



**UCL**

# **Structural Dynamics of L1 and L2 $\beta$ -lactamases**

## **PhD Thesis**

---

*by*

**Zhuoran Zhao**

**Supervised by Prof Shozeb Haider**

*at*

**School of Pharmacy**

**University College London**

# Plagiarism Statement

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# Abstract

*Stenotrophomonas maltophilia* is a Gram-negative bacterium, found in several different environments, such as soil, water and hospital. It can cause multiple infections but also has strong resistance to many antibiotics such as cephalosporins, carbapenems, and aminoglycosides. *S. maltophilia* confers antibiotic resistance through expression of two different  $\beta$ -lactamases: L1-metallo- $\beta$ -lactamase (L1 MBL) and L2  $\beta$ -lactamase. L1 MBL is a class B3  $\beta$ -lactamase and is the only known tetrameric  $\beta$ -lactamase known to humans. L2 is a class A  $\beta$ -lactamase which has been recently identified.

In L1 MBL, there are, two loops ( $\alpha$ 3- $\beta$ 7 and  $\beta$ 12- $\alpha$ 5) known as the gating loops, that enclose the active site. The “open” and “close” conformations of these two loops were observed in the molecular dynamic simulation. These two conformations allow the gate loops have the ability of controlling the volume of the zinc binding pocket. The pocket size affects the substrate binding and further influence the catalytic activity of the whole protein. Therefore, gating loops are thought to have an important role in substrate binding and catalysis. In this thesis, the dynamics of the gating loops is explored through Markov state models. The “open” and “closed” conformations are defined and three key interactions (salt bridge between R236 and D150c, the  $\pi$ - $\pi$  stacking between H151 and Y227 and the orientation of P225) were identified that play an important role in controlling the conformation of the gating loops. Furthermore, as a tetramer, the correlation between the four subunits was also explored through CVAE-based deep

learning and network analysis. The results revealed a ‘dimer of dimer’ dynamics in L1 MBL.

The second part of the thesis focuses on exploring the dynamics of L2  $\beta$ -lactamase family consisting of L2a, L2b, L2c and L2d enzymes. Homology modelling, MDLoFit, Markov state models, dynamic cross-correlation analysis and CVAE-based deep learning were employed for identifying potential key interactions and dynamic correlations between each subtype. Two dynamic combinations regions were revealed ( $\alpha$ 1 helix/N-terminal,  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop, hinge region, and C terminal,  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 8- $\beta$ 9 loop) which exist in all four L2  $\beta$ -lactamases. Stabilising these two combinations could possibly help inhibit the function of L2  $\beta$ -lactamases. Besides, several potential key residues which result in high dynamic regions were also identified. Since very few research targeted on L2  $\beta$ -lactamases, this work could be a starting point for the following-up work. The improved understanding of the dynamics of L1 and L2  $\beta$ -lactamases will help in the design of their inhibitors and discovery of novel resistance breakers.



# Impact Statement

In this thesis, I present the enhanced understanding of the dynamic mechanism of L1 and L2  $\beta$ -lactamases. L1 and L2  $\beta$ -lactamases together form the antibiotic resistance of *Stenotrophomonas maltophilia* (*S. maltophilia*). *S. maltophilia* is a nosocomial pathogen, which mainly infects immunocompromised hosts including cancer, transplant and cystic fibrosis patients. It can cause respiratory tract infection, along with clinical syndromes, such as catheter-associated and bloodstream. The dynamics of L1 and L2  $\beta$ -lactamases are still unclear. Understanding the dynamics pattern of these enzymes will help the investigation of L1 and L2  $\beta$ -lactamases inhibitors.

In L1  $\beta$ -lactamase work, I used adaptive sampling simulation method base on the ACEMD molecular dynamics engine. Markov State Models, CVAE-based deep learning and network analysis were employed to construct kinetic models and proposed dynamic mechanism. The “open” and “closed” conformations of L1  $\beta$ -lactamase has never been reported before. The reveal of these two conformations of gating loops will help the understanding of the L1  $\beta$ -lactamase binding pocket. In most case of L1  $\beta$ -lactamase dynamics, the gating loop stay in “closed” state. The “closed” state limits the volume of binding pocket and stop the substrates to get access to the binding site. By identifying the important interactions which control the gating loop conformations, it is possible to stabilize these loops into the “open” conformation. The “open” conformation of L1  $\beta$ -lactamase zinc binding site have a significant higher volume

which can allow much bigger substrate to get inside. The reveal of “open” and “closed” conformations not only improve the understanding of the L1  $\beta$ -lactamase dynamics but also give more possibility for developing L1  $\beta$ -lactamase inhibitors. Furthermore, a dimer-of-dimers dynamic mechanism was also found in L1  $\beta$ -lactamase tetramer. The conserve residues of signal transferring from one binding site to another were also identified through network analysis. L1  $\beta$ -lactamase tetramer has much higher catalytic activity compare with the monomer. Therefore, affect the signal transfer between each subunit in the tetramer can potentially make four subunits working individually and reduce the catalytic ability of L1  $\beta$ -lactamase. The dimer-of-dimers dynamic mechanism could help in the design of L1  $\beta$ -lactamase inhibitors.

In L2  $\beta$ -lactamase work, adaptive sampling simulations were carried out based on the homology modelling results. MDLofit, Markov State Models, Dynamic Cross-correlation and CVAE-based deep learning were employed for the analysis of L2  $\beta$ -lactamase dynamic mechanism. As no molecular dynamic research have been done before on L2  $\beta$ -lactamases, this work will help to improve the understanding of L2  $\beta$ -lactamases in this field. There are two dynamic combinations regions were found in this work:  $\alpha$ 1 helix/N-terminal,  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop, hinge region, and C terminal,  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 8- $\beta$ 9 loop. These two combinations exist in all four kinds of L2  $\beta$ -lactamases. Therefore, these two combinations could be important for the full function of L2  $\beta$ -lactamases. Stabilizing these two regions may help inhibit the enzymes. Furthermore, several key residues which responsible for the high dynamic regions were

also indicated. The conformational change of these residues could be important for the whole L2  $\beta$ -lactamase dynamics pattern. The findings can help the future work on the development of novel inhibitors of L2  $\beta$ -lactamase.

L1 and L2  $\beta$ -lactamases together form the antibiotic resistance of *S. maltophilia*. These two  $\beta$ -lactamases combined and produce antibiotic resistance to almost all the  $\beta$ -lactam antibiotics. This work focus on the dynamics research for L1 and L2  $\beta$ -lactamases. By identifying the important combinations, regions and key residues, new direction of developing  $\beta$ -lactamase inhibitors could be revealed. The importance of this work is to explore the dynamic pattern of L1 and L2  $\beta$ -lactamases and reveal the mechanisms and potential targets that never reported before.

# UCL Research Paper Declaration Form

## referencing the doctoral candidate's own published work(s)

*Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:*

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

*This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.*

### 1. For a research manuscript that has already been published (if not yet published, please skip to section 2)

a) **What is the title of the manuscript?**

**Gating interactions steer loop conformational changes in the active site of the L1 metallo- $\beta$ -lactamase**

b) **Please include a link to or doi for the work**

**<https://elifesciences.org/articles/83928>**

c) **Where was the work published?**

**eLife**

d) **Who published the work? (e.g. OUP)**

**eLife**

e) **When was the work published?**

**Feb 24, 2023**

f) **List the manuscript's authors in the order they appear on the publication**

**Zhuoran Zhao, Xiayu Shen, Shuang Chen, Jing Gu, Haun Wang, Maria F Mojica, Moumita Samanta, Debsindhu Bhowmik, Alejandro J Vila, Robert A Bonomo, Shozeb Haider**

g) **Was the work peer reviewed?**

**Yes**

h) **Have you retained the copyright?**

**Yes**

i) **Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv). If 'Yes', please give a link or doi)**

<https://www.biorxiv.org/content/10.1101/2022.10.25.513716v1>

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

*I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

**2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**

**a) What is the current title of the manuscript?**

Click or tap here to enter text.

**b) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)**

Click or tap here to enter text.

**c) Where is the work intended to be published? (e.g. journal names)**

Click or tap here to enter text.

**d) List the manuscript's authors in the intended authorship order**

Click or tap here to enter text.

**e) Stage of publication (e.g. in submission)**

Click or tap here to enter text.

**3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)**

Zhuoran Zhao - Running the simulations, building MSMs, Deep Learning, Comparison, Discussions

Xiayu Shen – Initial concept of metallo-betalactamases

Shuang Chen – Deep learning

Jing Gu – pocket similarity search

Haun Wang – Interaction energies

Maria F Mojica – MIC experiments

Moumita Samanta – development of Deep learning code

Debsindhu Bhowmik – Supervision of deep learning code

Alejandro J Vila – Supervision of metallo-betalactamases

Robert A Bonomo – Supervision of experimental data

Shozeb Haider – Overall design of study, supervision

**4. In which chapter(s) of your thesis can this material be found?**

Chapter 1 Introduction

Chapter 2.1 Metallo- $\beta$ -lactamase L1 Hypothesis

Chapter 3 Methods

Chapter 4 Metallo- $\beta$ -lactamase Result

Chapter 5 L1 $\beta$ -lactamase discussion

Chapter 8.1 L1 Metallo- $\beta$ -lactamase Discussion

**5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)**

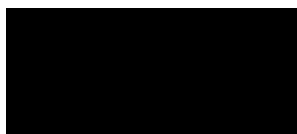
*Candidate*



*Date:*

**June 8th 2023**

*Supervisor/ Senior Author (where appropriate)*



*Date:* **June 8<sup>th</sup>, 2023**

# **Acknowledgements**

I want to express my gratitude to my parents first, they support me all the time; the love from them always supports me in continuing my work. Besides, my supervisor Professor Shozeb Haider helped me a lot, without whom this project would not have been possible and whose expertise, understanding and patience contributed considerably to my graduate experience.

# Table of Contents

<b>Plagiarism Statement.....</b>	<b>2</b>
<b>Abstract.....</b>	<b>3</b>
<b>Impact Statement.....</b>	<b>5</b>
<b>UCL Research Paper Declaration Form.....</b>	<b>8</b>
<b>Acknowledgements .....</b>	<b>11</b>
<b>Table of Contents .....</b>	<b>12</b>
<b>List of Figures and Tables.....</b>	<b>16</b>
<b>List of Abbreviations .....</b>	<b>24</b>
<b>1. Introduction.....</b>	<b>25</b>
<b>1.1 Antibiotics History .....</b>	<b>25</b>
<b>1.2 Mechanisms of Action of Antibiotics.....</b>	<b>28</b>
<b>Antibiotics inhibit cell wall synthesis .....</b>	<b>28</b>
<b>Antibiotics inhibit protein synthesis.....</b>	<b>29</b>
<b>Antibiotics inhibit DNA replication .....</b>	<b>30</b>
<b>1.3 Antibiotic Resistance .....</b>	<b>31</b>
<b>1.4 Mechanisms of Antibiotic Resistance.....</b>	<b>34</b>
<b>Chemical modification of antibiotics.....</b>	<b>34</b>
<b>Preventing antibiotics from reaching the target .....</b>	<b>34</b>
<b>Modification of the antibiotic binding sites .....</b>	<b>35</b>
<b>Global cell adaptations resistance.....</b>	<b>35</b>



1.5 $\beta$ -lactam and $\beta$ -lactamases .....	37
1.6 L1 and L2 $\beta$ -lactamases .....	47
L1 metallo- $\beta$ -lactamase.....	47
L2 $\beta$ -lactamases .....	59
2. Hypothesis and Objectives .....	63
2.1 L1 Metallo- $\beta$ -lactamase Hypothesis .....	63
2.2 L2 $\beta$ -lactamases Hypothesis .....	65
3. Methods.....	67
3.1 Molecular Dynamics .....	67
Force field .....	69
Periodic boundary conditions .....	75
Ensembles .....	75
3.2 Homology Modelling.....	80
3.3 Markov State Models (MSM) .....	83
3.4 CVAE-based Deep Learning.....	86
4. L1 Metallo- $\beta$ -lactamase Results.....	89
4.1 Set up and Running the Simulations.....	89
4.2 Markov State Models.....	92
4.3 CVAE-Based Deep Learning .....	102
4.4 Network Analysis .....	109
A-B paths .....	110
A-C paths .....	111

A-D paths .....	112
B-C paths .....	113
B-D paths .....	114
C-D paths .....	115
5. L1 $\beta$ -lactamase Discussion.....	116
“Open” and “closed” conformations.....	116
Dimer-of-dimers .....	130
6. L2 $\beta$ -lactamase Results .....	136
6.1 Homology Modelling.....	136
6.2 Set up and Running the Simulation .....	139
6.3 MDLofit Results .....	141
L2a $\beta$ -lactamase results .....	142
L2b $\beta$ -lactamase results .....	143
L2c $\beta$ -lactamase results .....	144
L2d $\beta$ -lactamase results .....	145
6.4 Markov State Models.....	146
L2a $\beta$ -lactamase results .....	147
L2b $\beta$ -lactamase results .....	153
L2c $\beta$ -lactamase results .....	159
L2d $\beta$ -lactamase results .....	165
6.5 Dynamic Cross-correlation Analysis.....	171
L2a $\beta$ -lactamase results .....	172

L2b $\beta$ -lactamase results .....	174
L2c $\beta$ -lactamase results .....	175
L2d $\beta$ -lactamase results .....	176
6.6 CVAE-based Deep Learning of L2 $\beta$ -lactamases.....	177
7. L2 $\beta$ -lactamases Discussion .....	187
L2a $\beta$ -lactamase discussion .....	188
L2b $\beta$ -lactamase discussion .....	196
L2c $\beta$ -lactamase discussion .....	199
L2d $\beta$ -lactamase discussion .....	205
Combined L2 $\beta$ -lactamase discussion .....	212
8. Conclusion .....	215
8.1 L1 Metallo- $\beta$ -lactamase Conclusions .....	215
8.2 L2 $\beta$ -lactamase Conclusions .....	218
8.3 Summary and Future Work .....	219
9. Reference .....	222

# List of Figures and Tables

<b>Figure 1:</b> The mechanisms of action of antibiotic. ....	29
<b>Figure 2:</b> The chemical structure of Penicillin. ....	38
<b>Figure 3:</b> Phylogenetic tree.....	39
<b>Figure 4:</b> Representative structures and their binding sites from the four classes of $\beta$ -lactamase .....	40
<b>Figure 5:</b> The proposed mechanism of catalysis for serine $\beta$ -lactamases (class A, C and D). ....	43
<b>Figure 6:</b> The catalytic mechanism of L1 metallo- $\beta$ -lactamase .....	44
<b>Figure 7:</b> The structure of L1 metallo- $\beta$ -lactamase. ....	49
<b>Figure 8:</b> The M175-A side chain and the hydrophobic pocket on subunit B surface. ....	50
<b>Figure 9:</b> The N-terminal extensions of subunit A and subunit C.....	52
<b>Figure 10:</b> L27, L30 and the surface cavity residues.....	52
<b>Figure 11:</b> The backbone of P166 (A-subunit) interacts with the side chain of R148 (C-subunit). ....	53
<b>Figure 12:</b> The interactions between subunit A and subunit D. ....	54
<b>Figure 13:</b> The zinc binding site of L1 metallo- $\beta$ -lactamase.....	57
<b>Figure 14:</b> The two elongated loops ( $\alpha_3$ - $\beta_7$ and $\beta_{12}$ - $\alpha_5$ ) form the gate entrance to the active site. ....	58
<b>Figure 15:</b> Sequence information of L2 $\beta$ -lactamase.....	61

<b>Figure 16:</b> The alignment of L2b $\beta$ -lactamase crystal structure (PDB number 5NE2) and a typical class A $\beta$ -lactamase (KPC-2 $\beta$ -lactamase, PDB number 2OV5) .....	<b>62</b>
<b>Figure 17:</b> Representation of different forms of potential energy in the system. ....	<b>70</b>
<b>Figure 18:</b> three approaches of incorporating metal atoms into force fields. ....	<b>72</b>
<b>Figure 19:</b> Diagrammatic representation of MSM. ....	<b>84</b>
<b>Figure 20:</b> Autoencoder architecture. ....	<b>87</b>
<b>Figure 21:</b> Different protonation type of histidine .....	<b>91</b>
<b>Figure 22:</b> Implied time scale plot of L1 $\beta$ -lactamase. ....	<b>93</b>
<b>Figure 23:</b> Chapman-Kolmogorov test plot of L1 $\beta$ -lactamase.....	<b>95</b>
<b>Figure 24:</b> L1 $\beta$ -lactamase energy landscape .....	<b>96</b>
<b>Figure 25:</b> The PCCA distribution plot for 7 states in L1 $\beta$ -lactamase .....	<b>97</b>
<b>Figure 26:</b> Net flux pathways plot of L1 $\beta$ -lactamase. ....	<b>100</b>
<b>Figure 27:</b> Evaluation of information loss over latent dimensions.....	<b>103</b>
<b>Figure 28:</b> Evaluation of information loss of different latent dimensions by epoch number. ....	<b>104</b>
<b>Figure 29:</b> Comparison of the original input file and reconstructed decoded file. ....	<b>104</b>
<b>Figure 30:</b> General conformation distribution plot for all 7 metastable states. ....	<b>107</b>
<b>Figure 31:</b> All 7 metastable states of L1 MBL conformation represented individually distribution plots. ....	<b>108</b>
<b>Figure 32:</b> The detailed view of paths from subunit A binding site to subunit B binding site. ....	<b>110</b>
<b>Figure 33:</b> Diagrammatic representation of the paths from subunit A binding site to	

subunit B binding site.....	110
<b>Figure 34:</b> The detailed view of paths from subunit A binding site to subunit C binding site. ....	111
<b>Figure 35:</b> Diagrammatic presentation of the paths from subunit A binding site to subunit C binding site.....	111
<b>Figure 36:</b> The detailed view of paths from subunit A binding site to subunit D binding site. ....	112
<b>Figure 37:</b> Diagrammatic representation of the paths from subunit A binding site to subunit D binding site.....	112
<b>Figure 38:</b> The detailed view of paths from subunit B binding site to subunit C binding site. ....	113
<b>Figure 39:</b> Diagrammatic representation of the paths from subunit B binding site to subunit C binding site.....	113
<b>Figure 40:</b> The detailed view of paths from subunit B binding site to subunit D binding site. ....	114
<b>Figure 41:</b> Diagram representation of the paths from subunit B binding site to subunit D binding site.....	114
<b>Figure 42:</b> The detailed view of paths from subunit C binding site to subunit D binding site. ....	115
<b>Figure 43:</b> Diagram representation of the paths from subunit C binding site to subunit D binding site.....	115
<b>Figure 44:</b> The position of the salt bridge, the $\pi$ - $\pi$ stacking, P225 and zinc ions.	117

<b>Figure 45:</b> Different status of salt bridge, $\pi$ - $\pi$ stacking and proline. ....	119
<b>Figure 46:</b> cis and tans configuration of proline. The $\omega$ angle is highlighted. ....	121
<b>Figure 47:</b> The $\omega$ angle analysis of P225 in all four subunits. ....	121
<b>Figure 48:</b> The interaction between A224 and Q310. ....	123
<b>Figure 49:</b> The dynamics of the F156 side chain .....	125
<b>Figure 50:</b> Superimposed structures of a state 7 monomer (red), a monomer from the apo state crystal structure (yellow, PDB entry 1SML) and a monomer from the inhibitor binding complex (green, PDB entry 5DPX). ....	126
<b>Figure 51:</b> CVAE-based deep learning results for clustering of D150c-R236 salt bridge, H151-Y227 $\pi$ - $\pi$ stacking and the in/out conformation of P225. ....	129
<b>Figure 52:</b> A comparison of the subunits A-B paths with subunits C-D paths. ....	132
<b>Figure 53:</b> The comparison of subunits A-C paths with subunits B-D paths. ....	133
<b>Figure 54:</b> The comparison of subunit A-D paths with subunit B-C paths. ....	134
<b>Figure 55:</b> The relationship of each subunit in L1 MBL. Subunit A (B) and D (C) are adjacent subunits. A-B, A-C, B-D and C-D are neighbour subunits. ....	135
<b>Figure 56:</b> Sequence alignment of L2a, L2B, L2c and L2d $\beta$ -lactamase .....	136
<b>Figure 57:</b> Ramachandran plots of L2a, L2c and L2d models. ....	138
<b>Figure 58:</b> 25% of $C\alpha$ RMSD was more than 0.9Å in L2a $\beta$ -lactamase. ....	142
<b>Figure 59:</b> MDLovofit result of L2a $\beta$ -lactamase.....	142
<b>Figure 60:</b> 3% of $C\alpha$ RMSD was more than 0.9Å in L2b $\beta$ -lactamase. ....	143
<b>Figure 61:</b> MDLovofit results of L2b $\beta$ -lactamase.....	143
<b>Figure 62:</b> 4% of $C\alpha$ RMSD was more than 0.9Å in L2c $\beta$ -lactamase. ....	144

<b>Figure 63:</b> MDLofit result of L2c $\beta$ -lactamase.....	144
<b>Figure 64:</b> 13% of C $\alpha$ RMSD was more than 0.9Å in L2d $\beta$ -lactamase. ....	145
<b>Figure 65:</b> MDLofit results of L2d $\beta$ -lactamase.....	145
<b>Figure 66:</b> Implied time scale plot for L2a $\beta$ -lactamase.....	147
<b>Figure 67:</b> Chapman-Kolmogorov test plot for L2a $\beta$ -lactamase.....	148
<b>Figure 68:</b> PCCA distribution plot for 5 states in L2a $\beta$ -lactamase .....	149
<b>Figure 69:</b> L2a $\beta$ -lactamase free energy landscape at lag time 1ns.....	150
<b>Figure 70:</b> Net flux pathway plot of L2a $\beta$ -lactamase.....	151
<b>Figure 71:</b> L2b $\beta$ -lactamase implied timescale plot. ....	153
<b>Figure 72:</b> Chapman-Kolmogorov test plot for L2b $\beta$ -lactamase. ....	154
<b>Figure 73:</b> L2b $\beta$ -lactamase PCCA distribution plot.....	155
<b>Figure 74:</b> L2b $\beta$ -lactamase free energy landscape at lag time 4ns.....	156
<b>Figure 75:</b> L2b $\beta$ -lactamase net flux pathway plot.....	157
<b>Figure 76:</b> L2c $\beta$ -lactamase implied timescale plot.....	159
<b>Figure 77:</b> Chapman-Kolmogorov test plot for L2c $\beta$ -lactamase.....	160
<b>Figure 78:</b> L2c $\beta$ -lactamase PCCA distribution plot.....	161
<b>Figure 79:</b> L2c $\beta$ -lactamase free energy landscape at lag time 20ns. ....	162
<b>Figure 80:</b> Net flux pathway plot of L2c $\beta$ -lactamase. ....	163
<b>Figure 81:</b> L2d $\beta$ -lactamase implied timescale plot .....	165
<b>Figure 82:</b> Chapman-Kolmogorov test plot for L2d $\beta$ -lactamase .....	166
<b>Figure 83:</b> L2d $\beta$ -lactamase PCCA distribution plot.....	167
<b>Figure 84:</b> L2d $\beta$ -lactamase free energy landscape at lag time 2ns.....	168



<b>Figure 85:</b> L2d $\beta$ -lactamase net flux pathway plot.....	<b>169</b>
<b>Figure 86:</b> Cross-correlation analysis results for L2a $\beta$ -lactamase .....	<b>173</b>
<b>Figure 87:</b> Cross-correlation analysis results for L2b $\beta$ -lactamase. ....	<b>174</b>
<b>Figure 88:</b> Cross-correlation analysis results for L2c $\beta$ -lactamase. ....	<b>175</b>
<b>Figure 89:</b> Cross-correlation analysis results for L2d $\beta$ -lactamase .....	<b>176</b>
<b>Figure 90:</b> Evaluation of information loss over latent dimensions.....	<b>178</b>
<b>Figure 91:</b> Evaluation of information loss of different latent dimensions by epoch number. ....	<b>179</b>
<b>Figure 92:</b> Comparison of the original input file and reconstructed decoded file.	<b>179</b>
<b>Figure 93:</b> Conformational distribution in 2D plot clustered using the t-SNE method. .....	<b>181</b>
<b>Figure 94:</b> Free energy distribution plot.....	<b>182</b>
<b>Figure 95:</b> Superimposed structures extracted from different low energy minim.	<b>184</b>
<b>Figure 96:</b> Comparison of extracted structures from the same protein. ....	<b>186</b>
<b>Figure 97:</b> RMSF plot for 5 states in L2a $\beta$ -lactamase.....	<b>188</b>
<b>Figure 98:</b> Distance between S21 and Q273 .....	<b>190</b>
<b>Figure 99:</b> The position of $\alpha$ 1, $\beta$ 7- $\beta$ 8 loop, $\beta$ 9- $\alpha$ 15 loop, T216, R243 and Y271 .	<b>192</b>
<b>Figure 100:</b> A closeup view of T216, R243 and Y271 position.....	<b>192</b>
<b>Figure 101:</b> The hydrogen bond formation between the side chain of R98 ( $\alpha$ 6) and the backbone of K115 ( $\alpha$ 7). ....	<b>193</b>
<b>Figure 102:</b> C-terminal affect $\beta$ 1- $\beta$ 2 loop and $\beta$ 8- $\beta$ 9 loop dynamics. ....	<b>195</b>
<b>Figure 103:</b> RMSF plot for 9 states of L2b. ....	<b>196</b>

<b>Figure 104:</b> $\phi$ angle rotation of G156 in $\alpha 10$ .....	<b>198</b>
<b>Figure 105:</b> RMSF plot for the 6 states of L2c $\beta$ -lactamase.....	<b>199</b>
<b>Figure 106:</b> $\psi$ angle rotation of G175 on $\Omega$ loop.....	<b>201</b>
<b>Figure 107:</b> The $\psi$ angle rotation of $\beta 7$ - $\beta 8$ loop.....	<b>203</b>
<b>Figure 108:</b> Rotation of D241 $\psi$ angle. ....	<b>204</b>
<b>Figure 109:</b> RMSF plot for 8 states of L2d $\beta$ -lactamase .....	<b>205</b>
<b>Figure 110:</b> The interactions between $\beta 1$ - $\beta 2$ loop, $\beta 2$ and $\beta 2$ - $\beta 3$ loop .....	<b>207</b>
<b>Figure 111:</b> The interactions between $\beta 2$ - $\beta 3$ loop, $\alpha 10$ - $\alpha 11$ loop and $\alpha 12$ . ....	<b>208</b>
<b>Figure 112:</b> The $\psi$ angle rotation occurs on $\alpha 14$ - $\beta 7$ loop in residue G227. ....	<b>210</b>
<b>Figure 113:</b> Rotation of the $\psi$ dihedral angle of residue G227.....	<b>211</b>
<b>Figure 114:</b> RMSF plot for L2a, L2b, L2c and L2d $\beta$ -lactamases. ....	<b>213</b>

**Table 1:** Mean first passage times between different states in L1  $\beta$ -lactamase.....**101**

**Table 2:** The population of different states ( $\pi$ ) and their free energy estimation. .**101**

**Table 3:** Mean first passage times between different states in L2a  $\beta$ -lactamase. ..**152**

**Table 4:** L2a  $\beta$ -lactamase 5 states population ( $\pi$ ) and their free energy estimation.  
.....**152**

**Table 5:** Mean first passage times between different states in L2b  $\beta$ -lactamase...**158**

**Table 6:** L2b  $\beta$ -lactamase 9 states population ( $\pi$ ) and free energy estimation. ....**158**

**Table 7:** Mean first passage times between different states in L2c  $\beta$ -lactamase. ..**164**

**Table 8:** L2c 6 states population ( $\pi$ ) and free energy estimation.....**164**

**Table 9:** Mean first passage times between different states in L2d  $\beta$ -lactamase...**170**

**Table 10:** L2d  $\beta$ -lactamase 8 states population ( $\pi$ ) and free energy estimation. ...**170**

# List of Abbreviations

## List of Abbreviation

Abbreviation	Definition
DD	D-alanyl-alanine
PBPs	penicillin binding proteins
RNA	Ribonucleic acid
mRNA	messenger Ribonucleic acid
DNA	Deoxyribonucleic acid
WHO	World Health Organization
US	The United States
CDC	Center for Disease Control and Prevention
MRSA	Methicillin resistant Staphylococcus aureus
UK	The United Kingdom of Great Britain and Northern Ireland
AMR	advanced microbial resistance
P.aeruginosa	Pseudomonas aeruginosa
OprD	outer membrane protein
E.coli	Escherichia coli
Tet(O)	elongation factor-like protein
Tet(M)	ribosomal protection protein
SBL	serine $\beta$ -lactamase
MBL	metallo- $\beta$ -lactamase
S. maltophilia	Stenotrophomonas maltophilia
PDB	Protein Data Bank
MSM	Markov State Models
CVAE	Convolutional Variational Auto Encoders
Amber	Assisted Model Building with Energy Refinement
MD	Molecular dynamics
ZAFF	Zinc AMBER force field
MTK	Modeling ToolKit
MK	Merz-Singh-Kollman
RESP	restrained electrostatic potential
NVT	particles in the system (N), system volume (V) and temperature (T) to be constant
NVE	particles in the system (N), system volume (V) and system total energy (E) to be constant
NPT	particles in the system (N), pressure (P) and temperature (T) to be constant
GPCR	G protein-coupled receptor
ACEMD	Accelerated Molecular Dynamics
3D	Three-dimensional
CK	Chapman-Kolmogorov
PCCA	Perron Cluster-Cluster Analysis
TPT	transition path theory
TST	transition state theory
RMSD	Root Mean Square Deviation
RMSF	root-mean-square fluctuation
NME	N-terminal methyl amide
ACE	C-terminal acetyl
tICA	Time-structure Independent Components Analysis
IC	Independent Component

# 1. Introduction

Antibiotics are one of the most important discoveries in modern history.<sup>1</sup> The benefits of antibiotic usage are not just limited to the medical field, but it is also used in biological science research, agriculture, animal husbandry and the food industry.<sup>2</sup> This makes it an essential product in our everyday lives. However, the misuse of antibiotics over the years has resulted in multi-drug resistance. One of the most common ways of developing resistance is via the expression of the  $\beta$ -lactamase enzyme.

## 1.1 Antibiotics History

An antibiotic is an antimicrobial substance active against bacteria. The first modern antibiotic drug was Salvarsan which was developed in 1910.<sup>3</sup> It was used for treating syphilis and African trypanosomiasis and was the first effective treatment for this disease. However, the history of treatments for preventing infection stretches back to 2000 years ago. People used various natural products, such as animal organs, honey and wood bark. Rome, Egypt, China and Greece have used mouldy bread to treat open wounds.<sup>4</sup> Besides, tetracyclines were also found in skeletons excavated in Nubia during the Rome domination period.<sup>5</sup> In 1676, microorganisms were identified by Antonie van Leeuwenhoek through his microscope.<sup>6</sup> However, due to unclear cognition and poor achievements, the findings did not attract real attention during the renaissance and enlightenment periods. Unscientific treatment was still widely administered at that time. The patients' conditions improved temporarily but often caused more serious side

effects. In 1871, penicillium glaucum inhibitory function was identified by Joseph Lister and used for curing nurses' injuries. During the same period, the inhibition of one kind of bacteria by species was also noticed by Louis Pasteur. He and his colleague Jules Francois Joubert reported that *bacillus anthracis* could be inhibited when cultivated with aerobic bacteria. The word "antibiosis" was defined as a result of this relationship, which meant "one living organism kills another to ensure its existence" by Jean Paul Vuillemin in 1889.<sup>2</sup> In the 1890s, Rudolf Emmerich and Oscar Löw found that the bacteria on patients' bandages could prevent infection. They grew this green bacteria in batches and isolated the supernatant as a drug, which is pyocyanase.<sup>7</sup> The treatment with pyocyanase had a mixed success, but it was considered a breakthrough for antibiotic discovery. In 1909, Salvarsan was discovered by Paul Ehrlich and his team and put on the market the year after that. This was the first modern antimicrobial agent and was finally replaced by penicillin in the 1940s.<sup>8</sup>

Penicillin was discovered in 1928 by Alexander Fleming. This was regarded as a milestone in the history of drug discovery.<sup>9</sup> Fleming, by accident, found that a fungus was inhibiting *Staphylococcus* growth. The purified inhibitory product from this fungus was penicillin. However, the industrial production of penicillin through *Penicillium chrysogenum* succeeded only in 1940 by Ernst Chain and Howard Florey.<sup>10</sup> It was for the first time that humans had an effective product for treating bacterial infection. The success of penicillin promoted the development of antibiotics. Selman Waksman performed systematic research on soil bacteria (especially Streptomycetes and

Streptomyces members). Waksman's platform was developed to mark the antagonisms among bacteria and discovered several major antibiotics drugs, such as clavacin, actinomycin, neomycin, and fumigation.<sup>11</sup> Since then, the antibiotic discovery has seen tremendous progress and has grown as a vast field.<sup>11</sup> Some popular antibiotics, including penicillin, polymyxin, quinolone, rifamycin, macrolides and tetracyclines, were discovered during that golden age.<sup>12</sup> The last new class of antibiotics was discovered in the 1980s.<sup>2</sup> The rising antibiotic resistance and low rate of investment returns have turned drug companies to gradually abandon the discovery of new antibiotics.<sup>13</sup> There were 20 pharmaceutical companies that did antibiotic discovery research in the 1980s, while only five companies are now left in this field in 2015.<sup>14</sup>

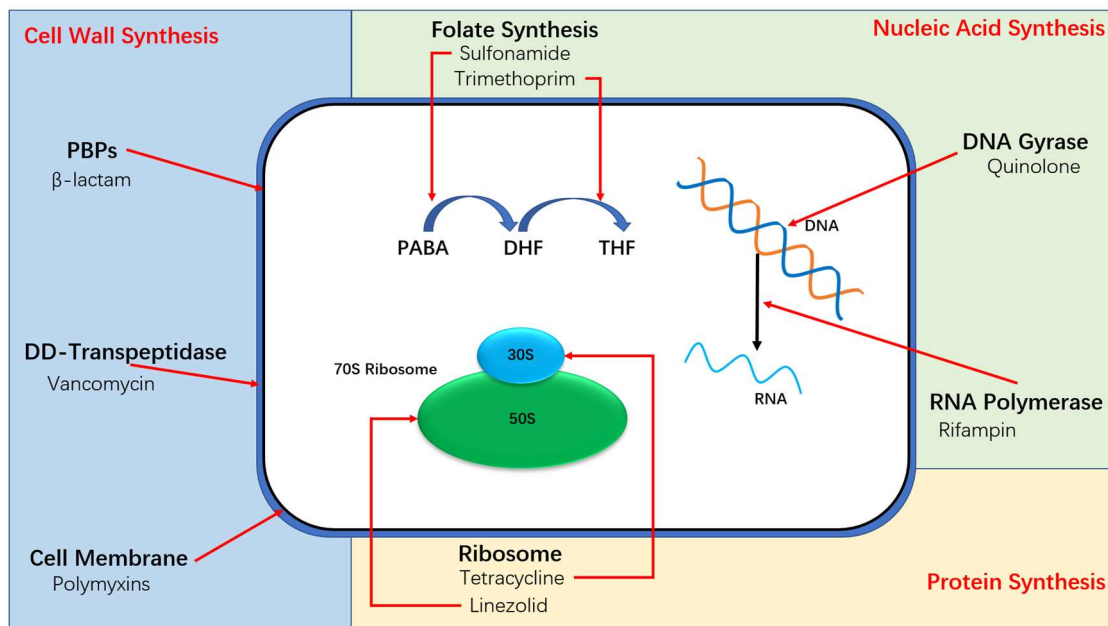
## 1.2 Mechanisms of Action of Antibiotics

Antibiotics can be classified based on their mechanism of action.<sup>15</sup> There are three main classes of antibiotic mechanisms: (1) inhibiting cell wall synthesis (2) inhibiting nucleic acid synthesis (3) inhibiting protein synthesis. (Figure 1)

### **Antibiotics inhibit cell wall synthesis**

Bacteria are surrounded by a cell wall, without which it cannot survive. The cell wall is made of peptidoglycan, a biopolymer consisting of sugars and amino acids. The peptidoglycan crosslinks peptide chains and sugars with the help of the enzymes transglycosidases.<sup>16</sup> The crosslinking makes the cell wall stronger and more stable. The D-alanyl-alanine (DD) portion of peptide requires penicillin binding proteins (PBPs) to crosslink glycine residues.<sup>17</sup>  $\beta$ -lactam and glycopeptides can inhibit the binding of PBPs, which results in the death of bacteria.<sup>18</sup> The  $\beta$ -lactam ring can mimic the DD- peptides and bind to PBPs. After  $\beta$ -lactam binds to PBPs, the new peptidoglycans cannot be crosslinked and synthesised leading to the synthesis of an imperfect cell wall. Glycopeptides antibiotics can bind to DD-peptidases of precursor peptidoglycan and result in the inhibition of PBPs. When the bacteria grows, the loss of new peptidoglycan synthesis will result in bacteria cell wall lysis and death.





**Figure 1:** The mechanisms of action of antibiotic. The main mechanisms of actions by antibiotics can be divided into three main pathways: inhibiting cell wall synthesis, inhibiting nucleic acid synthesis and inhibiting protein synthesis.

## Antibiotics inhibit protein synthesis

Proteins are synthesised in bacteria using messenger RNA (mRNA), on the ribosome via transcription and translation.<sup>18</sup> The ribosome in bacteria is 70S and is composed of 30S and 50S subunits. Antibiotics bind to and inhibit the function of 30S and 50S subunits.<sup>19</sup> Tetracyclines and aminoglycosides can interact with 16S rRNA of 30S subunit. This interaction leads to premature termination and misleading in mRNA translation.<sup>20</sup> Macrolides and chloramphenicol can interact with 23S of 50S subunit and causes premature termination and misleading in mRNA translation.<sup>21</sup>

## **Antibiotics inhibit DNA replication**

DNA replication requires several enzymes' participation, such as DNA helicase, DNA polymerase, DNA ligase and DNA gyrase. Among these enzymes, DNA gyrase is essentially required.<sup>18</sup> During DNA replication, polymerases bound to the helicase hexamer, unwind DNA and forces DNA to rotate.<sup>22</sup> The rotation causes an accumulation of twists in DNA.<sup>23</sup> DNA gyrase is responsible for relieving the tension caused by unwinding the DNA helix and it achieves the goal by adding negative supercoils.<sup>24</sup> The DN gyrase enzyme is a homodimer composed of 2 A subunits and 2 B subunits. During DNA replication, A subunit cuts the DNA strand while B subunit introduces negative supercoils. Then, the A subunit reseals the nicked strand.<sup>20</sup> Fluoroquinolones have high affinity binding to the A subunit of DNA gyrase and reduces its cutting and resealing efficiency. These further influence protein synthesis in bacteria.

## 1.3 Antibiotic Resistance

Antibiotics are essential drugs for humans. WHO and US data indicated that the proportion of deaths because of infections has decreased from 50% to 10% after widespread antibiotic usage.<sup>1</sup> However, antibiotic resistance has been reported to steadily rise in recent years. The antibiotic resistance pathogens increasing rate has exceeded the new antibiotic drug discovery rate. This could be the end of the unparalleled modern medicine age. US CDC (Center for Disease Control and Prevention) has declared 16 kinds of antibiotic resistance are in serious stage in 2019 and more than two million patients have suffered due to antibiotic resistant infections.<sup>25</sup>

Antibiotics have only been widely used for about 80 years. The wide use of antibiotics as therapeutics creates a selection for antibiotic resistance. For instance, methicillin is a  $\beta$ -lactam class antibiotic which was available in the market by 1959. Methicillin resistant *Staphylococcus aureus* (MRSA) was first found in 1961. The bacteria later became resistant to almost all  $\beta$ -lactam class antibiotics and was called as “super bacteria”. At first, MRSA was only found in hospitals. However, research indicated that MRSA has spread to patients not in previous contact with hospitals.<sup>26</sup> The O’Neill commission is an organisation established in 2014 which was commissioned by the UK government to investigate antibiotic resistance. Their report indicated that antibiotic resistance would result in a 7% loss of economy (up to 210 trillion US dollars) and about 10 million advanced microbial resistance (AMR) annual deaths.<sup>27</sup>

Antibiotics and antibiotic resistance can both be found as natural products in the natural environment. Most of the known antibiotics were derived from actinomycetes (soil-dwelling bacteria groups). Actinomycetes can produce specific antibiotic substances during metabolising, such as erythromycin, vancomycin and streptomycin.<sup>28</sup> When the antibiotic producer generates antibiotics, it must have the resistance to prevent it from being killed by its own metabolites. The genes from these organisms are responsible for antibiotic resistance in many patients.<sup>29</sup>

The reason why bacteria show multiple antibiotic resistance is because it can obtain resistance genes via both vertical and horizontal gene transfer.<sup>30,31</sup> Vertical gene transfer is obtaining genes from parents' gene pool during division, while through horizontal gene transfer the bacteria could obtain resistance from the environment.<sup>30,31</sup> Therefore, the environment becomes a massive source of acquiring and accumulating resistance genes.<sup>32</sup> After accumulating the resistance genes in the natural environment, horizontal gene transfer makes it able to spread to the genes in the microbial community members. Recent research has identified resistance genes to  $\beta$ -lactams, glycopeptides and tetracyclines from metagenomic samples thought to be over 30000 years old.<sup>33</sup> Besides, along with medical and agricultural use, the resistance accumulation in the biosphere is dramatically increasing.<sup>34</sup> There is recent research indicating that the sensitivity of antibiotic drugs is being gradually and steadily replaced by resistance in clinical use.<sup>35</sup>

The main non-therapeutic use of antibiotics originates from the increasing need for animal protein. After Second World War, scientists tried to introduce antibiotics into the food industry.<sup>36</sup> At that time, researchers tried vitamins, folic acid and other chemical substances to accelerate animal growth. In 1946, Moore et al. indicated that feeding with sulpha drugs, folic acid and streptomycin can promote chicken growth.<sup>36</sup> In 1949, Lederle Laboratory tried various diet types to identify the “animal protein factor”, and believed the answer was vitamin B12 along with some unknown substances.<sup>37,38</sup> The factor that promoted growth was soon identified as an antibiotic but not vitamin B12. This finding was applied on swine and cattle by to further research.<sup>39,40</sup> In 1951, people started to use Terramycin, Aureomycin and Penicillin in husbandry. Between 1963 to 1965, the *Salmonella typhimurium* outbreak reminded people the risk of widely using antibiotics in animals. UK government established a joint committee for “Antibiotics in Animal Husbandry and Veterinary Medicine”. After detailed research, the commission indicated the real risks of developing antibiotic resistance by using antibiotics to promote animal growth. However, by 2013, about 131,109 tons of antibiotics was being used in the animal industry. The consumption of antibiotics is projected to reach 200,235 tons by 2030.<sup>41</sup>

## 1.4 Mechanisms of Antibiotic Resistance

Along with the development of antibiotics, their wide misuse also promoted to the appearance and dissemination of antibiotic resistance. Bacteria can use multiple resistance mechanisms to overcome the effects of antibiotics.<sup>42</sup> The mechanisms of antibiotic resistance can be broadly categorised into four groups: 1) chemical modification of antibiotics, 2) preventing antibiotic from reaching its target, 3) modification of the binding site and 4) resistance based on global cell adaptation.<sup>42</sup>

### **Chemical modification of antibiotics**

Chemical modification of antibiotics is one of the most common and successful mechanism employed by the bacteria. Here, the bacteria expresses an enzyme that can destroy the function of antibiotic drugs. The enzyme can either add specific chemical alternations or completely destroy the antibiotic. For instance, aminoglycoside modifying enzymes can inhibit aminoglycoside molecules by modifying the amino groups or hydroxyl groups.<sup>43</sup> The  $\beta$ -lactam resistance is achieved through the breakdown of the  $\beta$ -lactam ring in the antibiotics by the  $\beta$ -lactamase enzymes.<sup>44</sup>

### **Preventing antibiotics from reaching the target**

Bacteria can either decrease the permeability of antibiotics or expunge the antibiotic through efflux pumps. Many antibiotics have intracellular or cytoplasmic membrane targets, which means the antibiotic molecule needs to have an acceptable permeability

to reach the target. For instance, *P.aeruginosa* can have an aberrant expression of OprD protein, which can decrease the permeability and result in resistance to carbapenems antibiotics.<sup>45</sup> The efflux pumps can also result in resistance via expunging antibiotic molecules completely out of the bacteria. In the 1980s, scientists recorded *E.coli* can pump tetracycline out of the cytoplasm of *E.coli*.<sup>46</sup>

## **Modification of the antibiotic binding sites**

The binding sites of antibiotics in proteins can also be modified, which results in antibiotic resistance subsequently. There are several mechanisms that generate this kind of resistance, such as protection of the targets, mutation of the target sites, enzymatic alteration of the target sites or bypassing of the binding sites. For instance: Tet(O) and Tet(M) can dislodge tetracycline from its binding sites by interacting with the ribosome,<sup>47</sup> while erythromycin ribosomal methylation can develop macrolide resistance.<sup>42</sup>

## **Global cell adaptations resistance**

Bacteria has its own mechanisms to cope with the external environment. In the host body, the bacteria need to compete for nourishment while it is constantly attacked by the host's immune system. Thus, the ability to adapt and cope with the stressful environment is essential. During thousands of years of evolution, bacteria has devised

complex mechanisms to keep the safety of pivotal cellular processes such as membrane homeostasis and cell wall synthesis. For instance, vancomycin resistance has developed using the global cell adaption response to the antibiotics.<sup>42</sup>

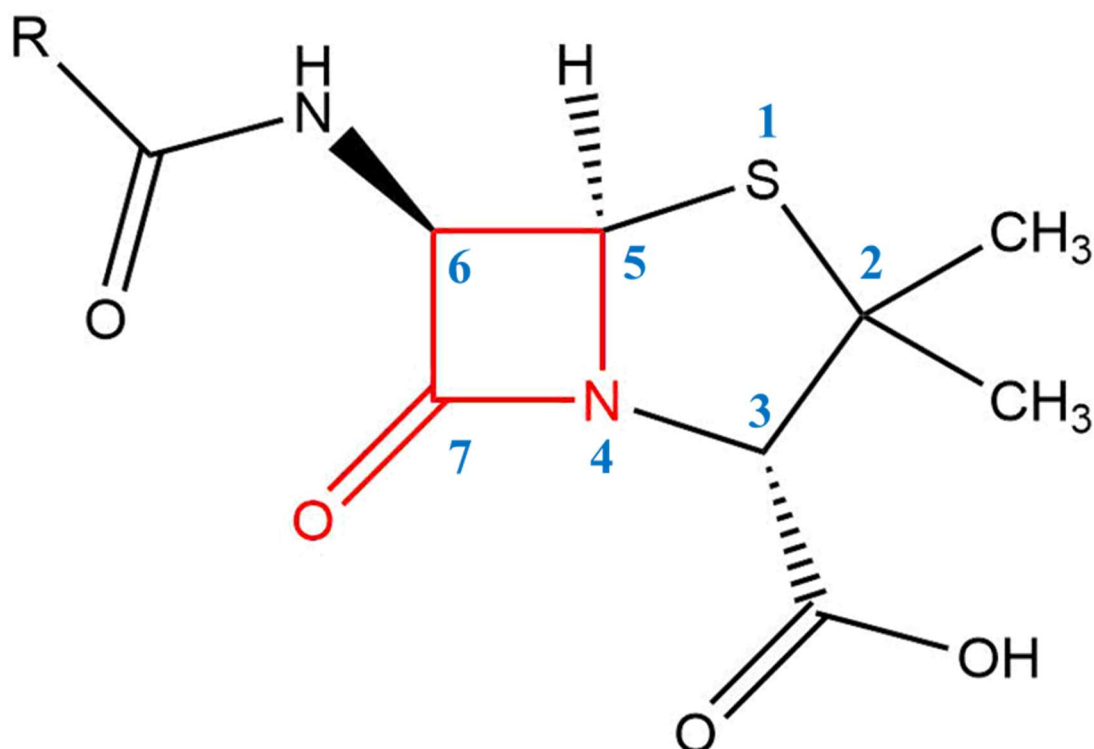


## 1.5 $\beta$ -lactam and $\beta$ -lactamases

The class of  $\beta$ -lactam antibiotics contain a  $\beta$ -lactam ring as their core structure (Figure 2).  $\beta$ -lactam antibiotics are the most commonly used antibiotics to treat infections. Since the discovery of penicillin in 1928, there are a number of  $\beta$ -lactam containing drugs have been discovered, including penicillin derivatives,  $\beta$ -lactam monobactams and cephalosporins and carbapenems.  $\beta$ -lactam has advantages in the treatment of serious bacterial infection based on its reaction mechanism. When  $\beta$ -lactam antibiotic gets into the human body, it can interact with penicillin-binding proteins (PBPs) and result in the prevention of bacteria cell wall formation.<sup>48</sup> Its safety and efficacy have been proven in clinical treatment through time. However, the widely use of  $\beta$ -lactam results in the frequent expression of the  $\beta$ -lactamase enzyme, which will be mentioned in the following sections.

$\beta$ -lactamase is an enzyme that can hydrolyse  $\beta$ -lactam functional group and subsequently destroy the antibiotic biofunction.<sup>48</sup>  $\beta$ -lactamase was first discovered in *Bacillus (Escherichia) coli*, even before the clinical application of penicillin.<sup>49</sup> Wide usage of  $\beta$ -lactam antibiotics has resulted in growing cases of antibiotic resistance. High use of  $\beta$ -lactam antibiotics results in selective pressure on the bacteria, where only the ones expressing  $\beta$ -lactamase survive.<sup>50</sup> There are about 2800  $\beta$ -lactamases that have been identified and investigated.<sup>51</sup> Due to the high rate of replication and mutation frequency, bacteria have the ability to adapt to new  $\beta$ -lactam antibiotics. As the resistance develops,  $\beta$ -lactam drugs gradually lose their potency against bacteria.<sup>52</sup> The

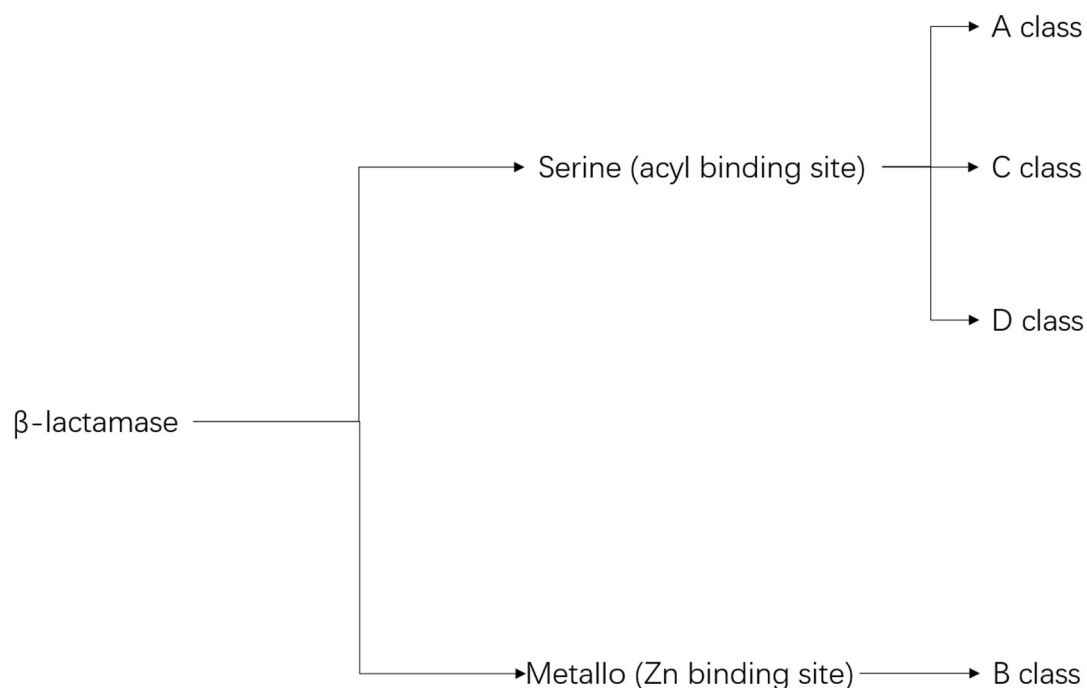
rapid evolution rate allows  $\beta$ -lactamase enzymes to make the new antibiotic redundant within a relatively short time.



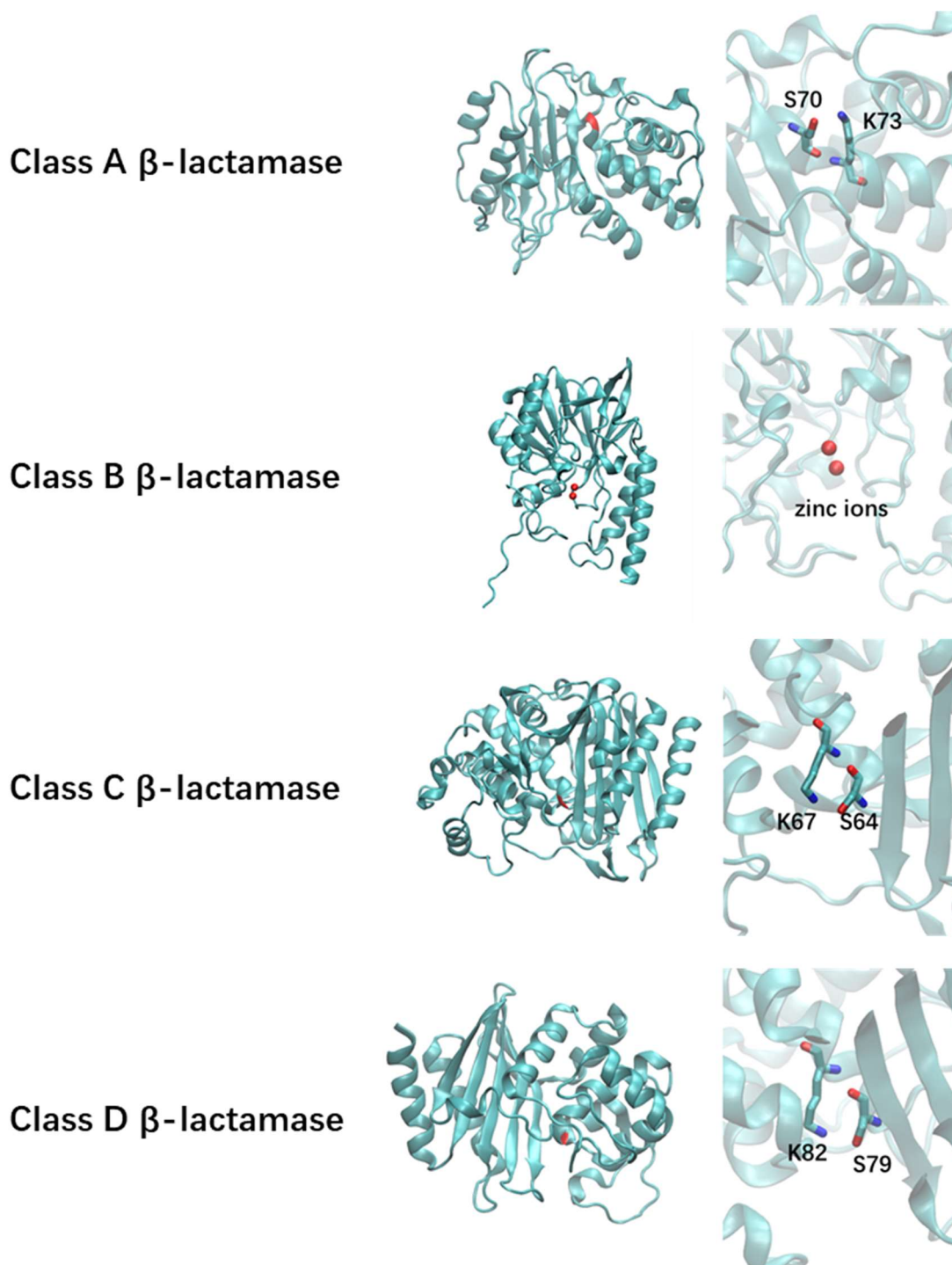
**Figure 2:** The chemical structure of Penicillin. The red ring is the  $\beta$ -lactam ring, the nitrogen atom attaches to the  $\beta$ -carbon atom relative to the carbonyl and is therefore named as “ $\beta$ -lactam”. The carbon number is highlighted with blue.

There are two major classifications of  $\beta$ -lactamases - the Ambler and Bush-Jacoby system.<sup>53</sup> The Ambler system categorises  $\beta$ -lactamases into class A to D based on protein sequence, namely A (serine penicillinases), B (metallo- $\beta$ -lactamases), C (cephalosporinases), D (oxacillinases).<sup>53</sup> The Bush-Jacoby system categorises  $\beta$ -lactamase into class 1 to 4 based on the substrate hydrolysis profile and inhibitor profile.<sup>54</sup> Researchers further subdivided the classification of  $\beta$ -lactamases into two

main groups based on different hydrolysis mechanisms (Figure 3): by generating serine active site of an acyl enzyme (Class A, C, and D) and hydrolysis reaction facilitated through one or two essential zinc ions in the active site (Class B, metallo- $\beta$ -lactamases).<sup>55</sup> The representative structures and their binding site from the four classes of  $\beta$ -lactamase are shown in Figure 4. The serine binding site and lysine general base are shown in class A, and C  $\beta$ -lactamases.



**Figure 3:** Phylogenetic tree. Showing the structural evolutionary similarity between the four classes of  $\beta$ -lactamase. Classification of the  $\beta$ -lactamase enzyme family. Class A, C and D are serine  $\beta$ -lactamases which have acyl binding sites. Class B is metallo- $\beta$ -lactamase which has zinc binding site.

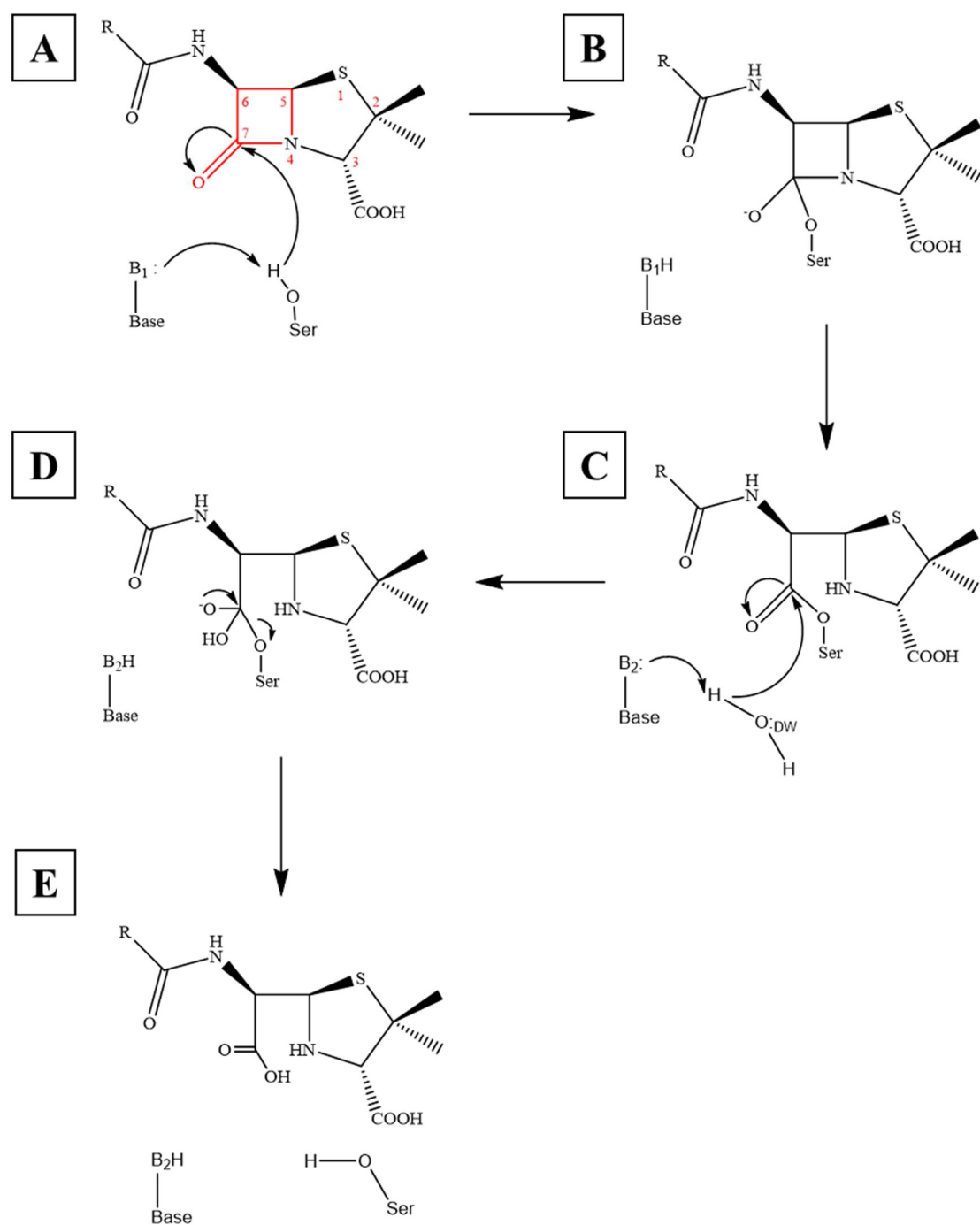


**Figure 4:** Representative structures and their binding sites from the four classes of  $\beta$ -lactamase. The red region is the serine binding residue. The serine binding site and lysine general base in the binding site are focused on the right. Class A KPC-2  $\beta$ -lactamase is from PDB 5UL8. Class B L1 Metallo- $\beta$ -lactamase is from PDB entry 1SML. Class C AmpC  $\beta$ -lactamase is from PDB entry 1KE4. Class D OXA-23  $\beta$ -lactamase is from PDB entry 4JF4.<sup>56-59</sup>

The serine  $\beta$ -lactamases (class A, C and D) mechanism of catalysis is well studied.<sup>52</sup> The three classes of SBLs (serine  $\beta$ -lactamases) displays low sequence identity. In spite of this, all of the SBLs exhibit the same acylation-deacylation mechanism of catalysis (Figure 5). During the acylation reaction, a general base B1 activates the nucleophilic serine (catalytic serine) and attacks the scissile carbonyl carbon of the  $\beta$ -lactam ring (Figure 5-A). The amide bond of  $\beta$ -lactam ring will break during the process via a tetrahedral oxyanion transition state (Figure 5-B). The intermediate state acylenzyme is formed as a result (Figure 5-C). The deacylation reaction occurs when a water molecule is activated by general base B2 for nucleophilic attack via tetrahedral deacylation (Figure 5-D). The water in this step is called the deacylating water (DW). A Penicilloate product is liberated after the nucleophilic attack (Figure 5-E).

The class B metallo- $\beta$ -lactamases (MBLs) proposed mechanism is also well documented (Figure 6).<sup>52</sup> The ground state MBLs zinc binding site is associated with five histidine and one aspartic acid. Two zinc ions (Zn1 and Zn2) are bridged by a water. The substrate binding allows the bridged water molecule to do a nucleophilic attack on the scissile carbonyl (Michaelis complex). In the transition state complex, the carbonyl carbon geometry is altered. The geometry changes from planar to tetrahedral, to enable the  $\beta$ -lactam ring to be broken. In product dissociation, zinc ions will be bridged by a new water molecule. In MBLs research, the catalytic mechanism is important and widely investigated.<sup>60</sup> The zinc ions in MBLs can be substituted by more active metal ions (such as cadmium and cobalt) while retaining the hydrolytic ability of the

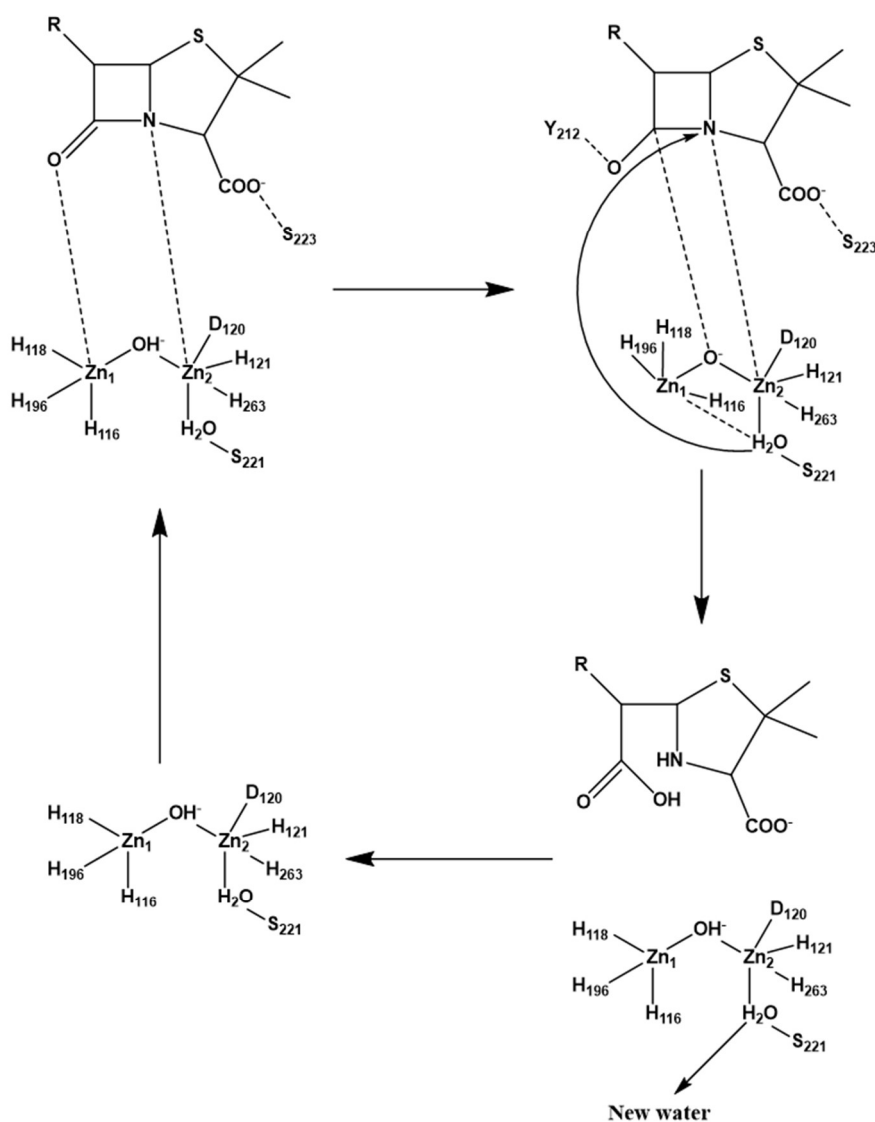
enzyme.<sup>61</sup> The zinc ions (Zn1 and Zn2) have different affinities in the zinc binding site.<sup>62</sup> The enzymes with only Zn1 position occupied can be isolated.<sup>63</sup> On the other hand, the Zn2 ion can coordinate with cysteine and have shown to be sensitive to oxidation.<sup>64</sup> Researchers have a consensus view for the catalytic mechanism of MBLs. The binuclear binding site has the highest catalytically ability and represents for most cases in physiological conditions.<sup>65</sup>



**Figure 5:** The proposed mechanism of catalysis for serine  $\beta$ -lactamases (class A, C and D). (A) general base  $B_1$  activates nucleophilic serine to do a nucleophilic attack on carbonyl carbon. (B) the geometry of carbonyl carbon changes from planar to tetrahedral. (C) the  $\beta$ -lactam ring is broken and generates a covalent acyl-enzyme. Deacylating water activates general base  $B_2$ . (D) nucleophilic attack on carbonyl carbon in tetrahedral deacylation transition state. (E) product dissociation.

## Michaelis complex

## Transition state complex



## Ground state

## Product dissociation

**Figure 6:** The catalytic mechanism of L1 metallo- $\beta$ -lactamase. In the ground state, two zinc ions in the binding site are bridged by a water molecule. These two zinc coordinate with five histidine residues and an aspartic acid. Zn2 also has an indirect bond with a serine residue through a water molecule. During binding to the substrate (Michaelis complex), Zn1 binds to the carbonyl oxygen, Zn2 binds to the  $\beta$ -lactam ring nitrogen, water bridge provides a nucleophilic attack on the carbonyl carbon. The nucleophilic attack on the carbonyl carbon changes its geometry, from planar to tetrahedral causing the Zn1-oxygen bond to be broken and forming a new bond with Y212. The water molecule bound to Zn2 is acidified by both zinc ions because of their proximity. The acidified water protonates the  $\beta$ -lactam ring nitrogen and bridges the two Zinc ions. In the product dissociation, another water from the solvent bridges Zn2 and S221.



