Exploring Ligand Binding in Glycine-Gated Ion Channels with Molecular Dynamics and Enhanced Sampling Methods

Clark, Jacob

Awarding institution:
King's College London

King’s College London

School of Natural and Mathematical Sciences – Department of Physics

Exploring Ligand Binding in Glycine-Gated Ion Channels with Molecular Dynamics and Enhanced Sampling Methods

Ph.D. in *Computational Biophysics*

Jacob Clark

Supervisor: *Professor Carla Molteni*

Second Supervisor: *Professor Lucia Sivilotti*

Contents

List of Figures **5**
List of Tables **13**
Abstract **14**
Chapter 1
Introduction **16**
1.1 The Nervous System 16
1.2 Pentameric Ligand Gated Ion Channels 20
1.2.1 pLGIC General Structure 20
1.2.2 pLGIC Functions 21
1.3 The GlyR 22
1.3.1 GlyR Function 23
1.3.2 GlyR Structure 25
1.3.3 GlyR Ligand Binding 27
1.3.4 Partial and Full Agonism 29
1.4 Mutants 32
1.5 Structural Biology 36
1.5.1 Imaging Techniques 36
1.5.2 Available Structures 37
1.6 Simulating Ion Channels 40
1.6.1 Simulations of GlyRs 41
1.6.2 Investigating GlyR Ligand Binding 42
Chapter 2
Methods **44**
2.1 Biomolecular simulations 44
2.1.1 Why Use Computational Methods 44
2.2 Molecular Dynamics 47
2.2.1 Initialisation 47
2.2.2 Forcefields 48
2.2.3 Bonded interactions 50
2.2.4 Non-bonded Interactions: van der Waals Interactions 50
2.2.5 Non-bonded interactions: Electrostatic Interactions 51
2.2.6 Ligand Parameterisation 52
2.2.7 Integrating the Equations of Motion 53
2.2.8 Temperature and Pressure within MD Simulation 54

2.2.9 Minimisation and Equilibration 56
2.3 Metadynamics 57
2.3.1 Well-tempered Metadynamics 59
2.3.2 Infrequent Metadynamics 60
2.3.3 Reweighting Free Energy and Time 61
2.3.4 Funnel Metadynamics 63
2.4 Simulation Analysis 67
Chapter 3
Modelling Ligand Binding in Glycine Receptors 70
3.1 Introduction 70
3.2 Modelling the Whole Glycine Receptor within a membrane 71
3.2.1 Glycine Parameterisation 73
3.2.2 Modifications 73
3.2.3 System Preparation 74
3.2.4 Minimisation to Production 75
3.3 Modelling the Glycine Receptor ECD 76
3.3.1 System Preparation 76
3.3.4 Minimisation to Production 78
3.4 Whole Receptor Simulation Results 78
3.4.1 The Structure and Stability 79
3.4.2 Hydrogen bonding 82
3.4.3 Cation- π Interactions 84
3.4.4 The Binding Mode 84
3.4.5 Conclusion 88
3.5 The ECD system Results 90
3.5.1 The Structure and Stability 90
3.5.2 Hydrogen bonding 92
3.5.3 Cation-π Interactions 93
3.5.4 The Binding Mode 94
3.5.5 Conclusions 95
3.6 Discussion 95
Chapter 4
Partial and Full Agonism in Glycine Receptors 101
4.1 Introduction 101
4.2 Modelling the GABA Bound Glycine Receptor 103
4.2.1 GABA Parameterisation 105
4.2.2 Simulation Details 105

..... 106
4.3 The Partial Agonist Bound GlyR system Results 108
4.3.1 The Structure and Stability 108
4.3.2 Hydrogen Bonding 110
4.3.3 Cation-π Interactions 113
4.3.4 The Binding Mode 114
4.3.5 Comparing The Glycine and GABA bound structures 116
4.3.6 Conclusion 119
4.4 Funnel Metadynamics 119
4.4.1 Exploratory Metadynamics 119
4.4.2 Simulation Details 120
4.4.3 Comparing Full and Partial Agonism 122
4.4.4 Stability + Flexibility 122
4.4.5 The Free Energy Surface 126
4.4.6 Reweighting the Free Energy Surface 127
4.4.7 Glycine Unbinding 129
4.4.8 GABA Unbinding 133
4.5 Discussion 135
4.5.1 Full Agonism in the GlyR 135
4.5.2 Partial Agonism in the GlyR 137
4.5.3 Comparing Full and Partial Agonism of GlyRs 137
4.5.4 GABA resistance to dieldrin receptor (RDL) 139
4.5.5 Conclusion 142
Chapter 5
The N46K Mutation in the Glycine Receptor and its Effects on Glycine Binding 143
5.1 Introduction 143
5.2 modelling the N46K mutant 147
5.2.1 Mutation Procedure 147
5.2.2 Simulation Details 147
5.3 The Mutant N46K GlyR system Results 148
5.3.1 Structure and Stability 148
5.3.2 The Binding Mode 151
5.3.3 Hydrogen Bonding 151
5.3.4 Cation-π Interactions 152
5.3.5 Comparing residue 46 154
5.3.5 Conclusions 157

5.4 Funnel Metadynamics 157
5.4.1 Simulation Details 157
5.5 Comparing Results of the Wild Type to the N46K mutant 158
5.5.1 Funnel Metadynamics Stability 158
5.5.2 Funnel Metadynamics Convergence 161
5.5.3 Funnel Metadynamics Convergence 162
5.5.3 Reweighting the Free Energy Surface 163
5.5.4 Glycine Unbinding from the N46K GlyR Mutant 164
5.5.5 The N46K Residue and Unbinding 166
5.5 Discussion 168
5.6.1 The Mutant Glycine Receptor system 168
5.6.2 Funnel metadynamics and the N46K Mutant 169
5.6.3 The Mechanism of Action of The N46K Mutant 169
5.6.4 Conclusion 171
Chapter 6
Conclusions 173
Acknowledgements 177

List of Figures

Figure 1.1: Simple diagram of the prototypical neurone. The prototypical neurone being an example of a neurone with the basic structures typically found in the cell type. Some structures like the nucleus and essential organelles are found in all neurones whereas myelin is a more specific structure found in specialised neuronal subtypes. 17

Figure 1.2: This graph represents the time evolution of an action potential with potential difference across the membrane on the y axis and time on the x axis[2]. The lower section of the diagram represents a neuronal membrane and what ionic movement occurs at each stage: A, a stimulus depolarises the membrane. B, once a threshold is reached, the rising phase starts with faster depolarisation. C, a second threshold is reached and polarising channels open, leading to a slowdown and reversal in the membrane potential. D, here equilibrium processes including active transport of sodium and potassium across the membrane results in the resting potential being reached at point E. 19

Figure 1.3: A cartoon representation of the alpha-1 glycine receptor captured with cryo-EM (PDB: 6PM5 21

Figure 1.4: A diagrammatic representation of the function of glycine release at an excitatory synapse where the glycine triggers chloride uptake via glycine receptors to polarise the post-synaptic membrane and supress action potential propagation 25

Figure 1.5: A cartoon representation of both the ECD and TMD from a single subunit of the desensitised **α** -1 glycine receptor a, the ECD from a single subunit, viewed from a 90 degree angle facing in to the[2] centre of the pore. b, The TMD with key annotations, in the ECD the main annotations of structural components are also shown. 26

Figure 1.6: Diagram of the binding pocket in the glycine receptor bound with the full agonist GLY, imaged, and taken from the 6PM5 structure 28

Figure 1.7: Diagram of the binding pocket in the glycine receptor bound with the partial agonist GABA, imaged and taken from the 6PLX structure 30

Figure 1.8: Functional data derived from electrophysiology experiments carried out on α-1 GlyRs. A, The peak current plotted against concentration of agonist, measured from whole cells with the EC50s being 190 ± 20 μM, 1050 ± 80 μM and 28.4 ± 0.9 mM for glycine, taurine and GABA respectively, the curves are normalised against the maximum glycine current recorded in each cell. B, the single channel recording data of GlyR openings elicited by 100 mM of taurine, 100mM of GABA and 10 mM glycine. The data and plots represented here were taken from Yu et al 2021 31

Figure 1.9: The dogma of molecular biology, showing the progression of DNA to RNA to polypeptide protein. Here the wildtype glycine receptor is shown on the left and N46K mutant on the right, with a cartoon representation of how the substitution would come about and the repercussions of that in structural form in the bottom images 33

Figure 1.10: Cartoon representation of the alpha-1 glycine receptor extracellular domain with several potent mutant variants that are of interest for investigating receptor function. 34

Figure 1.11: Functional data derived from electrophysiology experiments carried out on WT and mutant α-1 GlyRs, the WT in black and the N46K mutants in red. The maximal responses are normalised to the maximum glycine response.This figure was taken from Wilkins et al (2016).. 35

Figure 1.12: A comparative illustration of the aligned desensitised (blue), apo(orange) and open state cyan alpha-1 glycine receptor. In A, the structures shown from cryo-EM with the subunit pair across a binding pocket shown. In B, a closer look at the differences between the desensitised, open and apo binding pockets. the dotted lines represent interactions. 39

Figure 2.1: General applications of molecular modelling at various resolutions and timescales. The work presented here involves all-atom simulations on a nano and microsecond timescale. Diagram taken from Kmiecik et al 45

Figure 2.2: Schematic demonstrating the various terms involved in the Amber forcefield. Circles represent the atoms involved and black lines denote intramolecular bonds whilst the dotted blue lines show intermolecular forces. The left column covers the bonded interactions and the right column illustrates the non-bonded interactions. 49

Figure 2.3: Graph of van der Waals potential with the switching function (dotted line) and without (solid line) 51

Figure 2.4: Decreasing potential energy over time during the energy minimisation process using NAMD 2.14. This is taken from the minimisation of the 5CFB structure described in Chapter 3. 56

Figure 2.5: Diagram illustrating how across one CV (labelled x) the progression of time accumulates gaussian terms to explore multiple energy minima in the potential energy V(x)..... 59

Figure 2.6: Diagram of the parameters and shape of the funnel restraint used in funnel metadynamics. The binding pocket is represented in simplified form as a chain of amino acid residues which would likely have a more complex topographical arrangement. Point A represents the centre of the binding pocket, the position through which the funnel Z-axis bisects. 66

Figure 2.7: Graphic to illustrate how measurements are made from simulations. A demonstrates the measuring of distance between the centres of mass of two groups of atoms (signified by the yellow dot. B illustrates the formation of hydrogen bonds between an acceptor (in red) and a hydrogen atom in grey bonded to a donor (in green). C illustrates a cation-Pi interactions where the electronegative π cloud on the ring (in blue) interacts with a positively charged group. 69

Figure 3.1. A, the cartoon representation of the 5CFB GlyR embedded in a membrane viewed from above the pore and from the side in B. C, is the scoring chart that is used as a measure of structure quality; this specifically is the 5CFB scores. 72

Figure 3.2: Here is a depiction of the binding pocket, the colour scheme is split across the two subunits that make up the binding pocket with the complementary (-) subunit shown in orange and the principal (+) subunit in blue. The Arginine (ARG-65) residues from the original 5CFB is highlighted in yellow and the modified form overlayed and highlighted in green. 74

Figure 3.3: A and B are structural representations of the completed system and show the 6PM5 structure from a parallel view (A) and an orthogonal view (B). The original PDB validation results are also shown here (C). 77

Figure 3.4: RMSD of the whole protein backbone as a function of time during the production run. The raw data is plotted with a paler colour and a running average is calculated for every 1 ns and plotted for clarity. 79

Figure 3.5: RMSD of the whole protein backbone as a function of time during the production run. The error bars presented are measures of the SEM over the course of the simulation for the specified pocket. 80

Figure 3.6: Time evolution of RMSD values for each binding pocket in A, the whole subunit structures in B, and the Loop-C structures in C. 81

Figure 3.7: Profile of the total number of hydrogen bonds formed with the ligand over the course of the production. The pocket to which each ligand belongs to is coloured according to the key. 82

Figure 3.8: Profile of the types of hydrogen bonds formed over the production. A, a snapshot of the stable binding pocket D from the simulation with the residues and respective bonds labelled. B, a snapshot of the unstable binding pocket B with residues and interactions highlighted. C and D, here occurrence is used a metric to measure the strength of these interactions as a percentage of time present. In C, the graph represents the H-bonding profile of ligand in pocket D. D, shows the bonding profile of the weaker ejection prone pocket B. 83

Figure 3.9: π -Interactions recorded for each subunit with the specific residue involved on the X axis and the occurrence as a percentage of time on the Y. Only subunit A, B and D were found to have these interactions present. 84

Figure 3.10: A, Distance measured between the centre of mass at the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues. B, Distance measured between the centre of mass of the glycine ligand and the centre of mass of the binding pocket residues. The time covered is the full 1000 ns of the production run and the different subunits are coloured according to the key. 85

Figure 3.11: A bar chart representing the percentage of the simulation that ligand-water-protein bridges are present for, including the GLU (E157) and the SER (S158) residue. These were measured by calculating the presence of hydrogen bonds in cpptraj between the ligand, a water molecule in the pocket and the binding pocket residues occurring simultaneously. 86

Figure 3.12: A, Diagram of the binding pocket of the 5CFB structure. The water molecule pictured in the system is stabilised by hydrogen bonds to the oxygen groups of S158 and E157 in red as well as a hydrogen bond to the ammonium group of the glycine ligand in blue. The (+) subunit is shown in orange and the (-) subunit in blue.B shows A cartoon illustration of the binding pocket with water and the ligand within. The blue line marks the path the water takes into the pocket, travelling below the pocket and in past the C-loop. 87

Figure 3.13: Here both binding pockets are illustrated and superimposed, using the newer glycine bound(6PM5) structure in blue to compare with the strychnine bound system (5CFB) in orange. The superimposed structures are shown with the backbone protein visible in A and the key residues highlighted in B. 88

Figure 3.14: The volume of two binding pockets in the 5CFB system over the course of a 100 ns excerpt from the production. In blue, the unstable pocket B demonstrates much larger variability in volume prior to ejection later in the simulation. In Orange is the stable pocket D which whilst there is some fluctuation, it is much more consistent than the former pocket 89

Figure 3.15: A, RMSD of the ECD protein backbone over time during the production run. B, Average RMSD values for each of the three important structures, the whole subunits, Loop-C and the binding pocket residues. 90

Figure 3.16: Time-evolution of the RMSD values for each binding pocket in A, the whole subunit structures in B, and the Loop-C structures in C. 91

Figure 3.17: Profile of the total number of hydrogen bonds formed with the ligand over the course of the production. The subunit to which each ligand belongs to is coloured according to the legend. 92

Figure 3.18: Profile of the types of hydrogen bonds formed over the production. A, a snapshot of the stable binding pocket C from the simulation with the residues and respective bonds labelled. B, a plot of the types of interactions occurring and for what proportion of the production they are present. 93

Figure 3.19: Cation-π interactions recorded for each pocket with the specific residue involved labelled along with for what proportion of the simulation time the interaction was present. 93

Figure 3.20: A, Distance measured between the centre of mass at the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues. B, Distance measured between the centre of mass of the glycine ligand and the centre of mass of the binding pocket residues. The time covered is the full 1000 ns of the production run and the different subunits are coloured according to the key. 94

Figure 3.21: The sequence alignment between two ECD structures, the top row shows the protein sequence for an α-1 Zebrafish (6pm5)[26] sequence whereas the second row of each pair is the sequence for the human α-1 receptor (8DN4)[173]. 98

Figure 3.22: The Root Mean Square Fluctuation for both the 5CFB (In orange) and 6PM5 (In blue) for each residue averaged across all five subunits, with the average structure calculated as the reference. 99

Figure 4.1: A schematic showing the basic arrangement of residues within the pocket and how structural data has shown the C loop to move from an uncapped to capped pose when glycine binds, changing the structure from the Apo open/closed state. The C loop structure here is illustrated by a curved line capped with the appropriate residue threonine. 101

Figure 4.2: A, The structure of 6PLX with the cryo-EM map data overlayed from a horizontal view. B, the same structure with cryo-EM data overlayed from a top-down perspective. C, The cryo-EM data of the binding pocket of 6PLX along with the predicted interactions. 104

Figure 4.3: The ECD taken from the PDB data from a top-down view in A and a horizontal view in B. The validation metrics are shown in C. 106

Figure 4.4: The RMSD of the protein backbone as a function of time during the production simulation. All plotted RMSD values are shown with the raw data shaded and the moving average taken every 10 frames shown in solid colour. 108

Figure 4.5: The average RMSD values for each substructure for each of the 5 pockets, the error bars here are representing the standard deviation measured across the production for each value recorded. 109

Figure 4.6: The RMSD of key substructures measured from the first frame of production including the subunits over the production in A, and the RMSD of the C-loops in B. C, shows the RMSD values of the binding pockets residues. 110

Figure 4.7: The total number of hydrogen bonds for each pocket over the course of the production, measuring interactions between the ligands and their respective key binding pocket residues. 111

Figure 4.8: A, a snapshot of the pocket with the key binding pocket residues and the interactions drawn with dotted lines corresponding in colour to the plot in B, which illustrated the occurrence of the different hydrogen bonds present over the course of the production. 112

Figure 4.9: The RMSD of the protein backbone as a function of time during the production simulation. 113

Figure 4.10: A, The distance between the tip of the C-loop centre of mass (THR-204) and the centre of mass of the binding pocket using the principal binding pocket residues. B, a measurement of the distance between the centre of mass of the ligand to the centre of mass of the binding pocket, each pocket plotted over the course of the production. 114

Figure 4.11: Depiction of binding pocket B, showing the positioning of the C-loop in orange, Ligand as well as the movement of water into the pocket. 115

Figure 4.12: The volume of binding pockets B and C plotted over the course of the production, evaluated with the mdpocket tool 116

Figure 4.13: The exploratory metadynamics trajectories used to find a general direction for unbinding, the data here illustrates the many paths the ligand took overlayed with the structure of the protein. The dashed blue lines represent the trajectories of each of the 15 replicas used and the red solid line shows an average of these trajectories combined. 120

Figure 4.14: The 6PM5 ECD structure illustrated alongside the funnel representation. The view of the pocket and funnel from the top view is shown in A, the side on view in C and the view running along the funnel projected axis out of the pocket in D. Visualising of the pocket orientation is aided by B showing the principal residues alone with the funnel, colour coded for arginine (blue), phenylalanine (pink), threonine (red) and serine (orange) and glutamic acid (green). 121

Figure 4.15: A, The RMSD of the protein backbone as a function of time during the whole funnel metadynamics simulation with both the 6PM5 and 6PLX models plotted. The average RMSD values for the important substructures identified, where plot B represents the 6PM5 system values and C shows the values measured for the 6PLX model. 123

Figure 4.16: The RMSF values for the backbone atoms of residues in the 6PM5 system (blue) and the 6PLX system (orange), averaged across the subunits. The data collected here is taken from the funnel metadynamics simulation. 124

Figure 4.17: A, the CV fps.lp plotted along the funnel metadynamics trajectory of the 6PM5 model. B, the CV fps.lp plotted along the trajectory for the 6PLX model. The red line represents the point at which the ligand can be considered either unbound or bound. The binding free energy difference plots covering the last few hundred nanoseconds of the simulation with standard error plotted in red, 6PM5 in C and 6PLX in D. 125

Figure 4.18: A, The free energy surface of the glycine bound 6PM5 system as a function of the CVs. B, The free energy surface of the GABA bound 6PLX system as a function of the CVs. C, The

free energy profile projected along the funnel axis for both the 6PLX and 6PM5 models. The red bar illustrates where the cylindrical correction is applied. (needs rechecking as it seems a bit large) 126

Figure 4.19: The reweighted free energy maps showing the funnel metadynamics reweighted from fps.lp and fps.ld onto the ligand coordinates in cartesian space p.x p.y and p.z. A, the reweighted funnel of the 6PLX system with fps.lp CV on the X axis and p.z on the Y, the same data is plotted for the 6PM5 system in B, both providing a vertical cross-section of the funnels. C and D are the plots of the reweighted p.x and p.y values, of the 6PLX and 6PM5 system respectively. C and D provide a horizontal cross-section of the funnel free energy surface. 128

Figure 4.20: A snapshot taken from the funnel metadynamics trajectory showing a potential pre-binding pose where the ligand is flanked by protein residues ARG-65, ASN-42, GLY-174 and GLN-177. The C-loop is highlighted with orange as part of the illustrated protein structure. 128

Figure 4.21: The binding pocket is illustrated here with the C-loop in red and the two residues used to judge opening and closing of the C-loop highlighted in yellow, THR-204 and ARG-65..... 129

Figure 4.22: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white.. G, shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H, is a plot of distances both for the C-loop opening metric in blue and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for these analyses. 131

Figure 4.23: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white. G shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H is a plot of distances both for the C-loop opening metric in purple and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for this set of analyses. 134

Figure 4.24: The volume of the binding pockets shown over the course of the ejection trajectory (The first 10 ns of FmetaD) with the GABA bound system in green and the glycine bound system in purple. These measurements were made using the funnel metadynamics trajectory analysed with the mdpocket tool..... 138

Figure 4.25: Illustration of the the RDL receptor (blue) and the α-1 glycine receptor (orange) structures superimposed. In A, the protein structures are represented as a cartoon illustration along with the key binding pocket residues of each system. In B a clearer representation of the binding pockets along with labels of important residues is shown with the GABA ligands in the centre. 140

Figure 4.26: The free energy landscapes produced from this work alongside the results gained in the Comitani et al paper[86]. A shows the past work with the RDL receptor with the funnel axis fps.lp on the x axis and the fps.ld collective variable on the y axis. There are two visible wells labelled A (the orthosteric binding site) and B indicating a pre-binding pose. B shows the results from this work previously discussed for GABA binding to the α-1 glycine receptor. 141

Figure 5.1: Lysine and Asparagine molecular structure on the left and right respectively. The structures shown are the complete molecules outside of the protein structure. The backbone atoms are highlighted by a red dotted line and the amine group that becomes positively charged at physiological pH is outlined with a blue dotted line. 145

Figure 5.2: Snapshots from during the production showing the location of the mutation, changing the residue from Asparagine to Lysine. The binding pocket loops A, B, D and F are also labelled with the principal (-) and complementary (+) subunits coloured cyan and orange respectively. 146

Figure 5.4: The average RMSD values for each substructure across the five pockets. The error bars here represent the standard deviation measured over the course of the production. 149

Figure 5.3: The RMSD of the protein backbone as a function of time during the production molecular dynamics simulation. 149

Figure 5.5: The RMSD values plotted over time for the key substructures. A, Subunits. B, pockets. C, C-loops. 150

Figure 5.6: The total number of hydrogens for each pocket over the course of the production, measured between the ligands and binding pocket residues of each respective pocket A-E. .. 151

Figure 5.7: A, the hydrogen bonding over the course of the production including the types of interactions formed along with the proportion of the simulation they were present for. B, a labelled illustration of the binding pocket taken from a frame of the simulation. 152

Figure 5.8: The cation-π interactions recorded for each binding pocket, plotted as a proportion of the simulation they were detected as present for. Binding pocket B showed no cation-pi interactions formed and therefore was left out of the plot. 153

Figure 5.9: A, The distance measured between the centre of mass of the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues over time. B, The distance measured between the centre of mass of the ligand and the centre of mass of the binding pocket residues over time. 154

Figure 5.10: A, an illustration of residue 46 in between the subunits below the binding pocket, the interacting residues highlighted. B, the hydrogen bonding profile of the ASN-46 residue during the production of the wild type model. 155

Figure 5.11: A, an illustration of residue 46 in between the subunits below the binding pocket, the interacting residues highlighted. B, the hydrogen bonding profile of the LYS-46 residue during the production of the wild type model. 156

Figure 5.12: A, The RMSD of the protein backbone as a function of time during the whole funnel metadynamics simulation with both the 6PM5 and N46K models plotted. The average RMSD values for the important substructures identified where plot B, shows the values measured for the N46K model. 159

Figure 5.13: The RMSF values for the backbone atoms of residues in the 6PM5 system (Blue) and the N46K system (orange). The data here has been averaged across the 5 subunits to demonstrate general differences between the models, standard deviation has been plotted as error bars. 160

Figure 5.14: Plots taken from the funnel metadynamics to measure convergence. A, The CV fps.lp shows the movement along the z axis of the funnel in Å measured over the course of the

simulation, the red line denotes the binding pocket position. B, The binding free energy over the course of the second part of the funnel metadynamics stabilising, standard error is plotted in red. 161

Figure 5.15: The free energy maps as a function of the CVs of the wild type, B, and mutant models,A. Projected free energy profiles as a function of the ligand position along the funnel axis, C, with the correction due to the cylindrical part of the funnel illustrated by a red bar. 162

Figure 5.16: The reweighted free energy maps showing the funnel metadynamics reweighted from the fps.lp and fps.ld CVs onto the position of the ligand along the x (p.x), y (p.y) and z (p.z) coordinates. A and B show the reweighted free energy surfaces for the N46K mutant whilst C and D show the reweighted plots for the wild type model. 163

Figure 5.17: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white. G, shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H, is a plot of distances both for the C-loop opening metric in purple and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for these analyses. 165

Figure 5.18: The wild type glycine receptor model, with the Asparagine targeted by the N46K mutation highlighted along with residues that it forms interactions with and the closest binding pocket residue for reference: PHE-63. A, during the ejection period in Figure 4.23. The hydrogen bonding profile for the N46 residue is plotted in B. 166

Figure 5.19: The mutant glycine receptor model, with the Asparagine mutated to Lysine by the N46K mutation highlighted along with residues that it forms interactions with in A. A second pose during the ejection process is also identified along with the key residue interaction highlighted in B. These results are taken from the ejection period in Figure 5.17. The hydrogen bonding profile for the K46 residue is plotted in C. 167

List of Tables

Table 1.1: A table of structures used in this work along with the key pieces of information. 38

Table 3.1: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was called to the subsequent value. 76

Table 3.2: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was called to the subsequent value. 78

Table 4.1: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was rescaled to the subsequent value. 107

Abstract

Ion channels are fundamental units of the nervous system responsible for mediating intercellular signals and propagating action potentials, this makes them critical in brain function, disease, and pharmacology. Pentameric Ligand Gated Ion Channels (pLGICs) are a subgroup of the proteins present at synapses responsible for modulating synaptic activity; the glycine receptor (GlyR) is a common variant responsible for inhibitory action. Understanding the activation of these channels requires detailed structural and functional profiling, which is limited by current experimental techniques. Structural data can be collected to a very high resolution of under 3 Å, and functional data can be generated with electrophysiology experiments to determine channel opening times and binding rates. However, the structure can only be observed in snapshots of time after being modified and processed in non-physiological conditions, where functional data and dynamics cannot be simultaneously collected. Atomistic simulations can overcome these limitations by providing atomic resolution dynamics on timescales of hundreds of ns to μs. In this PhD project, we investigated the binding of the glycine receptor through the application of these techniques on various systems. Three systems of the GlyR-α1 were developed to provide insights into the binding pocket dynamics of pLGICs, one bound with glycine to represent the endogenous ligand binding, a system with the partial agonist GABA bound and a system with the N46K mutation. Through simulating these systems with molecular dynamics and metadynamics, we can carry out comparative analysis across them, focusing on the binding profiles and binding/unbinding events. The aim of this PhD is to utilise this analysis to gain key insights into the binding dynamics of GlyRs with a focus on the binding mode, the differences that affect binding affinity so potently between partial and full agonists, and how mutants can impact the binding mode of pLGICs. The first chapter of this thesis will cover the essential context and required for understanding both the function and structure of pLGICs and specifically the GlyRs, as well as an introduction to partial versus full agonism, mutants of interest in GlyRs and the techniques used to study these channels experimentally and computationally. The second chapter will describe and detail the theoretical concepts and tools used for the work in later chapters covering all atom molecular dynamics, metadynamics and funnel metadynamics. The third chapter covers the development of the GlyR system and an investigation of the binding mode in unbiased simulations. This chapter includes a comparison of a whole protein membrane embedded system against an extracellular domain system along with an analysis of the starting conditions, quality, and stability of the binding during simulations. The fourth chapter provides a further analysis of the binding mode in the context of full and partial agonism, expanding upon unbiased simulations through the application of funnel

metadynamics to profile and compare the free energy surface and binding paths of full and partial agonists. The fifth chapter applies the same approach and methods but for comparing a wild type system versus the N46K mutant variant to establish an atomistic description of how this mutation impacts ligand binding. Overall, the work displayed and discussed in this thesis provides an expanded view into the dynamics of GlyR binding and consequently a clearer understanding of pLGICs.

Chapter 1

Introduction

1.1 The Nervous System

The study of the brain and the nervous system is referred to as the field of neuroscience, and for much of its history the field was dominated by experimental approaches. However, recently it has become more common for computational systems to play an important role in understanding the biology and physical phenomena at play. Neuroscience is the study of the nervous system at every level, encapsulating a broad variety of fields from the large scale involving cognitive science, anatomy, and physiology down to the small scale cellular and molecular disciplines. This diversity of subfields reflects the incredibly complex nature of neuroscience[1], a necessity, given the fact that what is being studied is widely considered the most complex entity in the known universe. The nervous system in simple terms is a communication network between cells.

The principal cells of the nervous system are the neurones, a specialised cell type with several unique properties and structures capable of mediating the rapid communication needed for complex behaviour, motor function and homeostatic control. The general structure of a typical neurone is illustrated in Figure 1.1. The main body of the neurone is the soma, a rounded 20 μm across cell body, filled with cytosol and containing the fundamental organelles needed for a cell to function. The nucleus, mitochondria, Golgi apparatus, rough/smooth endoplasmic reticulum and ribosomes are all present within the soma[2].

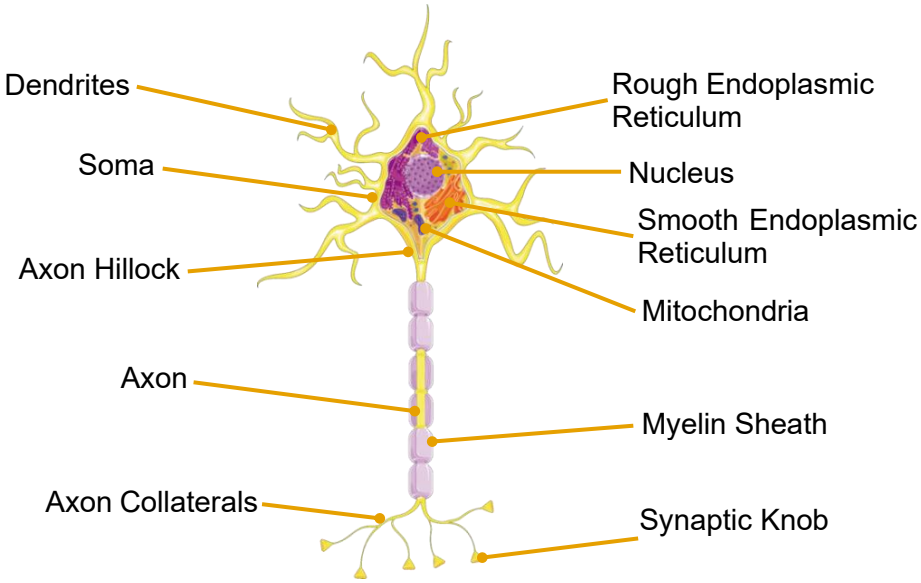

Figure 1.1: Simple diagram of the prototypical neurone. The prototypical neurone being an example of a neurone with the basic structures typically found in the cell type. Some structures like the nucleus and essential organelles are found in all neurones whereas myelin is a more specific structure found in specialised neuronal subtypes.[2]

The structures outside the soma are even more specialised for intercellular communication. Dendrites are branching structures connecting to the soma, covered in synapses, and function as points of contact for other neurones to transmit signals via neurotransmission. The axonal projections deals with the transmission of signals to other neurones. The start of the axon tapers out from the axon hillock connected to the soma. This structure extends, sometimes branching with axon collaterals, to the synaptic cleft of dendrites on other neurones. The axon ends with a synaptic terminal, which is packed with synaptic vesicles that contain neurotransmitter and mitochondria to meet the needs for signalling molecules and energy in neurotransmission. One of the most important and specialised parts of a neurone is the membrane itself. Whilst it is only around 5nm thick, it is peppered with clusters of proteins essential for mediating neurotransmission. Ultimately, understanding the biophysical properties of the neuronal membrane and how the proteins embedded within it function is vital to unravelling how the nervous system functions on a larger scale[2].

The movement of information between neurones and many other cell types gives rise to somatosensation, cognition, behaviour, and many essential homeostatic processes. A combination of cellular biology and advances in imaging techniques has led to an intense effort to map these communicating networks to functions, as well as understand the molecular mechanisms at play that mediate these processes[1]. However, the nervous circuitry is much more complex than the electrical counterpart it is often compared against. There are many forms of communication in the nervous system and even the most straightforward behaviour or nervous function requires a diverse range of cellular interactions to be mediated[3]. Within

each level of these communication networks, often every component can be modulated in some way.

Networks of neurones and glia can communicate directly through electrical coupling at gap junctions[4], or indirectly using neuropeptides or neurotransmitters in synaptic transmission. Synaptic transmission occurs at synapses, which are perhaps one of the most studied structures in neuroscience. Their importance is due to their role in highly organised systems of learned behaviour through synaptic reinforcement. The principal component in synaptic transmission is the ion channel. Ion channels are multimeric proteins that form pores in membranes and allow movement of ions across the membrane. In many circumstances this includes, but is not limited to, when a change in voltage is detected or a ligand/neurotransmitter binds to the channel[5,6].

The reason that controlled permeability across a membrane is so crucial is due to the electrochemical properties of the membrane and how nervous signal transduction works. The lipid membrane structure is impermeable to the ions in the extracellular fluid, and since ions carry a charge, both a high electrical and chemical gradient can be set up simultaneously across a membrane. In neurones, active and passive transporter proteins maintain a steep electrochemical gradient across their membranes, specifically, a high sodium concentration extracellularly and high potassium concentration intracellularly[2]. Certain stimuli can trigger channels to open, disturbing this equilibrium and triggering what is called an action potential (AP) illustrated in Figure 1.2.

Depending on the cell type, a stimulus triggers the opening of sodium and sometimes calcium channels (A, in Figure 1.2). This disturbs the membrane potential at a localised patch of the membrane and the positively charged ions move into the cell. Once a threshold magnitude of change in the electrochemical gradient is reached, voltage-gated sodium channels open, allowing a further and larger sodium influx into the cell. This mediates rapid depolarisation (b, in Figure 1.2) that is referred to as the rising phase.

