flexibility of the ring. Epock is sensitive to fluctuations in protein cavity ring environment and the protein matrix is more rigid than the solvation layer, leading to wider volume distributions.

The fact that contact with water widens the volume distribution is demonstrated in Figure 3.7, where the distribution of volumes for the buried residue F45 is narrow with a low average volume. The opposite is true for the dipeptide, where the distribution is wide with a high average volume. Residue F4, which is solvated and has more protein around it than the dipeptide, but less than F45, ends up in the middle. There are three different bumps in the volume distribution of the dipeptide, likely due to the three different χ_2 rotamers, which will have a different amount of protein atoms around them.

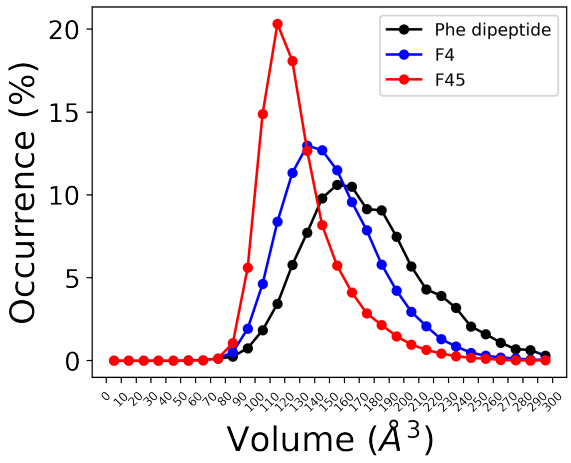


Figure 3.7: The volume distributions over all states in the Phenylalanine dipeptide, residue F4 and residue F45, shown as histograms.

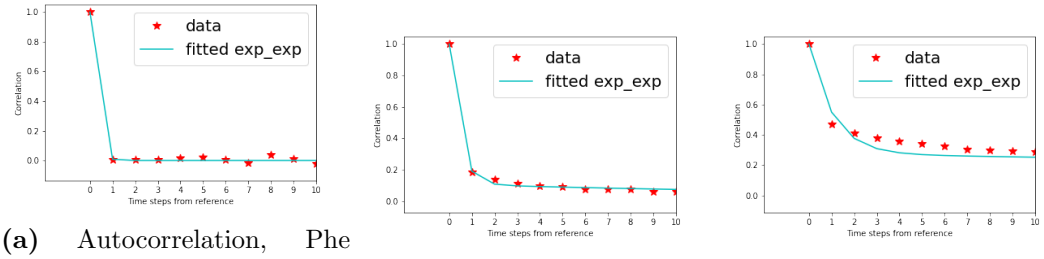
By comparing this to the area of the integral of the function fitted to the autocorrelation of the volumes, there is a trend confirming the image drawn by the volume distributions. The integral is much larger for F45 than F4, which in turn is larger

than that of the dipeptide. If the integral is large, it means that the volumes are correlated, and that implies that the fluctuations in the rings local environment is lower for F45 than F4, which is supported by the narrow distribution of it. The χ_2 integral however is not consistent with this view, as the dipeptide has the lowest area and F45 the highest, as demonstrated in Figure 3.8 and Table 3.5. There instead, the fitted parameters have that same relation.

Table 3.5: Parameters and integrals from the exponential functions fitted to the autocorrelation of χ_2 and V for the residue F45, F4 and the dipeptide.

Residue	a_{0,χ_2}	$\chi_2$integral	$a_{0,V}$, $a_{1,V}$, $a_{2,V}$	V integral
F45	182	183	10.4, 0. 724, 1120	403
F4	42.8	65.8	4.29, 0.893, 269	59.2
Phe dipeptide	10.5	252	1.6 1.04 1.81	49.0

Figure 3.8 shows the autocorrelation of the volume values for Residue F4, Residue F45 and the phenylalanine dipeptide, and Table 3.5 contains the a constants fitted to the exponentials describing the early decay. Based on the fitted exponentials, the more solvated residue decreases faster, but there was poor agreement between this and the volume integrals. One would expect the solvated residue to decrease faster and thus to have a lower integral, as a low correlation can be inferred from fast flips. In terms of χ_2, this means that the rings flip faster as the ring fluctuates between discrete intervals in the two periodic main states. The transitions between them result in traveling from a positive angle to a negative one (the angular space is defined to be from $-\pi$ to π radians) and there is little correlation between values from different main states.



(a) Autocorrelation, Phe dipeptide

(b) Autocorrelation, F4

(c) Autocorrelation, F45

Figure 3.8: The correlation and fitted double exponentials for the Phe dipeptide for the ten closest consecutive frames.

The autocorrelation of the volumes show that there is little predictability from the current volume to the next, suggesting either that the cavity is highly dynamic or that Epock is sensitive towards small changes in the local environment. Findings during method development showed that it is quite sensitive to tiny variations in the center of the inclusion sphere, and it may also be sensitive to small changes in the ring cavity border.

It seems like solvation may indeed be a contributing factor to flip rate and some theoretical support can be found, but a more systematic approach is necessary. One approach could be to quantify the number of water contacts and add this to the analysis.

3.5 Biasing cavity expansion

After running simulations with bias on the coordination number as defined in the Methods section and plotting χ_2 and CN against time, it is clear that the system visits certain confirmations with a low coordination number, and that it is connected to ring flips. The system visits certain low CN regions, at around 160 000 ps and 250 000-300 000 ps. There, χ_2 rapidly changes meaning that flips occur, shown in Figure 3.9. This connection indicates that the coordination number successfully creates space around the ring, as the χ_2 rotation is now permitted. Higher coordination numbers however do not permit flips.