Class A  $\beta$ -lactamase is serine dependent enzyme, including NMC-A (non-metallo-carbapenamase), KPC, IMI, and SME. The members of class A  $\beta$ -lactamase can hydrolyse aztreonam, cephalosporins, carbapenems and penicillins.<sup>66</sup> Class A  $\beta$ -lactamase is mainly found in *K. pneumoniae*, *Enterobacter cloacae* and *Serratia marcescens*. These bacteria often show multiple drug resistance and narrow the treatment options. Class A  $\beta$ -lactamase is widely spread which makes the research of it more clinically important.<sup>57</sup>

Class B metallo- $\beta$ -lactamase show wide sequence diversity. There is only 25% sequence identity among some of the enzymes.<sup>67</sup> However, structures from multiple metallo- $\beta$ -lactamases have shown conserved structure around the binding sites. The enzyme has a special sandwich fold, with active sites on the interface. Up to 6 residues at the active sites can coordinate with one or two zinc ions in the centre of the binding site and hydrolyse a number of  $\beta$ -lactam antibiotics. The metallo- $\beta$ -lactamase has recently attracted a lot of attention.<sup>68-70</sup> In spite of the immense work that has been carried out to investigate metallo- $\beta$ -lactamases, there is still no metallo- $\beta$ -lactamase inhibitor available for clinical use.<sup>69,71,72</sup>

Class C  $\beta$ -lactamase are serine cephalosporinases and the first identified  $\beta$ -lactamase.<sup>52</sup> Many most important Gram-negative species also have these genes to encode class C  $\beta$ -lactamase such as *Enterobacter cloacae*, *Citrobacter freundii*, *Morganella morganii*, *Serratia marcescens* and *Pseudomonas aeruginosa*.<sup>73</sup> These genes are distributed on a

sequence labelled as AmpC. The AmpC genes are inhibited by upstream sequence labelled AmpR and are not expressed under normal condition.<sup>74</sup> When the bacteria is exposed to a  $\beta$ -lactam, cell wall metabolite and intracellular incorporation will happen. The cell wall metabolite is processed by the enzyme AmpD.<sup>75</sup> The interaction between substrate AmpD and AmpR reverses the repression, resulting in the AmpC  $\beta$ -lactamase to be synthesized.<sup>73</sup> The class C  $\beta$ -lactamase usually show resistance to cefotaxime, cefoxitin, cefotetan and ceftriaxone. This kind of  $\beta$ -lactamase also has a low hydrolysing rate for cefepime and could be inhibited by aztreonam, cloxacillin and oxacillin.<sup>53</sup>

Class D  $\beta$ -lactamases are serine oxacillinases. These enzymes get their name based on their high rate of hydrolysing oxacillin (at least 50% hydrolyse benzylpenicillin)<sup>76</sup>. The class D  $\beta$ -lactamases have the most diversity in all four classes of  $\beta$ -lactamase and have very poor understanding in many respects.<sup>52</sup> When the class D  $\beta$ -lactamase was first found, it only shows hydrolysed function to penicillins.<sup>77</sup> However, class D  $\beta$ -lactamases also show activity against carbapenems and cephalosporins now.<sup>78</sup> These OXA enzymes are found in many Gram-positive bacteria species in recent research, the widely spread of this kind of  $\beta$ -lactamase increase the clinical significance.<sup>79</sup>

## 1.6 L1 and L2 $\beta$ -lactamases

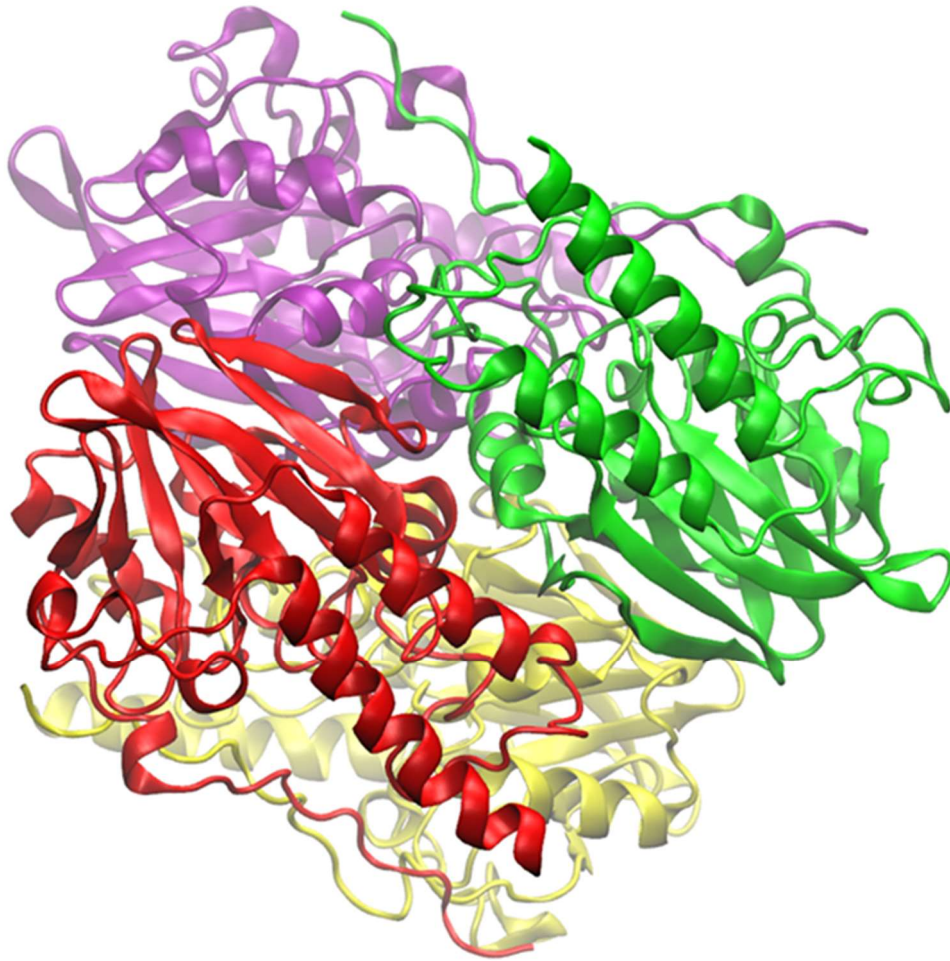
### **L1 metallo- $\beta$ -lactamase**

*Stenotrophomonas maltophilia* (*S. maltophilia*) is a nosocomial pathogen, which mainly infects immunocompromised hosts such as transplant, cancer and cystic fibrosis patients.<sup>80</sup> *S. maltophilia* causes respiratory tract infection, along with clinical syndromes, such as catheter-associated and bloodstream.<sup>81</sup> Although *S. maltophilia* is not a highly virulent pathogen, however, it has a crude mortality rate from 14% to 69% in bacteraemia patients.<sup>82,83</sup> Treating *S. maltophilia* infection is rather difficult because it has low susceptibility to the commonly used antibiotics.<sup>84</sup> However, the low susceptibility of *S. maltophilia* is not because of the acquisition through antibiotic resistance or mutations.<sup>85</sup> Thus, there is an argument on how *S. maltophilia* infects humans. Some researchers believe that *S. maltophilia* infecting people is not a result of pathotype adaption to human hosts.<sup>86,87</sup> On the other hand, recent research indicates the infection of *S. maltophilia* is the result of how bacteria adapts to the human hosts (fibrosis patients).<sup>88</sup> Comprehensive research is still needed for a more specific understanding of *S. maltophilia* and the infections it causes.

The antibiotic resistance of *S. maltophilia* is caused by two co-ordinately expressed chromosomal proteins: L1 metallo- $\beta$ -lactamase (class B MBL) and L2  $\beta$ -lactamase (class A SBL).<sup>89</sup> The expression of these two  $\beta$ -lactamases is introduced by  $\beta$ -lactam antibiotics and both controlled by AmpR regulator. AmpR is a master regulator which

controls hundreds of genes expression.<sup>90</sup> The AmpR regulated genes involved in the  $\beta$ -lactam resistance. For *S. maltophilia*, research has indicated the AmpR regulator is necessary for L1 and L2  $\beta$ -lactamases production.<sup>91</sup> To be more specific, AmpR is activator for L1  $\beta$ -lactamase production with or without an inducer.<sup>92</sup> Besides, AmpR is an essential factor for L1  $\beta$ -lactamase production even in a basal level. For L2  $\beta$ -lactamase, AmpR is a repressor without an inducer and an activator with an inducer.<sup>92</sup> In contrast, L2  $\beta$ -lactamase basal expression is not AmpR-dependent. Therefore, L1 and L2  $\beta$ -lactamases seem to be regulated in different ways through AmpR. L1 and L2  $\beta$ -lactamases combine to generate the resistance observed in *S. maltophilia*. However, the relationship between the structure and function are not fully known.

L1  $\beta$ -lactamase is a unique  $\beta$ -lactamase. It has several features that make it special among all  $\beta$ -lactamases - It is the only known  $\beta$ -lactamase enzyme that exists as a tetramer (Figure 7). Each monomer of L1  $\beta$ -lactamase has 10  $\alpha$  helix and 12  $\beta$  sheets. The tetramer is maintained through three discrete sets of intra- and intermolecular interactions which make the monomers interact with the other three monomers.<sup>59</sup> The interactions between monomers are mainly hydrophobic, each monomer has a buried non-polar side-chain surface area of 1300  $\text{\AA}^2$  for forming the tetramer.<sup>59</sup>

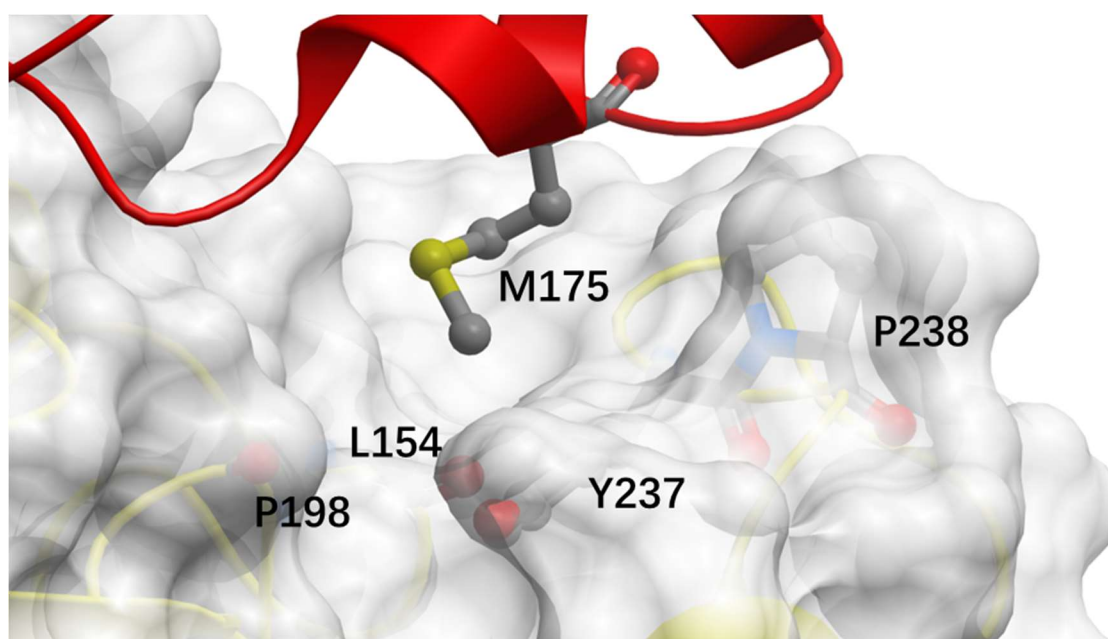


**Figure 7:** The structure of L1 metallo- $\beta$ -lactamase. L1 MBL has four identical subunits. Different subunits are coloured in distinct colour. Subunit A is coloured with red, subunit B is coloured with yellow, subunit C is coloured with green and subunit D is coloured is coloured with purple. L1 MBL is the only known tetramer in all of the  $\beta$ -lactamase family.

There is no evidence that can prove that the allostery affects the resistance activity, however, all other known  $\beta$ -lactamase, including both serine and metallo- $\beta$ -lactamase, are monomeric.<sup>93</sup> The proposed mechanism of L1 metallo- $\beta$ -lactamase hydrolysing  $\beta$ -lactam antibiotic has been shown above in Figure 6.

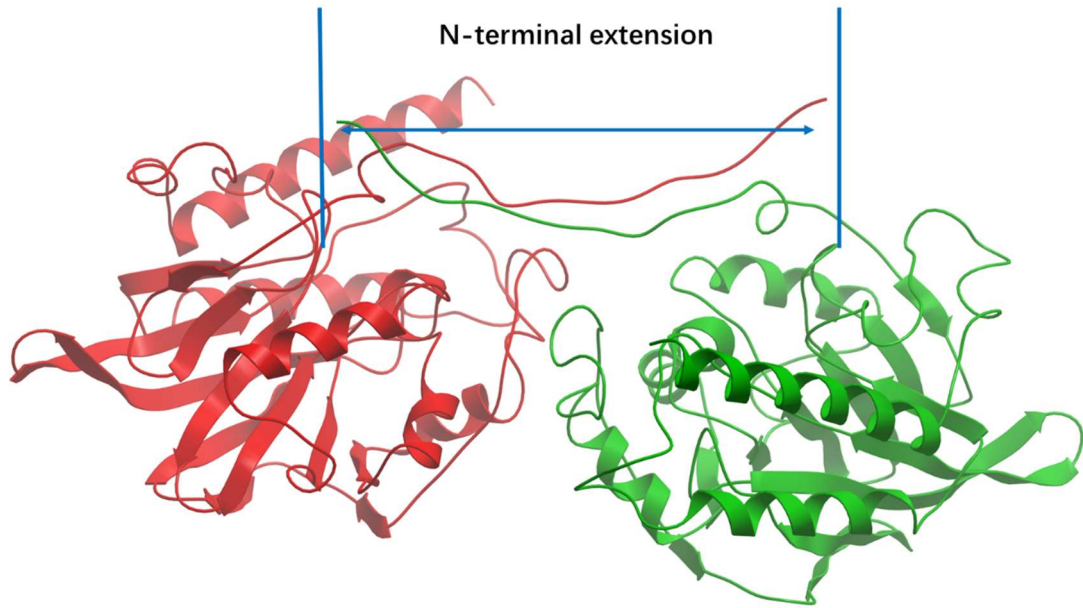
There are several interactions that exist between subunits in L1  $\beta$ -lactamase, with which the monomer interact with other three monomers. First is the interaction between

subunit A and subunit B (and also between C and D subunits). The A and B subunits have similar interactions as C and D. The side chain of M175 of subunit A is surrounded by residues L154, P198, Y237 and Pro238 from subunit B, forming hydrophobic interactions (Figure 8).<sup>59</sup> M175 (subunit A) interaction with the hydrophobic pocket in subunit B (repeated in C and D) is essential for tetramer stabilisation. A disruption of this interaction prevents tetramer formation of L1  $\beta$ -lactamase.<sup>94</sup> The interactions in this pocket are stable and no clinical strains have been identified that show variations affecting these interactions.<sup>81</sup>

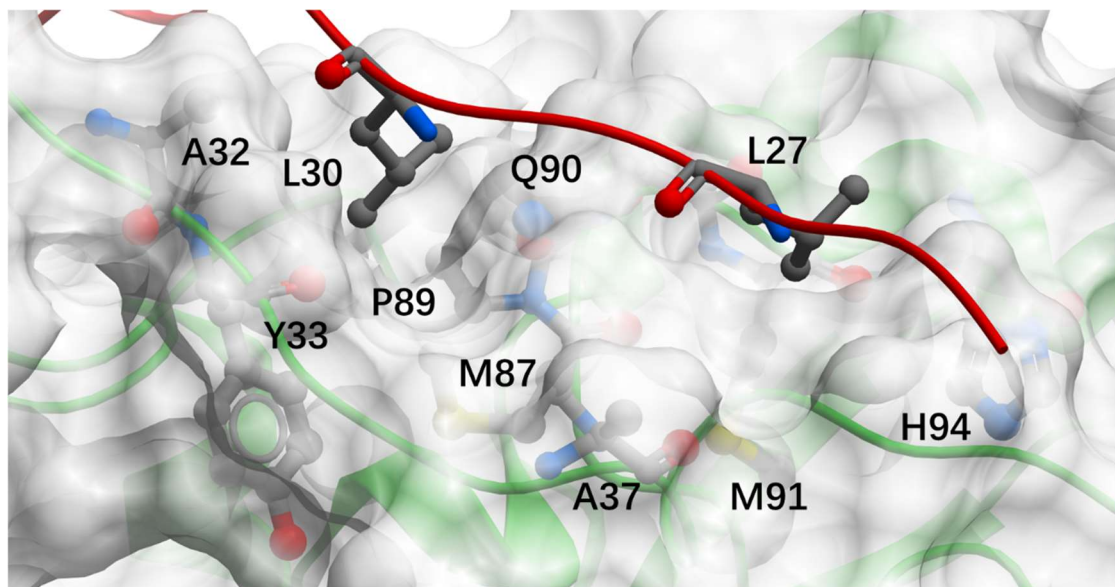


**Figure 8:** The M175-A side chain and the hydrophobic pocket on subunit B surface. Subunit A is coloured in red and subunit B is coloured in yellow. The side chain of M175 penetrates into the hydrophobic pocket on subunit B surface.

There are interactions between A and C subunits (B and D subunits) as well. The N-terminals of each subunit forms an extended U-shaped structure (Figure 9) which is stabilised by mutual hydrophobic and electrostatic contributions.<sup>59</sup> Two prominent leucine residues, L27 and L30 are positioned into the polar cavities on the opposing subunit surface. The surface cavities (Figure 10) are formed by the distal end (A32, Y33 and A37) and N-terminal end of  $\alpha 1$  (M87, P89, Q90, M91 and H94).<sup>81</sup> The area accounts for 600 Å<sup>2</sup> of buried hydrophobic surface in each monomer of L1  $\beta$ -lactamase. Interactions between M87 and H94 are reported to be important as well.<sup>95</sup> The 15 amino acid N-terminal extension is also responsible for  $\beta$ -lactam binding.<sup>95</sup> The N-terminal deletion results in a much slower hydrolysis rate and different substrate profile. Furthermore, the N-terminal is not necessary for  $\beta$ -lactam hydrolysis and tetramer stabilisation, but it is essential for negative cooperativity.<sup>94</sup> Besides, the interaction of P166A with P166C and R148C can be further formed (Figure 11).<sup>81</sup>

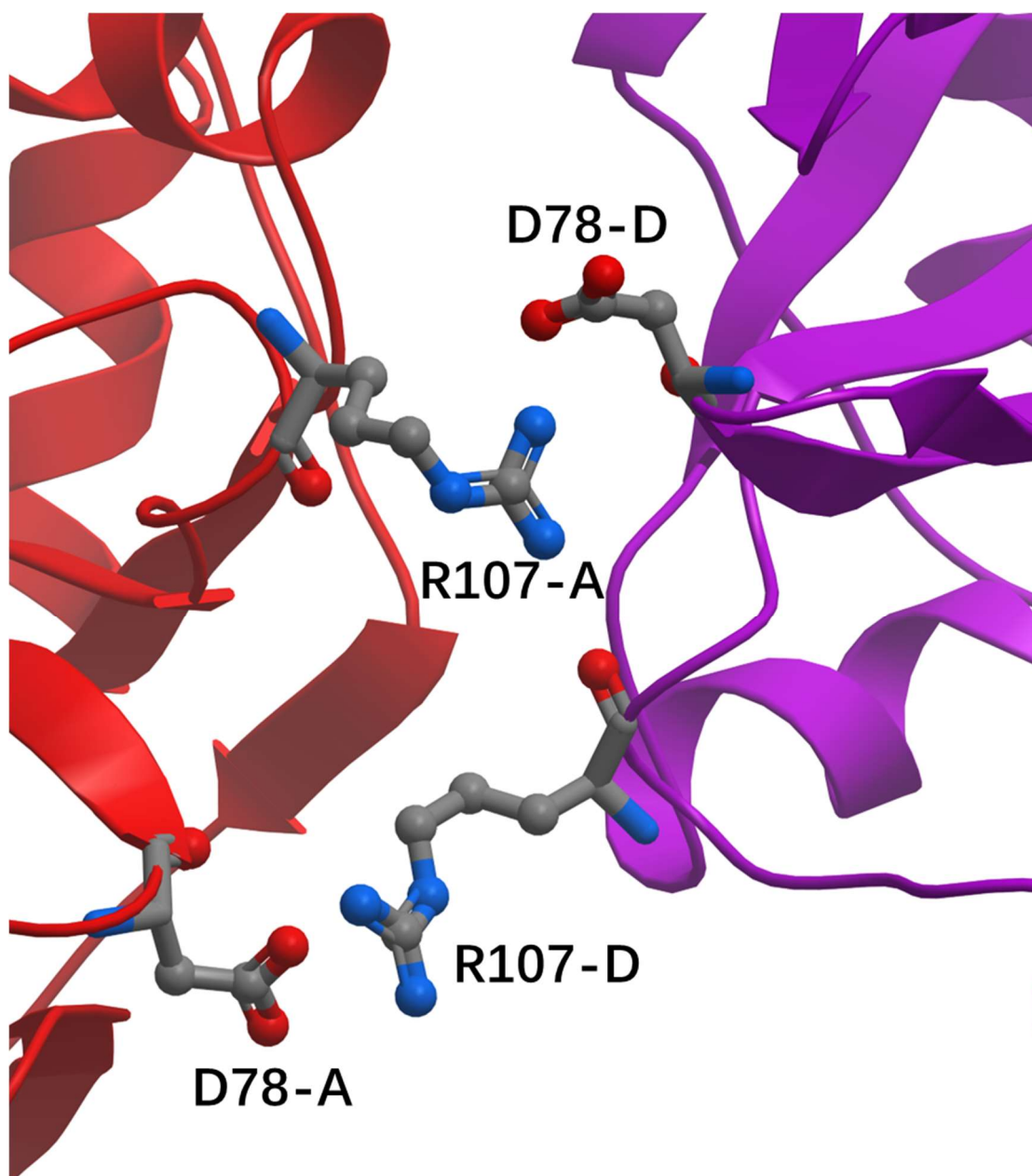


**Figure 9:** The N-terminal extensions of subunit A and subunit C. Each N-terminal extension is 15 amino acids long. The interactions between the two chains stabilized each other. The A subunit is colour in red and C subunit is coloured in green.



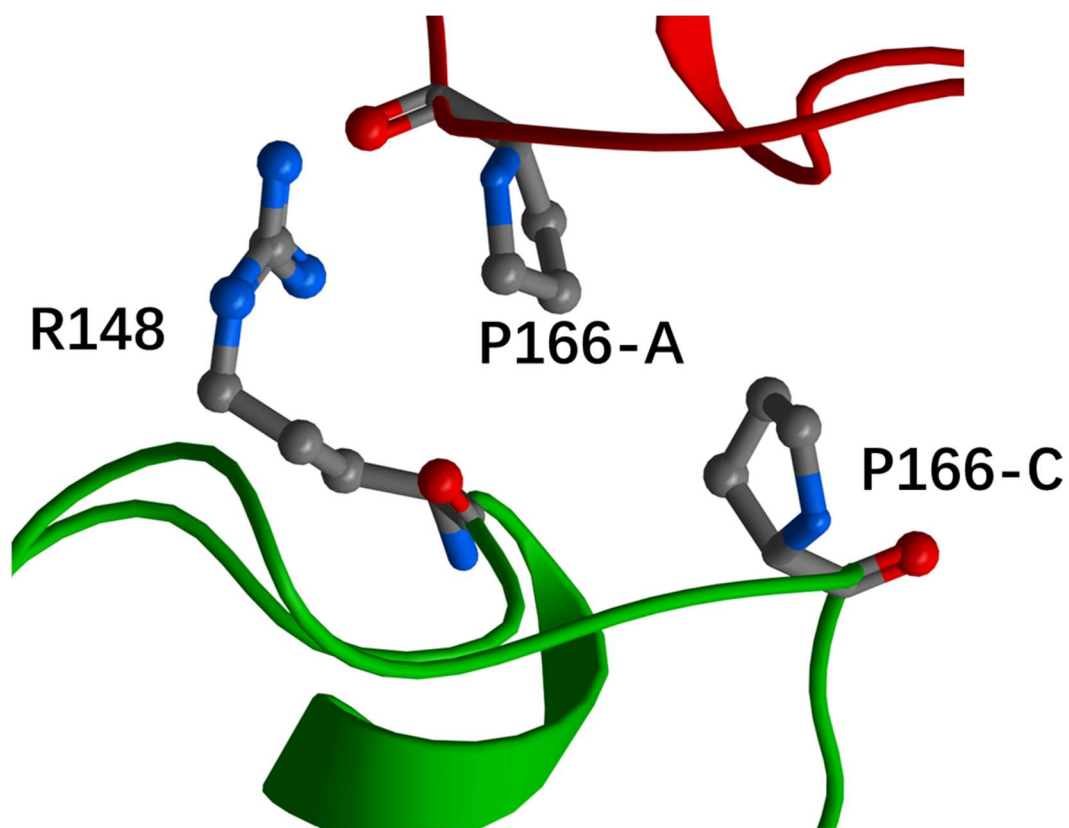
**Figure 10:** L27, L30 and the surface cavity residues. Two leucine residues (L27 and L30) penetrate into the surface cavities formed in the adjacent subunit. The cavities consist of residues A32, Y33, A37, M87, P89, Q90, M91 and H94. Subunit A is coloured in red and subunit C is coloured in green.





**Figure 11:** The backbone of P166 (A-subunit) interacts with the side chain of R148 (C-subunit). Subunit A is coloured in red and subunit C is coloured in green.

There are only a few interactions between subunit A and subunit D (or the analogous subunit B and subunit C). The hydrogen bonds between subunit A and D are mainly formed between D78 and R107 (Figure 12).<sup>59</sup>



**Figure 12:** The interactions between subunit A and subunit D. D78 and R107 are the main contributors forming a salt bridge interaction. The subunit A is coloured in red and subunit D is coloured in purple.

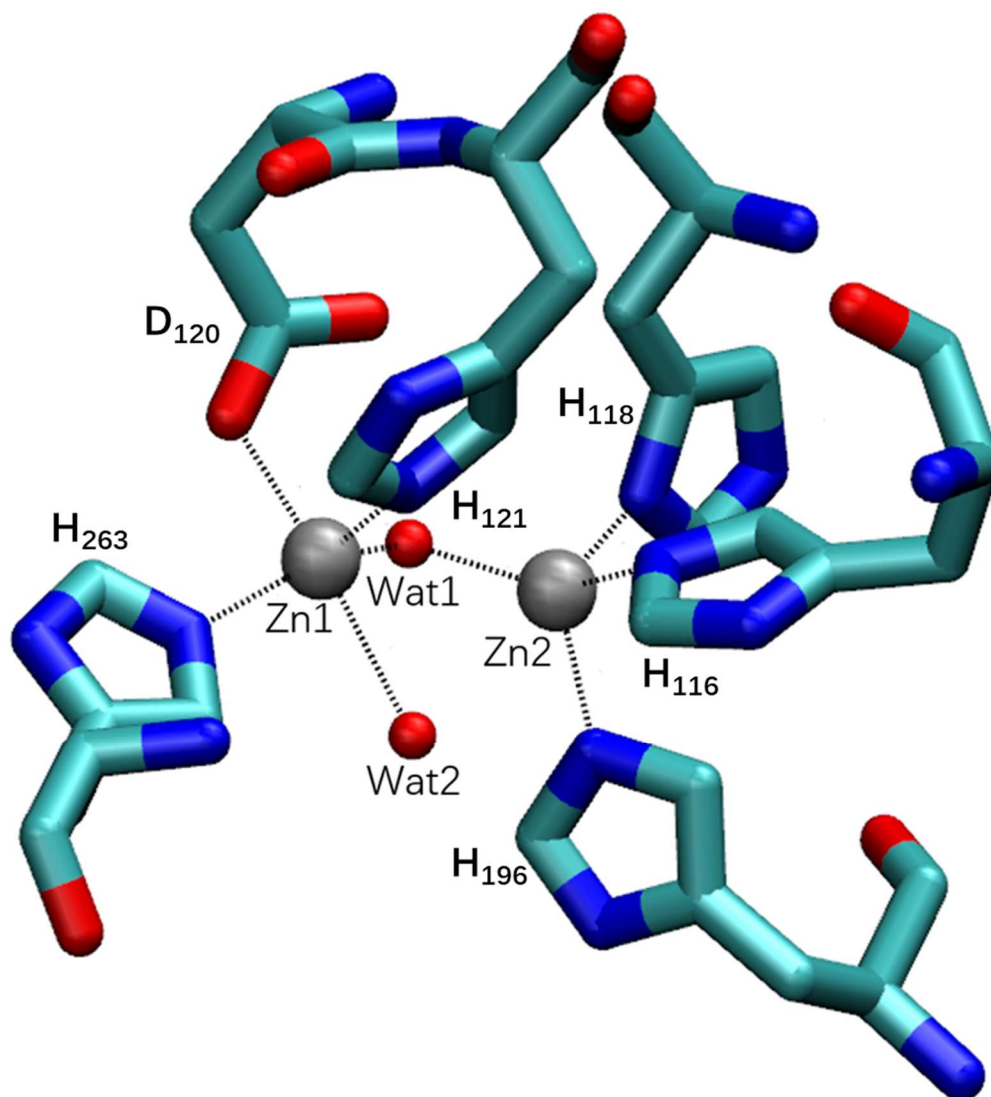
The two zinc ions in the binding site are bridged by a water molecule. The water molecule plays an important role in the hydrolysis reaction and present as a nucleophile. Zn1 is bound in trigonal bipyramidal coordination and Zn2 is bound in tetrahedral coordination (Figure 13). The geometry and residues involved in the ligation of Zn1 were first time observed in metallo- $\beta$ -lactamases.<sup>59</sup> Recent research indicates that Zn1 ion is essential for catalysis reaction and Zn2 ion is essential for proper folding of L1  $\beta$ -lactamase.<sup>96</sup>

A disulphide bridge and two elongated loops connect to the active site (Figure 14). Typically, the disulphide bond is the strongest bond in the proteins. It is responsible for stabilising the globular structure and maintaining protein conformation. Thus the disulphide bond plays an important role in protein folding and stabilization.<sup>97</sup>

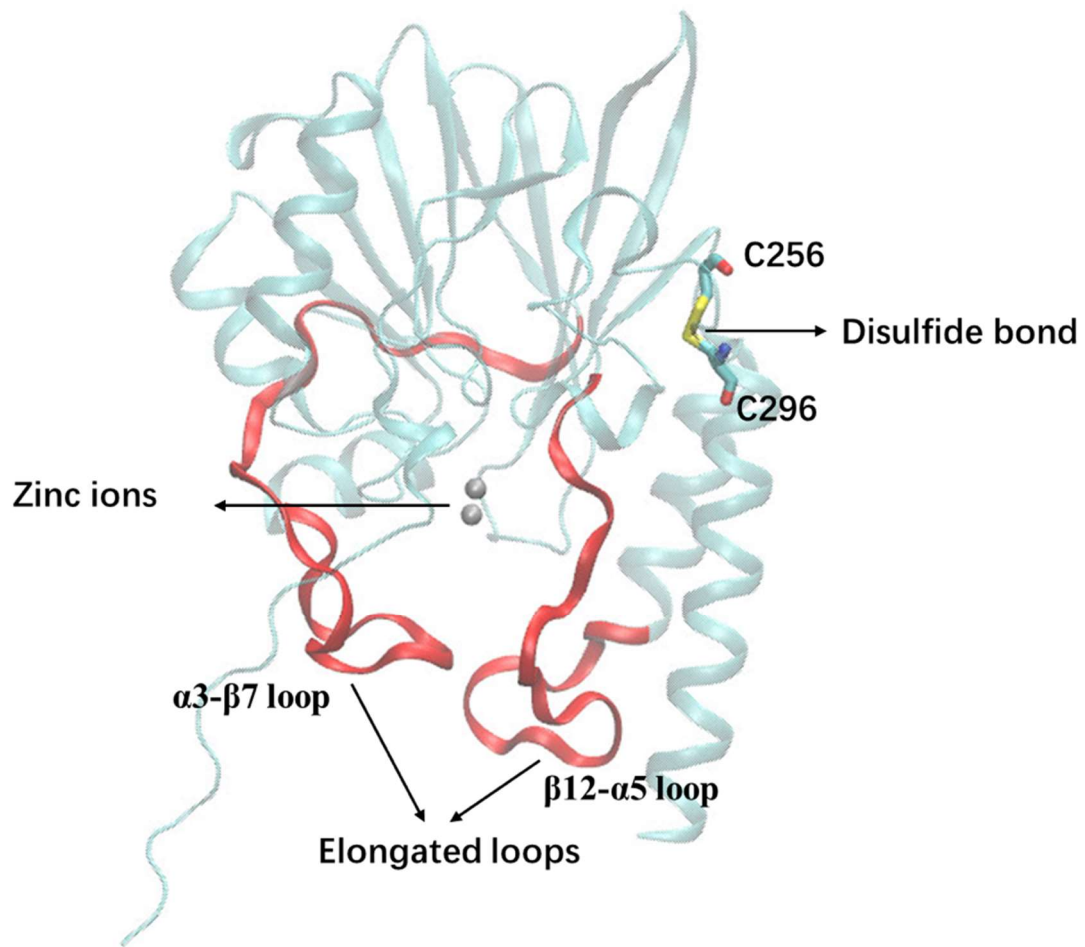
There are two elongated loops present in each subunit. These two protruding loops,  $\alpha_3$ - $\beta_7$  and  $\beta_{12}$ - $\alpha_5$  surround the binding site (metal centre). These loops can have hydrophobic interactions with  $\beta$ -lactam at the C2 and C6 positions and further influence the spectrum of substrates.<sup>59,98-100</sup> Besides, two novel substitutions were found on this area: G195A and (G/E)198D.<sup>81</sup> It is important because the substitution (G195A) can affect the positioning of Phe191 by decreasing the flexibility of the region.<sup>81</sup> Residue F191 has been proven to be important for catalysis and substrate binding through several studies such as molecular modelling and kinetic study.<sup>98,99</sup> Similarly, the substitution of (G/E)198D can also change the efficiency of enzyme catalytic. Residue

S221 on  $\beta$ 12- $\alpha$ 5 loop forms an indirect interactions with Zn<sup>2+</sup> and S225 is considered as a second-shell residue of H196.<sup>59</sup> There are four novel substitutions have been reported recently: (P/A)235V, Q230R, (K/Q)232R, G233D. Although none of these novel substitutions directly interact with the substrate, however, it is possible that these substitutions can influence the overall conformation of the  $\beta$ 12- $\alpha$ 5 loop and further affect the catalysis efficiency.<sup>81</sup> For instance, residues S221 and S225 are on the loop and have been proven important to the catalysis reaction. The substitutions can change the H-bonding network strength or alter its electrostatic characteristics, which will result in the changing of overall conformation. The overall conformation change could affect substrate binding and catalysis.

There are several sites which are thought to be essential for metallo- $\beta$ -lactamase function. However, L1  $\beta$ -lactamase has different substitutions at these sites. At position 219, an aspartic acid replaces cysteine, which is thought to be important for zinc binding; at position 222, a lysine is substituted by a serine. The lysine in this position is thought to be critical for substrate binding.<sup>89</sup> The focus of this research on L1  $\beta$ -lactamase will look into the key dynamic aspects of the L1  $\beta$ -lactamase structure.



**Figure 13:** The zinc binding site of L1 metallo-β-lactamase. Two zinc ions coordinate with six residues and two water molecules. Zn1 coordinates with residue D120, H121, H263. Zn2 coordinates with residue H116, H118, H196. Zn1 and Zn2 are bridged by a water molecule.



**Figure 14:** The two elongated loops ( $\alpha_3$ - $\beta_7$  and  $\beta_{12}$ - $\alpha_5$ ) form the gate entrance to the active site. These two loops surround the zinc binding site and control the volume of binding pocket. The disulphide bond is between residues C256 and C296. The loops are identical in all four subunits.

## L2 $\beta$ -lactamases

*Stenotrophomonas maltophilia* (*S. maltophilia*) produces antibiotic resistance through L1 and L2  $\beta$ -lactamase. These two  $\beta$ -lactamases can generate resistance to almost all  $\beta$ -lactam antibiotics.<sup>101</sup> L2  $\beta$ -lactamase is a class A  $\beta$ -lactamase which is inhibited by clavulanate.<sup>102</sup> Compared to L1, the discovery of L2  $\beta$ -lactamase is relatively new and has much less research work carried out on it. The literature mostly focuses on the function of AmpR, which regulates L2  $\beta$ -lactamase expression. As detailed above, AmpR was identified as a regulator, upstream of L2  $\beta$ -lactamase genome sequence. It has also been reported that AmpR is necessary for the catalytic function of both L1 and L2  $\beta$ -lactamase.<sup>103</sup> Besides,  $\beta$ -lactamase inhibitors for L2 also have been investigated, such as avibactam and bicyclic boronate.<sup>104</sup> In this thesis, computational methods were used to study the conformational dynamics of four members of the L2  $\beta$ -lactamase family namely L2a, L2b, L2c and L2d. Each L2  $\beta$ -lactamase has 15  $\alpha$  helices and 9  $\beta$  sheets (Figure 15). The sequence information was generated from PDBsum.<sup>105</sup>

The biggest difference between L2  $\beta$ -lactamase and other class A  $\beta$ -lactamase is that L2 has an extra  $\alpha$  helix at the beginning of the sequence (Figure 16). Compared with other families of class A  $\beta$ -lactamase such as KPC, TEM, SME and SHV, the elongated  $\alpha$  helix stacks on top of the enzyme and is far away from other secondary structural elements in the protein. This suggests that  $\alpha$ 1 helix, untethered at one terminal end is more flexible and therefore has more chance to influence the dynamics observed in the L2  $\beta$ -lactamase family. Besides, L2  $\beta$ -lactamases have no disulphide bond like in some

other families of class A  $\beta$ -lactamase (Figure 16). One of the CYS residue is replaced by ASN in L2  $\beta$ -lactamase. The disulphide bond is the strongest bond in the protein. It is responsible for stabilising the globular structure and maintaining protein conformation and therefore plays an important role in protein folding and stabilization.<sup>97</sup> As a result the corresponding position on L2  $\beta$ -lactamase,  $\beta$ 3- $\alpha$ 3 loop and  $\beta$ 7- $\beta$ 8 loop will be more dynamic. As a consequence of the loss of the disulphide bond, the adjacent loops are more relaxed and can adopt multiple conformations during dynamics. As the understanding of the dynamics of L2  $\beta$ -lactamase is sparse, the research described in this thesis aims to focus on the structural dynamics of the L2  $\beta$ -lactamases.



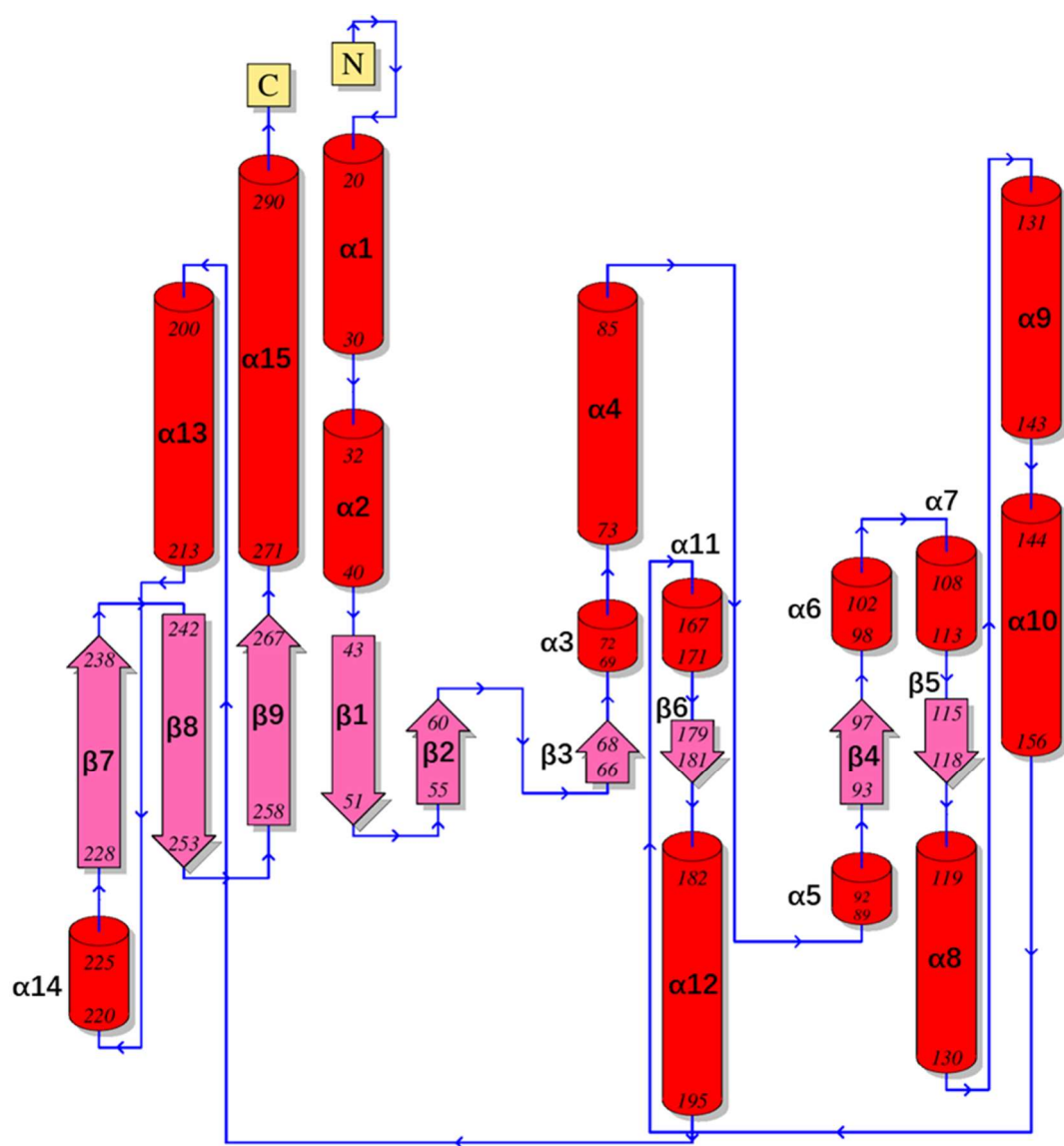
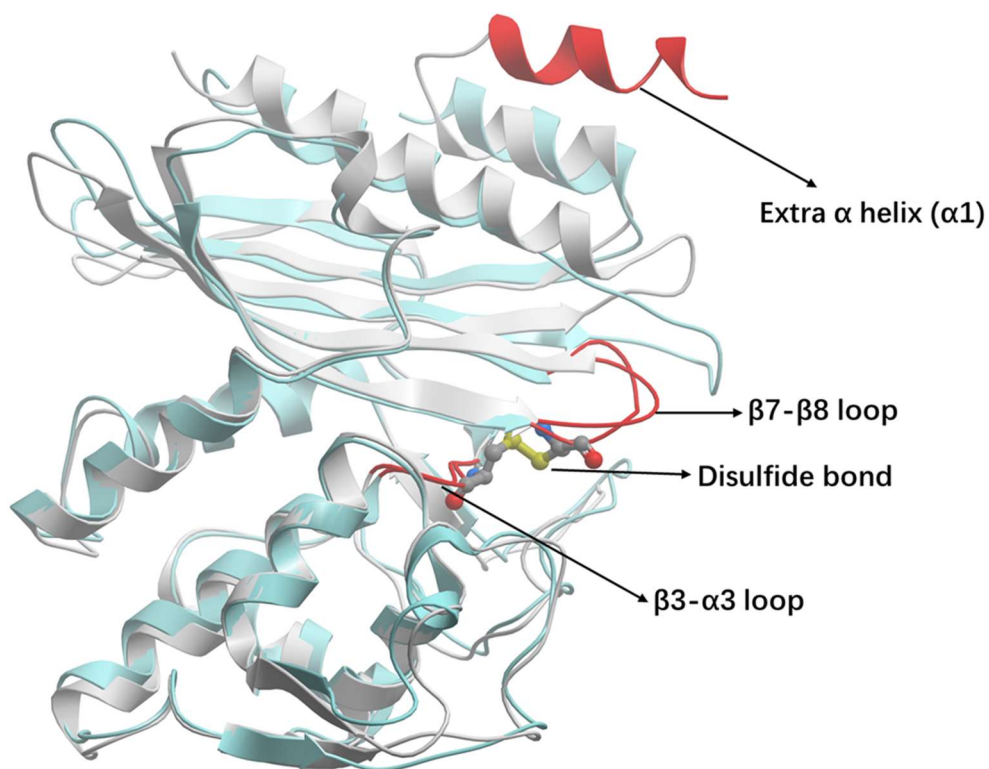


Figure 15: Sequence information of L2  $\beta$ -lactamase.<sup>105</sup>



**Figure 16:** The alignment of L2b  $\beta$ -lactamase crystal structure (PDB number 5NE2) and a typical class A  $\beta$ -lactamase (KPC-2  $\beta$ -lactamase, PDB number 2OV5). L2b enzyme is coloured in white and KPC-2 is coloured in light blue. The  $\alpha$ 1 helix and disulphide bond have been highlighted. KPC and other class A  $\beta$ -lactamases do not have the extra  $\alpha$ 1 helix that present in the L2 family. The  $\alpha$ 1 helix of L2  $\beta$ -lactamase is present at the C-terminal end of L2  $\beta$ -lactamase. The disulphide bond position of KPC is on the corresponding position of  $\beta$ 3- $\alpha$ 3 loop and  $\beta$ 7- $\beta$ 8 loop in L2  $\beta$ -lactamase.<sup>104,106</sup>

## 2. Hypothesis and Objectives

### 2.1 L1 Metallo- $\beta$ -lactamase Hypothesis

**Background:** The occurrence of antibiotic resistance in bacterial infection has increased in recent years. L1  $\beta$ -lactamase, a class B3 metallo- $\beta$ -lactamase (MBL), is unique, as it is the only functional tetramer amongst the large  $\beta$ -lactamase family. L1 MBL exhibits intrinsic resistance to almost all  $\beta$ -lactam antibiotics. Therefore, finding ways to inhibit MBL is crucial to overcome antibacterial resistance. One of the distinct structural features in the enzyme is the presence of two elongated loops between  $\alpha_3$ - $\beta_7$  and  $\beta_{12}$ - $\alpha_5$  that have been implicated to be directly involved in substrate binding and catalysis. The  $\alpha_3$ - $\beta_7$  and  $\beta_{12}$ - $\alpha_5$  loops are present at the interface between subunits and are in direct contact with M175, a residue that has been shown to have a direct effect on the catalytic activity of the enzyme.<sup>94</sup> Furthermore, the loops are directly linked with the substrate binding sites and residues in the direct vicinity of the loops directly coordinate with the Zinc atoms.<sup>59</sup>

**Hypothesis:** The working hypothesis is to explore and understand the structural dynamics of the  $\alpha_3$ - $\beta_7$ ,  $\beta_{12}$ - $\alpha_5$  loops and the correlation information between four subunits in the tetramer. Since the loops can control the volume of the binding pocket and thus affect the catalytic activity of the enzyme. Finding ways to stabilise them might be a route to inhibit L1 MBL. The signal transferring is also important for the full function of L1  $\beta$ -lactamase. Finding the correlation between subunits in the tetramer

will indicate the potential target to affect the signal transferring and further reduce the catalytic activity of L1  $\beta$ -lactamase.

**Aims:** The aim of this study is to use enhanced sampling molecular dynamics simulations to investigate:

- (a) What is the conformational dynamics observed in the elongated loops?
- (b) How the four subunits structurally communicate within the tetramer?
- (c) Identification and conformational analysis of the metastable states.

**Methodology:** Molecular dynamics simulations have been run with adaptive sampling approach and data has been collected from: 2090 trajectories, each run for 50 ns. Conformational analysis has been carried out using Markov State Models (MSM) and Convolutional Variational Auto Encoders (CVAE).

## 2.2 L2 $\beta$ -lactamases Hypothesis

**Background:** L1 and L2  $\beta$ -lactamases combine to induce antibiotic resistance in *Stenotrophomonas maltophilia*. L2  $\beta$ -lactamase, a class A  $\beta$ -lactamase, has limited research reported in the literature. The only experimental observations are that L2  $\beta$ -lactamase enzymes are inhibited by clavulanate and AmpR can regulate its function. To further understand the L2  $\beta$ -lactamase family, structural dynamics is needed to highlight the potential key residues and important regions that can have an impact on substrate binding and catalytic activity.

**Hypothesis:** The working hypothesis is to explore and understand the structural dynamics of four L2  $\beta$ -lactamases (L2a, L2b, L2c and L2d). L2  $\beta$ -lactamases should have important regions that influences the dynamics of the system. In spite of the high sequence identity between the four L2  $\beta$ -lactamases, there are still considerable differences exist between these proteins, the key residue difference will result in different dynamics pattern in the simulation. Such as the relatively low identical rate of L2d sequence compare with other three enzymes and the only stabilized  $\alpha$ 1 helix in L2b  $\beta$ -lactamase. The four  $\beta$ -lactamases still have various mechanisms. Identifying both common and unique features in these enzymes will help in further investigating the structural dynamics of L2  $\beta$ -lactamases.

**Aims:** The aim of this study is to use enhanced sampling molecular dynamics simulations to investigate:

- (a) What are the highly dynamic regions in the L2  $\beta$ -lactamase family.
- (b) How these high dynamic regions function and are correlated.
- (c) The similarities and differences in the dynamic regions of different L2  $\beta$ -lactamases.

**Methodology:** Molecular dynamics simulations have been run and data has been collected. Each L2  $\beta$ -lactamase has at least 1500 trajectories; each trajectory has been run for 50ns. Conformational analysis was carried out using Markov State Models (MSM), MDLofit, Bio3d and Convolutional Variational Auto Encoders (CVAE).

## 3. Methods

### 3.1 Molecular Dynamics

Molecular dynamics (MD) is a computational method that has been developed to study the motion of atoms and molecules. In 1957, the first MD simulation was carried out on primitive gases, and the first macromolecule simulation through MD was performed in 1977.<sup>107,108</sup> The 2013 Nobel Prize in Chemistry recognised the basics of MD simulation work as part of the achievements.<sup>109,110</sup> Nowadays, MD simulations play an indispensable role in the drug discovery process. A number of approved drugs have been identified using this approach, including captopril, ritonavir, saquinavir, indinavir and tirofiban.<sup>111,112</sup> The success in drug discovery and drug development makes MD an essential tool in the pharmaceutical industry.<sup>113</sup>

MD simulation is based on the principle of solving the Newton's equation of motion. The input is the position of atoms (coordinates). The method then calculates the forces on the atoms through the Newton's law ( $F = ma$ ). In the simulation, input molecules are regarded as four-dimensional substances.<sup>114</sup> The dynamic motions of the protein is an essential descriptors to describe the molecule movement.<sup>114</sup> The dynamic motions of the protein regulates the processing activities, such as catalytic activity and communications.<sup>115</sup> Then the force applied to the atoms generates motions and predicts the spatial position of each atom as it evolves through time.<sup>116</sup> In the dynamic systems, the time scale of the events occurring can range from femtoseconds to seconds. The

microsecond to millisecond time scale motions are considered as the most relevant to study function, as most conformational changes occur on this timescale. For example the slow backbone dynamics or the domain rearrangements can indicate the global translation of the protein.<sup>114</sup> The picosecond to nanosecond time scales motion are usually chosen for describing the fast motions, usually associated with side chain dynamics.<sup>114</sup> In the simulations, once the timescale is selected, the process is repeated and the new positions of atoms are calculated and updated. Summing up all the positions of atoms over time generates a trajectory. The potential energy of the system is determined by molecular mechanics methods.<sup>117</sup>

An MD simulation permits to study how each atom will move in a molecule or a protein. This reveals some important motions at atomistic level, such as protein folding, pocket opening and closing and ligand binding.<sup>118</sup> The conformational changes also lead to understanding a biomolecule response, which includes ligand interactions, stable intra- and intermolecular interactions and protein-protein interactions. Based on some of these capabilities, MD simulations have been shown to complement experiments. Starting from experimental data, MD simulations can be used to refine and optimise Xray derived structures.<sup>119</sup> The computational method based on the annealing protocol helps to fit structural data and keep it stable at the same time.<sup>120</sup> MD simulations have been found useful in detecting ambiguous ligand density in cryo-EM structures.<sup>121</sup> Beyond adjusting the existing states, MD is also widely used in building states.<sup>117</sup> MD simulation has been used to assess and estimate ligand binding. For instance, when the



ligand binding pose is more stable in simulations, the accuracy of it is more than the unstable ones.<sup>122</sup> Besides these, MD simulations are routinely used to calculate structural dynamics in macromolecules. The dynamic motion of water and ions can also reveal important aspects of protein function and ligand binding.<sup>123</sup>

## Force field

In MD simulation, atoms are represented by mass, potential energy and charges. These parameters are used to describe bonded ( $E^{bonded}$ ) and non-bonded ( $E^{nonbonded}$ ) interactions in mathematical form. In 1969, Levitt and Lifson proposed the empirical expression for the potential energy of a system.<sup>110</sup> The system potential energy ( $E^{system}$ ) at position  $x$  is the sum of bonded energy and non-bonded energy. A force field contains the sufficient details of potential energy function which include the functional forms and terms. Base on the information contained in the force field, the relative energy of the system is calculated and will be used to describe the atom position.<sup>124</sup> The covalent connections between atoms are called as “bonded” interactions, the interactions are used to describe the different dynamic forms of the molecule.

$$E^{system}(x) = E^{bonded}(x) + E^{nonbonded}(x)$$

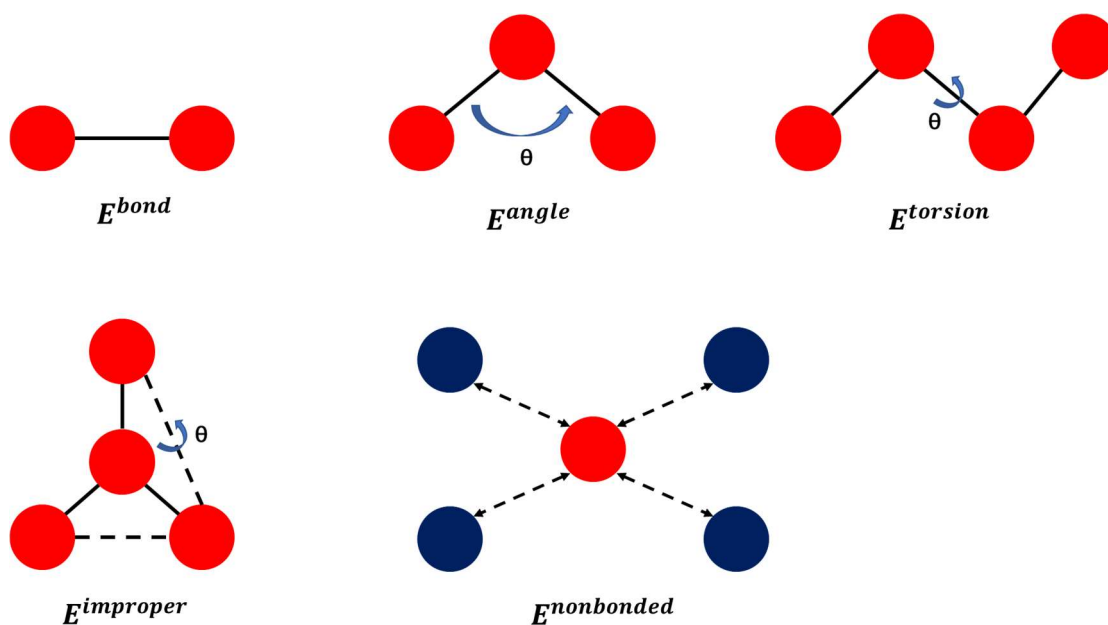
The bonded energy ( $E^{bonded}$ ) is the sum of bond stretching energy ( $E^{bond}$ ), angle bending energy ( $E^{angle}$ ), proper torsions energy ( $E^{torsion}$ ) and improper dihedral bending energy ( $E^{improper}$ ).

Non-bonded interactions ( $E^{nonbonded}$ ) consist of Coulomb electrostatics ( $E^{ele}$ ) and van der

Waals interactions ( $E^{VDW}$ ). However, some other terms can also describe the interactions of hydrogen bond and polarisable sites.<sup>117</sup> Figure 17 shows the representation of bonded and non-bonded interactions which are used to describe the forcefield interactions definition. An appropriate force field is essential for the simulation.<sup>124</sup>

$$E^{bonded}(x) = E^{bond}(x) + E^{angle}(x) + E^{torsion}(x) + E^{improper}(x)$$

$$E^{nonbonded}(x) = E^{ele}(x) + E^{VDW}(x)$$

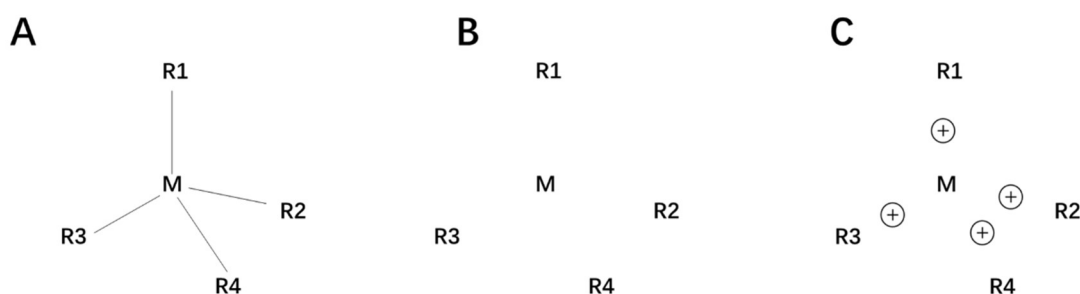


**Figure 17:** Representation of different forms of potential energy in the system. The potential energy of the system is the sum of bonded interactions and nonbonded interactions. Spheres represent the atoms, covalent bonds are represented as solid lines, arrows represent the relevant interactions.

In this work, zinc Amber force field (ZAFF) was employed for L1  $\beta$ -lactamase and Amber ff14SB force field was employed for L2  $\beta$ -lactamases.<sup>125,126</sup> The full name of Amber is “Assisted Model Building with Energy Refinement.”<sup>127</sup> Amber contains a set of force fields and also includes packages of molecular simulation programs.<sup>128</sup> The force fields and programs work together which make Amber become a powerful framework for computational calculation.<sup>129</sup> It was first developed by Peter Kollman's group at the University of California, San Francisco.<sup>128</sup> The Amber programs were used for performing and analysing molecular dynamics simulations, especially for proteins, nucleic acids and carbohydrates.<sup>127</sup>

As mentioned above, an appropriate force selection is essential for simulation. It should be emphasized here, there is no completely right or wrong in the force field selection, different force field suit different occasions. Force fields are used to describe and simulation the molecular dynamic process in the real environment. The real environment is complicated and force fields work on one protein probably will have poor results on the other proteins, especially for the innovation work. And that is the reason why molecular dynamic tools need updating. For instance, Amber has updated to Amber 23 version, each version will contain new features to describe the molecules including sampling method, new restrains parameter, etc.<sup>130</sup> The force field selection is in the “set up simulation” step, which means we can only check the results after generating the trajectories. Therefore, the force field selection is mostly based on researchers' publications and experience.

L1  $\beta$ -lactamase has a zinc binding site, therefore, a zinc force field need to be employed to describe the binding site. ZAFF was selected for this work. ZAFF force field used bonded plus electrostatics model approach to incorporate metal atoms into force field.<sup>126</sup> There are three main approaches to incorporate metal atoms into force fields: bonded model (Figure 18-A), nonbonded model (Figure 18-B) and cationic dummy atom model (Figure 18-C).<sup>126</sup> Bonded model describes the bonds and angles between metal atoms and ligands.<sup>131</sup> Bonded plus electrostatics model includes the electrostatic potential (ESP) changes more than bonded model.<sup>132</sup> The nonbonded model place integer charge on metal atoms while no bonds are added.<sup>133</sup> The cationic dummy atom model is related to nonbonded model. Instead of putting integer charge, it adds cationic dummy atoms around the metal atoms to mimic the valence electrons.<sup>134</sup>



**Figure 18:** three approaches of incorporating metal atoms into force fields. (A) bonded model. Including bonds and angles between metal atoms and ligands. (B) nonbonded model. Including integer charge on metal atoms. (C) cationic dummy atom model. Including cationic dummy atoms around metal atoms. Figure adapted from reference 128.

ZAFF is designed for four-coordinated zinc metal centre. It is a bonded plus electrostatics model. The bonded angle is decided by Hooke's law which mean it needs a force constant number. The force constant (spring constant) of ZAFF was calculated

based on Cartesian Hessian matrix (equation 1) through Seminario method. Seminario method was developed by Jorge Seminario in 1996.<sup>135</sup> It is developed to parametrize bond and angle parameters base on Hessian matrix. The Hessian matrix is the second derivation of energy with respect to coordinates.  $[k]$  is the 3N x 3N Hessian matrix of the molecule. Force constants  $\lambda_i$  and the normal mode energy  $\nu_i$  is obtained basing on the eigen analysis from Modeling ToolKit++ (MTK++) (equation 2).  $k_{AB}$  is the interatomic force constant between atom A and atom B (equation 3).  $\delta$  is the displacement of atoms and molecules. The calculation of force constant is to avoid defining internal coordinates. Electrostatics charge of ZAFF is obtained from the Merz–Singh–Kollman (MK) and restrained electrostatic potential (RESP).<sup>136,137</sup>

$$[k] = k_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j} \quad (1)$$

$$F_i = -[k]\nu_i \delta d = -\lambda_i \nu_i \delta r \quad (2)$$

$$\delta F_A = [k_{AB}] \delta r_B \quad (3)$$

There are several different force fields describe zinc atoms. The reason of choosing ZAFF is because it works well on other zinc binding site researches and it give good results after we generating the trajectories. The nonbonded method was also employed such as cationic dummy atom method which not worked well on L1  $\beta$ -lactamase.

For L2  $\beta$ -lactamases, ff14SB force field was employed. Compare with L1  $\beta$ -lactamase, L2  $\beta$ -lactamases are serine  $\beta$ -lactamases which does not have zinc ions binding site. The ff14SB force field selection is based on experience. It was updated from ff99SB force field which is one of the most widely used force field and have been used for more than 10 years.<sup>138</sup> The development of ff14SB is intended to work with TIP3P water model and it also improved the accuracy of protein backbone and side chain parameters.<sup>125</sup> ff14SB force field updated all the dihedral parameters of side chain.<sup>125</sup> The backbone parameters are based on alanine and glycine, the side chain parameters are based on wider and bigger molecules.<sup>125</sup> Both of backbone and side chain parameters are based on large amount of force field training base on quantum theory.<sup>125,139</sup>

## **Periodic boundary conditions**

Periodic boundary conditions are a series of boundary conditions. In molecular simulation, a macroscopic system contains moles of molecules which makes it large enough and approximates infinite space.<sup>119</sup> The bulk system needs periodic boundary conditions to be better described and approximated. A small box called “unit cell” is employed. The unit cell geometry is regarded as perfect two-dimensional tiling.<sup>140</sup> When an object get through one side of the unit cell in a certain velocity, it will appear on the opposite side of the unit cell in the same velocity.<sup>123</sup> The macroscopic system is approximated by infinite number of unit cells. In molecular dynamics simulation, one of these unit cells is the original one and the other unit cells are called copy images. Therefore, only the features of original unit cell need to be recorded, replicated and propagated.<sup>141</sup>

## **Ensembles**

Molecular dynamics simulation needs a specific ensemble at a specific time to describe a molecular condition.<sup>142</sup> Ensemble is a physical concept, to be more specific, a statistical mechanics concept.<sup>143</sup> It is an idealization which contains a large amount of a system’s virtual copies.<sup>143</sup> Each virtual copy represents a possibility of the system.<sup>144</sup> The concept of ensemble was set up in 1902 by J. Willard Gibbs.<sup>144</sup> A thermodynamic ensemble is a particular branch of the statistical ensemble and important to the molecular dynamics simulation. It is considered in statical equilibrium and employed to investigate the details of the thermodynamic system.<sup>142</sup> There were three important

ensembles defined by J. Willard Gibbs: microcanonical ensemble (NVE ensemble), canonical ensemble (NVT ensemble) and grand canonical ensemble ( $\mu$ VT).<sup>144</sup> Microcanonical ensemble assumes three primary macroscopic variables as constant number: particles in the system (N), system volume (V) and system total energy (E). Instead of total energy of system (E), the NVT ensemble assume particles in the system (N), system volume (V) and temperature (T) to be constant. The grand canonical ensemble assumes chemical potential ( $\mu$ ), system volume (V) and temperature (T) to be constant. There are also other important ensembles exist, such as isothermal–isobaric ensemble (NPT). Isothermal–isobaric ensemble assumes particles in the system (N), pressure (P) and temperature (T) to be constant. Because the experiments react at constant pressure and temperature under laboratory conditions, NPT is also important for observing more realistic variations.<sup>142</sup>

The MD simulation has its unique advantages when we compare it with experimental methods. First, the simulation will generate the coordinates and dynamic details of every atom in the system through time, which is very hard to achieve with any other experimental technique.<sup>119</sup> Second, the conditions in a molecular simulation experiment are precisely known and controlled such as the detail of protein structure, mutations, modifications made or other molecules present in the system. It is easy to compare the simulation results under different experimental conditions.



MD simulation in drug discovery provides an interesting example of how simulations can control and drive experiments.<sup>145</sup> The investigations in structural biology have found many key targets; for example in neuroscience drug discovery, such as transporters, GPCRs, ion channels, etc. The structure-based drug design requires the evaluation of dynamic properties in these proteins. In the generation of lead compounds, MD simulations can provide important information to improve the design. At a qualitative level, the efficacy and other properties that change will guide the optimisation. Simulations can predict the key interactions between pockets and ligands, the dynamic change of protein structural properties will guide the rearrangements of ligands.<sup>146</sup> At a quantitative level, simulations allow the estimation of ligand affinities accurately.<sup>147</sup>

Recently, MD simulations have attracted a lot of attention. The structural breakthrough in the resolution of a number of membrane proteins, such as GPCRs (G protein-coupled receptors) and neurotransmitter transporters, considered essential class molecules in neuroscience, have been researching accelerators.<sup>148</sup> Such proteins were historically considered difficult targets. The breakthrough in relevant fields provided MD simulation with its initial data as a starting point.

MD simulations have also become popular due to hardware and software technological advances. The computers available today are small and affordable. The graphics processing units (GPUs) make computers much more powerful than ever before and

allow researchers to run a simulation locally.<sup>149,150</sup> The hardware improvement further leads to a software supply explosion, such as Amber, ICM and Gromacs, which are designed to manipulate structural macromolecules. In this work, Amber was used to set up the simulation, ACEMD engine was used to generate trajectories, SWISS model was used to build homology models, PyEMMA was used to build Markov state models, MDLovofit was used to identify high dynamic regions, WISP was used to generate network analysis results and Bio3d was used to generate dynamic cross-correlation analysis. These methods were employed to generate the results and give more specific information about the proteins. The detail of these methods will be mentioned later in the results part. For analysing the results, VMD was the mainly used software.<sup>151</sup> After generating the Markov state models, VMD allowed us to load all the models with all the frames at the same time. By superimposing these models and focused on the high dynamic regions, important conformational change can be revealed. And the details of residue orientation change can also be easily observed. Furthermore, the RMSD, atoms distance calculation and dihedral angle plot are also generated from VMD. ICM was employed for analysis as well. It gave a clearer way to visualise the alignment of sequences and easy to select the residues base on the alignment.

These research productivity tools make MD simulations more efficient and accurate. Along with the trend of the times, the MD simulation threshold for entry is lowered and more widely used. The applications have become faster, cheaper and more accurate. The theoretical and practical advantages offer MD simulations a great developing

opportunity.

## 3.2 Homology Modelling

Homology modelling is a technique that uses high sequence identity to build atomistic models of “target” proteins based on a “template” protein. To be more specific, homology modelling is a structure prediction method in computational biology. It can help determine the 3D structure of proteins from their sequence. Before homology modelling was developed, the first crystal structure myoglobin was solved in 1960 through experimental methods such as NMR or X-ray.<sup>152</sup> These experimental methods have their own limitations. For NMR method, it is traditionally limited to small proteins.<sup>153</sup> The recent breakthrough of NMR based on isotope-labelling and pulse sequence techniques enable NMR to process large proteins which have hundreds of kilodaltons molecular mass.<sup>154,155</sup> And for running an X-ray method, the proteins need to be crystallized first.<sup>153</sup> Both of these two methods are time consuming because of their long experimental process. The homology modelling can help to overcome this. Homology modelling method is comprised on two premises. First, the protein sequence determines the homology. Second, the protein structure is conserved and therefore changes much slower compared with the sequence during evolution.<sup>156</sup> Thus, the similar homologous sequences of proteins will fold into similar 3D structures even when they have low sequence identity but have a reasonably high similarity. Homology modelling is considered as the most accurate computational structure prediction method.<sup>157</sup> It can provide high accuracy 3D structures of proteins with lower price and less time consumption.<sup>158</sup> Thus homology modelling is widely used for generating high quality 3D structures of proteins if there is a suitable template. However, homology

modelling also has limitations. The accuracy of homology models depends on the sequence identity rate between targets and templates. Over 50% of identity sequence generally leads to acceptable models while lower than 50% of identical sequence can result in serious error.<sup>158</sup> Therefore, the validation of homology modelling is always needed.

There are several steps for building a reliable homology model. 1) identification and selection templates. In this step the input sequence will query the PDB database (such as RCSB), templates will be searched and selected. The templates selection is based on several factors other than just sequence identity. For instance, templates from similar phylogenetic tree are also considered as an important factor.<sup>152</sup> 2) Sequence alignment and correction. The input sequence will be aligned with the template sequence after the template is selected. The correct alignment is crucial in homology modelling. 3) Building the model. Based on the template, there are several ways to build models. The model building could be classified in four ways: segment matching method, artificial evolution method, rigid-body assembly method and spatial restraint method.<sup>152</sup> The segment matching method is extracting a cluster of atomic position from template as the leading position. The selection of the segment is based on the energy, sequence identity and structure geometry.<sup>158</sup> The leading segments are used as a pillar to build the whole structure. The rigid body assembly method separates proteins into three parts: basic conserved core regions, side chains and loops.<sup>152</sup> The rigid bodies are aligned with the template and put together to build new model. The artificial evolution method

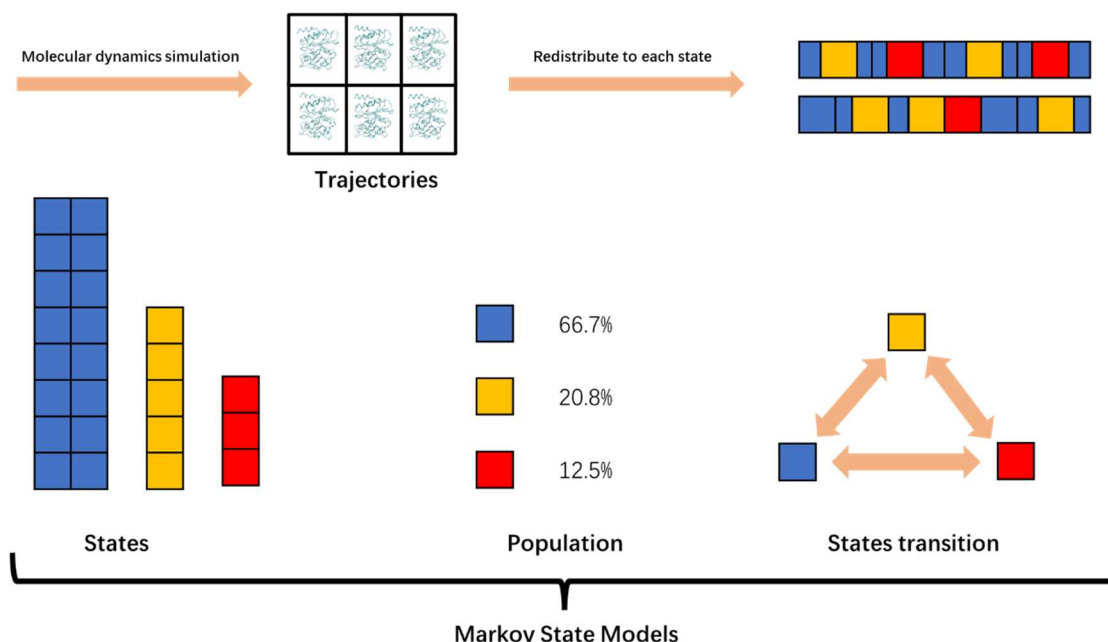
combines rigid body assembly method with stepwise template evolutionary mutations together.<sup>159</sup> The mutation will simulate until the template sequence is the same as the input sequence. The spatial restraint method use the restraints from templates to build the model.<sup>160</sup> The restraints contain bond angle, contact distance, bond length and dihedral angle. 4) Loop modelling. Loops are not conserved compared with the other secondary structures. Therefore, loops structure can be varied even without deletion or insertaion.<sup>160</sup> The loops can determine the specificity of proteins but they are more difficult to predict due to their flexibility. There are two main methods for building loops: database search and conformational search.<sup>160</sup> Database search will browse protein structures from the PDB database and detect optimal segments build loops. The conformational search is based on a score function optimization.<sup>158</sup> The possible conformation number will increase rapidly when loops get longer. This will make the conformational searching more time consuming. 5) Side chain modelling. In this step, the side chains are presented as low energy structures which are called as rotamers. The rotamers are determined by the backbone coordinates from the template and thereby determine the side chain model.<sup>152</sup> 6) Model optimization. After the model is built, the model needs to be optimized to eliminate errors. There are several ways to optimize model such as energy minimization, using more precise force field. 7) Model validation. The accuracy of model is important for further applications. Therefore, verification and validation are an essential step in homology modelling. The model could be checked by several ways, such as stereochemistry or Ramachandran plot.<sup>152</sup>

### 3.3 Markov State Models (MSM)

In 1983, Zwanzig indicated if the dynamics of metastable states could be sensibly chosen and sufficiently described, the process to reach certain states would not be recorded in the system. This could be a simple way to calculate the transition rate between different systems.<sup>161</sup> The idea was widely discussed in Van Kampen meeting and presented the frame work based on several key papers around the year 2000.<sup>162-165</sup> As the computational techniques and hardware rapidly develop, several present researches have benefited from it. This finally has led to the appearance of Markov State Models (MSM) around the year 2010.<sup>166-169</sup> In recent years, the method has been developed as a subfield of kinetic analysis and referred to as Markov State Modelling.<sup>170</sup> With the combination of variational approach, states selection and the support of open-source software, MSM can be employed in multiple field analysis such as protein-protein association, protein ligand binding and the dynamics of protein folding.

Combined with the molecular dynamics method, MSM was used to analyse the changes in the dynamics of the systems. MSM is a framework which means it has complicated mathematical functions to describe the entire dynamics of a system. In molecular dynamics simulations, MSM can also be used to identify underrepresented conformations which can be used as starting conformations to simulate dynamic systems. MSM is a transition probability matrix. Assuming the system is in a thermodynamic equilibrium condition; the configuration space is divided into several

discrete spaces of the system. Different states can represent a different dynamic process in the system. These states are separated by a chosen lag time. The transition probability, population and other parameters can be generated from the system. Therefore, appropriate states need to have appropriate lag time. The lag time needs to be long enough to describe the slow processes like conformational changes involving backbone dynamics and short enough to resolve fast dynamic motions employed by the side chains. In practice, the molecular dynamics simulation gives the input files (trajectories), MSM assigns the simulations to different states based on their different dynamic characteristics (Figure 19). The states generation process will also give useful information such as the population of states, how long does it take from one state to another and the free energy barrier between each state.