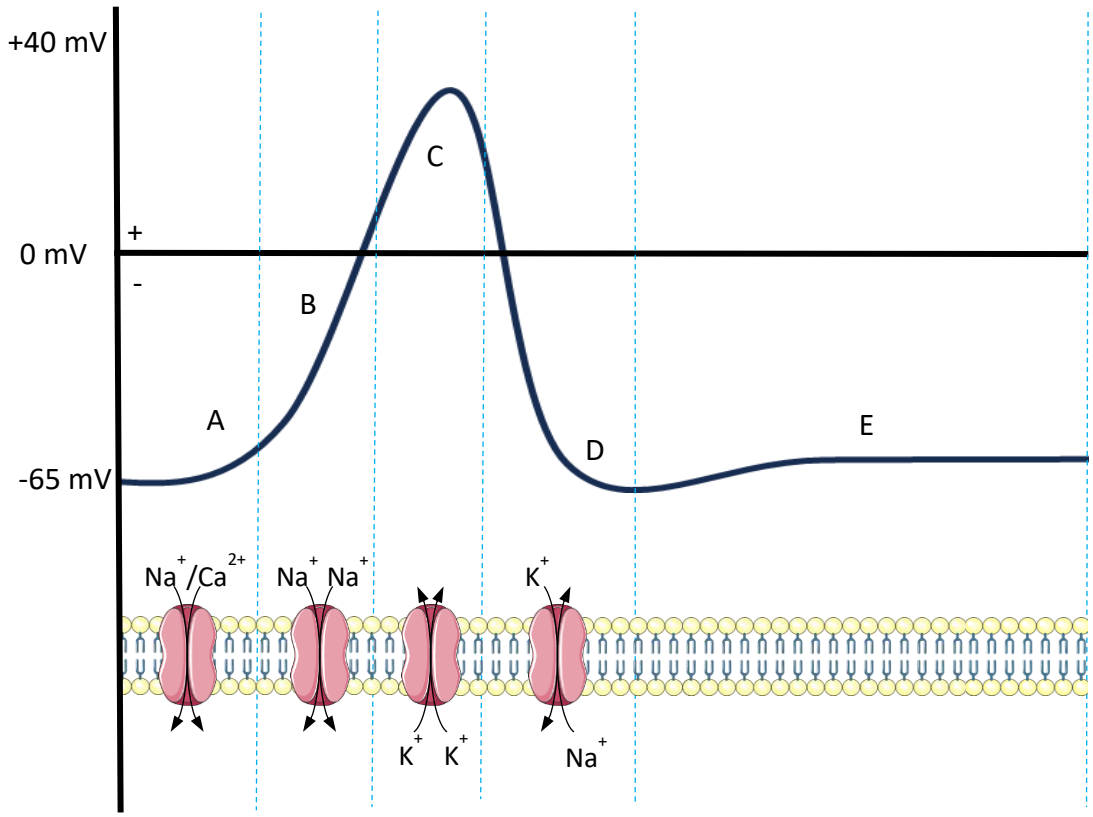

Figure 1.2: This graph represents the time evolution of an action potential with potential difference across the membrane on the y axis and time on the x axis[2]. The lower section of the diagram represents a neuronal membrane and what ionic movement occurs at each stage: A, a stimulus depolarises the membrane. B, once a threshold is reached, the rising phase starts with faster depolarisation. C, a second threshold is reached and polarising channels open, leading to a slowdown and reversal in the membrane potential. D, here equilibrium processes including active transport of sodium and potassium across the membrane results in the resting potential being reached at point E.

The rapid depolarisation from opened voltage-gated sodium channels is enough to further open adjacent channels, meaning that the depolarisation spreads like a wave along a membrane. This allows the AP to move rapidly across a neurone to another synapse where it can trigger more neurotransmitter release. A second threshold is reached (c, in Figure 1.2), triggering voltage-gated potassium channels to open, allowing efflux of these ions. This polarises the membrane below the original resting potential (d, in Figure 1.2) where equilibration mechanisms bring the membrane back to the original state (e, in Figure 1.2)[7].

This sequence of events is typically triggered and perpetuated at the synapse by the activation of ligand-gated ion channels (LGICs)[8,9]. The AP is a fundamental part of neuronal signalling and is vital to fast synaptic transmission; the LGIC is the principal component for mediating it. Understanding the simpler mechanical functioning of these proteins is crucial for attempting to gain insights into the more complex systems at play built from these proteins. LGIC function is the focus of this thesis.

1.2 Pentameric Ligand Gated Ion Channels

LGICs have subgroups based on their structure, function, and genetic homology. One such group is the pentameric ligand gated ion channel (pLGIC), which includes the Cys-loop family of ligand-gated ion channels. This group of ion channels has an extensive catalogue of research published and is one of the most well-documented LGIC group[10]. The channels are of critical physiological importance, highly conserved in the animal kingdom and ubiquitous throughout the human nervous system[11]. This is further evidenced by dysfunction leading to a range of pathologies[12]. Another big reason for such interest in pLGICs is due to how progress in understanding how these channels function has set the foundation for developing therapeutics including anaesthetics, anticonvulsants, and anxiolytics[13].

1.2.1 pLGIC General Structure

The pLGIC superfamily share many architectural similarities, the key parts illustrated in Figure 1.3. Notably, the biggest features shared includes five distinct subunits and a large N-terminal extracellular domain that contains a loop held together by an internal cysteine disulphide bridge (the Cys-loop)[13]. The large extracellular domain (ECD) contains the orthosteric ligand-binding sites at the interface between subunits, the two subunits form a cavity in the ECD where residues such as arginine, threonine, phenylalanine, serine and glutamic acid can form interactions with a ligand when inside the pocket. Connected to this domain is the bundle of alpha-helical transmembrane domains (TMDs) with an intracellular loop between the third and fourth TMD[14,15.]

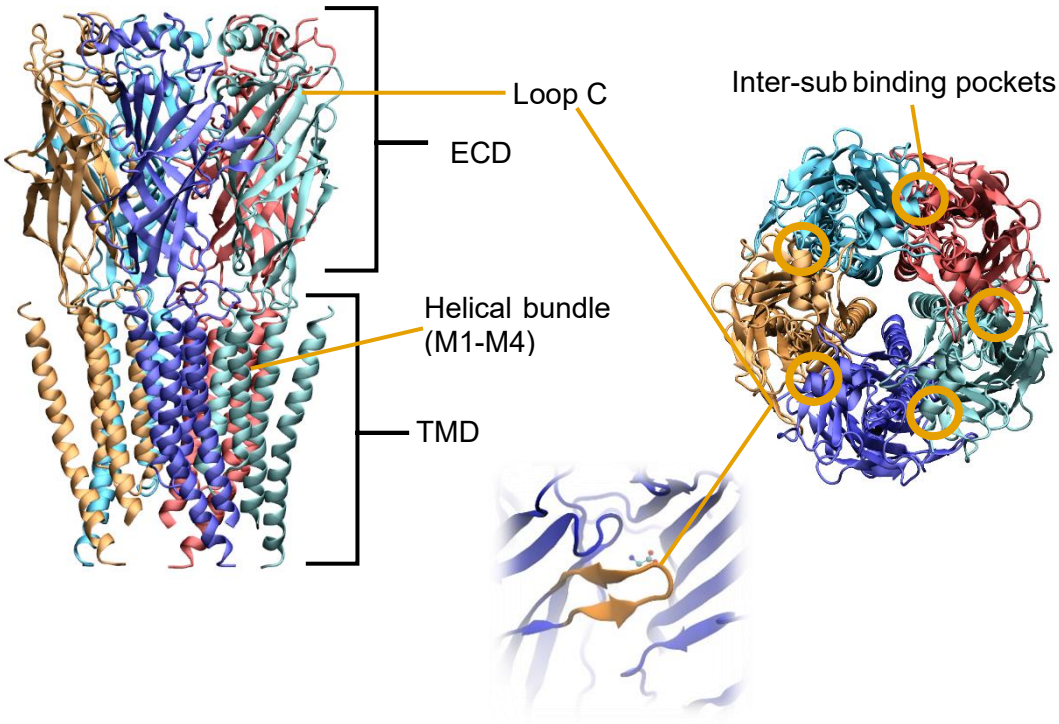

Figure 1.3: A cartoon representation of the alpha-1 glycine receptor captured with cryo-EM (PDB: 6PM5[26]). On the left the full structure is shown, excluding the intracellular domain which was not captured in the method used. The extracellular domain (ECD) and Transmembrane domain (TMD) are shown complete on the left unobscured by membrane which would wrap around the TMD region, each of the five subunits are shaded with a different colour.

The transmembrane domain (TMD) is made up of the M1, M2, M3 and M4 helical structures) with the intracellular domain (ICD) connecting at the C-terminus. The ICD is a highly variable loop region that links the M3 and M4 helixes together, it is typically removed to allow structural imaging and so there is very little structural information for this region. Another vital component is loop C, a Tyrosine loop in the extracellular domain (Figure 1.3), which 'caps' the binding pocket when the neurotransmitter is bound, shrinking, the binding pocket and potentially occluding it site from water[16]. Depending on the localisation and expression of these neuroreceptors, their opening and closing can have a broad range of physiological functions.

1.2.2 pLGIC Functions

The pLGICs can be functionally described in simple terms: they permit the flow of ions across the membrane through an ion selective pore only when conditions such as binding of a ligand are met. In mammals, the pLGIC family includes the nicotinic acetylcholine (nAChR), serotonin type 3 (5-HT_3R), γ-aminobutyric acid type A ($GABA_A$R) and glycine receptors (GlyR)[17]. The

nAChRs are spread across both the central and peripheral nervous system with a diverse range of roles in each. They are important in skeletal muscles for signal transduction from the brain where they also mediate synaptic transmission. Other than these broad well documented roles, nAChRs also have specialised roles in sensory, metabolic, and immune tissues[18]. With regards to the binding pocket itself, the key residues cited are the tryptophan (W143) and tyrosine (Y185) where the tryptophan forms the base of the binding pocket and tyrosine stabilises the bound state[19]. In simulations, water has been seen in the binding pocket suggesting a role in the binding of ligands to these receptors[20].

The 5-HT_3R is unique when compared to other serotonin receptors as it is part of the cys-loop ligand gated ion channel. These receptors present in both the peripheral and central nervous system and have been implicated as playing a role in the gastrointestinal tract, emesis, cognition, and psychological disorders[21,22]. The $GABA_A$R is reportedly present in a number of brain structures and modulates excitatory activity in the CNS and can lead to severe neurological disorders if disrupted [23]. This modulatory action has also made them prime targets for pharmaceuticals such as benzodiazepines and anaesthetics[24]. Like the nAChR, the principal binding residues include a number of aromatic amino acids including tryptophan (W90, W183) and phenylalanine (F226). The GlyRs mediate synaptic inhibition in the spinal cord, brainstem, and a range of other parts of the central nervous system; their key roles are in the regulating the excitability of motor, sensory and pain fibres but also have roles in auditory and visual processing [25].

1.3 The GlyR

The work presented in this thesis focuses on the GlyR with the goal of uncovering insights into the ligand binding properties. pLGICs like the GlyR have been at the centre of neuroscience for over half a century and the GlyR is referred to as an ideal vehicle for understanding pLGICs and phenomena like agonist efficacy. Early work by Kats and Thesleff suggested a rigid scheme of gating where a receptor would transition between closed/resting, open/active and closed/desensitised states. In this context a resting receptor is one that is not permitting ion flow. More recent work however suggests a more nuanced process where the receptor exists in a range of intermediate states between closed, open, and desensitised forms. Structural and electrophysiological experiments have been carried out to probe into the mechanism of gating and partial agonist action in GlyRs[26]. The work here captured partial agonist-bound closed states which are purported to be intermediate states that fit in between the resting and

open/desensitised states supporting this idea that a landscape of intermediates exists between states. As well as this, details about partial agonist binding such as the link between contraction of the binding pocket and efficacy was made, providing a theory for how partial agonists function in GlyRs and pLGICs in general. Another outcome of this work was a slew of high-resolution, high-quality structures of the GlyR with glycine and partial agonists clearly defined and accurately placed in the pockets, a highly valid starting point for any simulations to be carried out. Whilst the experimental work has provided many insights into the GlyR structure, function and partial agonism there still exists a lot of questions, specifically what is the binding path for full and partial agonists, how does the binding differ between agonists and what are the most important and immediate changes that happen upon ligand binding. The more dynamic aspects of GlyR and pLGIC binding and these questions that can be more thoroughly probed in computational work as is shown in the simulations and analysis presented in this thesis.

1.3.1 GlyR Function

GlyRs are present across the nervous system but have a more well-known role in spinal circuits and have also been identified within brain regions such as the cortex, cerebellum, thalamic and brainstem nuclei. The importance of GlyRs in the brain has been made clear by observations in cases of GlyR dysfunction which contribute to a pathology such as temporal lobe epilepsy[27]. Along with general pathological observations, mutation of the GlyRs in mice leads to a spasmodic phenotype and the human disease hyperekplexia, characterised by heightened muscle tone and startle reflexes. These investigations led to the observation of reduced inhibitory postsynaptic potentials (IPSPs). This predictably mimics the hyperexcitability generated from poisoning by strychnine, a GlyR competitive antagonist. Muscle tone is not the only physiological role for the receptor. It has been well established that GlyR is important in hippocampus-dependent higher brain function and neuropsychiatric disorders, as well as imbalances in glycinergic neurotransmission being linked closely to chronic pain, cerebral ischemia and neurodegeneration [28].

Synapses are not simple excitatory systems; they are modulators and can inhibit or encourage signal transduction. A significant part of inhibitory action at synapses in the nervous system is the action of glycine. Glycine activates the GlyR to permit chloride ion influx, disturbing the electrochemical balance across the membrane and impacting the typical biophysical phenomena that take place[29]. Ultimately, GlyR modulates neuronal excitability, controlling how

and when neurones activate and preventing extended states of excitability as will be explained in the following sections.

The importance of and the larger systemic role of GlyRs has been investigated through extensive study at the molecular and cellular level as well as observations made of mutant forms and pathologies involving the protein. Understanding how the GlyR impacts larger systems requires a fundamental understanding of how the protein functions within neurones. GlyRs act at the synapses and functions via a process called fast-synaptic neurotransmission, specifically the inhibitory form.

Fast-synaptic inhibitory neurotransmission starts with an action potential reaching the terminal of a glycinergic neurone, this depolarisation carried with the AP leads to subsequent release of vesicular glycine out of the pre-synaptic membrane. This glycine can then bind to the GlyRs on the post-synaptic neuronal membrane making a selectively permeable pore open in the centre of the GlyRs. Chloride is negatively charged and therefore moves into the cell down its concentration gradient, causing hyperpolarisation of the postsynaptic membrane.

This is the inhibitory post synaptic potential (IPSP), which makes the membrane harder to depolarise by pushing the potential difference across the membrane away from the AP threshold, thus reducing the impact of sodium influx from excitatory postsynaptic potentials (EPSPs)[27]. The whole process including the AP alongside the IPSP is illustrated in Figure 1.4. This inhibitory action modulates glutamatergic fast synaptic activity and therefore impacts glutamatergic modulation of motor and sensory reflexes, muscle function, pulmonary ventilation, and sensation[30].

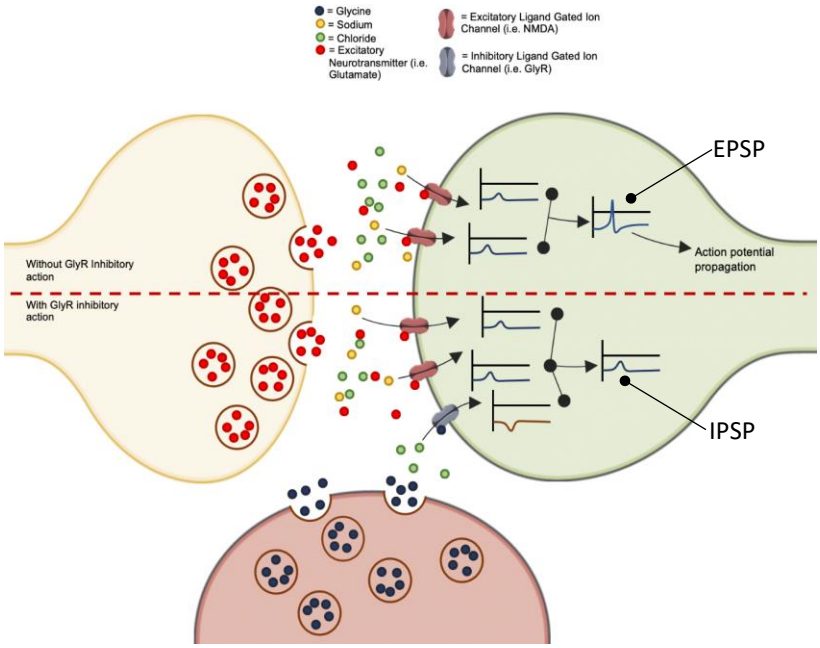

Figure 1.4: A diagrammatic representation of the function of glycine release at an excitatory synapse where the glycine triggers chloride uptake via glycine receptors to polarise the post-synaptic membrane and supress action potential propagation[2]. The presynaptic terminal from which neurotransmitter is released is on the left and the postsynaptic neurone where the ligand gated ion channels reside is on the right. Glycine is often released by a third, modulatory neurone shown at the bottom of the diagram.

To understand both neuronal physiology and pathologies related to GlyRs, we should understand what the dynamics and mechanisms at play on the molecular level are. Identifying key residues in the binding pocket, how they interact with various endogenous and synthetic ligands is critical for understanding and ultimately treating GlyR pathologies. Until recently, this higher resolution understanding was out of reach due conventional imaging techniques not reaching atomic level resolution. This has nearly been fulfilled in recent decades with the development of high-fidelity X-ray crystallography and Cryo-EM techniques, providing us now with near atomic resolution insights into the structures of GlyRs [31].

1.3.2 GlyR Structure

GlyRs share the same general structure as other pLGICs with the ECD, TMD and intracellular domains (Figure 1.5). Whilst the Intracellular domain is less well understood, the TMD and ECD have been very well profiled with both X-ray and cryo-EM methods. The ECD consists of an N-terminal alpha helix followed by a series of 10 beta strands, a second alpha helix is

also present between β strand 3 and 4. The TMD beyond the end of the ECD is made up of 4 helices per subunit, the M2 of these being the helix that lines the important channel pore.

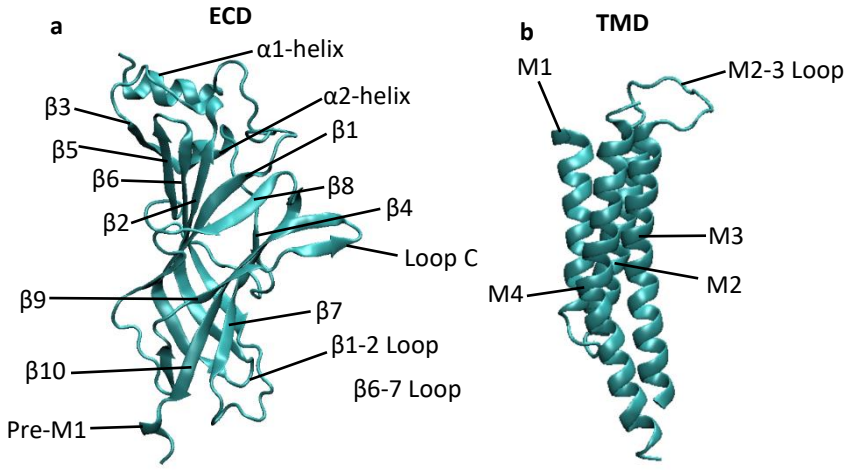

Figure 1.5: A cartoon representation of both the ECD and TMD from a single subunit of the desensitised α -1 glycine receptor **a,** the ECD from a single subunit, viewed from a 90 degree angle facing in to the[2] centre of the pore. **b,** The TMD with key annotations, in the ECD the main annotations of structural components are also shown.

There are also a number of GlyR subunit isoforms where minor variations in this structure exist, each with distinct impacts on binding, gating and ultimately, their physiological role. The α-1 subunit is the most common isoform and makes up most of GlyR subunits expressed at CNS synapses. α-2 subunits are primarily expressed during development, at embryonic and neonatal stages where it mostly replaces by α-1 postnatally. Whilst α-2 subunits have been found to still express in the adult CNS, their exact physiological role is not fully understood beyond having regulating sensory pathways. α-3 subunits are a minor adult isoform with similar expression to α-1. It has been found to inhibit nociceptive signal propagation to higher brain regions as well as also being found to regulate sensory pathways[25,32] . The α-4 isoform is the least understood type; however, in humans it is an unusable pseudogene. As well as these α-subunits there is a β-isoform which differs structurally more so than the α-subunits from themselves which typically share 90% sequence similarity. The GlyRs can form homomeric α-GlyRs and heteromeric αβ-GlyRs with the latter being the predominant type in adults[33]. Heteromeric receptors share a similar structure to the homomeric counterparts with particular conservation of the binding pocket, retaining the key principal residues that interact with glycine to maintain high affinity. The β-subunit allows postsynaptic scaffolding protein gephyrin to associate and localise the receptor at synapses. Whilst understanding how differences in this subunit affects the receptors, because of the high conservation in the binding pocket itself[34], the study of homomeric receptors in the context of binding is still very relevant to provide general understanding of the pocket dynamics and binding process.

Whilst the arrangement, shape, and properties of residues in the binding pockets of cys-loop receptors are very similar[35] there are important differences that result in such a high affinity for glycine binding. The architecture and content of a binding site determine key properties of a receptor and mechanically how it activates and the conditions influencing such events. It's important to note the ways in which a receptor can be impacted by such differences, including affinity, efficacy and potency. Affinity, as already mentioned and will be discussed many times through this thesis, is the strength at which a ligand binds to a given target. Efficacy of a ligand is the power of a ligand to evoke an effect at the target, typically this is measured in terms of a ligands ability to elicit a maximum response. The potency is a more comparative term typically associated with a measurement comparing ligands. For example, if the measurement used was EC_{50} (the minimum concentration needed to produce a 50% maximal effect) then a ligand with a greater EC_{50} than another would have a lower potency.

1.3.3 GlyR Ligand Binding

When glycine is bound to the GlyR, in zwitterionic form, it is situated between subunits and almost completely isolated from any external solvent. As shown in the desensitised structure binding pocket (Figure 1.6), the carboxylate moiety of glycine forms hydrogen bonds with the side chains of ARG-65 and SER-129, and THR-204. The ammonium of glycine is involved in a hydrogen-bond interaction with the main chain carbonyl of PHE-159 and a cation-π interaction with PHE-207. The ammonium is also hydrogen-bonded to a side chain of Glu157 and a main chain carbonyl of SER-158 through a water molecule. SER-129, ARG-65, THR-204, PHE-207, SER-158 and PHE-159 are the principal residues at the binding site[26,36].

Under biological conditions, the ligands that bind to the GlyR are highly efficacious, with the channel being open 98% of the time when bound. In Figure 1.6, glycine can be seen along with the surrounding residues suggested to interact with it as it binds[29,36]. The process for ligand binding starts when the ligand enters the binding pocket and a number of these residues such as SER-129, PHE-159 and ARG-65 to form hydrogen bonds and cation-pi interactions.

In many biomolecular processes, including ligand binding, these types of interactions can be potent with hydrogen bond strength ranging from 5 to 10 Kcal/mol[37] and cation-pi interactions having similar strength or even much stronger depending on the ligand [38–40].

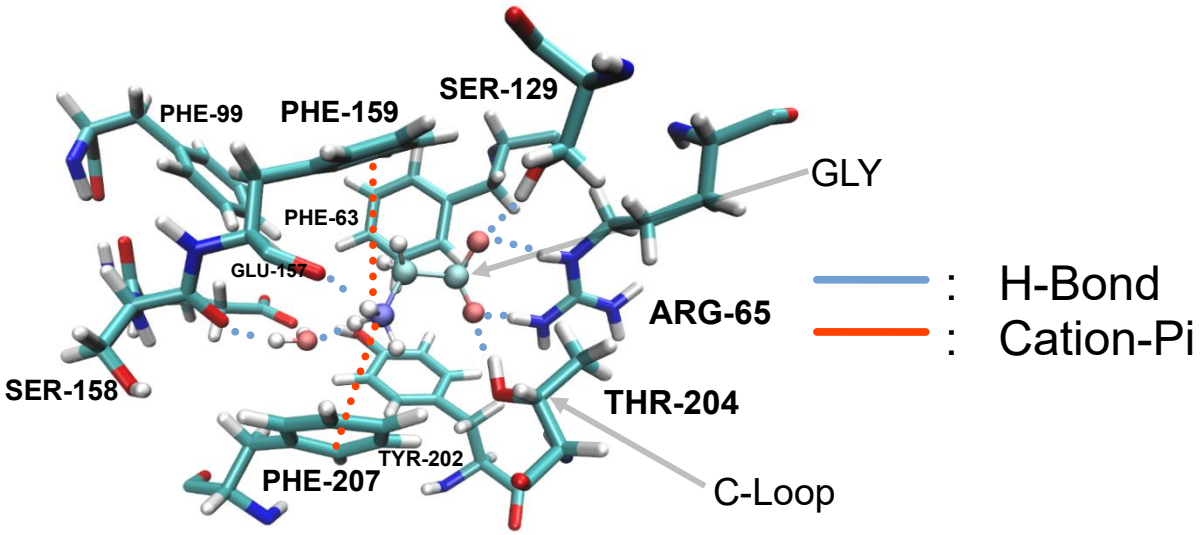

Figure 1.6: Diagram of the binding pocket in the glycine receptor bound with the full agonist GLY, imaged, and taken from the 6PM5 structure[26]. Nitrogen is coloured blue, oxygen red, hydrogen white and carbon teal, the dotted lines represent interactions. The principal binding pocket residues are also labelled appropriately.

Comparisons between Apo (unbound) structures and the agonist-bound structures have provided insights into what changes upon binding, specifically, the pocket undergoes a contraction along with movement of Loop C from an “open” position away from the pocket to a “closed” position. It is hypothesised that this contraction of residues around the bound ligand leads to conformational changes that travel down the protein and mediate gating. The β8-9, β1-2 loops and pre-M1 region are repositioned and the M2-M3 loop acts as a bridge between the TMD and ECD. At the TMD itself, each subunit rotates counterclockwise by 8.6˚ which in turn expands the pore, opening the channel for ion permeation[26].

Whilst the more common states including the open/desensitised conformational states have been profiled, the finer molecular-scale details of the ligand binding and dynamics between these states is less clear. Recent evidence suggests that upon binding, the GlyR enters intermediate states where the channel is still shut in a position along the path of conformational changes required to open[41]. Whilst the open and desensitised states have both been analysed via electron microscopy, neither the intermediate states or the specifics of the changes that occur between activation and gating are clearly defined due to limits of structural imaging[29].

1.3.4 Partial and Full Agonism

There are a wide array of compounds that can affect glycine receptors, many of which that do not involve the binding pocket that glycine interacts with. These allosteric binding sites allow compounds like cannabinoids, zinc, steroids and other drugs to bind and modulate the behaviour of the receptor[42]. Glycine, antagonists and partial agonists however, impart their effect through the binding region at the interface of subunits.

The typical endogenous ligand that binds to the GlyR being neurotransmitter glycine. Both a binding pocket and a ligand can have multiple targets/specific compounds they efficaciously bind to. The development of pharmaceuticals depends on this and struggles to deal with its more negative consequences, i.e dealing with off-target effects[43]. The GlyR is no different, having a range of both synthetic and endogenous ligands capable of binding to it. Agonists are those ligands that elicit a response from a target when bound to the orthosteric pocket; in the case of pentameric ligand gated ion channels like the GlyR this involves a conformational change, rotation of subunits and opening of the channel[26]. A key part of understanding these receptors is knowing what makes a potent agonist, i.e. how a ligand binding is made strong and then how it instigates a conformational change and the intended functional consequences. This is a particularly pertinent line of inquiry in drug development where the desired end goal is likely going to be a ligand that has both a high affinity and efficacy.

Glycine is a strong, efficacious, full agonist. This means that it has a high affinity for the binding pocket and triggers a strong response when it does bind. The interactions it forms with the residues depicted above are likely critical in either the binding process, holding the ligand in the pocket, mediating conformational changes, or likely all three. What is poorly understood is the dynamics of this process, and in particular which elements of the dynamics make for a potent agonist like glycine. Comparisons to partial agonists is a commonplace method for gaining insights into receptors. This has been done in pLGICs, particularly with experimental work [44], where structural imaging has revealed key elements of the gating mechanism in the form of intermediate structure. Theoretical work using simulations has also been applied to similar problems, where molecular dynamics is used to investigate the binding under dynamic physiological conditions between agonists and partial agonists[45], but this has not yet been done in GlyRs.

Whilst there are several partial agonists available to study[26], a common choice of partial agonist to look to for comparisons with the binding of the GlyR is the other endogenous ligand, γ-aminobutyric acid (GABA). GABA is a partial agonist, and its action is considerably weaker than glycine, but is still capable of binding and inducing a response. The key interactions suggested by the structural work are illustrated in Figure 1.7; this was obtained in a similar way to that of glycine in Figure 1.6, from structural snapshots utilising the cryo-EM technique at around a 3Å resolution. GABA is a larger molecule than glycine and prevents the contraction of the binding pocket as much as glycine does when it binds. Chemically, GABA is very similar to glycine, possessing similar hydrogen bonding and salt bridge formation with the arginine and glutamic acid residues, but with additional CH2 groups and therefore a perfect choice of molecule to compare against[46].

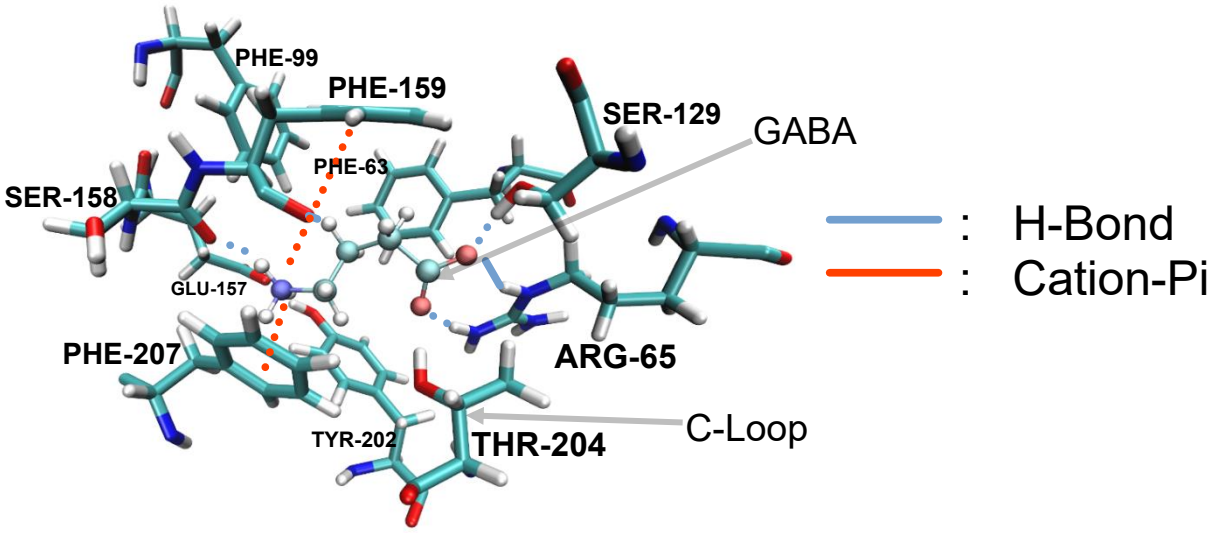

Figure 1.7: Diagram of the binding pocket in the glycine receptor bound with the partial agonist GABA, imaged and taken from the 6PLX structure[26]. Nitrogen is coloured blue, oxygen red and carbon teal, the dotted lines representing interactions. The key binding pocket residues are also labelled.

Stepping beyond the structural snapshots of the pockets, many functional experiments have been carried out to provide a clearer understanding of what these structural changes mean on a cellular level. Both whole cell and Single channel electrophysiology recordings have been carried out for glycine and the partial agonists GABA, taurine, and β-alanine to demonstrate the efficacy and affinity of these ligands to the GlyR. Functional data has found that glycine has a much higher maximum opening probability (Po) of 98 ± 1% compared to taurine (66 ± 3%) and GABA (39 ± 3%). The effective concentrations (EC_{50}) values also show the high

efficacy of glycine (190 ± 20 μM) compared to taurine and GABA at 1.05 ± 0.08 mM and 28.4 ± 0.9 mM respectively (Figure 1.8)[26].

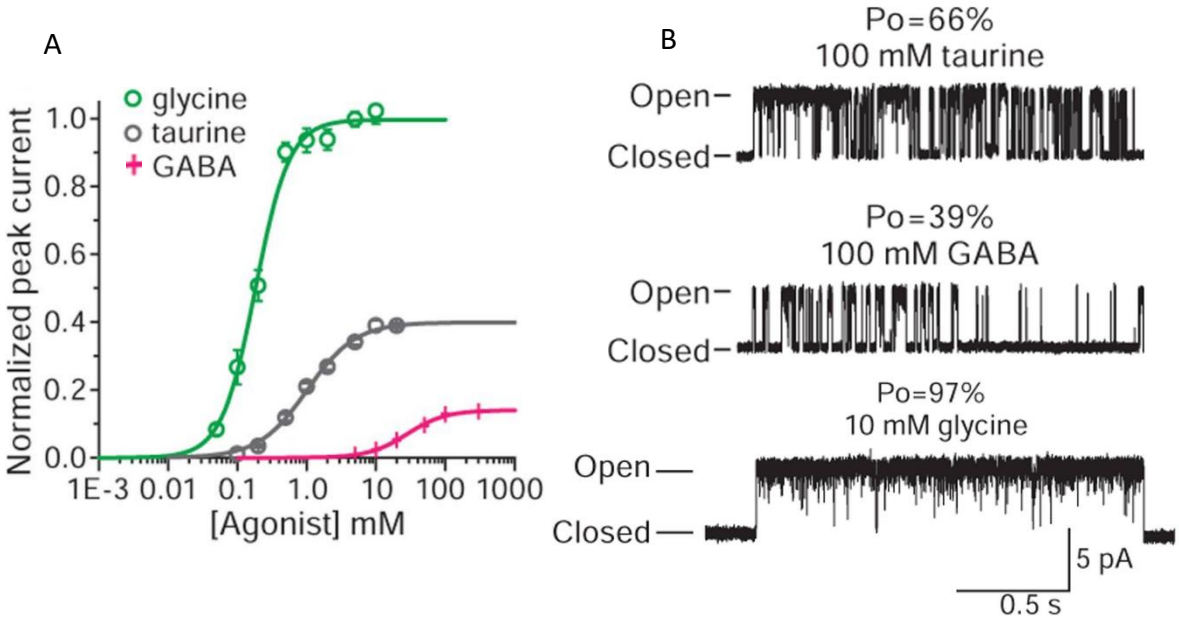

Figure 1.8: Functional data derived from electrophysiology experiments carried out on α-1 GlyRs. A, The peak current plotted against concentration of agonist, measured from whole cells with the EC50s being 190 ± 20 μM, 1050 ± 80 μM and 28.4 ± 0.9 mM for glycine, taurine and GABA respectively, the curves are normalised against the maximum glycine current recorded in each cell. B, the single channel recording data of GlyR openings elicited by 100 mM of taurine, 100mM of GABA and 10 mM glycine. The data and plots represented here were taken from Yu et al 2021[26].

Partial agonists have in the past been shown to be limited in efficacy by a conformational state following binding that remains closed for a period of time slowly transitioning to a 'flipped' pre-open state. Full agonists like glycine however, rapidly transition and can open much faster[41]. Structures have now been resolved of this pre-open conformational state that has not been seen in glycine bound receptors due to the short lived nature of the state being difficult to capture in cryo-EM. From this data, it has been hypothesised that the closed states captured of partial agonist bound structures are the states that preclude the 'flipping' change seen in single-channel recordings. Aligning with the efficacy of GABA and taurine as partial agonists, the fraction of particles in an open/desensitised state are smaller for taurine compared to glycine and even less with GABA bound receptors.

The structural data collected specifically in relation to the binding pocket shows a contraction between the closed and open states that is more extreme in glycine bound receptors when compared to the partial agonists. The general interactions holding the ligands however remains very similar with agonist carboxylate and ammonium groups forming multiple hydrogen bonds and cation-π interactions with the binding pocket residues[26].