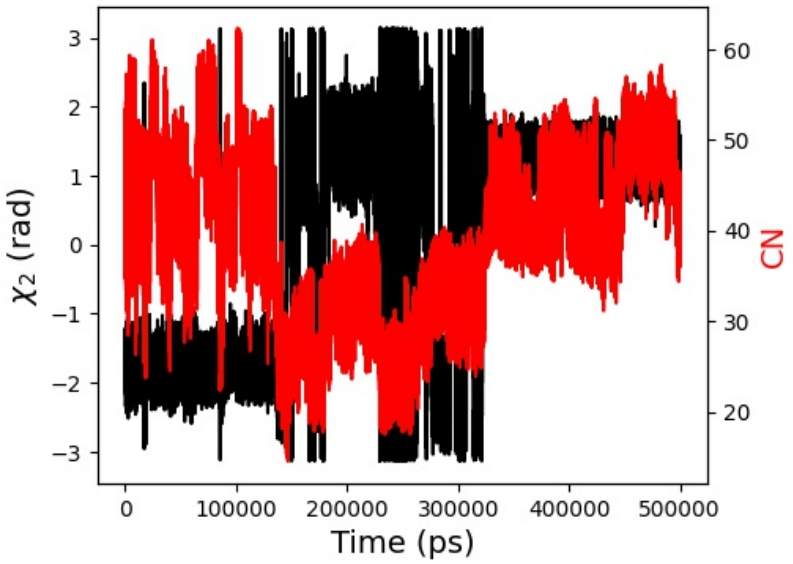
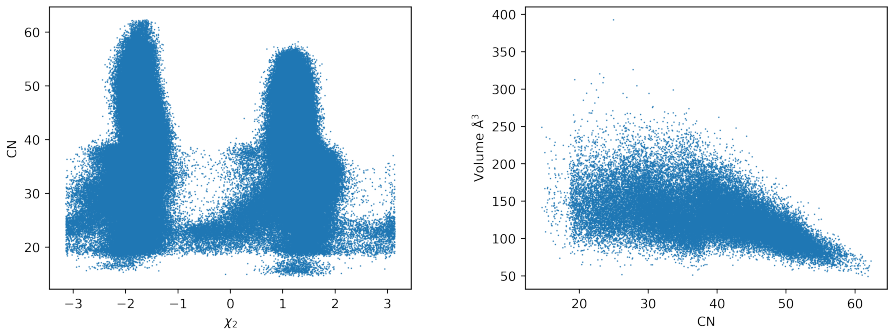


Figure 3.9: The coordination number and χ_2 plotted against time for Residue Y23 with bias on CN. The black is χ_2 and the red is CN. At around 160 000 ps and 250 000-300 000 ps, there are many transitions between the χ_2 intervals, corresponding to the main states. The transitions between the discrete χ_2 intervals, the flips, are seen only when the coordination number is low.

further evidence for this relation is provided when plotting χ_2 against CN, shown in Figure 3.10. The width of the main state fluctuations at around -2 and 1 rad along the χ_2 axis gives a further indication that the two are connected, as an increasing width is associated with the ring having more space to rotate, and a decrease in CN with the protein atoms being at a greater distance. Below CN = 40, the flips begin to happen and below 25, the flips are frequent and the entire angular space is sampled.

Plotting volumes against the coordination number, there seems to be a linear, al-

though the line is broad, relationship from CN 40 and higher. Below 40, the volumes are more widely distributed, which seems consistent with the results and theory previously discussed, around 40 is the specific region where flips can start to happen. Around CN = 25, there seems to be a constant interval, with the minimum value around 100 Å^3, as compared to the minimum values at around CN = 60 at almost 50 Å^3.

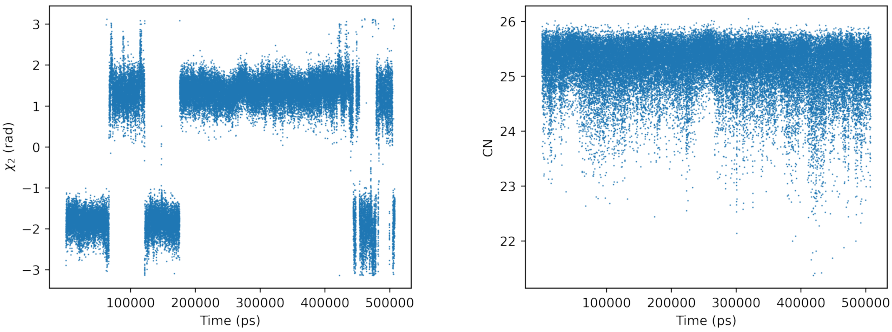


(a) CN against χ_2 for Residue Y23 with bias on CN

(b) χ_2 against CN for Residue Y23 with bias on CN

Figure 3.10: a) shows the relation between CN and χ_2 and b) shows the relation between CN and volume. At about CN 25 and 40, there seems to be changes occuring to both χ_2 and volume that can be argued to be causal effects of changing the CN.

The simulations with upper walls at CN 25 and 40 build on the observations from Figure 3.9. No flips happened with CN 40, but there were some with CN 25 as shown in Figure 3.11. The CN is quite stable around the wall, indicating that the system is trying to escape the low values which it has been constrained to. Comparing to Figure 3.9, there is a similar number of flipping events, but here, there are occasional single flips as opposed to the regions with a high frequency of flips.