**Figure 19:** Diagrammatic representation of MSM. The trajectories generated from simulation will be assign to different states base on the selected structural features. In this process, useful information could be extracted such as popularity and state transitions. Figure adapted from reference 109.



The MSMs also promote the experimental methods development. Researchers have fine-tuned MSM to make experimental results observable. If the coarser experimental results could be consistent with the models generated from MSM, the structural details from MSM becomes more reliable. The consistency between simulation-based model and experimental results provide richer information of the dynamic system and also help to improve the accuracy of the dynamic system.<sup>171,172</sup> Besides, MSM has also been confirmed to identify the hidden binding sites (cryptic) in proteins.<sup>173</sup> Even without a proposed ligand, the hidden binding site could be revealed.<sup>170</sup>

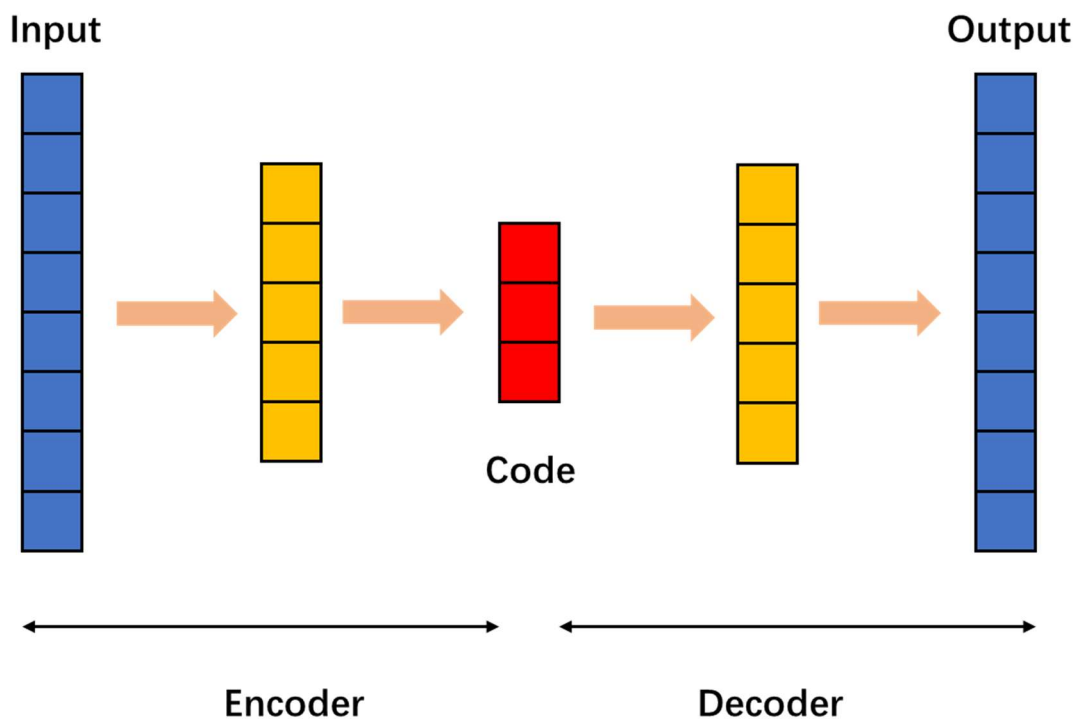
In this thesis, MSM will be used to explore different conformational states of both L1 and L2  $\beta$ -lactamases. Based on sufficient sampling, different states will be used to describe a variety of dynamic characteristics arising during the simulation. Other useful information will also be extracted to help further the research such as populations, free energy distribution and kinetic transition from one state to another. These models and other incidental details extracted from the simulations will help to improve the understanding the dynamic systems for further research.

### 3.4 CVAE-based Deep Learning

CVAE or Convolutional Variational Autoencoder is an artificial neural network-based unsupervised deep learning technique. To understand what CVAE is, autoencoders (AE) need to be explained first. The autoencoders responsible for the main function of CVAE, C (convolutional) and V (variational) are the extension method base on autoencoders. Autoencoder is an unsupervised deep learning technique. It can identify the important information such as underlying correlation base on the neural network models. The generalised architecture of the autoencoder is shown in Figure 20. When the input file is received by the encoder, data will be compressed into a lower dimension space. The lower dimension data storage location is termed the latent space. In the latent space, an appropriate number of nodes need to be chosen to determine the dimensionality of the data. After selecting the dimensionality, the data will be put into the decoder for reconstruction. In most cases, the decoder is a mirror image of the encoder, the input data for the decoder only comes from the latent space. During the encoding and decoding, the autoencoder can be trained by minimising the reconstruction error. Therefore, the autoencoder can learn a lower dimensional representation of the original data. The potential correlations and important information can be found through neural network states in latent space, while insignificant information could be ignored during the training.

Variational Autoencoder (VAE) is an extension based on the autoencoder..<sup>174</sup> VAE provides a continuous and smooth latent space representation by using a different

method to describe the latent space. Instead of using a single value to describe each dimension in the latent space, VAE use probability distribution to describe each latent distribution in the encoder, mean and standard deviation of a Gaussian in most cases. An input vector is generated for the decoder, the vector selection is based on a random sample from each distribution. The samples in each distribution have similar values, therefore, the vector will help the decoder to give an accurate reconstruction.



**Figure 20:** Autoencoder architecture. The squares indicate the dimensions in the system. The code gets in the system and making a dimensionality reduction through encoder then put into decoder to do a reconstruction.

Convolutional Variational Autoencoder (CVAE) is a variation of VAE, where the convolutional layers have been added to the neural network states which arranges the data into a normal distribution in latent space. After the input matrix is entered in the

convolutional layers, the convolutional filter will be employed for each slice of the matrix. This is followed by the generation of a feature map. The feature map contains the strength and location information of detected features in the input file. The convolution filter will repeat in the convolution layers and allows the proteins with similar structure and energy characteristic to be clustered.<sup>175</sup> Through clustering, the CVAE can provide a more efficient way to capture the key information in the system, especially for the secondary and tertiary structures.

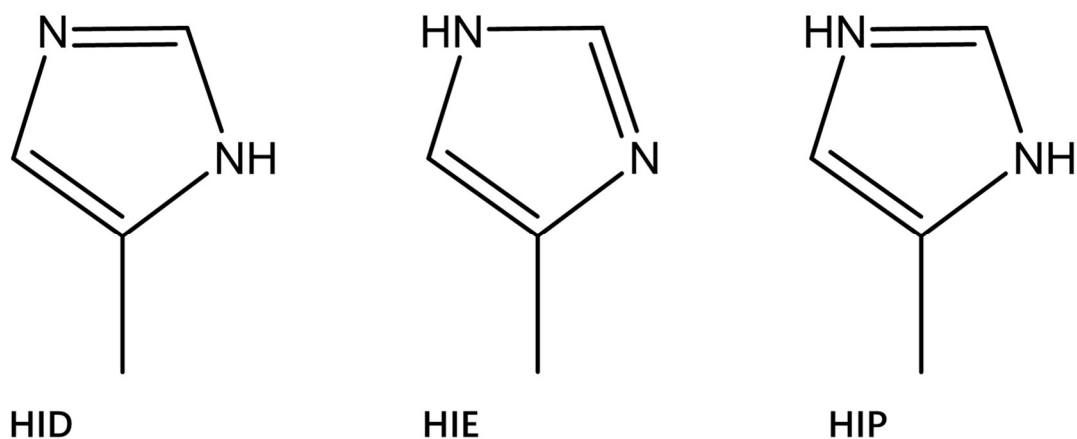
In this research work, CVAE will be employed to study both L1 and L2  $\beta$ -lactamases. More specifically, the CVAE will be used to investigate the co-operativity of the four subunits in the L1  $\beta$ -lactamase homo-tetrameric enzyme. Understanding the dynamic patterns of co-operation between identical subunits of L1  $\beta$ -lactamase, can help in developing effective strategies for drug design, leading to the inhibition of the enzyme. In L2  $\beta$ -lactamase family, four different L2  $\beta$ -lactamase proteins (L2a, L2b, L2c and L2d) will be combined into one single CVAE model. To investigate the distribution of structures and extract the key conformations, a comparison of the dynamics between different enzymes of the L2  $\beta$ -lactamases family will be made. The results will be put into both parallel and vertical comparison to investigate the important regions in individual proteins and the differences in same region between each protein. The resulting analysis will help to improve the understanding of the dynamic systems of the L2  $\beta$ -lactamase family.

## 4. L1 Metallo- $\beta$ -lactamase Results

### 4.1 Set up and Running the Simulations

The structure file of L1 metallo- $\beta$ -lactamase (PDB entry 1SML) was obtained from the protein data bank.<sup>59</sup> The tetramer was generated using the protein quaternary server (PQS).<sup>176</sup> The tetramer file needed to be fixed before assigning the protonation states to the side chains. The HETATM record in the PDB file was changed to ATOM, all hydrogens and inconsistent naming were removed, TER were added between each chain, water molecules and zinc atoms. The fixed file was imported into H++ for protonation. H++ is a software which calculates ionizable groups at a particular pH value.<sup>177</sup> H++ can also add missing hydrogen atoms based on the environment pH value.<sup>178</sup> The pH was set at 7.0 in L1 simulation. H++ added all the missing hydrogen atoms to the protein during the process and generated the protonated output file. Amber20 was used to process the output file from H++ to generate the coordinate file (.inpcrd file) and topology file (.prmtop file)<sup>126</sup>. Zinc Amber force field (ZAFF) was employed to run the simulations.<sup>179</sup> Before parsing the output file from H++, the file was fixed into ZAFF format. All the histidine residue names (HIS) in the binding site of each subunit were changed to the protonation type name (HID, HIE and HIP) (Figure 21). Histidine has three types of protonation type, HID, HIE and HIP. The protonation type can change depend on pH value.<sup>180</sup> When pH is around 7, two neutral protonation type HIE and HID exist. HID has hydrogen on the  $\delta$  nitrogen. HIE has hydrogen on the  $\epsilon$  nitrogen. When pH is low, positively charged protonation type HIP exists. HIP has

hydrogens on both nitrogens. When pH is high, imidazolate ion will be produced. Since the simulation have already set pH at 7, only HID and HIE will exist in the molecule. The protonation type can affect the results in the molecular dynamic simulation.<sup>181,182</sup> Therefore, a correct protonation type of histidine is important. Residue names of cysteine of disulphide bond in each subunit were changed from CYS to CYX, two zinc ions residue name in each subunit (Zn1 and Zn2) were changed to ZN6 and ZN9. Therefore, Zn1, H116, H118 and H196 were changed into ZN6, HD4, HE2 and HD5, respectively. Zn2, D120, H121 and H263 were changed into ZN9, AP2, HD8 and HD9, respectively. To load the ZAFF format file, some atom types were added to the ff14SB force field in the following format: ZN6—{{“ZN” “Zn” “sp3”} {“N5” “N” “SP3”} {“N6” “N” “sp3”} {“N7” “N” “sp3”}}, ZN9—{{“ZN” “Zn” “sp3”} {“NQ” “N” “SP3”} {“NP” “N” “sp3”} {“E2” “N” “sp3”} {“D2” “O” “sp3”}}. Atomic ions library file(atomic\_ions.lib) and parameter modification file (frcmmod.ions, 1lsm\_hfe\_tip3p, ZAFF.prep, ZAFF.frcmod) were eventually loaded using the tleap command. After loading the input file, the system was solvated using the TIP3P water box. The box edge to the closest protein atom was at least 12 Å.



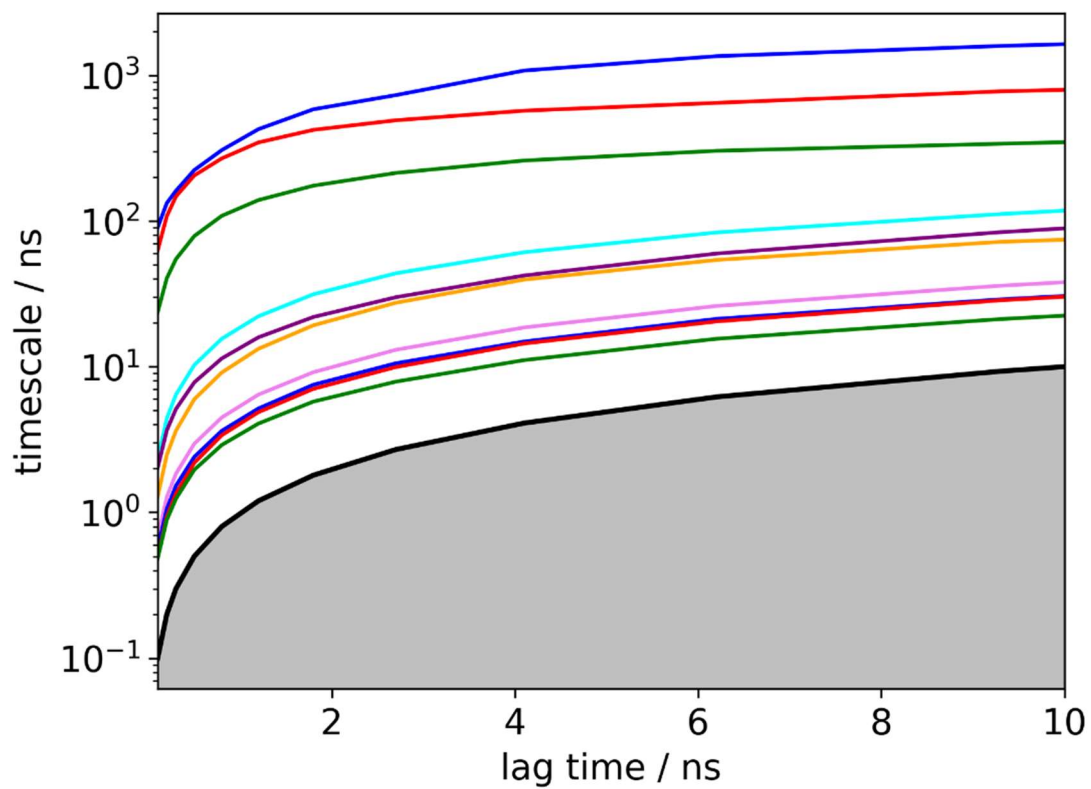
**Figure 21:** Different protonation type of histidine. HID: Histidine with hydrogen on the  $\delta$  nitrogen. HIE: Histidine with hydrogen on the  $\epsilon$  nitrogen. HIP: Histidine with hydrogens on both nitrogens. HIP is positively charged.

The Amber ff14SB force field with ZAFF parameters was used to generate the topology and coordinate files. The electrostatic interaction distance was set at  $\leq 9 \text{ \AA}$ . During the simulations, the systems were first energy minimised for 1000 iterations of steepest descent. Then an equilibration protocol was run for 5ns at 1 atmospheric pressure, 300K temperature using the NPT ensemble. After equilibration, Markov State Models based adaptive sampling simulations was run for 50 ns. Instead of sampling from same conformational regions, adaptive sampling is designed to explore more under-sampled regions.<sup>183</sup> After one simulation is done, the following simulations will utilize the information from all previous simulations to determine the starting point.<sup>184</sup> In this case, it will focus on the most under-sampled regions in all previous simulations. The ACEMD molecular dynamics engine was employed for the simulation.<sup>185,186</sup> The frames were saved every 0.1ns in the production run. 2090 trajectories were generated, each trajectory contained 500 frames.

## 4.2 Markov State Models

PyEMMA v2.5.7 was employed to build the Markov State Models (MSM).<sup>187</sup> All trajectories were loaded, backbone torsion and RMSD of all residues in gating loops (G149-D171 and A219-H239) from each subunit of the tetramer were selected as the input feature. The frame reading stride was set at 5, the tICA (Time-lagged independent component analysis) method was employed for data reduction, all the data was allocated into 3 ICs (independent components), the lag time was set at 1.0ns and the cluster number was set at 30. 7 states were built based on the implied timescale plot (Figure 22). The lines of the implied timescale plot were the maximum likelihood results of L1  $\beta$ -lactamase simulation. The different colour lines indicate the maximum likelihood results. The thick black line over the grey area indicates when lag time and timescale were converged. The grey area under the black line describes an area where MSM cannot solve processes. The lag time need to be selected when all the results are converged. This is when the lines become flat, 5.0ns lag time was deemed acceptable and selected for subsequent calculations.

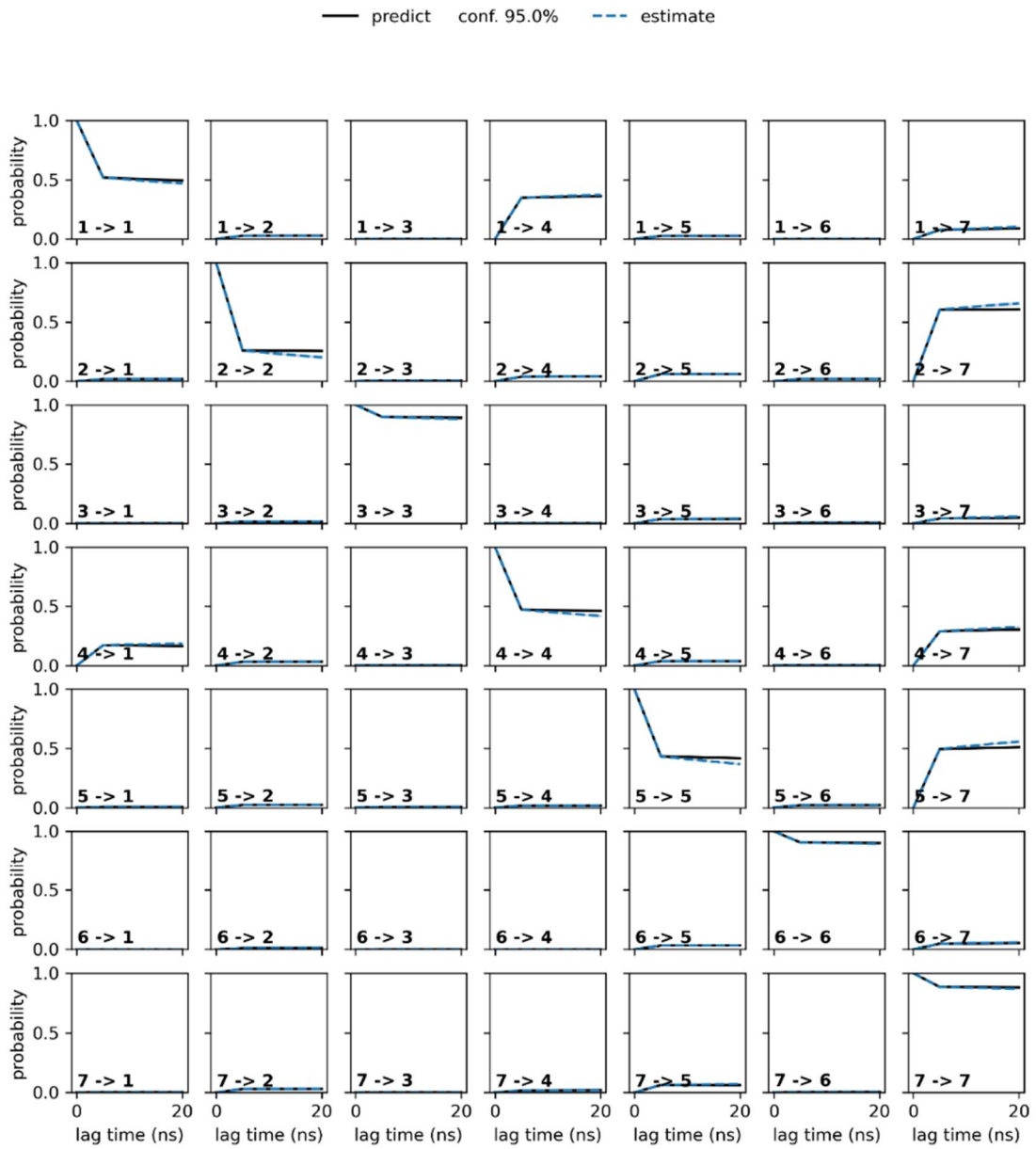




**Figure 22:** Implied time scale plot of L1  $\beta$ -lactamase. The different colour lines indicate the maximum likelihood results which show the converged behaviour of implied timescale. The lines become flat at 5.0ns which indicate the results are converge at 5.0ns.

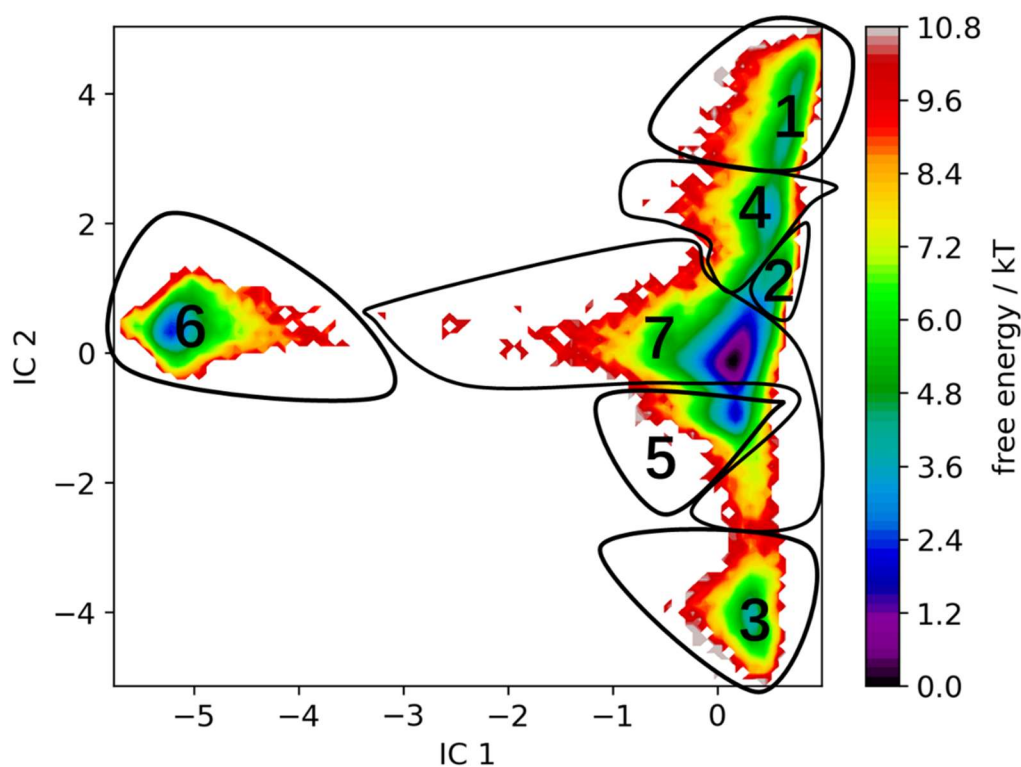
CK test plot (Chapman-Kolmogorov test plot) is an essential step in MSM validation. The MSM simulation can only continue after passing the CK test plot. In the last step, implied time scale can provide information to select an appropriate lag time to resolve the dynamics. However, it only focuses on eigenvalues and ignores the eigenvectors. The eigenvectors contain more details of the simulation.<sup>188</sup> Therefore, CK plot is employed for a stronger validation. The CK test is based on the assumption that at a given lag time  $\tau$ , the system will resolve into a number of states (models) that should explain the dynamics of the system. The prediction will be compared with another independent estimation. The standard CK test calculates the transition possibility between each metastable states based on that lag time.<sup>188</sup>

The 7 state MSM model generated need to pass CK test. Figure 23 shows the CK test plot result. The CK plot of 7 metastable states with a lag time of 5.0ns illustrates an estimated (solid line) and a predicted (dashed line) projection within confidence intervals. If these two lines overlap with each other, it indicates the states pass the Bayesian sampling procedure with 95% confidence intervals. The results of L1  $\beta$ -lactamase CK plot indicate the estimated and the predicted values are within the range of confidence interval.



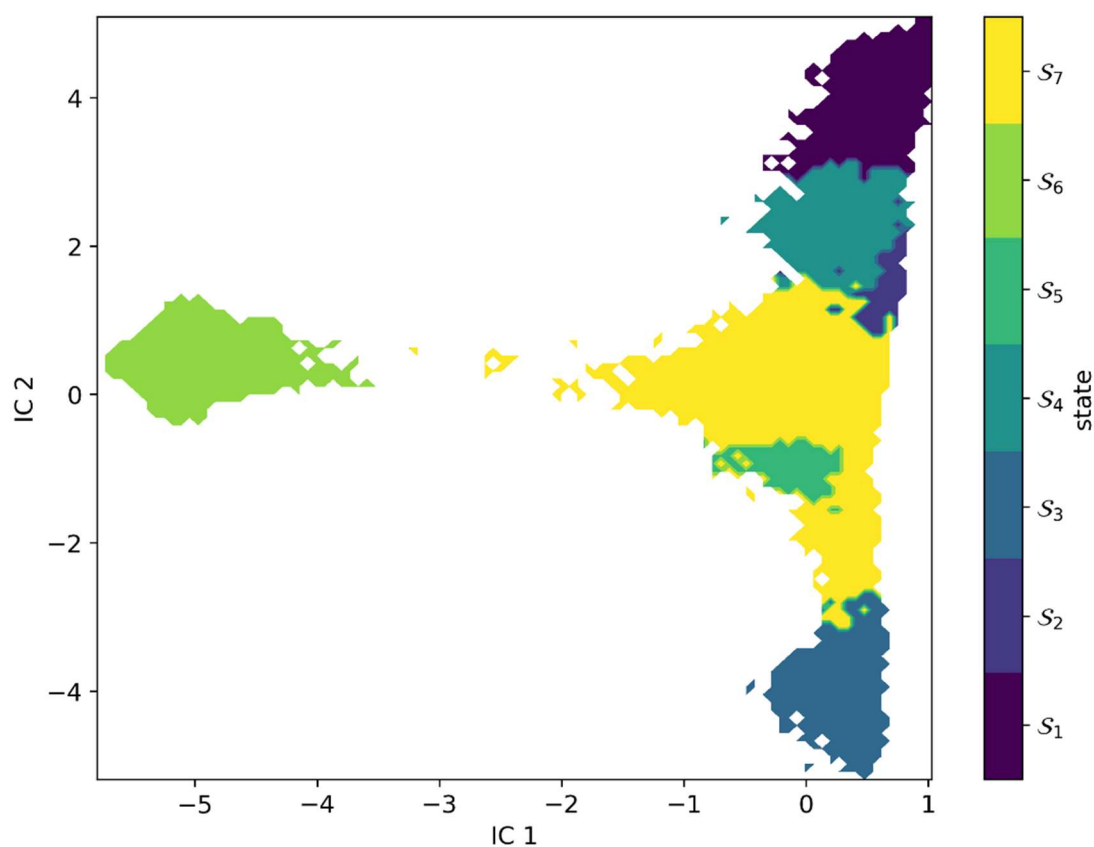
**Figure 23:** Chapman-Kolmogorov test plot of L1  $\beta$ -lactamase. All 7 states passed the CK test within confidence intervals.

The free energy landscape at lag time 5.0ns was projected onto first 2 independent components. Multiple free energy minima can be observed in this plot (Figure 24). The PCCA distribution separates different metastable states into different energy minima. The two main minima are observed for state 6 and 7. The other sub minima are present for states 1 to 5. The metastable states need to be clustered clearly on the distribution plot to make sure the conformations in each state have the same features.



**Figure 24:** L1  $\beta$ -lactamase energy landscape. Multiple energy minima could be observed in the plot. The structures distributed in each energy minima had the same features. The plot could be used to verify if the PCCA distribution plot is reasonable or not.

The Perron Cluster-Cluster Analysis (PCCA) distribution plot highlights the location of the 7 metastable states (Figure 25). A clean distribution indicates clear conformational clustering, i.e. the conformations with same features are clustered and distributed into the same state. The first two independent components were used to demonstrate and identified 7 defined metastable states.

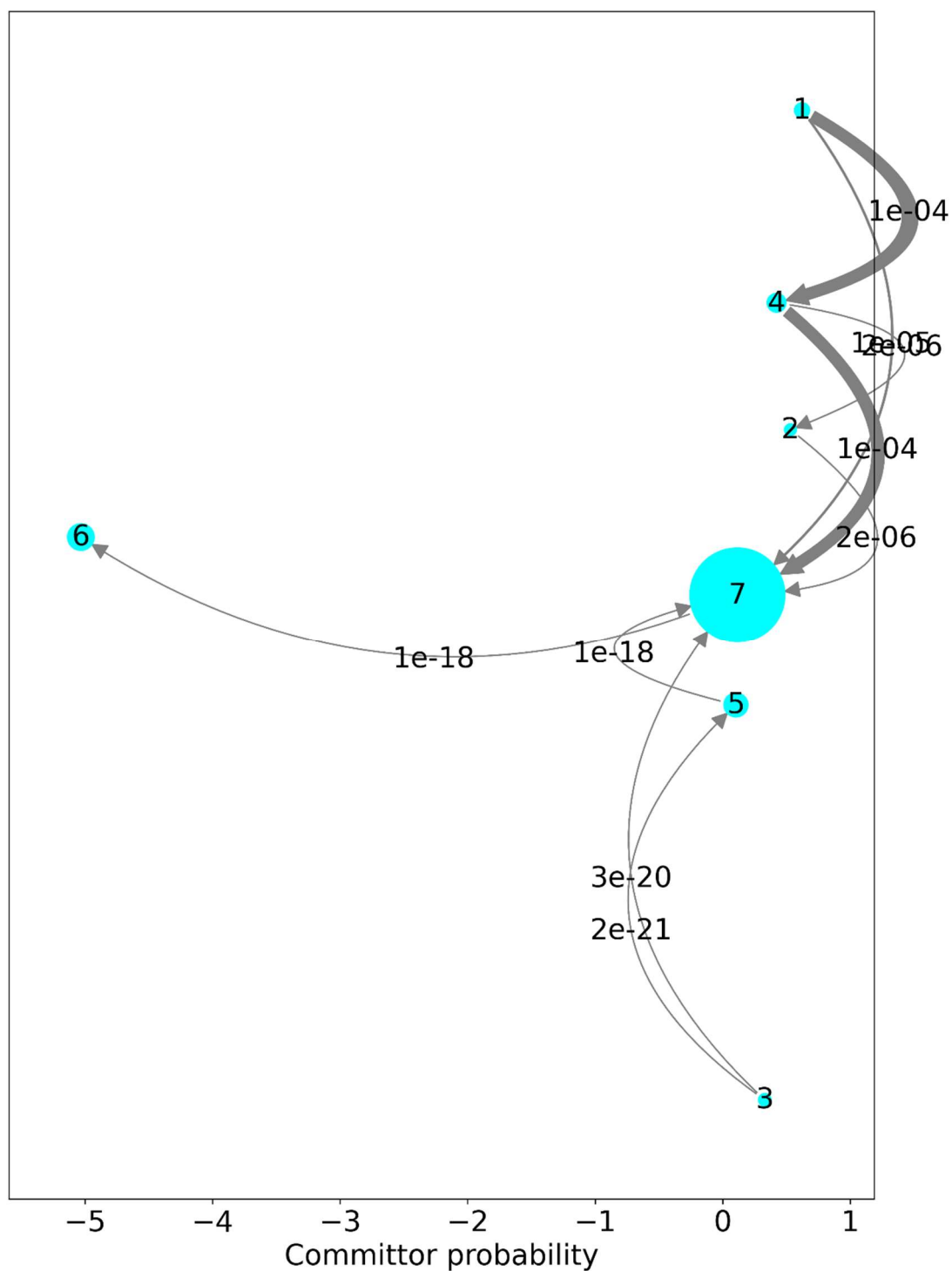


**Figure 25:** The PCCA distribution plot for 7 states in L1  $\beta$ -lactamase. Different colours in the distribution represented different conformational states in simulation. The conformations in the same state display similar features.

To understand the net flux pathways, the transition path theory (TPT) needs to be mentioned first. TPT is a theoretical tool for analysing the transition path ensemble.<sup>189</sup> To be more specific, researchers want to use TPT to solve rugged energy landscapes.<sup>189</sup> Before TPT is developed, transition state theory (TST) and transition path sampling (TPS) was used to solve the rugged energy landscapes.<sup>190-192</sup> TST is a well-known theory used as the cornerstone for chemical physics.<sup>191-193</sup> It provides the intuition, language and foundation for developing computational tools.<sup>191</sup> The TST has its limitation on solving rugged energy landscape, the complex situation makes the notion of transition state become not clear and result in the inaccuracy.<sup>189</sup> Different from TST, TPS focuses on the transition path ensembles instead of transition states.<sup>194</sup> However, TPS also has limitation on solving the complex system. Focusing on rare event occurs will make the transition paths themselves become too complicated to analyse.<sup>190</sup> Therefore, TPT was developed. TPT focus on the transition state ensemble and transition tubes.<sup>189</sup> It can solve the following questions: (1) what is the probability distribution (2) what is the transition paths probability (3) what is the transition rate.<sup>195</sup> TPT allows a precise and more complete picture regardless of the complicated system.<sup>189</sup>

The net flux is generated by TPT object in MSM. It calculates the productive recrossing between pairs of states. To be more specific, net flux is the difference between value of the fluxes in both directions. The net flux analysis provides the kinetic mechanism of interconversion between conformations during the simulation. It shows the pathways

of system which go from A to B and also provide the contribution of these pathways to the whole system. Net flux pathways plot of L1  $\beta$ -lactamase is shown in Figure 26. In L1  $\beta$ -lactamase, most structures went from state 1 to the other states. State 7 is the most populated state. 1-4-7 (89.27%) is the dominant transition pathway of conversion between the metastable states. The thickness of arrows between states indicated the probability of transition. The thicker arrow indicates the higher possibility for transition between the two states.



**Figure 26:** Net flux pathways plot of L1  $\beta$ -lactamase. Each cyan circle with certain number represented the corresponding metastable state. The bigger the circle, the more populated the state. Arrows between each state represented the possibility of each transition. The thicker the arrow, the higher the possibility. Most structures went from state 1 via state 4 and stayed at state 7.



Mean first passage time is an important parameter in MSM. It is an average timescale for a random event to first occur.<sup>164</sup> Mean first passage time between different states of L1  $\beta$ -lactamase was shown in Table 1. The number shows different timescale of transition from one state to another. The larger timescale implies a higher energy barrier and lower transition possibility between states. The larger the number, the fewer the occurrences of the conformational change. 0.00 indicates there is no energy barrier when states make no transition. Table 2 is the population and free energy estimation for the 7 metastable states of L1  $\beta$ -lactamase. State 7 represented for over 80% of all the structures in the simulation.

**Table 1:** Mean first passage times between different states in L1  $\beta$ -lactamase. Most structures went from state 1 to state 4 and transit from state 4 to state 7.

State/(ns)	1	2	3	4	5	6	7
1	0.00	2829.89	47074.03	114.57	1559.23	17125.10	269.20
2	7796.47	0.00	46768.51	4373.55	1262.80	16838.53	30.69
3	8718.04	3405.80	0.00	5297.76	1581.42	17465.74	631.38
4	3142.98	2750.64	46995.67	0.00	1480.80	17046.95	192.47
5	8124.86	2822.16	46503.00	4704.48	0.00	16855.85	5666.00
6	9153.89	3861.98	47850.15	5733.95	2318.63	0.00	1087.37
7	7977.68	2677.59	46674.18	4557.67	1141.23	16772.15	0.00

**Table 2:** The population of different states ( $\pi$ ) and their free energy estimation.

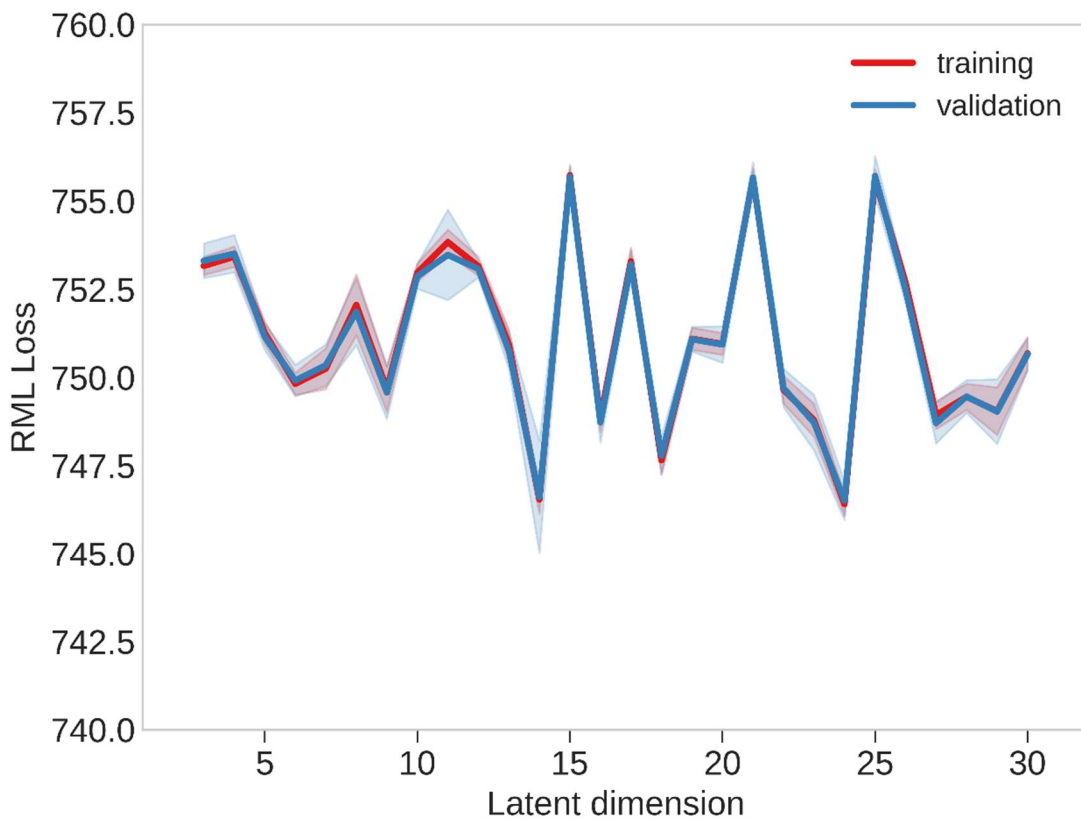
state	$\pi$ (%)	G/kT (J)
1	2.0389	3.892736
2	1.2309	4.397427
3	1.3944	4.272701
4	3.1653	3.452921
5	5.1502	2.966125
6	6.3325	2.759479
7	80.6877	0.214584

### 4.3 CVAE-Based Deep Learning

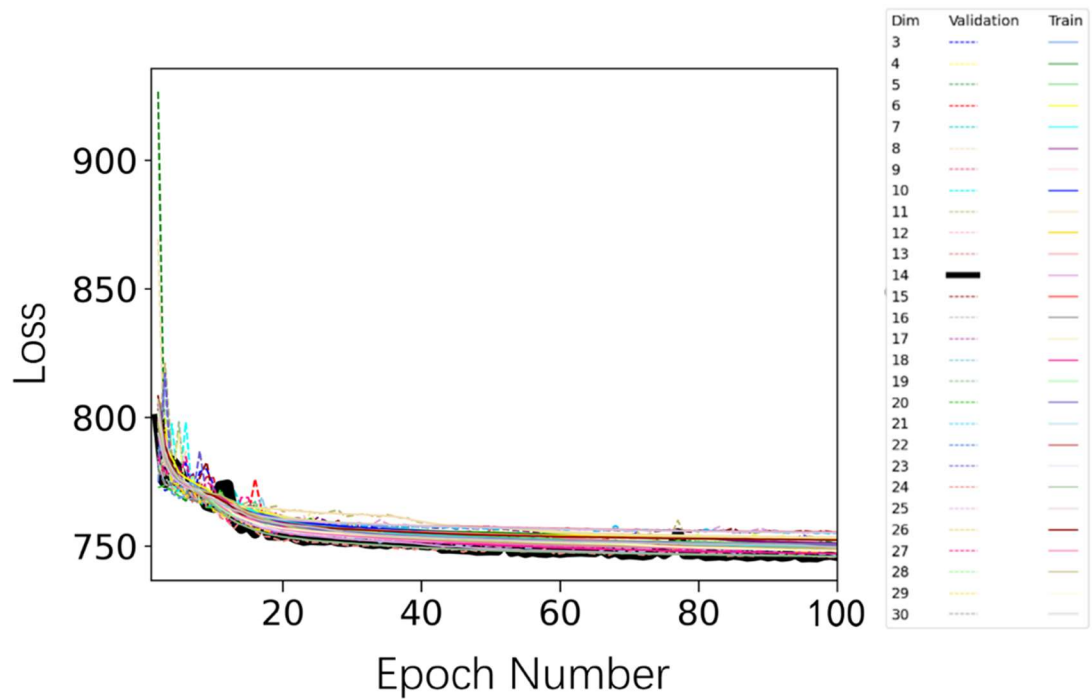
CVAE was applied to analyse the cooperativity between the four subunits in L1  $\beta$ -lactamase. For running CVAE, 3 steps need to be employed. First is the pre-processing of trajectories. In this step, RMSD, native contacts and distance matrices needed to be generated from the trajectories. Distance matrix does not contain any information on the translational and rotational motions of the protein. Translational and rotational motions are unwanted parameters from the dynamics in most cases.<sup>175</sup> RMSD, native contacts and distance matrices files were combined into a single system file. In the work presented here, the system file was generated for each subunit individually to investigate the dynamic pattern of each individual subunit. Then, the four subunits system files were combined into a single system file that contained features of all 4 subunits albeit individually.

The second step was to run the CVAE and select an appropriate latent dimension number. The purpose of this step was to minimise information loss through the training epochs. In this experiment, CVAE latent loss is calculated at each dimension from 3-30 (Figure 27). Each dimension runs 100 epochs to ensure that the training and validation loss is converged (Figure 28). After running CVAE, an optimal dimension needs to be selected. The dimension with the lowest loss is selected to post process the data. The CVAE results were inspected by repeating the training process with optimal latent dimension. The calculated loss from latent dimensions was shown in Figure 27. The red and blue lines represent the training and validation respectively. The lowest

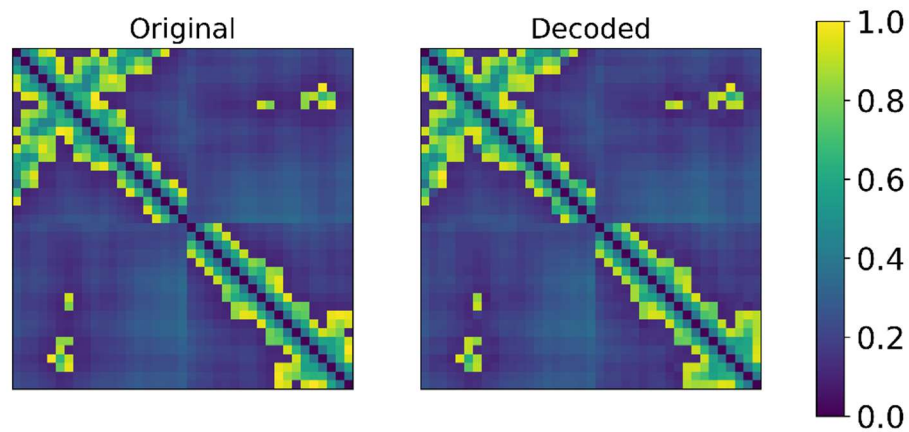
loss was identified to occur in dimension 14. The loss observed in latent dimensions by epoch numbers is shown in Figure 28. The most optimal selected dimension 14 is highlighted with black. Then the output from dimension 14 was compared with the original data to ensure that the decoded result had no significant loss of information and was a good representation of the encoded data (Figure 29). The comparison between original encoded data and reconstructed decoded data verified that the compression of dimension 14 did not result in a significant information loss and can be used in the next step.



**Figure 27:** Evaluation of information loss over latent dimensions. Red line was the training loss, the blue line was the validation loss. Latent dimension 14 show the best result with the lowest information loss.



**Figure 28:** Evaluation of information loss of different latent dimensions by epoch number. The lowest loss dimension was highlighted with black colour.



**Figure 29:** Comparison of the original input file and reconstructed decoded file, the compression of dimension 14 did not result in a significant information loss.

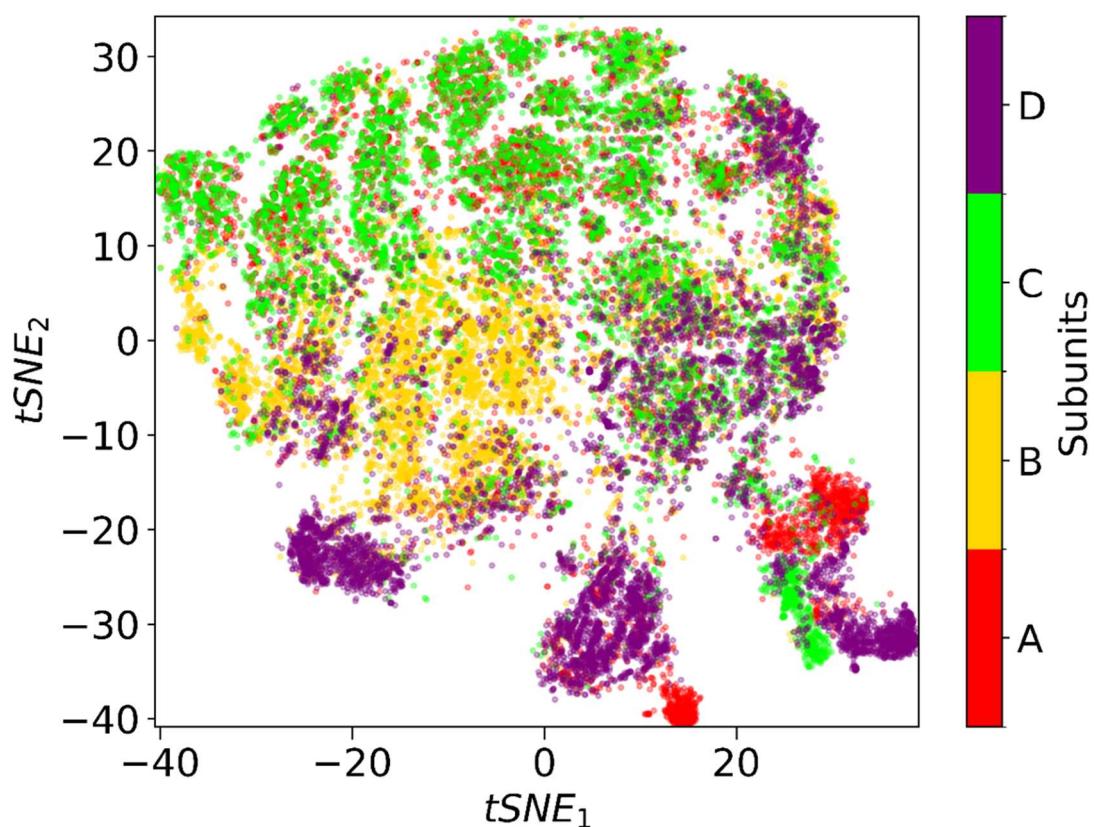
After selecting dimension 14 as the optimal CVAE reduced dimension, the last step was to post process and visualise the data. In order to make the result easier to understand and analyse, the simulation results were first projected into a 3D latent space representation and then projected to a 2D plot for conformation extraction. Each conformation from the simulation was represented as a corresponding coordinate. In this case, t-SNE (t-distributed stochastic neighbour embedding) method was employed to reduce the data.<sup>196</sup> For easier understanding of the CVAE latent dimension, conformations were also projected onto 2D plot with the t-SNE method.

In L1 MBL CVAE experiment, all frames were put into the system to generate a general conformation distribution plot for all 7 states (Figure 30). The different subunits are represented with different colours, (subunit A-red, subunit B-yellow, subunit C-green and subunit D-purple.) Then the seven states' structures which was generated from MSMs were projected on the general plot separately. This step was to check how four subunits cooperate in each metastable state.

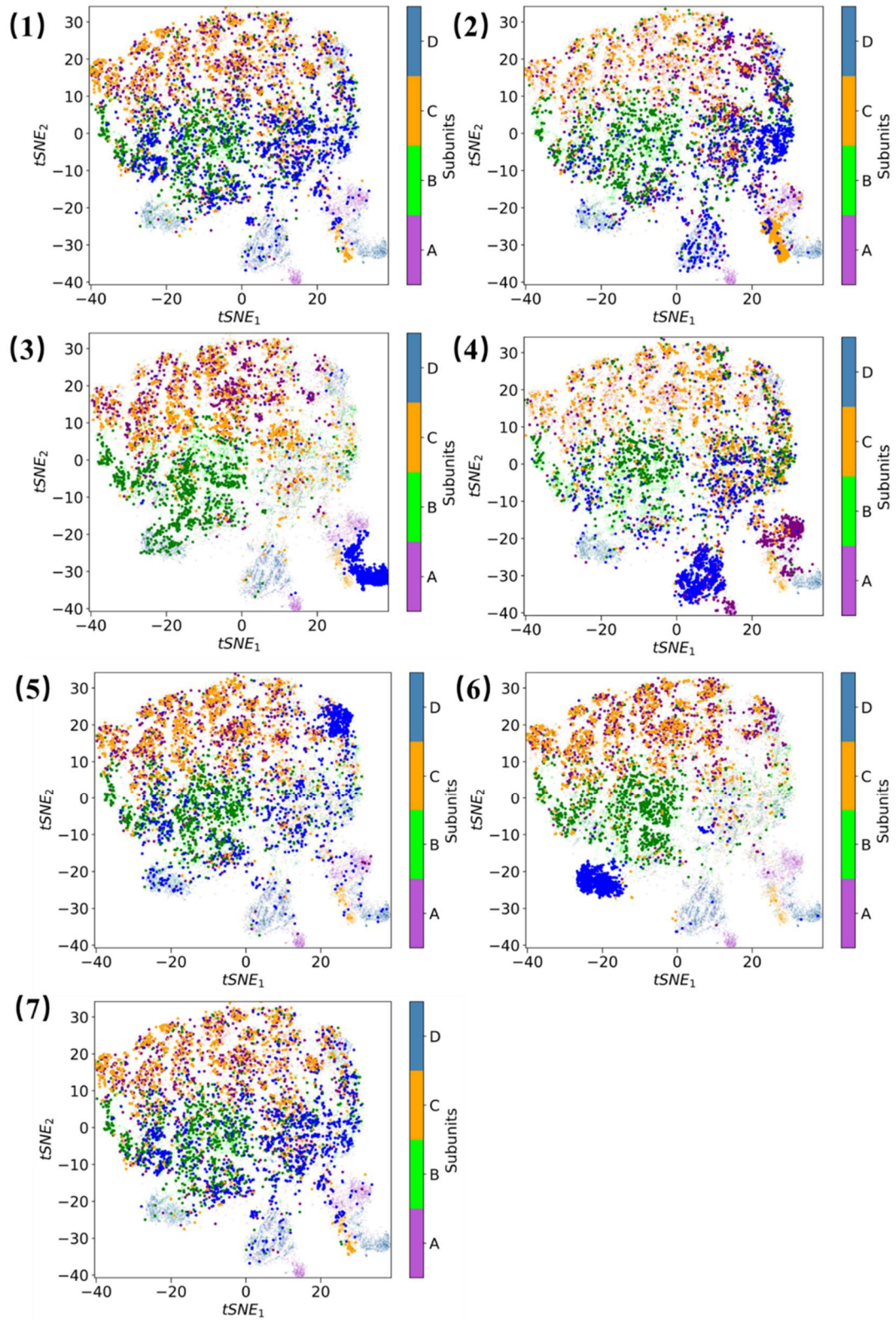
A combined general conformation distribution plot for all 7 metastable states of L1 MBL is shown in Figure 30. The separated conformation distribution plot for each metastable state is shown on Figure 31. Each state provided 1000 structures as representation. A total of 7000 structure distribution are shown on this plot. The different subunits were given distinct colours, namely subunit A-red, subunit B-yellow, subunit C-green and subunit D-purple. All subunits sample similar motions, as the conformations are clustered based on comparable dynamics. Subunits A and C share similar dynamics, while subunits B and D are clustered in the same tSNE space. Each metastable state provided the same number of structures rather than based on their population. The less populated state will have the same number of structures as the dominant state on the general plot. Therefore, the general plot cannot show the population of metastable states and looks congested.

In the CVAE distribution, all subunits sampled similar motions. The conformations clustered base on the comparable dynamics. Based on the conformation distribution plot shown in Figure 30, most subunit A and C structures clustered together. Only a small cluster of subunit C in state 2 and a similar cluster distribution of subunit A in state 4 separated. The mixed distribution of subunit A and C was observed in all 7 metastable states and represented for an overwhelming majority. This indicates that the subunit A and C displayed similar dynamics during the simulation. In contrast, subunit B and D clustered together in state 3, 4, 5 and 6. However, these states only represented for about 16% of all the structures. The majority of subunit B and D showed mixed

distribution. Therefore we can conclude that subunit B and D also showed a similar dynamic patterns during the simulation. Comparing the dynamics, the correlation between subunit A and C was stronger than the correlation between subunit B and D.



**Figure 30:** General conformation distribution plot for all 7 metastable states. The results show that A and C subunits show similar dynamics, while B and D subunits cluster together.



**Figure 31:** All 7 metastable states of L1 MBL conformation represented individually distribution plots.



## 4.4 Network Analysis

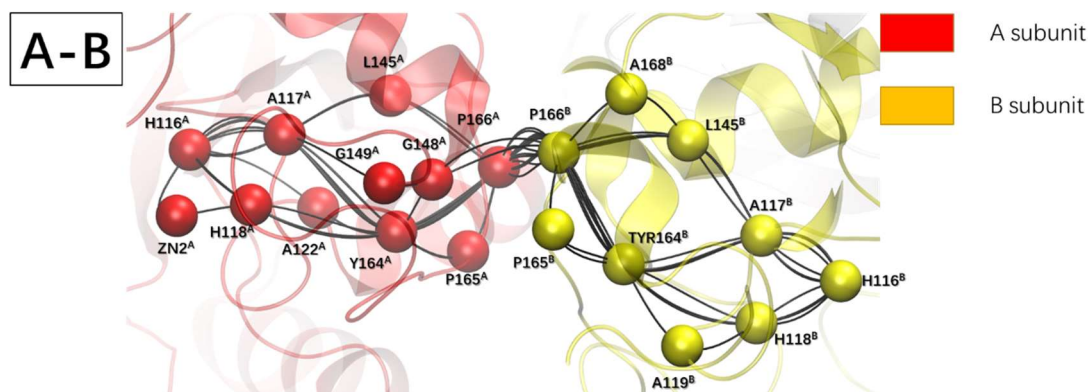
The network analysis was also employed to investigate how the signals transferred between each subunit of the L1 MBL tetramer. Weighted Implementation of Suboptimal Paths (WISP) was used to study the communication between subunits.<sup>197</sup>

The network analysis is employed to understand how conformational change in one part of the protein can influence other far away sites in other regions of the protein. In traditional network analysis, the computational methods focus on the most optimal paths from the allosteric site to the binding site. WISP analysis also includes the suboptimal paths which could also impact the regulation of active site.<sup>197</sup>

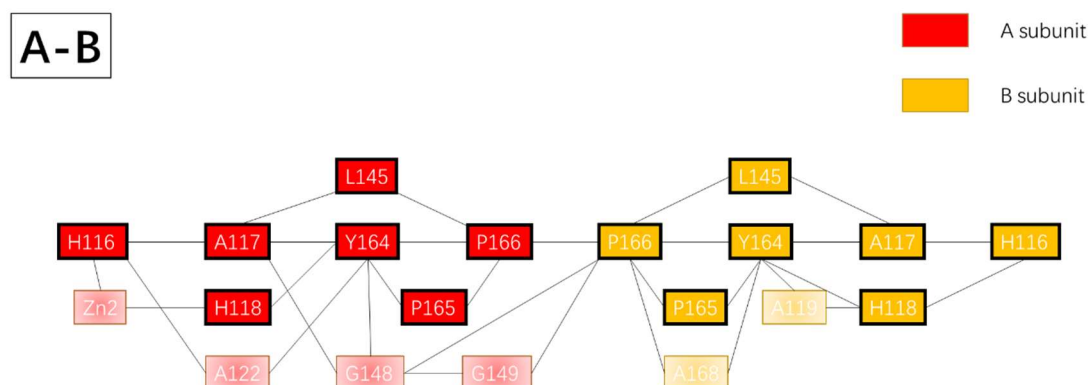
In L1 MBL, network analysis focused on how the four subunits communicate in the tetramer. Since the structure is symmetric, the binding site histidine (H116) was chosen as the start node in one subunit and also as the end node in the adjacent subunit. The most optimal paths between the two residues were then identified. The paths from one subunit to another gave the information of how one binding site influences another. The study of nodes and paths will help to improve the understanding of dynamic patterns of structural communication in the L1 MBL. As there are four subunits in the tetramer, 6 plots were generated in total. For example paths from subunit A to B, A to C, A to D, B to C, B to D, C to D). The network analysis results will be illustrated first by a general overview to show the paths in the complete tetramer and then following by the detailed network paths between adjacent subunits. A diagrammatic network result is also illustrated for a clear explanation.

## A-B paths

In subunits A-B network analysis result, A subunit is coloured red and B subunit is coloured yellow, other subunits remain white. The paths between two subunits are shown in Figure 32. The network is also shown in a diagrammatic way in Figure 33 for a clear explanation. In Figure 33, the identical nodes in subunit A and B are highlighted (A117, H118, L145, Y164, P165 and P166). Among these conserved nodes (same nodes in both subunit A and B), P166 is essential for signal transmission between these two subunits, all signals have to pass via P166 to reach the nodes in the other subunit.



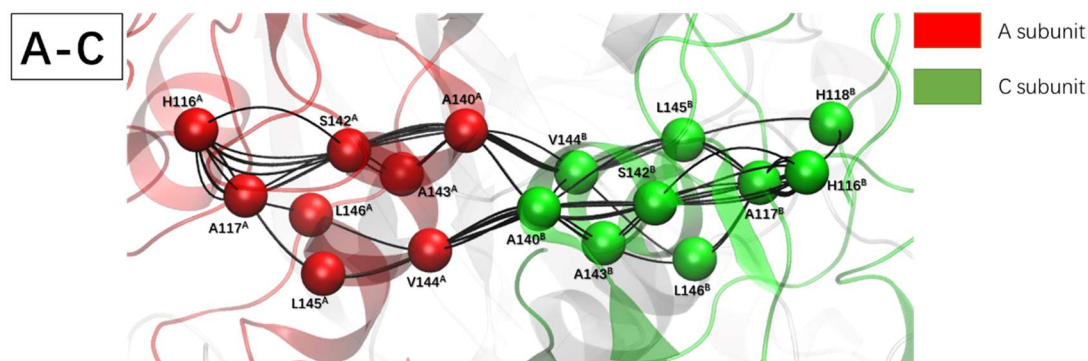
**Figure 32:** The detailed view of paths from subunit A binding site to subunit B binding site.



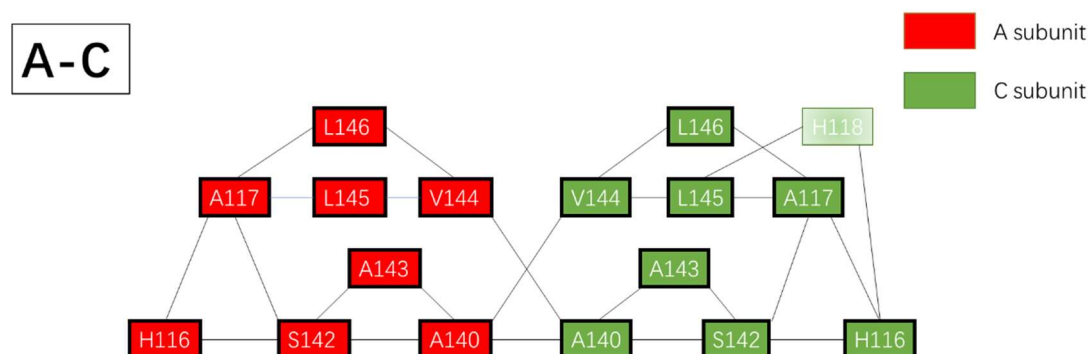
**Figure 33:** Diagrammatic representation of the paths from subunit A binding site to subunit B binding site. Identical nodes between two subunits are highlighted in a black box. The non-identical nodes between two subunits are coloured lighter.

## A-C paths

In A-C network analysis, A subunit is coloured red and C subunit is coloured green, other subunits remain white. The paths between two subunits are shown in Figure 34. The network is also shown as a diagrammatic representation in Figure 35 for a clear explanation. The conserved nodes between subunit A and C are A117, A140, S142, A143, V144, L145 and L146. Among these conserved nodes, the signal transmission between two subunits passes via A140 or V144 to reach the other subunit. A140 is more important in the transmission. All signals pass from one subunit via residue A140 or V144 but will enter the adjacent subunit via A140 as the first node in the other subunit.



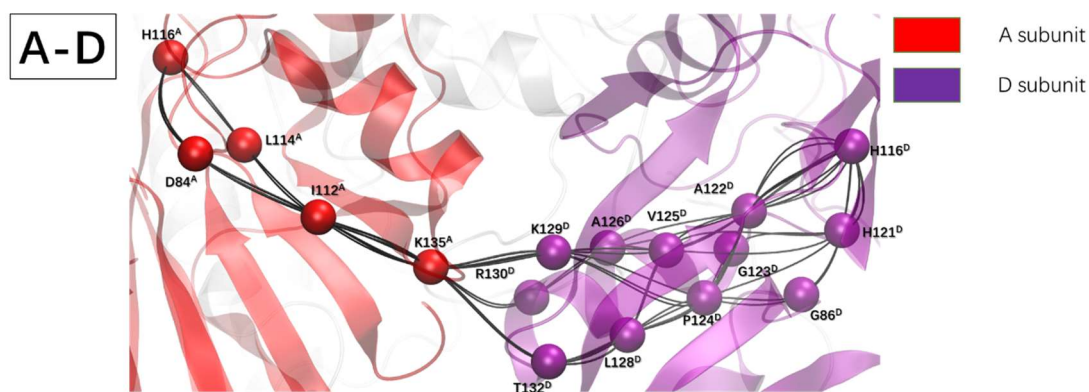
**Figure 34:** The detailed view of paths from subunit A binding site to subunit C binding site.



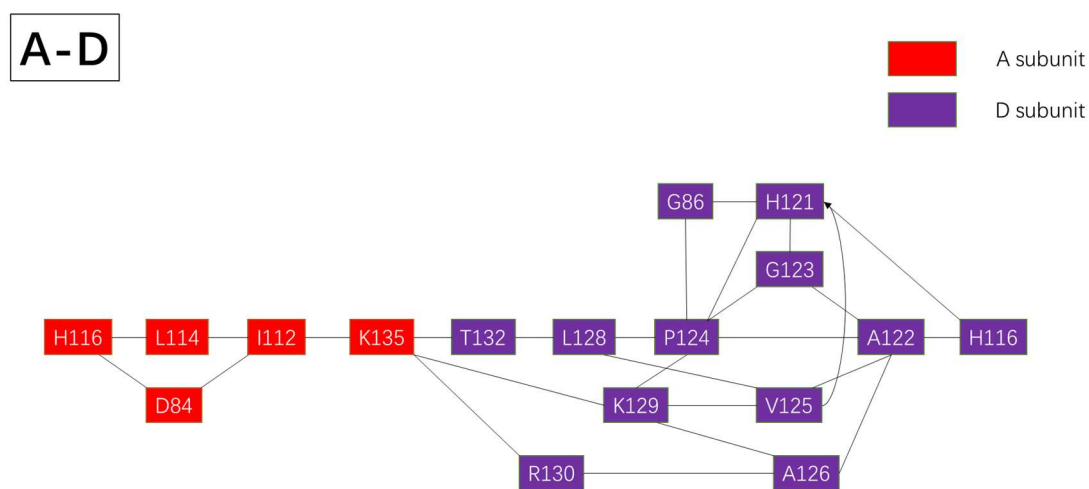
**Figure 35:** Diagrammatic presentation of the paths from subunit A binding site to subunit C binding site. Identical nodes between two subunits are highlighted in a black box. The non-identical nodes between two subunits are coloured lighter.

## A-D paths

In the subunit A-D network analysis, A subunit is coloured red and D subunit is coloured purple, other subunits remain white. The paths between two subunits are shown in Figure 36. The network is also shown as a diagram in Figure 37 for a clear explanation. There is no conserved node between subunit A and D. All the signal transmission from one subunit to the other needs to go via residue K135 from subunit A.



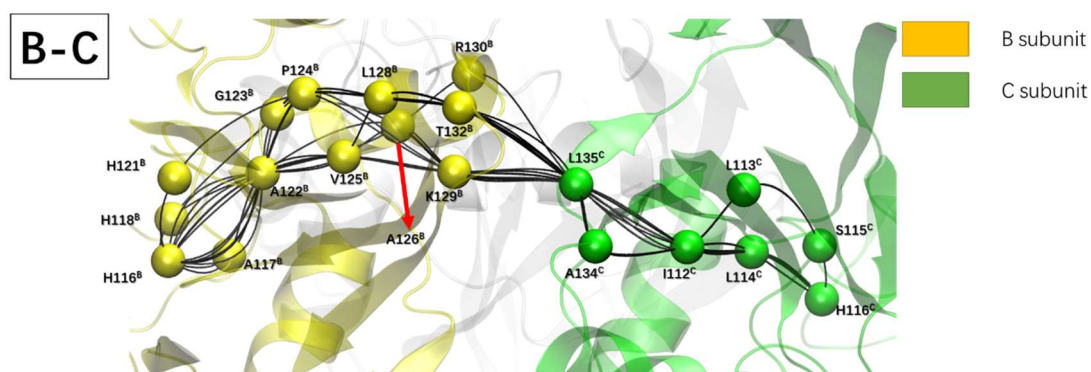
**Figure 36:** The detailed view of paths from subunit A binding site to subunit D binding site.



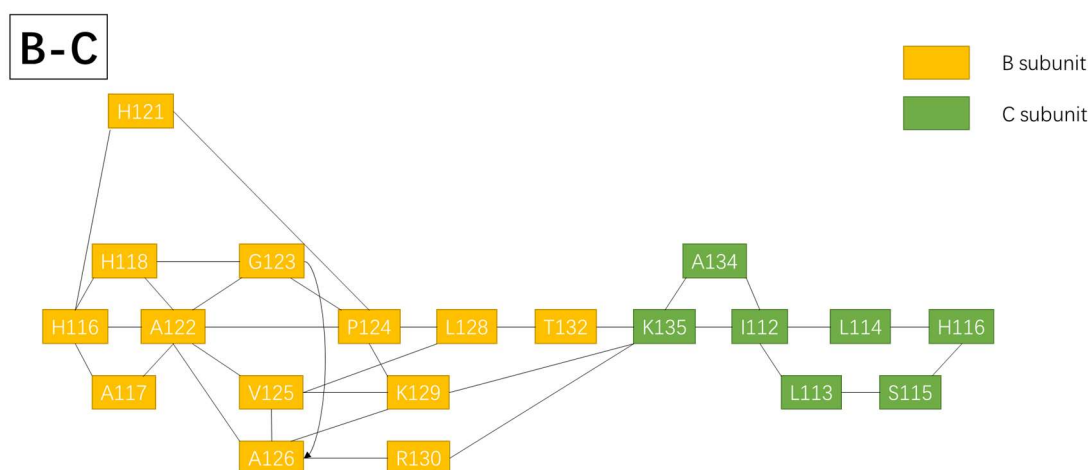
**Figure 37:** Diagrammatic representation of the paths from subunit A binding site to subunit D binding site. There is no identical node existing between these two subunits.

## B-C paths

In subunit B-C network analysis, B subunit is coloured yellow and C subunit is coloured green, other subunits remain white. The paths between two subunits are shown in Figure 38. The network is also shown as a diagrammatic representation in Figure 39 for a clear explanation. There is no conserved node between subunit B and C. All the signal transmission from one subunit to the other is transmitted via K135 from subunit C.



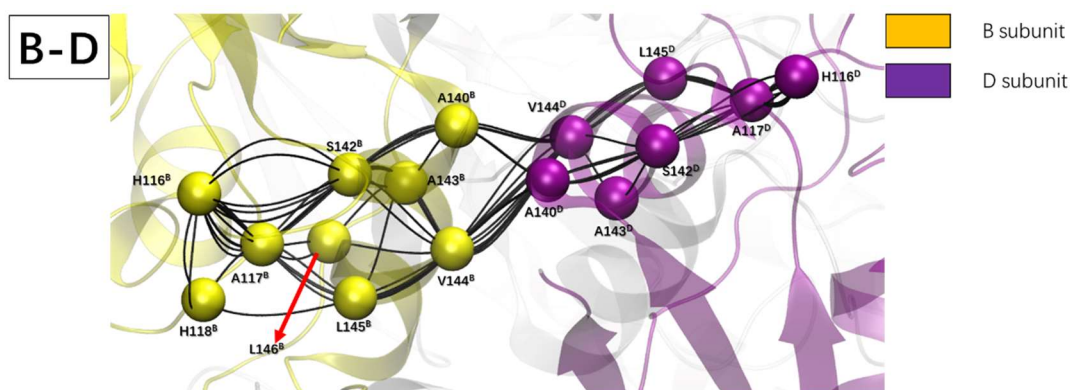
**Figure 38:** The detailed view of paths from subunit B binding site to subunit C binding site.



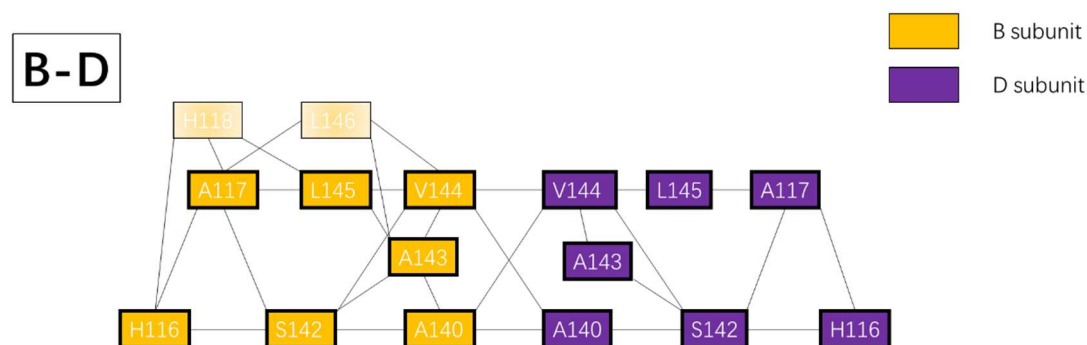
**Figure 39:** Diagrammatic representation of the paths from subunit B binding site to subunit C binding site. There are no identical nodes between these two subunits.

## B-D paths

In the B-D network analysis, B subunit is coloured yellow and D subunit is coloured purple, other subunits remain white. The paths between the two subunits are shown in Figure 40. The network is also shown as a diagram in Figure 41. The conserved nodes between subunit B and D are highlighted (A117, A140, S142, A143, V144 and L145) in a black box. The nodes and paths are almost identical compared with the network result observed in subunit A-C. In the B-D network analysis, A140 and V144 has a direct path with A140 and V144 in the other subunit. Here the signal has two routes to transfer from one subunit to another.