Beyond single snapshots and functional experimental data illustrating the large changes between conformations and the differences in affinity and efficacy, it is still poorly understood what mechanisms occur when a ligand binds to mediate changes and how the process differs to yield such drastic functional disparities between the partial and full agonists. Illuminating what a full and partial agonist does when it binds and what specific residues are impacted in the proximity of the binding pocket could lend substantial insights into what makes a full agonist have such a high efficacy and affinity, concepts critical for both general understanding of pLGICs and in developing pharmaceuticals.

1.4 Mutants

Whilst the study of how differences in a ligand can provide us insights into the mechanics of ligand-protein interactions, alterations in the chemical moieties in the protein can also be used for mechanistic insights. According to the central dogma of molecular biology, and the vast amount of evidence for the process, DNA holds the information required for protein synthesis in a three-letter codon system; this is transcribed into RNA via the action of RNA polymerase and translated at ribosomes to produce polypeptide chains in the order required for constructing proteins like the $GlyR^{2}$.

Mutations are what happens when the code at the DNA level is changed; this can be either silent/neutral or more impactful resulting in a change to the polypeptide sequence and structure ultimately formed. The general process of protein synthesis as well as how a mutation may come about is shown, simply presented, in Figure 1.9. There are many ways that mutations can have an impact and can include a range of stages through the pathway of expression from DNA to protein. For our purposes, we are mostly interested in mutations such as single base-pair substitutions and the gene-associated polymorphisms. In nature, these single base-pair substitutions commonly occur as a result of spontaneous deamination[47]. When a nucleotide is switched, it will produce different RNA and then potentially result in a different residue in the polypeptide chain.

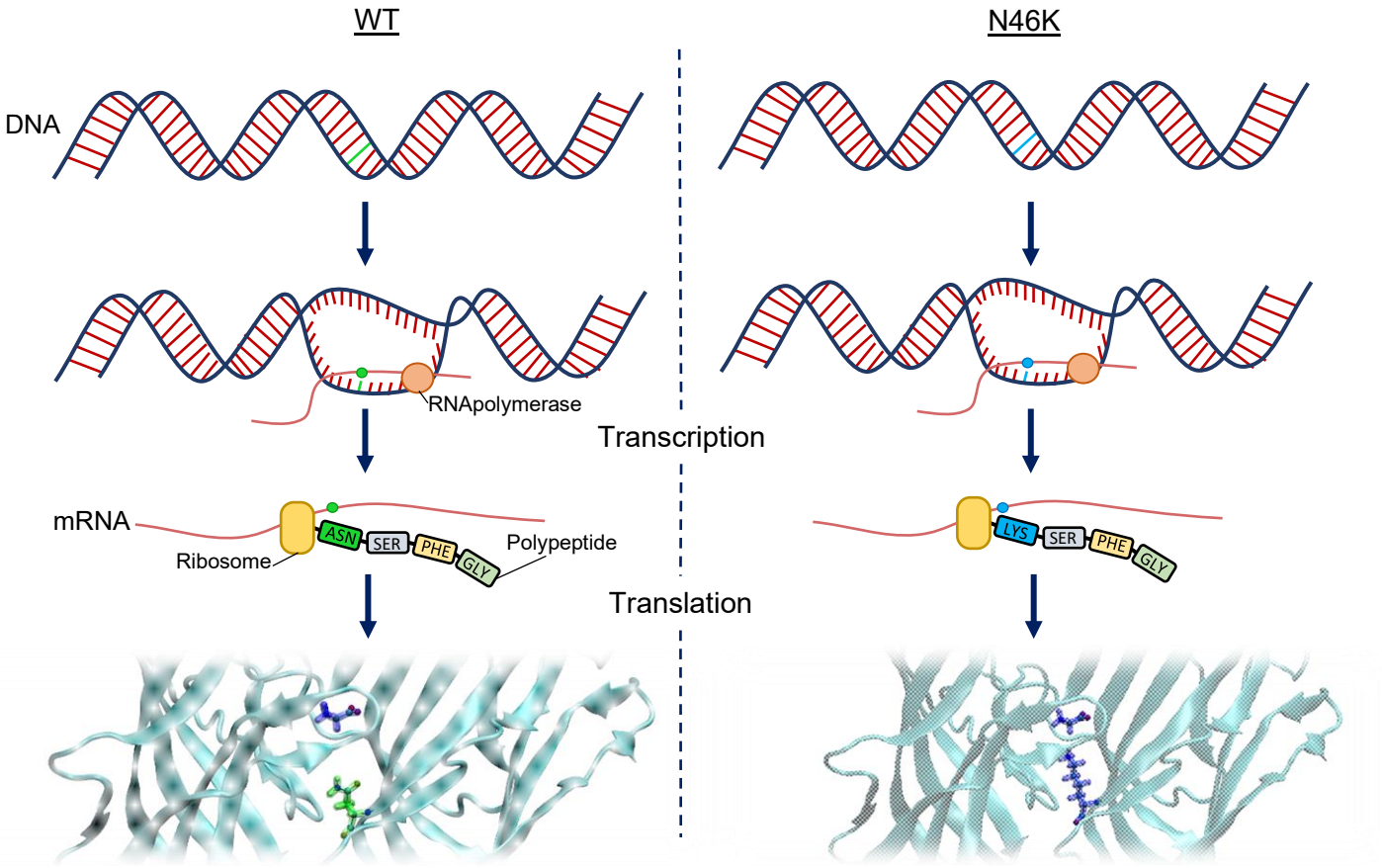

Figure 1.9: The dogma of molecular biology, showing the progression of DNA to RNA to polypeptide protein. Here the wildtype glycine receptor is shown on the left and N46K mutant on the right, with a cartoon representation of how the substitution would come about and the repercussions of that in structural form in the bottom images.

Both natural and synthetically induced mutants are extremely useful for investigating processes like ligand binding and gating of ion channels. Whilst many of the natural mutants become known of via identification of symptoms first such as with hyperekplexia in the case of GlyRs[48], the more systematic and experimental implementation of site directed mutagenesis has been applied effectively to probe the function and structural role of residues[49]. The utility of such mutations comes from making direct comparisons with the wild type form, when only a single residue changes in this way, we can see precisely what its role is with both structural and functional observations.

Mutations in GlyRs that lead to dysfunction of the channel can result in inhibited glycinergic neurotransmission. One phenotype that arises from these dysfunctional mutant isoforms is hyperekplexia/startle disease, a serious neurological condition affecting newborn children and is characterised by hypertonia, exaggerated startle response and intellectual disability[50]. In the 1990s, hyperekplexia was shown to have hereditary mutations in the GLRA1 gene that

encodes the GlyR in humans. Since then, the functional effects of mutations on GlyRs and the links with hyperekplexia have been studied in much greater depth, providing information on a much wider range of mutants[50,51].

In the case of the GlyR, mutations have been studied across the length of the protein. Several of these mutants that have links to hyperekplexia are highlighted in Figure 1.10 along with their location on the ECD [52]. Typically, the region investigated corresponds to the area of functionality that is of interest. The intracellular and transmembrane domain mutations often impact the efficacy of a receptor, whereas mutations in the extracellular domain are more likely to alter the affinity. To make claims about how a specific mutation may impact the efficacy or affinity of a given ligand on a protein, the two must be clearly separated experimentally which is both difficult and rarely done.

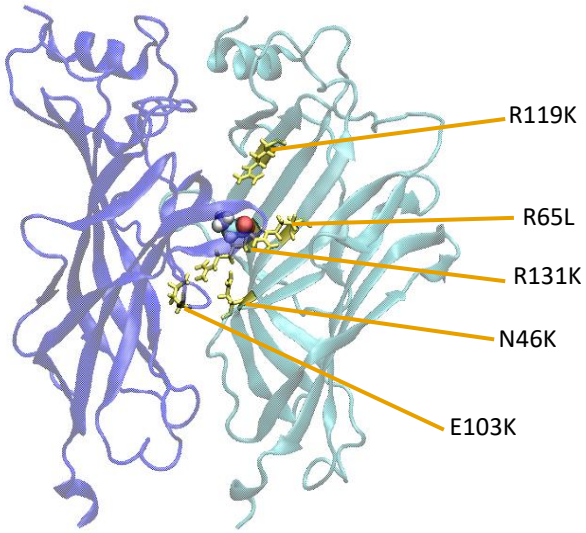

Figure 1.10: Cartoon representation of the alpha-1 glycine receptor extracellular domain with several potent mutant variants that are of interest for investigating receptor function.

To understand how various residues may play a role in ligand binding and the most upstream elements of mediating a conformational change, mutants in and around the binding pocket are key candidates to study. In the case of the GlyR, these include mutants such as the spasmodic, oscillator, Cincinnati and Nmf11 mutants[44]. In the spasmodic form, a missense mutation leads to reduced glycine sensitivity and co-operativity of binding. The oscillator and Cincinnati mutations involves an exon microdeletion or exon duplication leading to complete, severe, loss of function. The Nmf11 mutant is also lethal, involving a missense mutation of asparagine to lysine. Investigation of the N46K mutant is therefore an opportunity to highlight

the role of a lesser-known component of GlyR ligand binding and potentially yield greater insights as to its role in ligand binding[44].

Electrophysiology experimental work measuring EC50 values for GlyR wild types and the N46K mutant has demonstrated a substantial jump from 41 μM in the wild type to 372 μM in the mutant N46K. However, the maximal glycine activated current was found to be unchanged by the mutant. This is typically a strong indicator that the mutant is not acting via impeding gating or the function of the receptor opening but instead it is altering the binding. Substitution of the N46 residue for those without an amide moiety led to much more severe effects, leading to the suggestion being made that the amide may interact with residues on adjacent loops[44,52].

This N46 residue is located below the binding pocket, beneath residue F63. The illustration in Figure 1.11 shows what changes in the N46K mutant. The potency of this N46K mutation is of particular interest as it is not a principal binding residue of the binding site and is only in proximity to the binding site near the subunit-subunit interface, not in any binding loop or transduction pathway. The impact of this mutant has been studied experimentally, with work pointing towards mutation reducing affinity. The data used for this conclusion involved electrophysiology recordings made with glycine and partial agonist bound wild type (WT) and mutant receptors, specifically observing how the mutation shifts the EC50 of agonists vs partial agonists. If a mutation is impacting binding efficacy, it will lead to a rightward shift of a concentration-response curve in full agonists and a reduction in the maximal response of partial agonists. This is not evident in Figure 1.11 where the maximal responses are not significantly changed in the mutant form. Experiments like these have demonstrated that N46 may have an important role in the binding affinity of GlyRs.

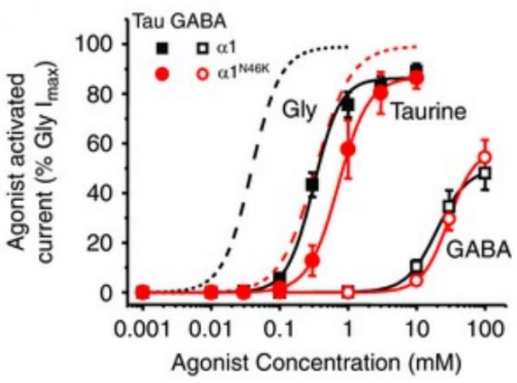

Figure 1.11: Functional data derived from electrophysiology experiments carried out on WT and mutant α-1 GlyRs, the WT in black and the N46K mutants in red. The maximal responses are normalised to the maximum glycine response.This figure was taken from Wilkins et al (2016)[44]

Studying the dynamics of this mutant variant could provide specifics as to how this residue acts, whether it is providing stability to the pocket via interactions with principal binding residues or whether it has some other structural/functional role to play in binding.

1.5 Structural Biology

A key aspect of understanding how proteins function lies within the structure and the dynamics at a molecular and atomic level. Whilst functional experiments can provide us great insights into what a protein does and what similarities proteins share, experiments have limited capacity to uncover how exactly they work. Clear and precise mechanical details of the dynamics occurring are necessary for a comprehensive understanding of protein function. In the case of GlyRs for example, how a ligand binds with a high affinity and elicit such potent effects require full topographical details of the binding pocket, ligand path and binding/unbinding dynamics. The foundation to gaining such an understanding starts with structural biology and the utilisation of experimental imaging techniques.

1.5.1 Imaging Techniques

X-ray crystallography, in its first iteration recognisable by modern techniques for biological systems, was carried out in the early 1930s on protein crystals; since then, it has grown to become a staple technique for protein structure determination. Fundamentally, X-ray crystallography relies on interactions between the X-rays directed through a protein and the electrons within the atoms of the structure. The resulting output is an X-ray diffraction pattern that can be analysed to derive the electron density map, from this atoms are located and interpreted into an atomic structure which can then be deposited in the protein data bank[53] where the average resolution for membrane proteins is 2.8 Å[54]. This all relies on Bragg's law of X-ray diffraction by crystals and so requires the protein sample to be in a crystal state. This preparation phase of crystalising protein can be both difficult, time-consuming, and costly depending on the protein of interest. This includes a requirement for additional modifications to be made to a structure for it to crystallise, causing the structure to deviate from the biologically relevant state[55]

Structural characterisation of membrane problems remains a substantial challenge, primarily due to difficulties in getting high yields synthesised, their poor stability and issues associated with membrane extraction and solubilisation. This is evident in the fact that only 3% of all deposited structures in the PDB are membrane proteins with an average resolution of 0.6Å lower than soluble protein. However, the regular improvement of newer single-particle cryogenic electron microscopy (Cryo-EM) has led to this being viewed as the optimal method for membrane protein structure derivation. Cryo-EM utilises high energy electrons fired through thin samples to discern structure. Whilst currently the average resolution of membrane proteins with cryo-EM is lower at 3.4 Å, the strong interaction of electrons with atoms can lead to a higher resolutions than X-ray crystallography, at near-atomic levels even reaching 2.45 Å when imaging protein in a micelle[54]. This approach requires much less sample, less preparation, with samples for Cryo-EM being prepared directly from biological material by flash freezing. The main advantages include: The method is not limited by the need to form protein crystals, conformational states can be inducted in physiological-like conditions, conformational heterogeneity can be observed and bilayers can be mimicked in the process. [54]

Accurate structures are particularly important for simulations with kinetics, and binding of ligands is involved. Glycine is one of the smallest ligands in biological systems, and it can be assumed that when designing an atomistic simulation, even small disruptions in the binding pocket may lead to openings for dissociation. For this reason, the highest resolution and most biologically accurate binding pocket is desirable; therefore, Cryo-EM structures are the optimal choice if high resolution structures of the desired state are available.

1.5.2 Available Structures

The availability of structures of good resolution for pLGICs is relatively recent and improving. For the GlyR, there are several available structures, including both cryo-EM resolved and X-ray crystallography structures. Choosing a structure depends on the relevance of what functional state the structure is presumed to be in and the quality of the deposited data. The assessment of data quality is based on a scoring system integrated into the protein data bank and is what we used to guide our decision on a structure.[56].

The clash score is a rating of how many pairs of atoms are unrealistically close to one another and are hence clashing in space and is reported in terms of how many clashes are present per thousand atoms. The Ramachandran outliers is a recording of the percentage of the system that has atypical Phi and Psi torsion angles. The Sidechain outliers are similar to

Ramachandran outliers but are a measure of how many protein side chains are abnormal in their torsional arrangement in space[56]. These metrics are all calculated with Molprobity[57].

Early in this project, the newest and reliable structure available was an X-ray structure bound with strychnine at a 3 Å resolution and was reported with a clash score of 4, Ramachandran outliers at 0.5% and sidechain outliers at 2.8%, which are all considered above average for similar structures at this resolution[15]. This original structure had a few issues that are detailed in later chapters but was the best system at the time and chosen on the basis of the quality of the structure derived, resolution and the presence of a ligand in the binding pocket that could be replaced with glycine. This structure was an alpha-3 glycine receptor bound with strychnine and captured with X-ray crystallography (5CFB[15]). As the project progressed new structures were also released which were better suited for investigating partial and full agonism. The structures chosen for this work are presented in table 1.1. The SMA is a tool for detergent free solubilisation of membrane proteins using styrene-maleic acid copolymers to function as a mimic of a real membrane for the protein \to embed into, common in Cryo-EM preparations.[15] Generally, resolution is an important indicator of the quality of a structure; this is not universal and therefore the other various scoring systems were considered when choosing appropriate structures for setting up systems. There are multiple states that these receptors exist in; in fact, the original ideas surrounding the mechanism of gating suggested a multistep process. Recent evidence building on this idea indicates there may be a complex landscape of one or more pre-open intermediates. The original steps have been structurally profiled in experiments, including the closed (resting), open (active), closed (desensitised)[26,58,59]. The 'intermediate' states that exist between each of these conformations have been more difficult to capture structurally, similarly to the glycine bound-closed structure, as they are likely extremely short-lived and thus far poorly characterised. From these experimental studies, a large body of information regarding the GlyRs has been formulated, enough for a mechanistic system of how the various changes between states occur. The species and genetic variant of the channel are other important consideration as differences can include alternate residues which behave differently and therefore makes comparative analysis difficult[60].

PDB ID	Uniprot AC	Organism	Resolution (Å)	Ligand	State	Method	Membrane-mimic	Ref
5CFB	O75311	Human	3.00	Strychnine	Closed	X-Ray	NA	15
6PM5	O93430	Zebrafish	2.90	Glycine	Desensitised	Cryo-EM	SMA	26
6PLX	O93430	Zebrafish	3.20	GABA	Desensitised	Cryo-EM	SMA	26

Table 1.1: A table of structures used in this work along with the key pieces of information.

The factors described were the main driving force for leading us choose the desensitised (PDB entry: 6PM5) structure for our second glycine bound system[26] that was intended to serve as a control against the mutant system and partial agonist systems that could be set up with the

same structure. As well as this, glycine has not been captured bound with the receptor in a closed/resting state and to ensure we had other structures bound to compare against we chose one with as many available ligands already bound and imaged. The desensitised structure and open structure are extremely similar, especially in the area around the binding pocket is, and given that the focus of this work is on the binding process and mechanics, the desensitised structure is a reasonable choice. Ultimately, the important aspect was to maintain the same state between system comparisons and to maintain consistency as all states would be viable options for observing binding/unbinding events. Other cryo-EM structures were released alongside the 6PM5 PDB, including a glycine bound super-open state and the apo (unbound) state. The super open state was disregarded, as it has a pore diameter too large to be considered a physiologically relevant structure[26]. The apo structure does not contain a glycine ligand and so would have required docking, an unnecessary step given the availability of the bound 6PM5. The analysis of these structures in previous work has given rise to a better understanding of the mechanism of binding and how conformational changes transition between apo and bound states, specifically with regards to the movement of the C-loop capping the binding pocket and the constriction of the pocket[26]. Here in figure figure 1.12 the differences are highlighted via superimposing the cryo-EM structures across each other.

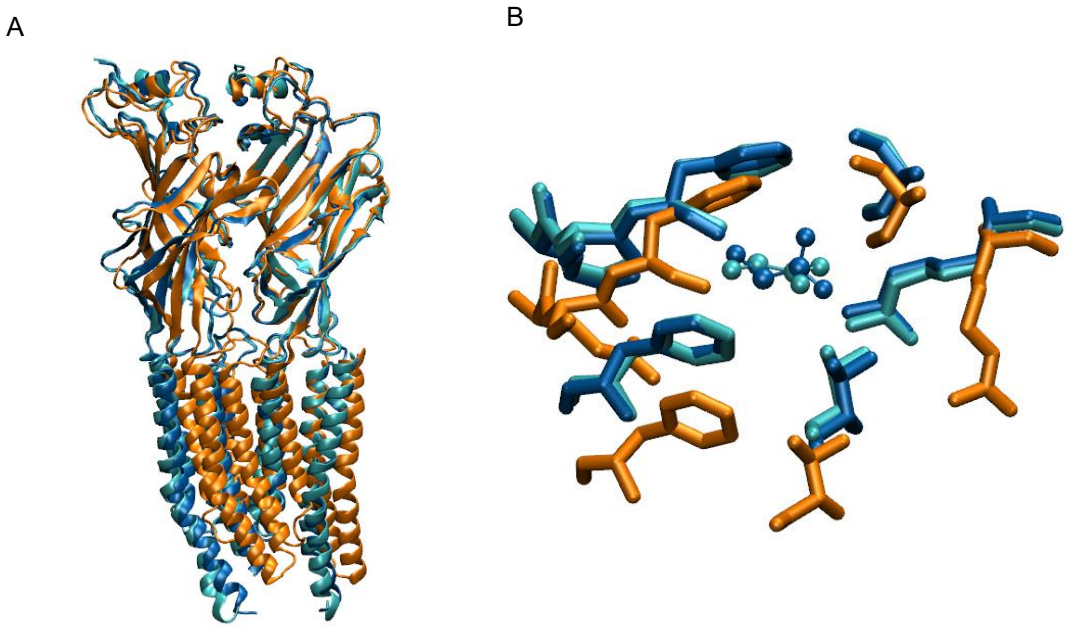

Figure 1.12: A comparative illustration of the aligned desensitised (blue), apo(orange) and open state cyan alpha-1 glycine receptor. In A, the structures shown from cryo-EM with the subunit pair across a binding pocket shown. In B, a closer look at the differences between the desensitised, open and apo binding pockets. the dotted lines represent interactions.

The differences in the figure are very minimal between the desensitised and open states whereas there are more considerable changes in the arrangement of the apo pocket. The receptor populates both the desensitised and open state when glycine is present, the closed state occurring transiently when glycine is first present to bind. The desensitised and open states are very similar, with the most notable difference being the constriction of the pore region of the transmembrane domain, with a reported pore diameter of 3 Å indicative of its non-conductive state[26]. The biggest structural differences arise from the Apo structure where we see the constriction of the binding pocket and capping of the C-loop when comparing the bound states to the apo structure. Given the similarities in the binding pocket, the lack of a transmembrane domain removing most of the bigger differences structure-wide between the open and desensitised states, the desensitised structure was chosen as the bases for the work in this thesis. These and other structures have been used to elucidate a general idea of how glycine receptors respond to ligand binding to activate and produce functional effects. The process of activation is understood to start with binding of the ligand, this stabilises the ECD in a contracted high-affinity state where conformational changes then propagate down to the pore where twisting of the pore-lining helices widens the channel in the transmembrane region, allowing ions to flow[61]. This mechanism is an extremely complex and sophisticated process that's not fully understood with a landscape of open, pre-open and closed intermediate states[26].

The additional PDB structures chosen were those from cryo-EM work carried out in Yu et al[26], 6PLX for the GABA bound structure in the desensitised state to function as a system of a partial agonism. The 6PM5 glycine system was also modified and used as the basis for the N46K mutant system. The limit of experimental approaches is linked to the temporal and spatial resolution of these techniques; structural imaging can only characterise a structure in one specific frozen state, whereas the functional studies can provide high-quality temporal resolution data but without the structural aspects. To study the receptor in greater detail beyond the limitations of conventional imaging techniques, atomistic simulations are a very powerful tool.

1.6 Simulating Ion Channels

As for what simulation work has been carried out with GlyRs and pLGICs in general, a range of approaches and techniques have been used. Prior to the development of cryo-EM and more refined X-ray crystallographic structures, homology modelling was a viable option to describe

the structures of GlyRs exploiting the similarities with the very limited structural information with other pLGICs and, for the ECD, with the acetylcholine binding protein. A structure would be created via homology modelling combined with molecular dynamics to also validate the system stability.[62,63]. The newer, more accurate structures, provide unprecedented opportunities to build reliable starting systems for atomistic simulations. These newer structures combined with improved simulation methods and computer power, can make significant contributions to the understanding of pLGIC working mechanisms at the molecular level. This could include but is not limited to: exploring the ligand binding path, characterising the binding kinetics of novel ligands, or even simulating the chain of conformational changes that lead to gating.

1.6.1 Simulations of GlyRs

Many atomistic simulations have been carried out on pLGICs, focussing on various aspects, like ion permeation[64], potential gating mechanisms[65], interactions with the cell membranes[66] etc... Work that focused specifically on GlyRs has used computational methods to aid in understanding, validation and physiological relevance of the various structural states identified in cryo-EM and X-ray crystallography studies[67,68]. For example, the work by Dämgen et al[67] utilised molecular dynamics to refine the experimental structure and identify regions structurally unclear. Their work focused on leucine residues that are poorly resolved and simulated how these components of the structure could be configured to influence the function of the channel, something that is very difficult to accomplish with classic experimental approaches. Similarly, the work Gibbs et al applied simulation methods to explore pore function specifically in various structural conformations, something also very difficult to achieve experimentally [68].

Similar approaches have even gone a step further and used methods to compare computational and experimental functional recordings. By measuring conductance and comparing the values to electrophysiological experiments done in vitro, the conclusions made about a structure in simulations can be supported and validated. Work done by Cerdan et al[69] used computational electrophysiology and polyatomic anion permeation simulations to validate GlyR structural systems from experimental data sets[69]. They were able to conclude which structures were physiologically relevant by identifying and correcting for structural artefacts by simulating the protein in a physiological and native environment. This work exemplifies what can be achieved by combining experimental datasets with simulation work to go beyond the scope of conventional experimental methods.

Some of the most recent work that has been carried out with GlyRs that utilised this approach of combining cryo-EM and MD simulations was carried out with three separate structures, including GlyR at rest, a wide-open glycine bound structure, as well as a PTX and glycine bound structure[70].Here the work similarly focused on the differences in conductance and the shape of the pore between the three states. However, this work was carried out on relatively short timescales of 200 ns and below, as well as with considerable restraints across the full structures. Whilst the results gained from this work did fit well with previous work, long equilibration and production times can provide greater validity to computational systems by demonstrating the structure is stable and covers all conformational space[71–73].

1.6.2 Investigating GlyR Ligand Binding

Whilst the glycine set in the pocket has been simulated[67], what has not been probed in depth in GlyRs is the ligand binding and unbinding process and the corresponding paths. The most recent work to investigating binding events is by Shi et al[65] , where molecular dynamics was employed to measure structural stability, flexibility as well as docking rating of open, closed and apo structures, albeit using strychnine. Besides molecular dynamics, there are other methods which can provide greater insights into ligand binding of GlyRs. These include enhanced sampling methods, particularly metadynamics, to accelerate rare events and evaluate free energies.

Funnel metadynamics[74] is a specific method useful for ligand-protein systems which efficiently explore the ligand-binding free energy landscape, by limiting the region to be explored within the receptor and in the solvent. This method has already been used with pLGICs successfully, specifically for the insect GABA-activated RDL receptor (wild-type and selected mutants), providing a protocol for studying the binding of ligands in pLGICs further[46]. A key aspect of ligand binding and unbinding to proteins is the kinetics of the system, which is very challenging to evaluate computationally with a current effort focused on developing methods to do this. Computer simulations may allow for an estimation of binding and unbinding rate values that can, like conductance, be compared to experimental data. Metadynamics may be used to evaluated kinetics in some cases and under certain conditions, as exemplified by an application on p38 MAP kinase[75] . If applicable to GlyRs, it could yield valuable insights into the binding/unbinding pathways and rates, which can be compared with experimental data using electrophysiology methods[58]. In fact, among pLGICs, GlyRs are suitable for experimental kinetics measurements.

Using the methods and insights gained from previous work here, a more comprehensive and in depth understanding of the mechanics of GlyRs, and by extension pLGICs, can be gained. The most recent structural data sets of the GlyR can be used to develop an atomistic system of the protein in a dynamic environment, bound with various agonists or altered by important mutations. These systems can then have enhanced sampling techniques applied to them to produce trajectories including the unbinding/binding events. Observations made with the free energy landscapes derived from such work along with quantifying the changing interactions throughout the process being simulated would provide a wealth of insights to further develop current understanding of ligand binding. As previously mentioned, the differences between partial and full agonists and the dynamics of various mutant isoforms have not been fully investigated, especially with such high spatial resolution like that of atomistic simulations. Investigating agonism and mutants in this way provides a level of understanding that so far experimental approaches have limited scope for and would provide insights not only into these specific phenomena but also the functioning of GlyRs and pLGICs in general.

This is the aim of the work presented here, to expand the understanding of partial/full agonism in GlyRs, the mechanism of action of mutants and in more general terms the role of various components within and around the binding pocket, the solvent, and the C-loop structure[26,36,67]. The next chapter discusses the details of the methodologies used in this thesis, along with justification for the approaches taken. Chapter 3 outlines the development of the GlyR systems and discusses the results and insights gained from investigating the binding pocket using molecular dynamics. Chapter 4 and 5 discuss the results of applying funnel metadynamics to the GlyR systems, specifically in the context of studying partial agonism in Chapter 4 and the mutant isoform N46K in Chapter 5. Finally, Chapter 6 includes a conclusion of the results garnered over the entirety of the thesis as well as the potential moving with these results and how the findings can be applied.

Chapter 2

Methods

2.1 Biomolecular simulations

Despite the computational demands and modern hardware required for condensed matter simulations at the atomic/molecular level, the methodology is not recent. The two key fundamental aspects of molecular simulations, Monte Carlo sampling (MC) and Molecular Dynamics (MD), were in fact introduced in the 1950s. MC is used to estimate the average value of a property in a system by sampling across a group of configurations randomly generated with a probability distribution that matches an appropriate statistical mechanics ensemble. MD is even simpler: it involves solving Newton's equations of motion for atoms within a system, allowing for the observation of their dynamic evolution[76]. Some of the first simulations conducted were carried out as early as 1950 and the first MD simulation of a fluid was published in 1957[77]. These original methods contained most of the elements of modern-day simulation albeit on different system scales and with added and more refined algorithms. Since these early ventures, molecular simulations have been firmly accepted as a key method for scientific inquiry, with a steadily growing increase in publications combining simulation tools with experimental methods[76,77]. The work presented in this thesis is no different and in this chapter the specifics of which molecular simulation methods were applied will be explored.

2.1.1 Why Use Computational Methods

There are several reasons why computational simulations have had such a steady growth in appeal; however, it is important to understand what drives the need for these methods, which is the limitations of experimental methods. Imaging techniques such as X-ray crystallography and Cryo-EM are invaluable tools for deriving the structural details of biomolecules but they are insufficient for understanding the dynamics at play on the atomic level. For a long time, there has been a consensus that proteins do not exist in a single rigid structure but in a distribution of possible structures that is dependent on the free energy needed to shift between states. It has been suggested that when a ligand binds, it alters the conformational equilibrium[78]. The fundamental nature of proteins is dynamic; therefore, the methods used to gain insights into their functionality and mechanisms must be able to capture this nature.

Whilst experimental methods have provided unparalleled insights into the structure of proteins, to go further and understand the nuanced changes that occur when a ligand binds or unbinds and how it may be acting on the conformational equilibria requires greater spatial and near real-time temporal resolution. The various scales of both temporal and spatial resolution that can be reached in molecular modelling can be seen illustrated in Figure 2.1.

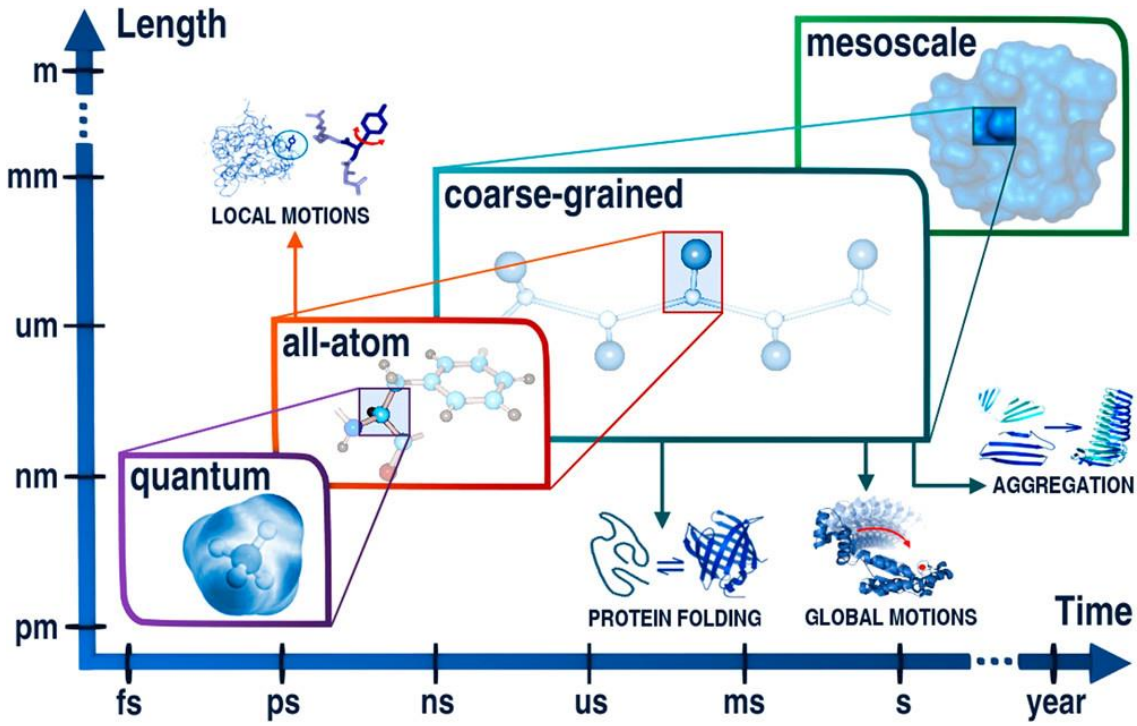

Figure 2.1: General applications of molecular modelling at various resolutions and timescales. The work presented here involves all-atom simulations on a nano and microsecond timescale. Diagram taken from Kmiecik et al[172]

There is a range of computational methods available, each with their own specific applications. Deciding which tool to use depends on both the questions being investigated and the capabilities of each method. The range of tools can be broken down based on the scale they function at. Ab initio (quantum-mechanical based) methods are the most detailed, allowing the explicit simulation of electrons in a system. Classical all-atom molecular dynamics is the next step up in scale, which uses forcefields to system interatomic interactions. Then, there are coarse-grained simulations where multiple atoms are grouped together and simulated as a single entity or “bead.” Ab initio methods would be the ideal choice for any simulation as they provide the closest emulation to real world physics; however, they come with a rather large caveat. Ab initio techniques are incredibly computationally intensive and so large protein systems are completely impractical to simulate in this way. modelling chemical reactions and relatively small systems however is what ab initio methods excel at in terms of how accurately they can simulate interactions. Another popular approach is the utilisation of hybrid quantum

mechanics/molecular mechanics (QM/MM). QM/MM utilises quantum chemistry theory in a localised region where a given chemical process of interest takes place whilst the rest of the simulation space is dealt with using a molecular mechanics force field. This approach is useful for large systems where a localised chemical reaction is of interest such as in functioning enzyme systems[79]. Classical molecular dynamics still use quantum mechanics simulations but only to create the empirical forcefields that the atomic interactions are based upon.[80] This reduction in computational demand makes forcefield based molecular dynamics much quicker and allows for larger systems like biomolecules to be practically simulated and still reproduce experimental observations[75,81,82].