(a) χ_2 against frame for residue Y23 with upper walls at CN 25

(b) CN against frame for residue Y23 with upper walls at CN 25

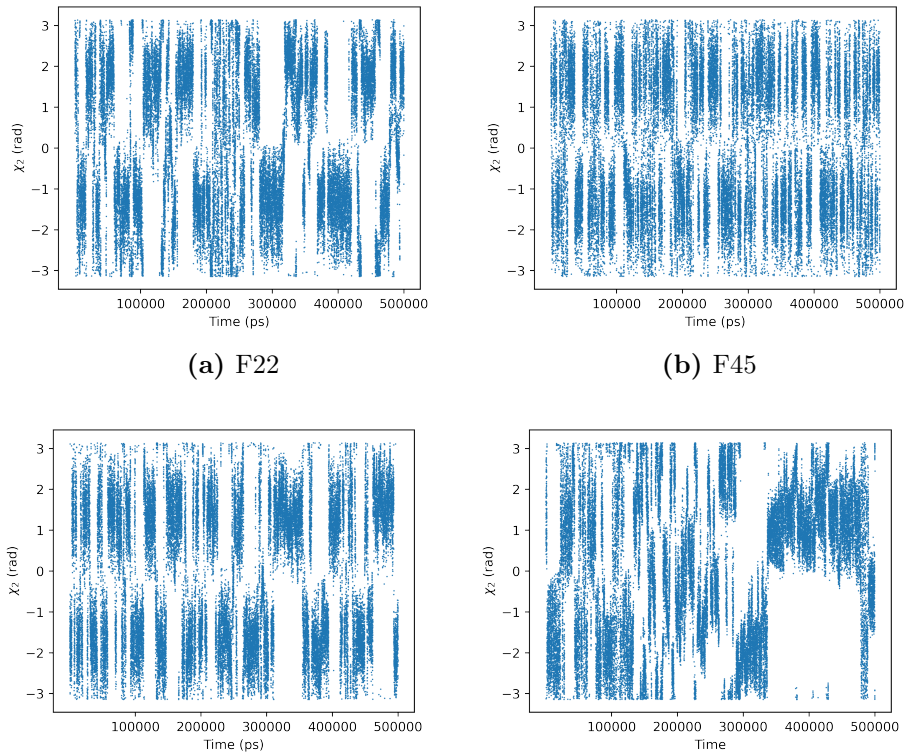
Figure 3.11: In the simulations with upper wall on CN 25 Residue Y23, some flips occured. The CN was mostly around 25, the upper wall, which indicates that the system tries to escape the constraint that has been put on it. It does not seem like any variations in the coordination number is connected to the flips.

This suggests that it is not enough to reach a low coordination number, and that the variability introduced by the time-dependent bias is required for the flips to occur. The

coordination number does not really report specific things on the local environment of the ring, only that the carbon atoms are at a certain distance from the ring, and fluctuations may be required to sample states which allow for flips, irrespective of the specific value of the coordination number.

One has to be careful reading too much into these results, as they are still generated by applying a metadynamics bias and are as such unphysical, but the results are remarkable in that they generate ring flips in a similar timescale as [8] without directly biasing χ_2. A further complication is that the simulation did not converge, and that similar coordination number can represent different conformations. The coordination number only reports on how close carbon atoms is to the ring.

Finally simulations with bias on the coordination number and χ_2 show that this combination can produce trajectories with frequent flips, in Figure 3.12. Residue F22, Y23 and Y35 approached but did not fully reach convergence in the simulation times, but residue F45 did. It is not clear if the other simulations would have converged, but the results still show that this CV combination generates frequent flips. It creates some doubts as to which CV combination should be used for determining the flip kinetics, if many combinations can generate flips. The doubt is increased when calculating the autocorrelation for χ_2 in residue F45 and calculating $a_0 = 93.4$ and the integral to 84.1 ps, which are different to the values from Table 3.5, indicating that the rate is different, although a more rigorous method of calculating the flip rate is necessary for certainty, such as the method used by Kulkarni and Söderhjelm [8].



(a) F22 **(b)** F45

(c) Y23 **(d)** Y35

Figure 3.12: χ_2 versus frame for the simulations with bias on χ_2 and CN. There is frequent flipping in all residues and F45 converged.

4 Conclusions and Future Work

Six out of the seven studied rings have positive volume differences associated with the transition from main state through transition state 1, but the transitions associated with intermediate states can be problematic. The ring cavity volumes were shown to be compressible, and the activation volume compressibility, transition state compressibility and the contribution to the free energy barrier as a result of increased pressure under isothermal conditions were determined, with the two former being on the same order of magnitude as previous experimental determinations, however, they were under variable pressure and temperature. There is quite little autocorrelation for the volumes determined with Epock, and solvated residues seem to have the lowest. Biasing a coordination number based on ring atoms and every carbon atom can generate ring flips, and biasing the coordination number and χ_2 together can generate a converged trajectory with frequent flips for residue F45.

Future work may include further verification of the transition state ensemble, ensuring that there are no ensembles within the ensembles that are the actual transition states. New methods could be implemented for the purpose of determining this, such as transition path sampling [25]. The methods presented here could also be applied on other proteins which have experimental data on ring flips, such as GB1, as well as other proteins of scientific interest. For instance, the PET degrading enzyme PETase has a conserved phenylalanine in its active site [26] and there are many aromatic rings in the binding pocket of the β1-adrenergic receptor [27]. There are many exciting proteins to be studied indeed. Developing a method to calculate the compressibility coefficient from the volume variance would be an improvement, as this would require less computation. A rigorous method for analyzing if solvation has an impact on water could evaluate the potential of it as a flip rate predictor. The coordination number could be used in the infrequent metadynamics method used to calculate the flip rate [8] It is still impossible to say whether the ring flip in itself drives the expansion of the local volume around it or if local volume is created by some other process, allowing for the flip to occur, and developing a method that gives a definitive answer remains a challenge.

Bibliography

[1] Stephani Joy Y. Macalino, Vijayakumar Gosu, Sunhye Hong and Sun Choi. "Role of computer-aided drug design in modern drug discovery". In: *Archives of Pharmacal Research* 38.9 (2015), 1686–1701. DOI: 10.1007/s12272-015-0640-5.