**Figure 40:** The detailed view of paths from subunit B binding site to subunit D binding site.

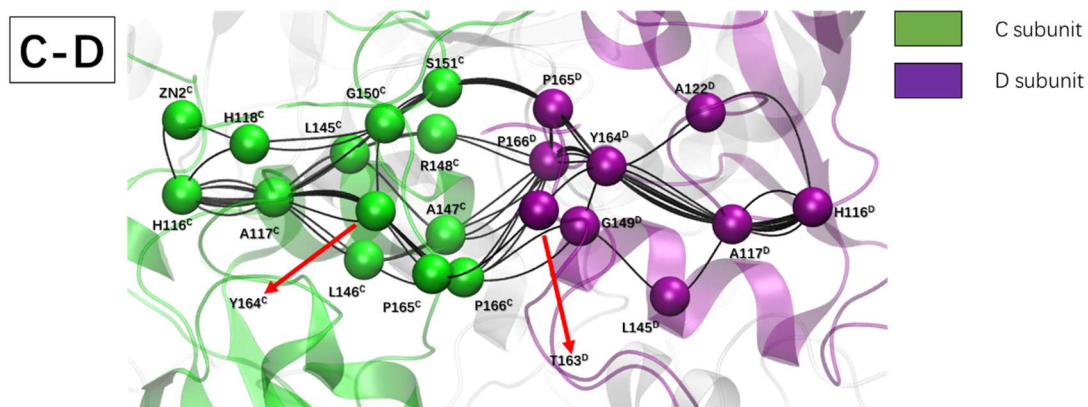


**Figure 41:** Diagram representation of the paths from subunit B binding site to subunit D binding site. Identical nodes between two subunits are highlighted in a black box. The non-identical nodes between two subunits are coloured lighter.

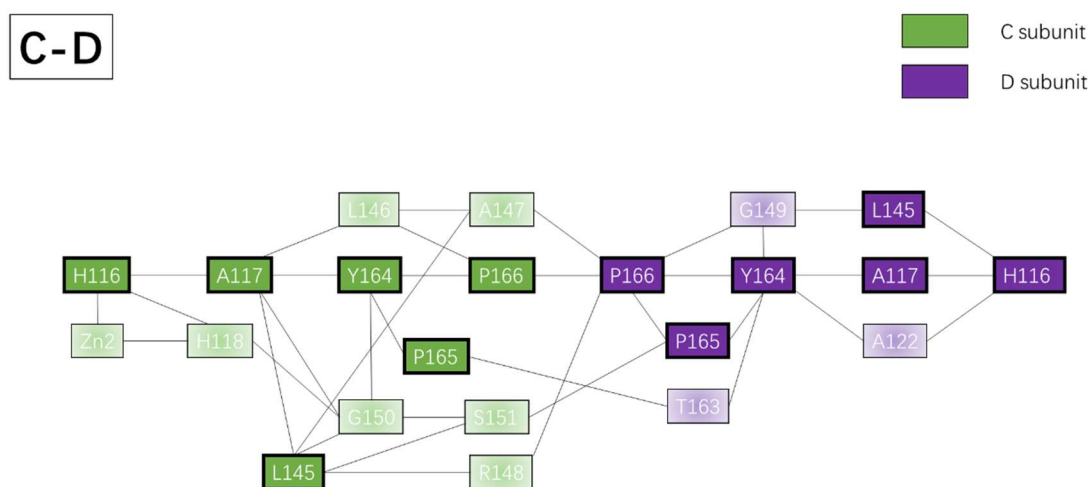


## C-D paths

In the C-D network analysis, C subunit is coloured green and D subunit is coloured purple, other subunits remain white. The paths between the two subunits are shown in Figure 42. The network is also shown as a diagram in Figure 43. The conserved nodes are highlighted (A117, L145, Y164, P165 and P166). The nodes and paths are similar to the network result of A-B subunits. P166 is still the essential node for signal transmission from one subunit to another.



**Figure 42:** The detailed view of paths from subunit C binding site to subunit D binding site.



**Figure 43:** Diagram representation of the paths from subunit C binding site to subunit D binding site. Identical nodes between two subunits are highlighted in a black box. The non-identical nodes between two subunits are coloured lighter.

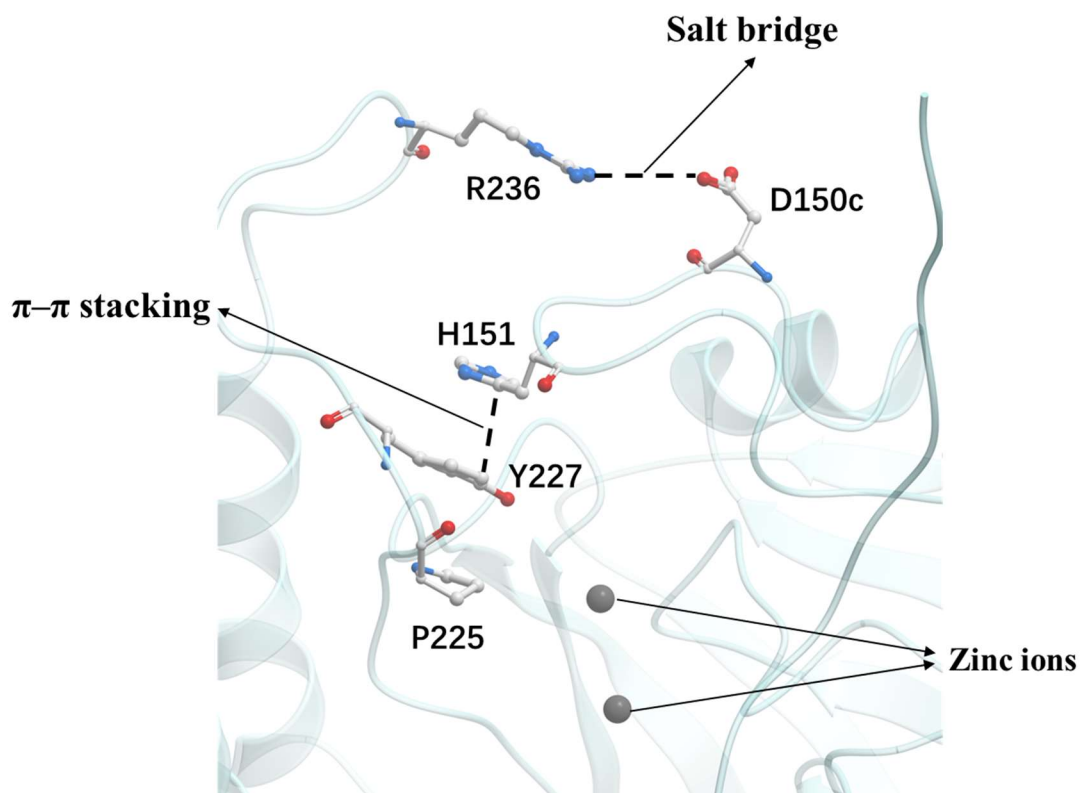
## 5. L1 $\beta$ -lactamase Discussion

L1  $\beta$ -lactamase have already been studied on two main directions: the zinc binding site mechanism and the potential L1  $\beta$ -lactamase inhibitors. For those who focus on the double zinc binding site, it has been proved that even with only one zinc L1  $\beta$ -lactamase can also carry out the catalytic activity with less efficiency.<sup>198,199</sup> Some of them focus on finding the inhibitors of L1  $\beta$ -lactamase. There are several compounds were found to be the potential inhibitor for L1  $\beta$ -lactamase, such as withanolide R and amino acid thioesters.<sup>200,201</sup> However, there is no inhibitor has been approved for clinical use.<sup>202</sup> In this work, L1  $\beta$ -lactamase analysis was carried out with molecular dynamic simulations to find out the conformational change of gating loops and the correlations in the tetramer. As a result, “open” and “closed” conformations were revealed and a dimer of dimer subunits correlation in the L1  $\beta$ -lactamase tetramer was also indicated.<sup>203</sup>

### “Open” and “closed” conformations

There were 7 states generated from the MSM analysis. Based on these states, three different conformations were identified that can describe the dynamics of the two gating loops: open, intermediate and closed. The conformations were defined by a combination of three structural features: (1) the salt bridge between residue D150c and residue R236; (2) the  $\pi$ - $\pi$  stacking between residue H151 and residue Y227; and (3) the conformation adopted by P225 (Figure 44).



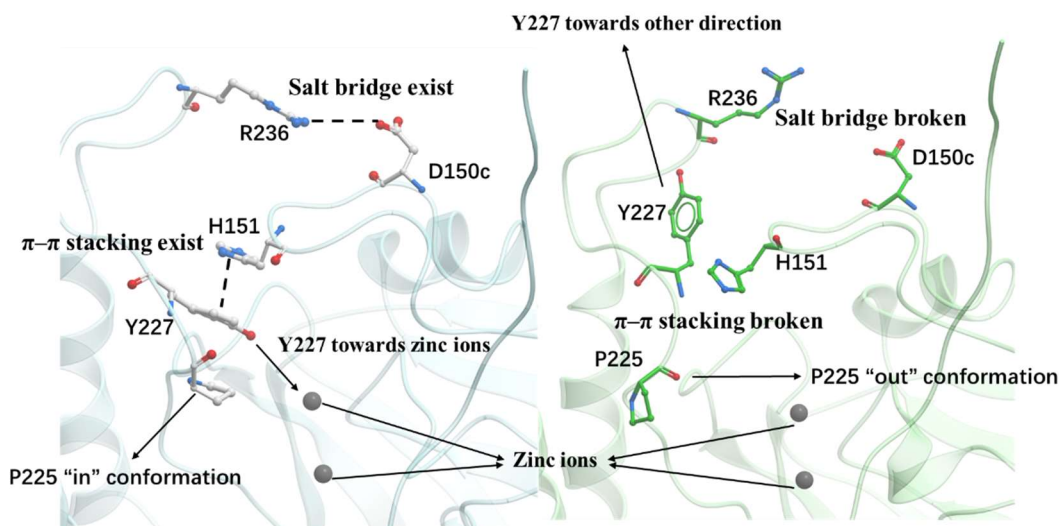


**Figure 44:** The position of the salt bridge, the  $\pi$ - $\pi$  stacking, P225 and zinc ions. D150c and H151 are on the  $\alpha$ 3- $\beta$ 7 loop, R236, P225 and Y227 are on the  $\beta$ 12- $\alpha$ 5 loop. The salt bridge acts like a gate to control the opening and closure of the zinc binding site.

The salt bridge between residue D150c and residue R236 is important for the L1 MBL catalytic activity.<sup>94</sup> Residue D150c is positioned on  $\alpha$ 3- $\beta$ 7 loop and R236 is present in the  $\beta$ 12- $\alpha$ 5 loop. The salt bridge from these two elongated loops act as a gate for the opening and closure of the zinc binding site.<sup>204</sup> In these two loops,  $\beta$ 12- $\alpha$ 5 loop displayed high flexibility during the dynamics. The  $\alpha$ 3- $\beta$ 7 loop was relative stable, due to the direct proximity to the zinc binding site. The formation and the loss of this salt bridge interaction regulated the “open” and “close” conformation adopted by the  $\alpha$ 3- $\beta$ 7 and  $\beta$ 12- $\alpha$ 5 loops. The gating loop interactions are driven by the conformations adopted by the side chain of R236 from the  $\beta$ 12- $\alpha$ 5 loop., The varied conformations from the

side chain of R236 can be observed in the different metastable states. On the other hand, the side chain conformations of D150c from the  $\alpha$ 3- $\beta$ 7 loop remains relatively stable across all metastable states. In a previously reported work on TEM-1  $\beta$ -lactamase (class A  $\beta$ -lactamase), the formation and loss of an important salt bridge interaction between key loops would change the orientation of the  $\Omega$  loop and thus affect antibiotic resistance.<sup>205</sup> This research provides evidence to infer that the gating loop salt bridge interactions could be important for the stability of zinc binding site and antibiotic resistance.<sup>205</sup>

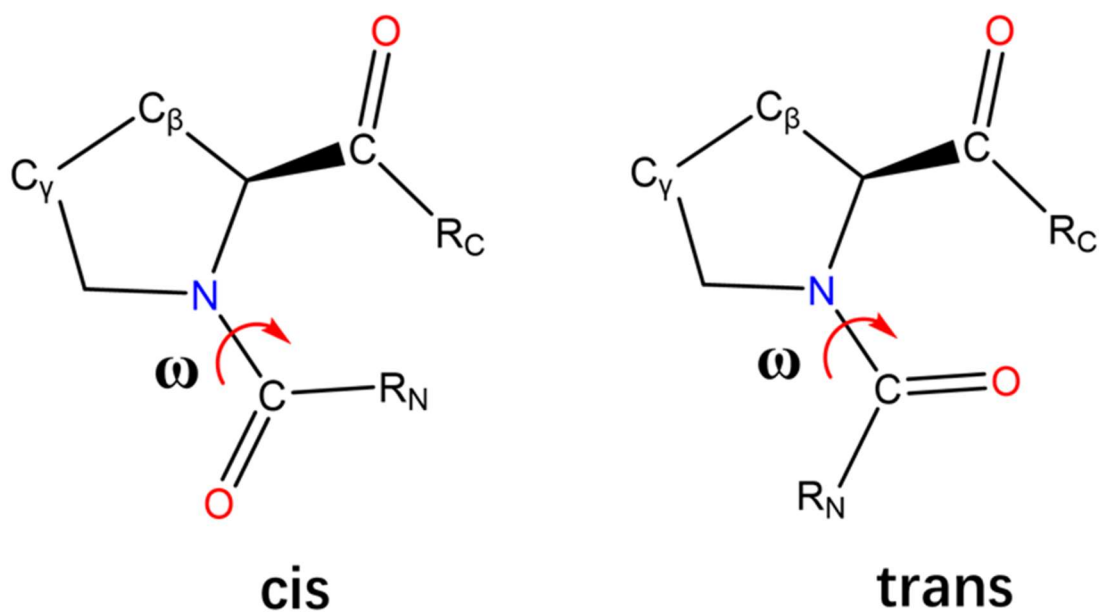
The  $\pi$ - $\pi$  stacking between residue H151 and residue Y227 is important because it can increase the stability of zinc binding site topology and help the recognition of antibiotics.<sup>206,207</sup> In L1 MBL, the  $\pi$ - $\pi$  stacking remained stable in the “closed” state. When the salt bridge between two gating loops was broken, the volume round the zinc binding site increases significantly. This allows the side chain of Y227 to change orientation from pointing towards zinc atoms to residue R236 (Figure 45). The rotation changes the  $\pi$ - $\pi$  stacking conformation and leads to the loss of the interaction. Therefore, the formation and loss of the  $\pi$ - $\pi$  stacking between residue H151 and residue Y227 is directly influenced by the salt bridge between the gating loops.



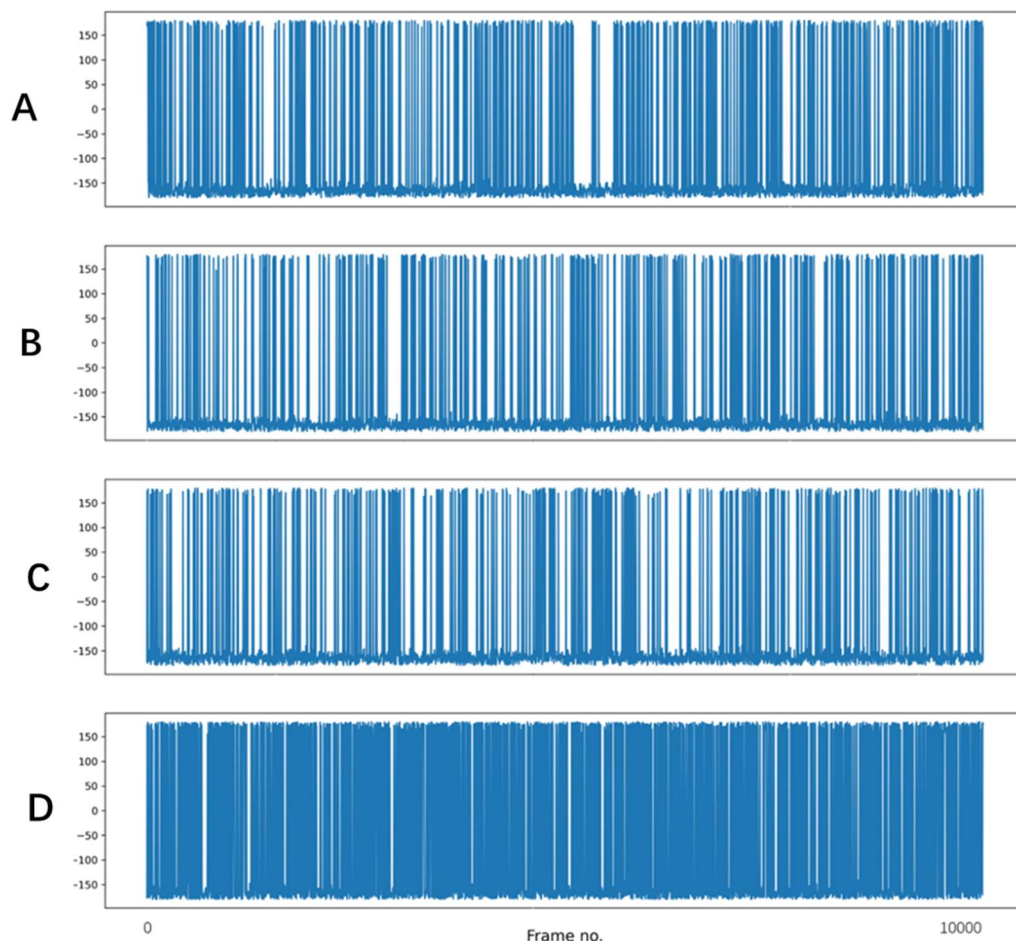
**Figure 45:** Different status of salt bridge,  $\pi$ - $\pi$  stacking and proline.

The conformation of residue P225 is also a unique structural feature in the L1  $\beta$ -lactamase. Proline can exist in two configurations: cis and trans (Figure 46). These two configurations are defined by the  $\omega$  angle of proline (cis configuration  $\omega$  angle close to 0 degree, trans configuration  $\omega$  angle close to 180 degree).<sup>208</sup> The side chain of P225 in L1 exists in the trans configuration. This is in contrast to the cis configuration that proline residues adopt among most other  $\beta$ -lactamases.<sup>208</sup> Published work on TEM-1  $\beta$ -lactamase (class A  $\beta$ -lactamase) suggested that when the proline exists in a cis configuration, it can help in the correct orientation of the binding site and protein folding.<sup>209,210</sup> The conformational change of proline around binding site would result in the structural misalignment of the binding site leading to the dysfunction of the protein.<sup>210</sup> The  $\omega$  angle of all P225 were analysed to confirm the trans configuration (Figure 47). The  $\omega$  angle stays at -180 degrees while often transiting to 180 degrees as well. This verifies the trans configuration of P225 in L1  $\beta$ -lactamase. When the C $\gamma$  atom of P225 is oriented towards the zinc binding site (Figure 45), it is defined as the “in”

conformation. The distance between C $\gamma$  atom of P225 and centroid of the two zinc atoms is calculated at 6.5 Å in the crystal structure.<sup>59</sup> All other orientations of P225 C $\gamma$  atom are defined as the “out” conformations (Figure 45). The difference between “in” and “out” conformation in different states was found to be regulated by coupled coordination of the gating loops salt bridge (D150c-R236) and  $\pi$ - $\pi$  stacking (H151-Y227). These two interactions could be the main reason why P225 stays in the trans configuration in L1  $\beta$ -lactamase. Besides, the “in” and “out” conformation could also affect substrate binding.

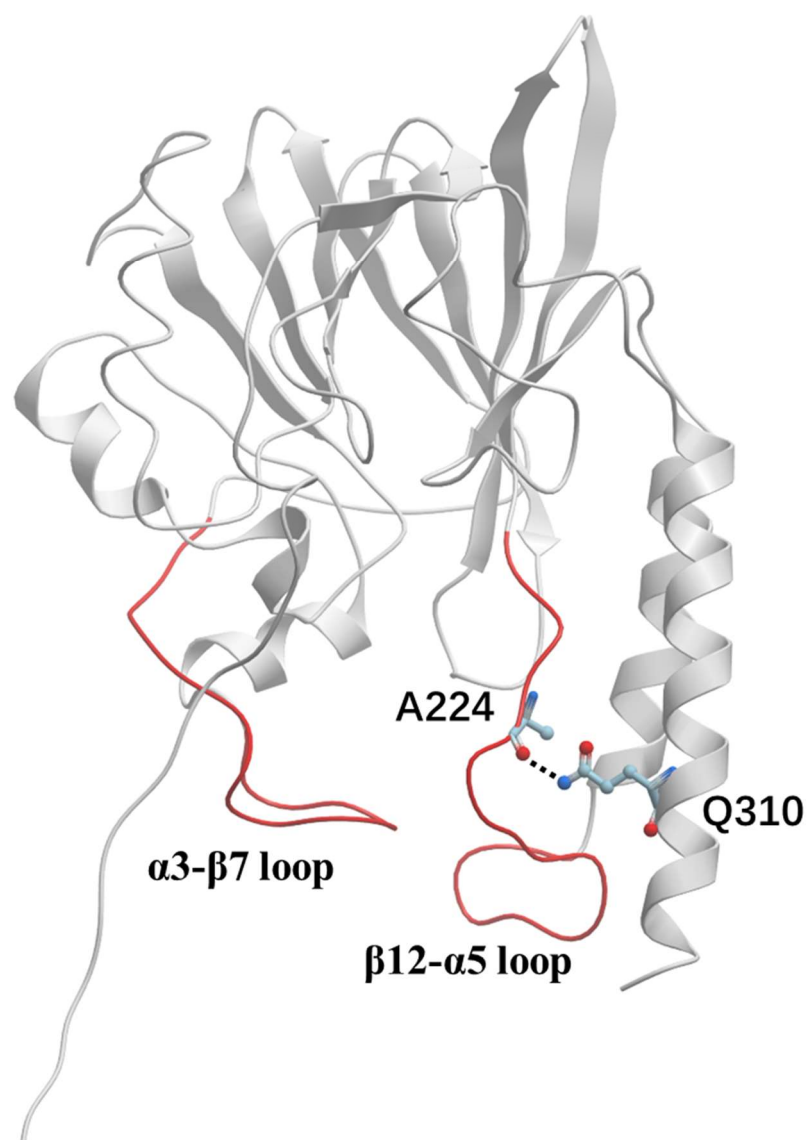


**Figure 46:** cis and trans configuration of proline. The  $\omega$  angle is highlighted.



**Figure 47:** The  $\omega$  angle analysis of P225 in all four subunits. The stride was set at 100 frames. The proline cis configuration  $\omega$  angle is close to 0 degree and the trans configuration  $\omega$  angle close to 180 degree. In L1 MBL, the  $\omega$  angle of P225 transition from -180 degree to 180 degree in all four subunits. Subunit D displayed highest flexibility in the tetramer.

To explain the conformational dynamics more clearly, the “closed” conformation of L1  $\beta$ -lactamase is discussed first. The “closed” conformation is defined by a combination of three structural conformations: the salt bridge formed between residue D150c and residue R236, the  $\pi$ - $\pi$  stacking formed between residue H151 and residue Y227, P225 present in the “in” conformation. Besides, there was an extra interaction present between the main chain of A224 ( $\beta$ 12- $\alpha$ 5) and side chain of Q310 ( $\alpha$ 6 helix). This interaction helps in the stabilization of the  $\beta$ 12- $\alpha$ 5 loop (Figure 48) and further stabilized both gating loops indirectly. State 3, 5, 6 and 7 represents the “closed” states. State 7 is the most populated state and represented over 80% of all the structures. State 7 is also the lowest energy state which mostly defined the “closed” conformation of L1  $\beta$ -lactamase. State 3, 5 and 6 were the sub-states of the “closed” conformation. There were minor differences found in gating loops and the distal end of the active site. Therefore, the “closed” conformation was the dominant conformation in the L1  $\beta$ -lactamase.



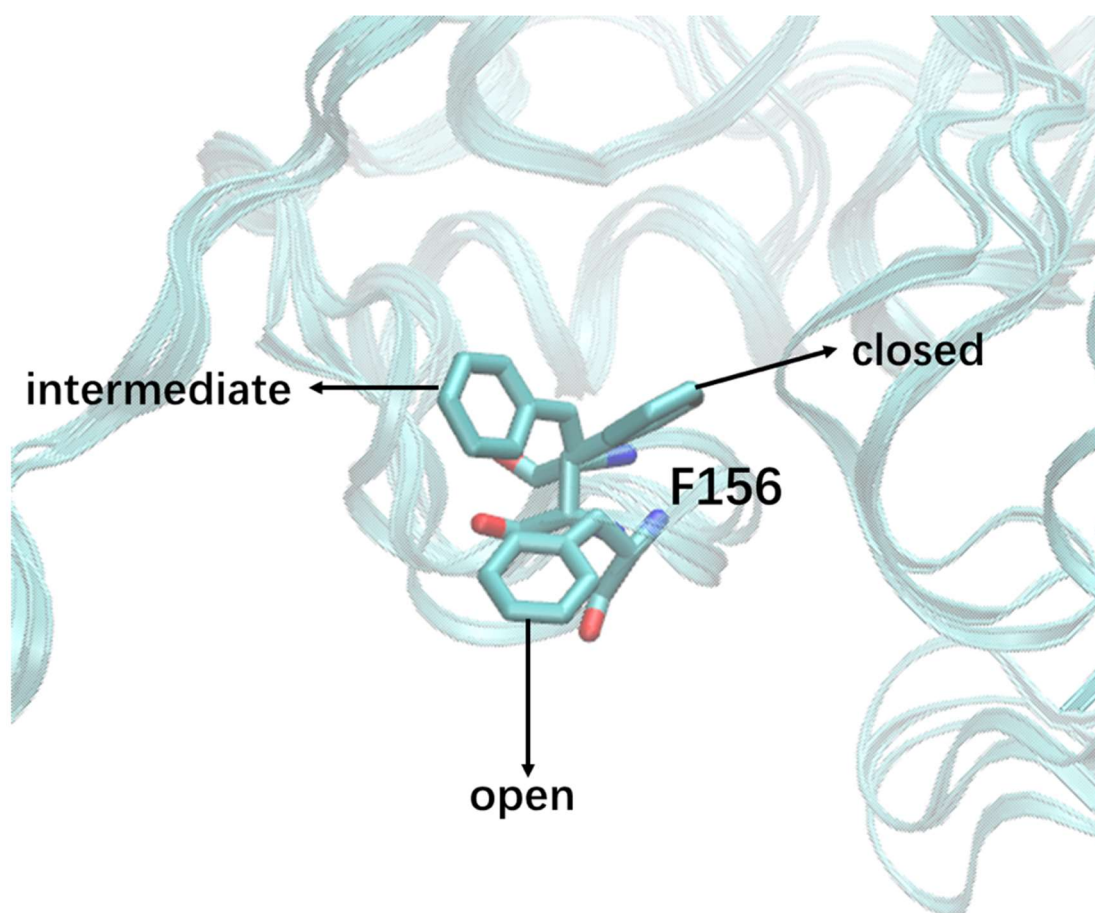
**Figure 48:** The interaction between A224 and Q310. Two elongated loops were highlighted with red colour. Q310 and A224 interaction help stabilized the  $\beta 12$ - $\alpha 5$  loop.

States 2 and 4 were the intermediate states. In state 2, the  $\pi$ - $\pi$  stacking between residue H151 and residue Y227 was still formed, and P225 was in the “in” conformation. However, the salt bridge between residue D150c and R236 was not formed. In state 4, two out of the three structural features were in the “open” conformation (salt bridge between residue D150c and R236 was not formed, no  $\pi$ - $\pi$  stacking between residue H151 and residue Y227, P225 present in the “out” conformation). Besides, the interaction between the main chain of A224 ( $\beta$ 12- $\alpha$ 5) and side chain of Q310 ( $\alpha$ 6 helix) was also lost, resulting in the  $\alpha$ 6 helix to move away from the zinc binding site.

State 1 represents the “open” state. In contrast to the “closed” state, the “open” state lost all three structural features, namely the salt bridge between residue D150c and R236 was not formed, there was no  $\pi$ - $\pi$  stacking between residue H151 and residue Y227, and P225 was present in the “out” conformation (Figure 45). As a result of the loss of all these interactions the  $\beta$ 12- $\alpha$ 5 loop rotates away from zinc binding site. This results in a much larger pocket being formed around the zinc atoms, which would allow the substrates access the substrate binding site. Compare with “closed” state, a larger pocket allows bigger inhibitor ligand and a higher probability of inhibition. Therefore, “open” state can let drug designer achieve maximum probability of inhibition.

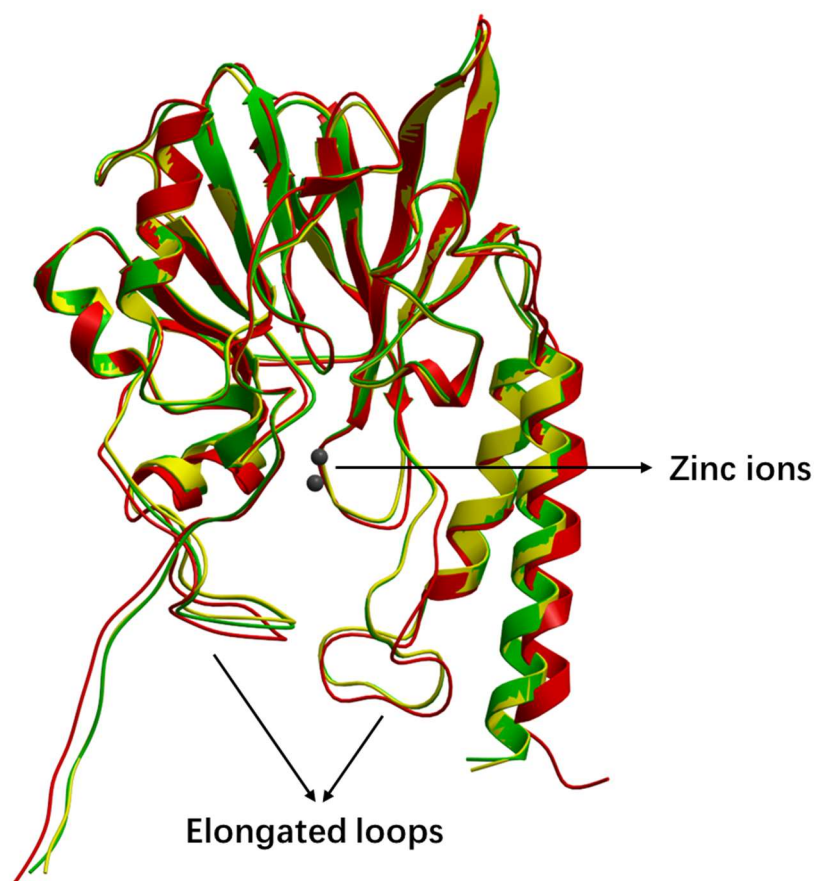


Besides, the side chain of residue F156 was also observed to have conformational differences in different states (Figure 49). The  $\chi_1$  dihedral angle of F156 also displayed large rotation in different states. Three snapshots were extracted from the “open”, “intermediate” and “closed” states and superimposed to give a more direct impression of the rotation of the side chain (Figure 49). It could be easily identified that the side chain of F156 has rotated as it makes transitions between different states. The side chain of F156 has also been shown experimentally to play an important role in L1 MBL substrate binding.<sup>100,211</sup> Therefore, a direct link between F156 could also be established with the dynamic system of the two elongated loops.



**Figure 49:** The dynamics of the F156 side chain. Three conformations were extracted from the “open”, “intermediate” and “close” states and superimposed.

The other L1  $\beta$ -lactamase crystal structures with ligand/substrate bound to the protein were also used for comparison with the states identified from the MSM (Figure 50).<sup>59,100,211,212</sup> A state 7 monomer (red), a monomer from apo state crystal structure (yellow, PDB entry 1SML) and a monomer from inhibitor binding complex (green, PDB entry 5DPX) were superimposed to compare the conformation of the elongated loops. All these structures were identified to be in a “closed” conformation. This indicated that the “closed” conformation is extremely stable and the predominant conformations, thus confirming our observations.



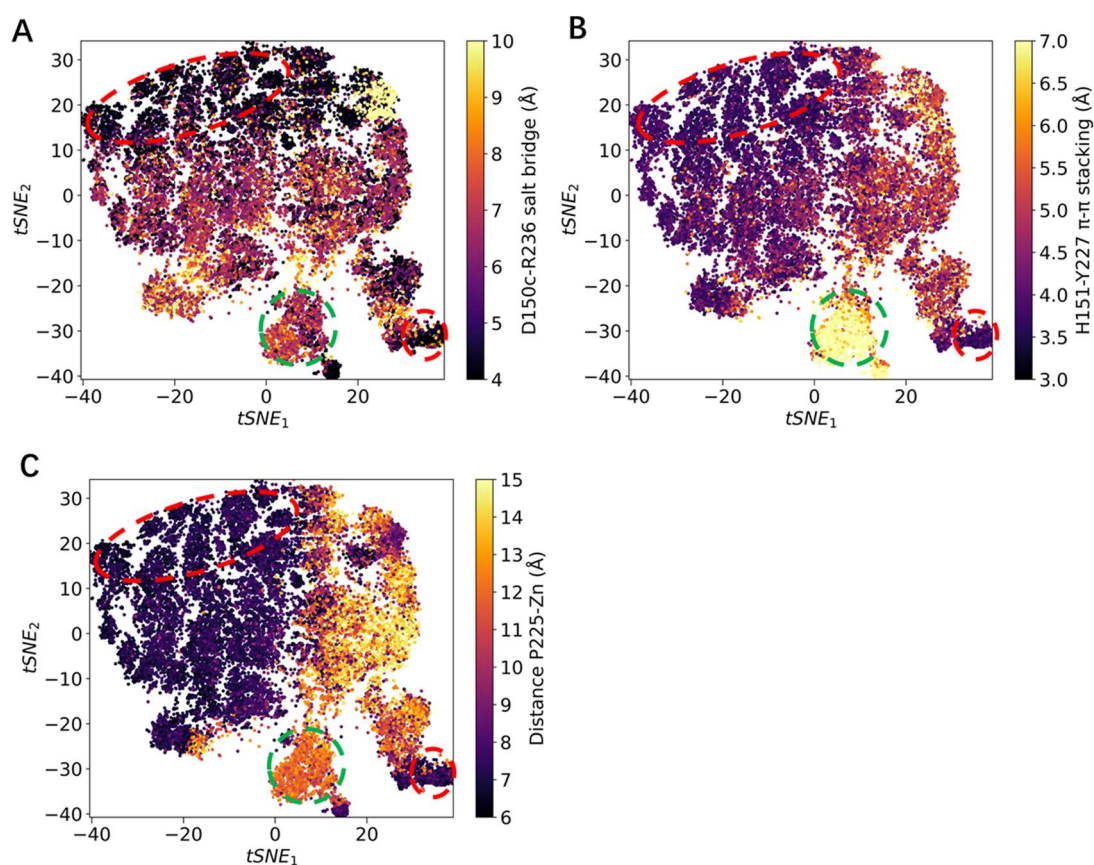
**Figure 50:** Superimposed structures of a state 7 monomer (red), a monomer from the apo state crystal structure (yellow, PDB entry 1SML) and a monomer from the inhibitor binding complex (green, PDB entry 5DPX). Zinc ions were also included and are illustrated as silver spheres.

Based on the analysis of the net flux plot of L1  $\beta$ -lactamase (Figure 26), most structures went from state 1 to state 7 and transit via state 4. This path had the least mean first passage time in the flux pathways (Figure 27), which could explain why the path accounts for 89.7% of all the kinetic structures. Besides, there were also two less populated paths: from state 1  $\rightarrow$  state 7 (9.1%) and states 1  $\rightarrow$  4  $\rightarrow$  2  $\rightarrow$  7 (1.2%). However, these two paths took longer mean first passage times. The longer mean first passage times indicates a higher free energy barrier. Therefore, the conformational changes between state 1, 4 and 7 could explain the dynamics of the L1  $\beta$ -lactamase best. In each of these states, all the important conformational changes around zinc binding site could be related to the two elongated gating loops.

To get a more detailed and a deeper understanding on the coordination of the three structural features (D150c-R236 salt bridge, H151-Y227  $\pi$ - $\pi$  stacking and the in/out conformation of P225) CVAE-based deep learning was employed.<sup>203</sup> The interaction between D150c-R236 salt bridge, the distance between H151-Y227  $\pi$ - $\pi$  stacking, and the distance between centroid of two zinc ions and P225 were used as CVAE calculation features.

The deep learning results are illustrated in Figure 51. The “closed” state conformations are highlighted in a red circle while the “open” state conformations are highlighted in a green circle. The “open” and “closed” conformations required all three structural features to be present or lost simultaneously (all open or all closed). Only three clustered

regions meet the requirement and have been highlighted in Figure 51. All other structures outside of these clustered regions were intermediate conformations. As detailed above, conformations from all 7 metastable states were clustered on a general conformation distribution plot (Figure 30). Subunit D in state 1 described the “open” conformation. The clustered structures are present on the right bottom of Figure 31-1, and are same as that encircled in green in Figure 51. Subunit B of state 2 (Figure 31-2) described the intermediate “closed” conformation (where two of the three features exist). Subunit A of state 4 (Figure 31-4) described the intermediate “open” conformation (where two of the three feature are lost). Subunit C of state 7 (Figure 31-7) was selected to describe the “closed” state. The results of CVAE-based deep learning verified the coordination of the structural features in the three conformations.



**Figure 51:** CVAE-based deep learning results for clustering of D150c-R236 salt bridge, H151-Y227  $\pi$ - $\pi$  stacking and the in/out conformation of P225. The tsne based clustering for (A) D150c-R236 salt bridge, (B) H151-Y227  $\pi$ - $\pi$  stacking and (C) in/out conformation of P225. The “open” conformations clustered region is highlighted in green circle while the “closed” conformation clustered region is highlighted in a red circle. All other regions represent the intermediate state structures.

In general, the conformational changes of D150c-R236 salt bridge, H151-Y227  $\pi$ - $\pi$  stacking, the in/out conformation of P225 and the F156 side chain rotation could explain the dynamics of the two elongated loops,  $\alpha$ 3- $\beta$ 7 loop and  $\beta$ 12- $\alpha$ 5 loop. The  $\beta$ 12- $\alpha$ 5 loop has higher flexibility compared with the  $\alpha$ 3- $\beta$ 7 loop. The  $\beta$ 12- $\alpha$ 5 loop acts similar to a lid for the zinc binding site. In the “open” state, D150c-R236 salt bridge broken, H151-Y227  $\pi$ - $\pi$  stacking lose, P225 adopt to “out” conformation, these changes allow the  $\beta$ 12- $\alpha$ 5 loop to move away from the active site. The “gate” when

open, increases the volume of the active site permitting the substrates to access the zinc binding site. The “open” state, however, is a high energy state. As a result the  $\beta 12-\alpha 5$  loop is drawn back to the “closed” state, which is lower energy and much more stable. This “closed” state is what is observed in all L1 MBL crystal structures including the conformation with inhibitors.<sup>59,211</sup> These changes in the conformational dynamics will have an obvious impact to the topology of the zinc binding site and influence the substrate binding and catalytic function in L1 MBL.

## **Dimer-of-dimers**

The CVAE-based deep learning results highlighted that the subunit A and subunit C had similar dynamic patterns while subunit B and subunit D had similar dynamic patterns (Figure 30). In state 2, 3, 4, 5, 6 and 7, the similar dynamics of subunit A and subunit C could be clearly observed (Figure 31). The similar dynamics of subunit B and subunit D were mostly observed in state 5 and state 7. However, as state 7 accounts for more than 80% of all the structures in the simulation, subunit B and subunit D dynamics could still be considered as highly similar.

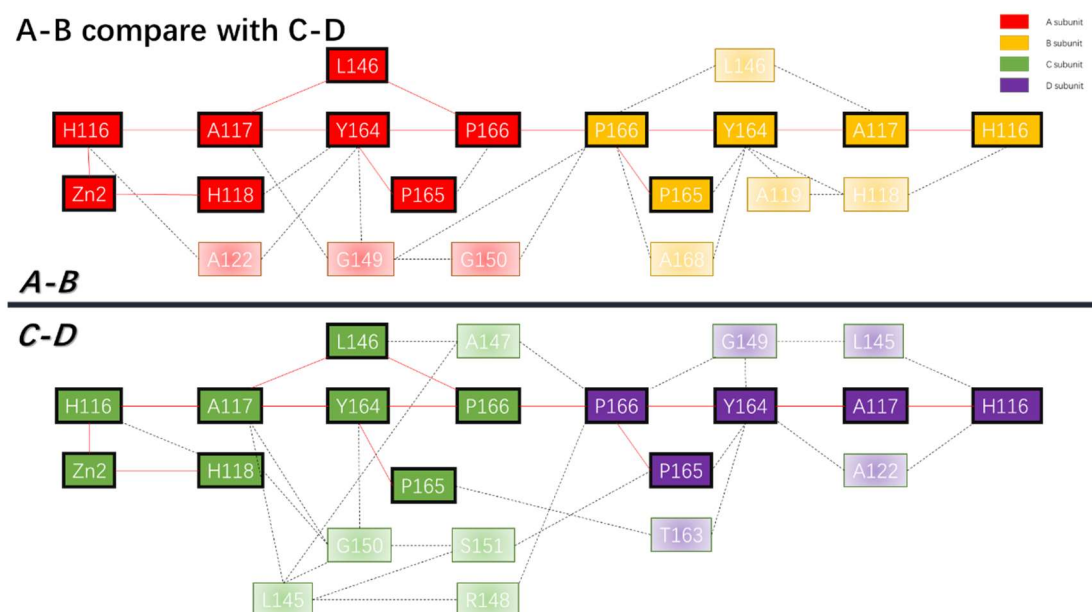
From the network analysis results, the signalling paths are also symmetric. Taken together, these results imply to a dimer-of-dimers dynamic pattern in L1 MBL. If the dimer-of-dimers motion is real, then the nodes along the signal paths in subunit A and C should be identical and subunit B and D nodes should be similar as well. As the

parallel comparisons were done (Figure 32-43), a vertical comparison would give a more straightforward understanding into this.

The network analysis results between each two subunits in the L1 MBL were compared. In the following figures (Figure 52-54), the identical nodes are shown in black boxes while non-identical nodes from different subunits were given different colour. Subunit A was red, subunit B was yellow, subunit C was green and subunit D was purple. The comparison highlighted the presence of identical nodes between the two compared network analysis (such as between A-B subunits compared with C-D subunits). The same paths between identical nodes are shown with red solid line and other paths are showed with black dot line. The comparison tries to find out the important residues and the similarities between the network results.

The network path from subunit A to subunit B was compared with the path from subunit C to subunit D (Figure 52). As mentioned above, residue P166-P166 bridges the signal transmission between subunit A and B (C and D as well). The conserved residues between subunits A-B and C-D network results also illustrate a similar result. All conserved residues signal transfer from one subunit to the other need to via the path of P166-P166. Therefore, P166 could be seen as an important residue which bridges the signal between the A and B (C and D) subunits. Besides, the conserved paths and nodes in subunit A (and C) are highly identical with subunit B (and D). The path H116-A117-Y164-P166-P166-Y164-A117-H116 is highly conserved and shows a symmetric

pattern. This path could be a main route for signal transfer from subunit A (and C) to subunit B (and D).

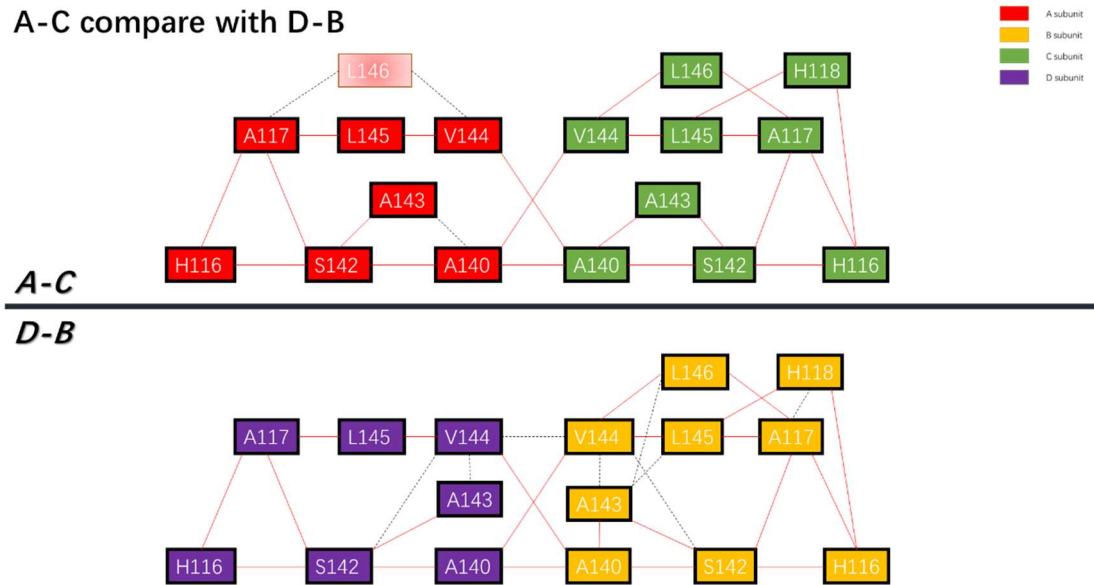


**Figure 52:** A comparison of the subunits A-B paths with subunits C-D paths. Residues in subunit A (red), B (yellow), C (green) and D (purple) are illustrated. The distinct nodes in different subunits that were not on symmetric paths have been illustrated in a lighter colour; red solid lines between nodes that are between symmetric paths (between same nodes in different subunits), while black dot lines highlight alternative paths between nodes in the vertical comparison. The signal paths in subunit A and C, subunit B and D follow similar nodes.

The network path from subunit A to subunit C was compared with the path from subunit B to subunit D (Figure 53). The nodes and paths are highly identical in subunits A-C and subunits B-D network analysis. More than 90% of the nodes are identical. Among these conserved nodes and paths, residue A140 bridges the signal transmission between the two subunits. The signal from one subunit can be transmitted via V144 or A140 but will eventually get into the adjacent subunit via A140. This indicates that A140 is an important residue to bridge the signal transmission between subunit A and C (or



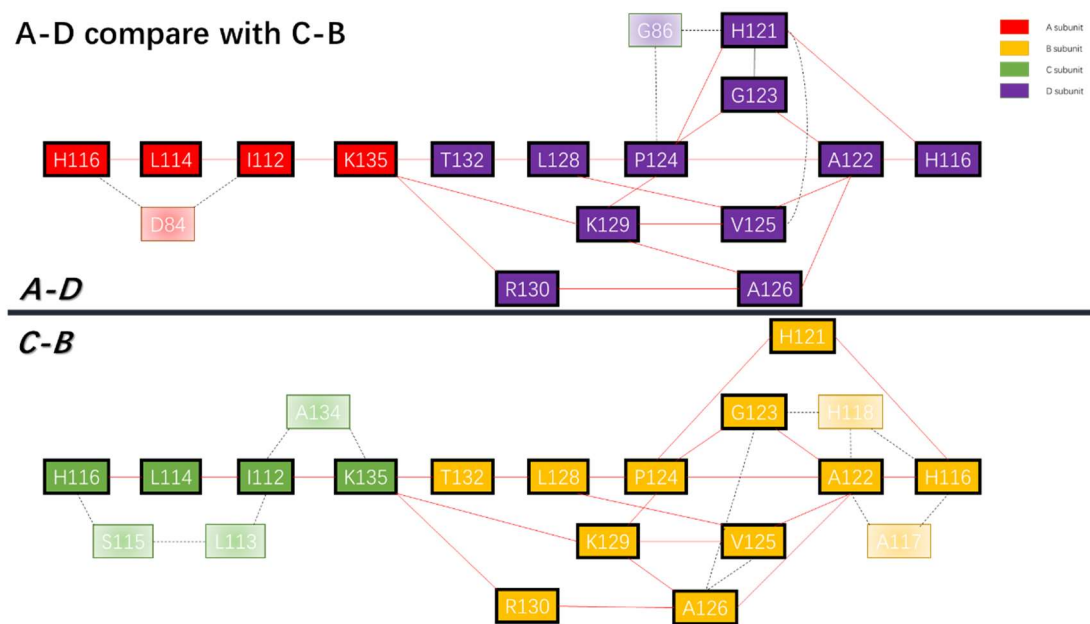
subunits B and D). Besides, the network analysis results in subunit A-C (or subunits B-D) are also highly symmetric.



**Figure 53:** The comparison of subunits A-C paths with subunits B-D paths. Subunit A is red, subunit B is yellow, subunit C is green and subunit D is purple. The distinct nodes in different subunits that were not on symmetric paths have been illustrated in a lighter colour; red solid lines between nodes are between symmetric paths (between same nodes in different subunits), while black dot lines highlight alternative paths between nodes in the vertical comparison. The signal paths in subunit A and D, subunit B and C follow similar nodes.

The network path from subunit A to subunit D was compared with the path from subunit B to subunit C (Figure 54). The nodes and paths are highly identical between these two network analysis results. The K135 (from A and C subunit) bridges the signal transferring between subunit A and D (B and C). The paths in subunit A (C) part are much simpler than D (B) part which makes the paths in subunit A (C) more important. The path H116-L114-I112-K135 in subunit A (C) could be the main route between subunit A (C) and D (B) signal transmission, the missing of each node on this path

could affect the whole signal transferring. Different from the results of A-B (C-D) and A-C (B-D), the nodes and paths are not symmetric (no symmetric residue) in A-D and B-C results.



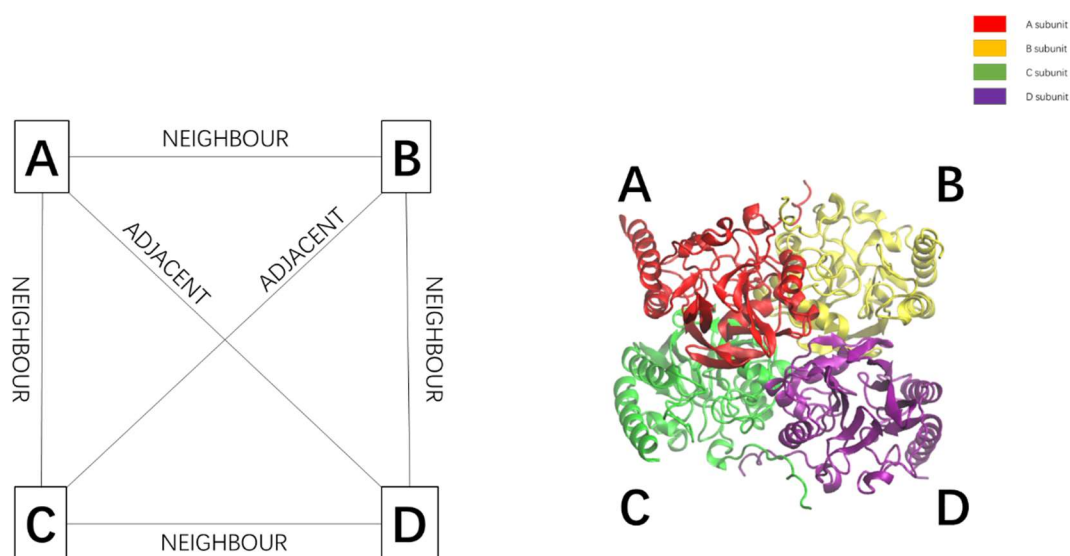
**Figure 54:** The comparison of subunit A-D paths with subunit B-C paths. Residues in subunit A (red), B (yellow), C (green) and D (purple) are illustrated. The distinct nodes in different subunits that were not on symmetric paths have been illustrated in a lighter colour; red solid lines between nodes are between symmetric paths (between same nodes in different subunits), while black dot lines highlight alternative paths between nodes in the vertical comparison. The signal paths in subunit A and C, subunit B and D follow similar nodes.

The results from the network analysis supported the “dimer-of-dimers” dynamics.<sup>213,214</sup>

In the L1 MBL tetramer each subunit has 2 neighbours. For example, subunit A is neighbour with subunit B and C. Similarly, B and C are neighbours to subunit D. (Figure 55). This explained why A-B, B-D, A-C and C-D paths are symmetrical. The symmetry in the tetramer results in the stochastic dynamics but follows structural communication involving similar path nodes. However, the dynamics of adjacent

subunits was different as they did not exhibit any symmetric nodes between them. Taken together, the CVAE-based deep learning and the network analysis results confirm a dimer-of-dimer dynamics for the L1 MBL. The correlations between subunits help in the understanding the dynamics of the L1 MBL tetramer.

The importance of dimer of dimers result is to indicate the potential key residues and correlation pattern of the tetramer. If the signal transferring could be reduced or inhibited, the tetramer could work similarly to four monomers or two dimers. As we mentioned above, L1 MBL has the highest function when it works as a tetramer. The conserve residues in signal transferring and on the protein-protein surface could be important targets for future drug design.



**Figure 55:** The relationship of each subunit in L1 MBL. Subunit A (B) and D (C) are adjacent subunits. A-B, A-C, B-D and C-D are neighbour subunits.

## 6. L2 $\beta$ -lactamase Results

### 6.1 Homology Modelling

In all four L2  $\beta$ -lactamases, only the crystal structure of L2b  $\beta$ -lactamase is available (PDB number 5NE2).<sup>215,216</sup> The sequences of L2a, L2c and L2d were downloaded from the  $\beta$ -Lactamase Database (L2a-Y08562, L2c-AJ251817, L2d-AJ272110).<sup>217</sup> Due to the relative high sequence identity between L2b and L2a, L2c and L2d (69%) shown in Figure 56, homology modelling was used to obtain the atomistic models. The colour of alignment is based on consensus strength, a method for high quality alignment.<sup>218</sup>

```

      .##...AA#...A.DFA.LEKA..GR#GVT#L.T..GRRI.GHRQDERFPMCSTFKS#L#A.
L2a  21 SAPTDAALTAATDFATLEKACAGRLGVTLLDTASGRRI-GHRQDERFPMCSTFKSVLAAT
L2b  21 NAPTDAAITAASDFAALEKACAGRLGVTLLDTASGRRI-GHRQDERFPMCSTFKSMLAAT
L2c  21 SAPTDAALTAATDFAALEKAVRGRFGVTLLDTASGRRI-GHRQDERFPMCSTFKSVLAAT
L2d  21 IPAANAAIAGAADFAALEKASGGRLGVTVLTNGRRI GGHRQDERFPMCSTFKSMLVAH

      VLS.A-...#.LD.RVP#...DLLS#APV.RRH#GKD#TVRDLCR.T#.TSDNTAANLL#
L2a  80 VLSQAERRPALDTRVPVRDADLLSHAPVTRRHAGKDMTVRDLCRATIIITSDNTAANLLF
L2b  80 VLSQAERMPALDRRVPVGEADLLSHAPVTRRHAGKDMTVRDLCRATIIITSDNTAANLLF
L2c  80 VLSQAERQPALDTRVPVRDADLLSHAPVTRRHAGKDMTVRDLCRATIIITSDNTAANLLF
L2d  81 VLSLADAGRVSLDTRVPIAGKDLLSYAPVARRHV GKDLTVRDLCRGTLTTSNTAANLLL

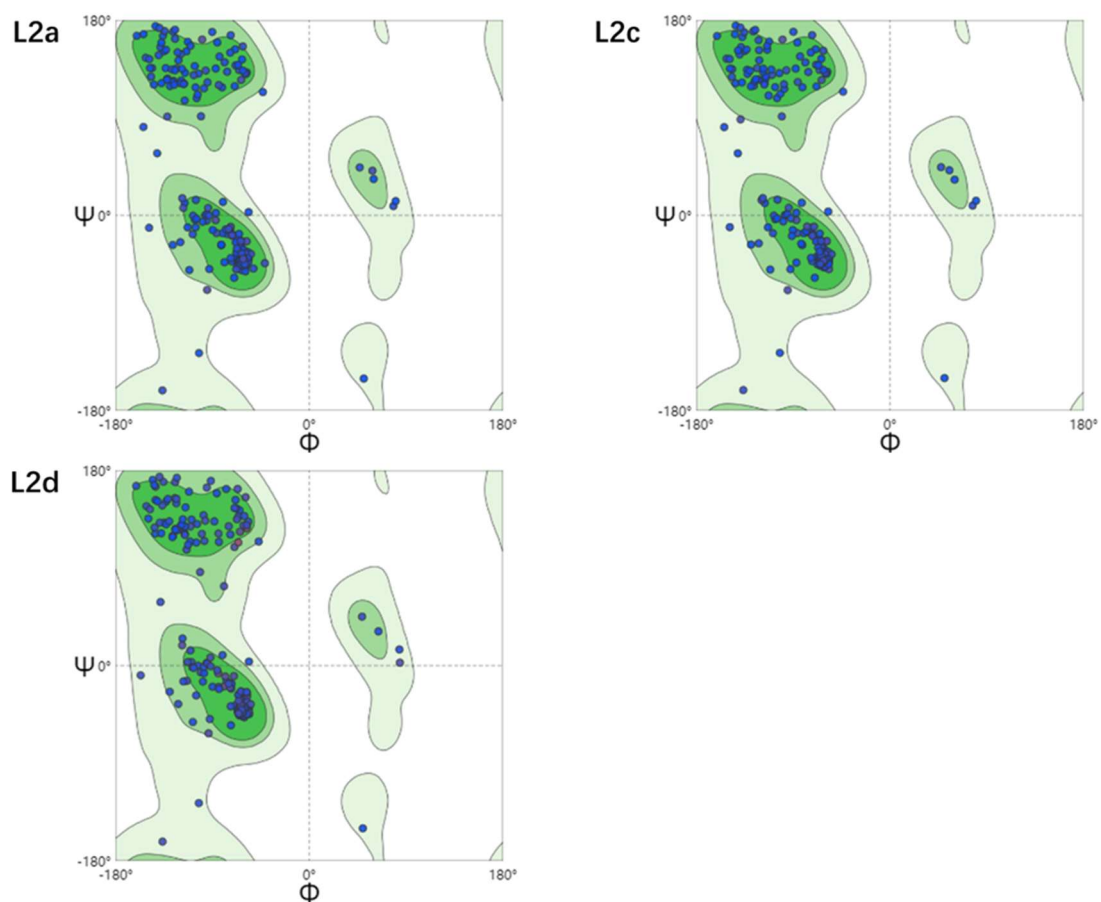
      .VVGGP.A#TAFLR..GD.#.R.DR.EP-#N.FAKGDPDRTT.P.AMA..L.R##G..L
L2a  140 GVVGGFPAVTAF LRTSGDAVSRSDRLEPELNSFAKGDPRDRTTTFGAMAATLQRVVLGEVL
L2b  140 GVVGGFPAVTAF LRSAGD TVSRSDRLEPELNSFAKGDPRDRTTTF AAMAATLQRVVLGEVL
L2c  140 GVVGGFPAVTAF LRSIGDAVSR TDRLEPELNSFAKGDPRDRTTTF AAMAATLQRVVLGEVL
L2d  141 EVVGGFSALTAFLRGQDSITRNDRNEPDVNLFAKGDPRDRTTDF AAMATSLARFAVGNGL

      Q#ASR.Q#ADWLIDN.TGDACLRAGLGK.WRV.DKTGSNG-D.RNDIAVLWP#AGG.#WV
L2a  200 QPASRQQLADWLIDNETGDACLRAGLGKRWVGDKTGSNGEDARNDIAVLWPFVAGGAPWV
L2b  200 QPASRQQLADWLIDNETGDACLRAGLGKRWVGDKTGSNGEDARNDIAVLWPFVAGGAPWV
L2c  200 QLASRQQLADWLIDNETGDACLRAGLGKLRVDRKTGSNGEDARNDIAVLWPFVAGGAPWV
L2d  201 QFASRRQFADWLIDNQTGDACLRAGLGKRWVGDKTGSNGDDTRNDIAVLWPHAGGTAVV

      #TAYLQ...#...-QRA.VLA.VG.#AD.#IG
L2a  260 LTAYLQAGAISYEQRATVLAQVGRIADRLIG
L2b  260 LTAYLQAGAISYEQRASVLAQVGRIADRLIG
L2c  260 LTAYLQAGAISYEQRATVLAQVGRIADRLIG
L2d  261 VTAYLQGASVDDDQRAAVLARV GALADAMIG
  
```

**Figure 56:** Sequence alignment of L2a, L2B, L2c and L2d  $\beta$ -lactamase. Conserved residues are highlighted with green. Dark green colour residues have higher consensus strength than the light green ones (lower RMSD). Conserve substitutions are highlighted with yellow. Nonconserved residues are coloured with white. The L2  $\beta$ -lactamases binding sites are highlighted with red box.

Out of the four L2  $\beta$ -lactamases, only L2b structure is available. The SWISS-MODEL website was used to generate the models of the other three proteins (L2a, L2c and L2d).<sup>219-221</sup> SWISS-MODEL is an automatic comparative modelling server for building 3D protein structures.<sup>222</sup> It is a free web based automated modelling tool and most widely used in this field.<sup>222</sup> Only the sequence need to be provided to build the model. The template selection, alignment and model building are automatically finished by the server.<sup>222</sup> In this case, L2a, L2c and L2d sequences were input into SWISS-MODEL. To consider the relationship between these four L2 enzymes, L2b  $\beta$ -lactamase (PDB number 5NE2) was selected as the template for L2a, L2c and L2d homology modelling. There is a D-glutamic acid bind to L2b  $\beta$ -lactamase in 5NE2. There are two type of glutamic acid exist: L-glutamic acid and D-glutamic acid. D-glutamic acid mainly exist as an essential component in the cell wall of certain bacteria.<sup>223</sup> In L2b  $\beta$ -lactamase, the D-glutamic acid is a crystallisation artefact due to the crystallisation buffer has excess amino acid.<sup>104</sup> Therefore, the D-glutamic acid should not be considered in the simulation. The models were built automatically by the server. Ramachandran plot was employed for model validation (Figure 57). In L2a Ramachandran plot, 96.64% of residues are in the Ramachandran favoured region, no residue in the disallowed region (Figure 57-L2a). L2c Ramachandran plot has 96.62% residues in the Ramachandran favoured region, no residue in the disallowed region (Figure 57-L2c). L2d model's Ramachandran plot has 95.47% residues in the Ramachandran favoured region, no residue in the disallowed region (Figure 57-L2d). All three models pass the validation and will be used in the next step.



**Figure 57:** Ramachandran plots of L2a, L2c and L2d models. The dark green region is the most favoured area. Green region is the additional allowed area. Light green region is the generously allowed area. White region is the disallowed area. The sum of dark green region and green region is the Ramachandran favoured region. No residue is in the disallowed area.

## 6.2 Set up and Running the Simulation

After the four L2  $\beta$ -lactamase models were ready, the next step was to set up the models to run the simulations. All hydrogen atoms were removed using PyMOL visualisation software.<sup>224</sup> The purpose of this step was to remove inconsistent naming, and nonstandard or incorrectly added hydrogens during the modelling process.<sup>225</sup> Besides, the PDB files were edited for correct reading into the H++ server. All HETATM records were changed to ATOM, TER record was added between each chain, ion and water molecules. After editing, the PDB model files were uploaded to H++ server to obtain the correct protonation state for each residue. The H++ server, based on the protonation state added the hydrogen atoms.<sup>177</sup> The parameters set in H++ were as follows: solvent box topology/coordinate files (AMBER) were set to be “Cubic”, water model was “TIP3P” and the box edge distance from the solvent was set at “9.0Å”.

The output file from H++ server is a PDB file that can be directly read into the Amber software, which then generates the coordinate (.inpcrd) and the topology file (.prmtop). The first step involves loading the force field, library files and parameter modification files for ions and residues in xleap. In this case, the Amber force field ff14SB, was loaded.<sup>125</sup> The second step was to add counterions such that the net charge on the system is zero. Next the system needed to be solvated with TIP3P water and the box edge was set to 9 Å from the nearest protein atom. After finishing the steps above, the topology file (.prmtop) and the coordinate file (.inpcrd) were generated. The NME (N-terminal methyl amide) and ACE (C-terminal acetyl) neutral groups were used for capping chain

termini.

During the simulations, the systems were first energy minimised for 1000 iterations of steepest descent. Then an equilibration protocol was run for 5ns at 1 atmospheric pressure, 300K temperature using the NPT ensemble. The system was equilibrated by keeping the backbone atoms fixed and allowing the side chain atoms, ions and water to move. The simulation time step was set at 4fs and the periodic boundary conditions were applied. The conventional simulation ran for 5ns at 1 atmospheric pressure. After the equilibration, the production run was carried out as multiple short MSM (Markov State Models)-based adaptive sampling simulations. Each trajectory ran for 50 ns (500 frames) in total.<sup>226</sup> The simulation generated at least 1500 trajectories for each protein system. ACEMD molecular dynamics engine was used to run the MD simulations.<sup>227</sup>



## 6.3 MDLofit Results

MDLofit was employed for analysing the RMSD, RMSF and related regions. <sup>228</sup>

MDLofit had its advantage in analysing the mobility of different backbone regions.

<sup>228</sup>In L2 research, the cut off for C $\alpha$  atom RMSD was set at <0.9Å. The cut off need to be high enough to avoid regions of less importance and low enough to make each L2  $\beta$ -lactamase display the dynamic regions. After trying several cut off values, 0.9Å was selected for the analysis.

The root-mean-square deviation (RMSD), was used to assess the average distance between atoms after superimposing them. RMSD help in studying the stability of the protein and RMSF help in identifying the flexibility of the structures.

The RMSD is calculated as:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2}$$

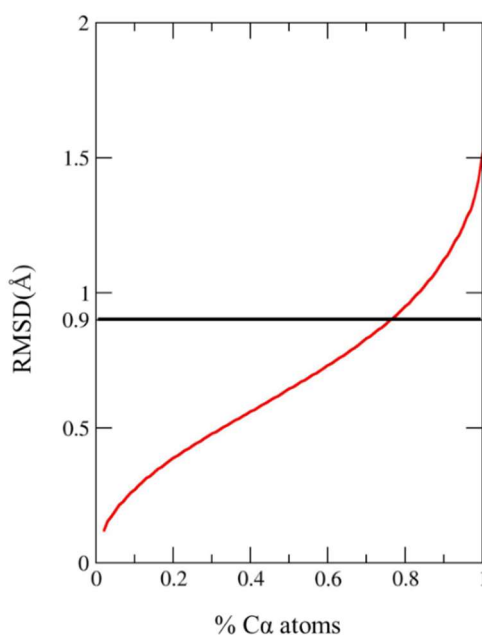
$\delta_i$  was the distance between two atoms  $i$  from different but superimposed proteins,  $n$  was the number of structure. The RMSF was the time average of RMSD. It was calculated as:

$$p_i = \sqrt{(x_i - \langle x_i \rangle)^2}$$

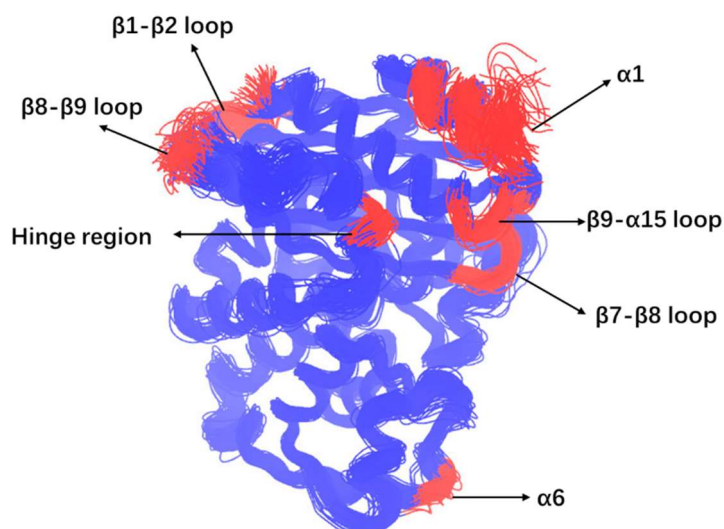
$x_i$  was the coordinate of particle I,  $\langle x_i \rangle$  was the average position of particle  $i$ .

## L2a $\beta$ -lactamase results

In L2a  $\beta$ -lactamase simulation, about 25% of  $C\alpha$  RMSD was more than  $0.9\text{\AA}$  (Figure 58). Several regions display  $C\alpha$  RMSD  $> 0.9\text{\AA}$  residues, and include:  $\alpha 1$ ,  $\beta 1$ - $\beta 2$  loop,  $\alpha 6$ ,  $\beta 7$ - $\beta 8$  loop,  $\beta 8$ - $\beta 9$  loop,  $\beta 9$ - $\alpha 15$  loop (Figure 59).



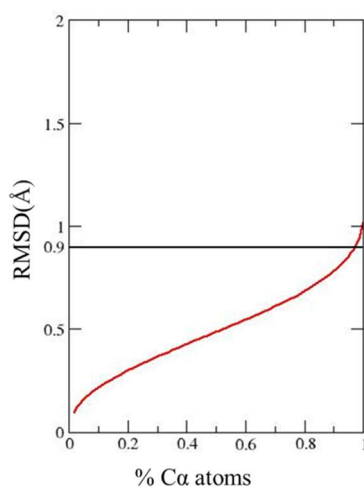
**Figure 58:** 25% of  $C\alpha$  RMSD was more than  $0.9\text{\AA}$  in L2a  $\beta$ -lactamase.



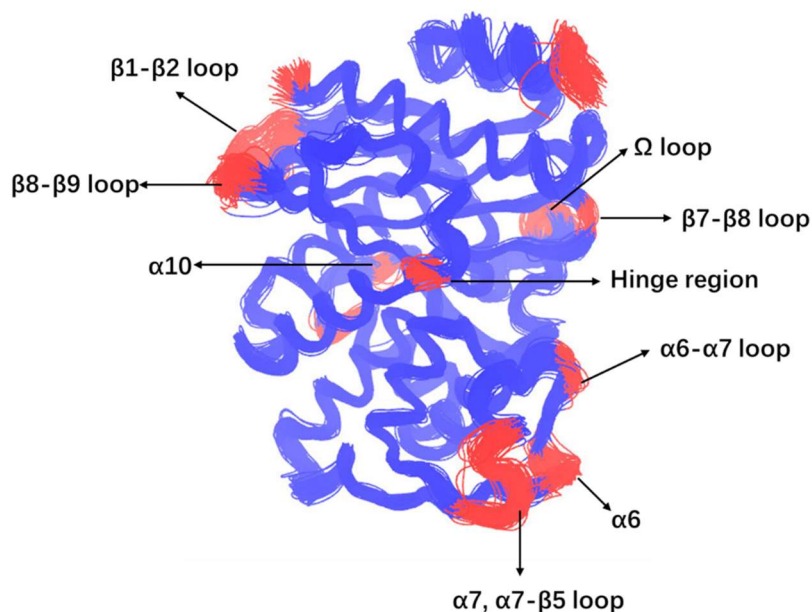
**Figure 59:** MD Lovofit result of L2a  $\beta$ -lactamase. Red colour highlights the high dynamic residues, where  $C\alpha$  RMSD was more than  $0.9\text{\AA}$ . Several regions contain these residues, including:  $\alpha 1$ ,  $\beta 1$ - $\beta 2$  loop,  $\alpha 6$ , hinge region,  $\beta 7$ - $\beta 8$  loop,  $\beta 8$ - $\beta 9$  loop and  $\beta 9$ - $\alpha 15$  loop.

## L2b $\beta$ -lactamase results

In L2b  $\beta$ -lactamase, only 3% of C $\alpha$  RMSD was more than 0.9Å (Figure 60). There were multiple regions had C $\alpha$  RMSD > 0.9Å residues in L2b  $\beta$ -lactamase, include:  $\beta$ 1- $\beta$ 2 loop,  $\alpha$ 6,  $\alpha$ 6- $\alpha$ 7 loop,  $\alpha$ 7,  $\alpha$ 7- $\beta$ 5 loop,  $\Omega$  loop,  $\alpha$ 10,  $\alpha$ 10- $\alpha$ 11 loop,  $\alpha$ 13- $\alpha$ 14 loop,  $\beta$ 7- $\beta$ 8 loop and  $\beta$ 8- $\beta$ 9 loop (Figure 61).



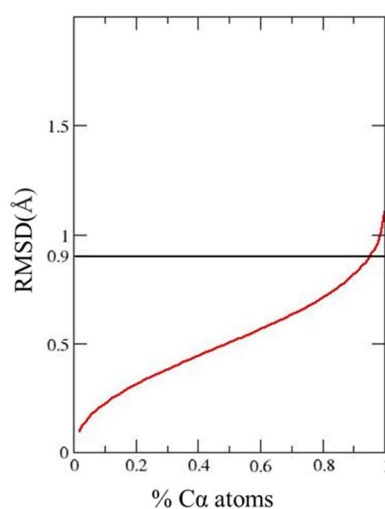
**Figure 60:** 3% of C $\alpha$  RMSD was more than 0.9Å in L2b  $\beta$ -lactamase.