All-atom molecular dynamics allow simulations of whole proteins in emulated physiological conditions where biological phenomena can be observed on nano/microsecond timescale. There are several specific areas that computational methods can be employed in to gain insights into the functional mechanisms of proteins. By observing key elements of a biomolecule interacting in a dynamic system over time, understanding can be gained as to the function of a given molecule within a larger biological context. In the case of pLGICs for example, observations can be made of region-specific flexibility, blooming, and twisting angles of the structural domains[65]. Another area where insights may be gained is in understanding and characterising the binding mode of a given ligand ligand; this is especially useful for complex ligands that may have multiple conformational states within a pocket or for observations of the role of water in the binding pocket, as this cannot be captured with structural imaging techniques[67,83]. A method that can be coupled with molecular dynamics is metadynamics[84], an enhanced sampling method which allows the free energy landscape of a system to be explored as a function of a few collective variables; this is particularly useful for identifying important states that a protein or complex may reside outside of the static state observed with methods like cryo-EM. In the case of ligand binding, metadynamics can be used to identify multiple binding modes or even pre-binding regions along a binding path that aids in the identification of important residues not in the immediate proximity of the binding pocket or part of the principal binding residues[46,85,86]. The accuracy of utilising molecular simulations in this way has been validated multiple times by reproducing experimental values. An interesting example is in Casasnovas et al, where the kinetics of unbinding for a ligand was derived using molecular dynamics and metadynamics and was in good agreement with experimental data[75].

The work presented here uses molecular dynamics and metadynamics to focus on both the binding mode of ligands and the exploration of the free energy landscapes. In this chapter

both the theory and methodological choices will be explained, from the initialisation of molecular dynamics through to the process of using enhanced sampling methods to derive the free energy landscape and the analysis required to collect relevant information.

2.2 Molecular Dynamics

Molecular dynamics has been used as a technique for simulating molecular systems for over 50 years and, with the growing power of computer hardware, has become a staple for investigating protein structure and function[87]. Fundamentally, MD involves the trajectories of atoms being determined by solving Newton's equations of motions. Typically, the starting positions are set by experimental data taken from imaging with cryo-EM or X-ray crystallography, particularly when biomolecules are involved. The forces that dictate the movements are calculated using forcefields which describe the interactions between atoms in empirical form, with parameters typically derived from experimental data or quantum mechanical calculations. The equations of motion are integrated numerically at discrete 'timesteps,' typically a few femtoseconds, with specific algorithms like the velocity Verlet one which will be described later and is used by the software implemented in this work. This process is carried out for many steps until a period appropriate for the phenomena of interest has been reached. The general architecture of the MD process is as follows: the initial positions of atoms and starting velocities are taken, the forces acting on all the atoms are calculated, Newton's equations of motion are integrated for the time step and then this cycle repeats[88].

2.2.1 Initialisation

The initialisation stage is the first part of this process, where starting positions of atoms and velocities are set. In the case of biomolecules, positions come from structural experiments to make the coordinate system of the atoms. The structure taken from the PDB however does not include all the information, hydrogen atoms are missing, there can be residues missing, solvent and ions also needed to be added prior to carrying out molecular dynamics. The PDB functions as a starting point for molecular simulations, a preparation phase is required to utilise the raw data. The most computationally intensive part of MD follows this, the force calculation stage, where the forces acting are calculated.

2.2.2 Forcefields

The atoms in all atom MD are represented by points in space with a mass, charge and force-field parameters depending on their type, allowing the dynamics to be characterised by Newton's equations of motion. The interactions between atoms are determined by utilising forcefields, which are potential energy functions. The charge can be fixed in these forcefields, a compromise to avoid excessively high computational demand that would make complex simulations impractical. Each atom is given an atom type, depending on the element, its environment and associated bonding, that determines its properties; for example 'H2' would represent an amino hydrogen in an NH_2 group. The forcefields and their parameters, including changes, bond angle, bond length, dihedral, torsional and van der Waals parameters, are originally determined through fitting of the energies of multiple conformations of simple peptides to ab initio calculations. The forcefields used in this work include Amber ff14sb and GAFF [89,90]. The GAFF forcefield was designed to be highly compatible with existing Amber forcefields and provide parameters for most organic molecules comprising of Hydrogen, Carbon, Nitrogen, Oxygen, Sulphur, Phosphorous and halogens. The nonbonded parameters are taken directly from the traditional Amber forcefields and partial charges assigned using a restrained electrostatic potential fit system (RESP). The bond length and angle parameters are derived from both experiments and ab initio calculations, with the force constants are optimised to reproduce experimental and ab initio vibrational frequencies. Torsion parameters were derived with the intention of reproducing rotational profiles found in experimental or ab initio data. Part of the resulting product is 35 basic atom types including five carbon, eight nitrogen, three oxygen, five sulphur, four phosphorus, six hydrogen and chlorine, iodine, bromine, and fluorine types. These atom types are defined by orbital hybridisation, aromatic properties, and chemical environments. This means that one atomic element such as carbon will have multiple atom type names, each with different chemical environments, types of orbital hybridisation or aromatic properties[89]. The ff14sb forcefield used for proteins was developed with the intention of improving upon the previous Amber ff99sb forcefield, with a heavy focus on the weakness of side chain rotamers and backbone secondary structure preferences[90].

Water is a critical component to biological simulations, as its unique set of properties and capacity for hydrogen bonding makes it a vital molecule to effectively emulate in molecular dynamics. Many water systems have been proposed. Some of them, including the one used here, treat water as a rigid molecule; their forcefield includes Lennard-Jones and Coulombic terms but the exact parameters differ from system to system and the differences in water

properties have been found to be significantly different as a result[91]. The system used here is the Transferable Interaction Potential 3-point system (TIP3P) system of water[91]. The system used here is the Transferable Interaction Potential 3-point system (TIP3P) system of water[92]. The geometry of this system is based on gas-phase water and is frequently used in simulations with Amber and OPLS forcefields; in particular, the ff14sb forcefield is optimized to be used with the TIP3P water system. Whilst updated and more computationally expensive systems that reproduce experimental properties of water have been developed, simpler systems like TIP3P are still commonly used due to their low computational cost and relative accuracy[93]. These force fields are designed to allow any interactions to be evaluated across the conformational states of a given system and therefore are invaluable for studying biological systems where important interactions can, in many cases, be purely non-covalent in nature[94]. The diverse range of interactions characterised by the Amber forcefields are illustrated in Figure 2.2.

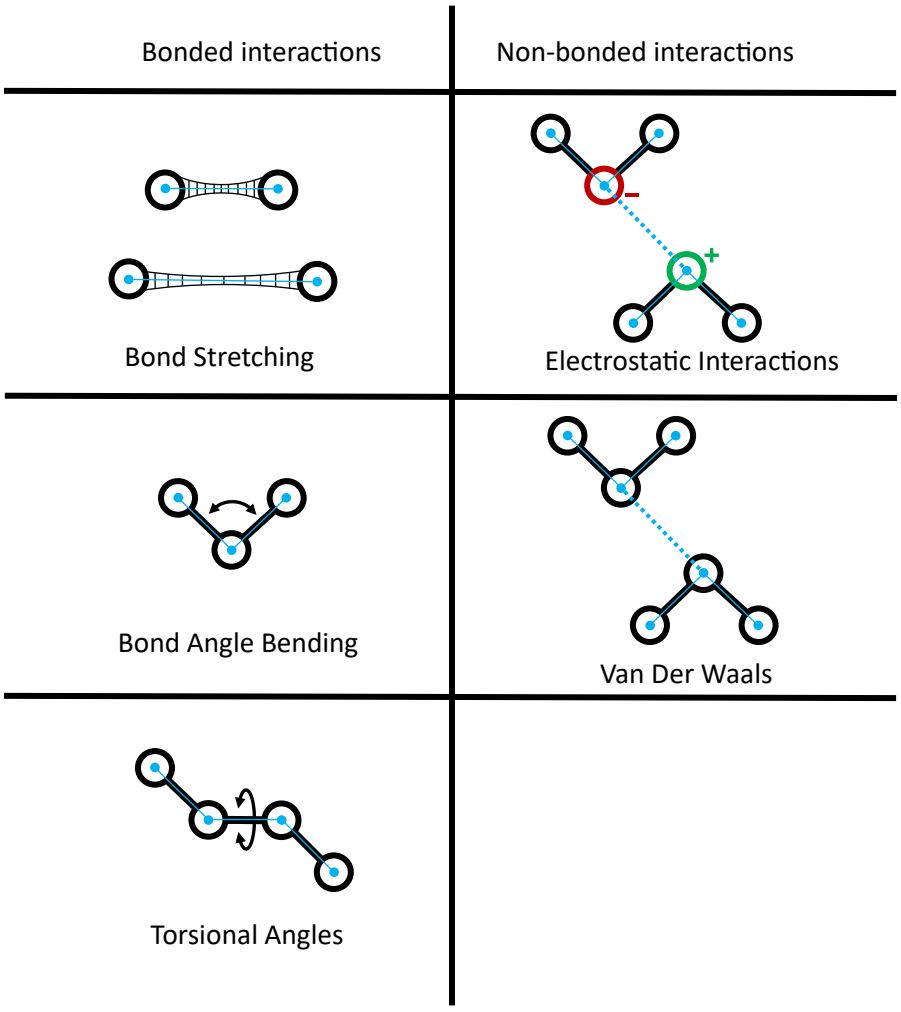

Figure 2.2: Schematic demonstrating the various terms involved in the Amber forcefield. Circles represent the atoms involved and black lines denote intramolecular bonds whilst the dotted blue lines show intermolecular forces. The left column covers the bonded interactions and the right column illustrates the non-bonded interactions.

2.2.3 Bonded interactions

For the work carried out here the Amber ff14sb, GAFF and TIP3P water system were used. The simplest functional form of the Amber forcefield 'bonded terms' is illustrated in equation 2.2.3.1 and represents bond stretching, bond angle bending and torsions. r and θ represent the bond length and angles, respectively, with $_{eq}$ denoting equilibrium values. ϕ represents the dihedral angle and $\mathrm{V_n}$ is the dihedral barrier height, n is the periodicity and γ the phase angle. The values $\mathrm{r_{eq}}$, θ_{eq}, $\mathrm{K_r}$ and $\mathrm{K_\theta}$ were all adjusted using ab initio and experimental data[95].

$$E_{bonded} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] \quad (2.2.3.1)$$

2.2.4 Non-bonded Interactions: van der Waals Interactions

The 'non-bonded' interactions include electrostatics and van der Waals interactions; the latter are calculated only under specific circumstances, between atoms of different molecules or in the same molecule when separated by three or more bonds. Van der Waals are relatively weak interactions and include the effects of transient shifts in the electron distribution of atoms. Temporary dipoles form and attract, contributing to the dynamics of a system and potential biological mechanisms at play. In MD, this phenomenon is modelled using the Lennard-Jones potential, shown in eq. 2.2.4.1, where $\mathrm{r_{ij}}$is the distance between two atoms, the term $\varepsilon_{\mathrm{ij}}$ is the depth of the potential well for the pair interaction and σ_{ij} is the distance where the potential becomes exactly zero; in other words, the point at which repulsive forces for the pair begins. The van der Waals interactions calculated with the Lennard-Jones potential have very small impacts on the atomic motion at long ranges but are also very computationally demanding.

$$V_{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.2.4.1)$$

For this reason, a cutoff is typically used, typically ranging between 8 Å and 12 Å in biological simulations. These cutoff values are often chosen to match those used in forcefield parameterisation and so in this thesis, with amber forcefields used, the cutoff is 10 Å. The cutoff balances an improvement in computational efficiency at the expense of simulation quality. A switching function[96] is often used that brings the Lennard-Jones potential down to 0 smoothly at the cutoff distance to reduce how abrupt the truncation is and give better calculations of non-bonded interactions, as illustrated in Figure 2.3[96], this was employed in our calculations with the NAMD software used in this thesis.

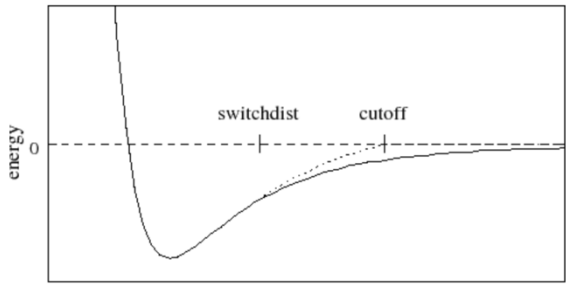

Figure 2.3: Graph of van der Waals potential with the switching function (dotted line) and without (solid line)[96].

2.2.5 Non-bonded interactions: Electrostatic Interactions

Electrostatic interactions are forces between charged atoms, which are fundamental for processes like ligand binding. The interactions are calculated using Coulomb's law for pairs of atoms. The Coulombic interactions are described by:

$$V^{q_i q_j}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.2.5.1)$$

Where $q_i q_j$ represents the charges on atoms i and j, with ϵ_0 describing the permittivity of free space[97]. The electrostatic interactions cannot be calculated easily and efficiently due to the equation itself poorly converging, especially for the long-range interactions. What we want to find out is the Coulomb contribution as this is crucial for calculating the forces of electrostatic interactions. This is notoriously difficult to calculate as the long-range interaction decays as $\frac{1}{r}$, making convergence difficult [98,99]. Ewald summation describes these long-range interactions for a system within periodic boundary conditions where the repeating space makes long distance interactions costly to calculate. The approach splits the coulombic interactions into a short-range element calculated in real space and a long-range element calculated in Fourier space. In the work presented in this thesis a specific method of computing the Ewald sum is

used called the particle-mesh Ewald (PME). PME is a highly efficient numerical method for the full electrostatic computations[94,96,97,100].

2.2.6 Ligand Parameterisation

Protein amino acids have been thoroughly studied and have established forcefields that can be used without modification, such as the Amber ff14SB used here. However, there are several cases where a molecule is not included, a particularly relevant example being with ligands such as the neurotransmitters glycine and GABA in zwitterionic form used in this work. These are classed as non-standard residues and need to be parameterised before they can be used, meaning their physical and chemical properties must be defined before use in a simulation. We used a protocol previously adopted for similar systems[46,86].

Firstly, the ligand structure is taken from a PDB or is generated using a molecular drawing tool like Avogadro[101]. From this we can calculate the partial charges of each atom using ab initio quantum mechanical calculations with a protocol consistent with the selected Amber forcefield. The first step involves geometry optimisation which is carried out to find a local energy minimum from the starting conformation using the Gaussian 09[102] software at the density functional theory level, with the B3LYP exchange and correlation potential and the 6-31G* basis set. This optimised structure can then be used for a single point energy calculation carried out at the Hartree Fock level with the 6-31G* basis set, where electrostatic potential (ESP) charges are determined. These partial charges reproduce the electrostatic potential, but have several flaws due to the statistical approach of the fitting. The ESP points used in the fit must be outside the van der Waals surface of a molecule and so problems arise with 'buried' charges that are not well characterised. Generally, intramolecular properties may be poorly characterised by ESP fitted charges, an important flaw if a process as sensitive as ligand-protein binding is being investigated. The ESP charges can be improved by carrying out what is called a RESP fit to get the restrained electrostatic potential charges. The RESP fit applies restraints to the non-hydrogen atomic charges towards a target charge resulting in an overall improvement in the quality of the charges derived[103]. This approach to charge parameterisation has worked effectively for previous work done by the group and is regarded as an effective method for determining charges for non-standard residues[46,86]. The parameters for the bonded and the van der Waals interactions were determined with GAFF[89].

2.2.7 Integrating the Equations of Motion

Once all the forces have been calculated, the simulation is advanced to the next time step: this is where Newton's equations of motion come into effect and are integrated. The time step itself is an important factor: too large a value and information are lost, too small, and the computational demand is too high. For biological systems, a two femtosecond timestep is commonly used[104] justified as capturing the timescale of important movements slower than the vibrational frequencies, restrained by tools like SHAKE which are typically used to constrain the fast vibrations of the bonds containing hydrogen atoms[105]. Several algorithms have been designed for integrating the equations of motion, such as the Verlet algorithm, Euler algorithm or the Beeman algorithm. For the simulations described here, the velocity Verlet algorithm is used[88].

The velocity Verlet algorithm is commonly used and is implemented in the NAMD software package. There are two main advantages to the velocity Verlet algorithm, its efficiency and simplicity which make it extremely useful when performing simulations for large systems[97]. In these equations:

$$1.\quad v_{n+\frac{1}{2}} = v_n + M^{-1}F_n \cdot \frac{\Delta t}{2},$$

$$2.\ r_{n+1} = r_n + v_{n+1/2}\Delta t,$$

$$3.\ F_{n+1} = F(r_{n+1}),$$

$$4.\ v_{n+1} = v_{n+\frac{1}{2}} + M^{-1} \cdot F_{n+1}\frac{\Delta t}{2}$$

(2.2.7.1)

where r refers to the positions of the atoms, t for the time, n being the step, v the velocity, F is the force and M is mass[97]. This method allows the position and velocity of the next step in the simulation r_{n+1}, v_{n+1}, to be calculated from the current timestep r_n, v_n, if the force has already been calculated.

2.2.8 Temperature and Pressure within MD Simulation

Macroscopic properties are important to consider, especially when simulating a biological system that is only going to function physiologically at a very narrow range of conditions. The integration of Newton's equations is therefore often coupled with algorithms capable of conserving selected macroscopic quantities such as temperature and pressure.

The natural statistical ensemble for MD simulations is the microcanonical (NVE) ensemble where the number of particles (N), the volume (V) and the energy (E) are conserved, due to the properties of the Newton's equations of motion. However, for biological systems, it is more realistic and common to perform simulations within the canonical ensemble (NVT) or the isothermal-isobaric (NPT) ensemble, where, beside the number of particles N, the temperature T and the volume V, or the temperature T and pressure P are conserved, respectively. Implementation of an ensemble like the NPT requires the use of a thermostat to control the temperature and a barostat to control the pressure.[88,97] The temperature in molecular dynamics simulations can be described with:

$$T(t) = \sum_{i}^{N} \frac{m_i v_i^2(t)}{k_B N_f} \qquad (2.2.8.1)$$

Where T is temperature, t represents time, m_i is the mass of the particle i, N is the number of particles with N_f being the degrees of freedom in the system. v_i^2 is the velocity of a particle and k_B is the Boltzmann constant. The total kinetic energy of a system fluctuates along with the instantaneous temperature. Multiple recipes for thermostats are available to control temperature, each with its own benefits and limitations. The thermostat we chose to use was the Langevin thermostat[106]. The Langevin thermostat mimics the coupling of the system being simulated to a virtual heat bath. This involves introducing a 'random force' term into the equations of motion, a stochastic element, as well as adding a deterministic element named 'frictional force' that is proportional to a given particle velocity. These extra terms can be thought of as a way of coupling the simulated particles with virtual particles of a heat bath through which collisions may occur. The Langevin equation utilises the equation of motion of Brownian particles according to Newton's laws with the assumption that a Brownian particle experiences these two forces:

$$m\frac{d^2r(t)}{dt^2} = \zeta\frac{dr(t)}{dt} + F(t) \qquad (2.2.8.2)$$

$F(t)$ represents that fluctuating 'random force' derived from a gaussian noise process with zero mean and variance. The 'friction term,' represented by ζ, is assumed to function through Stokes' law[106].

As with thermostats, there are multiple options for barostats. The Langevin piston is the type of barostat used in our simulations; it is a straightforward mechanism used to control pressure and eliminate the nonphysical vibrations other approaches have when attempting to use an NPT ensemble. The Langevin piston builds on the strengths of previous approaches, including the extended system method (Andersen barostat) and the weak coupling method (Berendsen barostat) whilst minimising the negative influences on dynamics. These two methods treat volume as a dynamic variable, controlling that to equalise the pressure. The argument for the Langevin piston is that these other two options either overdamp or underdamp the 'piston' that equalises the pressure. The Langevin piston approach allows partial damping by having the piston degree of freedom described through the Langevin equations[107].

$$\dot{r}_i = \frac{p_i}{m_i} + \frac{1}{3}\frac{\dot{V}}{V}r_i,$$

$$\dot{p}_i = F_i - \frac{1}{3}\frac{\dot{V}}{V}p_i, \qquad (2.2.8.3)$$

$$\ddot{V} = \frac{1}{W}[P(t) - P_{ext}] - \gamma\dot{V} + R(t)$$

here γ refers to the collision frequency, P is pressure, p is momentum, V is volume and W is the virtual mass of the piston. *R(t)* describes a random force pulled from a gaussian distribution with zero mean and variance:

$$\langle R(0)R(t)\rangle = \frac{2\gamma k_B T\Delta(t)}{W} \qquad (2.2.8.4)$$

It is worth noting that the cell that is used for molecular dynamics is implemented with periodic boundary conditions (PBCs). This is where a cubic box is replicated in space, infinitely forming a lattice of the cell. As a molecule leaves the central box, it will enter through the opposite face[97,107].

2.2.9 Minimisation and Equilibration

Minimisation is the first step in running simulations, following initialisation, and is carried out as a refinement of the original structure. It is key to getting a usable structure from raw data. The raw structure taken from cryo-EM and X-ray crystallography is an interpretation of electron density data obtained in an experiment and therefore is not a perfect representation of a bioactive molecule. The purpose of minimisation is to eliminate any abnormal conformational arrangements that may have come about because of experimental design, such as the crystallisation process or the snap freezing of a protein sample before imaging[108], as well as for the preparation of the system for simulations (e.g. addition of hydrogen atoms, missing residues, solvation etc.). Energy minimisation uses algorithms like the steepest descent algorithm and the conjugated gradient algorithm to bring the structure to a local energy minimum. In the NAMD software, the conjugate gradient algorithm is used.

The potential energy is the focus of the energy minimisation as shown in Figure 2.4; this containing all the different energetic components like bond bending, stretching, torsion etc.

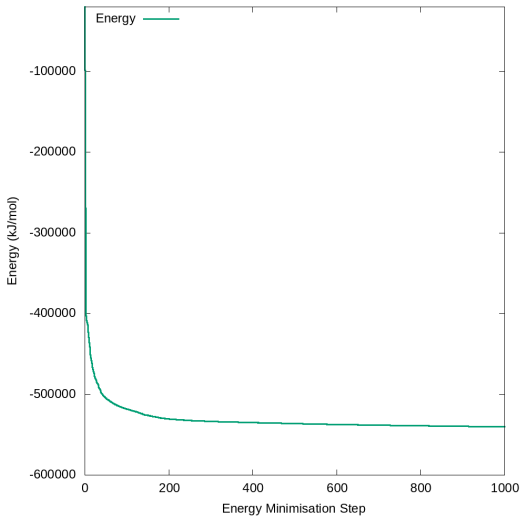

Figure 2.4: Decreasing potential energy over time during the energy minimisation process using NAMD 2.14. This is taken from the minimisation of the 5CFB structure described in Chapter 3.

An improved molecular arrangement is found in local energy minima where variations in structure result in small changes of energy. In the case of the pLGICs used here the minimisation process involve an initial minimisation of water and ions along with a fixed protein, followed by relaxing everything but the protein backbone and ligands, a minimisation with everything free but the ligands, and finally a minimisation of all atoms in the system without constraints.

Following minimisation, the structure will be in an energetically stable state and less likely to behave unrealistically in MD simulations. It is then ready for heating and general equilibration. The heating process involves incrementally increasing the temperature of the system in NPT until it reaches the biologically relevant temperature of interest. This is followed by a period of simulation in NPT with the relevant barostats and thermostats set up until the structure is stable at the desired temperature and pressure. Depending on the system, this could be all the way up to hundreds of nanoseconds and even have stages involving restraints being implemented and slowly released for enhanced stability.

2.3 Metadynamics

A major problem encountered when attempting to study a biological process with simulations is the timescale that these processes occur on. What is considered in biology to be a very rapid process in the range of hundreds of microseconds to milliseconds[71–73,109] is, in fact, extremely difficult to achieve in simulations. Current standard hardware, for example, can manage a microsecond of simulation in several months or a large membrane-bound protein, and so for a millisecond of simulation time, you would need to invest an absurd amount of time and resources to achieve ms timescales. For this reason, enhanced sampling methods are used, such as metadynamics, to drastically reduce the resource demands of sampling a rare event. The end goal of metadynamics is to push the system to visit all relevant conformational states, as well as to profile the free energy surface. It may also be useful to derive the rate at which events occur by using more advanced analysis techniques.

Metadynamics is a method first introduced in the early 2000s and has become a staple of studying rare events and hard to sample molecular systems in simulations[110]. The algorithm requires a set of collective variables (CVs) to be identified, which are used to define the specific process of interest to be sampled. Over the course of the simulation, a bias potential is then deposited along the points of the trajectory of these CVs already visited, growing over the time of the simulation. The bias potential is made up of gaussians, which are sets of repulsive terms

that deter a conformational state from being revisited once a gaussian has been placed there[111].

This 'bias potential' is history-dependent and a function of the CVs chosen over time. It is expressed as a sum of all the Gaussians that have been added along the CVs space during the simulation. With the CVs value at any given point written as s(q), q relating to the atomic coordinates.. The gaussian bias potential $V(s(q),t)$can be written as:

$$V(s(q),t) = \sum_{\kappa\tau} W(\kappa\tau) exp\left(-\sum_{i=1}^{d} \frac{\left(s_i - s_i\left(q(\kappa\tau)\right)\right)^2}{2\sigma_i^2}\right) \quad (2.3.1)$$

Here τ is the gaussian deposition stride, the width of the gaussian is expressed by σ and $W(\kappa\tau)$ represents the height of the gaussian at the time $\kappa\tau$, d is the total number of CVs that have been employed and s_i is the position of the CV at the i-th step, s(q) is the current CV value s, and the atomic coordinates being q.From the bias deposited, when a diffusive behaviour is reached, the free energy F of the system can be recovered by reversing the bias potential:

$$V(s, t \rightarrow \infty) = -F(s) + C \quad (2.3.2)$$

where C is a constant that will increase over time during the simulation. To best understand how this is impacting the system, we can consider a 2-dimensional plot of free energy versus as single collective variable[84,112], as shown in Figure 2.5 over the page. The process allows the system to move between the 'wells' of conformational space as they eventually fill up with gaussians. When complete, this provides a free energy surface from the total bias potential, which equates to negative free energy plus a constant.

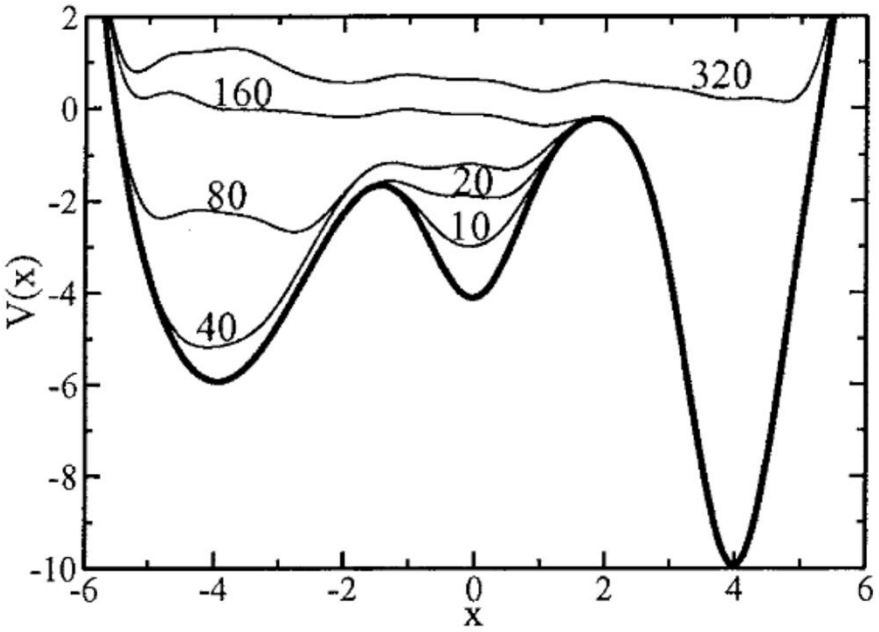

Figure 2.5: Diagram illustrating how across one CV (labelled x) the progression of time accumulates gaussian terms to explore multiple energy minima in the potential energy V(x). Each line is numbered 10,20,40… etc show the effect of deposition of 10, 20, 40 gaussians through a metadynamics simulation until all minima have been profiled.This diagram was taken from Laio and Parrinello (2002) [110]

2.3.1 Well-tempered Metadynamics

Whilst metadynamics has been a widely accepted and popular method, there have also been some adaptations and improvements which perform better or allow extra data collection. One of the issues is how difficult it can be to determine when to cease a metadynamics run: when too short you risk not exploring all the relevant minima, and when too long you risk forcing the system into states that are not physically relevant.

Well-Tempered metadynamics aims to remedy this issue by controlling convergence and setting extra parameters so that computational resources are dedicated only to the parameter space relevant to the system[113]. The changes made focus on addressing these problems by adding a rescaling technique where the height of the Gaussian decreases with simulation time as needed:

$$W(\kappa\tau) = W_0 \exp\left(-\frac{V(s(q(\kappa\tau)),\kappa\tau)}{\mathrm{k_B}\Delta T}\right) \quad (2.3.1.1)$$

Here W_0 is the initial height, and ΔT represents the dimension of temperature and can be tuned to speed up convergence. This rescaling of height ensures bias potential convergence is much smoother. The free energy with well-tempered metadynamics is calculated with this equation:

$$V(s, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(s) + C \quad (2.3.1.2)$$

where when $\Delta T = 0$ this equation functions the same as standard molecular dynamics. When used in PLUMED, the input parameter for using WT-MetaD is the biasfactor, as expressed here:

$$\gamma = \frac{T + \Delta T}{\Delta T} \quad (2.3.1.3)$$

For our simulations we used a biasfactor of 15, a typical value for biological systems[46].

2.3.2 Infrequent Metadynamics

Another adaptation is infrequent metadynamics, which is useful when attempting to capture the time it takes for conformational transitions. As with any metadynamics simulation, the CVs selected are of great importance for the correct application of bias to sample a rare event. However, when collecting rates and time-related data, it can be argued that the selection of the CVs is even more critical as the results will be very sensitive to poorly selected CVs (as well as to other factors, including the forcefield). The main technical difference in this approach to other metadynamics methods is the deposition of gaussians being less frequent. The reason for this is so that bias is deposited frequently enough to accurately converge on the

transition event but infrequent enough to avoid depositing bias directly on the transition state. Any excess bias applied to the transition state, in fact, would cause problems in the time reweighting process to recover the unbiassed time and thus reduce the accuracy of the measurements by artificially inflating the time recorded[114].

To get around potentially complex or multiple transition states, the procedure is kept simple, and no assumptions are made other than a transition region having a low residence time. The carefully selected CVs profile the end goal of the metadynamics, i.e. bound/unbound, and can be used to extract rates by using the reweighting algorithm. The frequency of gaussian deposition can also be adapted so to keep it high far from the transition state and low while approaching the transition state, in a procedure named frequency-adapted metadynamics[115].

2.3.3 Reweighting Free Energy and Time

As bias is applied to a system in metadynamics calculations, both the time and other unbiased CVs are distorted and therefore the maximum amount of information that can be discerned from the metadynamics run is limited. With ligand binding for example, understanding the kinetics and time-dependent properties of the binding becomes a big issue. However, reweighting methods have been developed to overcome these issues through the application of a few simple algorithms. Through these reweighting techniques, we can retrieve the time required for an event to occur in metadynamics and therefore calculate the rates for specific rare events such as binding/unbinding.

For reweighting the free energy landscape and regaining details about unbiased CVs, an algorithm is used based on the relation between biased probability distribution $P(\boldsymbol{q}, t)$ at a given time and the Boltzmann distribution $P_0(\boldsymbol{q})$:

$$P(\boldsymbol{q}, t) = e^{-\beta(V(s(\boldsymbol{q},t)+C(t))} \cdot P_0(\boldsymbol{q}) \tag{2.3.3.1}$$

Here the inverse of $\mathrm{k_B}T$ is expressed by β. At a given time, this equation relates the probability distribution in a metadynamics biased ensemble with the canonical Boltzmann distribution needed to get an accurate free energy landscape. Here the time-dependent bias offset is expressed as:

$$C(t) = -\frac{1}{\beta}\log\left(\frac{\int ds\, e^{-\beta F(s)}}{\int ds\, e^{-\beta\left(F(s)+V(s,t)\right)}}\right) \quad (2.3.3.2)$$

Whilst this addresses the reconstruction of the free energy surface, for insights into the timescales rare events occur on, the time needs to be reweighted. The addition of bias distorts real-time, making the timestep no longer an indicator of physical time. The physical time is invaluable for kinetics and gaining a more detailed picture of dynamics is of interest, and by reweighting time, we can get the rates of rare event barrier crossing like those involved in ligand binding[116].

The reweighting process is simple, and the relation of simulation time to physical time can be expressed as $\mathrm{t} = \alpha \cdot \tau_{\mathrm{METAD}}$, where τ is the simulation time and α is the acceleration factor calculated directly from the metadynamics bias $\mathrm{V}(\mathrm{s}, \tau_{\mathrm{METAD}})$:

$$a = \langle e^{\beta V(s,\tau_{METAD})}\rangle_V \quad (2.3.3.3)$$

Here the CVs are represented by s. This can be used to write a full equation for calculating the rescaled time at any given step of the simulation:

$$t = \sum_{i}^{nMETAD} dt e^{\beta V(s(t_{iMETAD}),t_{iMETAD})} \quad (2.3.3.4)$$

Here nMETAD refers to the number of steps in the simulation and the i in t_{iMETAD} being the metadynamics step at which the time is calculated.

Following this reweighting process, there are a few more steps to ensure that the assumptions of this calculation have been met. For rare events, multiple replicas are run to get an average time taken for an event to occur, reweighted through this method and then a Kolmogorov-Smirnov (KS) statistical test is applied. An empirical cumulative distribution function (ECDF), from the simulation times for an event to occur is created, and a theoretical cumulative distribution (TCDF) generated from many randomly generated values according to a cumulative distribution function of a homogeneous Poisson process with estimated real time τ:

$$TCDF = 1 - \exp\left(-\frac{t}{\tau}\right) \quad (2.3.3.5)$$

The KS test is then carried out with the null hypothesis being that ECDF and TCDF show no statistically significant differences and therefore share the same distribution. The α value is the predetermined threshold for a significant KS test output result, this is typically 0.05. From the KS test we get a p-value that represents the KS test statistic quantifying the difference between the ECDF and TCDF. If this p-value is larger than the threshold then the null is accepted suggesting no significant differences between the collected data (ECDF) and the TCDF. Rare events should demonstrate a Poisson distribution, and so if bias has been applied correctly along with successful time reweighting, a KS test will return with a high p-value. The null hypothesis being accepted reflects appropriate bias deposition in the metadynamics and that the time reweighting has been carried out correctly[75,114,117].

2.3.4 Funnel Metadynamics

Determining the thermodynamic properties of ligand binding to pLGICs and the mechanisms involved are key to having a comprehensive understanding of neurobiology and for the development of novel pharmaceuticals. There are a range of both experimental and computational methods employed to gain insights into ligand binding, deriving the binding affinities of ligands and key metrics like the EC50 which are crucial for guiding effective drug design. To effectively measure properties of ligand-protein interactions such as binding affinity experimentally, there are three typical approaches used including isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence polarization (FP). ITC utilises differential scanning calorimetry and provides measurements of the enthalpy and heat capacity of a protein-ligand complex, providing insights into the forces that drive binding and stabilize the interactions. SPR is an optical based technique, which provides highly reproducible affinity measurements in real-time. FP is another method that allows measurements of kinetics using fluorescent-labelled ligand molecules that become unpolarised and emit light faster in an unbound state; the shift in emission time caused by fluorescent ligands being displaced by unlabelled ligands is the key part of this method for measuring kinetics. Whilst these methods are invaluable, they all come with inherent limitations of their experimental design. ITC is a highly sensitive and complex method that requires many samples, is low throughput and time consuming. SPR requires protein immobilisation which then has a confounding impact on conformational/rotational entropies

and association rates. FP requires ligands to be fluorescently tagged which can have a range of impacts on binding behaviour that would amount to errors in measurements. All these issues with experimental approaches are compounded with a high cost, laborious nature and time required to complete.[118] For these reasons, an alternative approach of utilising computational simulations can be employed.