[2] Bertil Halle. "Protein dynamics". In: *Biophysical Chemistry*. Lund University, 2020, 9:1–9:29.

[3] K. Wüthrich and G. Wagner. "NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor". In: *FEBS Letters* 50.2 (1975), 265–268. DOI: 10.1016/0014-5793(75)80504-7.

[4] J. Andrew McCammon, Bruce R. Gelin and Martin Karplus. *Dynamics of folded proteins*. URL: https://www.nature.com/articles/267585a0.

[5] Matthias Dreydoppel, Britta Dorn, Kristofer Modig, Mikael Akke and Ulrich Weininger. "Transition-state compressibility and activation volume of transient protein conformational fluctuations". In: *JACS Au* 1.6 (2021), 833–842. DOI: 10.1021/jacsau.1c00062.

[6] Ulrich Weininger, Kristofer Modig and Mikael Akke. "Ring flips revisited: 13C relaxation dispersion measurements of aromatic side chain dynamics and activation barriers in basic pancreatic trypsin inhibitor". In: *Biochemistry* 53.28 (2014), 4519–4525. DOI: 10.1021/bi500462k.

[7] Laurence Lins and Robert Brasseur. "The hydrophobic effect in protein folding". In: *The FASEB Journal* 9.7 (1995), 535–540. DOI: 10.1096/fasebj.9.7.7737462.

[8] Mandar Kulkarni and Pär Söderhjelm. "Free energy landscape and rate estimation of the aromatic ring flips in basic pancreatic trypsin inhibitor using metadynamics". In: (2021). DOI: 10.1101/2021.01.07.425261.

[9] J. A. McCammon and M. Karplus. "Dynamics of activated processes in globular proteins." In: *Proceedings of the National Academy of Sciences* 76.8 (1979), 3585–3589. DOI: 10.1073/pnas.76.8.3585.

[10] Alessandro Barducci, Massimiliano Bonomi and Michele Parrinello. "Metadynamics". In: *WIREs Computational Molecular Science* 1.5 (2011), 826–843. DOI: 10.1002/wcms.31.

[11] R. Huber, D. Kukla, A. Ruehlmann, O. Epp, H. Formanek, J. Deisenhofer and W. Steigemann. "Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 angstroms resolution". In: B31 (1975), 238–250. DOI: 10.2210/pdb3pti/pdb.

[12] Protein Information Resource SIB Swiss Institute of Bioinformatics UniProt Consortium European Bioinformatics Institute. *Pancreatic trypsin inhibitor*. 2021. URL: https://www.uniprot.org/uniprot/P00974.

[13] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser and Carlos Simmerling. "FF14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb". In: *Journal of Chemical Theory and Computation* 11.8 (2015), 3696–3713. DOI: 10.1021/acs.jctc.5b00255.

[14] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffry D. Madura, Thomas J. Dick, Greg L. Hura and Teresa Head-Gordon. "Development of an improved four-site water model for biomolecular simulations: TIP4P-EW". In: *The Journal of Chemical Physics* 120.20 (2004), 9665–9678. DOI: 10.1063/1.1683075.

[15] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess and Erik Lindahl. "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (2015), 19–25. DOI: 10.1016/j.softx.2015.06.001.

[16] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni and Giovanni Bussi. "Plumed 2: New feathers for an old bird". In: *Computer Physics Communications* 185.2 (2014), 604–613. DOI: 10.1016/j.cpc.2013.09.018.

[17] Alessandro Barducci, Giovanni Bussi and Michele Parrinello. "Well-tempered metadynamics: A smoothly converging and tunable free-energy method". In: *Physical Review Letters* 100.2 (2008). DOI: 10.1103/physrevlett.100.020603.

[18] Benoist Laurent, Matthieu Chavent, Tristan Cragnolini, Anna Caroline Dahl, Samuela Pasquali, Philippe Derreumaux, Mark S.P. Sansom and Marc Baaden. "Epock: Rapid Analysis of Protein Pocket Dynamics". In: *Bioinformatics* 31.9 (2014), 1478–1480. DOI: 10.1093/bioinformatics/btu822.

[19] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey and Michael L. Klein. "Comparison of simple potential functions for simulating liquid water". In: *The Journal of Chemical Physics* 79.2 (1983), 926–935. DOI: 10.1063/1.445869.

[20] Yulian Gavrilov. *Private communication.*

[21] Calvin R Chen and George I Makhatadze. "Proteinvolume: Calculating molecular van der waals and void volumes in proteins". In: *BMC Bioinformatics* 16.1 (2015). DOI: 10.1186/s12859-015-0531-2.

[22] Douglas B. Kitchen, Lynne H. Reed and Ronald M. Levy. "Molecular dynamics simulation of solvated protein at high pressure". In: *Biochemistry* 31.41 (1992), 10083–10093. DOI: 10.1021/bi00156a031.

[23] Filip Persson and Bertil Halle. "Compressibility of the protein-water interface". In: *The Journal of Chemical Physics* 148.21 (2018), p. 215102. DOI: 10.1063/1.5026774.

[24] Y. Marcus and G.T. Hefter. "The compressibility of liquids at ambient temperature and pressure". In: *Journal of Molecular Liquids* 73-74 (1997), 61–74. DOI: 10.1016/s0167-7322(97)00057-3.

[25] Hendrik Jung, Roberto Covino, A Arjun, Peter G. Bolhuis and Gerhard Hummer. *Autonomous artificial intelligence discovers mechanisms of molecular self-organization in virtual experiments.* 2021. arXiv: 2105.06673 [physics.chem-ph].