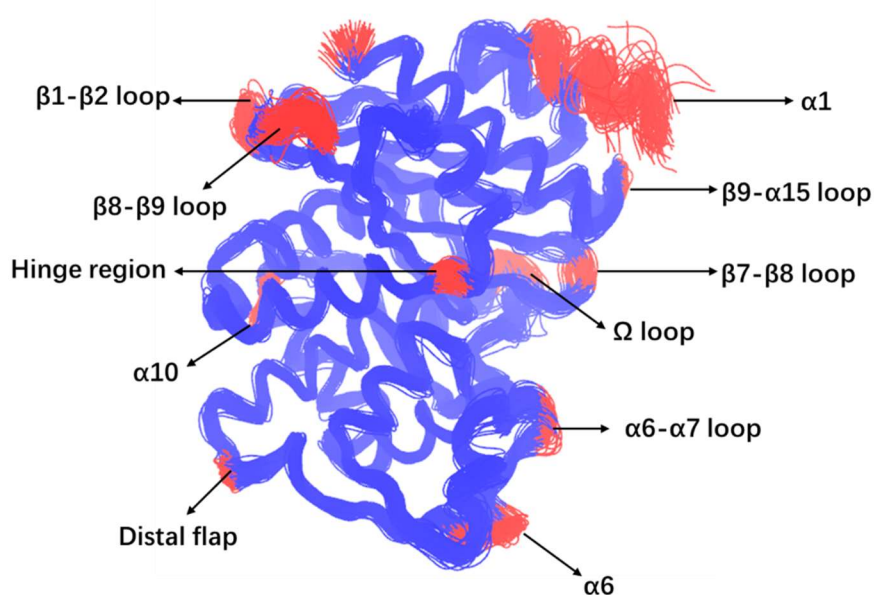
**Figure 61:** MDLovofit results of L2b  $\beta$ -lactamase. Red colour highlights the high dynamic residues where C $\alpha$  RMSD was more than 0.9Å. Several regions contain these residues, including:  $\beta$ 1- $\beta$ 2 loop,  $\alpha$ 6,  $\alpha$ 6- $\alpha$ 7 loop,  $\alpha$ 7,  $\alpha$ 7- $\beta$ 5 loop,  $\Omega$  loop,  $\alpha$ 10, hinge region,  $\beta$ 7- $\beta$ 8 loop and  $\beta$ 8- $\beta$ 9 loop.

## L2c $\beta$ -lactamase results

In L2c  $\beta$ -lactamase, 4% of C $\alpha$  RMSD was more than 0.9Å (Figure 62). Several regions show C $\alpha$  RMSD > 0.9Å residues, including:  $\alpha$ 1,  $\beta$ 1- $\beta$ 2 loop, distal flap,  $\alpha$ 6,  $\alpha$ 6- $\alpha$ 7 loop,  $\Omega$  loop,  $\alpha$ 10,  $\alpha$ 13- $\alpha$ 14 loop,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop (Figure 63).



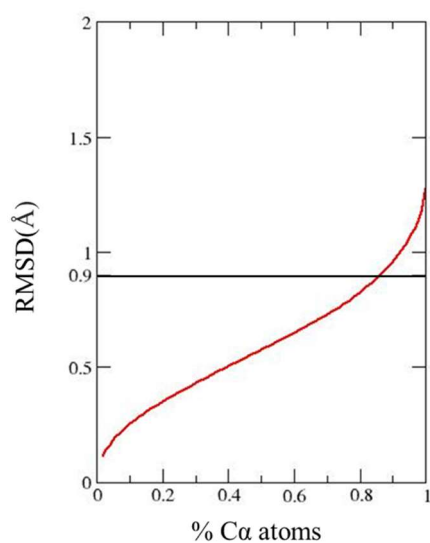
**Figure 62:** 4% of C $\alpha$  RMSD was more than 0.9Å in L2c  $\beta$ -lactamase.



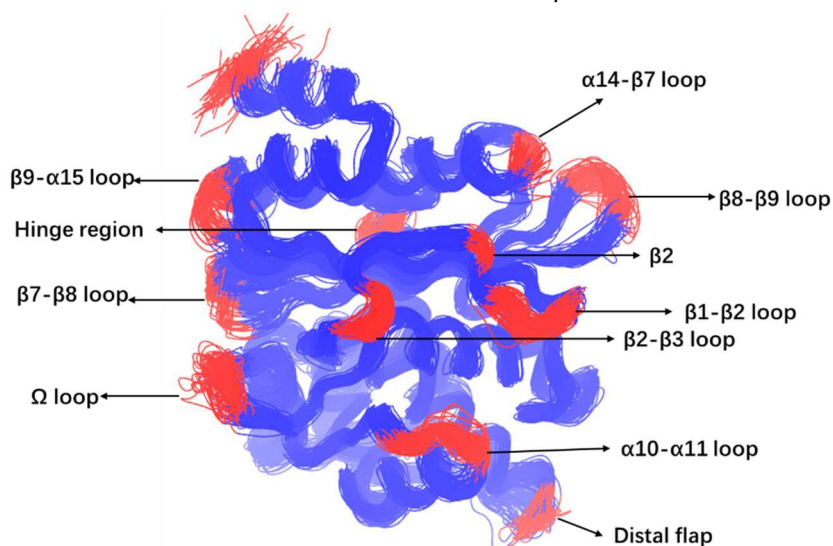
**Figure 63:** MDLovofit result of L2c  $\beta$ -lactamase. Red colour highlights the high dynamic residues, where C $\alpha$  RMSD was more than 0.9Å. Several regions contain these residues, including:  $\alpha$ 1,  $\beta$ 1- $\beta$ 2 loop, distal flap,  $\alpha$ 6,  $\alpha$ 6- $\alpha$ 7 loop,  $\Omega$  loop,  $\alpha$ 10,  $\alpha$ 13- $\alpha$ 14 loop,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop.

## L2d $\beta$ -lactamase results

In L2d  $\beta$ -lactamase, 13% of C $\alpha$  RMSD was more than 0.9Å (Figure 64). Several regions display C $\alpha$  RMSD > 0.9Å residues, including:  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 2,  $\beta$ 2- $\beta$ 3 loop, distal flap,  $\Omega$  loop,  $\alpha$ 10- $\alpha$ 11 loop,  $\alpha$ 13- $\alpha$ 14 loop,  $\alpha$ 14- $\beta$ 7 loop,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop (Figure 65).



**Figure 64:** 13% of C $\alpha$  RMSD was more than 0.9Å in L2d  $\beta$ -lactamase.



**Figure 65:** MD Lovofit results of L2d  $\beta$ -lactamase. Red colour highlight the high dynamic residues, where C $\alpha$  RMSD > 0.9Å. Several regions contain these residues, including:  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 2,  $\beta$ 2- $\beta$ 3 loop, distal flap,  $\Omega$  loop,  $\alpha$ 10- $\alpha$ 11 loop, hinge region,  $\alpha$ 14- $\beta$ 7 loop,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop.

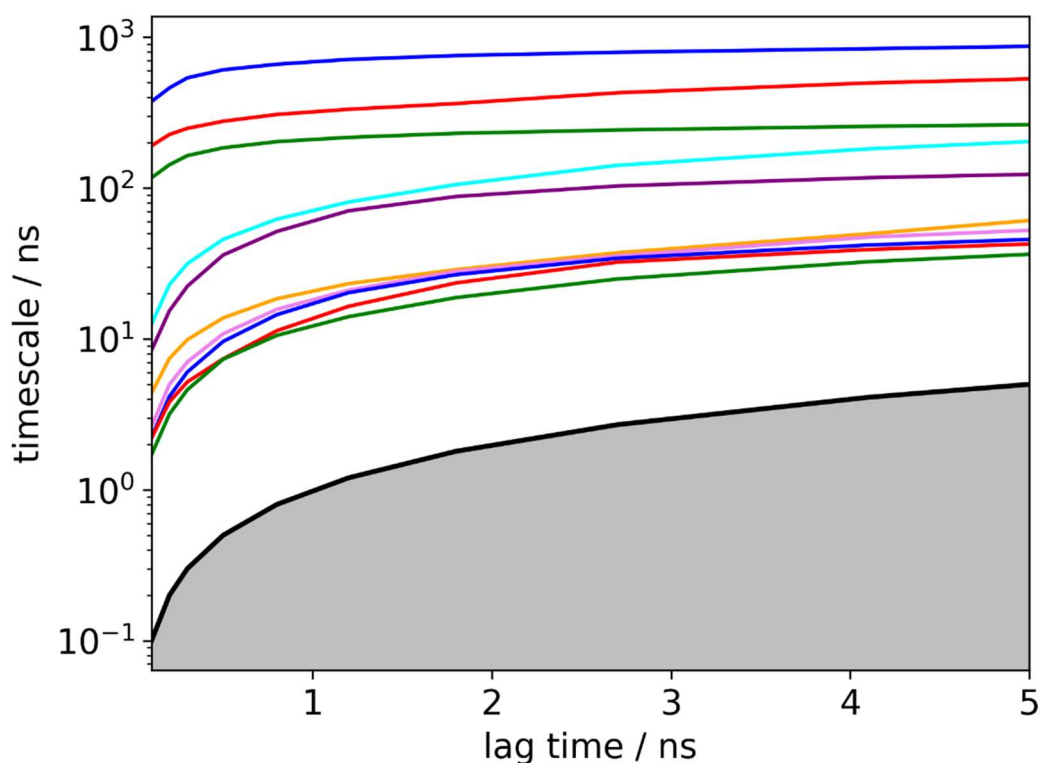
## 6.4 Markov State Models

PyEMMA v2.5.7 was used to build the MSM (Markov State Models).<sup>187</sup> All trajectories from 4 systems were loaded into PyEMMA separately. With selected features, the MSMs permit the resolution of metastable states and highlight structural changes during the kinetic transition between different states. To build the L2  $\beta$ -lactamase MSMs, the selected features were backbone torsions, RMSD and side chain torsion of hydrophobic nodes residues.<sup>229</sup> Hydrophobic nodes are consist of stretches of 3 to 9 highly conserved hydrophobic residues. Each stretch is called as a hydrophobic node.<sup>230</sup> These nodes exist in all class A  $\beta$ -lactamases.<sup>230</sup> These hydrophobic nodes were found to have local contact with each other and also help stabilizing the tertiary structures.<sup>229</sup> The class A  $\beta$ -lactamases show a dynamic similarities on these nodes and also important for the whole structure dynamics.<sup>229</sup>

After features were selected, all trajectories were loaded into the PyEMMA software with a stride of 1. The tICA method was employed to reduce the dimensionality of the data and was allocated into 3 components (ICs). The selected features were kept identical in all four L2  $\beta$ -lactamases to make comparisons reliable. Based on the MSM results, by comparing the sampled structures from different states led to significant understanding of the kinetic changes. The important structural dynamics can be observed during the conformation transitions from one state to the other.

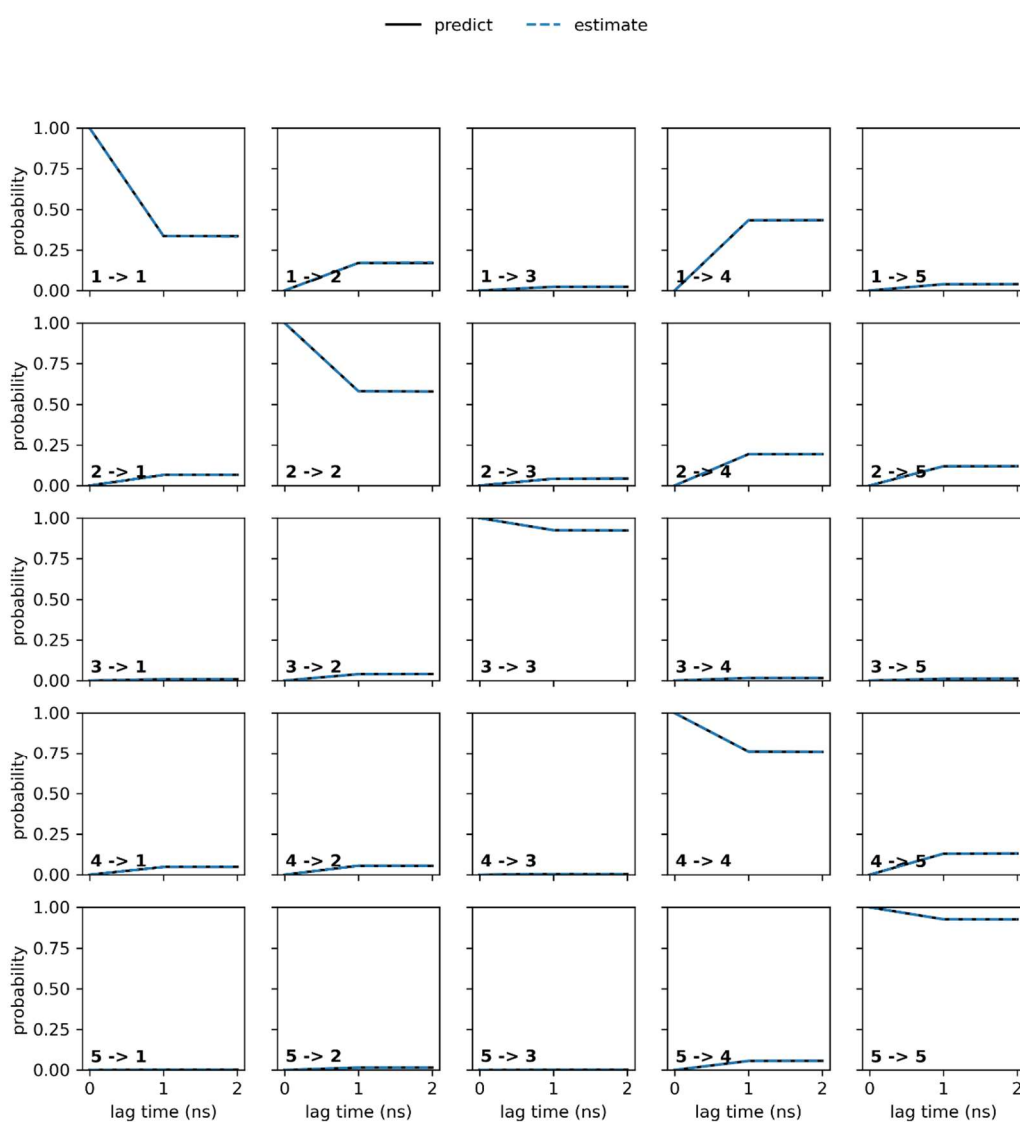
## L2a $\beta$ -lactamase results

To build a L2a  $\beta$ -lactamase MSM, the lag time was set to 1ns, the cluster number was 80, and 5 states were selected. To verify the convergence of the MSMs, an implied timescale plot was generated and is shown in Figure 66. The implied timescale refers to the relaxation timescales of input molecules.<sup>188</sup> The timescale is estimated by a MSM transition matrix at a certain lag time. The implied timescale plot confirmed the convergence behaviour of the process at 1ns lag time. The different coloured lines indicate the maximum likelihood results, the thick black line over the grey area indicates when lag time and timescale were equal. The black line with the grey area under it describes an area where the MSM cannot resolve processes.



**Figure 66:** Implied time scale plot for L2a  $\beta$ -lactamase. The total implied time scale was set at 5ns to show the converged processes. Different colour lines indicate the maximum likelihood results. Flat lines at 1ns indicates that 1ns is an appropriate lag time selection.

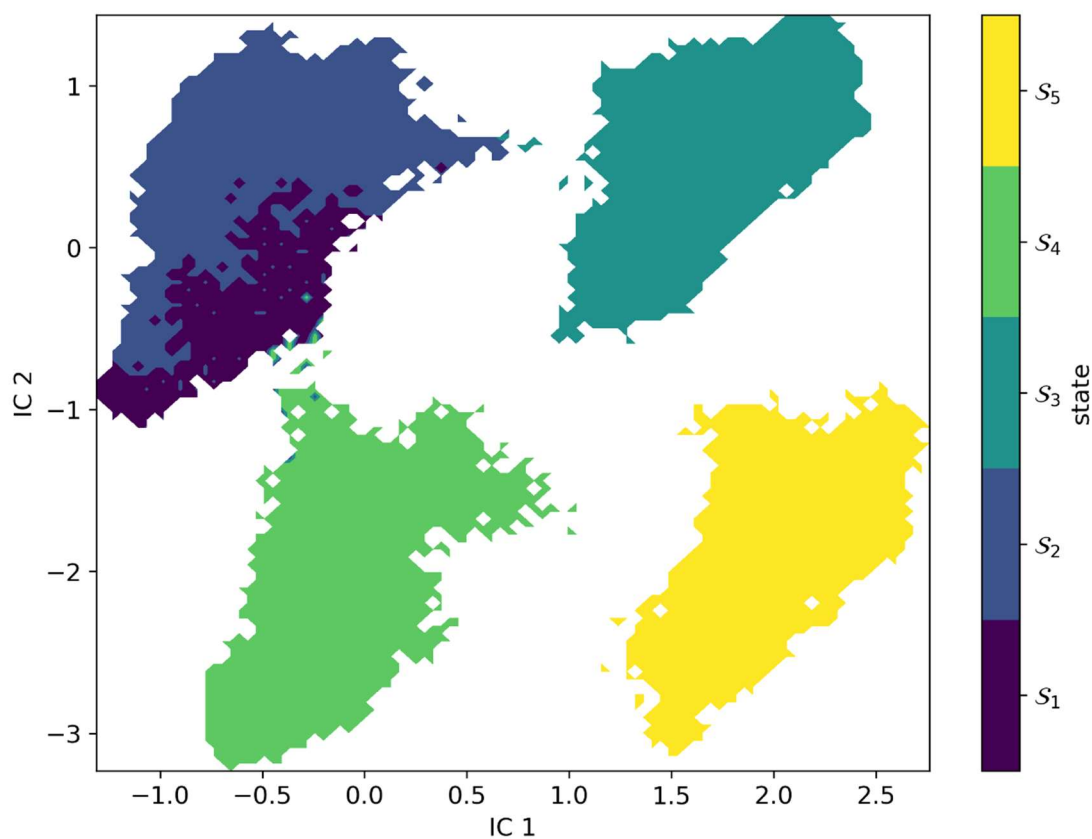
The CK plot of 5 metastable states with a lag time of 5.0ns illustrates an estimated (solid line) and a predicted (dashed line) projection within confidence intervals (Figure 67). If these two lines overlap with each other, it indicates the states pass the Bayesian sampling procedure with 95% confidence intervals. The results of L1  $\beta$ -lactamase CK plot indicate the estimated and the predicted values are within the range of confidence interval.



**Figure 67:** Chapman-Kolmogorov test plot for L2a  $\beta$ -lactamase. All five states passed the CK test plot.

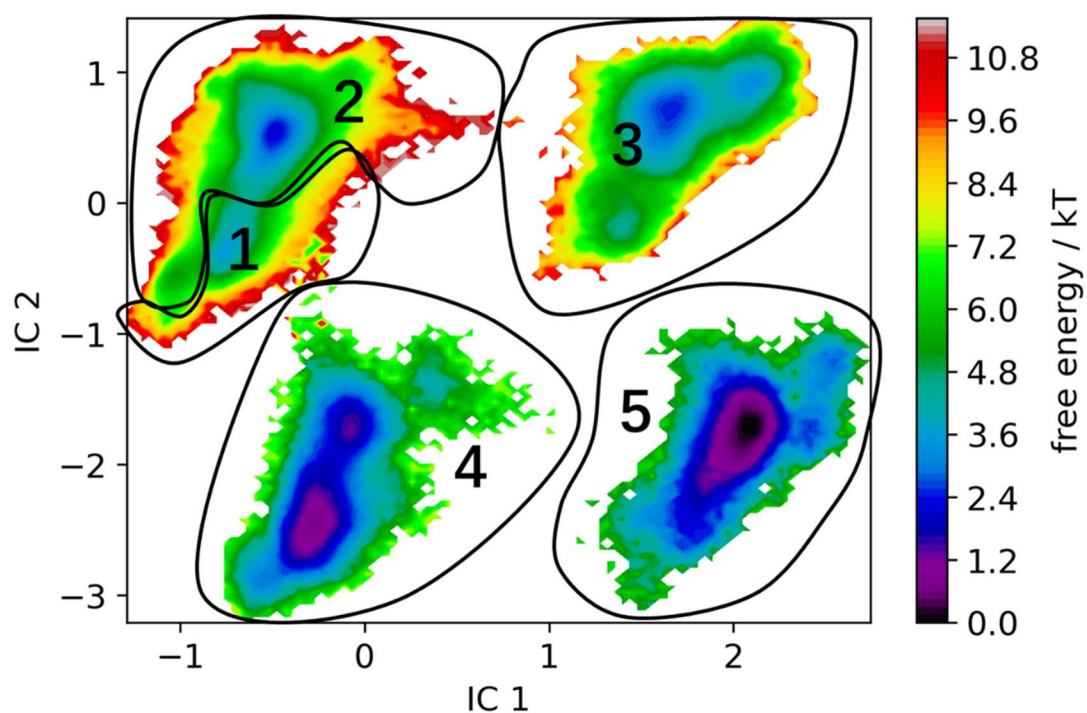


The PCCA distribution probability plot (Figure 68) highlights the clusters of different metastable states. A clean distribution indicates a clear conformational clustering. The conformations with the same feature are clustered into the same state. The first two independent components were used to identify 5 defined metastable states.



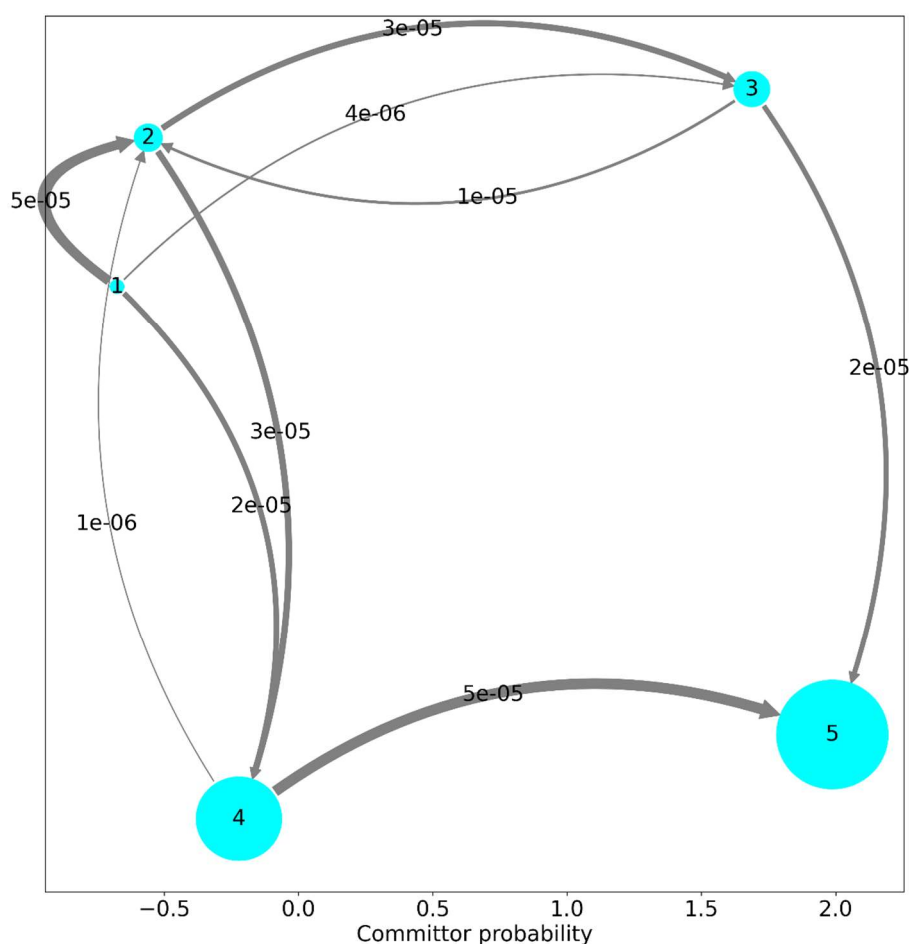
**Figure 68:** PCCA distribution plot for 5 states in L2a  $\beta$ -lactamase. Different colours in the distribution represented different states. The conformations in the same state display same features.

Free energy landscape shows the energy distribution at selected lag time. The free energy distribution at 1ns lag time is shown in Figure 69. The free energy (FE) calculated at lag time 1ns was projected into the first 2 independent components. Multiple free energy minima are observed in this plot. 4 main minima and several sub minima are shown. Each main minima correspond to a metastable state. Only the sub minima from upper-left corner can be separated and represent for state 1. The PCCA distribution tries to separate similar conformations representing metastable states into distinct free energy minima. The structures in the same energy minima are similar.



**Figure 69:** L2a  $\beta$ -lactamase free energy landscape at lag time 1ns. Multiple energy minima could be observed in the plot. The minima correspond to the metastable states in PCCA distribution plot.

The net flux pathways plot calculates the productive recrossing between pairs of states. The net flux plot of L2a is shown in Figure 70. The plot shows the transition pathways from state 1 (highest energy state) to the other states. State 4 and 5 are the most populated states in L2a  $\beta$ -lactamase. The arrow indicates the difference value of the fluxes in both directions. Therefore, state 4 structures will eventually go to state 5 and state 5 is most stable state in all 5 metastable states.



**Figure 70:** Net flux pathway plot of L2a  $\beta$ -lactamase. Each cyan circle with certain number represented the correspond state. The bigger the circle, the more populated the state. Arrows between each state represented the possibility of each transition. The thicker the arrow, the higher the possibility.

Mean first passage time is an important parameter in MSM. It is an average timescale for a random event to first occur.<sup>164</sup> The mean first passage times between different states in L2a  $\beta$ -lactamase is shown in Table 3. The number between each two metastable states indicates the timescale for the conformational change to first happen. Larger timescale implies a higher energy barrier and lower transition possibility between states. For instance, most structures transit from state 1 to state 4, because the energy barrier between these two states was lowest at 824 ns. Table 4 shows the population ( $\pi$ ) of each state and the free energy estimation. State 4 and 5 represents for about 90% of all the structures.

**Table 3:** Mean first passage times between different states in L2a  $\beta$ -lactamase. Most conformation transit from state 1 to state 4 and then transit from state 4 to state 5.

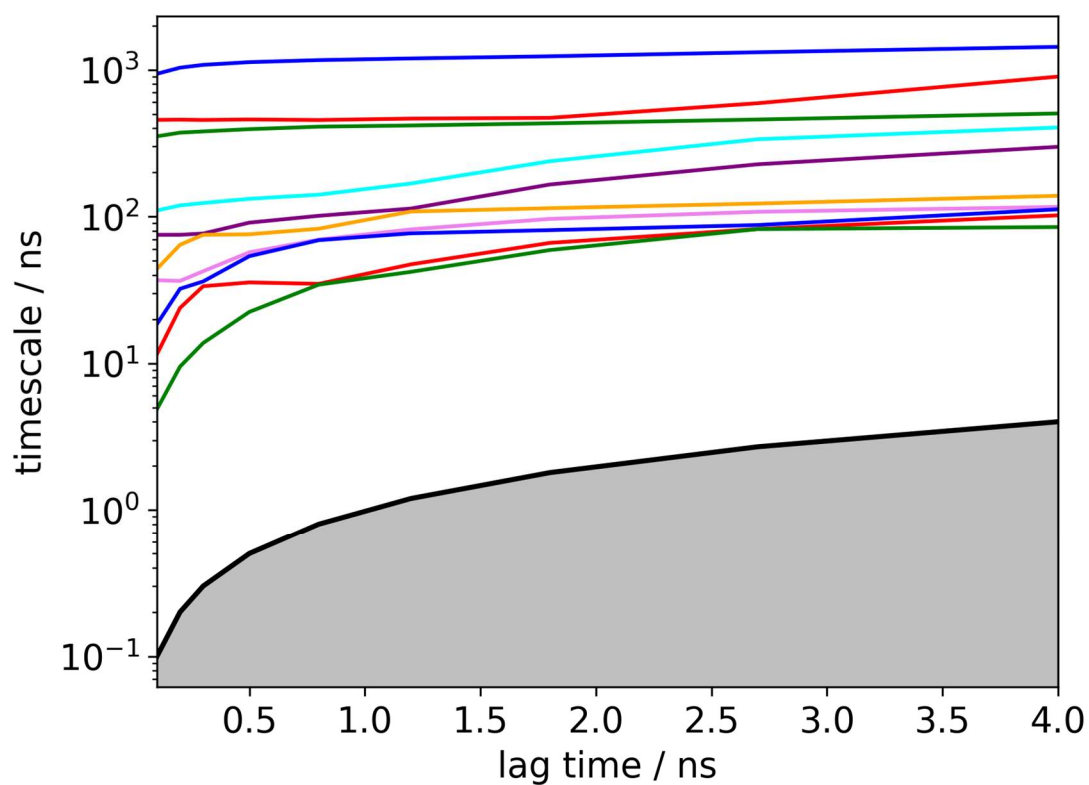
State/(ns)	1	2	3	4	5
1	0.00	1647.50	5733.05	824.48	1010.65
2	5246.50	0.00	4648.09	1002.49	1078.17
3	8579.77	4515.40	0.00	1334.94	1016.36
4	11839.44	8634.43	9292.10	0.00	521.15
5	12333.83	9039.02	9274.71	814.79	0.00

**Table 4:** L2a  $\beta$ -lactamase 5 states population ( $\pi$ ) and their free energy estimation.

State	$\pi$ (%)	G/kT (J)
1	0.8779	4.735404
2	3.5667	3.333517
3	5.8663	2.835938
4	33.1852	1.103066
5	56.5038	0.570862

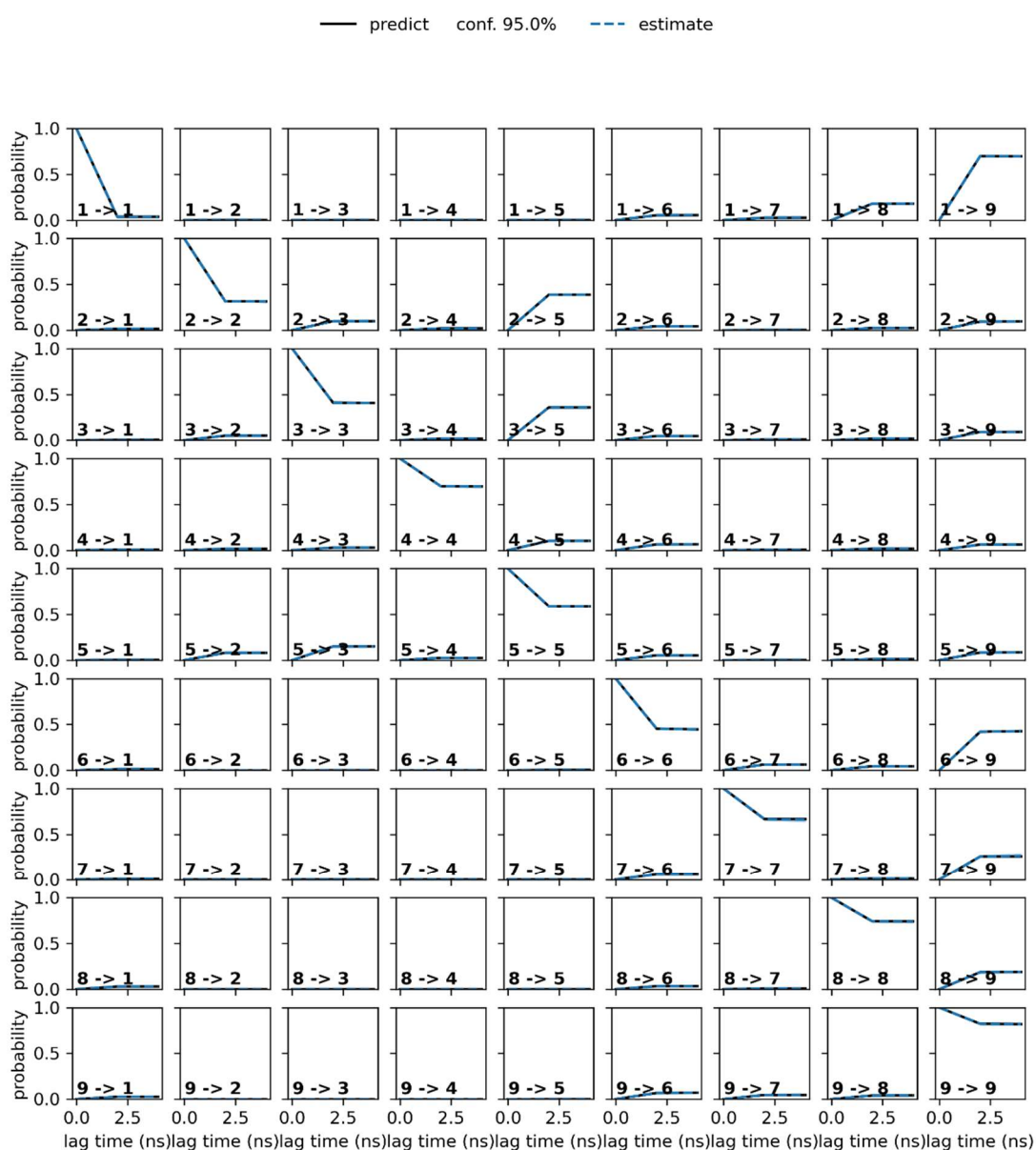
## L2b $\beta$ -lactamase results

In L2b  $\beta$ -lactamase MSM, the lag time was set at 4 ns, the cluster number was 100, and 9 states were built. Figure 71 is the implied timescale plot for L2b  $\beta$ -lactamase. The lines become flat and converged at 4ns lag time, which was deemed acceptable for the subsequent analysis.



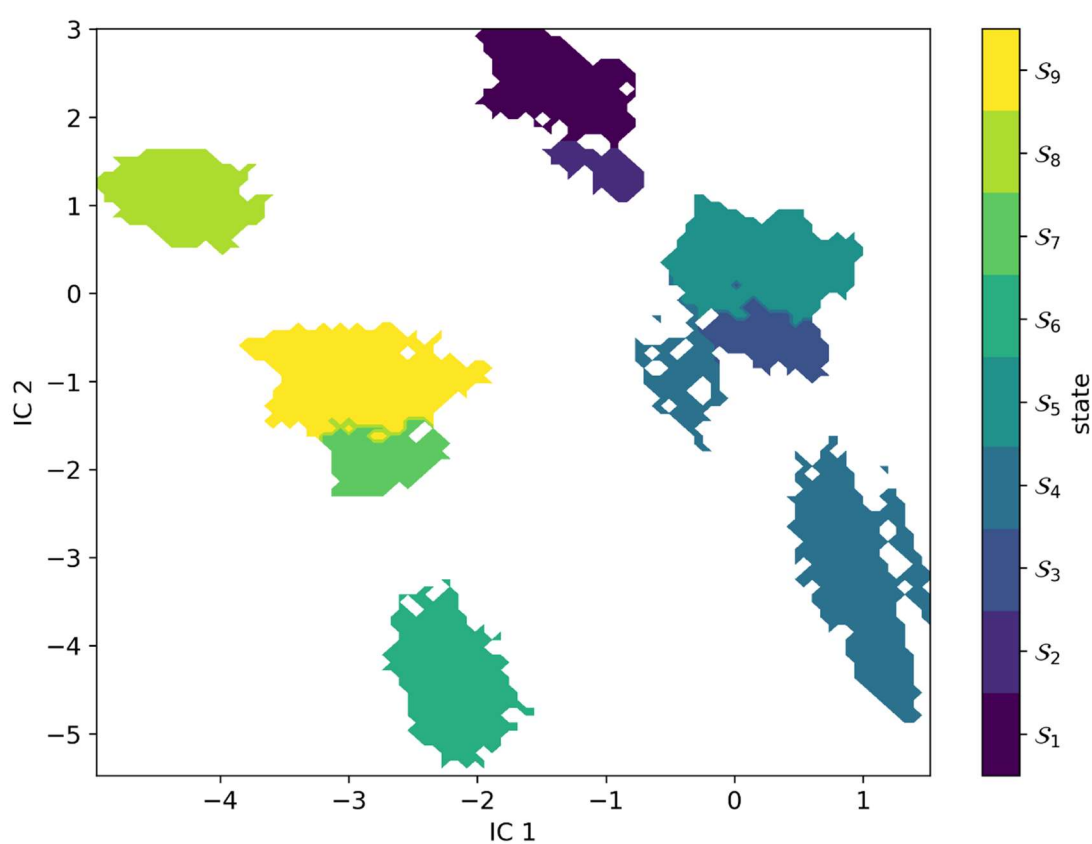
**Figure 71:** L2b  $\beta$ -lactamase implied timescale plot, lag time was set at 4ns. The converged lines indicate the 4ns is an appropriate lag time selection.

9 states were built in L2b  $\beta$ -lactamase MSM. The CK test plot for L2b  $\beta$ -lactamase is shown in Figure 72. The CK test verified 9 metastable states of L2b  $\beta$ -lactamase at 4ns lag time. The predicted line and estimate line overlaps with each other indicating 9 states were acceptable for L2b  $\beta$ -lactamase MSM and subsequent analysis.



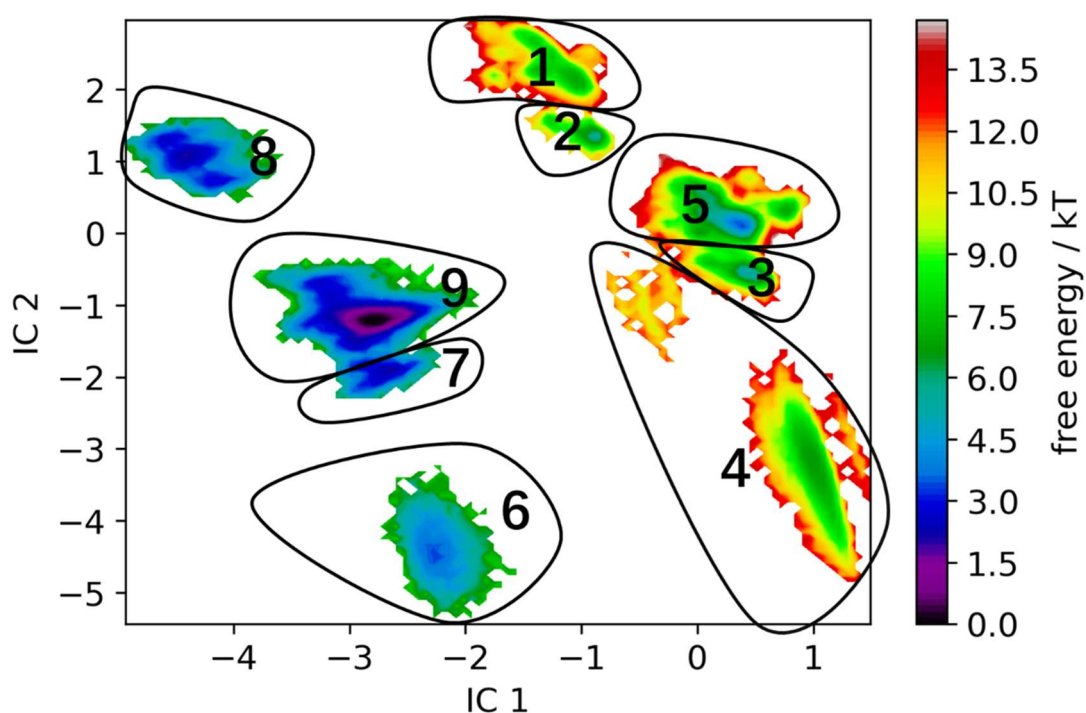
**Figure 72:** Chapman-Kolmogorov test plot for L2b  $\beta$ -lactamase. All 9 states passed CK test within confidence intervals.

There were 9 states built for L2b  $\beta$ -lactamase. The PCCA distribution plot for L2b  $\beta$ -lactamase is shown in Figure 73. 9 metastable states were well separated. State 4 was tried to split into two substates in order to resolve the clustering. However, more state number cannot give clean separation. Therefore, 9 states were built for the subsequent calculation.



**Figure 73:** L2b  $\beta$ -lactamase PCCA distribution plot. 9 metastable states were well separated. Different colours in the distribution plot represents different states in the simulation. The conformations in the same state exhibit same structural features.

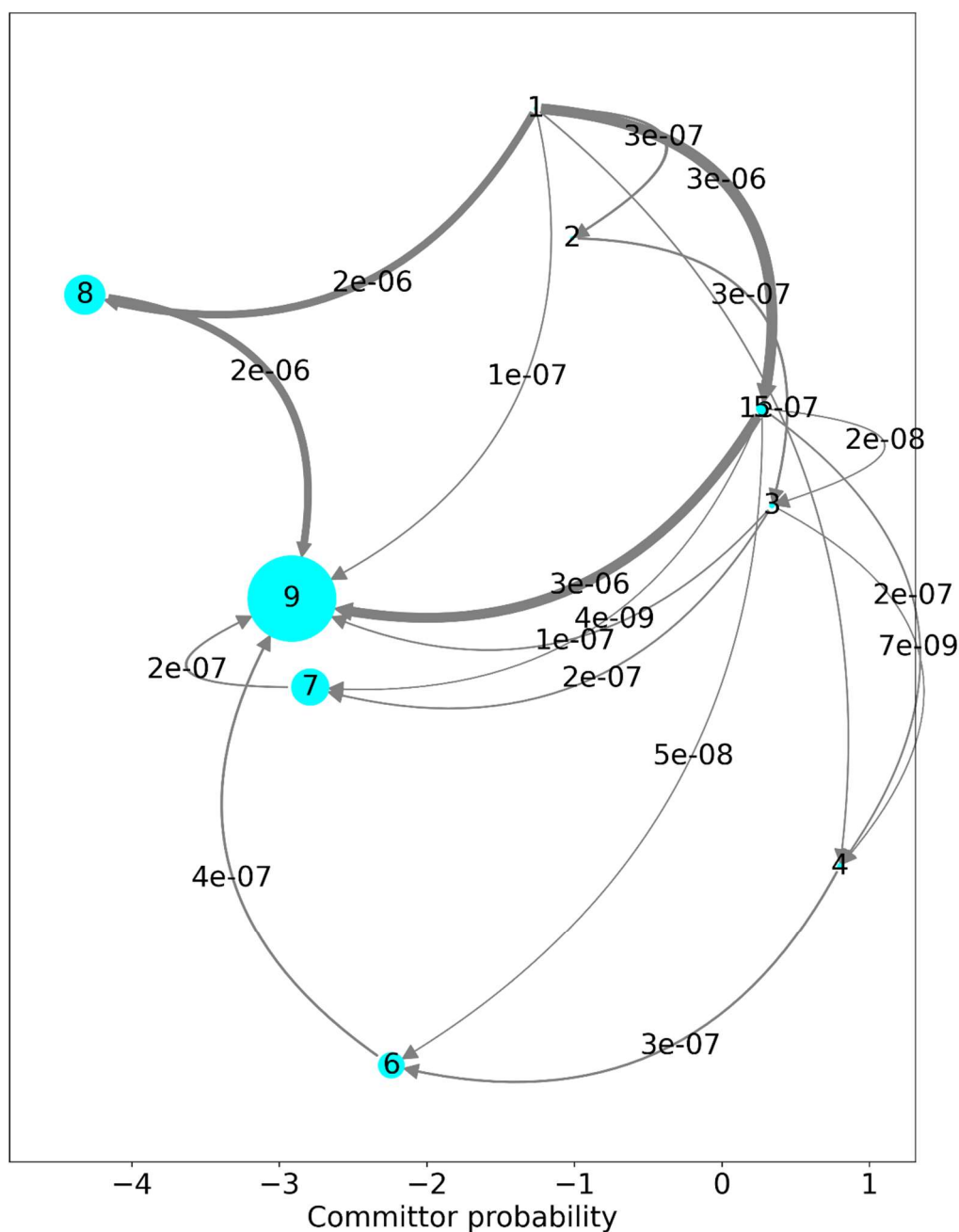
The free energy landscape plot can provide information on the low energy conformations and how they are distributed. The Free energy at lag time 4.0ns was projected into the first 2 independent components (Figure 74). Multiple free energy minima can be observed in this plot. There are two main minima and several sub minima observed. Each low energy region corresponds to a metastable state. State 4 in PCCA distribution plot cannot be separated into substates due to the high energy which could be observed in Figure 74.



**Figure 74:** L2b  $\beta$ -lactamase free energy landscape at lag time 4ns. Two main energy minima and several sub minima observed. The high energy of state 4 could be the reason why it is difficult to be separated into two states.



The net flux calculates the difference between the values of the fluxes in both directions. The net flux plot of L2b  $\beta$ -lactamase is shown in Figure 75. The plot shows the transition pathways from state 1(highest energy state) to the other states. State 9 was the most populated state in L2b  $\beta$ -lactamase. Most conformations take one of the two pathways 1-5-9 (49.42%) or 1-8-9 (35.70%).



**Figure 75:** L2b  $\beta$ -lactamase net flux pathway plot. Arrows between each state represented the probability of each transition. The thicker the arrow, the higher the probability of transition. The numbers on the arrows indicate the mean first passage times of transition.

Mean first passage time shows an average timescale for a random event to first occur.

The Mean first passage time between different states in L2b  $\beta$ -lactamase is shown in Table 5. The number shows different timescales of conformational transition from one state to another. A higher timescale between states implies a higher energy barrier and a lower possibility of the conformational change. Table 6 lists the population and free energy estimation for the 9 states of L2b  $\beta$ -lactamase.

**Table 5:** Mean first passage times between different states in L2b  $\beta$ -lactamase.

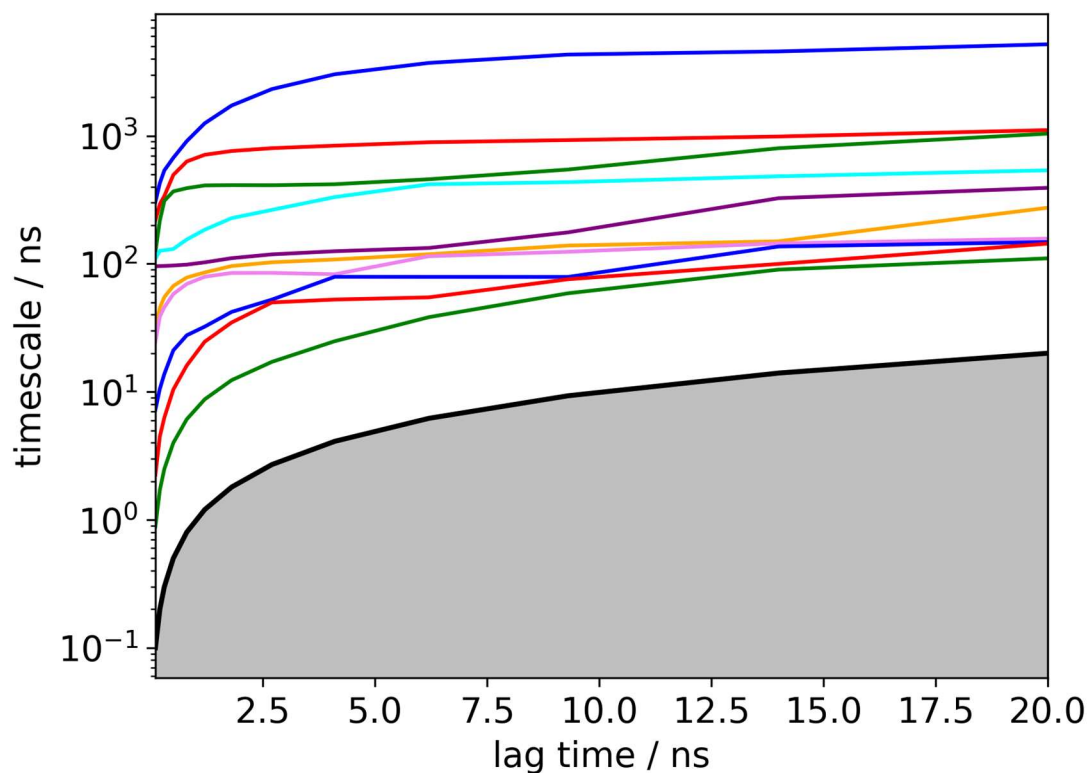
State/(ns)	1	2	3	4	5	6	7	8	9
1	0.00	442038.94	97467.79	265623.21	29438.54	2591.33	2324.49	2573.52	1200.56
2	87010.96	0.00	27573.15	260898.07	14288.12	3078.21	2775.51	3642.07	1704.71
3	122984.94	405007.63	0.00	259364.25	9117.95	2781.35	2462.92	3535.17	1412.59
4	138865.15	485852.16	105650.39	0.00	34324.01	2217.30	2511.58	3585.26	1379.74
5	122789.28	460173.03	76417.52	255486.22	0.00	2694.39	24444.35	3456.64	1335.39
6	185449.92	536977.55	158149.59	311187.68	90381.75	0.00	1266.73	2554.05	92.63
7	185576.13	537067.43	158225.46	311874.82	90523.38	1658.11	0.00	2612.01	159.16
8	184806.68	536888.63	158259.79	311914.69	90499.20	1927.91	1594.39	0.00	426.04
9	185466.89	537008.67	158185.28	311755.03	90426.64	1495.57	1173.76	2458.96	0.00

**Table 6:** L2b  $\beta$ -lactamase 9 states population ( $\pi$ ) and free energy estimation.

State	$\pi$ (%)	G/kT (J)
1	0.0738	7.211293
2	0.0827	7.097262
3	0.1597	6.43935
4	0.1769	6.337267
5	1.0041	4.601057
6	5.6597	2.871796
7	11.8615	2.131871
8	13.911	1.972488
9	67.0104	0.399427

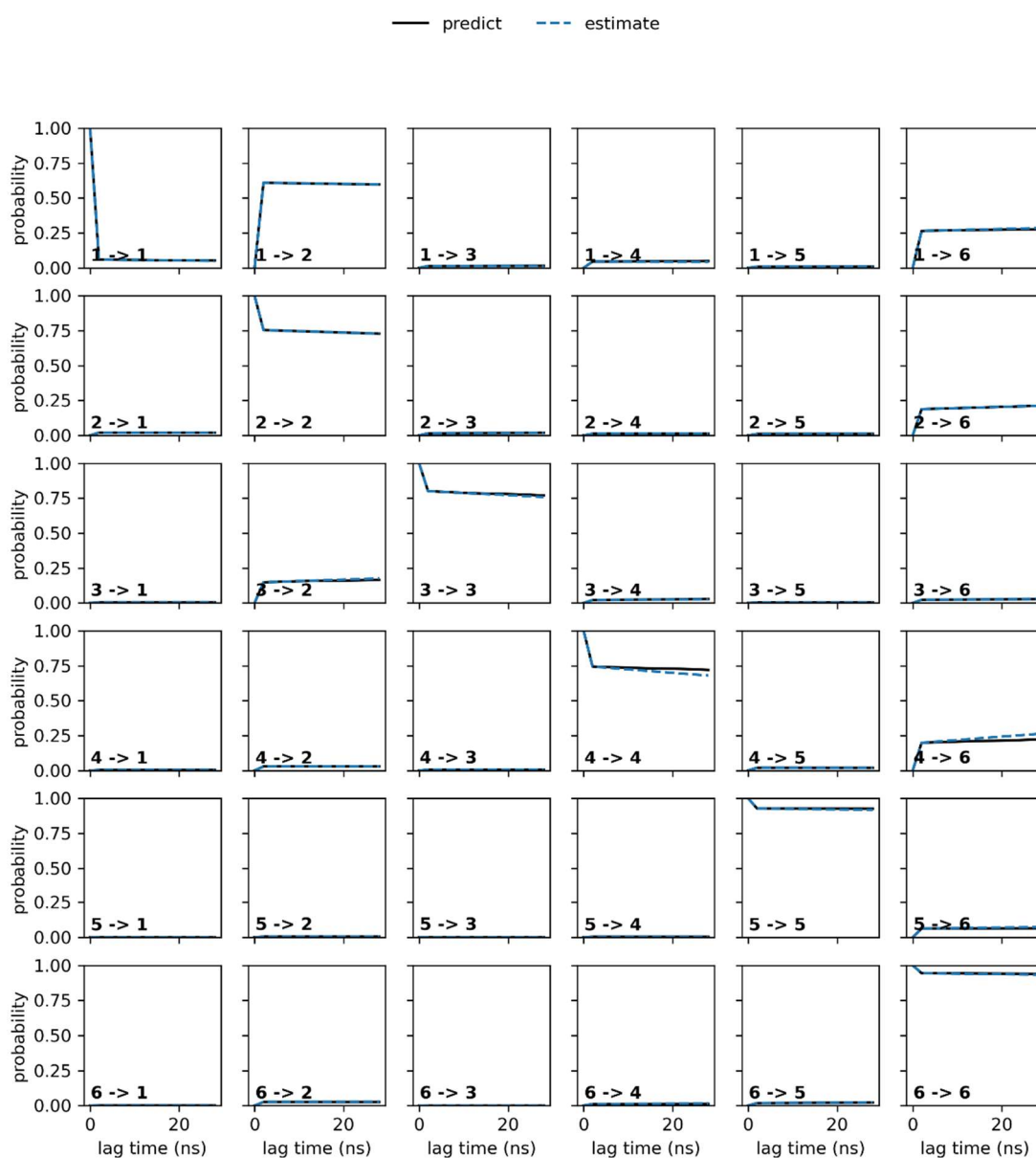
## L2c $\beta$ -lactamase results

For the L2c  $\beta$ -lactamase analysis, the lag time was set at 20ns, the cluster number was 100, 6 states were built. Figure 76 shows the implied timescale plot for L2c  $\beta$ -lactamase. Results were well clustered at 20ns, thus 20ns lag time was used for subsequent calculation.



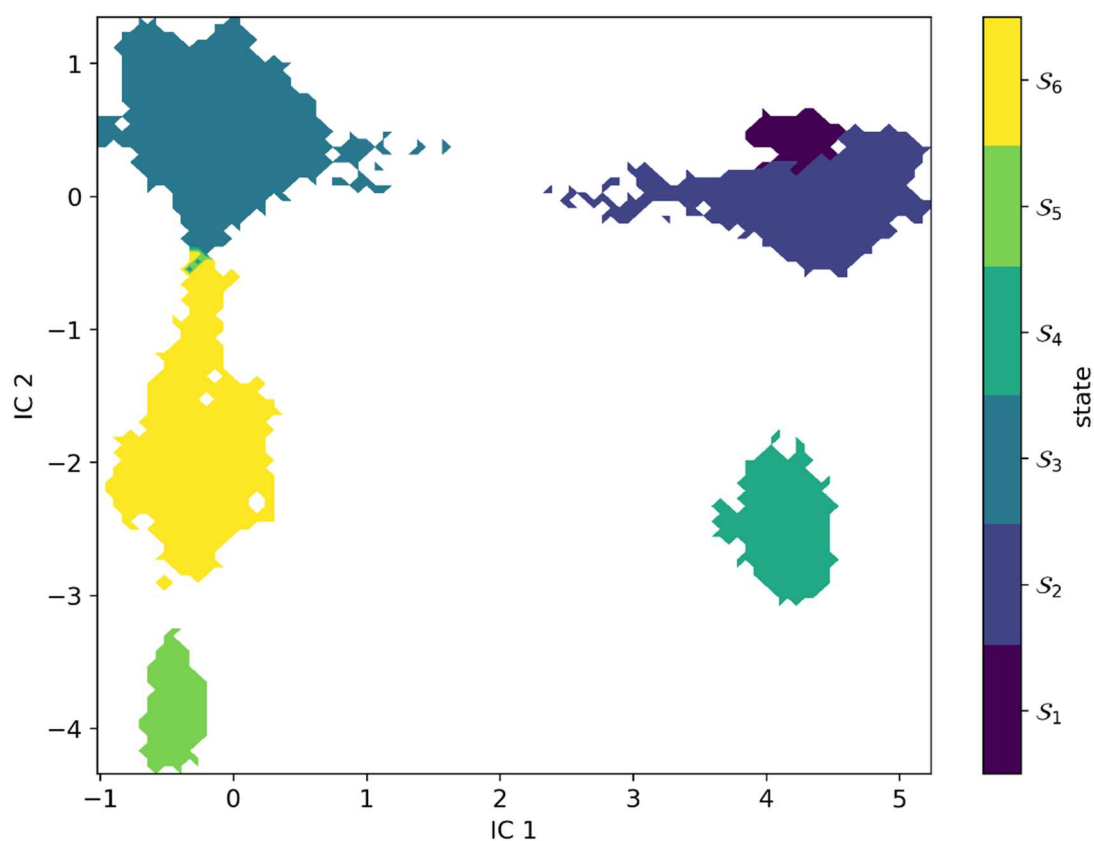
**Figure 76:** L2c  $\beta$ -lactamase implied timescale plot. The lag time was set at 20.0ns. Different colour lines indicate the maximum likelihood results. Flat lines indicate convergence at 20ns lag time.

6 states were built in L2c  $\beta$ -lactamase MSM. The CK test plot for L2c  $\beta$ -lactamase is shown in Figure 77. The CK plot assumed 6 states with 20ns lag time through MSM estimate and is shown as dot lines. The solid line is the prediction line which calculated the metastable states within confidence intervals. The predict line and the estimate line overlap with each other indicating 6 states were acceptable for L2c  $\beta$ -lactamase MSM subsequent calculation.



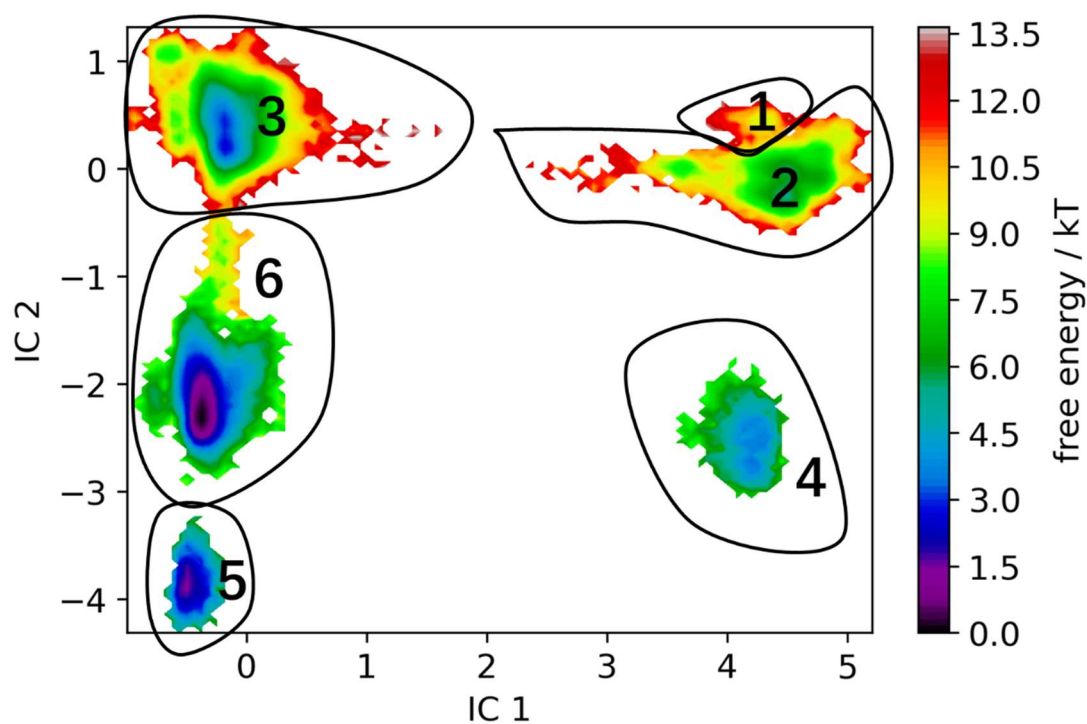
**Figure 77:** Chapman-Kolmogorov test plot for L2c  $\beta$ -lactamase. All 6 states passed CK test within confidence intervals.

The PCCA distribution plot shows how the metastable states are distributed. The plot for L2b  $\beta$ -lactamase is shown in Figure 78. The first two independent components were used to calculate distribution and identified 6 defined metastable states. 6 states were built and are well separated.



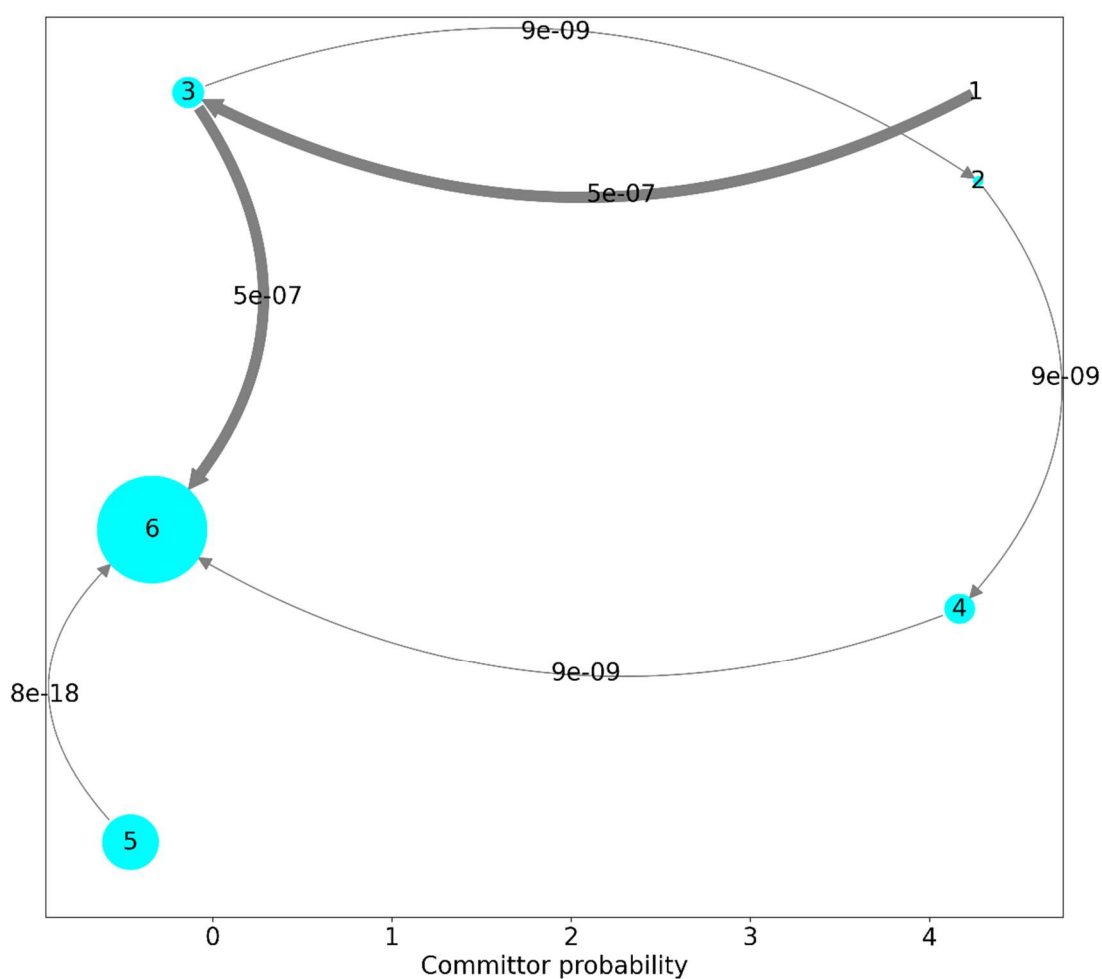
**Figure 78:** L2c  $\beta$ -lactamase PCCA distribution plot. 6 metastable states were well separated. Different colours in the distribution plot represented distinct states clustered from the simulation.

The free energy landscape plot shows the distribution of the free energy minima. The free energy landscape at 20ns lag time is shown in Figure 79. The free energy calculated at lag time 20 ns was projected onto the first 2 independent components. There are 4 main minima and several sub minima observed. The main minima correspond to the metastable states 4, 5 and 6 in PCCA distribution plot. The sub minima are separated into 3 parts and correspond to the metastable states 1,2 and 3.



**Figure 79:** L2c  $\beta$ -lactamase free energy landscape at lag time 20ns. 4 main minima and several sub minima are observed. The minima correspond to the metastable states in PCCA distribution plot.

The net flux plot is generated by TPT object in MSM to show the productive transition between pairs of states. Net flux plot of L2c  $\beta$ -lactamase is illustrated in Figure 80. As shown on the plot, state 6 represents for the most proportion (largest area). The net flux indicates all other states will eventually get into state 6. The path 1-3-6 (98.21%) is the dominant pathway taken by conformations during the simulation.



**Figure 80:** Net flux pathway plot of L2c  $\beta$ -lactamase. Most conformations went through 1-3-6. Each cyan circle with certain number represented the correspond state. The bigger the circle, the more populated the state. Arrows between each state represented the possibility of each transition. The thicker the arrow, the higher the possibility.

The Mean first passage times between different states in L2c  $\beta$ -lactamase is shown in Table 7. The timescale between 1-3-6 is less than that observed in the other pathways. This indicates the energy barrier between these states is much lower and conformational change is easier to happen. Table 8 shows the population ( $\pi$ ) and the free energy estimation of each state. State 5 and 6 represent for more than 85% of all the conformations. While state 1 and 2 represent for less than 1% of the structures. State 6 is the dominant state in L2c simulation.

**Table 7:** Mean first passage times between different states in L2c  $\beta$ -lactamase.

State/(ns)	1	2	3	4	5	6
1	0.00	78535.12	91.60	6925.35	11017.99	937.42
2	1916648.47	0.00	2352.59	6277.80	11260.96	1195.89
3	1912641.03	77168.03	0.00	6866.71	10895.64	830.92
4	1926144.51	87584.33	13173.65	0.00	10428.32	347.54
5	1928254.11	90532.24	15221.02	8445.04	0.00	2250.97
6	1925977.63	88271.17	12960.85	6164.06	10051.24	0.00

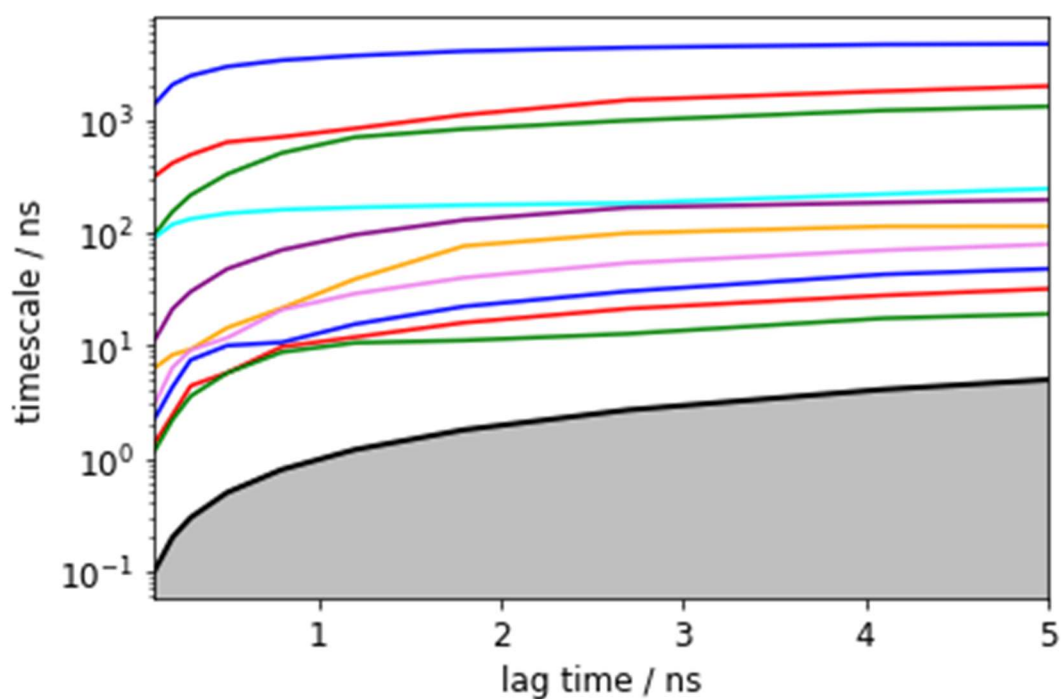
**Table 8:** L2c 6 states population ( $\pi$ ) and free energy estimation.

State	$\pi$ (%)	G/kT (J)
1	0.0049	9.929605
2	0.5085	5.281443
3	5.5731	2.887216
4	5.1581	2.964608
5	18.3193	1.697214
6	70.4361	0.350464



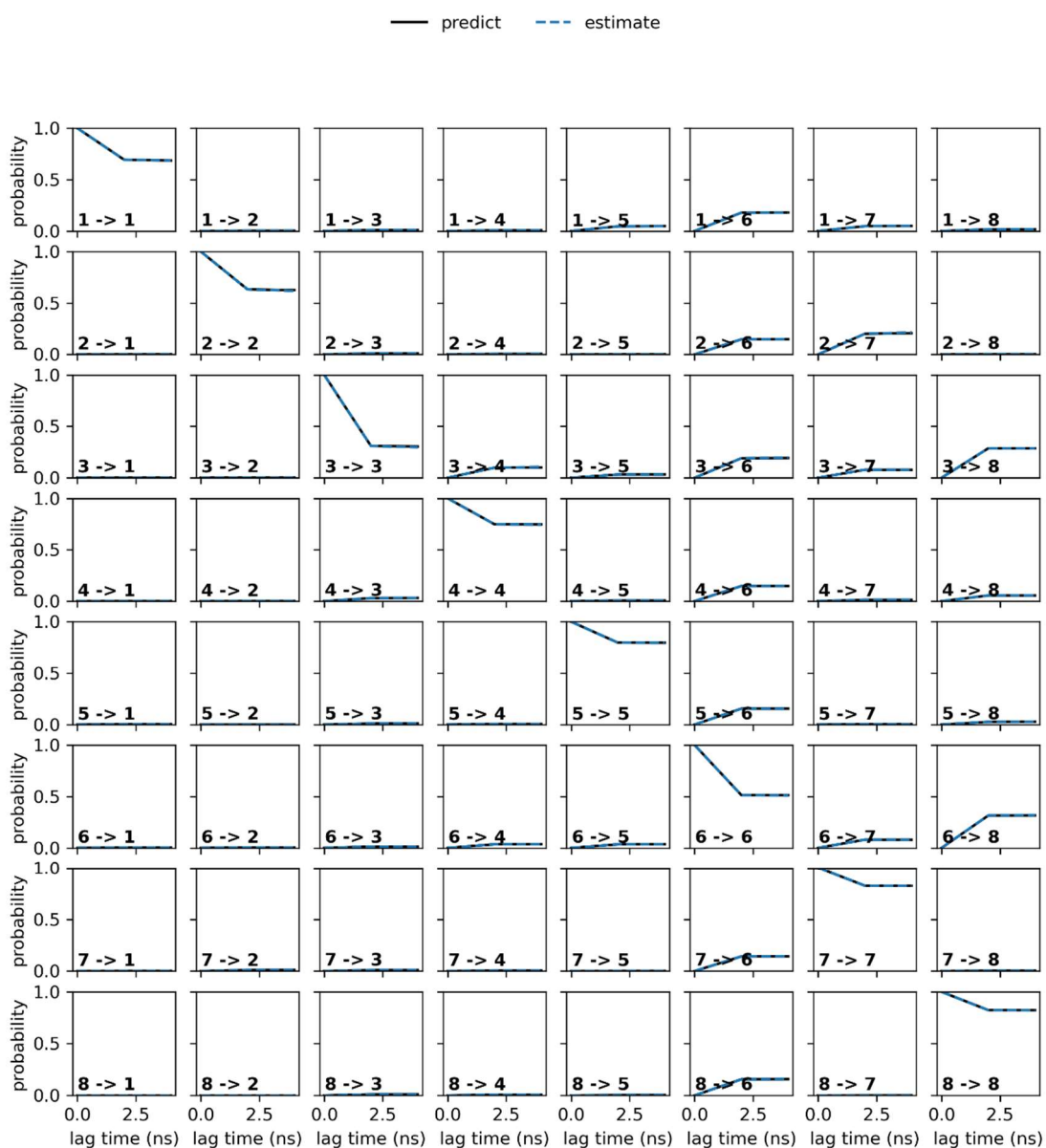
## L2d $\beta$ -lactamase results

In L2d  $\beta$ -lactamase MSM, the lag time was set at 2ns, the cluster number was 80, and 8 states were built. Figure 81 shows the implied timescale plot for L2d  $\beta$ -lactamase. The lines become flat and converged at 2ns lag time, therefore, 2ns was acceptable for the subsequent analysis.