A number of computational methods are available to probe the binding affinity and properties of ligand-protein systems. Free energy perturbation (FEP) for example, is considered a rigorous method. FEP applies statistical mechanisms to calculate the free energy differences from an MD simulation between two states. Typically, the potential energy for a state is measured with molecular mechanics or quantum mechanics over incremental time steps that aids in achieving better convergence of free energy calculation, convergence being a crucial element for FEP reliability and reproducibility. The difference in Gibbs free energies (ΔG) between each incremental state is limited to make for more manageable computation of free energy changes, also increasing the accuracy of the calculations. Other methods like MM-PB/SA or MM-GB/SA methods (molecular mechanics Poisson-Boltzmann or Generalized Born/Surface Area) estimate binding free energy of a ligand to a protein using a theoretical thermodynamic cycle which reduces the ligand-protein binding from an in-solution system to a gas-phase free energy and solvation free energy difference. Whilst this second approach is much faster than FEP it has had push back by researchers for lacking a solid theoretical foundation[119].

One approach that is used here is Funnel Metadynamics (FM)[74,120], a method initially introduced in 2013 which simulates binding of a ligand from the fully unbound state to the final binding site using metadynamics. One of the main limitations for applying metadynamics to ligand binding to a protein is the exploration of the bulk solvent in the unbound state. The way protein simulations are set up requires a large amount of the cell box to be filled by water and so when a ligand is outside of the binding pocket there is a vast amount of conformational space to be explored that is far away from and unrelated to the binding pocket. Whilst it is important to have some unbound conformational space considered when looking at the binding, a lot of the space is irrelevant and computational resource to explore it would be wasted. For this reason, the implementation of a funnel restraint is added in FM which limits this exploration to only the region around the binding pocket and along the likely/representative binding path.

The funnel restraint used here is made of two parts, a cone and a cylinder restraint which are placed to include the binding site at the base of the cone and a cylinder extending into the bulk

solvent to cover the ‘unbound’ region. For most of the simulation time within this space, the ligand is not influenced by the restraints but when it reaches the edge of the area defined by the funnel a repulsive bias is applied to dissuade and prevent exploration outside the set regions. The funnel is implemented in a way to not interfere with any binding residues and the binding pocket and so accelerates the metadynamics simulation of binding within without interfering with the process.

The funnel is set up using a series of defining parameters as illustrated in Figure 2.6, the first being the two points on which the funnel axis will sit: point A within the position in the binding pocket the funnel restraint will be directed towards, and point B in the solvent where the cylindrical section of the restraint will project into. The Zcc value is the switching point of the restraint where the funnel cone narrows to a point and becomes the cylinder. The cone amplitude section parameter is represented as α and Rcyl indicates the radius of the cylinder. All these parameters must be chosen in such a way that the shape of the cone restraint is not imposing on the binding pocket residues in any way. These parameters define the basic funnel but lower and upper walls are also applied at the ends of the funnel, the lower wall capping the base of the cone and the upper wall capping the cylinder so that the ligand cannot exit the funnel restraint area at the ends. These walls act in the same way as the funnel restraint cone and cylinder walls and so also need to be placed far enough away from any regions of interest to prevent biasing the binding mechanisms. The result and shape of the funnel is demonstrated in Figure 2.6.

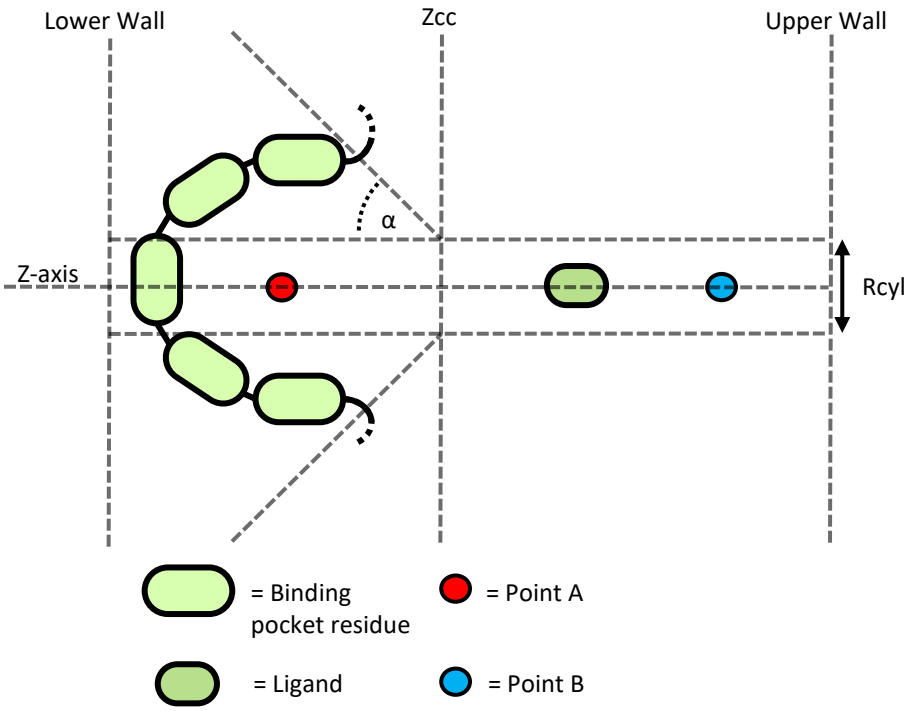

Figure 2.6: Diagram of the parameters and shape of the funnel restraint used in funnel metadynamics. The binding pocket is represented in simplified form as a chain of amino acid residues which would likely have a more complex topographical arrangement. Point A represents the centre of the binding pocket, the position through which the funnel Z-axis bisects.

The acceleration of the metadynamics by limiting ligand exploration of this area leads to multiple observations of binding and unbinding events, ultimately allowing for a well characterized binding free-energy and clear identification of the ligand binding mode at the lowest free-energy minimum. The absolute protein-ligand binding free energy is typically calculated with the following formula for ΔG_b^0:

$$\Delta G_b^0 = -\frac{1}{\beta} \ln(C^0 K_b) \qquad (2.3.4.1)$$

In this formula[74], $C^0 = \frac{1}{1661} \text{Å}^{-3}$ represents the standard concentration of 1M of all molecules reacting. $\beta = (K_b T)^{-1}$ with the Boltzmann constant represented by k_B and T being temperature. K_b is the equilibrium binding constant that is derived from finding the difference in free-energy between the bound and unbound states. The effect of the funnel restraint does need to be considered when deriving any results from this method, when calculating the binding constant in a system with the funnel restraint ΔG_b^0 can be derived from:

$$\Delta G_b^0 = \Delta G - \frac{1}{\beta} \ln(\pi R_{cyl}^2 C^0) \quad (2.3.4.2)$$

Here ΔG is the free-energy difference between the bound and unbound states calculated with metadynamics and πR_{cyl}^2 represents the cross-section area of the cylinder used as a restraint potential. Using this approach[74] the equilibrium binding constant used K_b can be defined as:

$$K_b = \pi R_{cyl}^2 C^0 \int_{Z_{min}}^{Z_{max}} dz \; e^{-\beta(w(z)-w_{ref})} \quad (2.3.4.3)$$

Where Z_{min} and Z_{max} represent the lower and upper limits of the funnel axis where binding can occur. The $w(z)$ is the one-dimensional potential mean force as a projection of the free energy along the funnel axis z. w_{ref} is any point along the axis defined as unbound. This equation considers the effects that the cylindrical section restraint imposes on the unbound conformations as a correction. It is possible for the restraining potential to be set up with a wide array of parameters and even different shapes to the one in figure 2.6, an example of this being a bell-shaped funnel if more clearance around the pocket within the restrained area is necessary[121].

2.4 Simulation Analysis

There are several analytical tools used to profile the structure and dynamics of an atomistic simulation. Here we briefly outline what was used to assess the stability of the GlyR in the simulations as well as how the binding pocket interactions were profiled.

RMSD

One of the basic analytical tools used is the root mean square displacement (RMSD). RMSD measures the deviation of a selection of coordinates from a set of reference coordinates would indicate a perfect match, here RMSD is defined by the equation:

$$RMSD = \sqrt{\sum_i^N \frac{(X_i - Y_i)^2}{N}} \quad (2.4.1)$$

Where N is the number of atoms, $\mathrm{X_i}$ the coordinate vector of the selected atom, $\mathrm{Y_i}$ is the reference coordinate vector, and i simply denotes which atom in the selection. This analysis

is carried out in the context of the evolution of time, where the first frame of a simulation is typically used as the reference set of atoms[122].

Distance

Another series of analyses carried out utilised the python package MDAnalysis[123] for measuring structural changes. For example, the centre of mass of two sets of atoms can be measured, and then the distance between them recorded over time, as illustrated in Figure 2.8.A. This can provide a metric for how much a ligand is moving from the centre of a binding pocket or how open/closed the C-loop structure is[123].

H-bonding

Hydrogen bonds (H-bonds) are formed when an electronegative atom (acceptor) is in the proximity of hydrogen bonded to another electronegative atom (donor) (Figure 2.7.B). This frequently happens in water, where the donor and acceptor are electronegative oxygen atoms in molecules. In proteins, hydrogen bonds are integral for ligand binding pocket interactions[124] and are important for glycine binding to GlyR[26].

To measure hydrogen bonding, the software cpptraj[122] or MDAnalysis[123] can be employed. The distance between a selection of atoms known to be H-bond donors/acceptors is evaluated, as well as the angle between donor-H-acceptor. A H-bond is then recorded if the distance is below a certain threshold and the angle above a certain threshold. Parameters must be chosen to define these distances and angles, in the case of cpptraj by default and this work, the threshold values are 3 Å[122] and 135°, respectively.

Cation-π Interactions

Biological systems commonly include a number of amino acids that contain aromatic rings, including Phenylalanine, Tyrosine and Tryptophan. These aromatic rings can interact with other residues and ligands via what are called cation-π interactions; these have been found to be important to a variety of processes, including secondary structure formation, protein-DNA complexes, and specifically of importance to this work ligand-protein complexes[39,40].

The theoretical basis for these weak intramolecular interactions is that the electron-rich π-cloud in the aromatic group of these amino acids attracts the electron-deficient cationic group of another residue[125,126]. In the case of glycine, the positively charged NH_3^+ group can act as

this cation and the multiple aromatic residues of the binding pocket (Phenylalanine). Using the cpptraj code, we considered the presence of a cation-π interaction based on the distance between the aromatic ring of the protein residues and the cationic group of the glycine, as well as the angle between the normal to the ring and the distance of the cation from the centre of the ring, as shown in Figure 2.7.C. The distance must be less than 6 Å, and the angle less than 45° or greater than 135° for the presence of the interaction to be recorded as present[127].

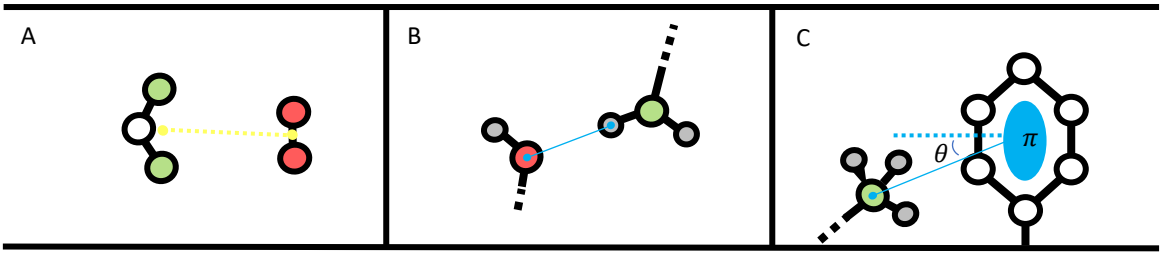

Figure 2.7: Graphic to illustrate how measurements are made from simulations. A demonstrates the measuring of distance between the centres of mass of two groups of atoms (signified by the yellow dot. B illustrates the formation of hydrogen bonds between an acceptor (in red) and a hydrogen atom in grey bonded to a donor (in green). C illustrates a cation-Pi interactions where the electronegative π cloud on the ring (in blue) interacts with a positively charged group.

Chapter 3

Modelling Ligand Binding in Glycine Receptors

3.1 Introduction

To gain insights into the binding mechanism of glycine receptors, a range of approaches can be employed including the utilisation of classical molecular dynamics and metadynamics. Classical molecular dynamics allows the interactions between a ligand and receptor to be observed in a dynamic environment at an atomic scale to get a clear picture of the binding mode or modes[128]. The making and breaking of hydrogen bonds[129,130] and cation-π interactions[131] can also be shown in detail over the course of an MD simulation; the prevalence of these interactions indicates their strength and the residues of the receptor which play crucial roles in the binding process are highlighted. The presence of water in the binding pocket and how it interfaces with a ligand is also an aspect of ligand binding that is difficult to study experimentally but can be investigated with classical molecular dynamics[67]. All this information is key to understanding how a ligand binds and what might follow to instigate the changes required for a receptor to activate.

The first step in modelling receptor-ligand interactions is to build a suitable system; this involves taking data collected from imaging experiments like cryo-EM or X-ray crystallography and preparing the structure for use through a series of stages explained in the next section. In the case of GlyRs, there is a variety of structures available currently that has rapidly grown in number over the past few years. The first system that was developed at the beginning of this work was made using X-ray crystallography data of a human α-3 homomeric glycine receptor at a 3.04 Å resolution (PDB: 5CFB)[15]. Whilst this structure ended up being suboptimal for our research goals, at the time it was the best candidate available in terms of structural data quality. Choosing a structure to start with depends on a number of factors, including the availability, the quality of the structure based on the PDB scoring system (Figure 1.12), the resolution and how close to the ideal set up the structure would be i.e. bound with a ligand or glycine. The 5CFB structure was published alongside work that sought to understand the inactivation mechanisms of GlyRs by highly specific antagonists like strychnine, and therefore was bound in all 5 pockets by that antagonist rather than our target of glycine. The strychnine

bound system required glycine to be switched into its places and led to some issues, the implications of this will be discussed with the results later in the chapter. The structure not being bound by glycine was a significant drawback but given the quality of the receptor itself and alternatives being less suitable it was chosen to move forward with. This first template was developed as a whole protein system including both the ECD and TMD, embedded in a system membrane, using the protocol described later in this chapter. The reasoning for creating the system within a lipid membrane was so that the real-world conditions could be emulated as best as possible and provide a platform to investigate lipid-protein interactions with the same system at a later date. As this system was set up, newer structures were published with a range of ligands bound, including glycine, providing an alternative that would not have the same issues of ligand retention seen in the first system.

The second system was prepared using cryo-EM data collected from a zebra fish homomeric α-1 GlyR at resolution 3.10 Å populated with glycine in all pockets (PDB: 6PM5[26]). The structure was published in work that was focused on understanding partial agonist action in GlyRs, including several other structures bound with partial agonists GABA and Taurine in a variety of conformational states. The 6PM5 system therefore required no alterations to be made to the ligands or the area around the pocket. In this system the TMD was removed and only the ECD restrained. The reasoning behind this choice considered the computational resource required for membrane and whole protein simulations, the distance of the binding pocket from the TMD-ECD interface and the assumed limited involvement of the membrane in the binding process and affinity, which was the focus of this work.

The main goal of the work described in this chapter was to develop a suitable system of GlyR binding that could be both a tool for investigating the binding pocket in a dynamic environment with classical molecular dynamics, probing key interactions and structural elements, and then providing a platform for metadynamics to be applied to observe and characterise binding and unbinding events in a wild type system, mutant system and with partial agonists. The next sections will go into detail on the protocols for preparing these systems, why certain choices were made and the results obtained for both systems being simulated with classical molecular dynamics.

3.2 Modelling the Whole glycine Receptor within a membrane

Both the full system and ECD-only system make use of the same setup; they were both run using NAMD on the ARCHER 2 UK supercomputer facility. Benchmarking was carried out on the ARCHER2 system to determine 4 nodes to be an optimal amount of computational resource for a system of this size using classical MD in NAMD, any more would yield little improvement in speed per core. Regarding the choice of software used, NAMD is a commonly used software package for running molecular dynamics simulations for biomolecules, which is well equipped for large scale parallel computing of large biological systems[132].

The whole protein structure was used to develop this system and the PDB scored high on the structural assessments generally agreed upon[133,134] for determining data quality. The initial structure was obtained with X-ray crystallography with a resolution of 3.04 Å by Huang et al[15,135]. The completed system is shown in Figure 3.1.A and 3.1.B along with the results of the PDB assessment criteria in Figure 3.1.C and demonstrate good scores relative to other similar structures for the Clash score, Ramachandran outliers and sidechain outliers. The structure contains the whole ECD and TMD but lacks a complete ICD which is instead replaced with a linker sequence of alanine, glycine, and threonine (AGT).

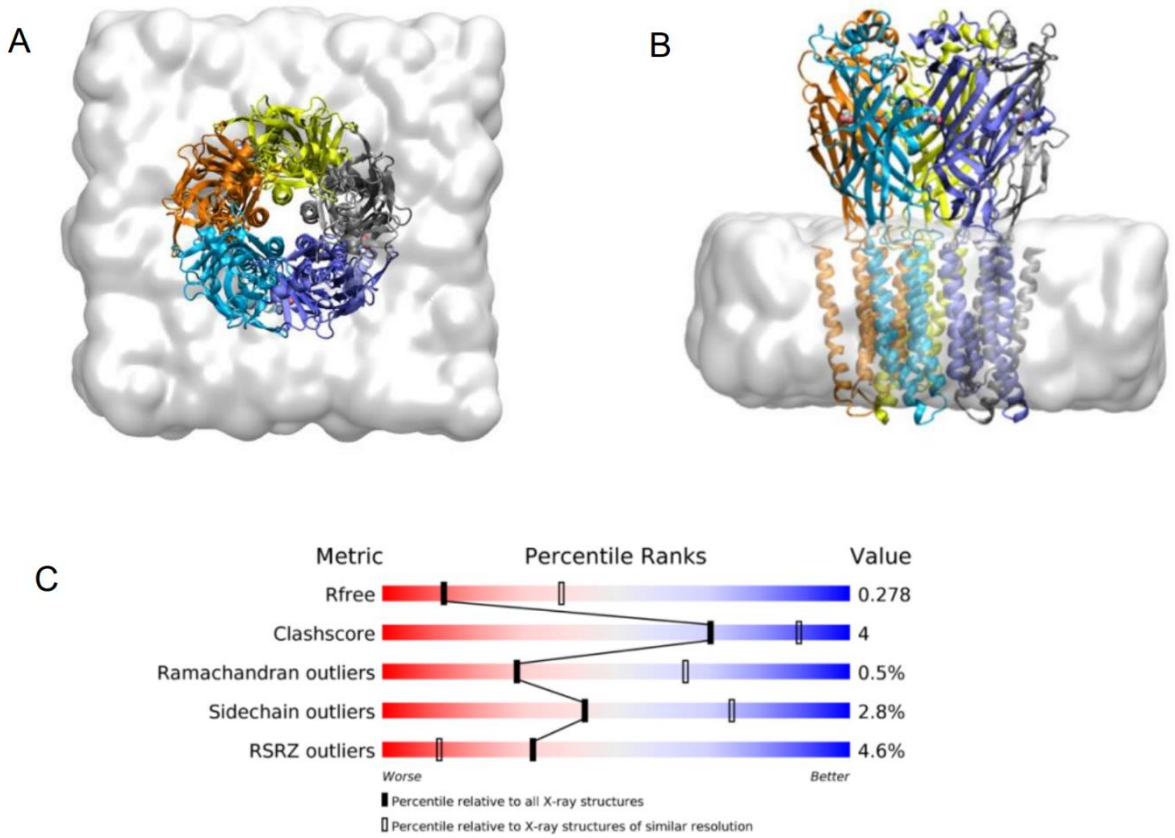

Figure 3.1. A, the cartoon representation of the 5CFB GlyR embedded in a membrane viewed from above the pore and from the side in B. C, is the scoring chart that is used as a measure of structure quality; this specifically is the 5CFB scores.

The lack of a full ICD domain would cause issues to the modulation of gating[136], but due to the timescales and computational resources needed for changes at the ICD to translate to the binding pocket, this is beyond the remit of our work and not a relevant concern.

3.2.1 Glycine Parameterisation

The glycine ligand in zwitterionic form is not a standard residue and therefore is not included within the Amber forcefields library, for this reason it needs to be parameterised. The parameterisation involves assigning the molecule atom types and defining the partial charges of the atoms. Due to the simplicity of the glycine structure, i.e., low number of possible conformations, the process of parameterisation was kept simple. Geometry optimisation was carried out on the ligand structure at the density functional theory level with B3LYP exchange and a 6-31G* basis set. To prevent deprotonation restraints were placed on the bond lengths for each hydrogen bond in the ammonium moiety. Following this, a single point energy calculation was carried out on the optimised structure at the Hartree Fock level using the same 6-31G* basis set for consistency with the Amber forcefield. A RESP fitting was then carried out to calculate the partial charges. The atom types and charges produced from this process were then assigned to the ligand and integrated into whole protein structure at the step of solvation and formatting of the PDB for Amber forcefields with the Leap tool.

3.2.2 Modifications

The 5CFB structure was bound to a strychnine ligand in each of the five binding pockets. The strychnine ligand is an antagonist and was used to stabilise the structure in a closed-bound state and therefore had to be modified prior to setup for our investigation into glycine binding. Using the VMD software, the strychnine was removed and replaced with a pre-parameterised zwitterionic glycine ligand in the centre of the binding pocket, the orientation of which was chosen based on prior works with the ligand bound in both experimental and computational works[16,36,137]. It is also worth noting the glycine ligand was given the residue name GZW to avoid clashing naming between GLY residues in the rest of the protein PDB.

Another change required before completing the setup and proceeding to prepare the structure for molecular dynamics was to address the binding pocket arrangement. Because strychnine is a much larger ligand than glycine, the binding pocket residues were not in a conformational arrangement that was suitable for glycine to stabilise in the binding pocket. This became apparent in early work where the ligands would rapidly eject in the equilibration period as soon as restraints were slowly released, lengthening this process yielded no improvement on ligand retention. Upon closer examination of the pockets, we found that the critical ARG-65 residue

was rotated away from glycine and was not forming the interactions necessary for the pocket to hold the ligand. To remedy this, the torsion angles were modified in the arginine residue to match current literature on glycine binding using ChimeraX with ISOLDE[138] to fit better with the established arrangements[16,26,137]. The required change consisted of altering the sidechain torsion angle of the arginine in the binding pocket initial structure so that the hydrogen groups were closer to glycine and matched other structures with glycine bound. This is highlighted in Figure 3.2 where the altered Arginine placement puts the terminus of the side chain in range of the glycine for hydrogen bonds to form consistently with the literature[26,139].

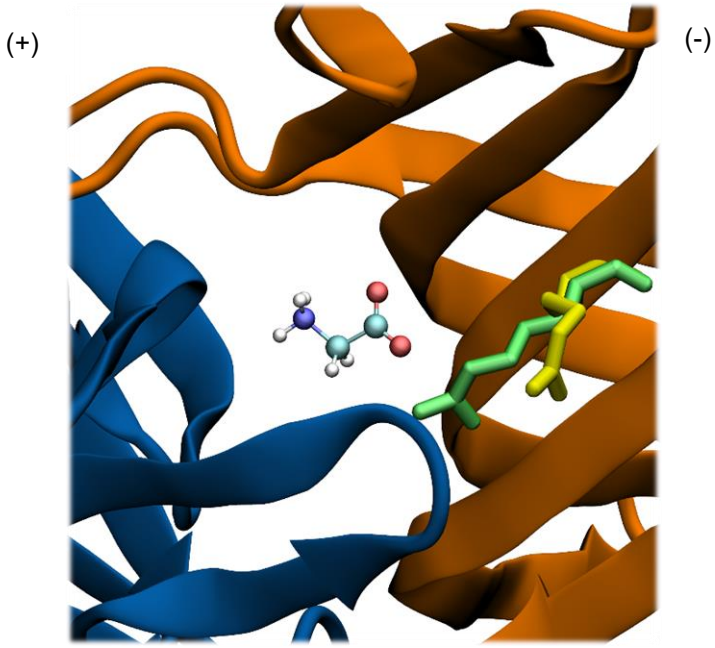

Figure 3.2: Here is a depiction of the binding pocket, the colour scheme is split across the two subunits that make up the binding pocket with the complementary (-) subunit shown in orange and the principal (+) subunit in blue. The Arginine (ARG-65) residues from the original 5CFB is highlighted in yellow and the modified form overlayed and highlighted in green.

The difference in 5CFB to the literature conformation of the binding pockets with glycine bound is likely due to the size of strychnine pushing residues further out than the smaller glycine would be. The system with GZW now in the binding pocket was then prepared for MD using the CHARMM-GUI[140] toolset, which allows proteins to be embedded in a lipid bilayer.

3.2.3 System Preparation

The modified PDB was protonated at a neutral pH and embedded in a lipid bilayer. As part of the CHARMM-GUI processing, the disulphide bond information was also added and checked between the relevant residues, in particular the disulphide bonds between cysteine 138 and 152 (The Cys-Loop disulphide bond) as well as between cysteine 198 and 208 (The GlyR specific disulphide bond). The cysteine disulphide bonds are suggested to be critical for

stabilisation of the structure, and therefore checking and implementing accurate disulphide bonds to be interpreted by NAMD is a key part of the setup[141,142].

Following checking and adding disulphide bond information, the orientation of the protein in the membrane was determined. For protein orientation, the OPM server[143] was used. The lipid membrane composition was made up of cholesterol (CHOL), 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) and 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) at a ratio of 35:30:30. The complexity of membrane that would exist in vivo based on reviewing literature suggesting these to be of critical importance to pLGIC function[66,144] as well as previous in-group work on pentameric ligand gated ion channels[145] showing these lipids to make stable membranes for MD simulations. Using leap[146], the complete protein and lipid structure were then solvated and formatted for use with NAMD and the AMBER forcefields ff14sb[90] and lipid14[147]. The protein was solvated using the TIP3P3[91,92] water system with a 15 Å buffer as well as 0.15 M of Na^+ and Cl^- ions to recreate a physiological-like environment in a periodically repeated orthorhombic cell. Following these steps, the system pre-minimisation consisted of 205,983 atoms, including the water, ions, lipid membrane, protein and GZW molecules.

3.2.4 Minimisation to Production

The minimisation step was the first to be undertaken; all simulations carried out, including this step, were done using NAMD 2.14[96] and the following parameters: the orthorhombic cell box size was 125 Å x 125 Å x 170 Å, the timestep for integration was set at 2 femtoseconds (fs), the SHAKE algorithm was used for constraining covalent bonds including including hydrogen atoms, Particle Mesh Ewald was employed for electrostatic interactions and for non-bonded interactions a 10 Å cutoff was imposed along with an added switching function[96]. The minimisation was carried out in 4 stages, with restraints on every atom excluding water molecules to begin with, followed by the removal of the restraints on the lipid and protein side chains, then removal of all protein restraints and finally, the removal of restraints on the GZW molecules. Heating was then carried out in incremental steps of 25 K until 310 K was reached with harmonic restraints applied to the backbone of the protein.

Equilibration, including the membrane, was carried out for 165 ns with harmonic restraints at 25 kcal/mol$Å^2$ that werescaled down over 65 ns to 0 kcal/mol$Å^2$; the specifics are detailed in table 3.1. The heating and equilibration were carried out whilst monitoring RMSD to ensure no large aberrant structural changes occurred, or ejection of ligands. Both the equilibration and production were carried out in the NPT ensemble at 310 K and 1 atm. The temperature

was controlled with a Langevin thermostat with the coupling coefficient set at 1 ps^{-1}, the pressure was regulated with a Langevin piston barostat setup with an oscillation restraints period of 200 fs and a damping time scale of 100 fs. The production was run for a total of 1000 ns using this system, with analysis carried out across the whole trajectory.

Restraints [Kcal/molÅ^2]	Timepoint [ns]
25	0
25	100
10	120
2.5	135
0.0025	155
0	165

Table 3.1: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was called to the subsequent value.

3.3 Modelling the Glycine Receptor ECD

For the second system, a newer glycine-bound structure was available for use (6PM5) and therefore no additional modifications were necessary before preparing the structure as the binding pockets already matched criteria for typical glycine binding. This structure was derived from Cryo-EM of the zebrafish α-1 glycine receptor, desensitised, and prepared using styrene-maleic acid copolymers (SMA)[26]. The resolution of this structure is 3.10 Å and has relatively high validation scores with 0 Ramachandran outliers, 0 sidechain outliers and a clash score of 5 (Figure 3.3.C). This structure was chosen for the high validation scores indicating high structural accuracy as well as the pockets being populated with the endogenous ligand glycine. The desensitised state was chosen to reduce any instability that might come from the dynamic nature of open pore structures and any potential reliance on a full length structure for stability[148], given the focus on the binding pocket itself and the use of only the ECD, a structure that is defined by differences in the TMD[26,149] would not be as appropriate.

3.3.1 System Preparation

The first step of preparation involved isolating the ECD from the rest of the structure in the original PDB file. The first 218 residues of each subunit, as well as the bound glycine ligands, were selected to produce a new ECD only PDB prior to any preparation; this can be seen in Figure 3.3 A and B. This PDB included the whole ECD as well as one residue into the M1 helix; the last five residues would serve as points where restraints could be applied to hold the

structure in position without needing a membrane. These last five residues were chosen as this is the area leading into the TMD that interfaces with the ECD and where signal transduction across domains occurs[16]. A 'wrist-like' cuff is what joins the ECD and TMD with the ECD buttressed by the pre-m1 linker[8] and so to limit any impacts of removing the TMD and to mimic the effects of this region being held in place by the TMD and membrane, these last five residues leading into the pre-m1 region were selected for where harmonic constraints would be applied.

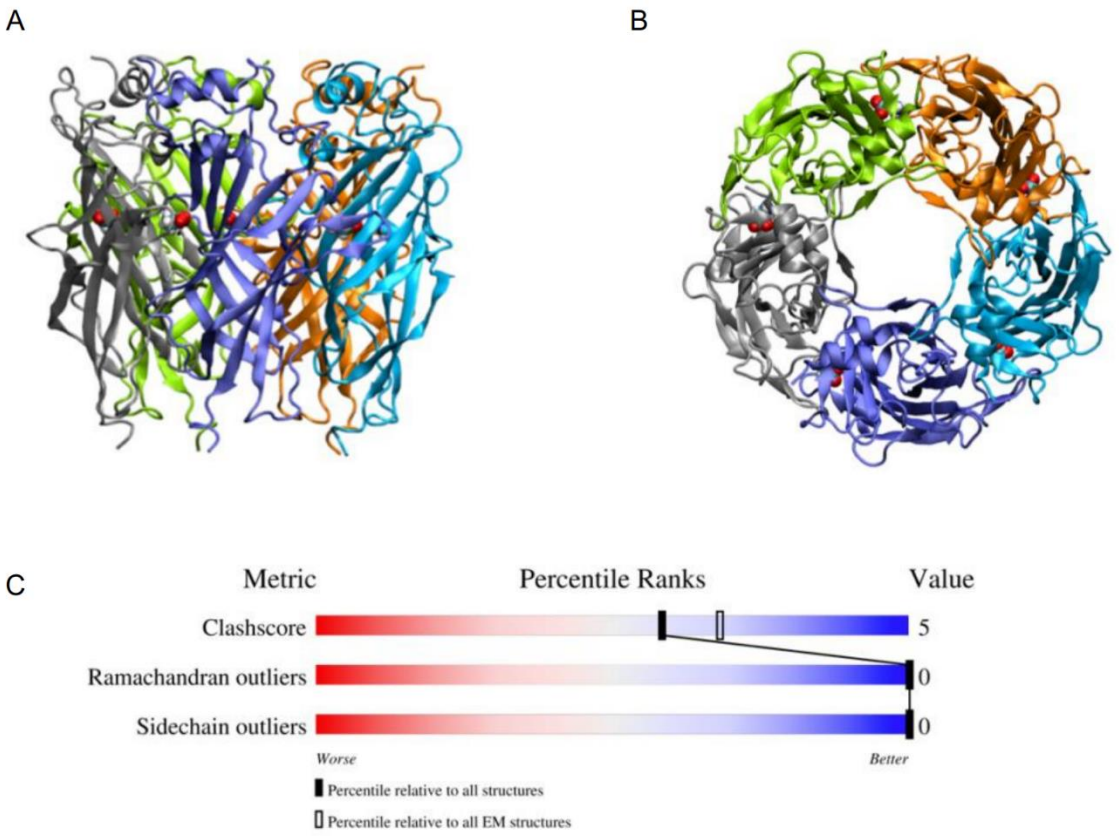

Figure 3.3: A and B are structural representations of the completed system and show the 6PM5 structure from a parallel view (A) and an orthogonal view (B). The original PDB validation results are also shown here (C).

Now that a structure had been assembled, the process of preparing for MD was carried out. The addition of hydrogen atoms (at neutral pH) not picked up by cryo-EM was carried out using the H++ tool[150] . The complete protein was then solvated and formatted for use with NAMD 2.14[96] and the amber forcefields ff14sb[90]. The protein was solvated using the TIP3P3 water system with a 15 Å buffer as well as 0.15 M of Na^+ and Cl^- ions. This step also included setting the correct disulphide bonds[141,142]. Following these steps, the final structure included a total of 123,299 atoms of protein, water, ions, and glycine ligands. The minimisation step was the first to be undertaken, and all simulations carried out, including this step, were done using NAMD 2.14. Other than the periodically repeated orthorhombic cell box size being 110 Å x

110 Å x 110 Å, all parameters, thermostats and barostats used were the same as that of the whole protein system.

3.3.4 Minimisation to Production

The minimisation was carried out in 4 stages, with restraints on everything but the solvent to begin with, followed by the removal of the restraints on the protein side chains, then removal of all protein restraints and finally, the removal of restraints on the GZW molecules. Heating was then carried out in incremental steps of 25 K until 310 K was reached, with harmonic restraints applied to the backbone of the whole protein.

Equilibration was carried out for 200 ns, and the harmonic restraints were carried over from heating at 25 kcal/molÅ^2 and slowly released over the course of the equilibration. The first 50 ns did not have any change in restraints, whilst the period of 50 ns to 150 ns involved the incremental reduction of restraints every 15 ns scaling down, as shown in the table table 3.2. For the duration of this, the restraints applied to the last five residues of each subunit were reduced to 1 kcal/molÅ^2. The production was run for a total of 1000 ns using this system, with analysis carried out across the whole trajectory.

Restraints [Kcal/molÅ^2]	Timepoint [ns]
25	0
25	50
10	65
1	80
0.4	95
0.04	110
0.016	125
0	140
0	190

Table 3.2: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was called to the subsequent value.

3.4 Whole Receptor Simulation Results

From the 1000 ns production simulation produced, a series of analyses were carried out to assess the stability of the systems as well as monitor the binding mode in each of the five pockets. Both systems were run for 1 μs during which: the RMSD of the structure was

evaluated for the binding pocket residues, the Loop-Cs, and each subunit. The data collected show the deviation of the protein over the course of production from the original positioning of the atoms following minimisation and heating. The binding mode was assessed by recording the interactions present, including hydrogen bonding, cation-π interactions and binding pocket volume. In this section the results from this analysis will be discussed as well as a comparison made between the two systems and their viability.