[26] Tobias Fecker, Pablo Galaz-Davison, Felipe Engelberger, Yoshie Narui, Marcos Sotomayor, Loreto P. Parra and César A. Ramírez-Sarmiento. "Active site flexibility as a hallmark for efficient pet degradation by I.nbsp;sakaiensis petase". In: *Biophysical Journal* 114.6 (2018), 1302–1312. DOI: 10.1016/j.bpj.2018.02.005.

[27] Andrew G Leslie, Tony Warne and Christopher G Tate. "Ligand occupancy in crystal structure of 1-adrenergic G protein–coupled receptor". In: *Nature Structural amp; Molecular Biology* 22.12 (2015), 941–942. DOI: 10.1038/nsmb.3130.

Appendix A

Volume distributions

Figures A.1-A.7 shows the volume distributions of the studied residues under ambient pressure. Full page width is necessary to distinguish the different lines, but since this would clog up the report, they were relegated here. The histograms give a more detailed view than ensemble averages, but do not change the discussion reasoning in the discussion.

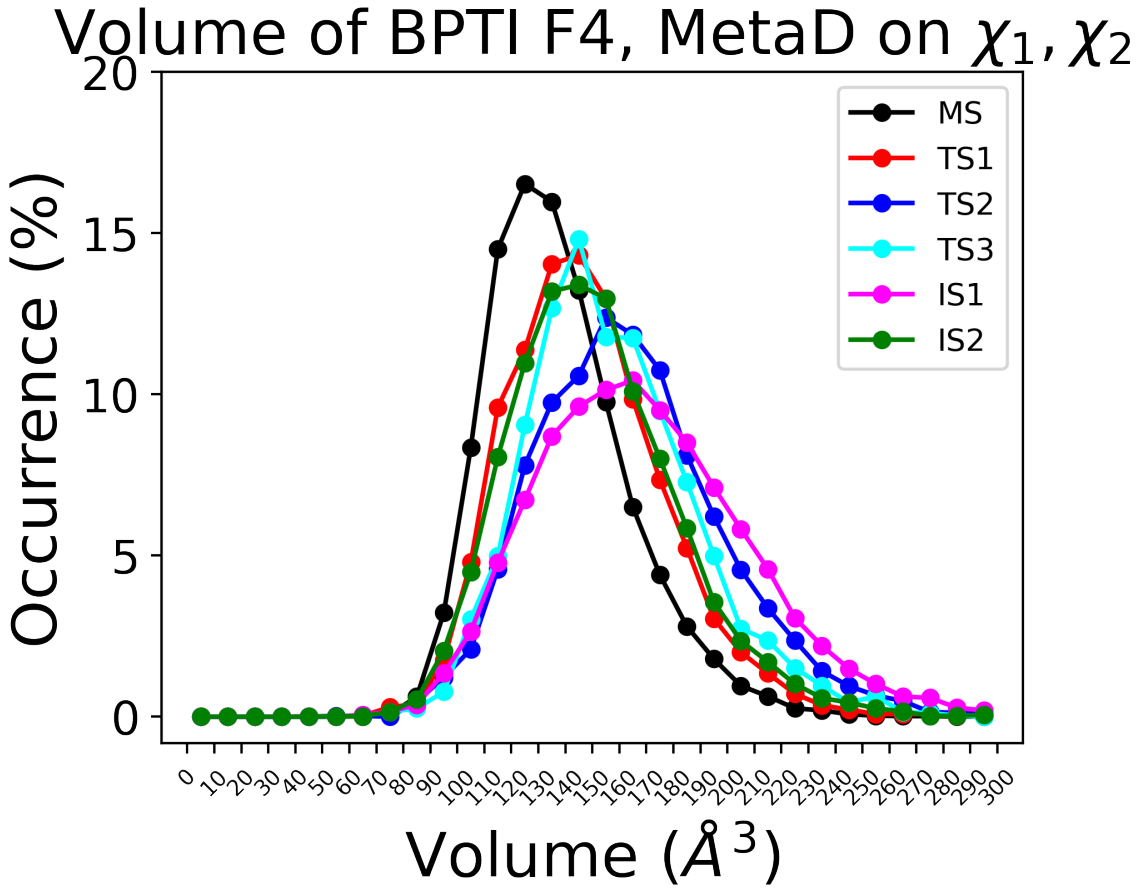


Figure A.1: The volume distributions of the relevant states in residue F4: MS, TS1, TS2, TS3, IS1 and IS2.