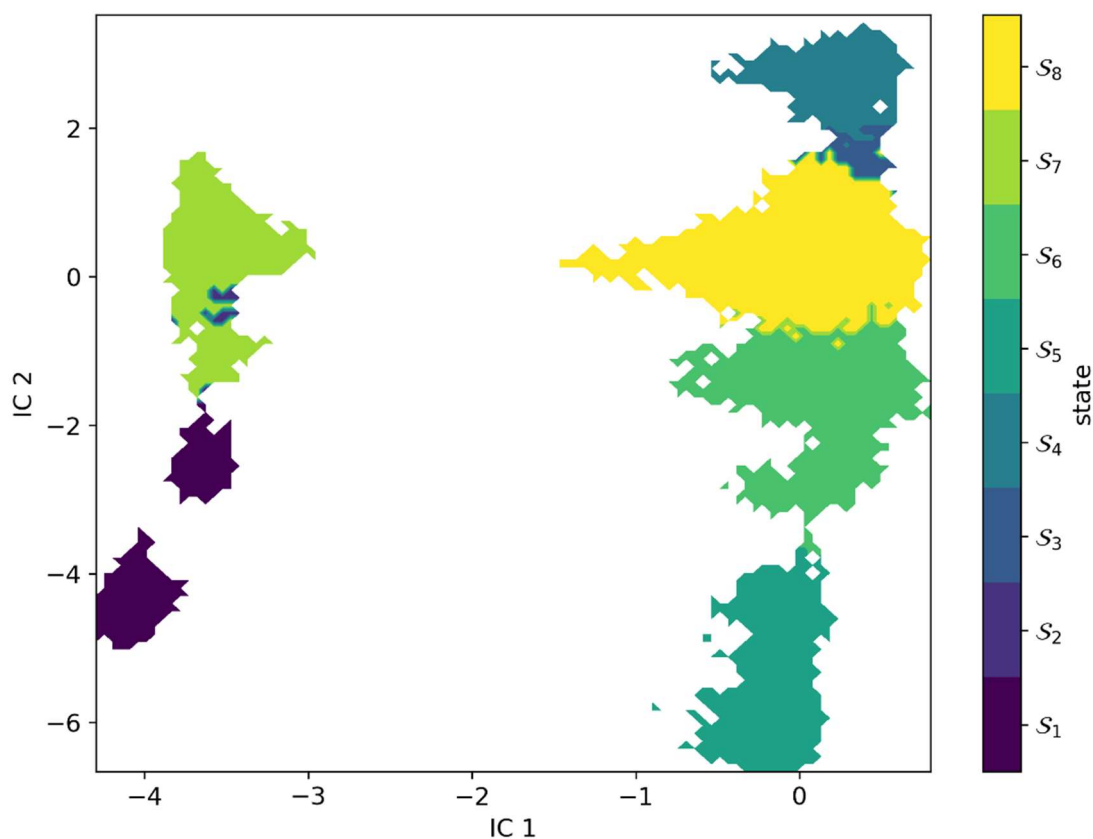
**Figure 81:** L2d  $\beta$ -lactamase implied timescale plot. Flat lines at 2ns indicates that this lag time was acceptable.

8 states were built for L2d  $\beta$ -lactamase MSM. The CK test was employed for the validation. The plot for L2d is shown in Figure 82. The prediction was given based on the model quantity and was compared with an independent estimation. The predict line and the estimation line superimpose with each other. Therefore 2ns is an acceptable lag time for subsequent analysis.



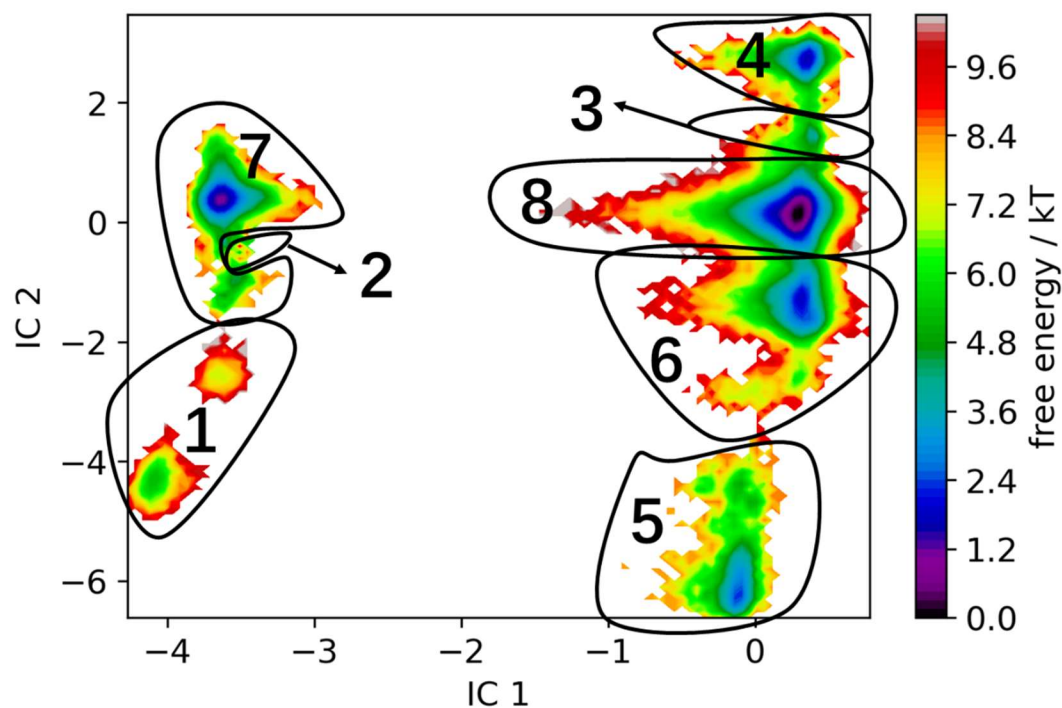
**Figure 82:** Chapman-Kolmogorov test plot for L2d  $\beta$ -lactamase. All 8 states were within confidence intervals and pass the CK test.

The PCCA distribution plot shows how 8 metastable states of L2d distributed during the simulation. The plot for L2d  $\beta$ -lactamase is shown in Figure 83. 8 metastable states were well separated. Higher states number was also tried to separate the state 1 and 6 but was not successful. Therefore, 8 states were chosen as the optimal state number.



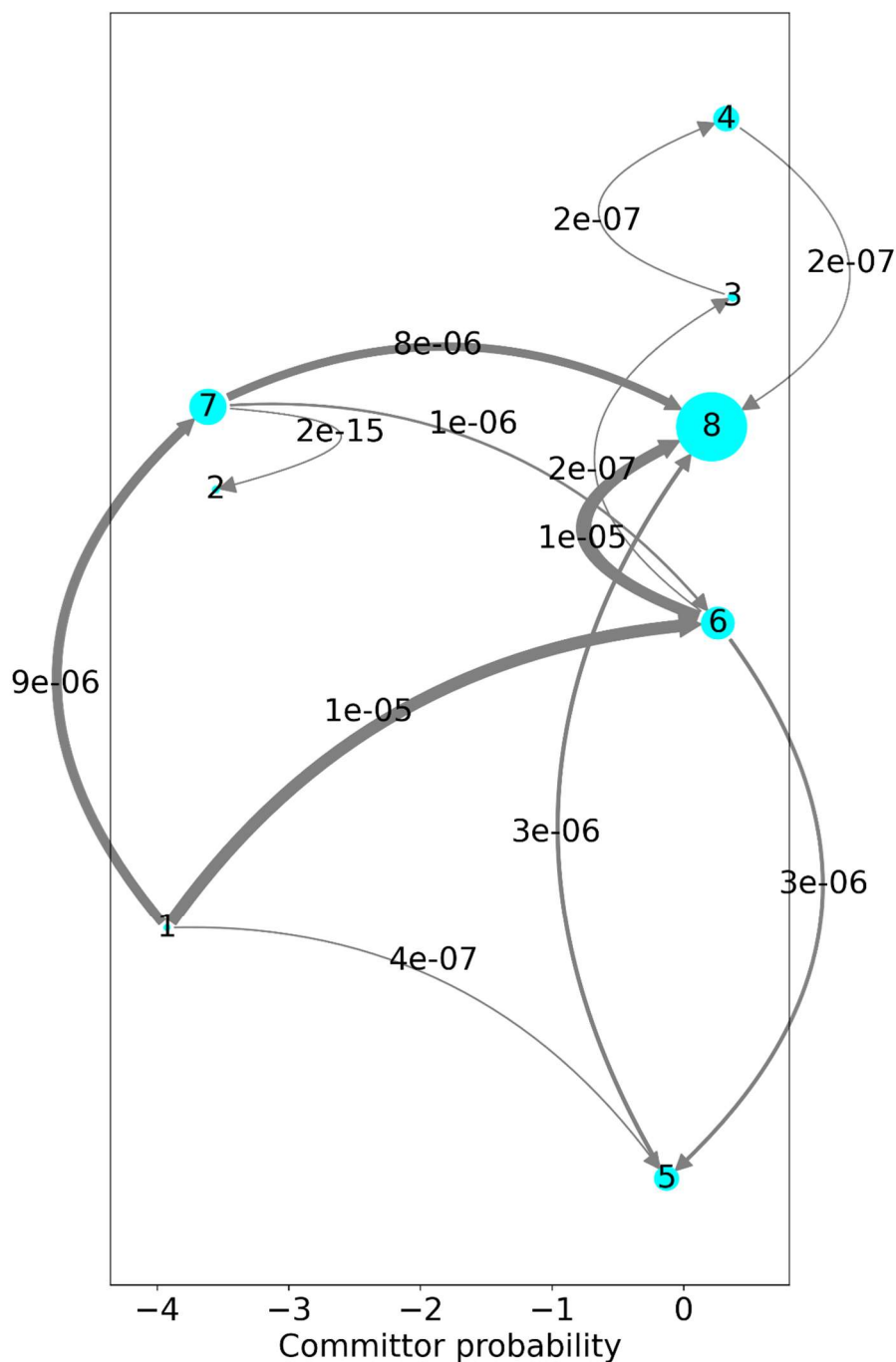
**Figure 83:** L2d  $\beta$ -lactamase PCCA distribution plot, 8 metastable states were well separated. Each different colour represents different state. The structures in the same states have similar features.

The Free energy distribution was shown in the free energy landscape. L2d free energy distribution at 2ns lag time is shown in Figure 84. There are 5 main minima and several sub minima observed. The high energy of the two regions of state 1 and 6 explain why it is difficult to separate these two states into more substate.



**Figure 84:** L2d  $\beta$ -lactamase free energy landscape at lag time 2ns. 5 main minima and several sub minima were observed. The minima correspond to the metastable states in PCCA distribution plot. State 1 has similar high energy on its edge. It can explain why it is hard to separate state 1 into the upper and lower parts. The state 6 is observed with similar high energy and difficult to be separated.

The netflux plot for L2d  $\beta$ -lactamase is shown in Figure 85. The dominant pathway in L2d  $\beta$ -lactamase conformational change are 1-6-8 (50.07%) and 1-7-8 (35.85%). Most structures will eventually turn into state 8.



**Figure 85:** L2d  $\beta$ -lactamase net flux pathway plot. Each cyan circle with certain number represented the correspond state. The bigger the circle, the more populated the state. Arrows between each state represented the possibility of each transition. The thicker the arrow, the higher the possibility.

Mean first passage times is an average timescale for a random event to first occur. It is an important parameter in MSM is because it implied the possibility of transimission between metastable states. The mean first passage time between different states in L2d  $\beta$ -lactamase is shown in Table 9. Table 10 is the population and free energy estimation for 8 states of L2d  $\beta$ -lactamase.

**Table 9:** Mean first passage times between different states in L2d  $\beta$ -lactamase.

State/(ns)	1	2	3	4	5	6	7	8
1	0.00	28019.86	19869.50	20681.75	9893.76	2195.96	13387.36	2483.54
2	35118.37	0.00	22051.00	22816.47	14768.75	4952.82	63.00	4489.08
3	45599.67	39065.57	0.00	3375.93	12599.54	1761.50	24459.14	1519.33
4	45878.28	39314.50	3182.34	0.00	12878.23	2113.70	24706.47	1728.19
5	41988.20	37895.81	18703.65	19510.50	0.00	1082.62	23286.00	1291.75
6	43996.03	37661.27	17279.88	18204.08	10996.39	0.00	23062.22	323.74
7	35075.27	14910.45	22043.40	22806.17	14756.66	4951.11	0.00	4471.19
8	44385.72	37704.84	17748.78	18464.98	11377.91	923.40	23090.34	0.00

**Table 10:** L2d  $\beta$ -lactamase 8 states population ( $\pi$ ) and free energy estimation.

State	$\pi$ (%)	G/kT (J)
1	0.3591	5.629193
2	0.4741	5.351449
3	0.8481	4.769877
4	7.1378	2.639763
5	6.6054	2.717279
6	12.3167	2.094213
7	15.1096	1.88984
8	57.149	0.559508

## 6.5 Dynamic Cross-correlation Analysis

Cross-correlation is a measurement that tracks the movement of two or more sets of relative data. In network analysis, cross-correlation was employed to find the correlation between different regions of the L2  $\beta$ -lactamase enzymes. These regions could be far away from each other, but the signal transfer during the simulation may influence the dynamics of each other. Bio3d software was used for residue correlation analysis.<sup>231</sup> The network analysis was based on different states which were built and described above in the MSM section. The analysis provides the correlation information of residues, both positive and negative. A positive correlation between two regions (or residues) is when one region becomes dynamic and influences the other region to also become dynamic. Their dynamic pattern is the same. A negative correlation between two regions (or residues) is when one region is dynamic while it influences the other region to become stable. Such a scenario is said to be invertly related.

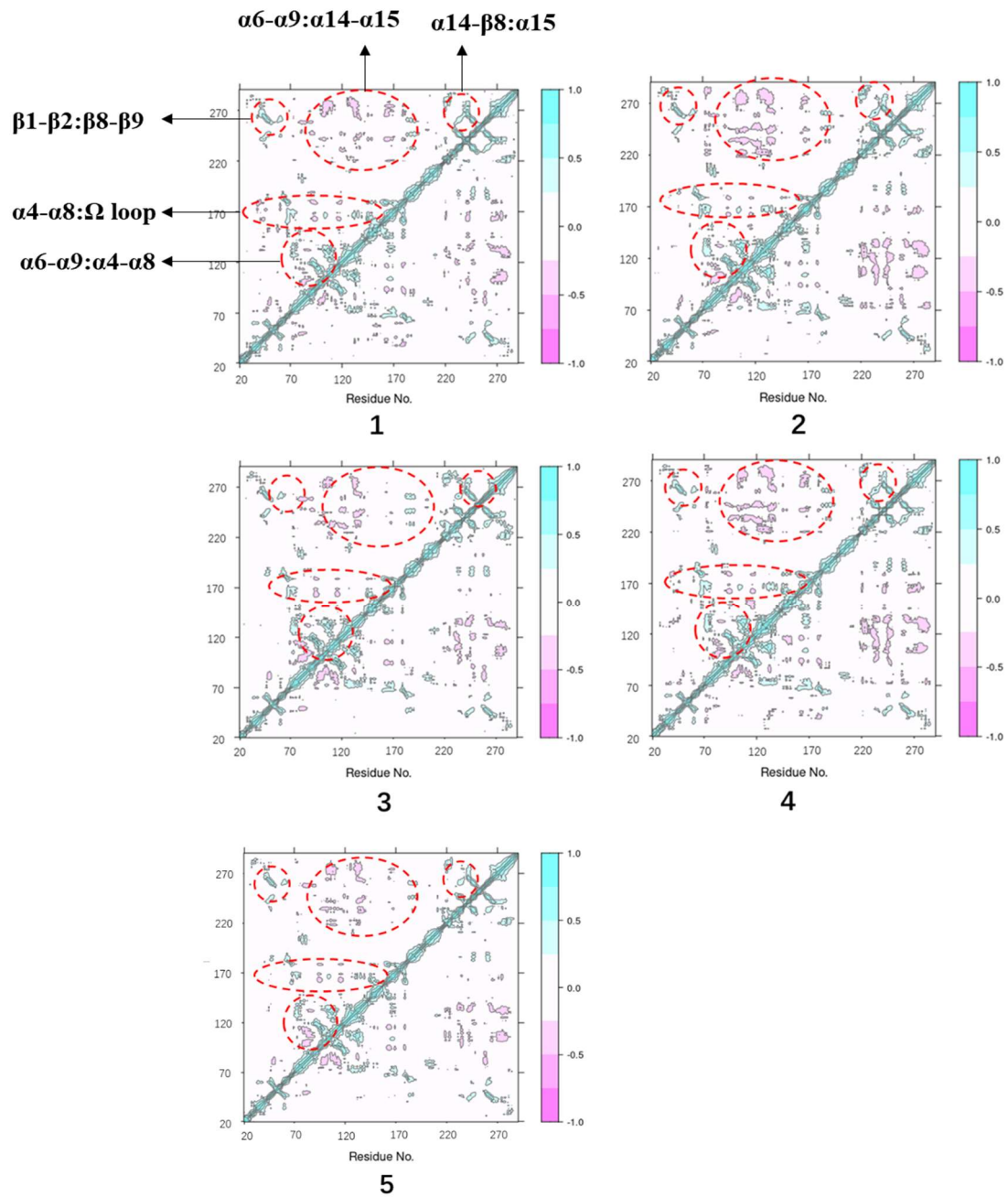
## L2a $\beta$ -lactamase results

In L2a  $\beta$ -lactamase, 5 states were built using MSM. Therefore, 5 cross-correlation plots were generated, each describing correlation between residues observed within each metastable state (Figure 86). The cyan area is the positive correlation and the pink area describes the negative correlation. The larger the area, the more residues involved in the correlation.

From the plots, the correlation occurs in three regions of the enzyme in  $\alpha$ 4- $\alpha$ 8,  $\Omega$  loop and  $\alpha$ 14- $\alpha$ 15. Between these three regions, the  $\alpha$ 4- $\alpha$ 8 correlation had an obvious expansion after state 1, with more residues being involved in the correlation. Most of the  $\alpha$ 4- $\alpha$ 8 correlations occurred within the two helices. The  $\Omega$  loop correlation had a slight increase in state 1 and 4 and remained the same in states 2,3 and 5. The  $\Omega$  loop correlation in state 1, 2 and 4 also occurred with  $\beta$ 7- $\beta$ 8 loop and  $\beta$ 9- $\alpha$ 15 loops.

The correlation between the two regions almost disappeared in the other 2 states. The  $\beta$ 1- $\beta$ 2 and  $\beta$ 8- $\beta$ 9 displayed a stable correlation and was observed in  $\beta$ 9. The dynamics correlated between  $\alpha$ 14- $\beta$ 8 and  $\alpha$ 6- $\alpha$ 9,  $\alpha$ 15. The  $\alpha$ 14- $\beta$ 8 and  $\alpha$ 6- $\alpha$ 9 area correlation had fewer residues involved at state 1,3 and 5. Especially in state 1 and 5, the correlation almost disappears. In contrast, the correlation involved much more residues in state 2 and 4. The correlation between  $\alpha$ 15 and  $\beta$ 4- $\alpha$ 9 was less dynamic. Compared to state 1, the correlation involved more residues in states 2 and 4 and involved fewer residues in state 3 and 5.

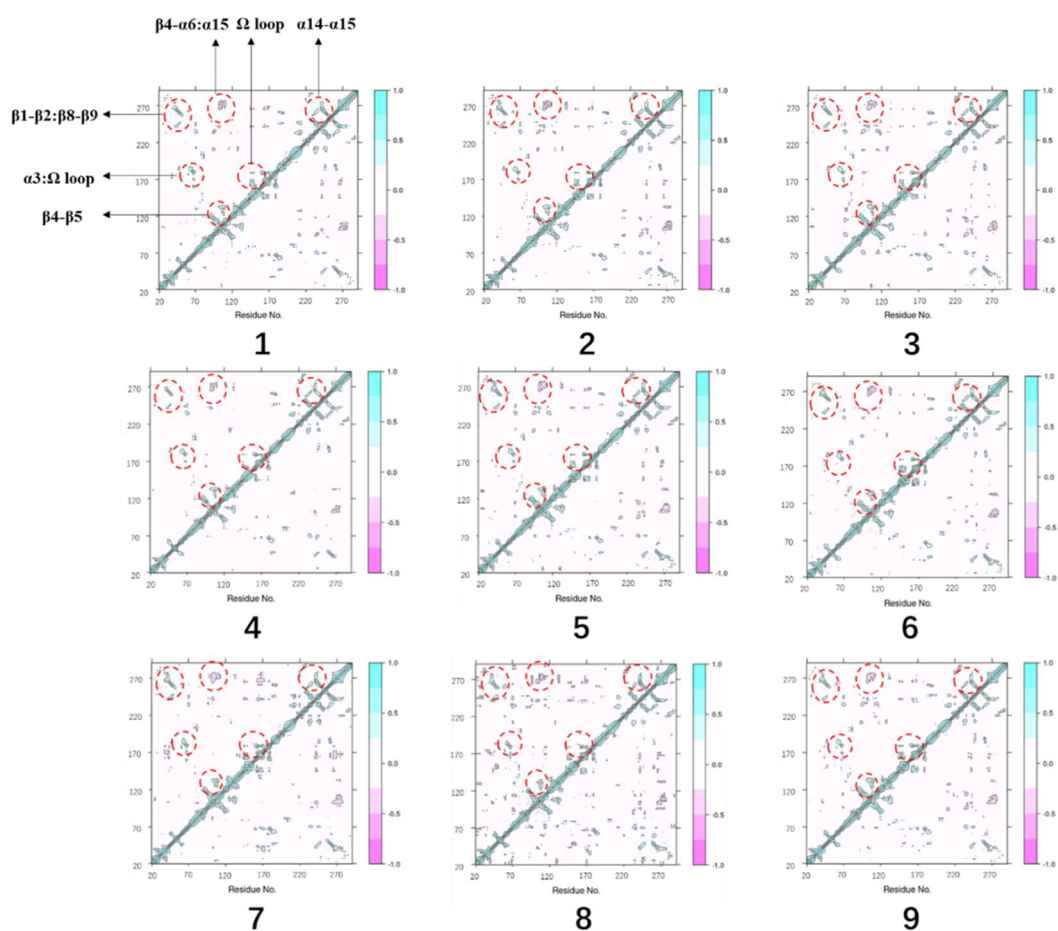




**Figure 86:** Cross-correlation analysis results for L2a  $\beta$ -lactamase. 5 plots were generated for 5 states built from the MSM. The cyan area shows positive correlation, the pink area shows negative correlation. The red circles indicate the regions involved in the cross-correlation.

## L2b $\beta$ -lactamase results

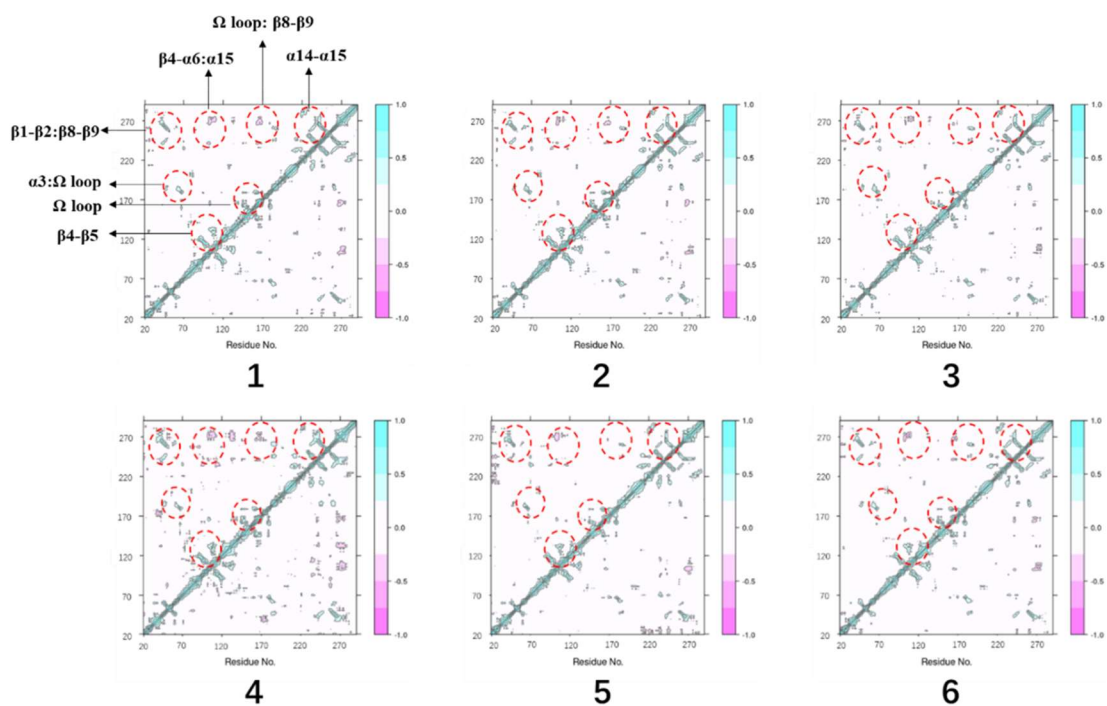
Compared with L2a  $\beta$ -lactamase, L2b  $\beta$ -lactamase displayed much less cross-correlation. 9 states were built in the MSM analysis. Therefore, 9 plots were generated (Figure 87), describing interactions within each metastable state.  $\beta$ 4- $\beta$ 5,  $\Omega$  loop and  $\alpha$ 14- $\alpha$ 15 had some minor correlation (red circle around the diagonal line). The correlation between  $\Omega$  loop and  $\alpha$ 3 remained consistent in all 9 states. In contrast,  $\Omega$  loop and  $\alpha$ 14- $\alpha$ 15 correlation was only observed in state 7 and almost disappeared in all other 8 states. 2 stable correlations were observed,  $\beta$ 8- $\beta$ 9 with  $\beta$ 1- $\beta$ 2 and  $\alpha$ 15 with  $\beta$ 4- $\alpha$ 6.



**Figure 87:** Cross-correlation analysis results for L2b  $\beta$ -lactamase. 9 plots were generated for 9 states built from the MSM. The cyan area shows positive correlation, the pink area indicates negative correlation. The red circles highlight the area of cross-correlation.

## L2c $\beta$ -lactamase results

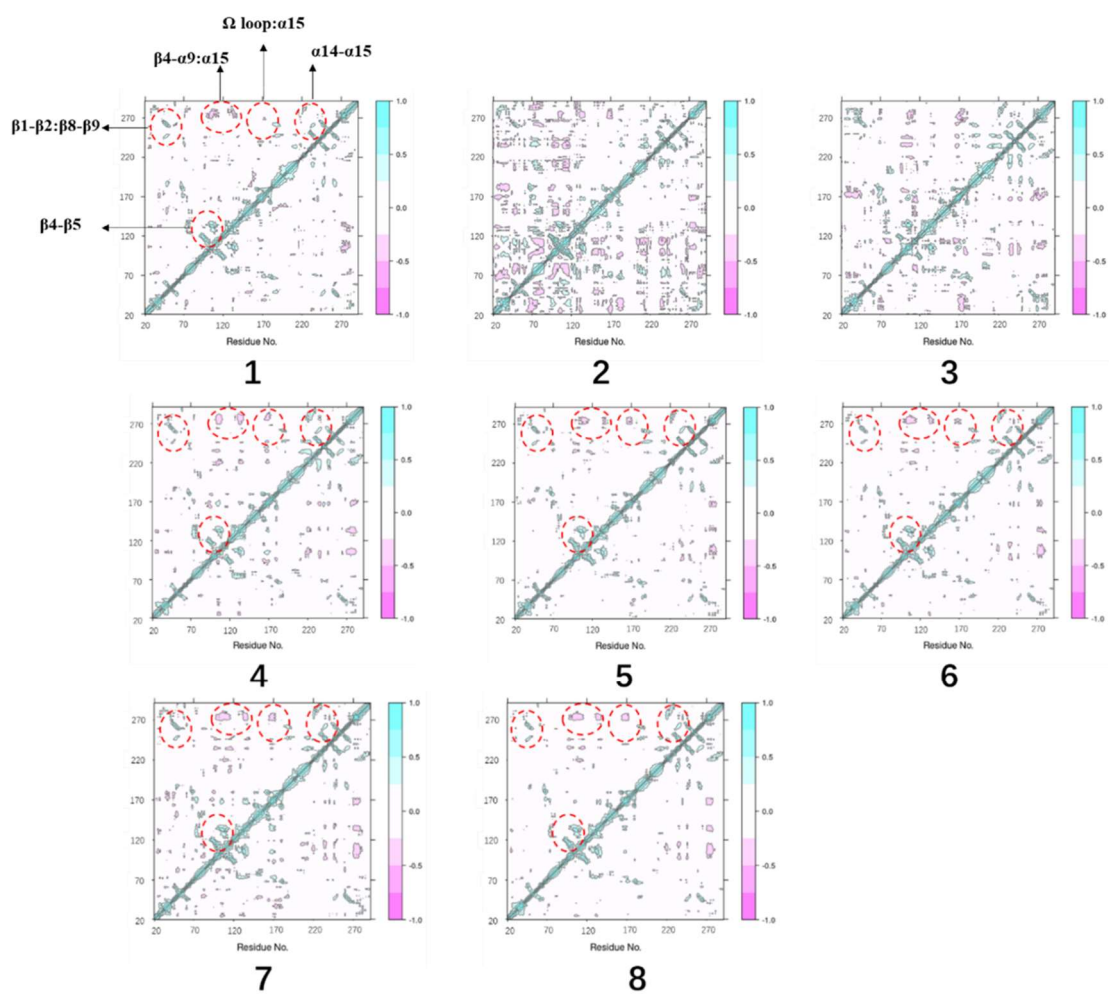
In L2c  $\beta$ -lactamase cross-correlation results, 6 plots were generated based on 6 states built in the MSM. Only a few correlations were observed (Figure 88).  $\beta$ 4- $\beta$ 5,  $\Omega$  loop and  $\alpha$ 14- $\alpha$ 15 had some inner correlation (red circle around the diagonal line). The correlation between  $\Omega$  loop and  $\alpha$ 3 involves several residues but exists in all 6 states.  $\Omega$  loop was also observed to have a correlation with  $\beta$ 8- $\beta$ 9 in state 1, 2 and 4. Only a few residues were involved in this correlation and it almost disappeared in state 3, 5 and 6. Besides, 2 stable correlations observed between  $\beta$ 8- $\beta$ 9 and  $\beta$ 1- $\beta$ 2 and between  $\alpha$ 15 and  $\beta$ 4- $\alpha$ 6. In state 5, some correlations were also observed between N terminal- $\alpha$ 1 and  $\alpha$ 14- $\alpha$ 15.



**Figure 88:** Cross-correlation analysis results for L2c  $\beta$ -lactamase. 6 plots were generated for 6 states built from the MSM. The cyan area shows positive correlation, the pink area indicates negative correlation. The red circles highlights the area of cross-correlation.

## L2d $\beta$ -lactamase results

In L2d  $\beta$ -lactamase, the cross-correlation analysis generated 8 plots for 8 states from the MSM (Figure 89). Among all 8 states, state 2 and 3 indicate highly dynamic results. The correlation in state 2 and 3 showed an obvious increase compared with the other 6 states. In state 1, 4, 5, 6, 7 and 8, several correlations were observed.  $\beta 4$ - $\beta 5$  and  $\alpha 14$ - $\alpha 15$  displayed some correlation (red circle around the diagonal line). There were several correlations observed in  $\beta 8$ - $\beta 9$ .  $\beta 8$ - $\beta 9$  with  $\beta 1$ - $\beta 2$ ,  $\alpha 15$  with  $\beta 4$ - $\alpha 9$ .  $\Omega$  loop also had a correlation with  $\alpha 15$  all states except state 1.



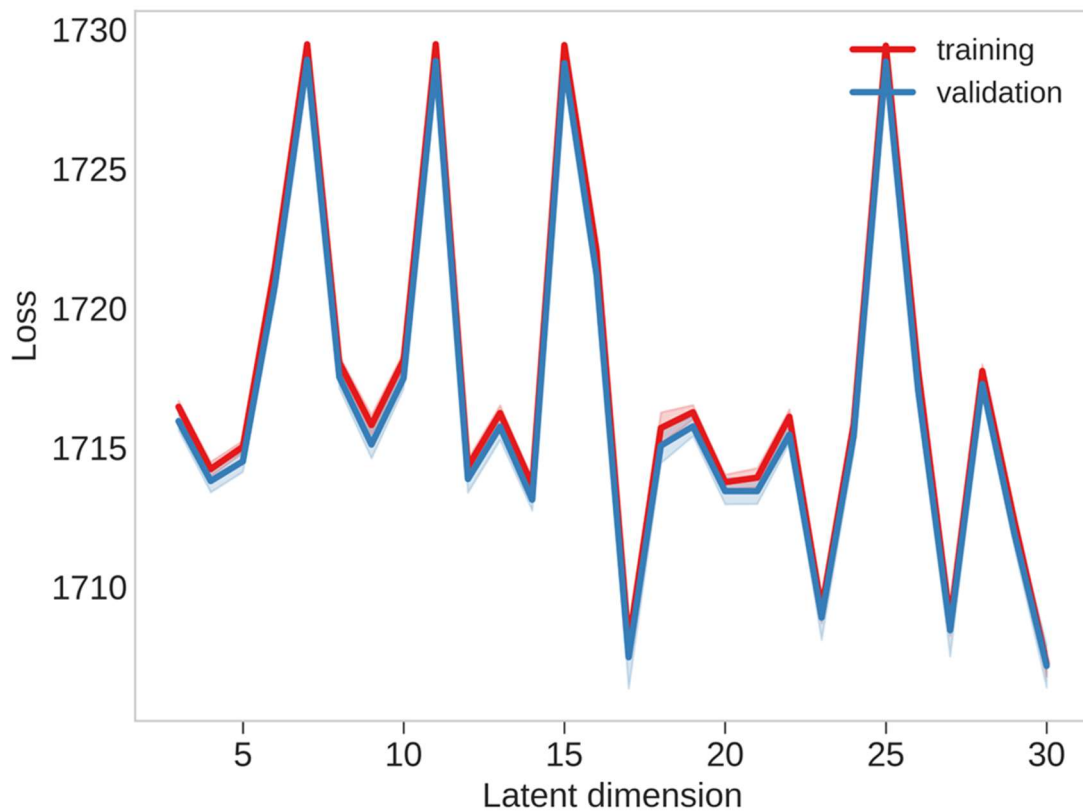
**Figure 89:** Cross-correlation analysis results for L2d  $\beta$ -lactamase. 8 plots were generated for 8 states from MSM. The cyan area represents positive correlation, the pink area corresponds to negative correlation. The red circles indicate the region of correlation.

## 6.6 CVAE-based Deep Learning of L2 $\beta$ -lactamases

CVAE-based deep learning was employed to analyse the trajectories of L2a, L2b, L2c and L2d  $\beta$ -lactamases. For running CVAE-based deep learning, 3 steps need to be employed. First was the pre-processing. In this step, RMSD, native contacts and distance matrices needed to be generated from the trajectories. Distance matrix does not contain the translational and rotational motion information.<sup>175</sup> The CVAE model was assessed by estimating the training and validation loss with the rising dimension numbers. RMSD, native contacts and distance matrices were compiled into a single system file which contains all information processed above. Following that, the four system files from L2a, L2b, L2c and L2d  $\beta$ -lactamases were combined into a single file system that contained a stack of the four individual systems.

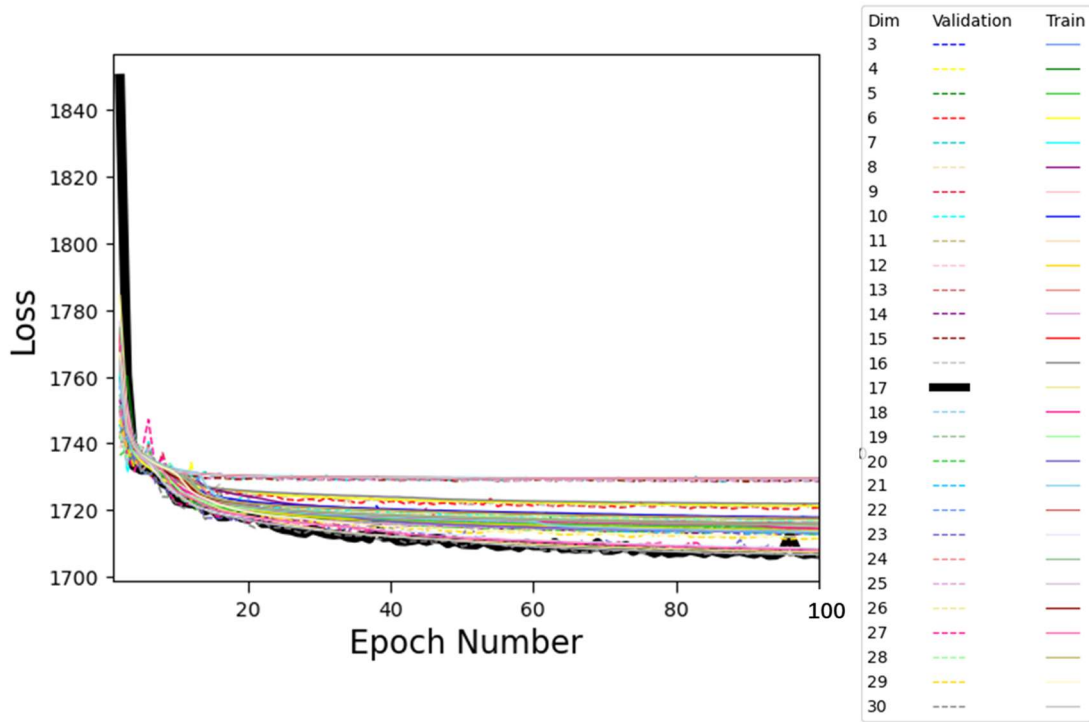
The second step was to run the CVAE and select an appropriate latent dimension number that displayed the lowest loss. The purpose of this step was to minimise information loss through the training epochs. In this experiment, CVAE latent dimensions were set at 3-30 (Figure 90) Each of the dimension ran 100 epochs to make sure the training and validation loss converged (Figure 91). After running CVAE, an optimal state needs to be selected. Therefore, the CVAE results were inspected by repeating the training process with different latent dimensions. The loss of latent dimensions is illustrated in Figure 90, where the red line represents the training loss and the blue line is the validation loss. The lowest loss was observed in dimension 17, where the mean validation loss was the minimum among all latent dimensions. The loss

of latent dimensions by epoch numbers is shown in Figure 91. The optimal selection is highlighted as a black line. Following this, the 17<sup>th</sup> dimension state was put into the reconstruction process and compared with the original data. A comparison of the encoded versus the decoded was done to ensure the decoded result had no significant loss of information (Figure 92).

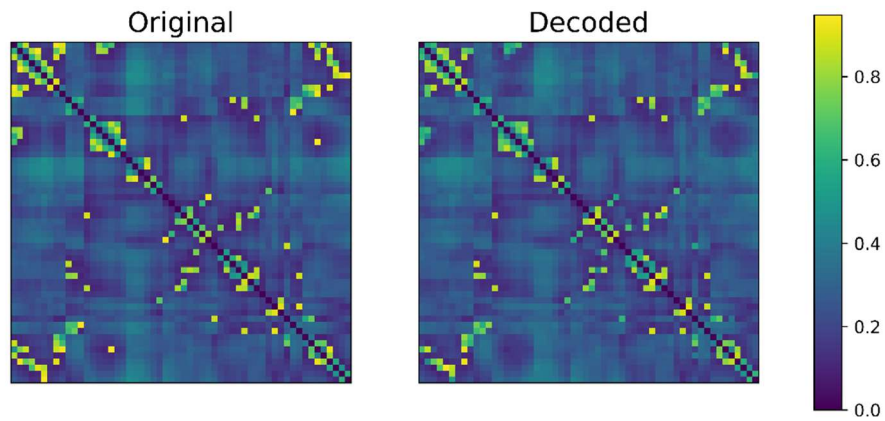


**Figure 90:** Evaluation of information loss over latent dimensions. Red line represents the training loss, the blue line is the validation loss. Latent dimension 17 show the best result with the lowest information loss. The comparison between original encoded data and reconstructed decoded data verified the compression of dimension17 did not result in a significant information loss and can be used in the next step.





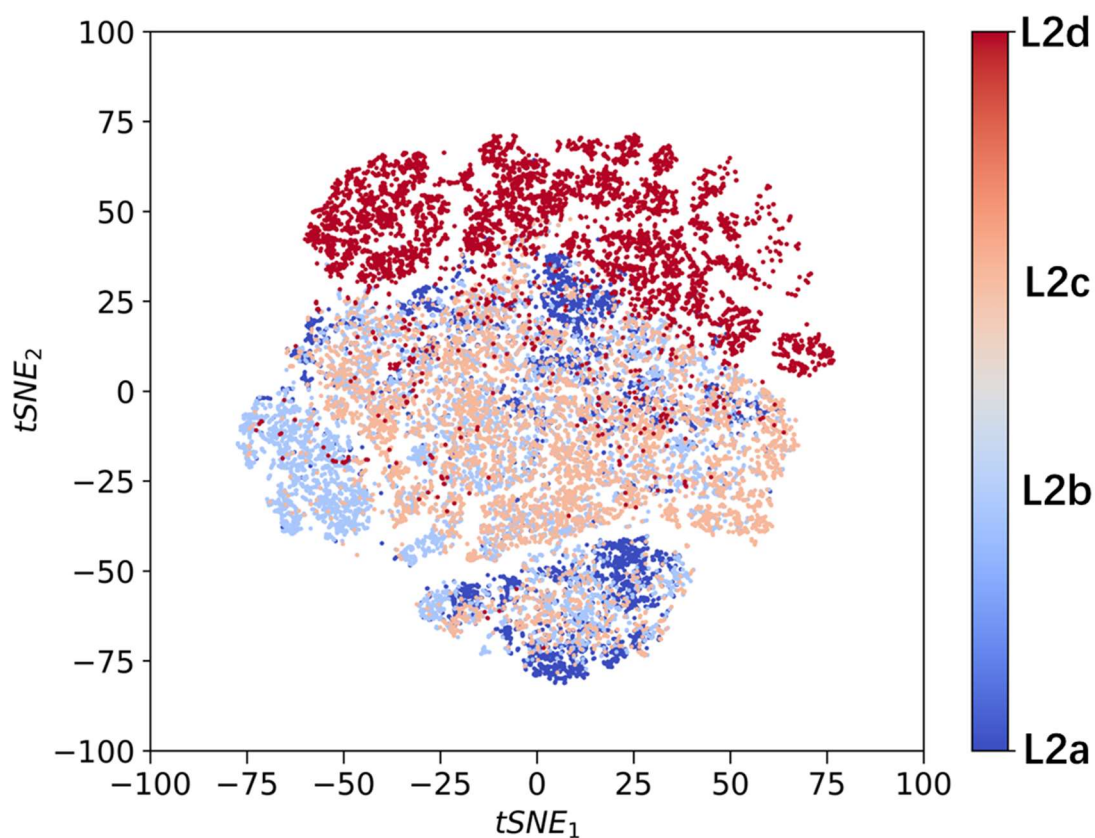
**Figure 91:** Evaluation of information loss of different latent dimensions by epoch number. The lowest loss dimension was highlighted with black colour.



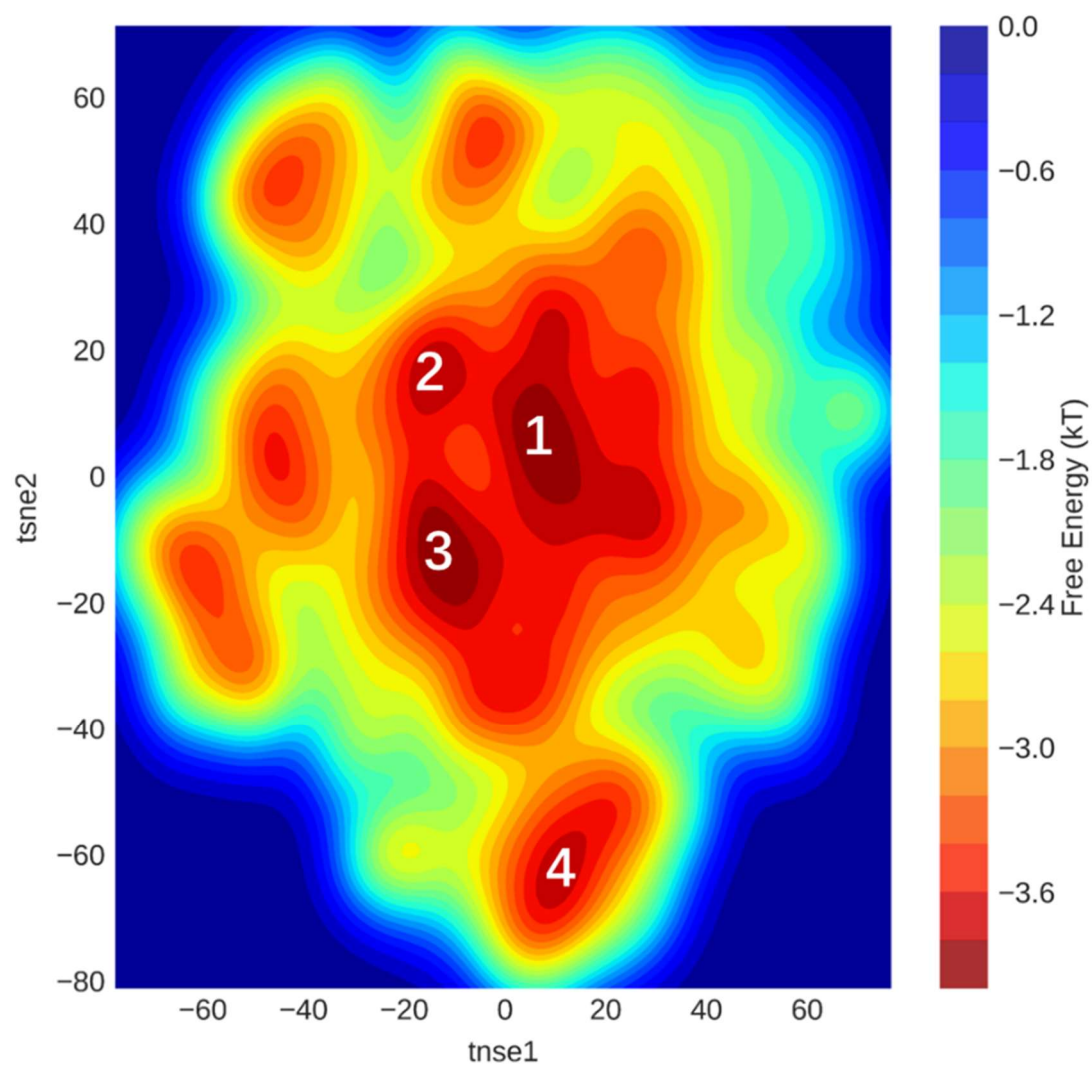
**Figure 92:** Comparison of the original input file and reconstructed decoded file, the compression of 17 dimensions did not result in a significant information loss.

After selecting 17th dimension as the optimal CVAE state, the last step was to visualise the CVAE results (Figure 93). In order to make the result easier to analyse and understand, the simulation results were first projected into a 3D latent space representation and then projected to a 2D plot for conformation extraction. Each conformation from the simulation was represented as a corresponding coordinate. In this case, t-SNE (t-distributed stochastic neighbour embedding) method was employed to reduce the data.<sup>196</sup> For easier understanding of the CVAE latent dimension, conformations were also projected onto 2D plot with the t-SNE method. Figure 93 shows the distribution of 17<sup>th</sup> latent dimension in a 2D plot. Different colours indicate different proteins from which the structure came from. Most of the L2d  $\beta$ -lactamase conformations clustered on the top, some L2b  $\beta$ -lactamase structures clustered on the left, others were mixed and distributed. The corresponding free energy plot was shown in Figure 94. There are several low energy regions. 4 lowest energy regions were selected and the structures of L2a, L2b, L2c, and L2d  $\beta$ -lactamase clustered in these regions were extracted for further analysis.



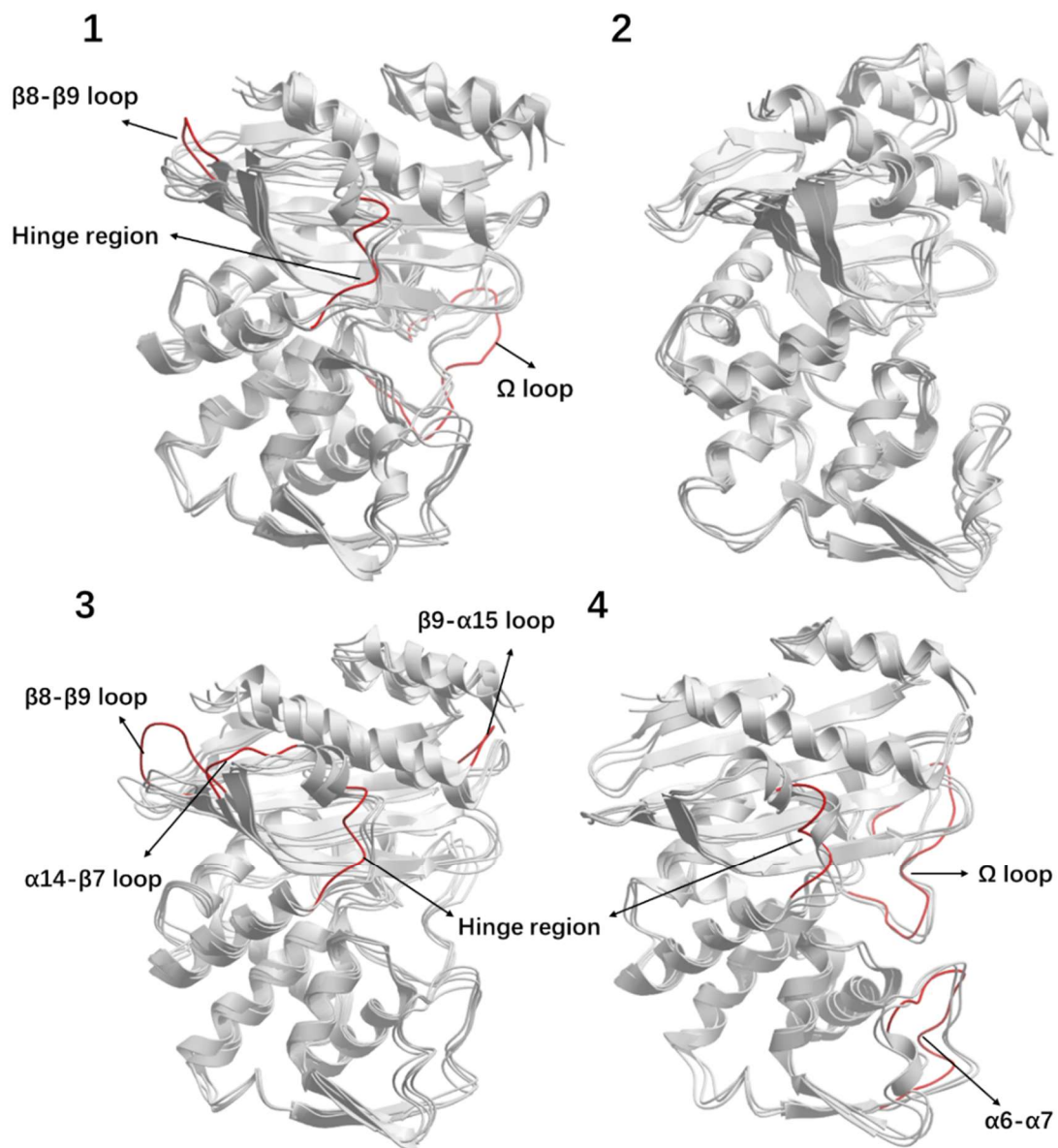


**Figure 93:** Conformational distribution in 2D plot clustered using the t-SNE method. Different protein conformations are shown in different colours, L2a (blue), L2b (light blue), L2c (orange), L2d (red). Most L2d  $\beta$ -lactamase conformations were clustered on the top of the plot, some L2b  $\beta$ -lactamase conformations were clustered on the left.



**Figure 94:** Free energy distribution plot. Four free energy minima were selected for conformation extraction.

In each low energy minima, four conformations from four enzymes with the most similar coordinates were selected for extraction. In region 1, 2 and 3, all four systems were observed. In region 4, only L2a, L2b and L2c  $\beta$ -lactamases conformations were observed and extracted. The conformations from the same regions were superimposed for comparison. Besides, the conformations from different low energy region but from same enzyme (such as four L2a extracted structures from four different regions) will also be superimposed and compared. The extracted conformations from the same regions were similar suggesting efficient clustering of the algorithm. The areas with enhanced dynamics and differences are highlighted with red colour (Figure 95). In structures extracted from region 1,  $\Omega$  loop, hinge region,  $\beta$ 8- $\beta$ 9 loop from L2a  $\beta$ -lactamase is different from other 3 proteins (highlighted with red in Figure 95-1). The loops on L2b, L2c and L2d were superimposed with each other while these loops on L2a  $\beta$ -lactamase is dissimilar from the other three enzymes' loops. The four structures from minimum 2 are similar. No significant difference was observed (Figure 95-2). In minimum 3, the hinge region and  $\beta$ 8- $\beta$ 9 loop of L2a  $\beta$ -lactamase,  $\alpha$ 14- $\beta$ 7 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop of L2d  $\beta$ -lactamase displayed some motion away from the others (highlighted with red in Figure 95-3).  $\alpha$ 6- $\alpha$ 7 of L2a  $\beta$ -lactamase and  $\Omega$  loop of L2b  $\beta$ -lactamase in minimum 4 show a slight difference (highlighted with red in Figure 95-4).



**Figure 95:** Superimposed structures extracted from different low energy minima, highlighting areas of dynamic differences (red).

Among the extracted structures, the conformations from the same protein were also investigated. This comparison gives more information about low energy structures during the simulation. The structures were similar but still showed some dynamic regions.

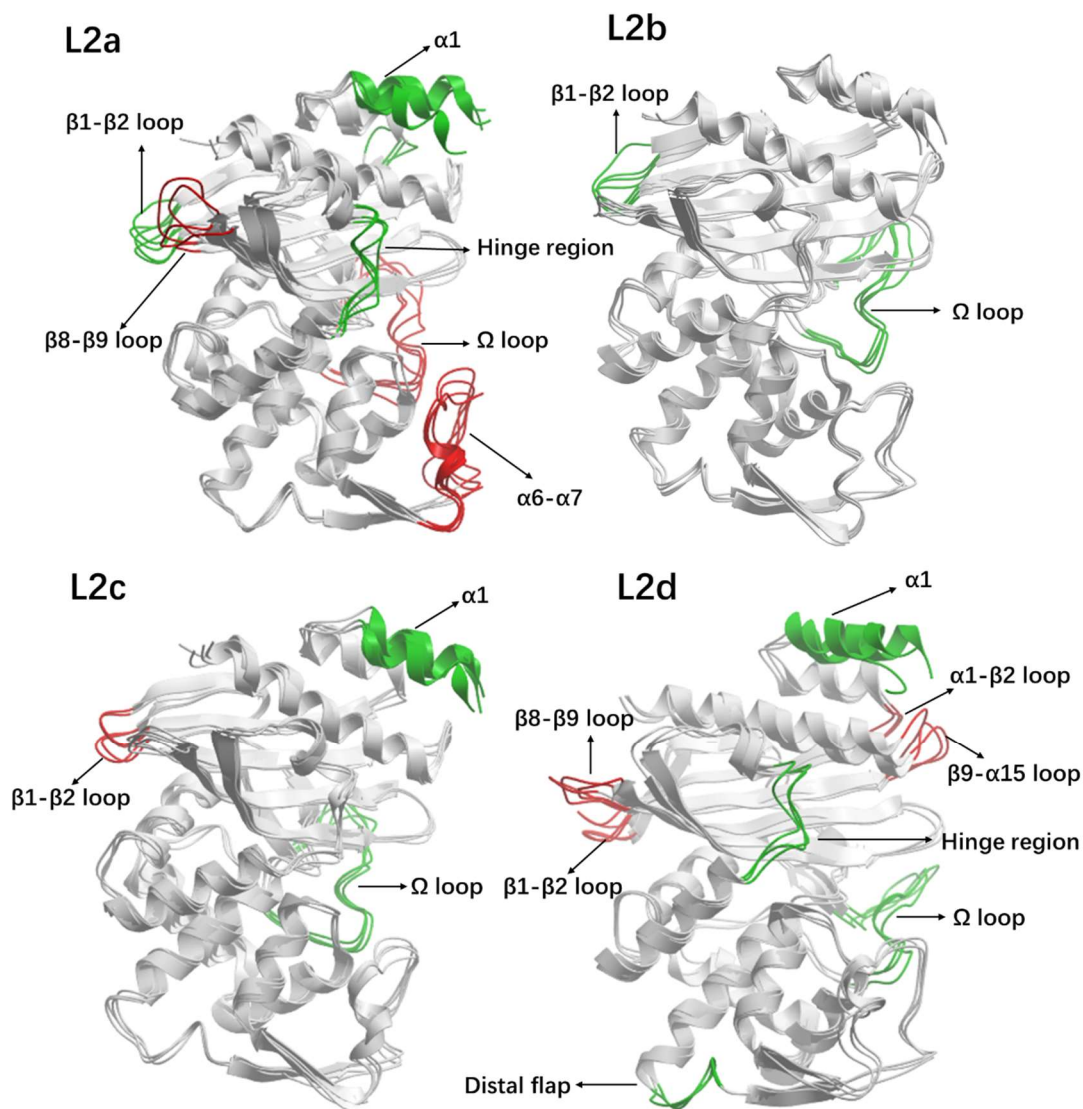
In L2a  $\beta$ -lactamase, the loops from different free energy regions show high dynamic motions. The loops that showed higher RMSD are  $\Omega$  loop,  $\beta 8$ - $\beta 9$  loop and  $\alpha 6$ - $\alpha 7$  (highlighted with red colour in Figure 96-L2a). Besides,  $\alpha 1$ ,  $\alpha 2$ - $\beta 1$ ,  $\beta 1$ - $\beta 2$  and hinge region also displayed high dynamics (highlighted with green colour in Figure 96-L2a).

The L2b  $\beta$ -lactamase structures were identical. Only  $\beta 1$ - $\beta 2$  loop and  $\Omega$  loop observed a conformational change in L2b (highlighted with green colour in Figure 96-L2b).

In L2c  $\beta$ -lactamase,  $\beta 1$ - $\beta 2$  loop from four structures extracted from different regions showed large differences. Other than superimposed with each other, four  $\beta 1$ - $\beta 2$  loops from four energy minima was divided into two groups (highlighted with red colour in Figure 96-L2c). The  $\alpha 1$  and  $\Omega$  loop showed conformational changes (highlighted with green colour in Figure 96-L2c).

For L2d  $\beta$ -lactamase, only three structures were extracted. There was no L2d  $\beta$  lactamases in minimum 4. The loops of these three structures had displayed motions away from each other.  $\alpha 1$ - $\beta 2$  loop,  $\beta 1$ - $\beta 2$  loop,  $\beta 8$ - $\beta 9$  loop and  $\beta 9$ - $\alpha 15$  loop had obvious

structural changes (highlighted with red colour in Figure 96-L2d). The  $\alpha 1$ , distal flap, hinge region and  $\Omega$  loop also had small structural differences.



**Figure 96:** Comparison of extracted structures from the same protein. The extracted structures from the same protein were aligned. The conformational changes are highlighted in red and green colours.

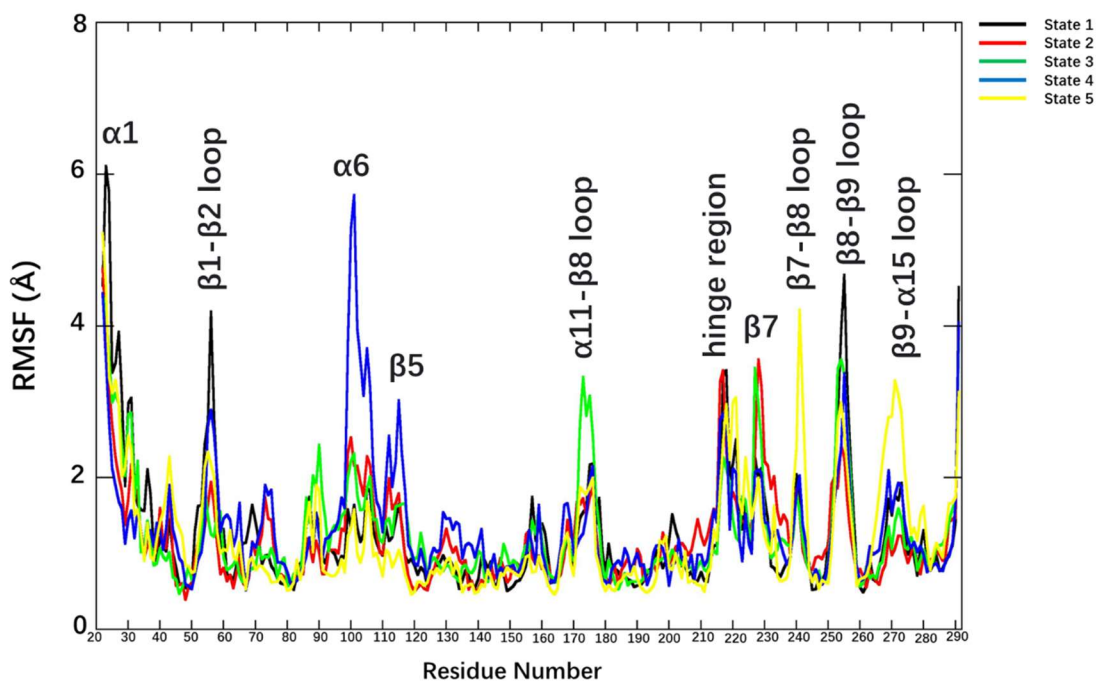
## 7. L2 $\beta$ -lactamases Discussion

L2 has few research on it, most of current L2 research focus on the AmpR regulator function.<sup>232–235</sup> L2  $\beta$ -lactamase has been proved that it is not AmpR dependent.<sup>92</sup> Some other publication focus on L2 inhibitors investigation, such as relebactam, clavulanate and avibactam.<sup>236,237</sup> There is no L2  $\beta$ -lactamase inhibitor has been found yet. On the other hand, as a class A serine  $\beta$ -lactamase, the mechanism of other class A serine  $\beta$ -lactamase has been widely explored. The key residue such as S70, K73, S130, N132, E166, T235 and T237 have already been revealed.<sup>52</sup> The important regions for catalytic regions are also indicated, such as hinge region,  $\Omega$  loop and hydrophobic nodes.<sup>229,238–241</sup> Furthermore, a number of class A  $\beta$ -lactamase inhibitors have also been found, such as clavulanate, sulbactam and tazobactam.<sup>242–244</sup> The widely research of class A  $\beta$ -lactamase allows L2 have a large database to refer.

As mentioned above, L2  $\beta$ -lactamase has an extra  $\alpha$  helix at the beginning of the sequence. The structural difference will result in potential different mechanisms. Since there is no dynamic research has been done on L2  $\beta$ -lactamase before, the understanding of L2  $\beta$ -lactamase dynamic pattern is important. In this work, four L2  $\beta$ -lactamase proteins were analysed. Therefore, they will be discussed first separately and then as a combined discussion.

## L2a $\beta$ -lactamase discussion

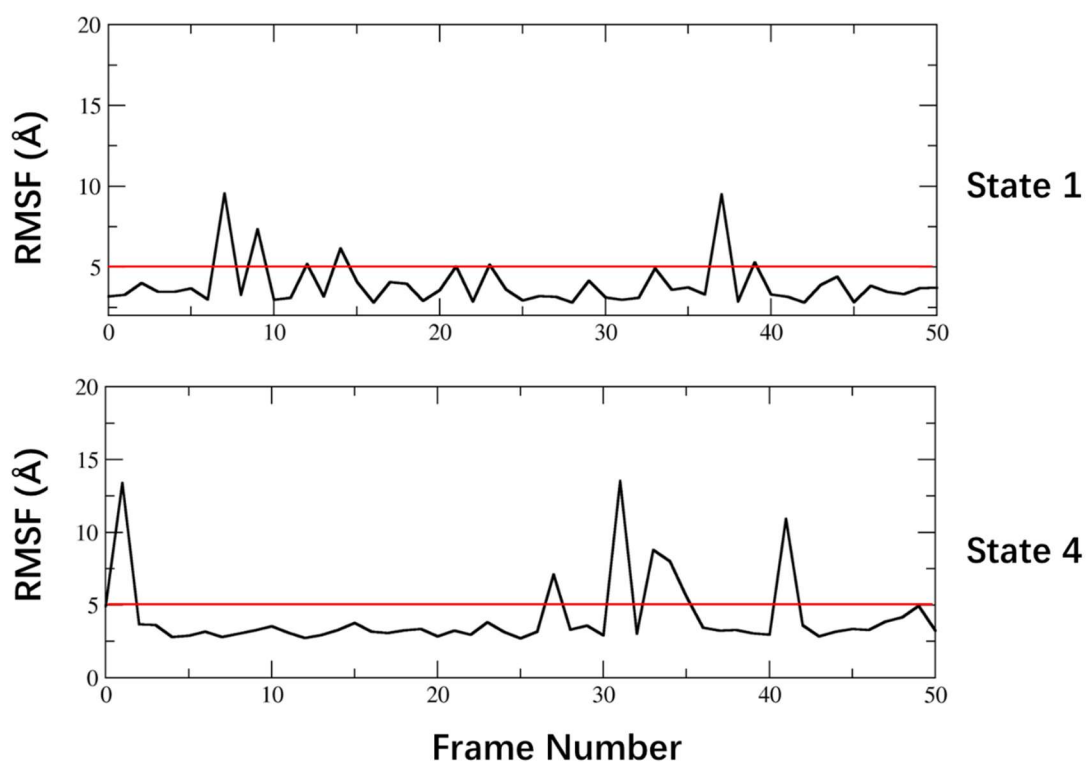
In L2a  $\beta$ -lactamase, 25% of C $\alpha$  RMSD was more than 0.9Å (Figure 58) which was higher when compared with the other 3 proteins. These included the  $\alpha$ 1,  $\beta$ 1- $\beta$ 2 loop,  $\alpha$ 6, hinge region,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop. The RMSF was calculated from each of the 5 metastable states (Figure 97). The comparison between the different metastable states from the same low energy minima gives a better indication of the dynamics observed in these minima.



**Figure 97:** RMSF plot for 5 states in L2a  $\beta$ -lactamase. Different colours represent different states in L2a  $\beta$ -lactamase.



All 5 metastable states show high dynamics in the complete  $\alpha 1$  helix. Some residues from  $\alpha 1$  interacted with the  $\beta 9$ - $\alpha 15$  loop. The interactions frequently disappeared during the course of the simulation. This makes this region highly dynamic. In this case, the highest RMSF for  $\alpha 1$  (state 1) was compared with the lowest RMSF state (state 4), to identify different interactions between these states. Among all the interaction between  $\alpha 1$  and  $\beta 9$ - $\alpha 15$  loop, S21(sidechain)-Q273(sidechain) showed the biggest difference. In state 1 the S21-Q273 interaction changes between formation and loss. In state 4, the interaction is stable and is seldomly lost. Besides, when interaction is broken, the two residues drift far from each other. The distance between residue S21 and residue Q73 had been calculated in Figure 98. The cut off of interaction length is set at 5Å. State 1 had 50% more changes between formation and loss of this interaction. Besides, when S21 get closer to Q273, other residues from  $\alpha 1$  also tend to form interactions with the  $\beta 9$ - $\alpha 15$  loop, such as T24 and D25. The less frequent changes between interactions may help stabilise structures in state 4. The residue dynamic cross-correlation plot (Figure 86) shows no correlation between these two regions in all 5 states. However,  $\alpha 1$  may indirectly stabilize  $\beta 9$ - $\alpha 15$ .

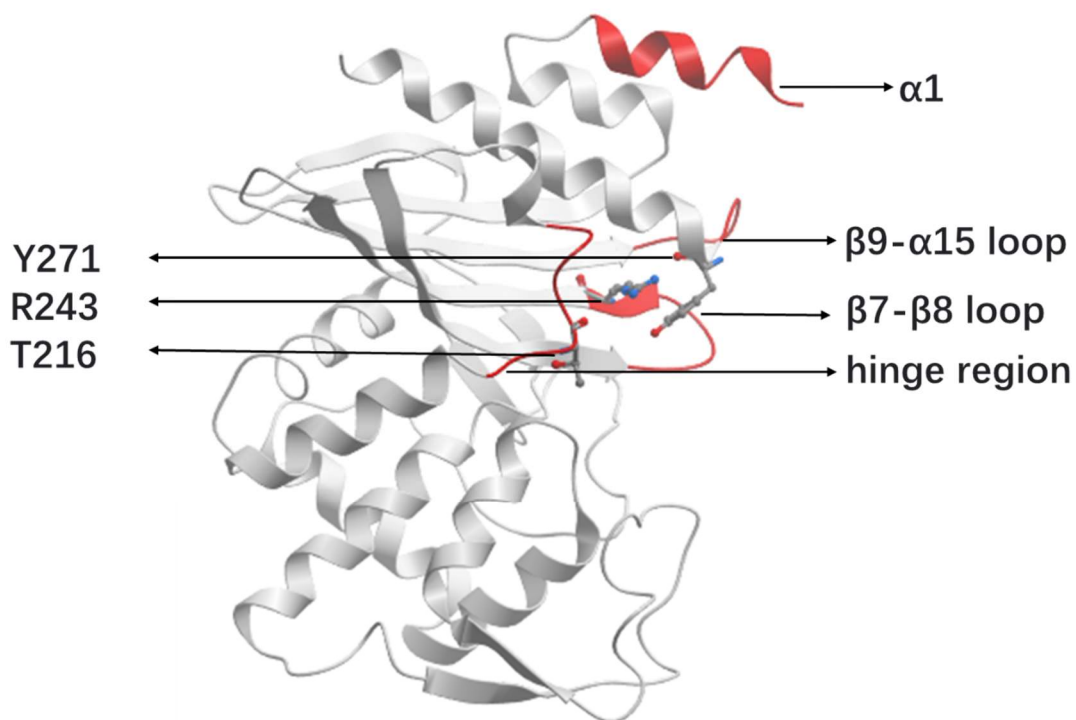


**Figure 98:** Distance between S21 and Q273. The cut off set at 5 Å (red line). The two states highlight the formation and loss of bond. The bond is lost 18% in state 1 and 12% in state 4.