3.4.1 The Structure and Stability

The RMSD of the whole protein (Figure 3.4) as a function of time provides a way of judging the stability of the structure and potentially the quality of the equilibration/setup procedure, a gradually increasing RMSD or an RMSD with erratic/high RMSD values would suggest instability and potential issues in the methodology. The backbone atoms of all protein residues were selected, and RMSD was calculated with cpptraj[122] to be plotted for one frame per nanosecond.

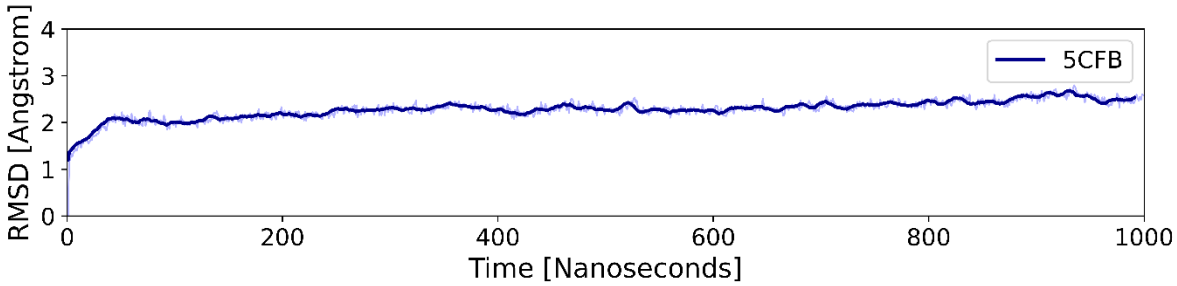

Figure 3.4: RMSD of the whole protein backbone as a function of time during the production run. The raw data is plotted with a paler colour and a running average is calculated for every 1 ns and plotted for clarity.

The plot in Figure 3.4 shows the RMSD climbing from the original frame to 2.5 Å and stabilising throughout the rest of the production. This is a particularly important step for the whole protein system as there are no restraints and the membrane plays a key role in maintaining the structural stability, a poorly made system therefore being more prone to drastic structural deformations, this measurement serves as a check of that process.

The RMSD averages shown in Figure 3.5 illustrate the deviation from the initial frame and therefore how much each selected group changes. The groups targeted included each whole subunit, the residues of Loop-C defined by the 4 residues either side of the tip of the Loop-C (THR-204), and the key binding pocket residues that were defined according to literature in the introduction[26]. Whilst there is very little difference in the RMSD of the subunits, there are identifiable differences between pockets for the RMSD of Loop-C and the binding pocket residues. Loop-C is an important component of GlyR binding, an increased RMSD value for this structure suggests reduced stability/increased flexibility. With Loop-C in an 'open' position or moved in any way from the 'closed' state typical for bound-state pLGICs, would result in a disrupted THR-GLY interaction and a more open pocket structure and therefore a reduced binding affinity. This is supported by the simulation results where binding pocket A and D have reduced RMSD values associated with Loop-C and retain their ligand for the duration of the simulation. The pockets with higher RMSD values for Loop-C however, fail to retain glycine in the pocket. Similar results can be seen for the RMSD values for the BP residues where A and D have lower results compared to the pockets that ultimately have their ligand ejected.

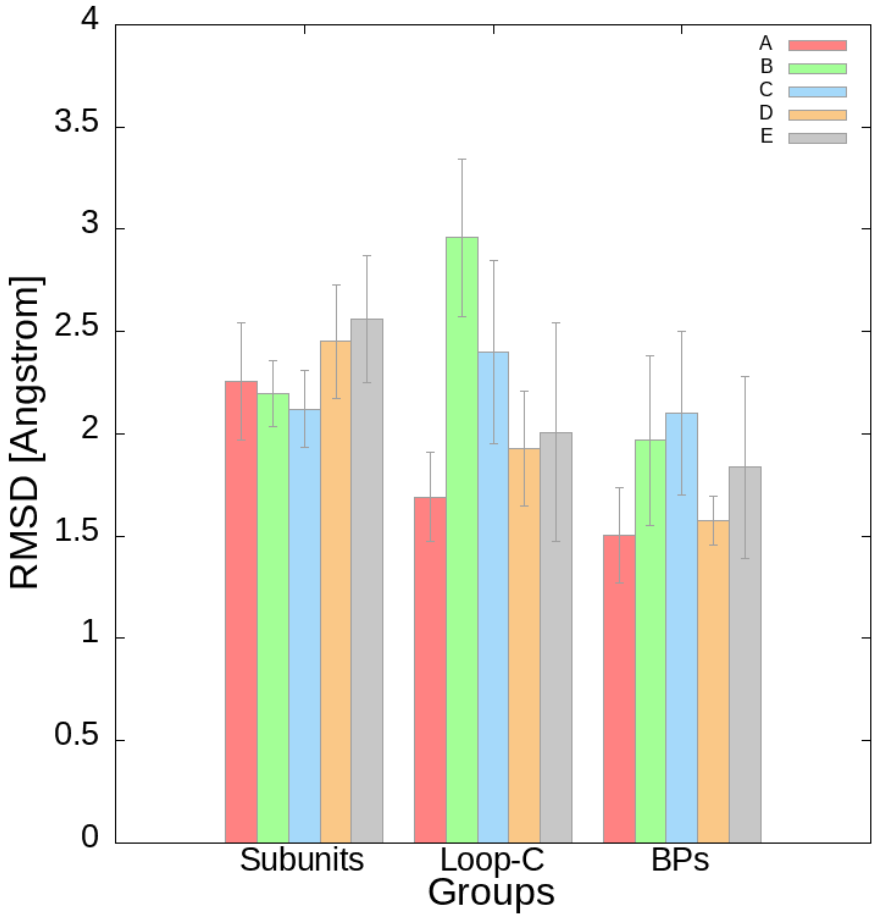

Figure 3.5: RMSD of the whole protein backbone as a function of time during the production run. The error bars presented are measures of the SEM over the course of the simulation for the specified pocket.

It is important to note that not all the pockets are required to be bound for GlyR activation[151]. The plot of RMSD over time gives greater insight into the changing stability in each group across the protein in real-time (Figure 3.6). These plots allow us to clearly observe the changing stability and movement of the different regions at specific timepoints and is useful when events such as the ejection of GZW molecules occur.

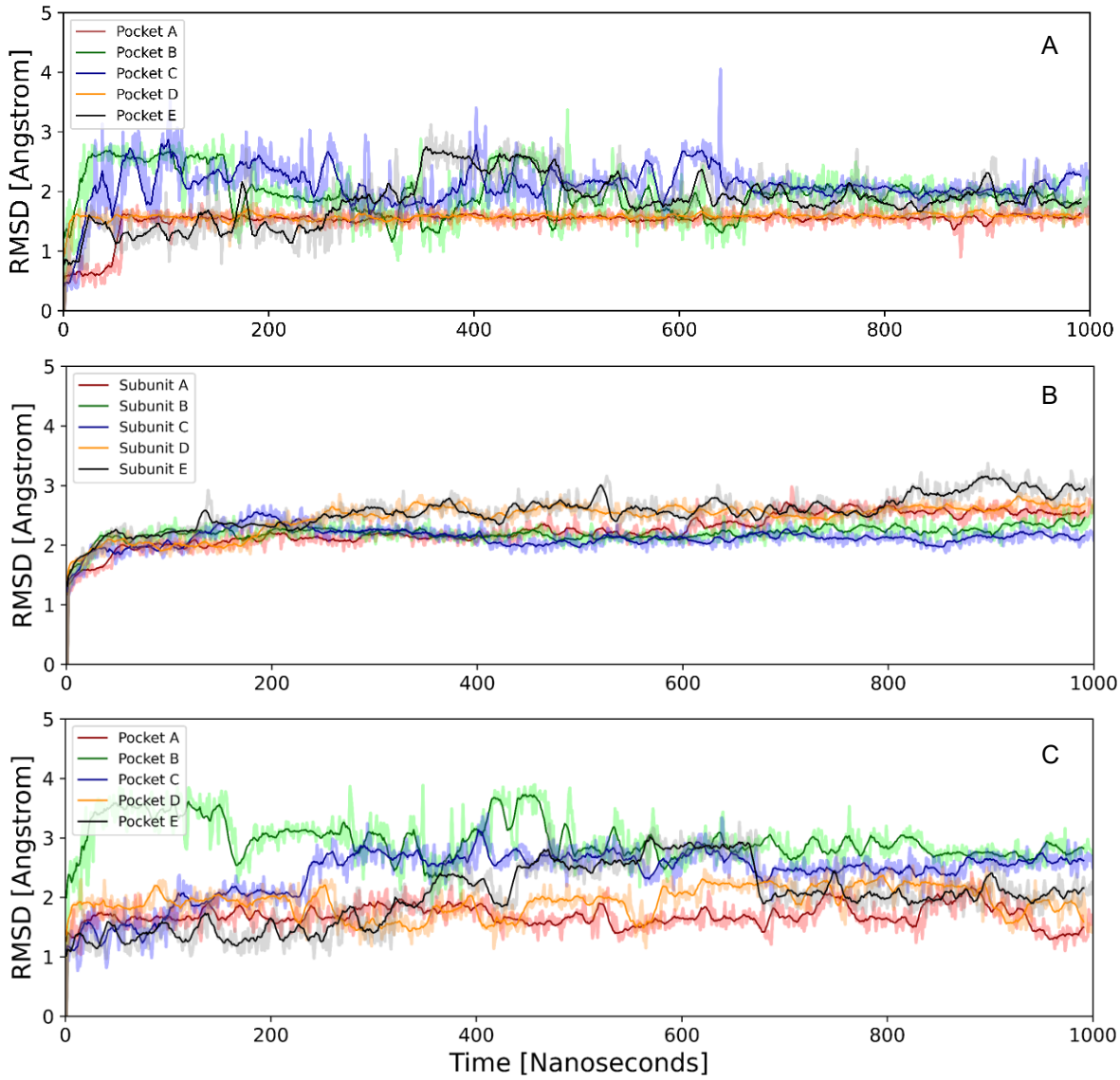

Figure 3.6: Time evolution of RMSD values for each binding pocket in A, the whole subunit structures in B, and the Loop-C structures in C.

The Loop-C structures show the most fluctuation, particularly in the pocket where ligand ejection occurs (A, C and E). The binding pocket RMSDs also highlight the stability of pockets B and D compared to the others where there is much more variability. The subunit RMSDs are mostly stable. Another aspect of the stability of the protein is in the binding mode; knowing a ligand is stable and bound correctly is critical for moving onto more advanced metadynamics

techniques. An unstable ligand could result in skewed free energy data or unreliable unbinding mechanisms.

3.4.2 Hydrogen bonding

A continuous plot of the total number of hydrogen bonds demonstrates how the occurrence of hydrogen bonding changes for each binding pocket through the production. In Figure 3.7, the binding pockets for the ligands which fail to remain in the pocket have a lower total number of hydrogen bonds to start with and even in binding pocket B, which remains bound for much longer, the number of hydrogen bonds never stabilises at 4/5 total hydrogen bonds which are seen for the stable ligands.

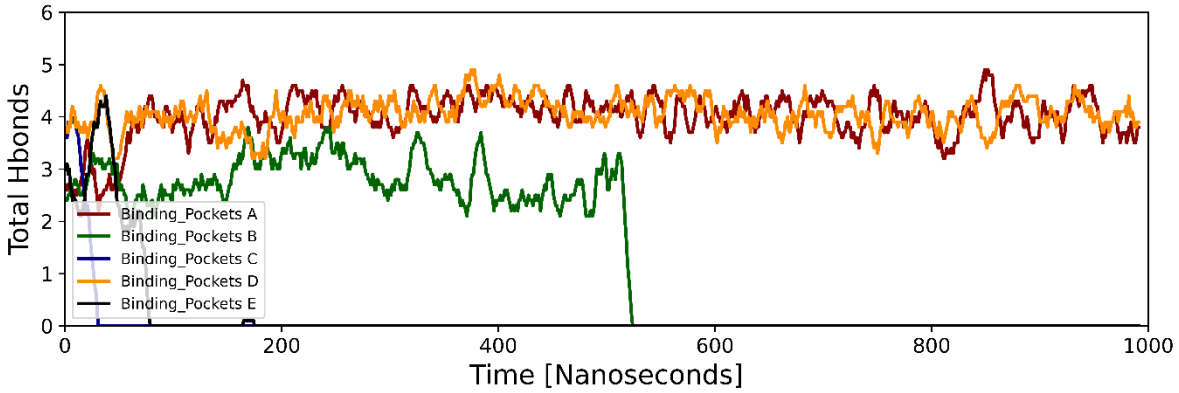

Figure 3.7: Profile of the total number of hydrogen bonds formed with the ligand over the course of the production. The pocket to which each ligand belongs to is coloured according to the key.

An example of a relatively stable binding of glycine can be seen in pocket D where there is a consistent set of hydrogen bonds between the ligand and protein (Figure 3.8) and the interactions observed match well with literature structural reports[26]. In Figure 3.8.C-D each hydrogen bond type present between the binding pocket and ligand is plotted with the fraction of the simulation it occurs for, prior to ejection of glycine in binding pocket B. Here, binding pocket D (Figure 3.8.A,C) has consistent interactions with ARG-65, SER-129, THR-204 and PHE-159, findings that are in line with structural imaging[26]. A water molecule can also be clearly seen fitting in between the ligand and residues GLU-157 and SER-158. By comparison, the interactions in pocket B (Figure 3.8.B,D), a pocket where glycine eventually ejected, are not as consistent or complete. In pocket B, whilst having much more variability in the total hydrogen bonding over time (Figure 3.7) the occurrence of each interaction is lowered compared to pocket D. Pocket B is missing any interaction with THR-204, PHE-159 and instead has hydrogen bonds form with the GLU-157 and SER-158 that are typically interacting with water in the stable pocket. From these results we can also see how much freedom each

ligand has for rotation within the pocket, in the case of the unstable pocket, the profile of hydrogen bonding shows that there is a preference for one oxygen over another to interact with ARG-65 whilst the stable pocket freely rotates between oxygen-Arginine hydrogen bond pairs.

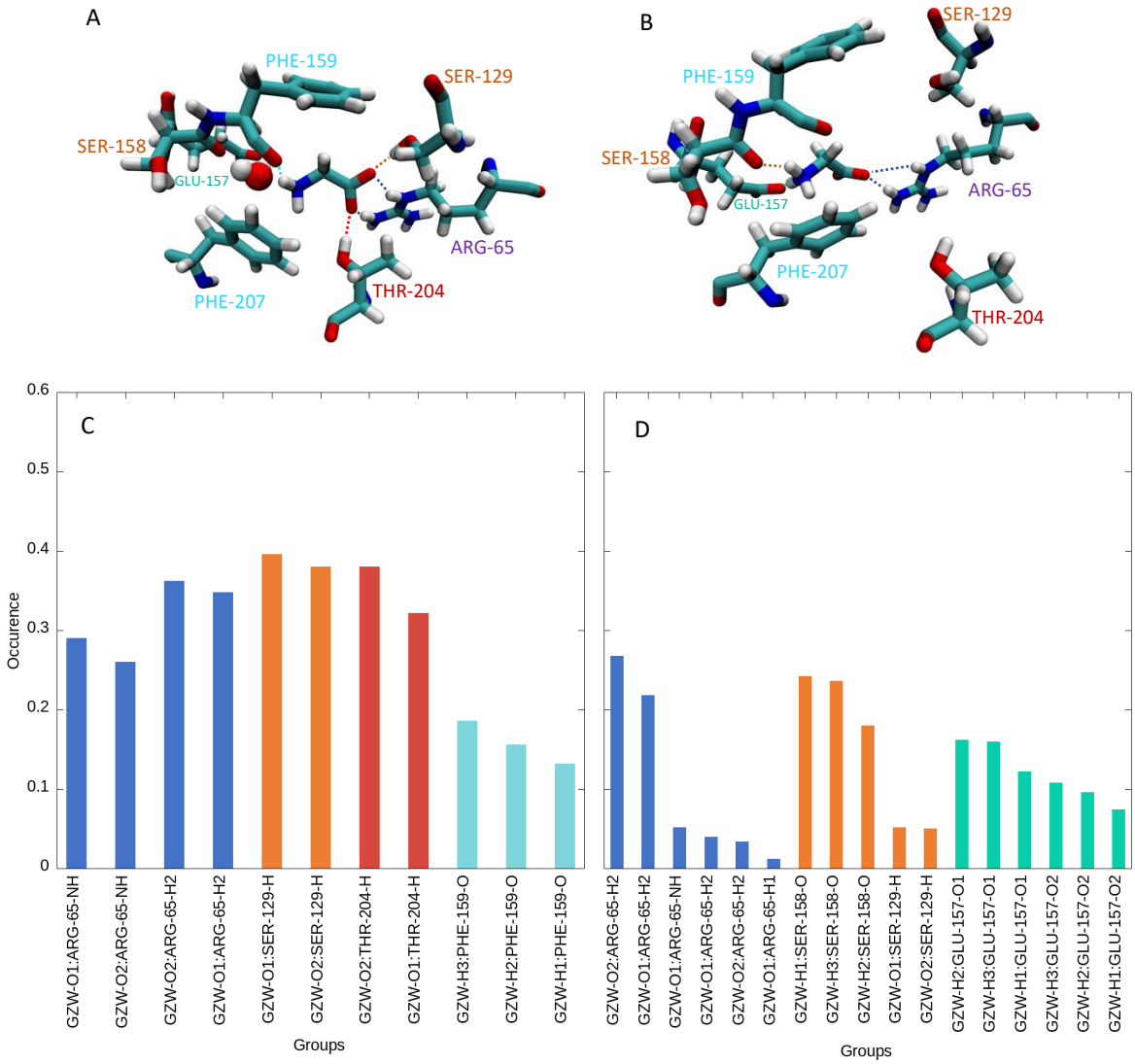

Figure 3.8: Profile of the types of hydrogen bonds formed over the production. A, a snapshot of the stable binding pocket D from the simulation with the residues and respective bonds labelled. B, a snapshot of the unstable binding pocket B with residues and interactions highlighted. C and D, here occurrence is used a metric to measure the strength of these interactions as a percentage of time present. In C, the graph represents the H-bonding profile of ligand in pocket D. D, shows the bonding profile of the weaker ejection prone pocket B.

3.4.3 Cation- π Interactions

As for other interactions, cation-π interactions can be considered where aromatic binding pocket residues are positioned to enable cation-π bonding. Here occurrence is used as a metric to measure the strength of these interactions as a proportion of time present. From the production, the occurrence of these interactions can be plotted (Figure 3.9) and shows that binding pockets E and C have no cation-π bonding to the ligand, whereas for A and D, there is constant interaction. Binding pocket B does have some cation-π interaction. However, this is much less than the stable A and D pockets and perhaps indicates some stability but not enough for retaining the ligand through the entire production run.

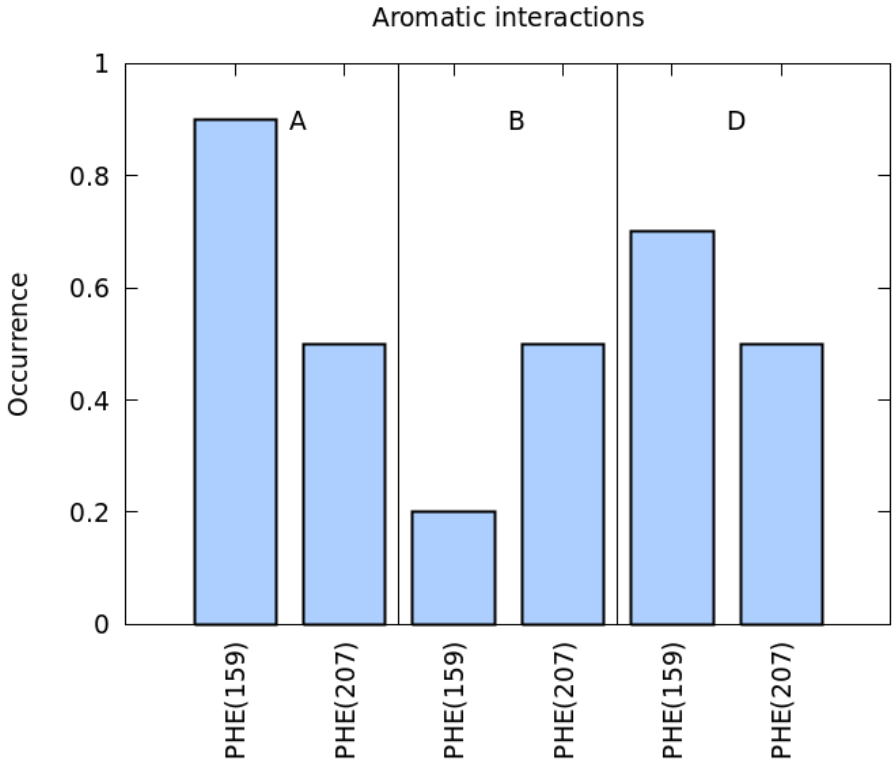

Figure 3.9: π -Interactions recorded for each subunit with the specific residue involved on the X axis and the occurrence as a percentage of time on the Y. Only subunit A, B and D were found to have these interactions present.

3.4.4 The Binding Mode

The binding mode of glycine in the receptor-binding pocket can be characterised with several parameters, the interactions as well as the conformational arrangement or position of glycine relative to other key residues. In the GlyR, the ligand is orientated so that the Arginine of the binding pocket is in contact with the carboxylate group of glycine. This orientation stated in literature is mirrored in the simulation work (Figure 3.8.A) where the position can be visualised as well as confirmed by the presence of specific hydrogen bonds. The differences in binding

pocket stability are further emphasised when considering which of these ligands remained in the binding pocket for the duration of the production run. The ligands of binding pockets A and D remained bound, whereas B, C and E were destabilised.

The relevance of Loop-C positioning discussed before also required a metric to define how ‘open’ or ‘closed’ it was. For this, the distance between the tip of the Loop-C and the centre of the binding pocket was recorded. By calculating this for each binding pocket over time, the results provide a general indicator of binding pocket validity and stability as a key element of the binding pocket is an interaction between glycine and the THR-204 in the Loop-C, displayed I figure 3.10.A. The Loop-C to binding pocket distance of the receptor is shown to be extremely stable and settle at around 6 Å with limited deviation for pockets D and A in comparison to the other subunits. These results agree with the RMSD values that showed binding pocket A and D both to have the most stable binding pockets and Loop-C structures. To illustrate the binding and ejection, a simple measure of the distance between the centre of the binding pocket and the centre of the respective ligand can be carried out shown in figure 3.10.B.

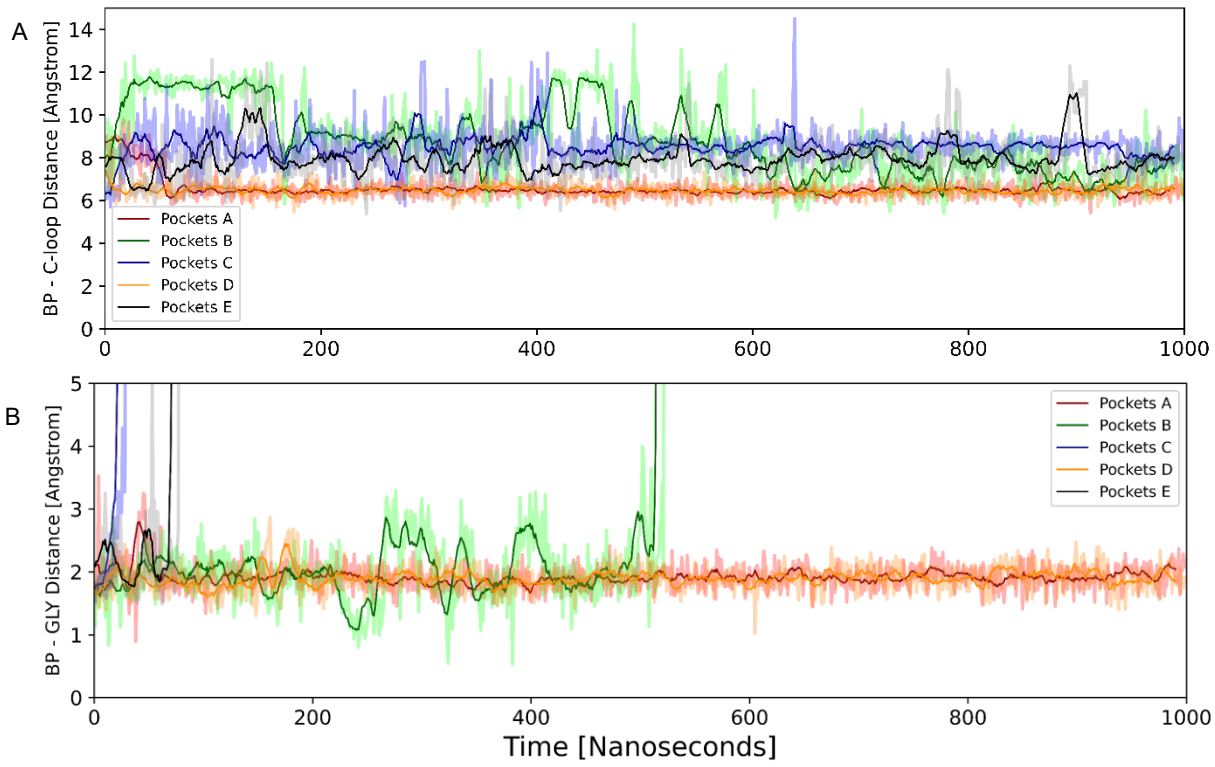

Figure 3.10: A, Distance measured between the centre of mass at the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues. B, Distance measured between the centre of mass of the glycine ligand and the centre of mass of the binding pocket residues. The time covered is the full 1000 ns of the production run and the different subunits are coloured according to the key.

Here the ligands for subunits B, C and E destabilise from the binding pocket and leave whereas the distance from the ligand to binding pocket in D and A remains constant (Figure 3.10), a result which also aligns with the drop off in hydrogen bonding in Figure 3.7.

Another important aspect of ligand binding often overlooked is the involvement of water molecules in the pocket for mediating interactions between a pocket and the ligand. In the case of GlyRs, this has been reported in both simulations[67] and structural imaging where the water is said to stabilise by hydrogens between GLU-157 and SER-158[152]. The importance of this water molecule has been probed in other simulations where a binding mode without water is seen but does not fit with previous mutagenesis results and therefore was dismissed[137], this water molecule is considered an important part of the ligand-protein complex. When this water molecule is not present, the binding pocket looks like that of Figure 3.8.B where the glycine ammonium moiety takes the place of the water, shifting its position in the pocket and reducing the stability or completely interfering with the known important ligand-protein interactions required for binding (Figure 3.8).

In these simulations with the 5CFB system, the water was not manually inserted but diffused into and stabilised within the pockets during the heating steps. Whilst the original waters would eventually move out of the pocket, they were always rapidly replaced, maintaining the solvent-bridge for much of the simulation. The consistency of the water bridges is illustrated in figure 3.11 where when water bridges were measured via hydrogen bond presence across through water between the glutamic acid (E157) and serine (S158).

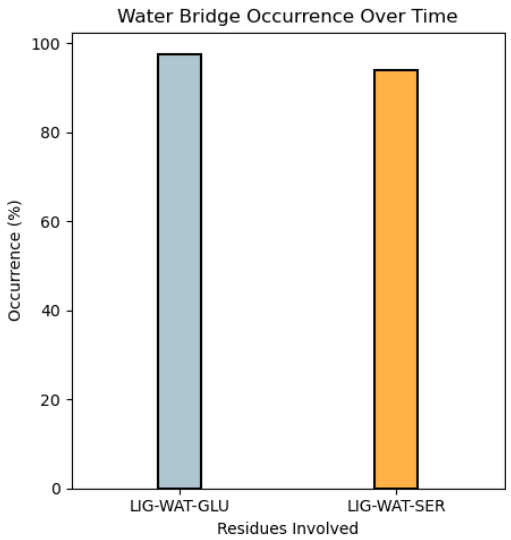

Figure 3.11: A bar chart representing the percentage of the simulation that ligand-water-protein bridges are present for, including the GLU (E157) and the SER (S158) residue. These were measured by calculating the presence of hydrogen bonds in cpptraj between the ligand, a water molecule in the pocket and the binding pocket residues occurring simultaneously.

The positioning of this water molecule along with the hydrogen bonds formed is illustrated in figure 3.12.A , showing the spatial arrangement of the key residues, glutamic acid and serine, with the ligand. The route through which water enters the pocket originally, is highlighted in figure 3.12.B, showing the path water takes from below the pocket and behind the C-loop.

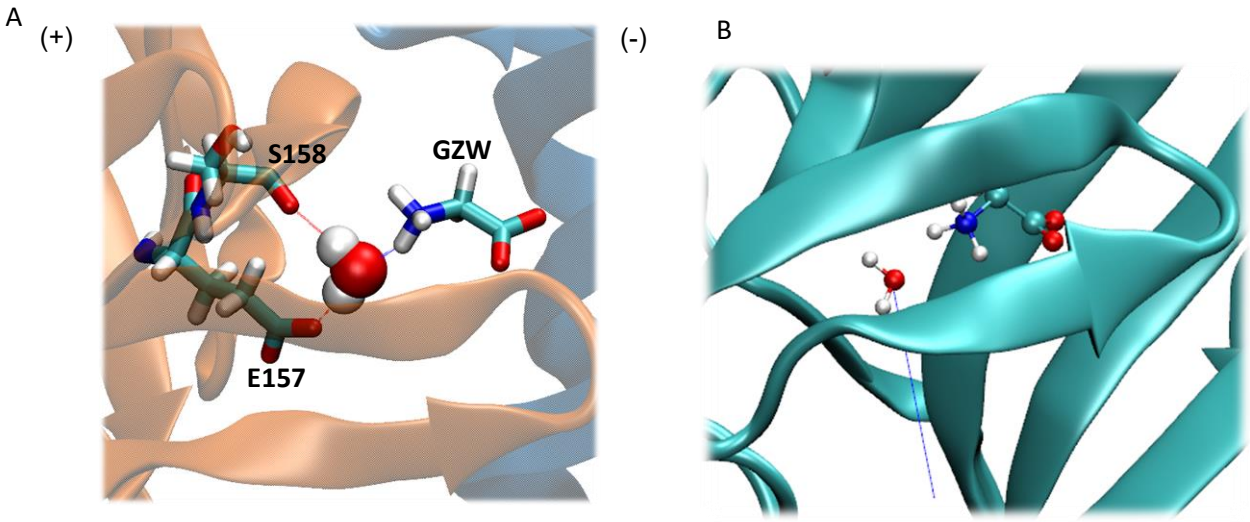

Figure 3.12: A, Diagram of the binding pocket of the 5CFB structure. The water molecule pictured in the system is stabilised by hydrogen bonds to the oxygen groups of S158 and E157 in red as well as a hydrogen bond to the ammonium group of the glycine ligand in blue. The (+) subunit is shown in orange and the (-) subunit in blue.B shows A cartoon illustration of the binding pocket with water and the ligand within. The blue line marks the path the water takes into the pocket, travelling below the pocket and in past the C-loop.

It has been shown that binding pocket contraction also has a role to play in ligand binding, where the full agonist bound structure has a smaller pocket volume compared with partial agonists. This contraction of the pocket has been suggested as part of a stabilisation process of the pocket interactions that is coupled to agonist-induced activation and therefore a more stable activation associated with the full agonist binding[26]. The pocket volume is therefore an important element to consider when discussing the validity and stability of a system as it would be expected to see this contraction of the binding pocket, especially following the removal of a larger ligand and replacement with glycine. The binding pocket volume was measured using the MDpocket tool[153] to track the individual pocket sizes through production. By comparing the stable and unstable pockets in Figure 3.14 we can see that the stable glycine retaining pocket was comparatively much more consistent than that of the pockets which eventually had glycine ejections which had much greater fluctuations in volume as illustrated in Figure 3.14. The mean volume was slightly higher in the stable pocket at 106.1 ± 31.2 $Å^3$ and in the unstable pocket volume was 96.4 ± 46.1 $Å^3$ (Where ± represents the standard deviation (SD)).

3.4.5 Conclusion

The 5CFB system production was finished at an early stage in the project and a new range of cryo-EM structures were released with glycine and multiple partial agonists bound. Several issues had become apparent when developing this system that presented challenges when moving forward. The biggest issues stemmed from the use of a system originally bound to strychnine, rather than glycine. Strychnine is a much larger molecule than glycine, resulting in a larger binding pocket volume and therefore requiring longer equilibration with glycine in the pocket for a stable state to be derived, the difference in ligand size is highlighted in Figure 3.13. The difference in the pockets is highlighted in the differences of backbone protein structure in figure 3.13.A, the clearer key residues highlighted in figure 3.13.B shows how the important binding pocket residues like arginine, threonine, serine and glutamic acid are shifted by the presence of strychnine in orange.

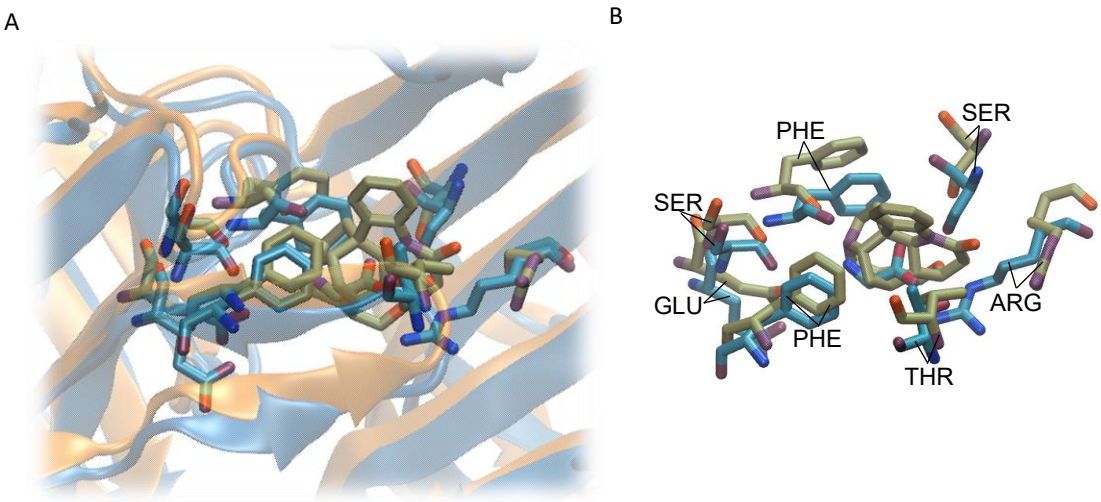

Figure 3.13: Here both binding pockets are illustrated and superimposed, using the newer glycine bound(6PM5) structure in blue to compare with the strychnine bound system (5CFB) in orange. The superimposed structures are shown with the backbone protein visible in A and the key residues highlighted in B.

We hypothesised that the larger starting ligand of strychnine may have expanded the pocket to a degree that was unable to constrict and stably bind glycine in the time frame used in equilibration. This is evidenced further in the results from characterisation of each pockets interactions showing that whilst stable pockets demonstrate cation-π interactions and hydrogen bonding that match with literature descriptions, most pockets however are unstable and do not form the predicted array of interactions, ultimately leading to the ejection of the ligands. As well as this, the volume of the binding pockets is much more variable, likely because of reduced interactions holding the pocket constricted. The contrast in behaviour of volume in the water filled and unfilled pocket is illustrated in figure 3.14.