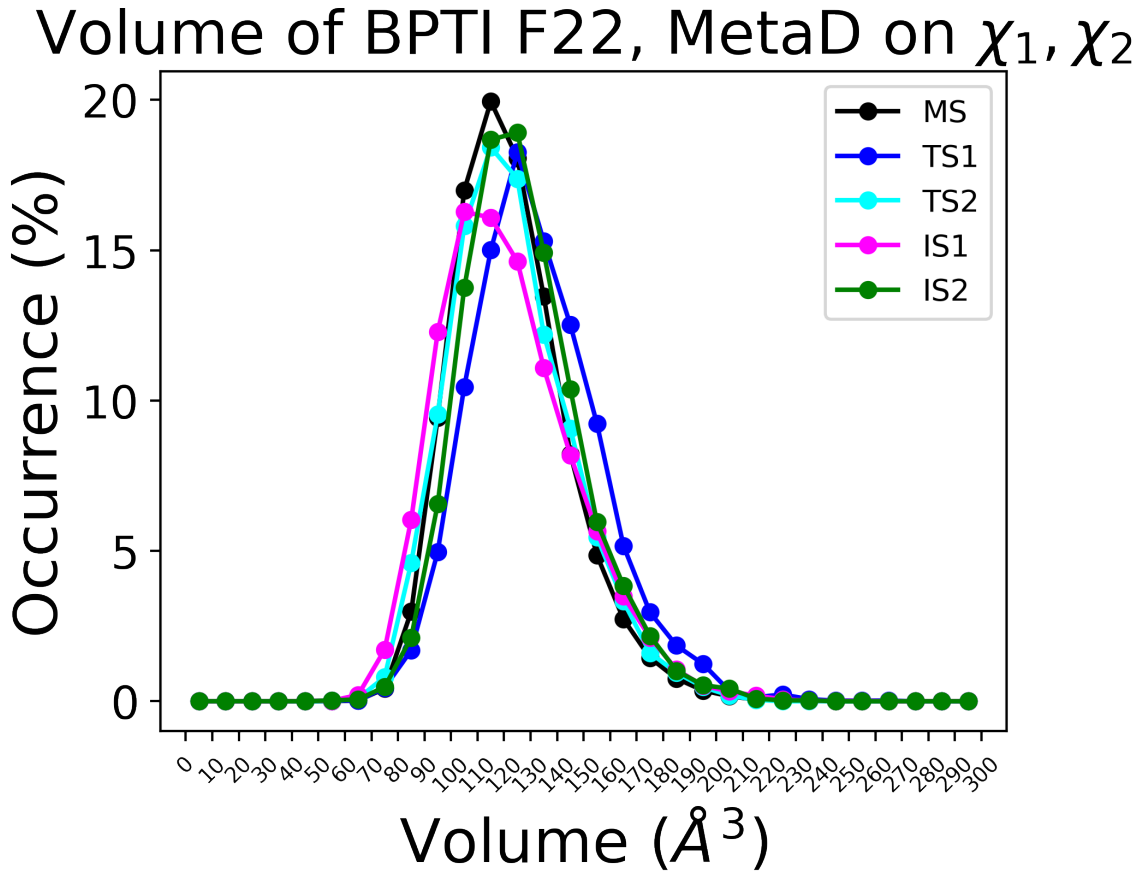


Figure A.2: The volume distributions of the relevant states in residue F22: MS, TS1, TS2, IS1 and IS2.

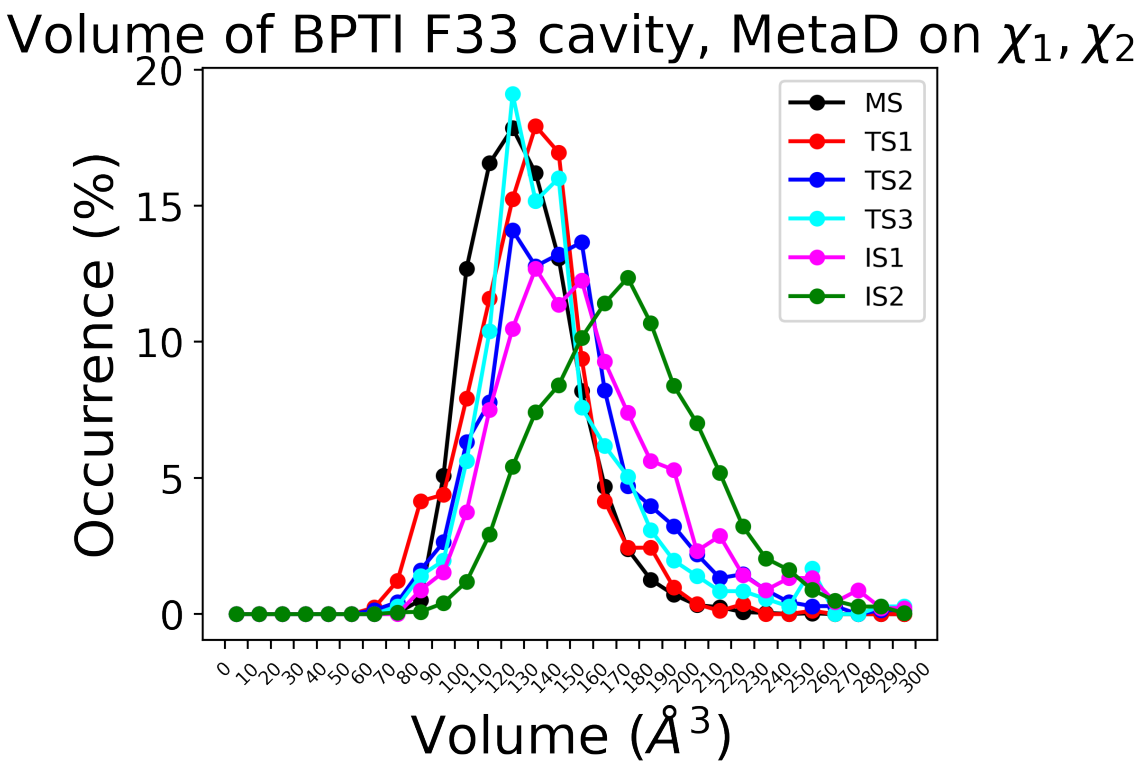


Figure A.3: The volume distributions of the relevant states in residue F33: MS, TS1, TS2, TS3, IS1 and IS2.