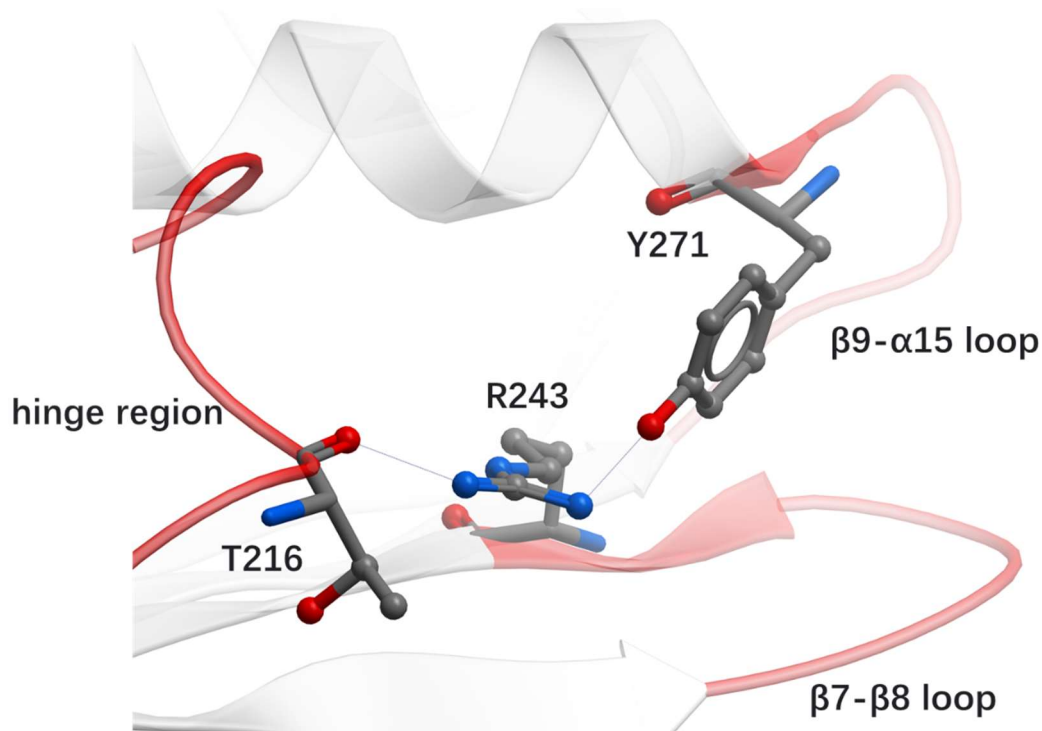
The  $\beta 9$ - $\alpha 15$  loop dynamic was not just influenced by  $\alpha 1$  but is also influenced by  $\beta 7$ - $\beta 8$  loop (Figure 99). State 5 displayed highest RMSF in the  $\beta 7$ - $\beta 8$  loop (Figure 97) while state 4 was the lowest. Comparing the dynamics in these two states, it was found that the long side chain of R243 (from  $\beta 8$  sheet, very close to  $\beta 7$ - $\beta 8$  loop) was responsible for the  $\beta 7$ - $\beta 8$  loop dynamics. R243 can attract and form hydrogen bonds with both the side chain of T216 (from hinge region) and the backbone of Y271 (from  $\beta 9$ - $\alpha 15$  loop). Y271 and T216 are on the opposite side of R243. If R243 get closer to each one residue, it will be far away from the other one (Figure 100). It should be emphasized that the bond formation and loss are not directly correlated but can be further influenced by the neighboring regions. When the R243 get attracted by Y271,

the  $\beta 7$ - $\beta 8$  loop and  $\beta 9$ - $\alpha 15$  loop get closer to each other. When R243 forms a bond with Y271, it drifts far away from T216. It is possible that R243 can form bonds with T216 and Y271 at the same time. The dynamics of  $\beta 7$ - $\beta 8$  loop,  $\beta 9$ - $\alpha 15$  loop and hinge region working together. Changes in each region can influence the dynamics and interactions the other two regions.

In state 5, R243 interacts with Y271. The high dynamics of  $\beta 9$ - $\alpha 15$  loop influenced  $\beta 7$ - $\beta 8$  loop and resulted in a relatively high RMSF in state 5 than in any other 4 states. In contrast, R243 in state 4 tended to stay in the middle or get closer to T271, while the bond between R243 and the backbone of Y271 was not broken. The interaction in two opposite direction makes the  $\alpha 1$  interactions with  $\beta 9$ - $\alpha 15$  loop and  $\beta 7$ - $\beta 7$  loops more stable. This can also explain the hinge region dynamics observed in all metastable states. R243 are attracted by both T216 and Y271, the bond forming and losing result in the dynamics in this region. When R243 was far away from T216 and stay with Y271 in state 5, the RMSF value of this region also showed a small amount of decrease.

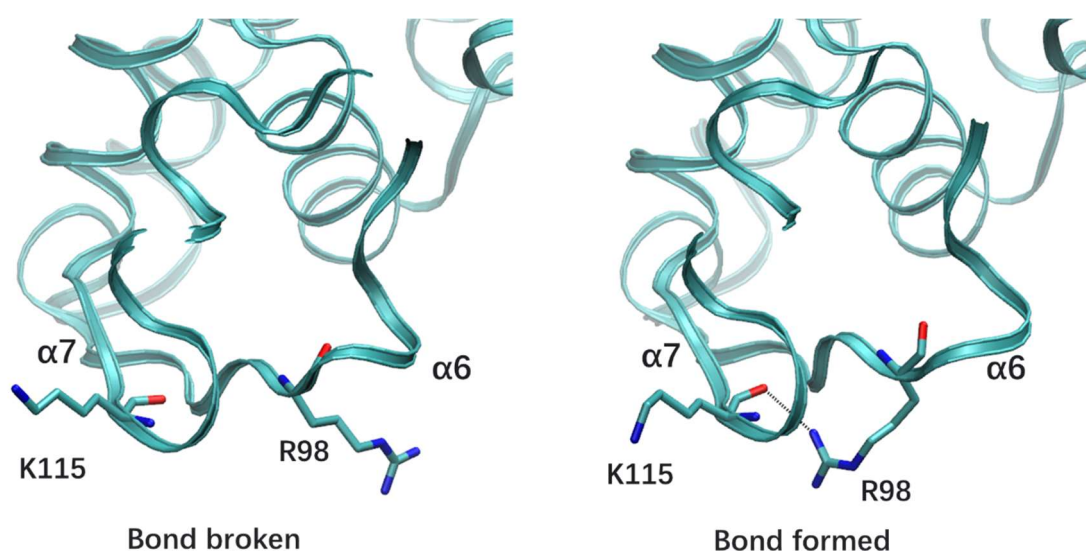


**Figure 99:** The position of  $\alpha 1$ ,  $\beta 7$ - $\beta 8$  loop,  $\beta 9$ - $\alpha 15$  loop, T216, R243 and Y271.  $\beta 9$ - $\alpha 15$  loop was in the middle of  $\alpha 1$  and  $\beta 7$ - $\beta 8$  loop, when it was in a highly dynamic state, it could be influenced by  $\alpha 1$  and  $\beta 7$ - $\beta 8$  loop. R243 has a long side chain which can form bonds with both T216 and Y271.



**Figure 100:** A closeup view of T216, R243 and Y271 position. R243 can form bonds with both T216 and Y271. When R243 get closer to one of the two residues, it will get far away from the other one.

From the DCCM plot (Figure 86), we observe the  $\beta 9$ - $\alpha 15$  loop had negative correlation with  $\alpha 6$  and  $\alpha 9$ . The stabilization of  $\beta 9$ - $\alpha 15$  loop in metastable state 4 may increase the dynamics of these two areas. Especially in  $\alpha 6$ , two RMSF peak was observed in  $\alpha 6$  and adjacent  $\alpha 7$ . It was found that  $\alpha 6$  tended to stay in a wider and more relaxed orientation in state 4. In this relaxed state, the interactions between residues within the  $\alpha 6$  loop decreased and these allowed interactions forming between  $\alpha 6$  and  $\alpha 7$  easier, especially between the side chain of R98 ( $\alpha 6$ ) and the backbone K115 ( $\alpha 7$ ) (Figure 101).

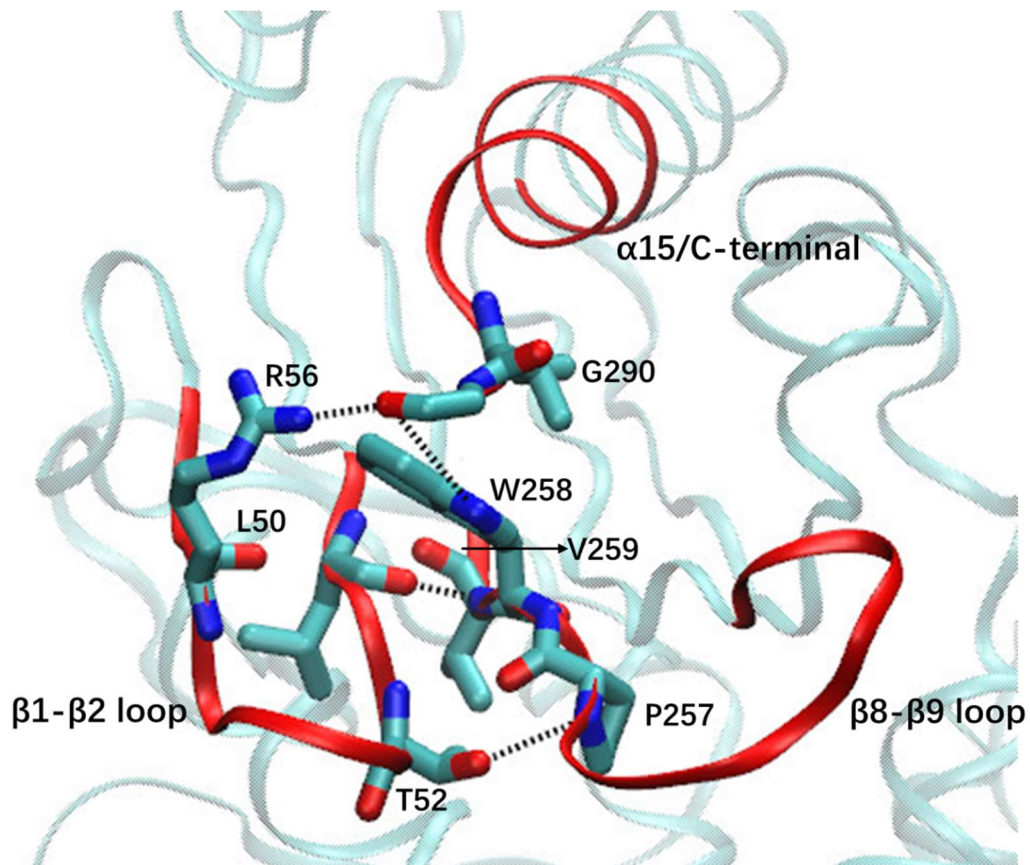


**Figure 101:** The hydrogen bond formation between the side chain of R98 ( $\alpha 6$ ) and the backbone of K115 ( $\alpha 7$ ). In order to form this interaction, a large rotation in the side chain of R98 was observed.

Furthermore, large dynamics was also observed in  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loops.  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loop are positioned parallel to one another. There were several stable interactions between them, such as L50 (bb)-V259 (bb) and T52 (sc)-P257(bb). Since there are no other loops close to  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loop, the dynamics of these loops

tend to influence each other. Comparing the high RMSF values of the  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loops, it might also be that the large dynamics could be further influenced by the long end of the C-terminal. When the C terminal gets close to the  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop, the C-terminal can make interactions frequently, such as R56 (sc)-G290(bb) ( $\beta$ 1- $\beta$ 2 loop) and W258-G290 ( $\beta$ 8- $\beta$ 9 loop). It is interesting to note that these two bonds could form at the same time (Figure 102). Therefore, the high dynamics of  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop could be a direct consequence of the flapping of the C-terminal frequently in the vicinity of  $\beta$ 1- $\beta$ 2 loop or  $\beta$ 8- $\beta$ 9 loop.

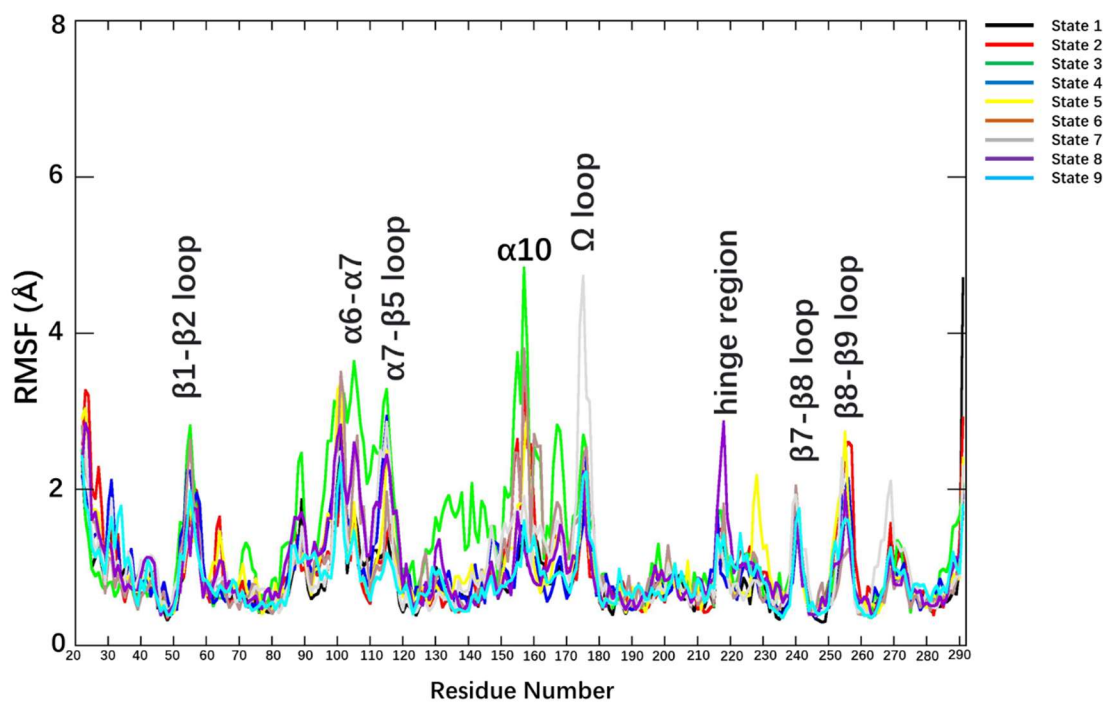
In L2a  $\beta$ -lactamase simulation, most structures will stay in metastable state 4 and 5 (Table 4). However, some regions show a quite opposite dynamic pattern in these two states ( $\alpha$ 1,  $\alpha$ 6,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop). In state 4  $\alpha$ 1,  $\beta$ 7- $\beta$ 8 loop and  $\beta$ 8- $\beta$ 9 loop are stable while these regions in state 5 are dynamic.  $\alpha$ 6 helix in state 4 is quite dynamic which is stable in state 5. Based on the DCCM plot, most negative correlation was also observed in these regions, which implied the L2a  $\beta$ -lactamase protein could have two patterns of dynamics.<sup>245</sup>



**Figure 102:** C-terminal affect  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop dynamics. C terminal got close to  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop.  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop had stable bonds forming between each other (such as L50-V259 and T52-P257). When the C terminal comes close to these two loops, it can form several bonds (such as R56-G290 and W258-G290). The high dynamic and unordered rotation of the C terminal will influence the dynamic of  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop.

## L2b $\beta$ -lactamase discussion

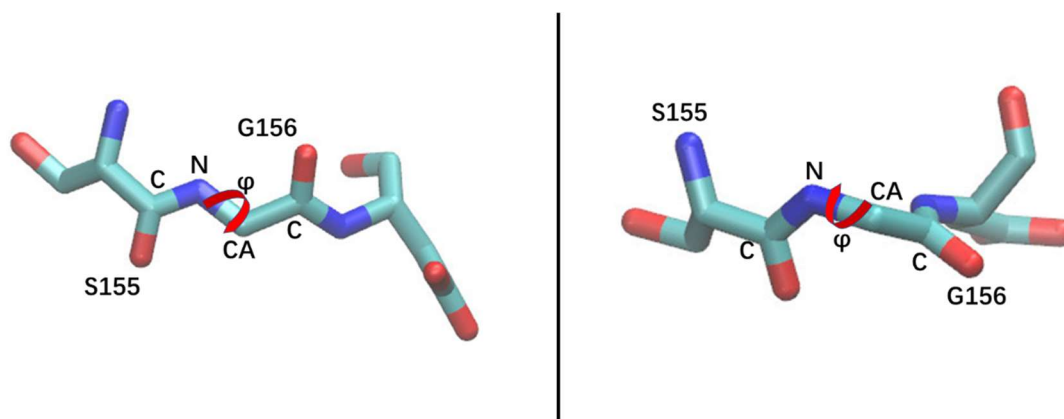
There was only 3%  $C\alpha$  RMSD over  $0.9\text{\AA}$  in L2b  $\beta$ -lactamase (Figure 60). However, there were numbers of highly dynamic regions, including:  $\beta$ 1- $\beta$ 2 loop,  $\alpha$ 6,  $\alpha$ 6- $\alpha$ 7 loop,  $\alpha$ 7,  $\alpha$ 7- $\beta$ 5 loop,  $\Omega$  loop,  $\alpha$ 10, hinge region,  $\beta$ 7- $\beta$ 8 loop and  $\beta$ 8- $\beta$ 9 loop. The high number of dynamic regions and low percentage of high  $C\alpha$  RMSD indicated the dynamic motions in L2b  $\beta$ -lactamase were relatively stable. There were 9 states built in L2b  $\beta$ -lactamase MSM analysis. Most of the structures are present in metastable state 7, 8 and 9 which represent more than 90% of all the structures. The RMSF plot for L2b  $\beta$ -lactamase is illustrated in Figure 103. State 3 was more dynamic than other states. Interestingly, state 3 only represented less than 1% of all conformations. It was not surprising that this state was a highly dynamic intermediate and a transitional state.



**Figure 103:** RMSF plot for 9 states of L2b. Different colours represent different states.



In L2b  $\beta$ -lactamase,  $\alpha 1$  was much more stable than that observed in L2a  $\beta$ -lactamase. Only the N terminal displayed disordered dynamics.  $\beta 9$ - $\alpha 15$  loop was stable in this case. In contrast,  $\alpha 6$ - $\alpha 7$  showed high dynamics in L2b  $\beta$ -lactamase. This could be explained by the stability of  $\beta 9$ - $\alpha 15$  loop. Based on the L2b  $\beta$ -lactamase residue in the DCCM plot (Figure 91), there is a negative correlation between  $\beta 9$ - $\alpha 15$  and  $\alpha 6$ - $\alpha 7$ . The stable state of  $\beta 9$ - $\alpha 15$  loop could increase the dynamics of  $\alpha 6$ - $\alpha 7$  and allow interactions to form more readily in this region. The  $\beta 7$ - $\beta 8$  loop comes closer to  $\Omega$  loop and can both form interactions with residues in the  $\Omega$  loop (N170-N238) and the hinge region (T216-R243) which results in the dynamic in these two areas. On the other side of the enzyme, the C terminal moves away from  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loop in L2b  $\beta$ -lactamase. The side chain of R283 moves closer to S276. This interaction pulls the entire  $\alpha 15$  helix away from  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$ . Without the influence of the N-terminal, the  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  are still highly dynamic regions but show less RMSF when compared with L2a  $\beta$ -lactamase. The interactions between these two loops, via L50-V259 and T52-P257, are the reason why they are positively correlated and can influence each other. The residues in  $\alpha 10$  helix were also highly dynamic. There was a  $\phi$  angle main chain rotation (around 100 degree) of G156 (Figure 104), which could change the localised conformation of several residues.<sup>230</sup>

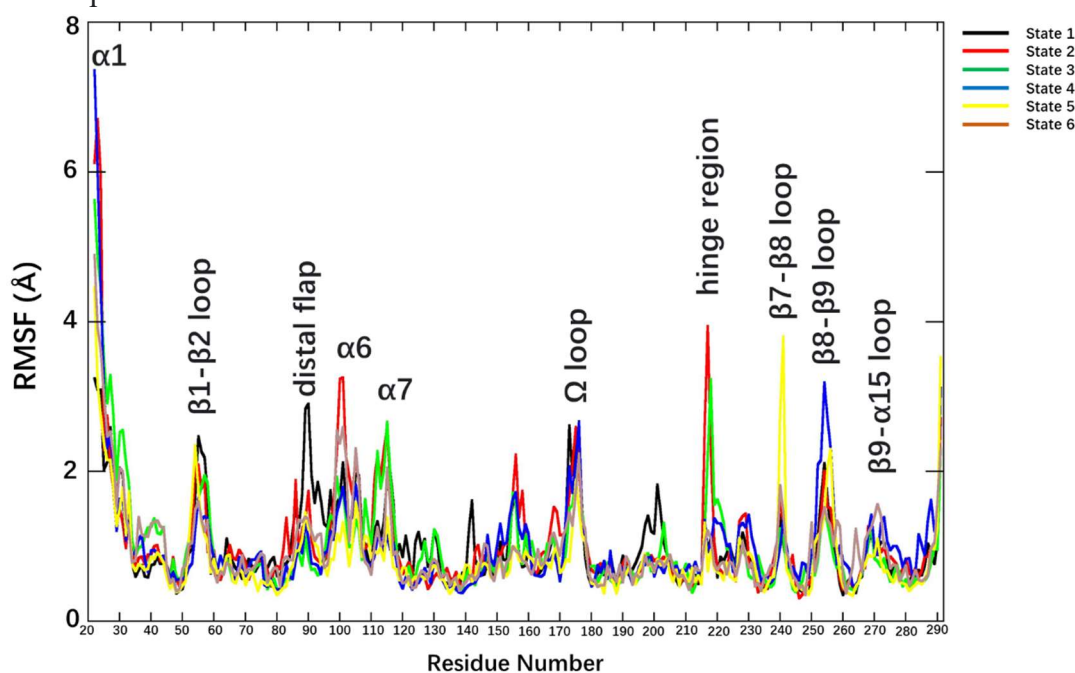


**Figure 104:**  $\phi$  angle rotation of G156 in  $\alpha 10$ . The main chain rotation resulted in the residue and its neighbours having a big conformational change.

In L2b  $\beta$ -lactamase, the cross-correlation was much apparent. The CVAE results of L2b  $\beta$ -lactamase in low energy basins were also very similar. The only differences were present in  $\beta 1$ - $\beta 2$  loop and  $\Omega$  loop. 9 states were built, almost all of them exhibited similar dynamic pattern. The most populated states had the same RMSF during the simulation. This could indicate that the L2b  $\beta$ -lactamase it is a much stable system.

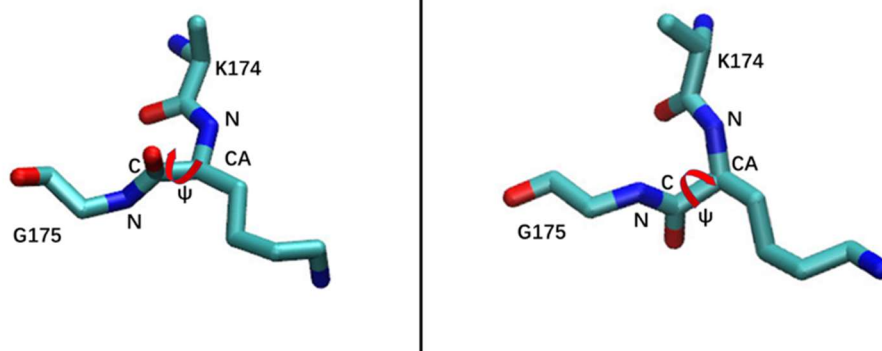
## L2c $\beta$ -lactamase discussion

There was only 4% C $\alpha$  RMSD over 0.9 Å in L2c  $\beta$ -lactamases (Figure 62). The high RMSD regions included  $\alpha$ 1,  $\beta$ 1- $\beta$ 2 loop, distal flap,  $\alpha$ 6- $\alpha$ 7,  $\Omega$  loop,  $\alpha$ 10, hinge region,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop. The RMSF plot was generated for 6 states in L2c (Figure 105). Based on the Table 8, most structures in the simulation will remain in state 6. State 5 was another populated state but is also a substate of state 6. The structures go from state 1 and reach state 6. After staying in state 6 some structures can change conformation between state 5 and state 6. Thus states 5 and 6 represent about 88% of all the structures. They exhibit much lower energy than the other metastable 4 states (Table 8). Only a few cross-correlations were identified in L2c  $\beta$ -lactamase; a little more than L2b  $\beta$ -lactamase but much less than L2a  $\beta$ -lactamase. The negative correlation between  $\beta$ 9- $\alpha$ 15 loop and  $\alpha$ 6 that was observed in L2a and L2b  $\beta$ -lactamases was still present.



**Figure 105:** RMSF plot for the 6 states of L2c  $\beta$ -lactamase. Different colour represents different states.

In the L2c  $\beta$ -lactamase, the  $\alpha 1$  helix,  $\beta 9$ - $\alpha 15$  loop,  $\beta 7$ - $\beta 8$  loop and the hinge region are highly dynamic, similar to L2a  $\beta$ -lactamase. The  $\alpha 1$  helix comes in the vicinity of  $\beta 9$ - $\alpha 15$  loop making interactions S21-Q273 and result in the elevated dynamics of these two regions. As mentioned above and showed in Figure 99, the hinge region,  $\beta 9$ - $\alpha 15$  loop and  $\beta 7$ - $\beta 8$  loop can affect each other's dynamics.<sup>229</sup> The  $\alpha 1$  helix and  $\beta 9$ - $\alpha 15$  loop dynamics will also result in the dynamics of hinge region and  $\beta 7$ - $\beta 8$  loop indirectly. However, the general dynamics of L2c  $\beta$ -lactamase was much stable than L2a  $\beta$ -lactamase. Due to the negative correlation that existed between  $\beta 9$ - $\alpha 15$  loop and  $\alpha 6$ , the patterns of motion in of  $\alpha 6$ - $\alpha 7$  was similar to L2b  $\beta$ -lactamase. In this case the  $\alpha 6$  and  $\alpha 6$ - $\alpha 7$  loop was in a relaxed state and can form interactions with neighboring residues. The  $\Omega$  loop is also in a highly dynamic state, where G175 main chain rotation is observed on the  $\Omega$  loop ( $\psi$  angle).<sup>230</sup> This occurs with the same frequency for all states which can explain the similar RMSF value in all states of the  $\Omega$  loop.<sup>229,230</sup> The main chain rotates by about  $\sim 200$  degrees rotation (Figure 106). Besides,  $\beta 7$ - $\beta 8$  loop can come closer to the  $\Omega$  loop and form bonds with the  $\Omega$  loop (N170-N238).<sup>246</sup> On the C terminal side, the combination of C terminal,  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loop was maintained. The C terminal was attracted by both  $\beta 1$ - $\beta 2$  loop and  $\beta 8$ - $\beta 9$  loop, resulting in a relatively high dynamic.

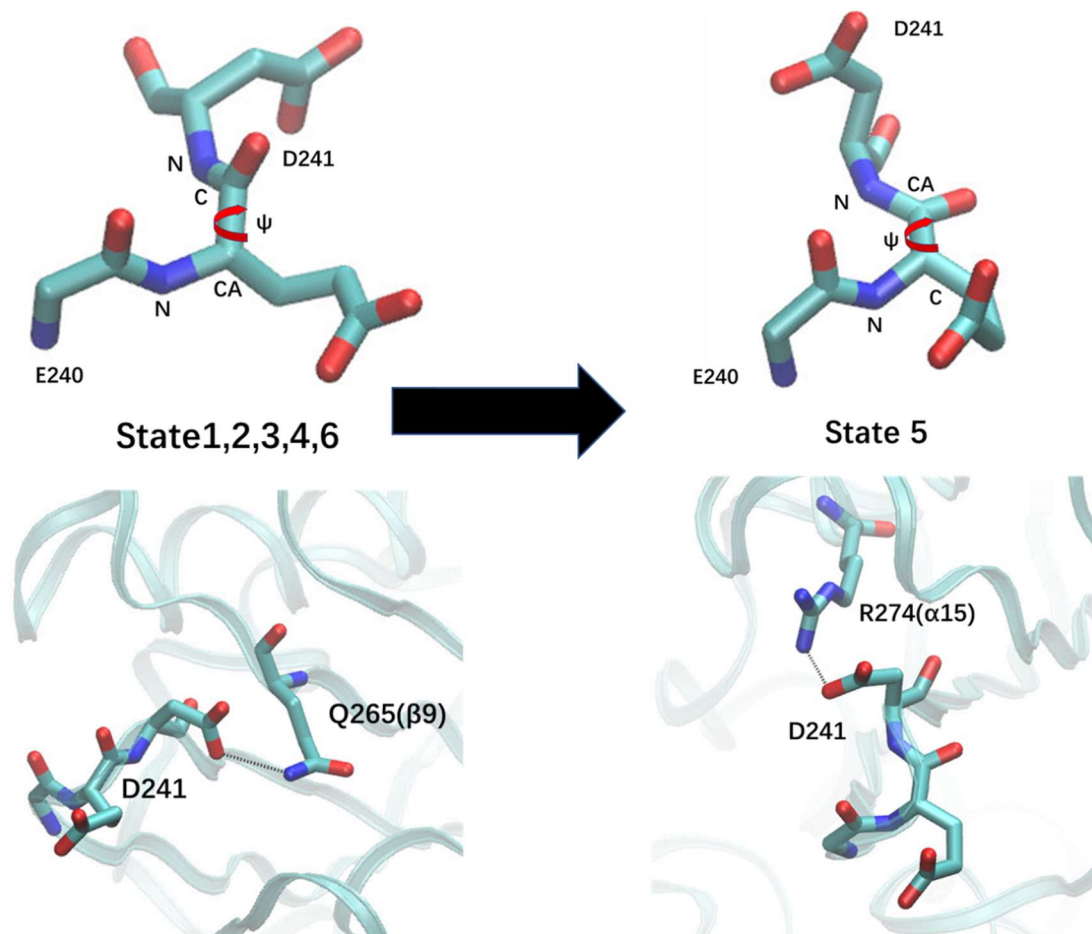


**Figure 106:**  $\psi$  angle rotation of G175 on  $\Omega$  loop. The main chain rotation will result in the residue and its neighbours having a big conformational change.

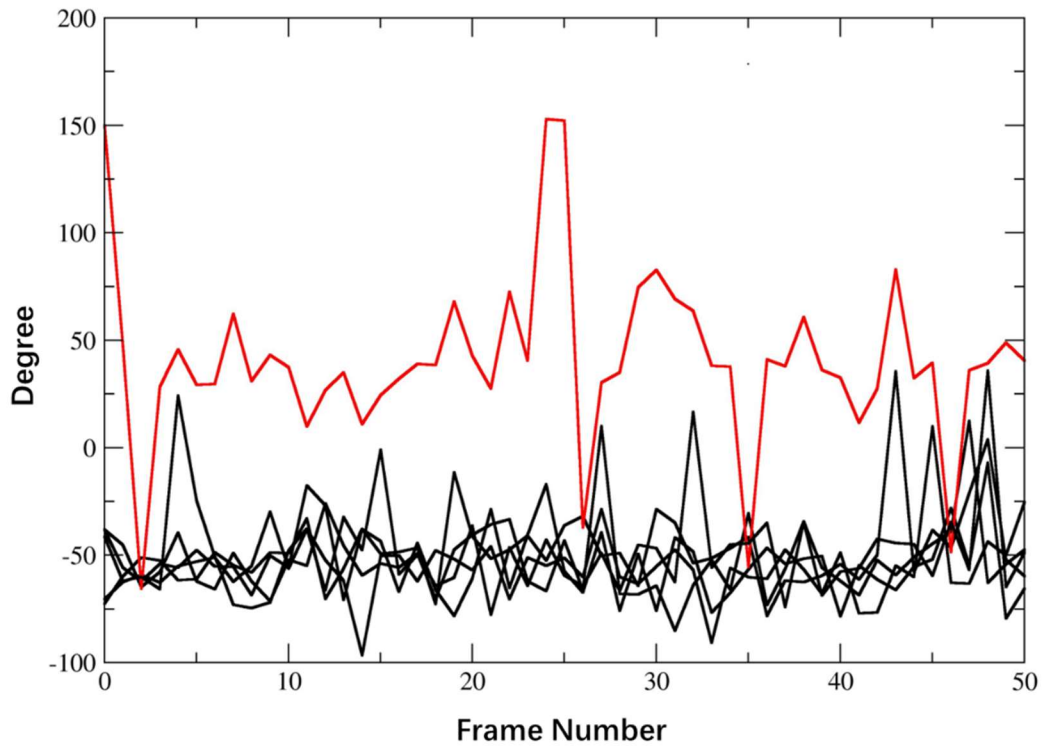
Based on the RMSF plot (Figure 105), a high RMSF region only existed in state 5 was observed, happened on  $\beta$ 7- $\beta$ 8 loop. It was found a main chain rotation mostly only happened in state 5 (Figure 107). The  $\psi$  angle of E240 was over 0 degree stay at around 25 degree in most structures (red line in Figure 108), while on all other 5 states, the  $\psi$  angle was mostly under 0 degree and stayed at around -50 degree (black lines in Figure 108). The main chain rotation made the stable bond D241-Q265( $\beta$ 9) change to stable bond D241-R274( $\alpha$ 15) (Figure 107). This main chain rotation could possibly explain why some structures reach state 6 and then change to state 5, because this was the only obvious dynamic difference between state 5 and state 6. This is also confirmed by the low mean first passage time between state 5 and 6 indicating a low energy barrier between the two states.<sup>188,247</sup> The distal flap was also observed to be highly dynamic but only included few residues (R86 and R87). The RMSF result highlights that this mostly happens in state 1 and 2 and becomes more stable in the following states. State 1 and 2 only represent for less than 1% of all conformations in total and the dynamics could be because of the R86 and R87 long side chain rotation. In general, the distal flap

was stable in most cases.

In L2c  $\beta$ -lactamase simulation, the CVAE extracted structures were quite similar.<sup>248,249</sup> Only  $\beta$ 1- $\beta$ 2 loop had obvious differences. The  $\Omega$  loop and  $\alpha$ 1 helix had small side chain rotations which were similar to L2b  $\beta$ -lactamase. In general, most L2c  $\beta$ -lactamase structures had similar dynamic pattern, from state 1 to state 6. Only state 5 had a main chain rotation which resulted in a differing RMSF peak. In spite of this state 5 was mostly similar to state 6. The main chain rotation could be the key factor of the different dynamics between state 5 and all other states. Very few cross-correlations were observed and not large differences between different states. This could imply a simple dynamic system which can be influenced by limited factors.<sup>245</sup>



**Figure 107:** The  $\psi$  angle rotation of  $\beta 7$ - $\beta 8$  loop. The rotation occurs in state 5. The main chain rotation breaks the stable D241-Q265( $\beta 9$ ) bond and made a new stable bond between D241-R274( $\alpha 15$ ). The rotation results in a high dynamics for D241.

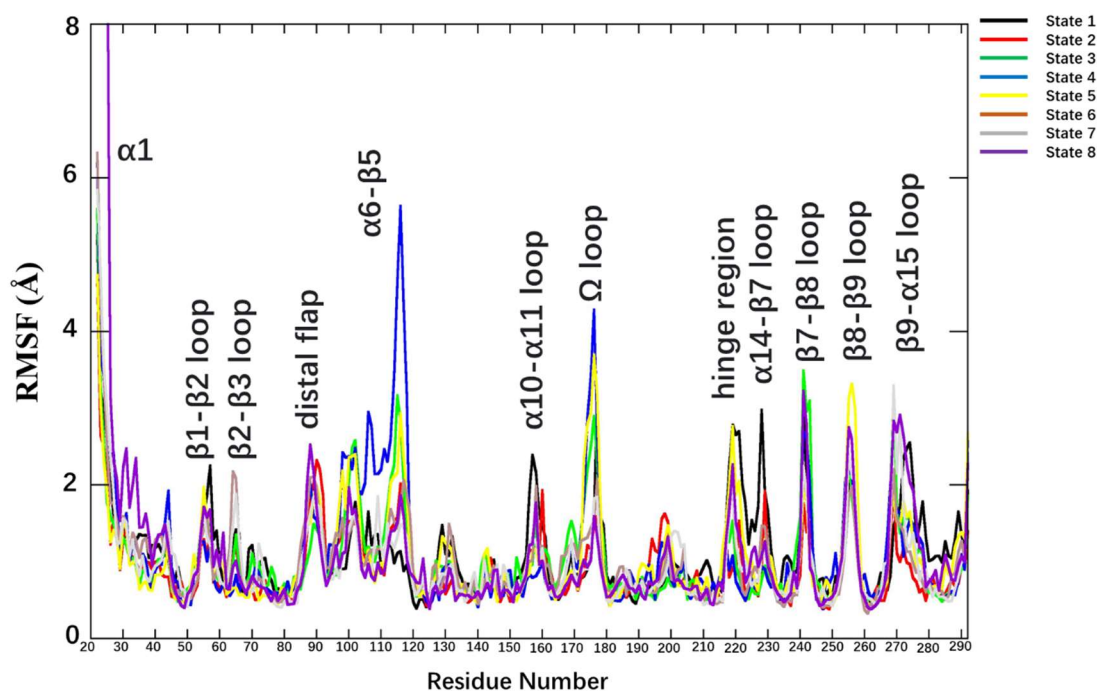


**Figure 108:** Rotation of D241  $\psi$  angle. The red line represents state 5, black lines represents state 1, 2, 3, 4 and 6. Only in state 5, the average value of  $\psi$  angle was about 25 degree. In all other states, the value of the  $\psi$  angle was around -50 degrees.



## L2d $\beta$ -lactamase discussion

In L2d  $\beta$ -lactamase, 13% of C $\alpha$  RMSD was more than 0.9Å (Figure 64). This is the most dynamic of all L2  $\beta$ -lactamase enzymes. The highly dynamic regions include:  $\alpha$ 1,  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 2- $\beta$ 3 loop, distal flap,  $\Omega$  loop,  $\alpha$ 10- $\alpha$ 11 loop, hinge region,  $\alpha$ 14- $\beta$ 7 loop,  $\beta$ 7- $\beta$ 8 loop,  $\beta$ 8- $\beta$ 9 loop and  $\beta$ 9- $\alpha$ 15 loop. L2d  $\beta$ -lactamase was the most different enzyme from the other three L2  $\beta$ -lactamase enzymes. It displays the least identical sequence, only around 69% (Figure 56). In contrast, L2a, L2b and L2c had over 90% sequence identity. 8 states were built for L2d  $\beta$ -lactamase. State 8 had 57% of structures while state 4 was 7%, state 5 was 6%, state 6 was 12% and state 7 is 15%. The structural distribution was more even in L2d  $\beta$ -lactamase. Therefore, the dynamic system could behave more varied which resulted in the large numbers of high dynamic regions. The RMSF plot was generated for L2d  $\beta$ -lactamase states and is illustrated in Figure 109.

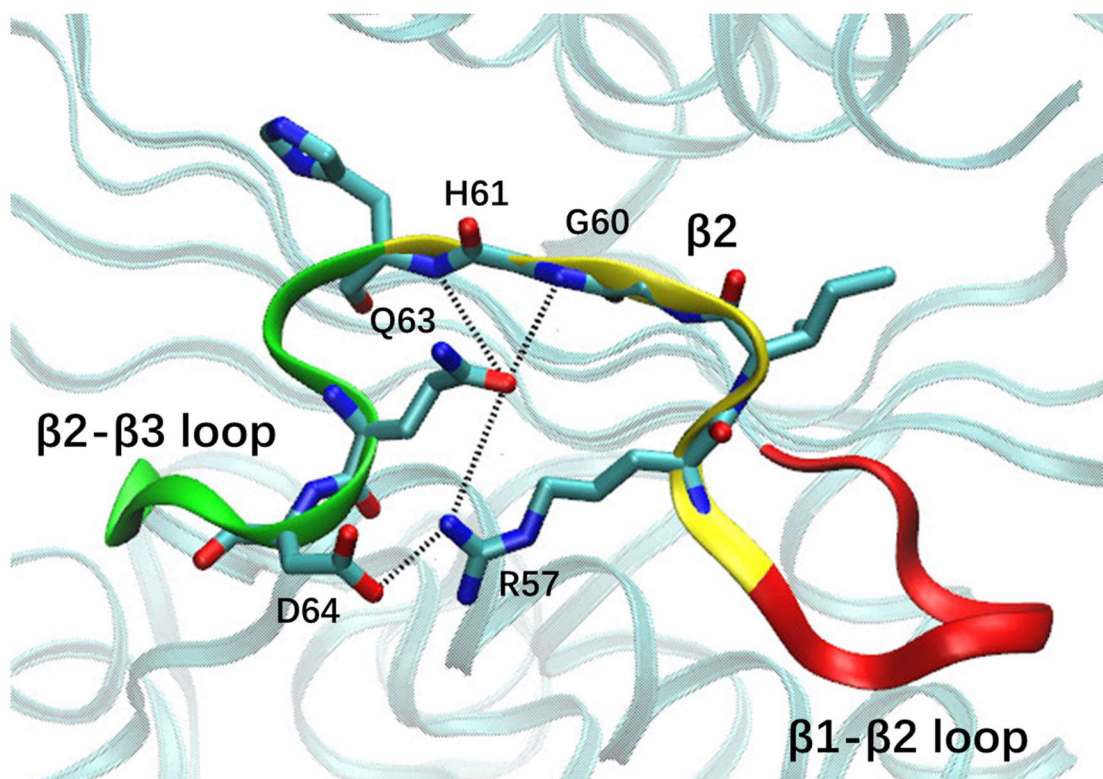


**Figure 109:** RMSF plot for 8 states of L2d  $\beta$ -lactamase. Different colour represents different states.

In L2d  $\beta$ -lactamase,  $\alpha$ 1 helix,  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop and hinge region are in close proximity (mentioned above and shown in Figure 99). The dynamics of each one of these four regions will affect the other three regions' dynamic.  $\alpha$ 1 was quite stable in most states, but in state 8 the RMSF was very high. It was found that in state 8, N terminal can rotate to a much wider region to make interaction with, such as with  $\beta$ 9- $\alpha$ 15 loop or  $\alpha$ 14 helix. The rotation of  $\alpha$ 1 can influence  $\beta$ 9- $\alpha$ 15 loop. As mentioned above, the bonds between T216-R243 and Y271-R243 control the dynamics of  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop and hinge region. In L2d, Y271 has been instead by an aspartic acid (D272. L2d had an extra residue at residue 40. When residue at the same position in the sequence alignment, their number need to plus one after residue 40 in L2d). The residue changes on Y271 made R244 in L2d easier to get attracted by C221. The change in sequence of L2d enzyme, did not alter the correlation between  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop and hinge region. Instead, it enhanced the dynamics and made it more varied.

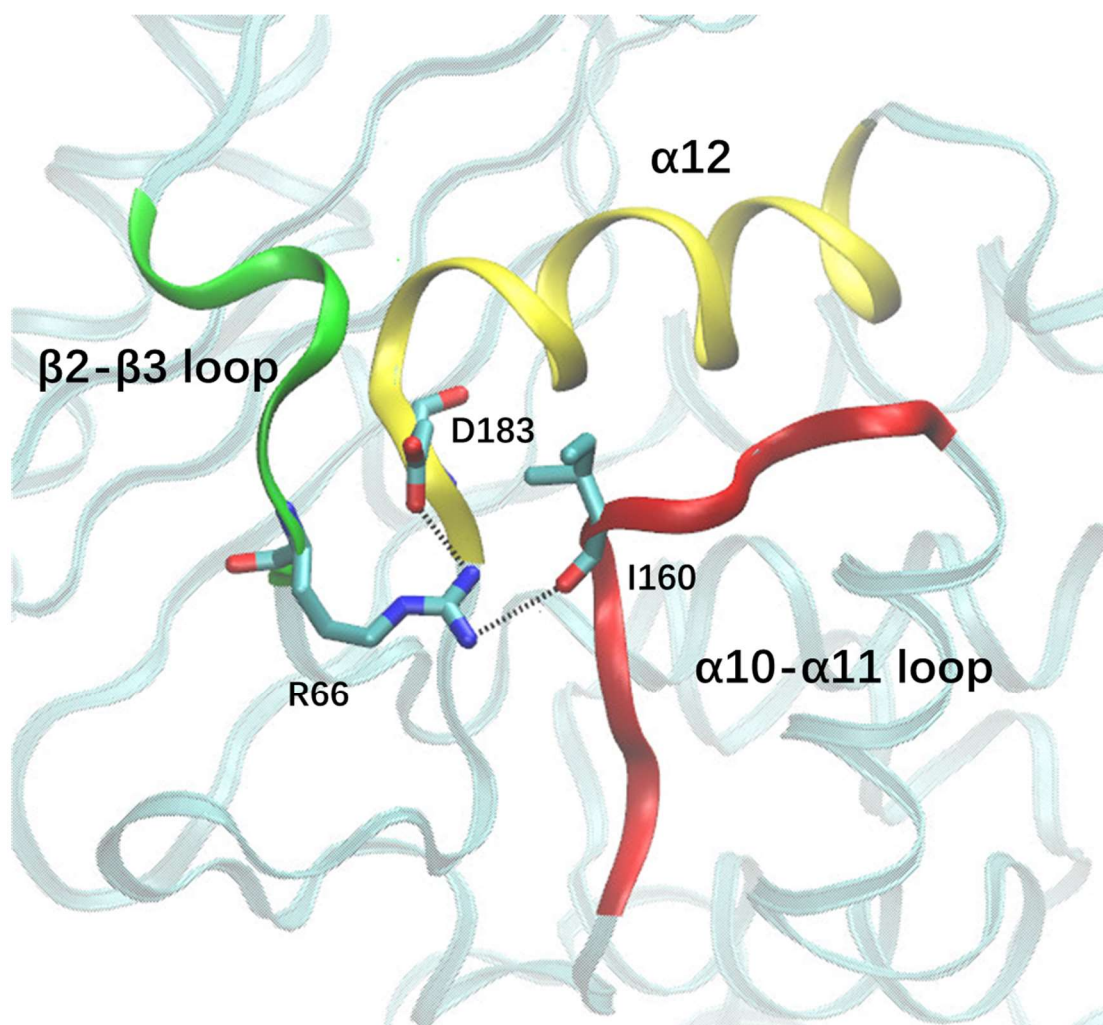
On the other side of the protein, the C terminal,  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop interactions were broken. The  $\alpha$ 15 helix showed greater stability. The C terminal remained in its position and could not come close to  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop. The bonds between  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop were still formed. Without the influence of the C terminal residues, it was much easier for  $\beta$ 1- $\beta$ 2 loop to make interactions with  $\beta$ 2- $\beta$ 3 loop (such as R57-Q63 and R57-D64, shown in Figure 110). Because of R57 influencing dynamics, the bond between H61 and  $\beta$ 2 loop (such as G60 and H61) was easier to break (shown in Figure 110). The formation and bond breakage among  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 2 and  $\beta$ 2- $\beta$ 3 loop

results in high dynamics of the three regions. Besides,  $\beta 8$ - $\beta 9$  loop could also be influenced as well via  $\beta 1$ - $\beta 2$  loop.



**Figure 110:** The interactions between  $\beta 1$ - $\beta 2$  loop,  $\beta 2$  and  $\beta 2$ - $\beta 3$  loop. The different coloured ribbons represent different regions, red for  $\beta 1$ - $\beta 2$  loop, yellow for  $\beta 2$ , green for  $\beta 2$ - $\beta 3$  loop. R57 could form bonds with both Q63 and D64. The presence of R57 will make the bonds between Q63 and  $\beta 2$  (such as H61 and G60) easier to break.

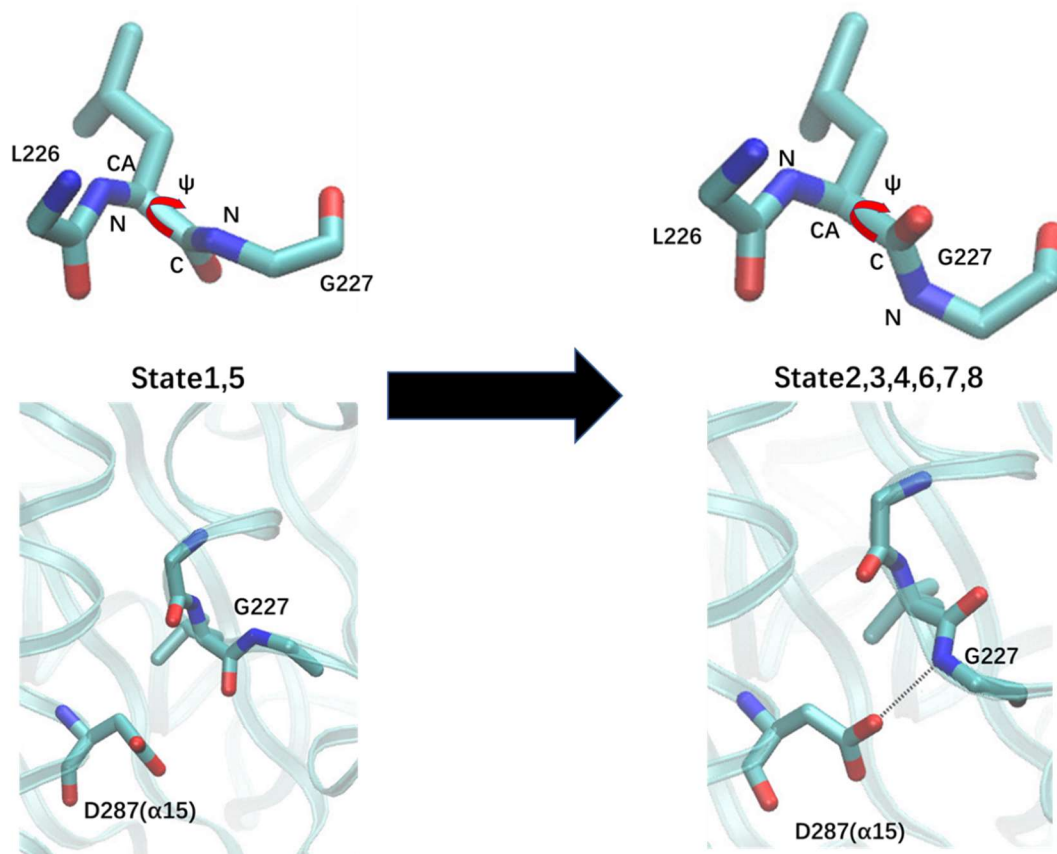
The dynamics of  $\beta 2$ - $\beta 3$  loop can also influence  $\alpha 10$ - $\alpha 11$  loop. The long side chain of R66 displayed bond formation with both  $\alpha 10$ - $\alpha 11$  loop and  $\alpha 12$  helix (D183). R66 was closer to  $\alpha 12$  helix, but in some cases, it can be attracted by  $\alpha 10$ - $\alpha 11$  loop and form bond with the backbone of I160. The bond formed between R66 and I160 did not affect the stability of  $\alpha 12$  helix very much. This could be because the secondary structure of the protein is harder to show a conformational change rather than loops.<sup>250,251</sup> On the other hand, the bond formation and loss with  $\alpha 10$ - $\alpha 11$  loop results in the  $\alpha 10$ - $\alpha 11$  loop to be highly dynamic (Figure 111).



**Figure 111:** The interactions between  $\beta 2$ - $\beta 3$  loop,  $\alpha 10$ - $\alpha 11$  loop and  $\alpha 12$ . The different colour ribbons represent different regions:  $\beta 2$ - $\beta 3$  loop (green),  $\alpha 10$ - $\alpha 11$  loop (red),  $\alpha 12$  helix (yellow). R66 can make interactions with  $\alpha 12$  (D183), and  $\alpha 10$ - $\alpha 11$  loop (I160). These interactions stabilise the loops in the region.

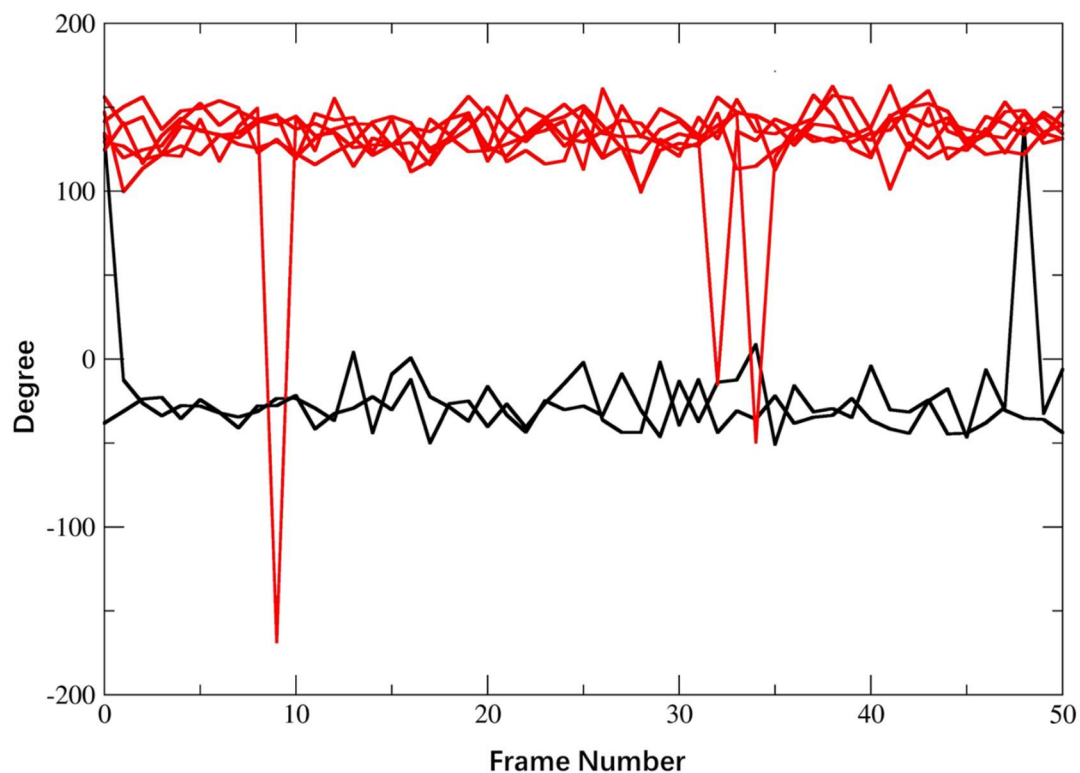
There was a  $\psi$  angle main chain rotation observed in L226 on  $\alpha 14$ - $\beta 7$  loop (about 180 degree) (Figure 112). The rotation occurs in state 2, 3, 4, 6, 7 and 8. It seldom happens in state 1 and 5. In state 1 and 5,  $\psi$  angle is stabilized at around -40 degree and became around 140 degree in state 2,3,4,6,7 and 8 (Figure 113). The rotation resulted in bond formation between G227 and D287( $\alpha 15$ ) which could help to stabilise the region (Figure 112). This is the reason why state 2,3,4,6,7 and 8 display a lower RMSF in this

region. Besides, distal flap showed a high dynamic in L2d  $\beta$ -lactamase.<sup>229,230</sup> The metastability of this region could imply an “open” and “close” dynamics. The open conformation of distal flap could be linked akin to a pocket opening to access the hydrophobic nodes, similar to that observed in KPC-2  $\beta$ -lactamase.<sup>230</sup> To investigate this loop further, the distance between V90(C $\alpha$ ) and F202(C $\alpha$ ) was measured. It was found that the structures stayed in an “open” state in most cases ( $\sim 18$  Å), but in some rare occasions, the distance will reach less than 15 Å.<sup>252</sup> No particular “closed” state was found in any of the 8 states. Only in some rare cases the closed conformation was observed. Such a low number of observed conformations is not statistically significant to label it as a closed state. The tendencies of this region result in the dynamic. The  $\Omega$  loop was also observed to be highly dynamic. Based on the cross-correlation plot (Figure 89), a negative correlation between  $\Omega$  loop and  $\alpha 15$  was identified.<sup>231,245</sup> Therefore, when  $\alpha 15$  becomes more stable, the negative correlation results in higher  $\Omega$  loop dynamics.



**Figure 112:** The  $\psi$  angle rotation occurs on  $\alpha14$ - $\beta7$  loop in residue G227. The rotation occurs in state 2, 3, 4, 6, 7, 8. The main chain rotation allows bond formation between G227 (bb) and the side chain of D287 ( $\alpha15$ ), helping to stabilise the loop. Therefore, state 2, 3, 4, 6, 7 and 8 show low RMSF in this region.





**Figure 113:** Rotation of the  $\psi$  dihedral angle of residue G227 .The black lines represent state 1 and 5, the red line represent state 2, 3, 4, 6, 7 and 8. In state 1 and 5, the  $\psi$  angle stayed at around -40 degree and was around 140 degree in states 2,3,4,6,7 and 8.

In L2d  $\beta$ -lactamase, the number of high dynamic regions was relatively higher than that of the other three enzymes. More main chain rotations were observed. The dynamic cross-correlation map also shows highly dynamic state 2. In CVAE, the extracted conformations show high diversity (Figure 95 and 96). This could be because of most of the L2d structures were clustered separately. Only a few conformations mixed with the other L2  $\beta$ -lactamases. The differences in the extracted conformations structures also implied a complicated dynamic system. These highly mobile regions could influence and affect the catalytic process in this enzyme.<sup>248,251</sup>

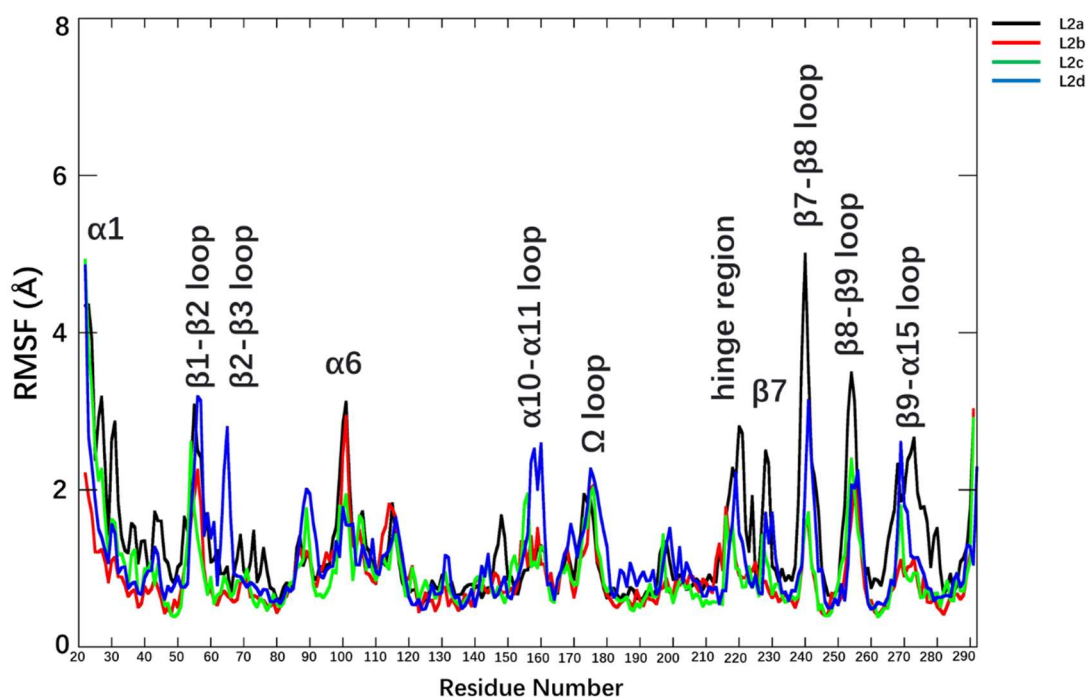
## Combined L2 $\beta$ -lactamase discussion

Among all four L2  $\beta$ -lactamase proteins, there were several common dynamic regions:  $\beta$ 1- $\beta$ 2 loop,  $\beta$ 7- $\beta$ 8 loop, hinge region and  $\beta$ 8- $\beta$ 9 loop. These regions can be clustered as two dynamic combinations observed in all four simulation. First, was the  $\alpha$ 1 helix,  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop and hinge region. The  $\alpha$ 1 helix shows high dynamics in the simulation. It is positioned at the untethered N-terminal end of the enzyme and does not make interactions with any other secondary structural elements. Thus, as a result of its flexibility this helix can come close to  $\beta$ 9- $\alpha$ 15 loop and affect  $\beta$ 9- $\alpha$ 15 loop dynamics.  $\beta$ 9- $\alpha$ 15 loop is positioned parallel to  $\beta$ 7- $\beta$ 8 loop; this allows bond formation between Y271 ( $\beta$ 9- $\alpha$ 15 loop) and R243 ( $\beta$ 7- $\beta$ 8 loop). However, the long side chain of R243 can also come close to T216 from the hinge region. Therefore, the R243 can also form bonds with both Y271 and T216. The bond formation and loss between these three residues influences the dynamics in  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop and hinge region (Figure 99).

The second region was the interactions between the C-terminal  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop region. C-terminal is also highly dynamic in most cases. The C-terminal can form interactions with  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loops individually or with both loops simultaneously. The interactions can vary, depending upon the sequence from different enzymes (such as G290-R56 in L2a). The high dynamics observed in the C terminal will influence the stability of  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop.



Different from other more investigated class A  $\beta$ -lactamase such as KPC, SME, SHV and TEM, L2  $\beta$ -lactamase has an extra  $\alpha 1$  helix, which could affect the dynamics of the entire system.<sup>210,229,243</sup> The  $\alpha 1$  helix is relatively stable in L2b and L2d  $\beta$ -lactamases. L2d  $\beta$ -lactamase has a sequence identity of around 69% with all other three enzymes. However, L2a, L2b, L2c have about 92% sequence identity, which makes it easier to identify key residues involved in the dynamics in  $\alpha 1$  helix (Figure 56 shows sequence alignment). In L2b  $\beta$ -lactamase  $\alpha 1$  helix, only residue N21 was different from L2a and L2c (residue S21). It was also observed that residue S21 correlated with the dynamics of  $\alpha 1$  helix in L2a and L2c enzymes. This indicates that residue S21 could be the key residue for stabilising the interactions in the  $\alpha 1$  helix.



**Figure 114:** RMSF plot for L2a, L2b, L2c and L2d  $\beta$ -lactamases. Different colours indicate different proteins.

L2a, L2b and L2c  $\beta$ -lactamases have high sequence identity.<sup>81,204</sup> However, the dynamic manifestation between these systems is quite different (Figure 114, RMSF plot for all four proteins). Based on the RMSF plot, even in spite of L2d  $\beta$ -lactamase having low sequence identity, it still shows similar dynamics.<sup>81,204</sup> This indicates there should be a conserved dynamic pattern for all four proteins. L2a  $\beta$ -lactamase had 25% of Ca RMSD  $>0.9\text{\AA}$  while L2b and L2c  $\beta$ -lactamase had 3% and 4% respectively. There are only few residues in L2a  $\beta$ -lactamase that are different from L2b and L2c enzymes, which can be a cause of this difference namely V41, R42, Q87, S154, I155, T162, L228 and R232. These residues mostly belong to the stable regions. This implies that some of them could be crucial in signal transfer. Besides, L2a  $\beta$ -lactamase was dynamic in  $\alpha 1$ - $\alpha 2$ ,  $\beta 9$ - $\alpha 15$  loop,  $\beta 7$ - $\beta 8$  loop which was different from all other 3 proteins. L2b and L2c  $\beta$ -lactamase show similar RMSF in these regions. This indicates that L2a  $\beta$ -lactamase key residues should be different from both L2b and L2c enzymes on the same position of the sequence. Due to the high sequence similarity between L2a, L2b and L2c, there were very limited residues that could meet the requirement. Residues V41 and R42 were the only two residues matching the selection. Besides, the RMSF peak at  $\beta 4$ - $\alpha 16$  loop could also use the same pattern of selection. The key residues in L2c  $\beta$ -lactamase in this region should be different from both L2a and L2b enzymes. R97 and D98 in L2c enzyme are the only ones that fit the criteria. These selected residues could be important nodes for signal transfer or the dynamic system.<sup>229,253</sup> Future research can target on these residues function, such as whether they join the catalytic activity or they can affect the substrate binding.<sup>253</sup>

## 8. Conclusion

### 8.1 L1 Metallo- $\beta$ -lactamase Conclusions

This study aimed to improve the understanding of L1  $\beta$ -lactamase using molecular dynamics simulations. The adaptive sampling dynamics simulation was employed, MSM models were built, CVAE were carried out, network analysis was used to investigate the dynamics of the elongated gating loops using various computational methods.<sup>188,197,226</sup>

More specifically, this work focuses on the dynamics of two gating loops and the relationship between the two elongated loops and the zinc binding site.<sup>204</sup> The results indicate that the two gating loops' ( $\alpha$ 3- $\beta$ 7 loop and  $\beta$ 12- $\alpha$ 5 loop) dynamics can affect the structural dynamics of the zinc binding site. Different conformations of these two loops results in different substrate binding pocket size and affects substrate binding. There are three main structural features that control the dynamics of the gating loops which were highlighted in different metastable states. The salt bridge between residue D150c and residue R236, the  $\pi$ - $\pi$  stacking between residue H151 and residue Y227, P225 in the "in" configuration classify the loops to be in open, intermediate or closed conformations. Besides,  $\beta$ 12- $\alpha$ 5 loop displays highly flexibility during dynamics while  $\alpha$ 3- $\beta$ 7 loop was relatively stable due to the direct proximity to the zinc binding site. Furthermore, the correlation between the four subunits was investigated by employing the CVAE-based deep learning and network analysis.<sup>174,175,248,249</sup> The results indicated

a dimer-of-dimers pattern of dynamics in the L1 MBL tetramer.<sup>213,214</sup> The CVAE results revealed subunit A has similar dynamics with subunit C while subunit B has similar dynamic pattern with subunit D. Network analysis also verified this through the parallel and vertical comparison of the network paths. The dynamic research of the two gating loops and the correlation between the tetramers will help the understanding of L1  $\beta$ -lactamase dynamics. The “open” and “closed” conformation can control the volume of the binding pocket. The pocket volume has a significant increase in the “open” state which can allow bigger substrates to get access. Therefore, the “open” state can give the maximum probability for L1 MBL inhibitor research. Furthermore, the key interactions that control the “open” and “closed” conformations are also revealed. Inhibiting or changing the function of these key interactions’ residues can possibility result in the maintain of “open” state and help the inhibitor design. Besides, a dimer of dimers dynamic pattern in L1 MBL tetramer has also been indicated. The conserve nodes on the protein-protein interactions between monomers could also be the potential target. Inhibiting these nodes function could possibility make the tetramer work similar to four monomers and reduce the dull catalytic function.

The work on L1 MBL has been published as: “*Gating interactions steer loop conformational changes in the active site of the L1 metallo- $\beta$ -lactamase*” Zhao, Z.; Shen, X.; Chen, S.; Gu, J.; Wang, H.; Mojica, M. F.; Samanta, M.; Bhowmik, D.; Vila, A. J.; Bonomo, R. A.; Haider, S. Gating Interactions Steer Loop Conformational

Changes in the Active Site of the L1 Metallo- $\beta$ -Lactamase. *eLife* 2023, 12, e83928.

<https://doi.org/10.7554/eLife.83928>.<sup>203</sup>

## 8.2 L2 $\beta$ -lactamase Conclusions

The aim of this study was to advance the understanding of the L2  $\beta$ -lactamase dynamics and investigate the dynamic similarities or differences in all four L2  $\beta$ -lactamase proteins through MSM analysis, MDLofit and CVAE-based deep learning.<sup>140,170,175,188,228</sup> The Markov State Models were built for analysing the diversity of dynamics between multiple metastable states.<sup>247</sup> Different metastable states are stabilised as different structural conformations. By analysing these structural changes in the same region from different metastable states help to understand the crucial conformational changes in the structures.<sup>170</sup> These high dynamic regions were explored, and several key conformational changes were revealed and explained through various analysis tools. Dynamic cross-correlation analysis was explored by Bio3d, deep learning was performed through CVAE, the free energy map helped in defining the low energy basins from which several structures were extracted.<sup>140,231</sup> The alignment of these structures, from the low energy basin, indicated different highly mobile regions and dynamic systems.<sup>140</sup> A parallel comparison within and between systems was also carried out. As initial research, the conformational changes and the dynamic patterns in the L2  $\beta$ -lactamase family is reported. The dynamic regions with high RMSF were explored and the structural role of  $\alpha$ 1 helix/N-terminal,  $\beta$ 9- $\alpha$ 15 loop,  $\beta$ 7- $\beta$ 8 loop, hinge region, and C terminal,  $\beta$ 1- $\beta$ 2 loop and  $\beta$ 8- $\beta$ 9 loop was revealed. It is proposed that stabilising these regions may help in influencing the structural dynamics of the L2  $\beta$ -lactamase.

## 8.3 Summary and Future Work

This work is focusing on the  $\beta$ -lactamase dynamic investigation. The  $\beta$ -lactamase problem is one of the most urgent topics for public health.<sup>52,242,254</sup> It can inhibit the function of the  $\beta$ -lactam antibiotic which is the most widely used antibacterial agent family.<sup>15,28,48</sup> *S. maltophilia* can produce two kinds of  $\beta$ -lactamases: L1 MBL and L2 SBL.<sup>88,89,232,255</sup> These two  $\beta$ -lactamases allow it to hydrolyse almost all the  $\beta$ -lactam antibiotic drug.<sup>88,232</sup> The mortality rate of *S. maltophilia* is from 14% to 69% for bacteraemia patients.<sup>82,83</sup> The more frequent appearance of *S. maltophilia* makes the investigation of L1 and L2  $\beta$ -lactamases become important. The more information we learn, the easier we can design corresponding  $\beta$ -lactamase inhibitors.

In this work, L1 gating loops dynamic and subunits correlation have been investigated. The “open” and “closed” conformations with the key interactions control these conformations were revealed.<sup>203</sup> As the “open” state can give the maximum probability for drug designer. The “open” state is the best target for inhibitor developing. Through inhibiting these key interactions will potentially result in the L1 MBL stay in “open” state. Besides, a dimer of dimers correlation pattern was also revealed, key residues were indicated.<sup>203,213</sup> The residues on the protein-protein interactions could be the potential targets for affecting the tetramer signal transferring and thus reduce the catalytic activity of the tetramer.

In L2 work, the important dynamic regions are investigated. Since there is few research on L2 SBL, this work is an early-stage study for the future investigation. Two high dynamic region combinations were revealed, several main chain dihedral angle change were indicated, the similarity and difference of four enzymes' dynamic pattern were studied. The two high dynamic regions combinations exist in all four kinds of L2  $\beta$ -lactamase. Thus they can be important for the whole dynamic system function. Target these regions and inhibit the dynamic combinations will potentially inhibit L2 SBL catalytic function.

There are also some limitations exist in this work. 1)The pH value was set at 7 for set up the L1 simulation. The histidine protonation type HIP only exists in lower pH value, which limits the simulation.<sup>182</sup> 2)We used tICA method for MSM model building which will ignore some fast motion in the simulation.<sup>256</sup> 3)The results of MSM are based on feature selection, multiple features and their combinations can be selected. It is possible that other features will give better results.<sup>165,173,188</sup> 4) In molecular dynamics simulation, the electronic effects are ignored.<sup>117,119,257,258</sup> 5) The parameters can be easily set in molecular dynamics simulation, but in the real environment, it could be difficult to reproduce the experiment.<sup>117-119</sup>

The future work can focus on the key residues, interactions and regions in this work. For L1, the investigation of how to inhibit the three key interactions of gating loops could be worth. Whether any of them can be inhibited, how much of “open” state will



represent for if one of these key interactions can be inhibited, what is the volume change of the binding pocket during the dynamic and which ligand could be possible fit this pocket. Furthermore, the conserve nodes of signal transferring between subunit could also be targets.<sup>197</sup> Whether these nodes' function can be inhibited. If the nodes can be inhibited, how will the tetramer work, four monomer, dimer of dimer or one monomer with trimer. Numbers of work can be done based on these results.<sup>213,214</sup> For L2, the two dynamic combinations could be important for the whole structure dynamic and it need to be tested and verified. Besides, will the extra  $\alpha$  helix affect the key residues function which has already been proved in other class A serine  $\beta$ -lactamases. The understanding of L2  $\beta$ -lactamase is still limited and need to be investigated.

## 9. Reference

- (1) Dodds, D. R. Antibiotic Resistance: A Current Epilogue. *Biochem Pharmacol* 2017, 134, 139–146. <https://doi.org/10.1016/j.bcp.2016.12.005>.
- (2) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic Discovery: History, Methods and Perspectives. *Int J Antimicrob Agents* 2019, 53 (4), 371–382. <https://doi.org/10.1016/j.ijantimicag.2018.11.010>.
- (3) Hutchings, M. I.; Truman, A. W.; Wilkinson, B. Antibiotics: Past, Present and Future. *Curr Opin Microbiol* 2019, 51, 72–80. <https://doi.org/10.1016/j.mib.2019.10.008>.
- (4) Haas, L. F. Papyrus of Ebers and Smith. *Journal of Neurology, Neurosurgery & Psychiatry* 1999, 67 (5), 578–578. <https://doi.org/10.1136/jnnp.67.5.578>.
- (5) Bassett, E. J.; Keith, M. S.; Armelagos, G. J.; Martin, D. L.; Villanueva, A. R. Tetracycline-Labeled Human Bone from Ancient Sudanese Nubia (A.D. 350). *Science* 1980, 209 (4464), 1532–1534. <https://doi.org/10.1126/science.7001623>.
- (6) Risse, G. B. The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present. *Med Hist* 1999, 43 (3), 396–397.
- (7) Schwartz, R. S. Paul Ehrlich's Magic Bullets. *N Engl J Med* 2004, 350 (11), 1079–1080. <https://doi.org/10.1056/NEJMp048021>.
- (8) Gould, K. Antibiotics: From Prehistory to the Present Day. *J Antimicrob Chemother* 2016, 71 (3), 572–575. <https://doi.org/10.1093/jac/dkv484>.