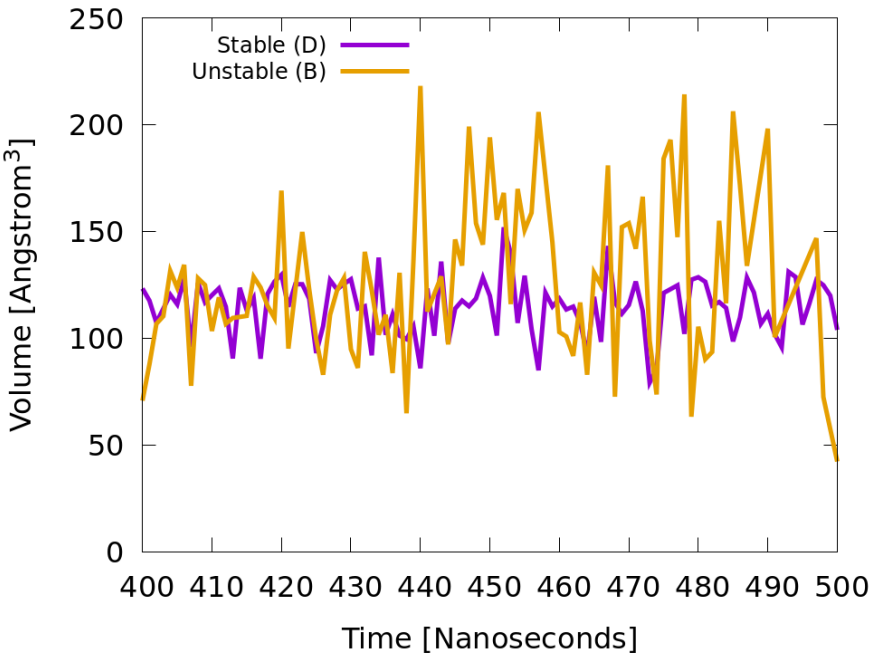

Figure 3.14: The volume of two binding pockets in the 5CFB system over the course of a 100 ns excerpt from the production. In blue, the unstable pocket B demonstrates much larger variability in volume prior to ejection later in the simulation. In Orange is the stable pocket D which whilst there is some fluctuation, it is much more consistent than the former pocket.

The elevated RMSD values for the Loop-C and BP residues support this idea, demonstrating how flexible the binding pocket and mobile key residues are. These differences are highlighted in Figure 3.8 where the profile of binding as well as location of the binding residues and ligand are not representative of the binding mode cited in literature[26].

In regards to the binding mode and interactions observed we can compare between the stable and unstable pockets to gain insights into the binding of glycine to this receptor. When comparing binding pockets, A and D with the unstable pockets before ejection we see a very different profile of interactions. In the unstable pockets the glycine ligand moves into the position that water typically takes and this shift away from the ARG-65 side of the pocket breaks the ARG-65 and SER-129 hydrogen bonds that would be expected to see along with the ligand carboxyl – ARG salt bridge. The displacement of this water and subsequent destabilisation of the pocket leading to ejection of ligands provides strong support for the critical role of this water bridge in glycine binding. The presence of water in the pocket seems to keep glycine from moving away from its central position in the pocket where the critical aromatic interactions and hydrogen bonds can form. This is likely a side effect of the destabilising and expanding effect strychnine removal had on the pockets but still demonstrates the important role this water molecule has in high affinity glycine binding.

3.5 The ECD system Results

3.5.1 The Structure and Stability

The assessment of stability and RMSD measurements were carried out identically to the previous section. The plot in Figure 3.15.A shows the RMSD of the whole protein backbone over time, the RMSD climbs from the start point before becoming stable between 2 Å and 2.5 Å.

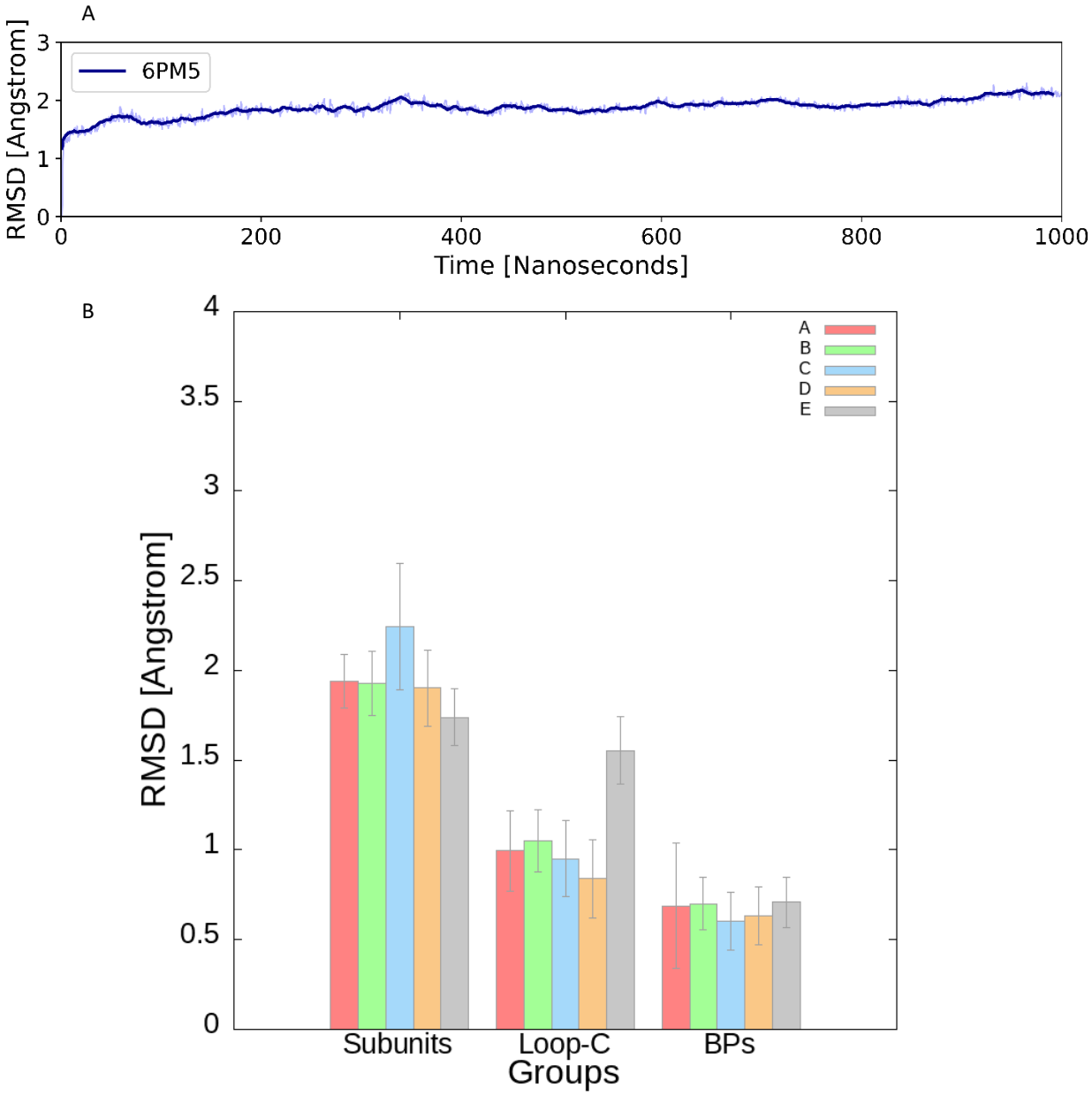

Figure 3.15: A, RMSD of the ECD protein backbone over time during the production run. B, Average RMSD values for each of the three important structures, the whole subunits, Loop-C and the binding pocket residues.

The RMSD averages shown here in Figure 3.15.B, unlike the whole protein system, have more uniformity across each subunit, especially across the general subunit structures. Loop-C in subunit E, however, does show a larger change than that on other subunits, whereas the binding pockets RMSD averages are uniform. Comparatively, the Loop-C and binding pocket RMSD values are much lower than that of the whole protein system by 1 Å or 2 Å, with the standard deviation also being reduced, suggesting a higher level of stability in the binding pockets.

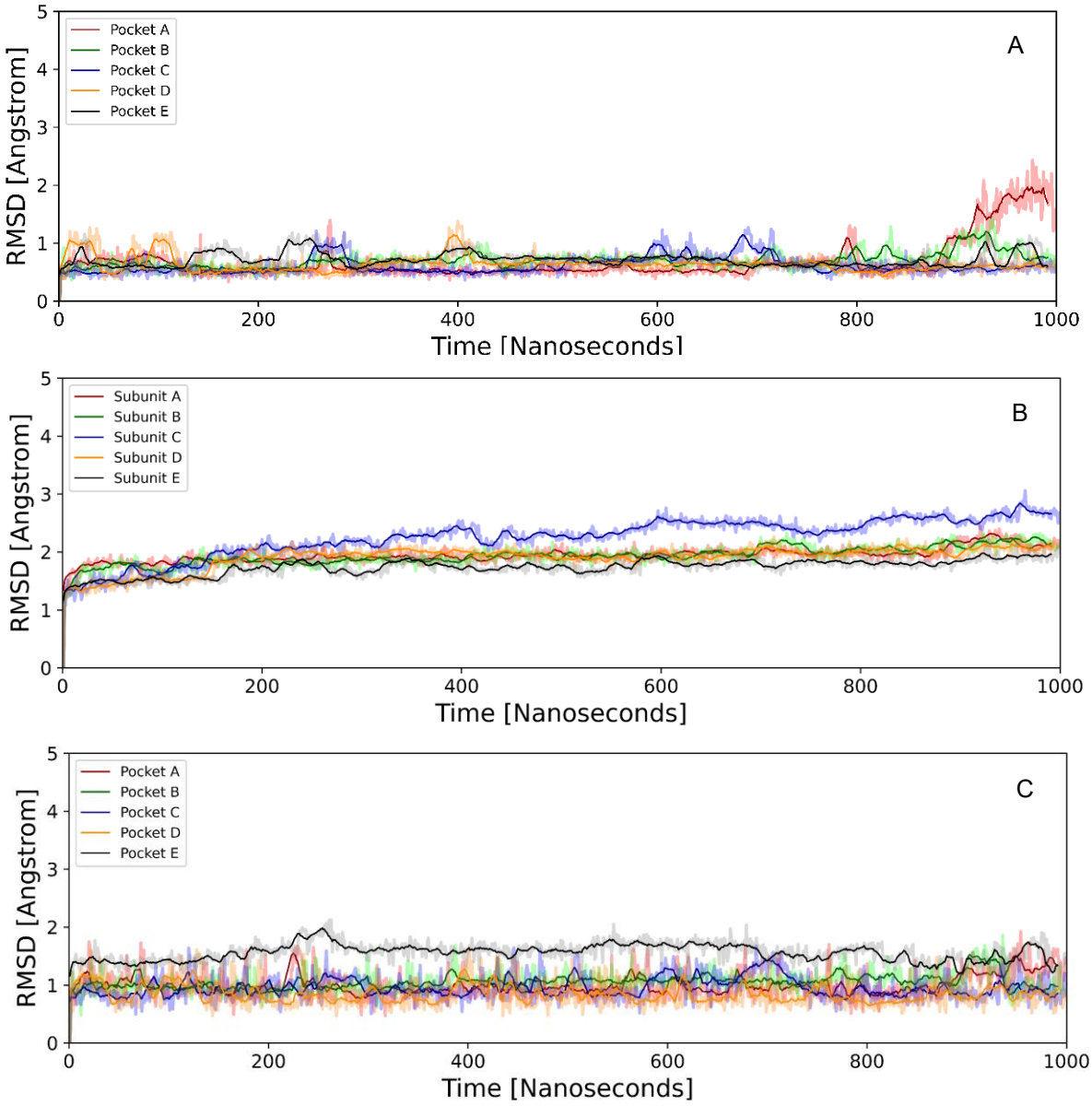

Figure 3.16: Time-evolution of the RMSD values for each binding pocket in A, the whole subunit structures in B, and the Loop-C structures in C.

The RMSD plots in Figure 3.16 show a higher level of similarity between the subunits over time and, in the case of the Loop-Cs, stabilises at a lower RMSD value than the whole protein system. Whilst there is a gradual climb in the whole subunits RMSD to a level similar to the full protein system measurements, the different subunits are grouped closer together and have a similar profile over the course of the production. There is some deviation in binding pocket A, specifically in the binding pocket RMSD where it climbs higher than the other pockets near the end of production, something not seen in Figure 3.15.B due to the short time span it occurs for.

3.5.2 Hydrogen bonding

The hydrogen bonding over time in Figure 3.17 shows between 3 and 5 consistent interactions over the production without loss of hydrogen bonding seen in the ejections of the full protein system. There is a drop in hydrogen binding in pocket A for a short period near the end of production; this fits with the slight destabilisation seen in the binding pocket RMSD also. This could suggest destabilisation of the ligand in that pocket but the other four pockets remain stable both in terms of the RMSD values and the hydrogen bonding.

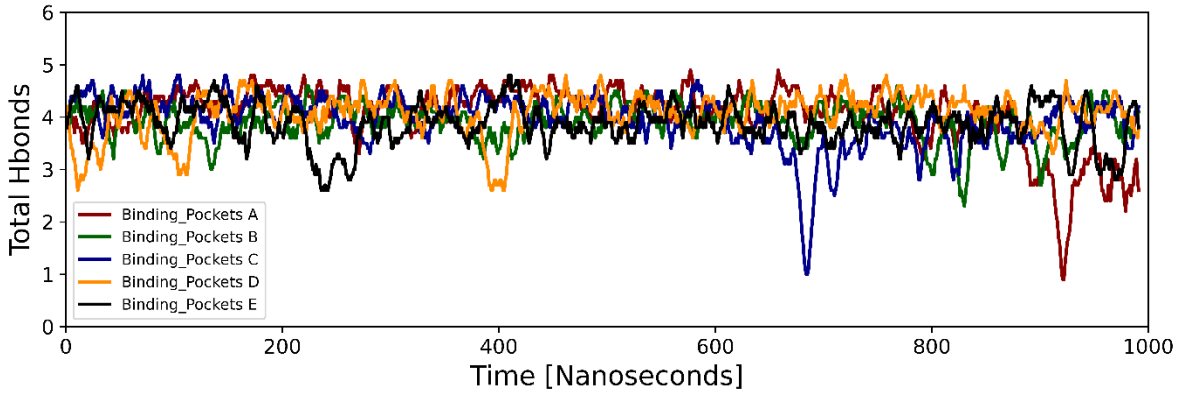

Figure 3.17: Profile of the total number of hydrogen bonds formed with the ligand over the course of the production. The subunit to which each ligand belongs to is coloured according to the legend.

Looking at an example of stable binding in this system (Figure 3.18.B), we can see a similar pattern of hydrogen bonding to that in the full protein system if not with more consistent bond formation. However, this profile of hydrogen bonds was present for all five binding pockets rather than just A and D, suggesting improved binding pocket conditions compared to the whole protein system.

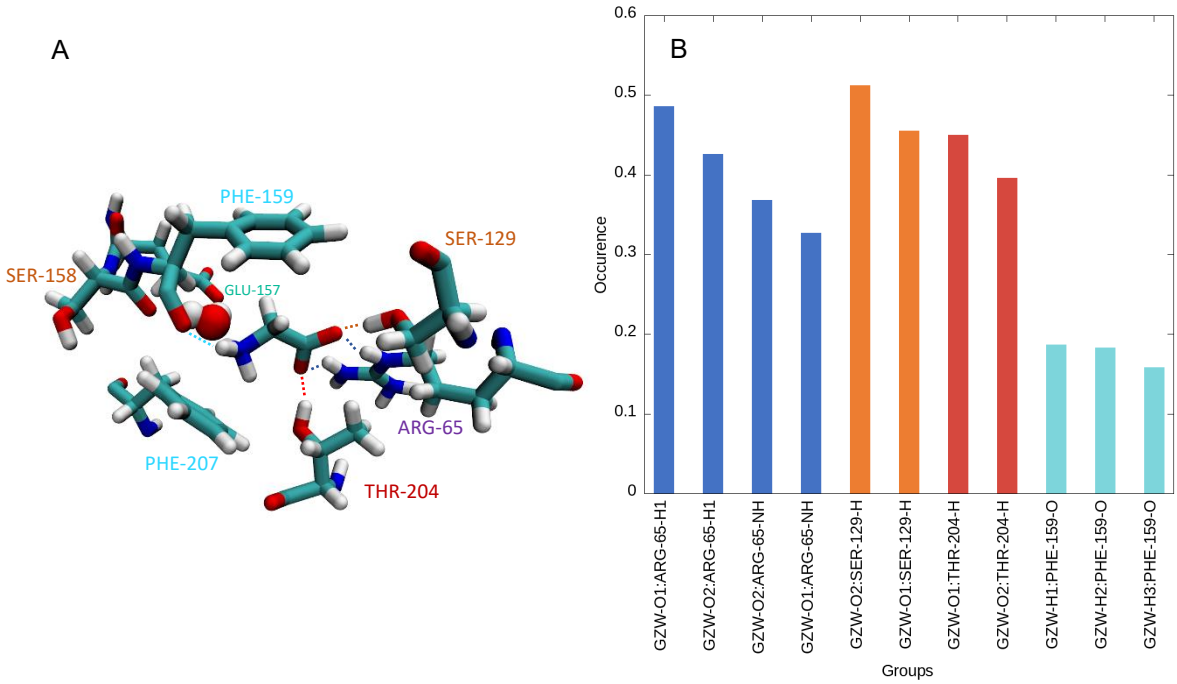

Figure 3.18: Profile of the types of hydrogen bonds formed over the production. A, a snapshot of the stable binding pocket C from the simulation with the residues and respective bonds labelled. B, a plot of the types of interactions occurring and for what proportion of the production they are present.

3.5.3 Cation-π Interactions

The cation-π interactions also act as a good indicator of the ligand binding in of the ECD-only system.

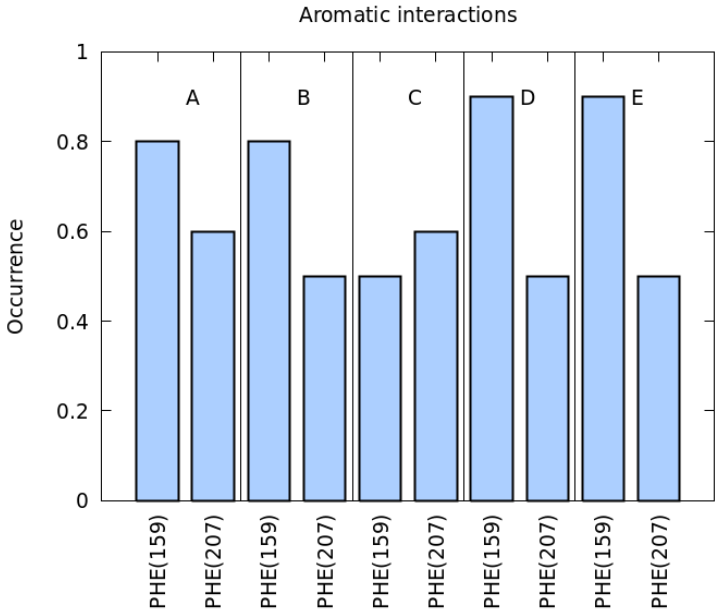

Figure 3.19: Cation-π interactions recorded for each pocket with the specific residue involved labelled along with for what proportion of the simulation time the interaction was present.

Where in the full protein system, there were only two pockets with regular π-Interactions, here in Figure 3.19, we can see that across all five binding pockets, there is consistent GZW to PHE-159 and PHE-207 interaction. The profile of interactions, as well as the other measurements shown here, provide a great indicator that the ECD system has five stabilised bound ligands. Whilst the full protein system has three binding pockets where glycine was ejected, the remaining ligands do show stable binding as indicated here and so are still usable for further experiments.

3.5.4 The Binding Mode

For the distances between THR-204 and the Loop-C in Figure 3.20, there are spikes of movement away from the pocket but not on the same magnitude of that in the whole protein system, and they are often brief. For most of the production, the Loop-C is stable at around 6 Å from the binding pocket centre. The largest opening of the Loop-C is seen in subunit A at the end of the production.

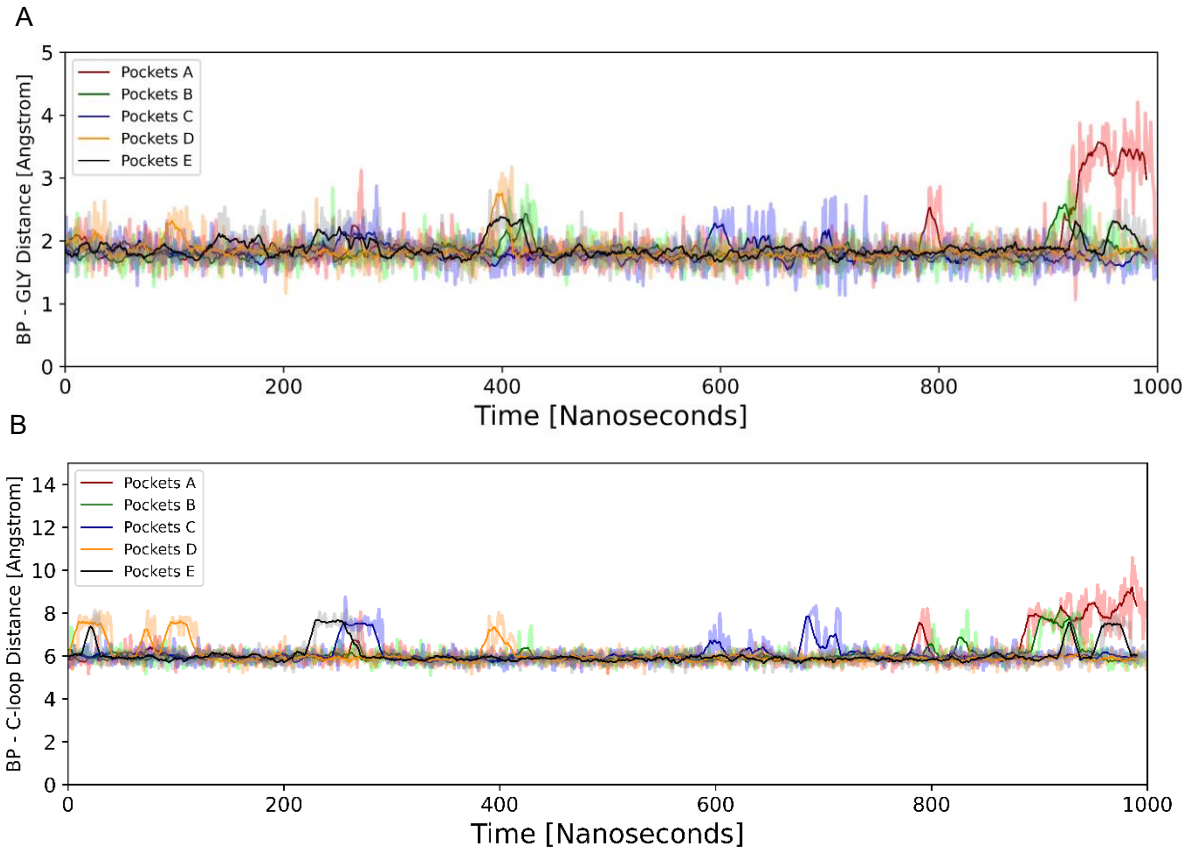

Figure 3.20: A, Distance measured between the centre of mass at the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues. B, Distance measured between the centre of mass of the glycine ligand and the centre of mass of the binding pocket residues. The time covered is the full 1000 ns of the production run and the different subunits are coloured according to the key.

The distance between glycine and the centre of the binding pocket is very similar to that of the whole protein system, stabilising between 1.5 Å and 2.5 Å, with the main difference being that ligands C, B and E remain stable at the same level as A and D in Figure 3.20. There is an increase in ligand A distance to the pocket centre; this suggests rearrangement of the pocket and movement of glycine which also aligns with the reduced hydrogen bonding seen also at the same time interval. The increasing distance of THR-204 from the binding pocket in A as well as the increased glycine to binding pocket could be attributed to a temporary fluctuation but given that the other binding pockets never reach this high of a value for the BP to THR-204 distance or Gly-BP distance it is likely that the pocket is destabilising. However, the other binding pockets are shown to be highly stable in this system from these distance-based measurements.

3.5.5 Conclusions

The 6PM5 system was developed with the newest structural data on glycine receptors. The aim of the analyses was to characterise the binding mode as well as verify that the methodology had not compromised the binding site in any way. The results displayed here demonstrate stability of the ECD and binding pocket as well as a range of interactions which has previously not been modelled in such a high quality structure before. The data collected has given us a strong justification for moving forward with the 6PM5 system and provides evidence that the structure is suitable for applying metadynamics and investigating binding without concerns of instability jeopardising the validity of any further data collected.

3.6 Discussion

In this chapter we focused on describing and analysing two systems of GlyRs for use in profiling ligand binding, including a breakdown of the stability of each system, the interactions present and how each system could be used in pursuit of the overarching aim of investigating the binding of these proteins.

Both systems have demonstrated adequate stability to be used in further experimentation but serve very different purposes moving forward. A full protein system embedded in a membrane has several extra uses that could be explored in future work. The primary difference between the systems is the presence of the membrane and transmembrane domain in the whole system as well as a lack of restraints. The interaction of lipids with ion channel proteins has

been explored in recent years several times; whilst lipid composition has already been discussed, the system built here could be used to profile and identify key interactions that are made with this channel. Lipid-protein interactions have been cited as being important for ion channel function in one of two ways, the non-specific effects from general lipid qualities where hydrophobic effects cause deformation and conformational change of the protein, altering both structure and function [66,145,154,155]. More specifically, binding and intercalating of lipids into the transmembrane domain has also been reported as influencing channel functioning. The full protein system is an ideal platform for profiling how lipids can interact with the glycine receptor and provide information that cannot be gained from the reduced system.

Whilst lipid-protein interactions can be profiled with this full system, the limitations come from the lack of binding stability. As previously mentioned, we predicted that the issue of binding stability stemmed from the use of a system that was originally bound with a much larger ligand, strychnine. The systems have been developed with the intention of carrying out binding and unbinding experiments to probe the binding event and gain insight into how glycine interacts with the binding pocket, the free energy surface of ligand binding and information on the binding path. The hydrogen bonding, π-interactions, RMSD measurements and binding pocket volume have shown that the 5CFB has unstable binding leading to three ligands being ejected during production. The 6PM5 system has consistent interactions with key residues identified in previous literature and retains all ligands throughout the production with reduced RMSD values for both the Loop-C and the key binding pocket residues. These interactions are what contribute to the free energy surface and free energy of binding that are to be explored with metadynamics and therefore a system like this will likely yield much more accurate results. The two remaining ligands in the 5CFB system do show a stable binding profile, but the three ejections limit our choice of replicas. As well as this, the increased RMSD of the BP residues and Loop-C suggests potential issues in ligand binding stability that may have a large impact on any metadynamics carried out. A longer equilibration could potentially remedy these issues of stability; however, the publication of the newer already glycine-bound higher quality structures including 6PM5, led us to the conclusion that a new system would be a more reliable approach. The development of this system however does highlight the importance of the interactions and binding pocket constriction in glycine retention once bound. This work also set up a framework for determining binding stability that would be utilised throughout the rest of the work presented here.

A further element taken into consideration with moving on with the 6PM5 ECD system was the complexity of using the whole 5CFB structure. The full protein system is very large and highly computationally demanding and so rebuilding it for new ligands, mutations, or any other

changes will take a lot of time and resources. For example, the 5CFB system with the membrane consists of 205,983 atoms whereas the 6PM5 system uses around 123,299 atoms. Moving to the 6PM5 system reduced the time required to run the simulations by over half, whilst still allowing the same intended results to be obtained.

Another key difference is the source of structural information. The project began before the 6PM5 structure was published and so the 5CFB structure was the best to use at the time. There are a few key benefits that the newer 6PM5 structure has over 5CFB worth considering. The X-ray crystallography process used to produce the 5CFB structure inherently has issues due to the crystallisation process requiring the protein to be in a non-physiological state to be imaged. The cryo-EM approach used for 6PM5, however, has the benefit of being imaged in a near-physiological state. In the case of 6PM5 specifically, this structure was imaged using styrene-maleic acid copolymer (SMA)82. Extraction of the protein using SMA conserved a patch of native endogenous lipid around the protein and therefore ultimately produced a structure in as native and physiologically relevant state as possible [15,53,55,156].

As well as the issue of resolving technique used, the 5CFB had to be modified as previously described to remove strychnine and add in a glycine ligand which likely led to a less stable binding conformation. Considering these issues with the starting structures leans to the 6PM5 system being more appropriately set up, especially for kinetics and metadynamics experiments. Whilst the full protein system has the benefit of a whole structure and membrane, which is much more physiologically valid, there is not a strong reason to include these extended structures when simulations focus solely on the binding process on a timescale that would not allow effects to propagate to the transmembrane domain. In fact, the speed at which gating occurs has been taken into consideration in the development of the extracellular domain system; experimental and theoretical opening rates are in the region of half a millisecond[157–159]. These time scales are far beyond the scope of the simulations carried out here, and so the conformational changes that are involved in open and closing events are highly unlikely to be captured in the timescales used in this work. The time needed for a binding event to influence regions of the protein in the transmembrane domain or vice versa, therefore, do not need to be considered when focusing only on the binding itself. We did predict a possibility that flexibility in the pre-M1 region where the 6PM5 system was restrained could have an impact on the Loop-C.

Native GlyRs have a very high affinity for the glycine ligand, this is due to the strong ligand-protein interaction and similarly how well the glycine ligand ‘fits’ and interacts with the pocket. The integrity of this strong interaction is sensitive to changes in the topology of the pocket.

Changes in binding pocket residues for example in mutated GlyRs results in large reductions in efficacy[160], or in the case of this work, the residues stay the same but the positioning changes compromising the binding stability. In the strychnine bound structure the repositioning of arginine leads to the pockets being largely unable to retain the bound ligands, indicating also a potential drop in the binding affinity that would likely lead to a drop in efficacy. The idea of the structure being highly specific and this topology being important for the efficacy is supported by comparisons across the binding pockets of various structures, both different subunit variants and across differing species. The heavily conserved nature of these channels has been documented previously[161]. An example of this can be illustrated by sequence alignment of the zebrafish and human GlyRs showing how the binding pocket structures are extremely conserved across species, illustrated in Figure 3.21.

Figure 3.21: The sequence alignment between two ECD structures, the top row shows the protein sequence for an α-1 Zebrafish (6pm5)[26] sequence whereas the second row of each pair is the sequence for the human α-1 receptor (8DN4)[173].

Utilising a structure with a lower resolution, poor validity scores or any structural rearrangements not expected in an endogenously bound state could break interactions, reduce affinity, and therefore undermine any observations made. The results seen with the strychnine bound structure simply highlights this, where the original unedited arginine resulted in glycine unbinding before equilibration had finished. Even after this was modified, some pockets still showed signs of destabilisation, the volume fluctuating and some ligands ejecting demonstrating the nuanced impact that small structural changes can have and the importance of using high validity, high resolution structures with little error to get stable binding as even a small deviation in the structure like the rotation of arginine can completely break stable binding. The ECD system utilised in this chapter and the structures used in later chapters avoid these issues by starting with structures captured in cryo-EM with the ligands bound and the binding pocket residues arranged appropriately. Each structure used also shares the same species and subunit composition and so the only functional differences that arise will be due to differences in ligands in chapter 4 and mutations in chapter 5.

To verify the restraints would not be an issue, the RMSF for both systems was measured and compared as illustrated in Figure 3.22. The RMSF values here show how much the structure fluctuates over the trajectory using the average structure positions for reference. Comparing both the 5CFB and 6PM5 systems we can identify the impact of using an ECD only system vs whole protein, if any. In the 5CFB system, the last 5 residues of the ECD unrestrained but leads on to the transmembrane domain residues. There is very little difference between these structures in terms of how much the structure changes at the end of the ECD with the addition of restraints and therefore we concluded that the ECD system was an appropriate choice moving forward given the negligible chance of any interference with the binding pocket residues and Loop-C structure.

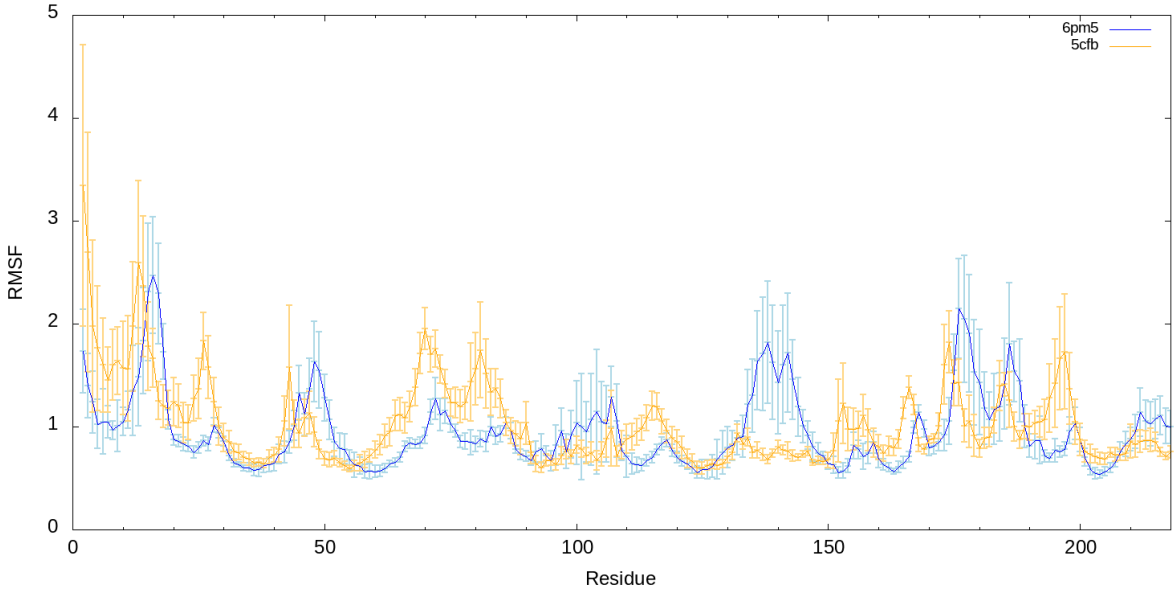

Figure 3.22: The Root Mean Square Fluctuation for both the 5CFB (In orange) and 6PM5 (In blue) for each residue averaged across all five subunits, with the average structure calculated as the reference.

In conclusion, when evaluating the differences between the systems, they both have potential use in future work. Both are stable and established, ready for the application of more advanced methods. Whilst the larger whole protein system has issues with regards to the stability of the binding present and a reduced number of pockets usable, we can investigate lipid interactions and broader changes. The ECD system has excellent stability, and all five ligands are ready for further experiments and the findings with regards to binding mode are the closest aligned to literature. In the next two chapters, we will discuss the experiments and findings from the use of the ECD system developed in this chapter to investigate ligand binding of GlyRs in the context of mutant variants of the receptor and partial versus full agonism. The next chapter will focus on the latter, where a description of the system set up using GABA as a partial

agonist was used along with the implementation of funnel metadynamics to probe the free energy landscape of the pocket through multiple binding and unbinding events.