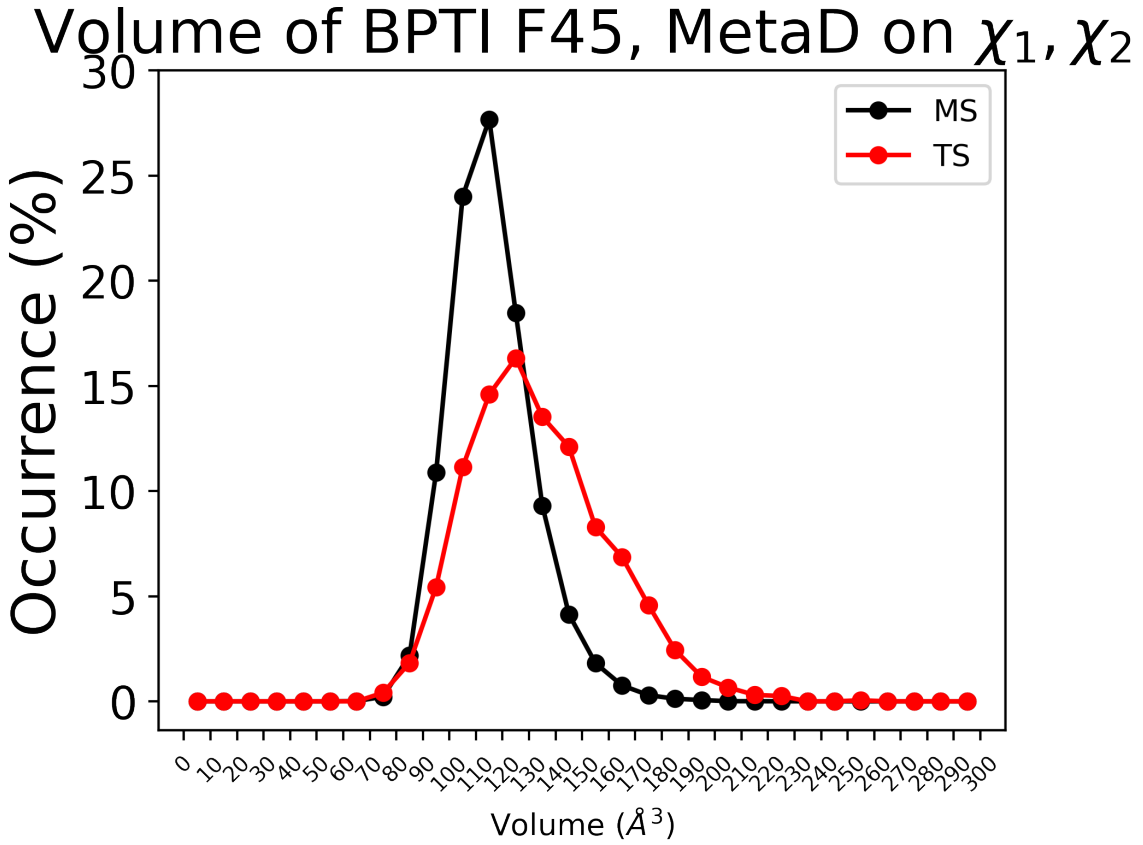


Figure A.4: The volume distributions of the relevant states in residue F45: MS and TS1.

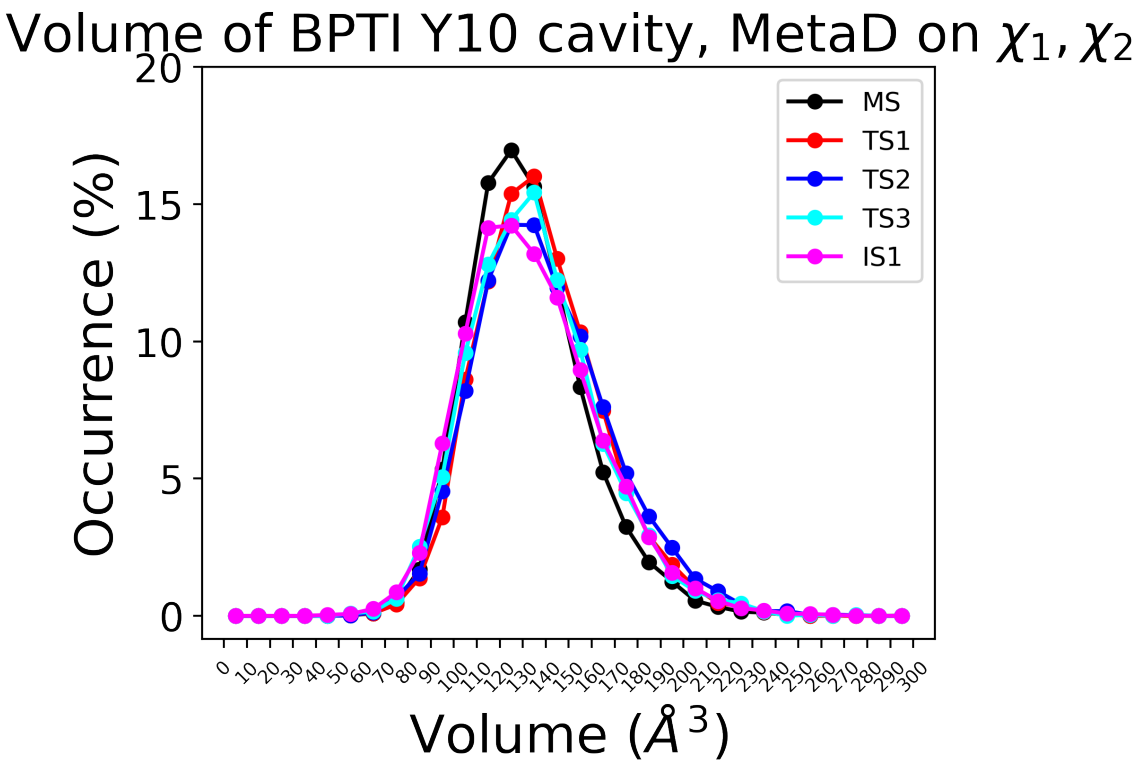


Figure A.5: The volume distributions of the relevant states in residue Y10: MS, TS1, TS2, TS3 and IS1.