- (9) Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzæ. *Br J Exp Pathol* 1929, 10 (3), 226–236.
- (10) *The Nobel Prize in Physiology or Medicine 1945*. NobelPrize.org. <https://www.nobelprize.org/prizes/medicine/1945/fleming/lecture/> (accessed 2021-03-24).
- (11) Lewis, K. Recover the Lost Art of Drug Discovery. *Nature* 2012, 485 (7399), 439–440. <https://doi.org/10.1038/485439a>.
- (12) Finberg, R. W.; Moellering, R. C.; Tally, F. P.; Craig, W. A.; Pankey, G. A.; Dellinger, E. P.; West, M. A.; Joshi, M.; Linden, P. K.; Rolston, K. V.; Rotschafer, J. C.; Rybak, M. J. The Importance of Bactericidal Drugs: Future Directions in Infectious Disease. *Clin Infect Dis* 2004, 39 (9), 1314–1320. <https://doi.org/10.1086/425009>.
- (13) Lewis, K. Platforms for Antibiotic Discovery. *Nature Reviews Drug Discovery* 2013, 12 (5), 371–387. <https://doi.org/10.1038/nrd3975>.
- (14) Blaskovich, M. A. T.; Butler, M. S.; Cooper, M. A. Polishing the Tarnished Silver Bullet: The Quest for New Antibiotics. *Essays in Biochemistry* 2017, 61 (1), 103–114. <https://doi.org/10.1042/EBC20160077>.
- (15) Kohanski, M. A.; Dwyer, D. J.; Collins, J. J. How Antibiotics Kill Bacteria: From Targets to Networks. *Nat Rev Microbiol* 2010, 8 (6), 423–435. <https://doi.org/10.1038/nrmicro2333>.
- (16) Kahne, D.; Leimkuhler, C.; Lu, W.; Walsh, C. Glycopeptide and

- Lipoglycopeptide Antibiotics. *Chem Rev* 2005, 105 (2), 425–448.  
<https://doi.org/10.1021/cr030103a>.
- (17) Reynolds, P. E. Structure, Biochemistry and Mechanism of Action of Glycopeptide Antibiotics. *Eur J Clin Microbiol Infect Dis* 1989, 8 (11), 943–950. <https://doi.org/10.1007/BF01967563>.
- (18) Peterson, E.; Kaur, P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Front Microbiol* 2018, 9. <https://doi.org/10.3389/fmicb.2018.02928>.
- (19) Vannuffel, P.; Cocito, C. Mechanism of Action of Streptogramins and Macrolides. *Drugs* 1996, 51 Suppl 1, 20–30. <https://doi.org/10.2165/00003495-199600511-00006>.
- (20) Kapoor, G.; Saigal, S.; Elongavan, A. Action and Resistance Mechanisms of Antibiotics: A Guide for Clinicians. *J Anaesthesiol Clin Pharmacol* 2017, 33 (3), 300–305. [https://doi.org/10.4103/joacp.JOACP\\_349\\_15](https://doi.org/10.4103/joacp.JOACP_349_15).
- (21) Yoneyama, H.; Katsumata, R. Antibiotic Resistance in Bacteria and Its Future for Novel Antibiotic Development. *Biosci Biotechnol Biochem* 2006, 70 (5), 1060–1075. <https://doi.org/10.1271/bbb.70.1060>.
- (22) Evans-Roberts, K. M.; Mitchenall, L. A.; Wall, M. K.; Leroux, J.; Mylne, J. S.; Maxwell, A. DNA Gyrase Is the Target for the Quinolone Drug Ciprofloxacin in *Arabidopsis Thaliana*. *J Biol Chem* 2016, 291 (7), 3136–3144. <https://doi.org/10.1074/jbc.M115.689554>.

- (23) Reece, R. J.; Maxwell, A. DNA Gyrase: Structure and Function. *Critical Reviews in Biochemistry and Molecular Biology* 1991, 26 (3–4), 335–375. <https://doi.org/10.3109/10409239109114072>.
- (24) Wigley, D. B.; Davies, G. J.; Dodson, E. J.; Maxwell, A.; Dodson, G. Crystal Structure of an N-Terminal Fragment of the DNA Gyrase B Protein. *Nature* 1991, 351 (6328), 624–629. <https://doi.org/10.1038/351624a0>.
- (25) *The biggest antibiotic-resistant threats in the U.S.* Centers for Disease Control and Prevention. <https://www.cdc.gov/drugresistance/biggest-threats.html> (accessed 2023-07-27).
- (26) Herold, B. C.; Immergluck, L. C.; Maranan, M. C.; Lauderdale, D. S.; Gaskin, R. E.; Boyle-Vavra, S.; Leitch, C. D.; Daum, R. S. Community-Acquired Methicillin-Resistant Staphylococcus Aureus in Children With No Identified Predisposing Risk. *JAMA* 1998, 279 (8), 593–598. <https://doi.org/10.1001/jama.279.8.593>.
- (27) O’Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*; Report; Government of the United Kingdom, 2016. <https://apo.org.au/node/63983> (accessed 2023-07-27).
- (28) Perry, J.; Waglechner, N.; Wright, G. The Prehistory of Antibiotic Resistance. *Cold Spring Harb Perspect Med* 2016, 6 (6). <https://doi.org/10.1101/cshperspect.a025197>.
- (29) Benveniste, R.; Davies, J. Aminoglycoside Antibiotic-Inactivating Enzymes in Actinomycetes Similar to Those Present in Clinical Isolates of Antibiotic-

- Resistant Bacteria. *Proc Natl Acad Sci U S A* 1973, 70 (8), 2276–2280.  
<https://doi.org/10.1073/pnas.70.8.2276>.
- (30) Stokes, H. W.; Gillings, M. R. Gene Flow, Mobile Genetic Elements and the Recruitment of Antibiotic Resistance Genes into Gram-Negative Pathogens. *FEMS Microbiol Rev* 2011, 35 (5), 790–819. <https://doi.org/10.1111/j.1574-6976.2011.00273.x>.
- (31) Wiedenbeck, J.; Cohan, F. M. Origins of Bacterial Diversity through Horizontal Genetic Transfer and Adaptation to New Ecological Niches. *FEMS Microbiol Rev* 2011, 35 (5), 957–976. <https://doi.org/10.1111/j.1574-6976.2011.00292.x>.
- (32) Wright, G. D. Antibiotic Resistance in the Environment: A Link to the Clinic? *Current Opinion in Microbiology* 2010, 13 (5), 589–594.  
<https://doi.org/10.1016/j.mib.2010.08.005>.
- (33) D’Costa, V. M.; King, C. E.; Kalan, L.; Morar, M.; Sung, W. W. L.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; Golding, G. B.; Poinar, H. N.; Wright, G. D. Antibiotic Resistance Is Ancient. *Nature* 2011, 477 (7365), 457–461. <https://doi.org/10.1038/nature10388>.
- (34) Knapp, C. W.; Dolfing, J.; Ehlert, P. A. I.; Graham, D. W. Evidence of Increasing Antibiotic Resistance Gene Abundances in Archived Soils since 1940. *Environ Sci Technol* 2010, 44 (2), 580–587.  
<https://doi.org/10.1021/es901221x>.
- (35) Chait, R.; Vetsigian, K.; Kishony, R. What Counters Antibiotic Resistance

- in Nature? *Nat Chem Biol* 2011, 8 (1), 2–5.  
<https://doi.org/10.1038/nchembio.745>.
- (36) Moore, P. R.; Evenson, A. Use of Sulfasuxidine, Streptothricin, and Streptomycin in Nutritional Studies with the Chick. *J Biol Chem* 1946, 165 (2), 437–441.
- (37) Stokstad, E. L. R.; Jukes, T. H.; Pierce, J.; Page, A. C.; Franklin, A. L. The Multiple Nature of the Animal Protein Factor. *Journal of Biological Chemistry* 1949, 180 (2), 647–654. [https://doi.org/10.1016/S0021-9258\(18\)56683-7](https://doi.org/10.1016/S0021-9258(18)56683-7).
- (38) Stokstad, E. L. R.; Jukes, T. H. Further Observations on the “Animal Protein Factor.” *Proceedings of the Society for Experimental Biology and Medicine* 1950, 73 (3), 523–528. <https://doi.org/10.3181/00379727-73-17731>.
- (39) Eyssen, H.; de Somer, P. The Mode of Action of Antibiotics in Stimulating Growth of Chicks. *J Exp Med* 1963, 117 (1), 127–138.
- (40) Gustafson, R. H.; Kiser, J. S. Nonmedical Uses of the Tetracyclines. In *The Tetracyclines*; Hlavka, J. J., Boothe, J. H., Eds.; Handbook of Experimental Pharmacology; Springer: Berlin, Heidelberg, 1985; pp 405–446. [https://doi.org/10.1007/978-3-642-70304-1\\_8](https://doi.org/10.1007/978-3-642-70304-1_8).
- (41) Van Boeckel, T. P.; Glennon, E. E.; Chen, D.; Gilbert, M.; Robinson, T. P.; Grenfell, B. T.; Levin, S. A.; Bonhoeffer, S.; Laxminarayan, R. Reducing Antimicrobial Use in Food Animals. *Science* 2017, 357 (6358), 1350–1352. <https://doi.org/10.1126/science.aao1495>.

- (42) Munita, J. M.; Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol Spectr* 2016, 4 (2). <https://doi.org/10.1128/microbiolspec.VMBF-0016-2015>.
- (43) Ramirez, M. S.; Tolmasky, M. E. Aminoglycoside Modifying Enzymes. *Drug Resist Updat* 2010, 13 (6), 151–171. <https://doi.org/10.1016/j.drup.2010.08.003>.
- (44) Bush, K.; Bradford, P. A. Interplay between  $\beta$ -Lactamases and New  $\beta$ -Lactamase Inhibitors. *Nat Rev Microbiol* 2019, 17 (5), 295–306. <https://doi.org/10.1038/s41579-019-0159-8>.
- (45) Quinn, J. P.; Dudek, E. J.; DiVincenzo, C. A.; Lucks, D. A.; Lerner, S. A. Emergence of Resistance to Imipenem during Therapy for *Pseudomonas Aeruginosa* Infections. *J Infect Dis* 1986, 154 (2), 289–294. <https://doi.org/10.1093/infdis/154.2.289>.
- (46) McMurry, L.; Petrucci, R. E.; Levy, S. B. Active Efflux of Tetracycline Encoded by Four Genetically Different Tetracycline Resistance Determinants in *Escherichia Coli*. *Proc Natl Acad Sci U S A* 1980, 77 (7), 3974–3977.
- (47) Connell, S. R.; Tracz, D. M.; Nierhaus, K. H.; Taylor, D. E. Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance. *Antimicrob Agents Chemother* 2003, 47 (12), 3675–3681. <https://doi.org/10.1128/aac.47.12.3675-3681.2003>.
- (48) Bush, K.; Bradford, P. A.  $\beta$ -Lactams and  $\beta$ -Lactamase Inhibitors: An Overview. *Cold Spring Harb Perspect Med* 2016, 6 (8).



<https://doi.org/10.1101/cshperspect.a025247>.

- (49) Abraham, E. P.; Chain, E. An Enzyme from Bacteria Able to Destroy Penicillin. *Nature* 1940, 146 (3713), 837–837. <https://doi.org/10.1038/146837a0>.
- (50) Massova, I.; Mobashery, S. Kinship and Diversification of Bacterial Penicillin-Binding Proteins and Beta-Lactamases. *Antimicrob. Agents Chemother.* 1998, 42 (1), 1–17. <https://doi.org/10.1128/AAC.42.1.1>.
- (51) Bush, K. Past and Present Perspectives on  $\beta$ -Lactamases. *Antimicrob. Agents Chemother.* 2018, 62 (10). <https://doi.org/10.1128/AAC.01076-18>.
- (52) Tooke, C. L.; Hinchliffe, P.; Bragginton, E. C.; Colenso, C. K.; Hirvonen, V. H. A.; Takebayashi, Y.; Spencer, J.  $\beta$ -Lactamases and  $\beta$ -Lactamase Inhibitors in the 21st Century. *J. Mol. Biol.* 2019, 431 (18), 3472–3500. <https://doi.org/10.1016/j.jmb.2019.04.002>.
- (53) Bush, K.; Jacoby, G. A.; Medeiros, A. A. A Functional Classification Scheme for Beta-Lactamases and Its Correlation with Molecular Structure. *Antimicrob Agents Chemother* 1995, 39 (6), 1211–1233.
- (54) Bush, K.; Jacoby, G. A. Updated Functional Classification of  $\beta$ -Lactamases. *Antimicrobial Agents and Chemotherapy* 2010, 54 (3), 969–976. <https://doi.org/10.1128/AAC.01009-09>.
- (55) Saini, A.; Bansal, R. Insights on the Structural Characteristics of NDM-1: The Journey so Far. *Advances in Biological Chemistry* 2012, 2 (4), 323–334. <https://doi.org/10.4236/abc.2012.24040>.

- (56) Smith, C. A.; Antunes, N. T.; Stewart, N. K.; Toth, M.; Kumarasiri, M.; Chang, M.; Mobashery, S.; Vakulenko, S. B. Structural Basis for Carbapenemase Activity of the OXA-23  $\beta$ -Lactamase from *Acinetobacter Baumannii*. *Chem Biol* 2013, 20 (9), 1107–1115. <https://doi.org/10.1016/j.chembiol.2013.07.015>.
- (57) Pemberton, O. A.; Zhang, X.; Chen, Y. Molecular Basis of Substrate Recognition and Product Release by the *Klebsiella Pneumoniae* Carbapenemase (KPC-2). *J Med Chem* 2017, 60 (8), 3525–3530. <https://doi.org/10.1021/acs.jmedchem.7b00158>.
- (58) Powers, R. A.; Shoichet, B. K. Structure-Based Approach for Binding Site Identification on AmpC Beta-Lactamase. *J Med Chem* 2002, 45 (15), 3222–3234. <https://doi.org/10.1021/jm020002p>.
- (59) Ullah, J. H.; Walsh, T. R.; Taylor, I. A.; Emery, D. C.; Verma, C. S.; Gamblin, S. J.; Spencer, J. The Crystal Structure of the L1 Metallo- $\beta$ -Lactamase from *Stenotrophomonas Maltophilia* at 1.7 Å Resolution. Edited by K. Nagai. *Journal of Molecular Biology* 1998, 284 (1), 125–136. <https://doi.org/10.1006/jmbi.1998.2148>.
- (60) Bush, K.; Bradford, P. A. Epidemiology of  $\beta$ -Lactamase-Producing Pathogens. *Clin Microbiol Rev* 2020, 33 (2), e00047-19. <https://doi.org/10.1128/CMR.00047-19>.
- (61) Meini, M.-R.; Llarrull, L. I.; Vila, A. J. Overcoming Differences: The Catalytic Mechanism of Metallo- $\beta$ -Lactamases. *FEBS Lett* 2015, 589 (22),

- 3419–3432. <https://doi.org/10.1016/j.febslet.2015.08.015>.
- (62) Llarrull, L. I.; Tioni, M. F.; Kowalski, J.; Bennett, B.; Vila, A. J. Evidence for a Dinuclear Active Site in the Metallo-Beta-Lactamase BcII with Substoichiometric Co(II). A New Model for Metal Uptake. *J Biol Chem* 2007, 282 (42), 30586–30595. <https://doi.org/10.1074/jbc.M704613200>.
- (63) Hawk, M. J.; Breece, R. M.; Hajdin, C. E.; Bender, K. M.; Hu, Z.; Costello, A. L.; Bennett, B.; Tierney, D. L.; Crowder, M. W. Differential Binding of Co(II) and Zn(II) to Metallo-Beta-Lactamase Bla2 from Bacillus Anthracis. *J Am Chem Soc* 2009, 131 (30), 10753–10762. <https://doi.org/10.1021/ja900296u>.
- (64) Carfi, A.; Duée, E.; Galleni, M.; Frère, J. M.; Dideberg, O. 1.85 Å Resolution Structure of the Zinc (II) Beta-Lactamase from Bacillus Cereus. *Acta Crystallogr D Biol Crystallogr* 1998, 54 (Pt 3), 313–323. <https://doi.org/10.1107/s0907444997010627>.
- (65) González, J. M.; Meini, M.-R.; Tomatis, P. E.; Medrano Martín, F. J.; Cricco, J. A.; Vila, A. J. Metallo-β-Lactamases Withstand Low Zn(II) Conditions by Tuning Metal-Ligand Interactions. *Nat Chem Biol* 2012, 8 (8), 698–700. <https://doi.org/10.1038/nchembio.1005>.
- (66) Queenan, A. M.; Bush, K. Carbapenemases: The Versatile β-Lactamases. *Clinical Microbiology Reviews* 2007, 20 (3), 440–458. <https://doi.org/10.1128/CMR.00001-07>.
- (67) Garau, G.; García-Sáez, I.; Bebrone, C.; Anne, C.; Mercuri, P.; Galleni, M.;

- Frère, J.-M.; Dideberg, O. Update of the Standard Numbering Scheme for Class B  $\beta$ -Lactamases. *Antimicrobial Agents and Chemotherapy* 2004, 48 (7), 2347. <https://doi.org/10.1128/AAC.48.7.2347-2349.2004>.
- (68) Boyd, S. E.; Livermore, D. M.; Hooper, D. C.; Hope, W. W. Metallo- $\beta$ -Lactamases: Structure, Function, Epidemiology, Treatment Options, and the Development Pipeline. *Antimicrob Agents Chemother* 2020, 64 (10), e00397-20. <https://doi.org/10.1128/AAC.00397-20>.
- (69) Crunkhorn, S. Inhibiting Metallo- $\beta$ -Lactamases. *Nat Rev Drug Discov* 2022, 21 (2), 98. <https://doi.org/10.1038/d41573-022-00009-1>.
- (70) Salahuddin, P.; Kumar, A.; Khan, A. U. Structure, Function of Serine and Metallo- $\beta$ -Lactamases and Their Inhibitors. *Curr. Protein Pept. Sci.* 2018, 19 (2), 130–144. <https://doi.org/10.2174/0929866524666170724160623>.
- (71) Behzadi, P.; García-Perdomo, H. A.; Karpiński, T. M.; Issakhanian, L. Metallo- $\beta$ -Lactamases: A Review. *Mol Biol Rep* 2020, 47 (8), 6281–6294. <https://doi.org/10.1007/s11033-020-05651-9>.
- (72) Jeffs, M. A.; Lohans, C. T. Inhibiting the Metallo- $\beta$ -Lactamases: Challenges and Strategies to Overcome Bacterial  $\beta$ -Lactam Resistance. *Future Medicinal Chemistry* 2022, 14 (14), 1021–1025. <https://doi.org/10.4155/fmc-2022-0097>.
- (73) Jdc; Slk; Gjd-h. Ralph D. Feigin, M.D.: April 3, 1938–August 14, 2008. In *Feigin and Cherry's Textbook of Pediatric Infectious Diseases (Sixth Edition)*; Feigin, R. D., Cherry, J. D., Demmler-Harrison, G. J., Kaplan, S. L., Eds.;

- W.B. Saunders: Philadelphia, 2009; p v. <https://doi.org/10.1016/B978-1-4160-4044-6.50002-9>.
- (74) Medeiros, A. A. Evolution and Dissemination of  $\beta$ -Lactamases Accelerated by Generations of  $\beta$ -Lactam Antibiotics. *Clinical Infectious Diseases* 1997, 24 (Supplement\_1), S19–S45. [https://doi.org/10.1093/clinids/24.Supplement\\_1.S19](https://doi.org/10.1093/clinids/24.Supplement_1.S19).
- (75) Jacobs, C.; Joris, B.; Jamin, M.; Klarsov, K.; van Beeumen, J.; Mengin-Lecreulx, D.; van Heijenoort, J.; Park, J. T.; Normark, S.; Frère, J.-M. AmpD, Essential for Both  $\beta$ -Lactamase Regulation and Cell Wall Recycling, Is a Novel Cytosolic N-Acetylmuramyl-L-Alanine Amidase. *Molecular Microbiology* 1995, 15 (3), 553–559. <https://doi.org/10.1111/j.1365-2958.1995.tb02268.x>.
- (76) Drawz, S. M.; Bonomo, R. A. Three Decades of Beta-Lactamase Inhibitors. *Clin. Microbiol. Rev.* 2010, 23 (1), 160–201. <https://doi.org/10.1128/CMR.00037-09>.
- (77) Leonard, D. A.; Bonomo, R. A.; Powers, R. A. Class D  $\beta$ -Lactamases: A Reappraisal after Five Decades. *Acc Chem Res* 2013, 46 (11), 2407–2415. <https://doi.org/10.1021/ar300327a>.
- (78) Evans, B. A.; Amyes, S. G. B. OXA  $\beta$ -Lactamases. *Clin Microbiol Rev* 2014, 27 (2), 241–263. <https://doi.org/10.1128/CMR.00117-13>.
- (79) Toth, M.; Antunes, N. T.; Stewart, N. K.; Frase, H.; Bhattacharya, M.; Smith, C. A.; Vakulenko, S. B. Class D  $\beta$ -Lactamases Do Exist in Gram-

- Positive Bacteria. *Nat Chem Biol* 2016, 12 (1), 9–14.  
<https://doi.org/10.1038/nchembio.1950>.
- (80) Denton, M.; Kerr, K. G. Microbiological and Clinical Aspects of Infection Associated with *Stenotrophomonas Maltophilia*. *Clin Microbiol Rev* 1998, 11 (1), 57–80.
- (81) Mojica, M. F.; Rutter, J. D.; Taracila, M.; Abriata, L. A.; Fouts, D. E.; Papp-Wallace, K. M.; Walsh, T. J.; LiPuma, J. J.; Vila, A. J.; Bonomo, R. A. Population Structure, Molecular Epidemiology, and  $\beta$ -Lactamase Diversity among *Stenotrophomonas Maltophilia* Isolates in the United States. *mBio* 2019, 10 (4). <https://doi.org/10.1128/mBio.00405-19>.
- (82) Adegoke, A. A.; Stenström, T. A.; Okoh, A. I. *Stenotrophomonas Maltophilia* as an Emerging Ubiquitous Pathogen: Looking Beyond Contemporary Antibiotic Therapy. *Front. Microbiol.* 2017, 8. <https://doi.org/10.3389/fmicb.2017.02276>.
- (83) Brooke, J. S. *Stenotrophomonas Maltophilia*: An Emerging Global Opportunistic Pathogen. *Clinical Microbiology Reviews* 2012, 25 (1), 2–41. <https://doi.org/10.1128/CMR.00019-11>.
- (84) Rello, J.; Kalwaje Eshwara, V.; Lagunes, L.; Alves, J.; Wunderink, R. G.; Conway-Morris, A.; Rojas, J. N.; Alp, E.; Zhang, Z. A Global Priority List of the TOP TEN Resistant Microorganisms (TOTEM) Study at Intensive Care: A Prioritization Exercise Based on Multi-Criteria Decision Analysis. *Eur J Clin Microbiol Infect Dis* 2019, 38 (2), 319–323.

- <https://doi.org/10.1007/s10096-018-3428-y>.
- (85) Gil-Gil, T.; Martínez, J. L.; Blanco, P. Mechanisms of Antimicrobial Resistance in *Stenotrophomonas Maltophilia*: A Review of Current Knowledge. *Expert Review of Anti-infective Therapy* 2020, 18 (4), 335–347. <https://doi.org/10.1080/14787210.2020.1730178>.
- (86) Alonso, A.; Rojo, F.; Martínez, J. L. Environmental and Clinical Isolates of *Pseudomonas Aeruginosa* Show Pathogenic and Biodegradative Properties Irrespective of Their Origin. *Environ Microbiol* 1999, 1 (5), 421–430. <https://doi.org/10.1046/j.1462-2920.1999.00052.x>.
- (87) Berg, G.; Martinez, J. L. Friends or Foes: Can We Make a Distinction between Beneficial and Harmful Strains of the *Stenotrophomonas Maltophilia* Complex? *Front Microbiol* 2015, 6, 241. <https://doi.org/10.3389/fmicb.2015.00241>.
- (88) Steinmann, J.; Mamat, U.; Abda, E. M.; Kirchhoff, L.; Streit, W. R.; Schaible, U. E.; Niemann, S.; Kohl, T. A. Analysis of Phylogenetic Variation of *Stenotrophomonas Maltophilia* Reveals Human-Specific Branches. *Front Microbiol* 2018, 9, 806. <https://doi.org/10.3389/fmicb.2018.00806>.
- (89) Walsh, T. R.; Hall, L.; Assinder, S. J.; Nichols, W. W.; Cartwright, S. J.; MacGowan, A. P.; Bennett, P. M. Sequence Analysis of the L1 Metallo-Beta-Lactamase from *Xanthomonas Maltophilia*. *Biochim Biophys Acta* 1994, 1218 (2), 199–201. [https://doi.org/10.1016/0167-4781\(94\)90011-6](https://doi.org/10.1016/0167-4781(94)90011-6).
- (90) Balasubramanian, D.; Kumari, H.; Mathee, K. *Pseudomonas Aeruginosa*

- AmpR: An Acute–Chronic Switch Regulator. *Pathog Dis* 2015, 73 (2), 1–14.**  
**<https://doi.org/10.1111/2049-632X.12208>.**
- (91) **Okazaki, A.; Avison, M. B. Induction of L1 and L2  $\beta$ -Lactamase Production in *Stenotrophomonas Maltophilia* Is Dependent on an AmpR-Type Regulator. *Antimicrob Agents Chemother* 2008, 52 (4), 1525–1528.**  
**<https://doi.org/10.1128/AAC.01485-07>.**
- (92) **Lin, C.-W.; Huang, Y.-W.; Hu, R.-M.; Chiang, K.-H.; Yang, T.-C. The Role of AmpR in Regulation of L1 and L2  $\beta$ -Lactamases in *Stenotrophomonas Maltophilia*. *Research in Microbiology* 2009, 160 (2), 152–158.**  
**<https://doi.org/10.1016/j.resmic.2008.11.001>.**
- (93) **Paton, R.; Miles, R. S.; Amyes, S. G. Biochemical Properties of Inducible Beta-Lactamases Produced from *Xanthomonas Maltophilia*. *Antimicrob Agents Chemother* 1994, 38 (9), 2143–2149.**
- (94) **Simm, A. M.; Higgins, C. S.; Carenbauer, A. L.; Crowder, M. W.; Bateson, J. H.; Bennett, P. M.; Clarke, A. R.; Halford, S. E.; Walsh, T. R. Characterization of Monomeric L1 Metallo- $\beta$ -Lactamase and the Role of the N-Terminal Extension in Negative Cooperativity and Antibiotic Hydrolysis\*. *Journal of Biological Chemistry* 2002, 277 (27), 24744–24752.**  
**<https://doi.org/10.1074/jbc.M201524200>.**
- (95) **Yang, K. W.; Crowder, M. W. Inhibition Studies on the Metallo- $\beta$ -Lactamase L1 from *Stenotrophomonas Maltophilia*. *Archives of Biochemistry and Biophysics* 1999, 368 (1), 1–6.**



<https://doi.org/10.1006/abbi.1999.1293>.

- (96) Hu, Z.; Periyannan, G.; Bennett, B.; Crowder, M. W. Role of the Zn1 and Zn2 Sites in Metallo- $\beta$ -Lactamase L1. *J. Am. Chem. Soc.* 2008, 130 (43), 14207–14216. <https://doi.org/10.1021/ja8035916>.
- (97) Saaranen, M. J.; Ruddock, L. W. Applications of Catalyzed Cytoplasmic Disulfide Bond Formation. *Biochem Soc Trans* 2019, 47 (5), 1223–1231. <https://doi.org/10.1042/BST20190088>.
- (98) Carenbauer, A. L.; Garrity, J. D.; Periyannan, G.; Yates, R. B.; Crowder, M. W. Probing Substrate Binding to Metallo- $\beta$ -Lactamase L1 from *Stenotrophomonas Maltophilia* by Using Site-Directed Mutagenesis. *BMC Biochemistry* 2002, 3 (1), 4. <https://doi.org/10.1186/1471-2091-3-4>.
- (99) Nauton, L.; Kahn, R.; Garau, G.; Hernandez, J. F.; Dideberg, O. Structural Insights into the Design of Inhibitors for the L1 Metallo- $\beta$ -Lactamase from *Stenotrophomonas Maltophilia*. *Journal of Molecular Biology* 2008, 375 (1), 257–269. <https://doi.org/10.1016/j.jmb.2007.10.036>.
- (100) Spencer, J.; Read, J.; Sessions, R. B.; Howell, S.; Blackburn, G. M.; Gamblin, S. J. Antibiotic Recognition by Binuclear Metallo- $\beta$ -Lactamases Revealed by X-Ray Crystallography. *J. Am. Chem. Soc.* 2005, 127 (41), 14439–14444. <https://doi.org/10.1021/ja0536062>.
- (101) Mojica, M. F.; Papp-Wallace, K. M.; Taracila, M. A.; Barnes, M. D.; Rutter, J. D.; Jacobs, M. R.; LiPuma, J. J.; Walsh, T. J.; Vila, A. J.; Bonomo, R. A. Avibactam Restores the Susceptibility of Clinical Isolates of

- Stenotrophomonas Maltophilia* to Aztreonam. *Antimicrob Agents Chemother* 2017, 61 (10), e00777-17. <https://doi.org/10.1128/AAC.00777-17>.
- (102) Walsh, T. R.; MacGowan, A. P.; Bennett, P. M. Sequence Analysis and Enzyme Kinetics of the L2 Serine Beta-Lactamase from *Stenotrophomonas Maltophilia*. *Antimicrobial Agents and Chemotherapy* 1997, 41 (7), 1460–1464. <https://doi.org/10.1128/AAC.41.7.1460>.
- (103) Okazaki, A.; Avison, M. B. Induction of L1 and L2  $\beta$ -Lactamase Production in *Stenotrophomonas Maltophilia* Is Dependent on an AmpR-Type Regulator. *Antimicrobial Agents and Chemotherapy* 2008, 52 (4), 1525–1528. <https://doi.org/10.1128/AAC.01485-07>.
- (104) Calvopiña, K.; Hinchliffe, P.; Brem, J.; Heesom, K. J.; Johnson, S.; Cain, R.; Lohans, C. T.; Fishwick, C. W. G.; Schofield, C. J.; Spencer, J.; Avison, M. B. Structural/Mechanistic Insights into the Efficacy of Nonclassical  $\beta$ -Lactamase Inhibitors against Extensively Drug Resistant *Stenotrophomonas Maltophilia* Clinical Isolates. *Mol Microbiol* 2017, 106 (3), 492–504. <https://doi.org/10.1111/mmi.13831>.
- (105) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural Summaries of PDB Entries. *Protein Sci* 2018, 27 (1), 129–134. <https://doi.org/10.1002/pro.3289>.
- (106) Ke, W.; Bethel, C. R.; Thomson, J. M.; Bonomo, R. A.; van den Akker, F. Crystal Structure of KPC-2: Insights into Carbapenemase Activity in Class A Beta-Lactamases. *Biochemistry* 2007, 46 (19), 5732–5740.

<https://doi.org/10.1021/bi700300u>.

- (107) Alder, B. J.; Wainwright, T. E. Phase Transition for a Hard Sphere System. *J. Chem. Phys.* 1957, 27 (5), 1208–1209. <https://doi.org/10.1063/1.1743957>.
- (108) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* 1977, 267 (5612), 585–590. <https://doi.org/10.1038/267585a0>.
- (109) Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and N-Alkane Molecules. *J. Chem. Phys.* 1968, 49 (11), 5116–5129. <https://doi.org/10.1063/1.1670007>.
- (110) Levitt, M.; Lifson, S. Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *Journal of Molecular Biology* 1969, 46 (2), 269–279. [https://doi.org/10.1016/0022-2836\(69\)90421-5](https://doi.org/10.1016/0022-2836(69)90421-5).
- (111) Van Drie, J. H. Computer-Aided Drug Design: The next 20 Years. *J Comput Aided Mol Des* 2007, 21 (10), 591–601. <https://doi.org/10.1007/s10822-007-9142-y>.
- (112) Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J. Non-Peptide Fibrinogen Receptor Antagonists. 1. Discovery and Design of Exosite Inhibitors. *J. Med. Chem.* 1992, 35 (24), 4640–4642. <https://doi.org/10.1021/jm00102a020>.
- (113) Muegge, I.; Bergner, A.; Kriegl, J. M. Computer-Aided Drug Design at Boehringer Ingelheim. *J Comput Aided Mol Des* 2017, 31 (3), 275–285.

<https://doi.org/10.1007/s10822-016-9975-3>.

- (114) Gibbs, A. C. Elements and Modulation of Functional Dynamics. *J. Med. Chem.* 2014, 57 (19), 7819–7837. <https://doi.org/10.1021/jm500325k>.
- (115) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* 2007, 450 (7172), 964–972. <https://doi.org/10.1038/nature06522>.
- (116) Im, W.; Roux, B. Ion Permeation and Selectivity of OmpF Porin: A Theoretical Study Based on Molecular Dynamics, Brownian Dynamics, and Continuum Electrodifffusion Theory. *Journal of Molecular Biology* 2002, 322 (4), 851–869. [https://doi.org/10.1016/S0022-2836\(02\)00778-7](https://doi.org/10.1016/S0022-2836(02)00778-7).
- (117) Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular Dynamics Simulations and Novel Drug Discovery. *Expert Opin Drug Discov* 2018, 13 (1), 23–37. <https://doi.org/10.1080/17460441.2018.1403419>.
- (118) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* 2002, 9 (9), 646–652. <https://doi.org/10.1038/nsb0902-646>.
- (119) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* 2018, 99 (6), 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>.
- (120) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Cryst D* 2012, 68 (4), 352–367.

<https://doi.org/10.1107/S0907444912001308>.

- (121) Koehl, A.; Hu, H.; Maeda, S.; Zhang, Y.; Qu, Q.; Paggi, J. M.; Latorraca, N. R.; Hilger, D.; Dawson, R.; Matile, H.; Schertler, G. F. X.; Granier, S.; Weis, W. I.; Dror, R. O.; Manglik, A.; Skiniotis, G.; Kobilka, B. K. Structure of the M-Opioid Receptor–G i Protein Complex. *Nature* 2018, 558 (7711), 547–552. <https://doi.org/10.1038/s41586-018-0219-7>.
- (122) Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* 2016, 12 (6), 2990–2998. <https://doi.org/10.1021/acs.jctc.6b00201>.
- (123) Wu, X.; Xu, L.-Y.; Li, E.-M.; Dong, G. Application of Molecular Dynamics Simulation in Biomedicine. *Chem Biol Drug Des* 2022, 99 (5), 789–800. <https://doi.org/10.1111/cbdd.14038>.
- (124) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* 2019, 1 (1), 5957–5957. <https://doi.org/10.33011/livecoms.1.1.5957>.
- (125) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* 2015, 11 (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.

- (126) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* 2010, 6 (9), 2935–2947. <https://doi.org/10.1021/ct1002626>.
- (127) CASE, D. A.; CHEATHAM, T. E.; DARDEN, T.; GOHLKE, H.; LUO, R.; MERZ, K. M.; ONUFRIEV, A.; SIMMERLING, C.; WANG, B.; WOODS, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* 2005, 26 (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.
- (128) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* 1995, 91 (1), 1–41. [https://doi.org/10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D).
- (129) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O’Hearn, K. A.;

Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.;  
Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.;  
Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.;  
Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber  
2021*; University of California, San Francisco, 2021.

(130) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.;  
Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cisneros, G. A.; Cruzeiro, V. W.  
D.; Darden, T. A.; Forouzes, N.; Giambasu, G.; Giese, T.; Gilson, M. K.;  
Gohlke, H.; Goetz, A. W.; Harris, J.; Izadi, S.; Izmailov, S. A.; Kasavajhala,  
K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T.; Li, P.;  
Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga,  
M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.;  
O’Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.;  
Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen,  
J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R.  
C.; Wang, J.; Wang, J.; Wei, H.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.;  
Zhao, S.; Zhu, Q.; Kollman, P. A. *Amber 2023*; University of California, San  
Francisco, 2023.

(131) Eshtiwi, A. A.; Rathbone, D. L. A Modified Bonded Model Approach for  
Molecular Dynamics Simulations of New Delhi Metallo- $\beta$ -Lactamase. *J Mol  
Graph Model* 2023, 121, 108431. <https://doi.org/10.1016/j.jmglm.2023.108431>.

(132) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Jr. Force Field Design for

- Metalloproteins. *J. Am. Chem. Soc.* 1991, 113 (22), 8262–8270.  
<https://doi.org/10.1021/ja00022a010>.
- (133) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C. L. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J Phys Chem B* 2014, 118 (16), 4351–4362.  
<https://doi.org/10.1021/jp501737x>.
- (134) Rahnamoun, A.; O’Hearn, K. A.; Kaymak, M. C.; Li, Z.; Merz, K. M.; Aktulga, H. M. A Polarizable Cationic Dummy Metal Ion Model. *J Phys Chem Lett* 2022, 5334–5340. <https://doi.org/10.1021/acs.jpcclett.2c01279>.
- (135) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *International Journal of Quantum Chemistry* 1996, 60 (7), 1271–1277. [https://doi.org/10.1002/\(SICI\)1097-461X\(1996\)60:7<1271::AID-QUA8>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W).
- (136) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry* 1995, 16 (11), 1357–1377. <https://doi.org/10.1002/jcc.540161106>.
- (137) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *Journal of Computational Chemistry* 1984, 5 (2), 129–145. <https://doi.org/10.1002/jcc.540050204>.
- (138) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple AMBER Force Fields and Development of



- Improved Protein Backbone Parameters. *Proteins* 2006, 65 (3), 712–725.  
<https://doi.org/10.1002/prot.21123>.
- (139) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E. I.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* 2011, 7 (9), 2886–2902.  
<https://doi.org/10.1021/ct200162x>.
- (140) Baig, M. H.; Ahmad, K.; Roy, S.; Ashraf, J. M.; Adil, M.; Siddiqui, M. H.; Khan, S.; Kamal, M. A.; Provazník, I.; Choi, I. Computer Aided Drug Design: Success and Limitations. *Curr Pharm Des* 2016, 22 (5), 572–581.  
<https://doi.org/10.2174/1381612822666151125000550>.
- (141) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational science (San Diego, Calif.); Academic Press: San Diego, 2002.
- (142) Heath Turner, C.; Brennan, J. K.; Lísal, M.; Smith, W. R.; Karl Johnson, J.; Gubbins, K. E. Simulation of Chemical Reaction Equilibria by the Reaction Ensemble Monte Carlo Method: A Review†. *Molecular Simulation* 2008, 34 (2), 119–146. <https://doi.org/10.1080/08927020801986564>.
- (143) Isaacs, A. *Oxford Dictionary of Physics*; 1996.
- (144) Bryan, G. H. Elementary Principles in Statistical Mechanics. *Nature* 1902, 66 (1708), 291–292. <https://doi.org/10.1038/066291a0>.
- (145) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations

- in Drug Discovery. *J Comput Aided Mol Des* 2012, 26 (1), 15–26.  
<https://doi.org/10.1007/s10822-011-9517-y>.
- (146) Spahn, V.; Vecchio, G. D.; Labuz, D.; Rodriguez-Gaztelumendi, A.; Massaly, N.; Temp, J.; Durmaz, V.; Sabri, P.; Reidelbach, M.; Machelska, H.; Weber, M.; Stein, C. A Nontoxic Pain Killer Designed by Modeling of Pathological Receptor Conformations. *Science* 2017, 355 (6328), 966–969.  
<https://doi.org/10.1126/science.aai8636>.
- (147) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; Baell, J. B.; Sexton, P. M.; Christopoulos, A.; Shaw, D. E. Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs. *Nature* 2013, 503 (7475), 295–299. <https://doi.org/10.1038/nature12595>.
- (148) Hilger, D.; Masureel, M.; Kobilka, B. K. Structure and Dynamics of GPCR Signaling Complexes. *Nature Structural & Molecular Biology* 2018, 25 (1), 4–12. <https://doi.org/10.1038/s41594-017-0011-7>.
- (149) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* 2013, 9 (9), 3878–3888. <https://doi.org/10.1021/ct400314y>.
- (150) Stone, J. E.; Hallock, M. J.; Phillips, J. C.; Peterson, J. R.; Luthey-Schulten, Z.; Schulten, K. Evaluation of Emerging Energy-Efficient Heterogeneous Computing Platforms for Biomolecular and Cellular Simulation Workloads.

- In *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*; 2016; pp 89–100.  
<https://doi.org/10.1109/IPDPSW.2016.130>.
- (151) William Humphrey (last); Andrew Dalke; Klaus Schulten. VMD -- Visual Molecular Dynamics. *Journal of Molecular Graphics* 1996, 14, 33–38.
- (152) Muhammed, M. T.; Aki-Yalcin, E. Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chemical Biology & Drug Design* 2019, 93 (1), 12–20. <https://doi.org/10.1111/cbdd.13388>.
- (153) Baker, M. Making Membrane Proteins for Structures: A Trillion Tiny Tweaks. *Nat Methods* 2010, 7 (6), 429–434.  
<https://doi.org/10.1038/nmeth0610-429>.
- (154) Boswell, Z. K.; Latham, M. P. Methyl-Based NMR Spectroscopy Methods for Uncovering Structural Dynamics in Large Proteins and Protein Complexes. *Biochemistry* 2019, 58 (3), 144–155.  
<https://doi.org/10.1021/acs.biochem.8b00953>.
- (155) Jiang, Y.; Kalodimos, C. G. NMR Studies of Large Proteins. *J Mol Biol* 2017, 429 (17), 2667–2676. <https://doi.org/10.1016/j.jmb.2017.07.007>.
- (156) Yalcin-Ozkat, G. Molecular Modeling Strategies of Cancer Multidrug Resistance. *Drug Resist Updat* 2021, 59, 100789.  
<https://doi.org/10.1016/j.drup.2021.100789>.
- (157) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discov Today* 2009, 14 (13–14), 676–

683. <https://doi.org/10.1016/j.drudis.2009.04.006>.
- (158) Werner, T.; Morris, M. B.; Dastmalchi, S.; Church, W. B. Structural Modelling and Dynamics of Proteins for Insights into Drug Interactions. *Adv Drug Deliv Rev* 2012, 64 (4), 323–343. <https://doi.org/10.1016/j.addr.2011.11.011>.
- (159) Petrey, D.; Xiang, Z.; Tang, C. L.; Xie, L.; Gimpelev, M.; Mitros, T.; Soto, C. S.; Goldsmith-Fischman, S.; Kernytsky, A.; Schlessinger, A.; Koh, I. Y. Y.; Alexov, E.; Honig, B. Using Multiple Structure Alignments, Fast Model Building, and Energetic Analysis in Fold Recognition and Homology Modeling. *Proteins* 2003, 53 Suppl 6, 430–435. <https://doi.org/10.1002/prot.10550>.
- (160) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* 1993, 234 (3), 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- (161) Zwanzig, R. From Classical Dynamics to Continuous Time Random Walks. *J Stat Phys* 1983, 30 (2), 255–262. <https://doi.org/10.1007/BF01012300>.
- (162) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *Journal of Computational Physics* 1999, 151 (1), 146–168. <https://doi.org/10.1006/jcph.1999.6231>.
- (163) Swope, W. C.; Pitera, J. W.; Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B* 2004, 108

- (21), 6571–6581. <https://doi.org/10.1021/jp037421y>.
- (164) Kampen, N. G. V. *Stochastic Processes in Physics and Chemistry*; Elsevier, 1992.
- (165) Singhal, N.; Snow, C. D.; Pande, V. S. Using Path Sampling to Build Better Markovian State Models: Predicting the Folding Rate and Mechanism of a Tryptophan Zipper Beta Hairpin. *J. Chem. Phys.* 2004, *121* (1), 415–425. <https://doi.org/10.1063/1.1738647>.
- (166) Krivov, S. V.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proceedings of the National Academy of Sciences* 2004, *101* (41), 14766–14770. <https://doi.org/10.1073/pnas.0406234101>.
- (167) Luty, B. A.; McCammon, J. A. Simulation of Bimolecular Reactions. *Molecular Simulation* 1993, *10* (1), 61–65. <https://doi.org/10.1080/08927029308022498>.
- (168) Karpen, M. E.; Tobias, D. J.; Brooks, C. L. I. Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2-Ns Trajectories of YPGDV. *Biochemistry* 1993, *32* (2), 412–420. <https://doi.org/10.1021/bi00053a005>.
- (169) Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *J. Chem. Theory Comput.* 2006, *2* (3), 840–857. <https://doi.org/10.1021/ct050162r>.
- (170) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science.

- J. Am. Chem. Soc.* 2018, 140 (7), 2386–2396.  
<https://doi.org/10.1021/jacs.7b12191>.
- (171) Rudzinski, J. F.; Kremer, K.; Bereau, T. Communication: Consistent Interpretation of Molecular Simulation Kinetics Using Markov State Models Biased with External Information. *J. Chem. Phys.* 2016, 144 (5), 051102. <https://doi.org/10.1063/1.4941455>.
- (172) Olsson, S.; Wu, H.; Paul, F.; Clementi, C.; Noé, F. Combining Experimental and Simulation Data of Molecular Processes via Augmented Markov Models. *Proceedings of the National Academy of Sciences* 2017, 114 (31), 8265–8270. <https://doi.org/10.1073/pnas.1704803114>.
- (173) Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S. Discovery of Multiple Hidden Allosteric Sites by Combining Markov State Models and Experiments. *Proceedings of the National Academy of Sciences* 2015, 112 (9), 2734–2739. <https://doi.org/10.1073/pnas.1417811112>.
- (174) Pinheiro Cinelli, L.; Araújo Marins, M.; Barros da Silva, E. A.; Lima Netto, S. Variational Autoencoder. In *Variational Methods for Machine Learning with Applications to Deep Networks*; Cinelli, L. P., Marins, M. A., Barros da Silva, E. A., Netto, S. L., Eds.; Springer International Publishing: Cham, 2021; pp 111–149. [https://doi.org/10.1007/978-3-030-70679-1\\_5](https://doi.org/10.1007/978-3-030-70679-1_5).
- (175) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinformatics* 2018, 19 (18), 484. <https://doi.org/10.1186/s12859-018-2507-5>.

- (176) Henrick, K.; Thornton, J. M. PQS: A Protein Quaternary Structure File Server. *Trends Biochem Sci* 1998, 23 (9), 358–361. [https://doi.org/10.1016/s0968-0004\(98\)01253-5](https://doi.org/10.1016/s0968-0004(98)01253-5).
- (177) Prejanò, M.; Romeo, I.; La Serra, M. A.; Russo, N.; Marino, T. Computational Study Reveals the Role of Water Molecules in the Inhibition Mechanism of LAT1 by 1,2,3-Dithiazoles. *J. Chem. Inf. Model.* 2021, 61 (12), 5883–5892. <https://doi.org/10.1021/acs.jcim.1c01012>.
- (178) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of PKs in Macromolecules. *Proteins* 2006, 63 (4), 928–938. <https://doi.org/10.1002/prot.20922>.
- (179) Yu, Z.; Li, P.; Merz, K. M. Jr. Extended Zinc AMBER Force Field (EZAFF). *J. Chem. Theory Comput.* 2018, 14 (1), 242–254. <https://doi.org/10.1021/acs.jctc.7b00773>.
- (180) Li, S.; Hong, M. Protonation, Tautomerization, and Rotameric Structure of Histidine: A Comprehensive Study by Magic-Angle-Spinning Solid-State NMR. *J Am Chem Soc* 2011, 133 (5), 1534–1544. <https://doi.org/10.1021/ja108943n>.
- (181) Kalani, M. R.; Moradi, A.; Moradi, M.; Tajkhorshid, E. Characterizing a Histidine Switch Controlling PH-Dependent Conformational Changes of the Influenza Virus Hemagglutinin. *Biophysical Journal* 2013, 105 (4), 993–1003. <https://doi.org/10.1016/j.bpj.2013.06.047>.

- (182) Kim, M. O.; Nichols, S. E.; Wang, Y.; McCammon, J. A. Effects of Histidine Protonation and Rotameric States on Virtual Screening of M. Tuberculosis RmlC. *J Comput Aided Mol Des* 2013, 27 (3), 235–246. <https://doi.org/10.1007/s10822-013-9643-9>.
- (183) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* 2016, 12 (4), 1845–1852. <https://doi.org/10.1021/acs.jctc.6b00049>.
- (184) Doerr, S.; De Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* 2014, 10 (5), 2064–2069. <https://doi.org/10.1021/ct400919u>.
- (185) Hargis, J. C.; Vankayala, S. L.; White, J. K.; Woodcock, H. L. Identification and Characterization of Noncovalent Interactions That Drive Binding and Specificity in DD-Peptidases and  $\beta$ -Lactamases. *J Chem Theory Comput* 2014, 10 (2), 855–864. <https://doi.org/10.1021/ct400968v>.
- (186) Langan, P. S.; Vandavasi, V. G.; Weiss, K. L.; Cooper, J. B.; Ginell, S. L.; Coates, L. The Structure of Toho1  $\beta$ -Lactamase in Complex with Penicillin Reveals the Role of Tyr105 in Substrate Recognition. *FEBS Open Bio* 2016, 6 (12), 1170–1177. <https://doi.org/10.1002/2211-5463.12132>.
- (187) Vanhove, M.; Raquet, X.; Palzkill, T.; Pain, R. H.; Frère, J. M. The Rate-Limiting Step in the Folding of the Cis-Pro167Thr Mutant of TEM-1 Beta-Lactamase Is the Trans to Cis Isomerization of a Non-Proline Peptide Bond. *Proteins* 1996, 25 (1), 104–111. [https://doi.org/10.1002/\(SICI\)1097-](https://doi.org/10.1002/(SICI)1097-)



0134(199605)25:1<104::AID-PROT8>3.0.CO;2-J.

- (188) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J Chem Phys* 2011, *134* (17), 174105. <https://doi.org/10.1063/1.3565032>.
- (189) E, W.; Vanden-Eijnden, E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annual Review of Physical Chemistry* 2010, *61* (1), 391–420. <https://doi.org/10.1146/annurev.physchem.040808.090412>.
- (190) Pratt, L. R. A Statistical Method for Identifying Transition States in High Dimensional Problems. *The Journal of Chemical Physics* 1986, *85* (9), 5045–5048. <https://doi.org/10.1063/1.451695>.
- (191) Wigner, E. The Transition State Method. *Trans. Faraday Soc.* 1938, *34* (0), 29–41. <https://doi.org/10.1039/TF9383400029>.
- (192) Eyring, Henry. The Activated Complex and the Absolute Rate of Chemical Reactions. *Chem. Rev.* 1935, *17* (1), 65–77. <https://doi.org/10.1021/cr60056a006>.
- (193) Horiuti, J. On the Statistical Mechanical Treatment of the Absolute Rate of Chemical Reaction. *BCSJ* 1938, *13* (1), 210–216. <https://doi.org/10.1246/bcsj.13.210>.
- (194) Bolhuis, P. G. Transition-Path Sampling of  $\beta$ -Hairpin Folding. *Proceedings of the National Academy of Sciences* 2003, *100* (21), 12129–12134.

<https://doi.org/10.1073/pnas.1534924100>.

- (195) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Illustration of Transition Path Theory on a Collection of Simple Examples. *The Journal of Chemical Physics* 2006, 125 (8), 084110. <https://doi.org/10.1063/1.2335447>.
- (196) van der Maaten, L. J. P.; Hinton, G. E. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 2008, 9 (nov), 2579–2605.
- (197) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.* 2014, 10 (2), 511–517. <https://doi.org/10.1021/ct4008603>.
- (198) Costello, A.; Periyannan, G.; Yang, K.-W.; Crowder, M. W.; Tierney, D. L. Site-Selective Binding of Zn(II) to Metallo- $\beta$ -Lactamase L1 from *Stenotrophomonas Maltophilia*. *J Biol Inorg Chem* 2006, 11 (3), 351–358. <https://doi.org/10.1007/s00775-006-0083-z>.
- (199) Heinz, U.; Adolph, H.-W. Metallo- $\beta$ -Lactamases: Two Binding Sites for One Catalytic Metal Ion? *CMLS, Cell. Mol. Life Sci.* 2004, 61 (22), 2827–2839. <https://doi.org/10.1007/s00018-004-4214-9>.
- (200) H, S. K.; Jade, D.; Harrison, M. A.; Sugumar, S. Identification of Natural Inhibitor against L1  $\beta$ -Lactamase Present in *Stenotrophomonas Maltophilia*. *J Mol Model* 2022, 28 (11), 342. <https://doi.org/10.1007/s00894-022-05336-z>.
- (201) Liu, X.-L.; Yang, K.-W.; Zhang, Y.-J.; Ge, Y.; Xiang, Y.; Chang, Y.-N.;

- Oelschlaeger, P. Optimization of Amino Acid Thioesters as Inhibitors of Metallo- $\beta$ -Lactamase L1. *Bioorg Med Chem Lett* 2016, 26 (19), 4698–4701. <https://doi.org/10.1016/j.bmcl.2016.08.048>.
- (202) Yang, Y.; Yan, Y.-H.; Schofield, C. J.; McNally, A.; Zong, Z.; Li, G.-B. Metallo- $\beta$ -Lactamase-Mediated Antimicrobial Resistance and Progress in Inhibitor Discovery. *Trends Microbiol* 2023, 31 (7), 735–748. <https://doi.org/10.1016/j.tim.2023.01.013>.
- (203) Zhao, Z.; Shen, X.; Chen, S.; Gu, J.; Wang, H.; Mojica, M. F.; Samanta, M.; Bhowmik, D.; Vila, A. J.; Bonomo, R. A.; Haider, S. Gating Interactions Steer Loop Conformational Changes in the Active Site of the L1 Metallo- $\beta$ -Lactamase. *eLife* 2023, 12, e83928. <https://doi.org/10.7554/eLife.83928>.
- (204) Mojica, M. F.; Humphries, R.; Lipuma, J. J.; Mathers, A. J.; Rao, G. G.; Shelburne, S. A.; Fouts, D. E.; Van Duin, D.; Bonomo, R. A. Clinical Challenges Treating *Stenotrophomonas Maltophilia* Infections: An Update. *JAC Antimicrob Resist* 2022, 4 (3), dlac040. <https://doi.org/10.1093/jacamr/dlac040>.
- (205) Vakulenko, S. B.; Tóth, M.; Taibi, P.; Mobashery, S.; Lerner, S. A. Effects of Asp-179 Mutations in TEM<sub>pUC19</sub> Beta-Lactamase on Susceptibility to Beta-Lactams. *Antimicrob Agents Chemother* 1995, 39 (8), 1878–1880. <https://doi.org/10.1128/AAC.39.8.1878>.
- (206) Langan, P. S.; Vandavasi, V. G.; Weiss, K. L.; Cooper, J. B.; Ginell, S. L.; Coates, L. The Structure of Toho1  $\beta$ -Lactamase in Complex with Penicillin

- Reveals the Role of Tyr105 in Substrate Recognition. *FEBS Open Bio* 2016, 6 (12), 1170–1177. <https://doi.org/10.1002/2211-5463.12132>.
- (207) Hargis, J. C.; Vankayala, S. L.; White, J. K.; Woodcock, H. L. Identification and Characterization of Noncovalent Interactions That Drive Binding and Specificity in DD-Peptidases and  $\beta$ -Lactamases. *J Chem Theory Comput* 2014, 10 (2), 855–864. <https://doi.org/10.1021/ct400968v>.
- (208) Huang, M.-C.; Chen, W.-H.; Huang, C.-W.; Huang, K.-Y.; Horng, J.-C.; Hayashi, M.; Chen, I.-C. Investigation of the Cis–Trans Structures and Isomerization of Oligoprolines by Using Raman Spectroscopy and Density Functional Theory Calculations: Solute–Solvent Interactions and Effects of Terminal Positively Charged Amino Acid Residues. *RSC Adv.* 2020, 10 (57), 34493–34500. <https://doi.org/10.1039/D0RA05746K>.
- (209) Vanhove, M.; Raquet, X.; Palzkill, T.; Pain, R. H.; Frère, J.-M. The Rate-Limiting Step in the Folding of the Cis-Pro167Thr Mutant of TEM-1  $\beta$ -Lactamase Is the Trans to Cis Isomerization of a Non-Proline Peptide Bond. *Proteins: Structure, Function, and Bioinformatics* 1996, 25 (1), 104–111. [https://doi.org/10.1002/\(SICI\)1097-0134\(199605\)25:1<104::AID-PROT8>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0134(199605)25:1<104::AID-PROT8>3.0.CO;2-J).
- (210) Vanhove, M.; Raquet, X.; Frère, J. M. Investigation of the Folding Pathway of the TEM-1 Beta-Lactamase. *Proteins* 1995, 22 (2), 110–118. <https://doi.org/10.1002/prot.340220204>.
- (211) Sevaille, L.; Gavara, L.; Bebrone, C.; De Luca, F.; Nauton, L.; Achard, M.;

- Mercuri, P.; Tanfoni, S.; Borgianni, L.; Guyon, C.; Lonjon, P.; Turan-Zitouni, G.; Dzieciolowski, J.; Becker, K.; Bénard, L.; Condon, C.; Maillard, L.; Martinez, J.; Frère, J.-M.; Dideberg, O.; Galleni, M.; Docquier, J.-D.; Hernandez, J.-F. 1,2,4-Triazole-3-Thione Compounds as Inhibitors of Zinc Metallo- $\beta$ -Lactamases. *ChemMedChem* 2017, 12 (12), 972–985. <https://doi.org/10.1002/cmdc.201700186>.
- (212) Crisp, J.; Connors, R.; Garrity, J. D.; Carenbauer, A. L.; Crowder, M. W.; Spencer, J. Structural Basis for the Role of Asp-120 in Metallo-Beta-Lactamases. *Biochemistry* 2007, 46 (37), 10664–10674. <https://doi.org/10.1021/bi700707u>.
- (213) Gur, M.; Cheng, M. H.; Zomot, E.; Bahar, I. Effect of Dimerization on the Dynamics of Neurotransmitter: Sodium Symporters. *J. Phys. Chem. B* 2017, 121 (15), 3657–3666. <https://doi.org/10.1021/acs.jpcc.6b09876>.
- (214) Haider, S.; Grottesi, A.; Hall, B. A.; Ashcroft, F. M.; Sansom, M. S. P. Conformational Dynamics of the Ligand-Binding Domain of Inward Rectifier K Channels as Revealed by Molecular Dynamics Simulations: Toward an Understanding of Kir Channel Gating. *Biophys J* 2005, 88 (5), 3310–3320. <https://doi.org/10.1529/biophysj.104.052019>.
- (215) Hinchliffe, P., Calvopina, K., Spencer, J. L2 Class A Serine-Beta-Lactamase. *Molecular Microbiology* 2017. <https://doi.org/10.2210/pdb5ne2/pdb>.
- (216) Calvopiña, K.; Hinchliffe, P.; Brem, J.; Heesom, K. J.; Johnson, S.; Cain, R.; Lohans, C. T.; Fishwick, C. W. G.; Schofield, C. J.; Spencer, J.; Avison,

- M. B. Structural/Mechanistic Insights into the Efficacy of Nonclassical  $\beta$ -Lactamase Inhibitors against Extensively Drug Resistant *Stenotrophomonas Maltophilia* Clinical Isolates. *Mol Microbiol* 2017, 106 (3), 492–504. <https://doi.org/10.1111/mmi.13831>.**
- (217) Naas, T.; Oueslati, S.; Bonnin, R. A.; Dabos, M. L.; Zavala, A.; Dortet, L.; Retailleau, P.; Iorga, B. I. Beta-Lactamase Database (BLDB) – Structure and Function. *Journal of Enzyme Inhibition and Medicinal Chemistry* 2017, 32 (1), 917–919. <https://doi.org/10.1080/14756366.2017.1344235>.
- (218) Prasad, J. C.; Comeau, S. R.; Vajda, S.; Camacho, C. J. Consensus Alignment for Reliable Framework Prediction in Homology Modeling. *Bioinformatics* 2003, 19 (13), 1682–1691. <https://doi.org/10.1093/bioinformatics/btg211>.
- (219) Bienert, S.; Waterhouse, A.; de Beer, T. A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—New Features and Functionality. *Nucleic Acids Research* 2017, 45 (D1), D313–D319. <https://doi.org/10.1093/nar/gkw1132>.
- (220) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo—Distance Constraints Applied on Model Quality Estimation. *Bioinformatics* 2020, 36 (6), 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>.
- (221) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.;

- Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* 2018, 46 (W1), W296–W303. <https://doi.org/10.1093/nar/gky427>.
- (222) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res* 2003, 31 (13), 3381–3385.
- (223) Dougherty, T. J.; Thanassi, J. A.; Pucci, M. J. The Escherichia Coli Mutant Requiring D-Glutamic Acid Is the Result of Mutations in Two Distinct Genetic Loci. *J Bacteriol* 1993, 175 (1), 111–116.
- (224) Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 2.4.
- (225) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- (226) Kleiman, D. E.; Shukla, D. Multiagent Reinforcement Learning-Based Adaptive Sampling for Conformational Dynamics of Proteins. *J Chem Theory Comput* 2022, 18 (9), 5422–5434. <https://doi.org/10.1021/acs.jctc.2c00683>.
- (227) Hargis, J. C.; Vankayala, S. L.; White, J. K.; Woodcock, H. L. Identification and Characterization of Noncovalent Interactions That Drive Binding and Specificity in DD-Peptidases and  $\beta$ -Lactamases. *J. Chem. Theory Comput.* 2014, 10 (2), 855–864. <https://doi.org/10.1021/ct400968v>.
- (228) Martínez, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation

- Analysis. *PLOS ONE* 2015, 10 (3), e0119264.  
<https://doi.org/10.1371/journal.pone.0119264>.
- (229) Olehnovics, E.; Yin, J.; Pérez, A.; De Fabritiis, G.; Bonomo, R. A.; Bhowmik, D.; Haider, S. The Role of Hydrophobic Nodes in the Dynamics of Class A  $\beta$ -Lactamases. *Front Microbiol* 2021, 12, 720991.  
<https://doi.org/10.3389/fmicb.2021.720991>.
- (230) Galdadas, I.; Lovera, S.; Pérez-Hernández, G.; Barnes, M. D.; Healy, J.; Afsharikho, H.; Woodford, N.; Bonomo, R. A.; Gervasio, F. L.; Haider, S. Defining the Architecture of KPC-2 Carbapenemase: Identifying Allosteric Networks to Fight Antibiotics Resistance. *Sci Rep* 2018, 8 (1), 12916.  
<https://doi.org/10.1038/s41598-018-31176-0>.
- (231) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* 2006, 22 (21), 2695–2696.  
<https://doi.org/10.1093/bioinformatics/btl461>.
- (232) Okazaki, A.; Avison, M. B. Induction of L1 and L2 Beta-Lactamase Production in *Stenotrophomonas Maltophilia* Is Dependent on an AmpR-Type Regulator. *Antimicrob Agents Chemother* 2008, 52 (4), 1525–1528.  
<https://doi.org/10.1128/AAC.01485-07>.
- (233) Hu, R.-M.; Chiang, K.-H.; Chang, Y.-C.; Yang, T.-C. Characterization of the Charge Variants of L2 Beta-Lactamase in *Stenotrophomonas Maltophilia*. *J Med Microbiol* 2009, 58 (Pt 3), 318–321.



<https://doi.org/10.1099/jmm.0.000380-0>.

- (234) Chang, Y.-C.; Huang, Y.-W.; Chiang, K.-H.; Yang, T.-C.; Chung, T.-C. Introduction of an AmpR-L2 Intergenic Segment Attenuates the Induced Beta-Lactamase Activity of *Stenotrophomonas Maltophilia*. *Eur J Clin Microbiol Infect Dis* 2010, 29 (7), 887–890. <https://doi.org/10.1007/s10096-010-0924-0>.
- (235) Avison, M. B.; Higgins, C. S.; Ford, P. J.; von Heldreich, C. J.; Walsh, T. R.; Bennett, P. M. Differential Regulation of L1 and L2 Beta-Lactamase Expression in *Stenotrophomonas Maltophilia*. *J Antimicrob Chemother* 2002, 49 (2), 387–389. <https://doi.org/10.1093/jac/49.2.387>.
- (236) Sharma, R.; Jade, D.; Mohan, S.; Chandel, R.; Sugumar, S. In-Silico Virtual Screening for Identification of Potent Inhibitor for L2- $\beta$ -Lactamase from *Stenotrophomonas Maltophilia* through Molecular Docking, Molecular Dynamics Analysis Study. *Journal of Biomolecular Structure and Dynamics* 2021, 39 (18), 7123–7137. <https://doi.org/10.1080/07391102.2020.1805365>.
- (237) Huang, H.-H.; Chen, P.-Y.; Hu, R.-M.; Lin, Y.-T.; Li, L.-H.; Yang, T.-C. Impacts of L1 Promoter Variation and L2 Clavulanate Susceptibility on Ticarcillin-Clavulanate Susceptibility of *Stenotrophomonas Maltophilia*. *Antimicrob Agents Chemother* 2018, 62 (11), e01222-18. <https://doi.org/10.1128/AAC.01222-18>.
- (238) Mathieu, V.; Fastrez, J.; Soumillon, P. Engineering Allosteric Regulation into the Hinge Region of a Circularly Permuted TEM-1 Beta-Lactamase.

- Protein Eng Des Sel* 2010, 23 (9), 699–709.  
<https://doi.org/10.1093/protein/gzq041>.
- (239) Kim, J. R.; Ostermeier, M. Modulation of Effector Affinity by Hinge Region Mutations Also Modulates Switching Activity in an Engineered Allosteric TEM1 Beta-Lactamase Switch. *Arch Biochem Biophys* 2006, 446 (1), 44–51.  
<https://doi.org/10.1016/j.abb.2005.11.014>.
- (240) Biswal, S.; Caetano, K.; Jain, D.; Sarrila, A.; Munshi, T.; Dickman, R.; Tabor, A. B.; Rath, S. N.; Bhakta, S.; Ghosh, A. S. Antimicrobial Peptides Designed against the  $\Omega$ -Loop of Class A  $\beta$ -Lactamases to Potentiate the Efficacy of  $\beta$ -Lactam Antibiotics. *Antibiotics (Basel)* 2023, 12 (3), 553.  
<https://doi.org/10.3390/antibiotics12030553>.
- (241) Egorov, A.; Rubtsova, M.; Grigorenko, V.; Uporov, I.; Veselovsky, A. The Role of the  $\Omega$ -Loop in Regulation of the Catalytic Activity of TEM-Type  $\beta$ -Lactamases. *Biomolecules* 2019, 9 (12), 854.  
<https://doi.org/10.3390/biom9120854>.
- (242) Docquier, J.-D.; Mangani, S. An Update on  $\beta$ -Lactamase Inhibitor Discovery and Development. *Drug Resist Updat* 2018, 36, 13–29.  
<https://doi.org/10.1016/j.drug.2017.11.002>.
- (243) Papp-Wallace, K. M.; Bethel, C. R.; Distler, A. M.; Kasuboski, C.; Taracila, M.; Bonomo, R. A. Inhibitor Resistance in the KPC-2 Beta-Lactamase, a Preeminent Property of This Class A Beta-Lactamase. *Antimicrob Agents Chemother* 2010, 54 (2), 890–897. <https://doi.org/10.1128/AAC.00693-09>.

- (244) White, A. R.; Kaye, C.; Poupard, J.; Pypstra, R.; Woodnutt, G.; Wynne, B. Augmentin® (Amoxicillin/Clavulanate) in the Treatment of Community-Acquired Respiratory Tract Infection: A Review of the Continuing Development of an Innovative Antimicrobial Agent. *Journal of Antimicrobial Chemotherapy* 2004, 53 (suppl\_1), i3–i20. <https://doi.org/10.1093/jac/dkh050>.
- (245) Yu, H.; Dalby, P. A. Chapter Two - A Beginner's Guide to Molecular Dynamics Simulations and the Identification of Cross-Correlation Networks for Enzyme Engineering. In *Methods in Enzymology*; Tawfik, D. S., Ed.; Enzyme Engineering and Evolution: General Methods; Academic Press, 2020; Vol. 643, pp 15–49. <https://doi.org/10.1016/bs.mie.2020.04.020>.
- (246) Parwana, D.; Gu, J.; Wang, Q.; Bethel, C. R.; Marshall, E.; Hujer, A. M.; Bonomo, R. A.; Haider, S. The Structural Role of N170 in Substrate-Assisted Deacylation in KPC-2  $\beta$ -Lactamase, 2023. <https://doi.org/10.1101/2023.05.22.541848>.
- (247) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Current Opinion in Structural Biology* 2014, 25, 135–144. <https://doi.org/10.1016/j.sbi.2014.04.002>.
- (248) Noé, F.; De Fabritiis, G.; Clementi, C. Machine Learning for Protein Folding and Dynamics. *Curr Opin Struct Biol* 2020, 60, 77–84. <https://doi.org/10.1016/j.sbi.2019.12.005>.
- (249) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine Learning Approaches

- for Analyzing and Enhancing Molecular Dynamics Simulations. *Curr Opin Struct Biol* 2020, 61, 139–145. <https://doi.org/10.1016/j.sbi.2019.12.016>.
- (250) Stollar, E. J.; Smith, D. P. Uncovering Protein Structure. *Essays Biochem* 2020, 64 (4), 649–680. <https://doi.org/10.1042/EBC20190042>.
- (251) Jisna, V. A.; Jayaraj, P. B. Protein Structure Prediction: Conventional and Deep Learning Perspectives. *Protein J* 2021, 40 (4), 522–544. <https://doi.org/10.1007/s10930-021-10003-y>.
- (252) Tooke, C. L.; Hinchliffe, P.; Bonomo, R. A.; Schofield, C. J.; Mulholland, A. J.; Spencer, J. Natural Variants Modify Klebsiella Pneumoniae Carbapenemase (KPC) Acyl–Enzyme Conformational Dynamics to Extend Antibiotic Resistance. *J Biol Chem* 2020, 296, 100126. <https://doi.org/10.1074/jbc.RA120.016461>.
- (253) Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.; Mojica, M. F.; Agarwal, P. K.; Tooke, C. L.; Gervasio, F. L.; Spencer, J.; Bonomo, R. A.; Mulholland, A. J.; Haider, S. Allosteric Communication in Class A  $\beta$ -Lactamases Occurs via Cooperative Coupling of Loop Dynamics. *eLife* 2021, 10, e66567. <https://doi.org/10.7554/eLife.66567>.
- (254) Bonomo, R. A.  $\beta$ -Lactamases: A Focus on Current Challenges. *Cold Spring Harb Perspect Med* 2017, 7 (1). <https://doi.org/10.1101/cshperspect.a025239>.
- (255) McManus-Munoz, S.; Crowder, M. W. Kinetic Mechanism of Metallo-Beta-Lactamase L1 from *Stenotrophomonas Maltophilia*. *Biochemistry* 1999, 38 (5), 1547–1553. <https://doi.org/10.1021/bi9826512>.

- (256) Schultze, S.; Grubmüller, H. Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *J. Chem. Theory Comput.* 2021, 17 (9), 5766–5776. <https://doi.org/10.1021/acs.jctc.1c00273>.
- (257) Naithani, A.; Taylor, P.; Eрман, B.; Walkinshaw, M. D. A Molecular Dynamics Study of Allosteric Transitions in *Leishmania Mexicana* Pyruvate Kinase. *Biophys J* 2015, 109 (6), 1149–1156. <https://doi.org/10.1016/j.bpj.2015.05.040>.
- (258) Śledź, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Curr. Opin. Struct. Biol.* 2018, 48, 93–102. <https://doi.org/10.1016/j.sbi.2017.10.010>.