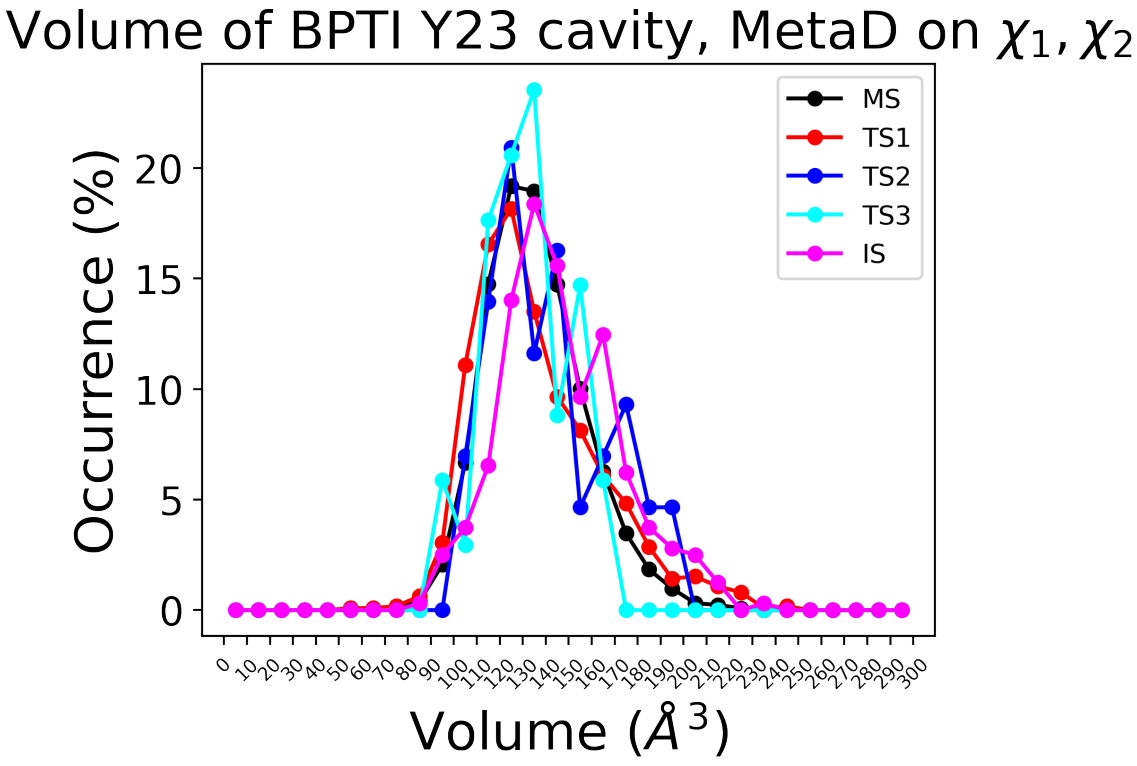


Figure A.6: The volume distributions of the relevant states in residue Y23: MS, TS1, TS2, TS3 and IS1.

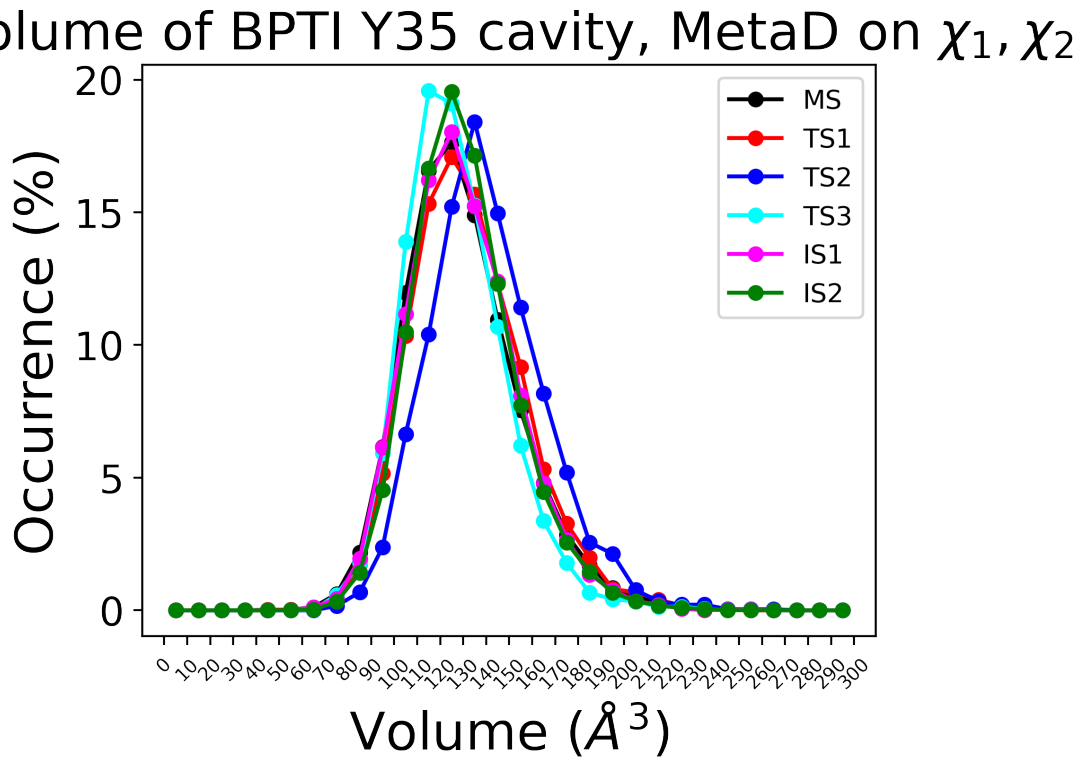


Figure A.7: The volume distributions of the relevant states in residue Y35: MS, TS1, TS2, TS3, IS1 and IS2.