Chapter 4

Partial and Full Agonism in Glycine Receptors

4.1 Introduction

Understanding ion channel biology is not a trivial task and it is an important one for two reasons. Firstly, ion channels are the second largest target for existing pharmaceuticals following only G-coupled protein receptors[162]. Secondly, also the reason for these proteins being such important targets, ion channels are the core units that mediate nervous system functionality and are therefore integral for greater understanding of neurobiology. The activation of LGICs is triggered by the binding of a ligand; in the case of the GlyR the endogenous ligand is glycine. Structural evidence suggests that between glycine binding and the activation of the channel the pocket undergoes contraction, involving the movement of loop C into a “capped” position and the movement of loop B towards the ligand, as in the schematic below (Figure 4.1)

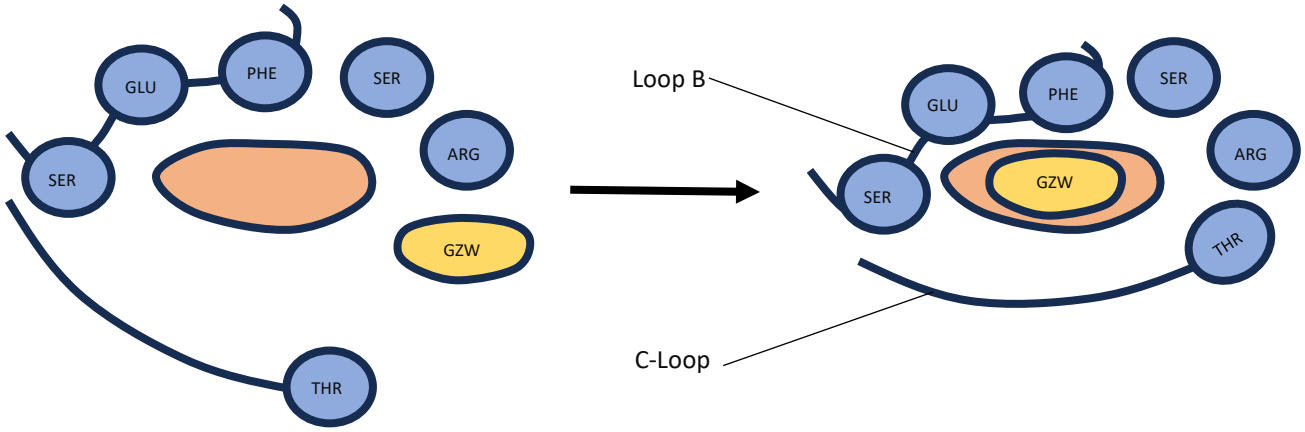

Figure 4.1: A schematic showing the basic arrangement of residues within the pocket and how structural data has shown the C loop to move from an uncapped to capped pose when glycine binds, changing the structure from the Apo open/closed state. The C loop structure here is illustrated by a curved line capped with the appropriate residue threonine.

Current research has been conducted using cryo-EM imaging of the various states the GlyR can occupy, however the dynamics and minute mechanisms that occur between these states

is poorly understood. Employing computational methods to further explore this process is the goal of this chapter, specifically by using metadynamics.

Metadynamics is a powerful tool for exploring the conformational space of complex systems such as proteins. It can be employed in a variety of scenarios with a particular use in accelerating rare events that cannot easily be studied within the limits of conventional molecular dynamics. Ligand binding and unbinding fit into this categorisation of rare events and therefore are ideal candidates for the application of this technique. The method allows for the estimation of the free energy of complex systems[163] and in the case of protein-ligand binding events, the ligand binding free energy should be obtainable. The binding free energy is a powerful metric used for understanding biological systems and in the process of drug development as it is commonly used to determine the affinity of a drug compound as well as an indicator of efficacy[164]. However, to determine the ligand binding free energy and effectively characterise the free energy landscape of a binding/unbinding path with metadynamics, several binding/unbinding events are required for the free energy calculation to converge and provide a well-characterised binding free energy[120]. To simulate multiple crossing events between the unbound and bound state, as discussed in Chapter 2, a funnel restraint can be applied to the system and used to keep the ligand within the appropriate area of conformational space to interact with the pocket and limiting the solvent exploration, crossing between the bound and unbound states in a practical length of simulation time. The funnel restraint is an important step as without it the ligand would likely use large amounts of computational resource and time exploring much of conformational space in the bulk solvent in poses completely unrelated to binding/unbinding. This approach has been used several times in the past, even specifically with pLGICs. Previous work applied the funnel restraint with the intention of simulating many binding/unbinding events and using the data collected to characterise the sequence of events in the binding process of an insect resistance to dieldrin (RDL) receptor, a GABA-activated pLGICs. In this work the free energy landscape was derived from wild type and mutant forms; from this the systems were compared by analysing at the interactions and dynamics of various components throughout the binding/unbinding process[86].

Glycine is a full agonist for the GlyR, meaning that it yields maximum open probabilities in channel proteins and therefore has very high efficacy when compared to any other compound. In a pharmaceutical context, these properties can be highly desirable as a maximal response of a given target may be needed for a therapeutic effect to be mediated, especially without substantial side effects from non-specific binding or ‘dirty pharmacology.’ In terms of the understanding of a given receptor, insight into how full agonists mediate such potent effects is crucial for a better understanding of the molecular mechanics at play. Full agonists are

potent because they interact with and change the protein, they are bound in the most efficient way to mediate a shift in conformation from one state to the next. Identifying these key interactions and the dynamics at play during binding of a full agonist could be used to inform pharmaceutical development of new compounds or to guide further investigations of processes that mediate conformational changes from binding to channel opening. Observations made of the binding/unbinding events and analysis of the free energy landscape could be used as a system to base further work. More specifically, by identifying which elements of the process are impeded by a partial agonist, we can form a greater understanding of the interactions and dynamics of the receptor and identify which components play more important roles in triggering conformational changes.

Both understanding the dynamics at play within a protein from its inactive to active state as well as being able to reproduce these effects with pharmaceutical agents rely on identifying the key characteristics that make full agonists so potent. In this chapter, we aimed to elucidate the differences between partial and full agonism in the glycine receptor. The next section will provide an overview of the system design and parameters followed by a discussion of the results taken from the simulations. Following on from this, the setup of the funnel metadynamics will be discussed and then a comparative analysis of the two systems in the context of ligand binding and partial versus full agonism of GlyRs will be presented.

4.2 Modelling the GABA Bound Glycine Receptor

Neurotransmitters like glycine for the GlyR are typically full agonist; however, there is a range of alternative compounds capable of binding at the orthosteric site that do not result in a maximal response; these are partial agonists. For the glycine receptor, the list of partial agonist includes compounds like Taurine, β-alanine, and GABA. In the case of the GABA agonist, this ligand can be co-released with glycine at synapses containing GlyRs where the ligands compete, thus modifying the time course of the postsynaptic response[26]. Previous work using Cryo-EM has produced structures that have GABA well defined within the binding pocket. The structure in Figure 4.2.A-B shows the Cryo-EM map produced[26] overlayed on the PDB used for this system. The cryo-EM map also highlights the capture of the orientation and positioning of GABA along with the suggested interactions, as shown in Figure 4.2.C.

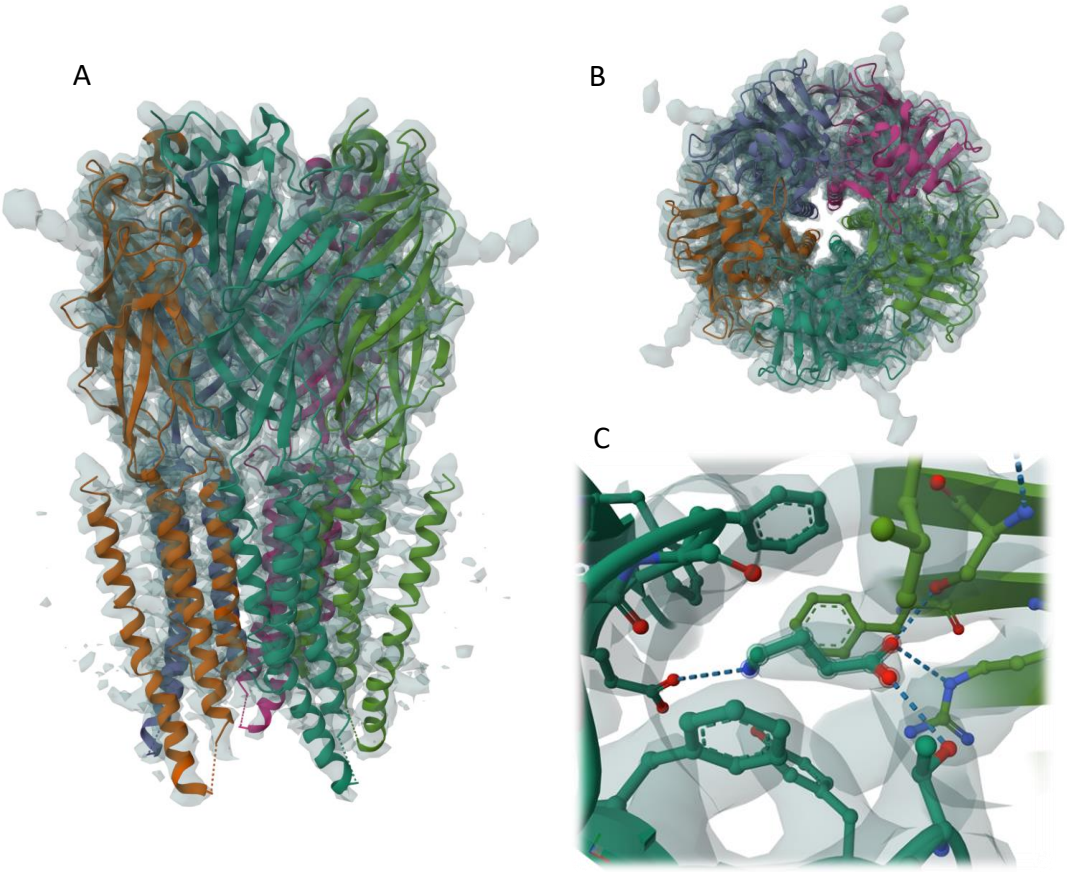

Figure 4.2: A, The structure of 6PLX with the cryo-EM map data overlayed from a horizontal view. B, the same structure with cryo-EM data overlayed from a top-down perspective. C, The cryo-EM data of the binding pocket of 6PLX along with the predicted interactions.

To system the glycine receptor with a partial agonist bound, this structure produced with GABA already bound was chosen to develop the system with. This structure was chosen with the same criteria as the glycine bound system discussed earlier but also with the intention of comparing the two; it was important to eliminate possible confounding effects of alternative methods to derive the structures and so choosing this specific system was appropriate given the experimental procedure was identical except for the ligand used. Details on the chosen system will be discussed in the following section.

4.2.1 GABA Parameterisation

The zwitterionic GABA ligand is not a standard residue included within the Amber forcefield library and so for this reason it needed to be parameterised. The protocol followed for the parameterisation was identical to that used for parameterising glycine, both because GABA is still a relatively simple ligand with similarities to glycine, but also to reduce any potential confounding effects of changing protocol between the two systems we wanted to compare. It is important to note that zwitterionic GABA is stable and therefore did not require the bond length constraints needed for glycine parameterisation. The geometry optimisation was carried out on the ligand structure at the density functional theory level using b3lyp exchange and correlation functional and the 6-31G* basis set with the Gaussian09 code[102]. Next, a single point energy calculation was carried out on the optimised structure at the Hartree Fock level with the same 6-31G* basis set, for compatibility with the Amber ff14SB forcefield. A RESP fitting was also carried out to calculate the partial charges. The atom types and charges produced from this process were then integrated into the system at the formatting and solvation step of PDB preparation for Amber forcefields with the Leap tool.

4.2.2 Simulation Details

From the work that published the 6PM5 glycine bound structure, there was also a GABA bound counterpart. This GABA bound structure (PDB 6PLX) was produced using the same methods as 6PM5, using a zebrafish alpha-1 glycine receptor, bound with GABA in a desensitised state and prepared with the same SMA copolymers[26]. The resolution was slightly higher, at 2.90 Å with similarly high validation scores of 0 Ramachandran outliers, 0 Sidechain outliers and a clash score of 6, as shown in Figure 4.3.C.

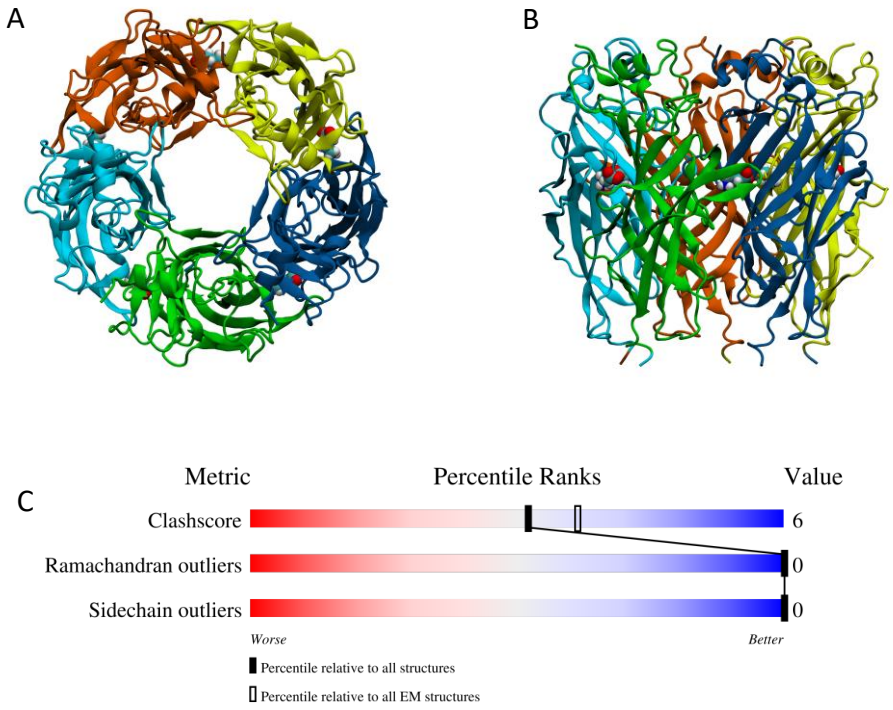

Figure 4.3: The ECD taken from the PDB data from a top-down view in A and a horizontal view in B. The validation metrics are shown in C.

The ECD was isolated in the same way, by taking the 6PLX PDB and removing all residues of the TMD and leaving the first 218 residues of each subunit intact, with the last five residues, including one residue of the M1 helix, serving as points for restraints to be applied for holding the structure together and imitating the anchoring effects of the TMD. The structure of the isolated ECD is illustrated in Figure 4.3.A-B. Using this PDB of the 6PLX ECD, the same procedure was followed as before to prepare the structure for use in molecular dynamics. The addition of hydrogen atoms was carried out at neutral pH using the H++ tool[150]. The structure was then solvated and formatted for use in NAMD 2.14[96], with the Amber forcefield ff14sb[89] and GAFF[165]. The protein solvation was carried out with the same parameters as for 6PM5, using the TIP3P water system and a 15 Å buffer including 0.15 M of Na+ and Cl- ions. The disulphide bonds were also re-checked to ensure consistency across the systems and literature[141,142]. Following this preparation, the final structure included a total of 122,261 atoms of protein, water, ions and the five GABA ligands.

The minimisation was then carried out, along with all subsequent steps, NAMD 2.14 was used for simulation. The periodically repeated orthorhombic box size was 115 Å x 115 Å x 110 Å and all other parameters including the thermostat and barostat were set up identically to the 6PM5 system. The simulation time step was 2 fs and bonds containing hydrogen were constrained using the SHAKE algorithm. Particle Mesh Ewald was implemented for electrostatic interactions using a 10 Å cutoff. Mirroring the glycine bound system, minimisation was carried out in 4 stages, starting with restraints on all components minus the solvent, then freeing of the side chains, the protein backbone and finally minimisation with all atoms unrestrained including the GABA ligand. Heating was then carried out using the same protocol as the 6PM5 system where incremental steps of 25 K were applied until 310 K was reached with the protein backbone under harmonic restraints of 25 Kcal/mol/Å^2.

Restraints [Kcal/molÅ^2]	Timepoint [ns]
25	0
25	50
10	65
1	80
0.4	95
0.04	110
0.016	125
0	140
0	190

Table 4.1: Table of the harmonic restraint rescaling protocol. The scaling rows provide the timescale across which the previous restraint was rescaled to the subsequent value.

Equilibration was carried out using the same protocol for a 190 ns period where the same harmonic restraints from heating were slowly scaled down from 25 Kcal/mol/Å^2 after 50 ns at small 15 ns increments for 150 ns (Table 4.1). The restraints on the last 5 residues of each subunit were again kept from scaling below 1 Kcal/mol/Å^2, the value which they would remain at for all further simulations. The production was carried out for a total of 1000 ns following equilibration, this being the simulation time used for carrying out the basic analysis of the system.

4.3 The Partial Agonist Bound GlyR system Results

4.3.1 The Structure and Stability

As previously carried out in Chapter 3, the stability of the system was measured through the production as a function of RMSD over time. The RMSD is calculated using cpptraj[122] and is measured across backbone atoms of the protein residues in relation to the starting frame of the simulation. The RMSD shows a gradual increase from the start of the production that stabilises at around 2.5 Å. An increasing RMSD or variable RMSD across the production would indicate instability in the system but what is shown in Figure 4.4 allows us to verify that the system developed does not have any large unexpected deviations from the equilibrated structure.

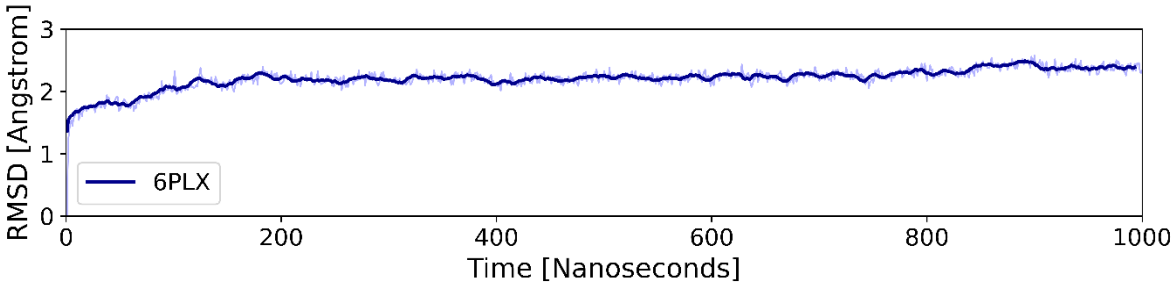

Figure 4.4: The RMSD of the protein backbone as a function of time during the production simulation. All plotted RMSD values are shown with the raw data shaded and the moving average taken every 10 frames shown in solid colour.

Whilst the RMSD of the whole protein is a useful metric for stability, it is also important to take a closer look at important structures as changes in specific regions may go unnoticed in measuring the RMSD across the full structure. Here we measure the RMSD averages across the production for specific sub structures including the individual subunits, the binding pockets, and Loop C (Figure 4.5).The RMSD of the individual subunits was also measured in relation to the first frame and the results show some discrepancies between the first (A) and fourth (D) subunit suggesting a degree of flexibility more in these chains than the others, although this difference is less than an Å. The RMSD of loop C is another important measurement that shows the flexibility of the loop of residues capping the binding pocket, large fluctuations in this structure could suggest that it is failing to cap the pocket and a vital element of GlyR binding suggested from structural data[16]. In this system however the RMSD of loop C fluctuates below 1 Å. The binding pocket residues were also grouped together per site and the RMSD across the backbone measured. The variation across binding sites with this metric was

again similar with a single exception in pocket B where the RMSD climbed around 0.5 Å compared to the other sites.

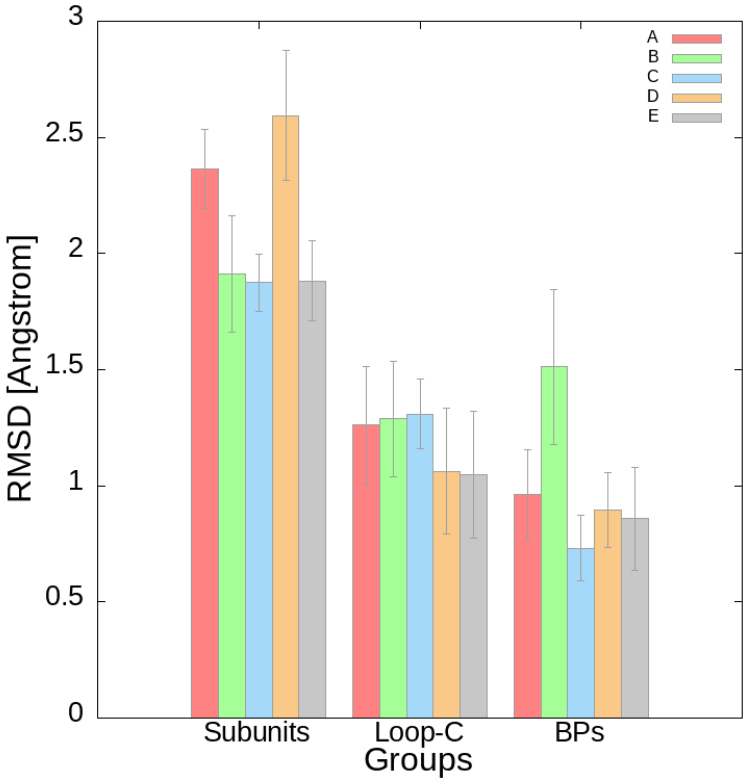

Figure 4.5: The average RMSD values for each substructure for each of the 5 pockets, the error bars here are representing the standard deviation measured across the production for each value recorded.

Whilst these averages do provide valuable indicators for instability and heightened flexibility, it is also useful to look at the changes over time to identify where and specifically when any differences might arise and whether the lack of uniformity in average RMSD values stems from individual spikes/events or a general increase in flexibility. For the individual subunits we can see that the increased average RMSD in Figure 4.6.A for subunit A and D is due to an overall increase in flexibility across the whole simulation, the difference however is so small at around half an Å it is unlikely to be problematic or indicate any issues with stability, especially due to how consistent this difference appears to be. The RMSD values over time for the Loop C structures in Figure 4.6.B give a strong indicator as to whether the cap is closed or open, more movement suggesting the latter. Here we can see that for the most part the 5 different pockets have a very similar degree of fluctuation between 0.5 Å and 1.5 Å. Finally, the binding pockets RMSD over time in Figure 4.6.C show an increased degree of movement for binding pocket B, where the value stabilises between 1.5 Å and 2 Å. This higher value of RMSD for binding pocket B could indicate some behaviour of loop-C that is unstable, making it a poor choice for further work using metadynamics.

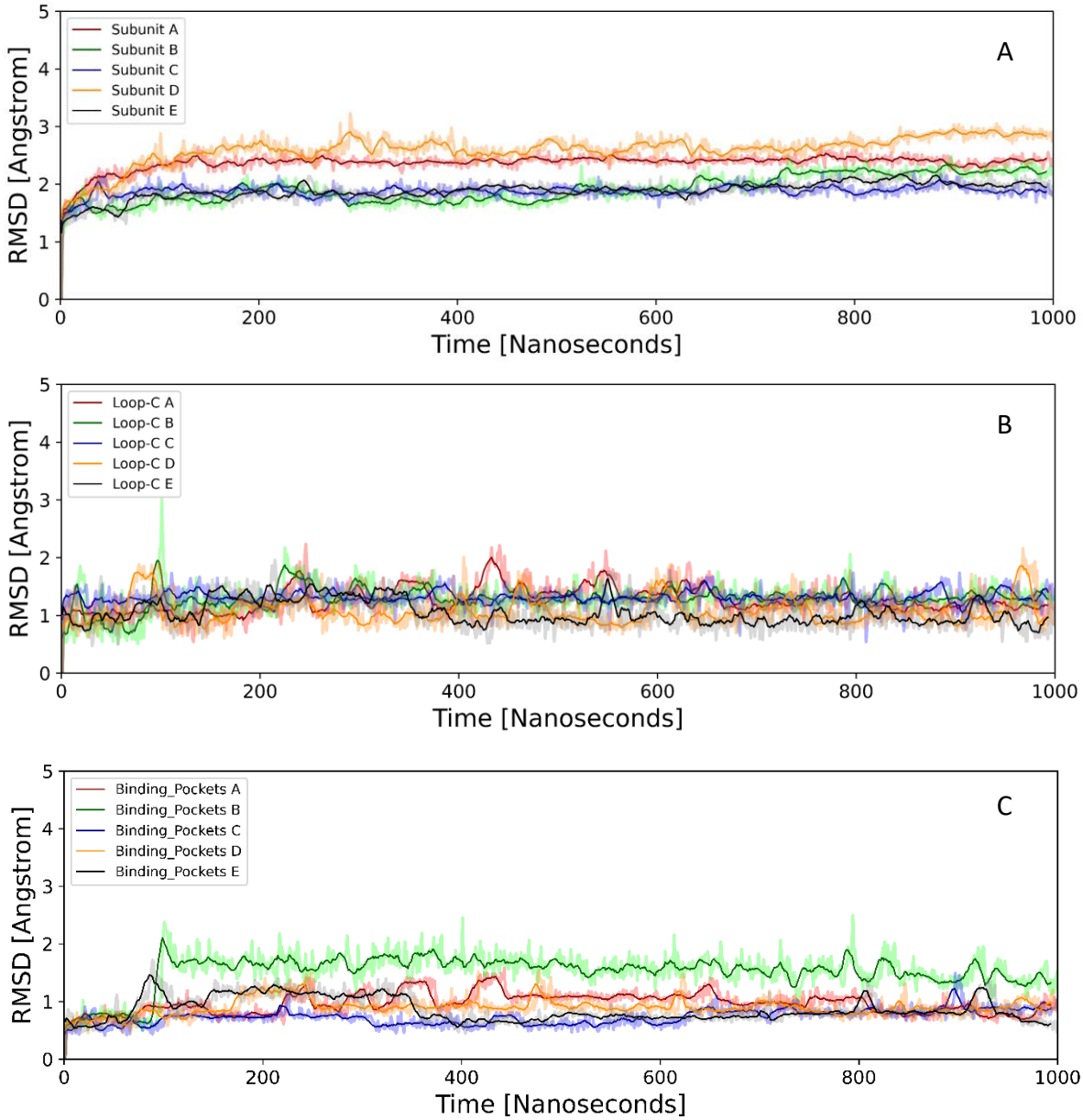

Figure 4.6: The RMSD of key substructures measured from the first frame of production including the subunits over the production in A, and the RMSD of the C-loops in B. C, shows the RMSD values of the binding pockets residues.

4.3.2 Hydrogen Bonding

To get a clearer understanding and verification that the system is accurately representing the binding pocket of GlyRs bound with GABA, the interactions taking place over the course of the simulation are important to measure. The same approach taken for the other systems was reproduced here to plot the hydrogen bonds that formed over time between the ligand and the key binding pocket residues.

The hydrogen bonds over time plot in Figure 4.7 shows that whilst most of the pockets are similar in terms of fluctuations and average, there is some reduction for binding pocket B where it more consistently seems to form only 2 or 3 hydrogen bonds compared to the other pockets that fluctuate around 4,5 and 6. From cryo-EM studies, there have been several key residues suggested to be integral for ligand binding of GlyRs including multiple serine and phenylalanine residues as well as threonine and arginine[16,26]. The five potential interactions identified from structural data aligns well with the system where A, C, D and E have a mean number of hydrogen bonds at 4.58 ± 1.13, 5.40 ± 0.84, 4.93 ± 1.03, 4.43 ± 1.25 respectively (with ± representing standard deviation), binding pocket B only has a mean number of hydrogen bonds of 3.18 ± 1.10. This suggests pocket B to be potentially weakly bound, an important result to consider going forward, especially when selecting a pocket to carry out metadynamics with.

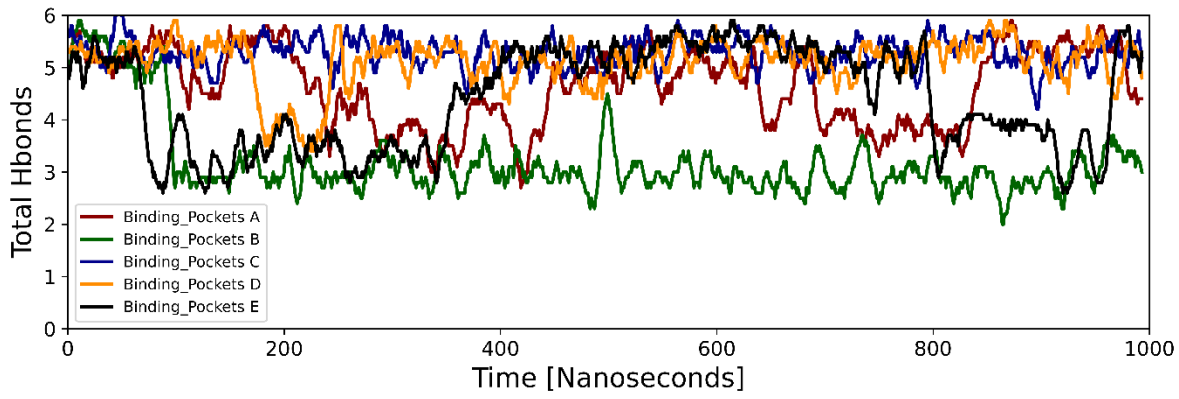

Figure 4.7: The total number of hydrogen bonds for each pocket over the course of the production, measuring interactions between the ligands and their respective key binding pocket residues.

Taking a closer look at one of the pockets, the well bound and stable pocket D, we can record the average occurrence and types of hydrogen bonds present through production. A snapshot of the pocket is shown alongside the plot of hydrogen bond types in Figure 4.8.A, over the page, where the orientation of GABA as well as its proximity to key binding residues aligns well with literature sources. The occurrence of interactions themselves shows that GABA has consistent bonding to ARG-65, SER-129 and THR-204 and to a lesser extent also SER-158 and GLU-157 (Figure 4.8.B).

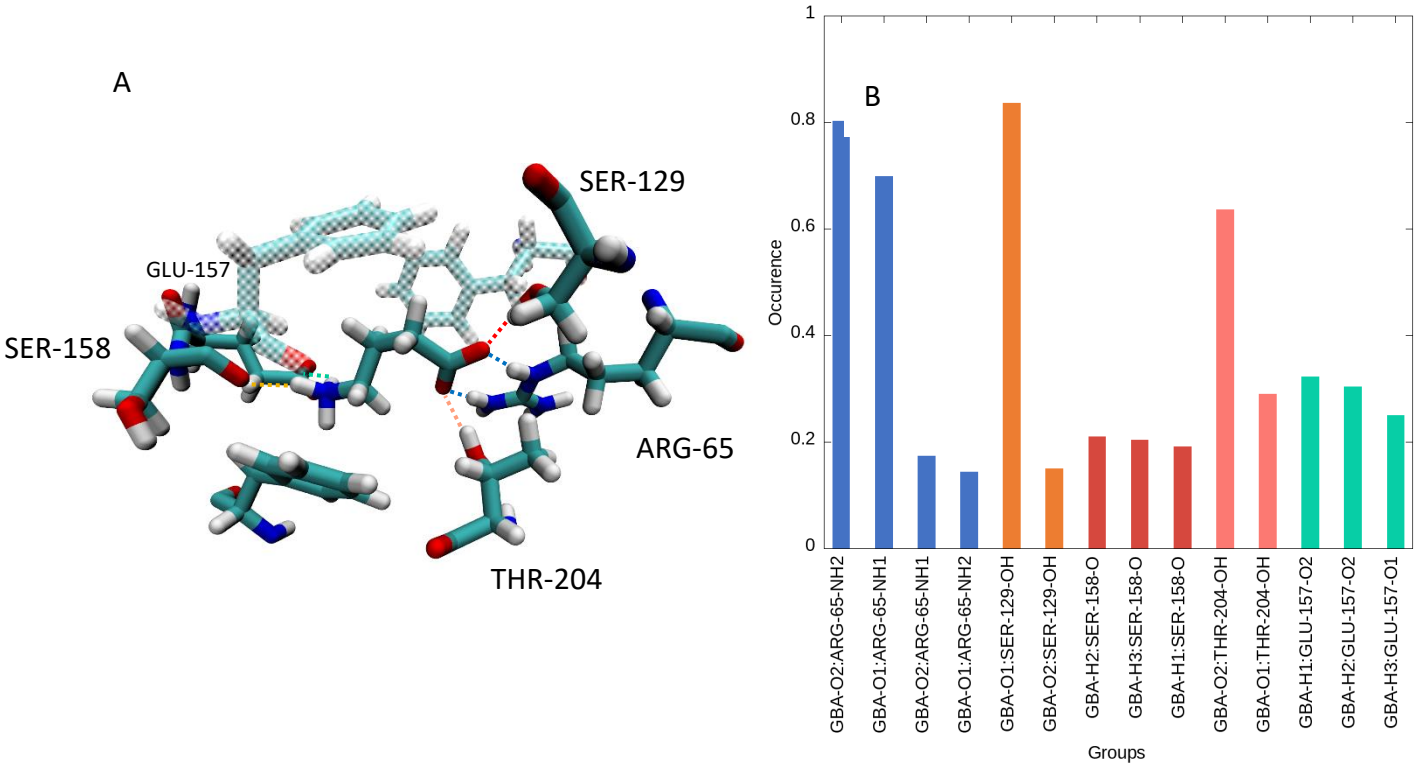

Figure 4.8: A, a snapshot of the pocket with the key binding pocket residues and the interactions drawn with dotted lines corresponding in colour to the plot in B, which illustrated the occurrence of the different hydrogen bonds present over the course of the production.

When comparing this to the glycine bound pocket in Figure 3.8, we see that the water in the pocket is completely removed with GABA binding. Instead of the amine moiety forming hydrogen bonds to a water molecule linking to the GLU-157 and SER-158, the GABA ligand forms hydrogen bonds directly with these residues. The same hydrogen bonds do form at the carboxyl moiety; however, instead of the occurrence being spread across the 2 possible combinations equally, there is a strong preference for GABA to stay in one arrangement with the O2 atom interacting with the NH2 group on ARG-65 and O1 forming a hydrogen bond with the NH group on ARG-65. This indicates that the carboxyl group is much more capable of rotating freely in the glycine bound pocket, whereas GABA appears much more fixed in its position. The same is seen in the interactions with SER-129 and THR-204 showing that GABA does not rotate freely in the pocket in the same way that glycine can do. This is likely due to the length of GABA and size of the binding pocket restricting torsional movement that glycine, with its smaller size, is free to do.

4.3.3 Cation-π Interactions

Similar to the hydrogen bond interactions, the cation- π show a similar profile for each pocket where an interaction is present between GABA and PHE-207 for most of the simulation and to a lesser extent PHE-159, shown in Figure 4.9. These interactions are consistent with literature sources that cite PHE-207 to be a key interacting residue for glycine. In the case of GABA, these residues have not been cited as interacting but in the case of the GABA binding to the the Acetylcholine binding protein (AChBP) that is structurally related to GlyRs, homology modelling pointed to two aromatic residues potentially interacting with the primary ammonium of GABA in a similar way to the interactions observed here[166]. From this work, we can conclude that these two aromatic groups may be serving a similar function in stabilising GABA within the pocket of the GlyR.

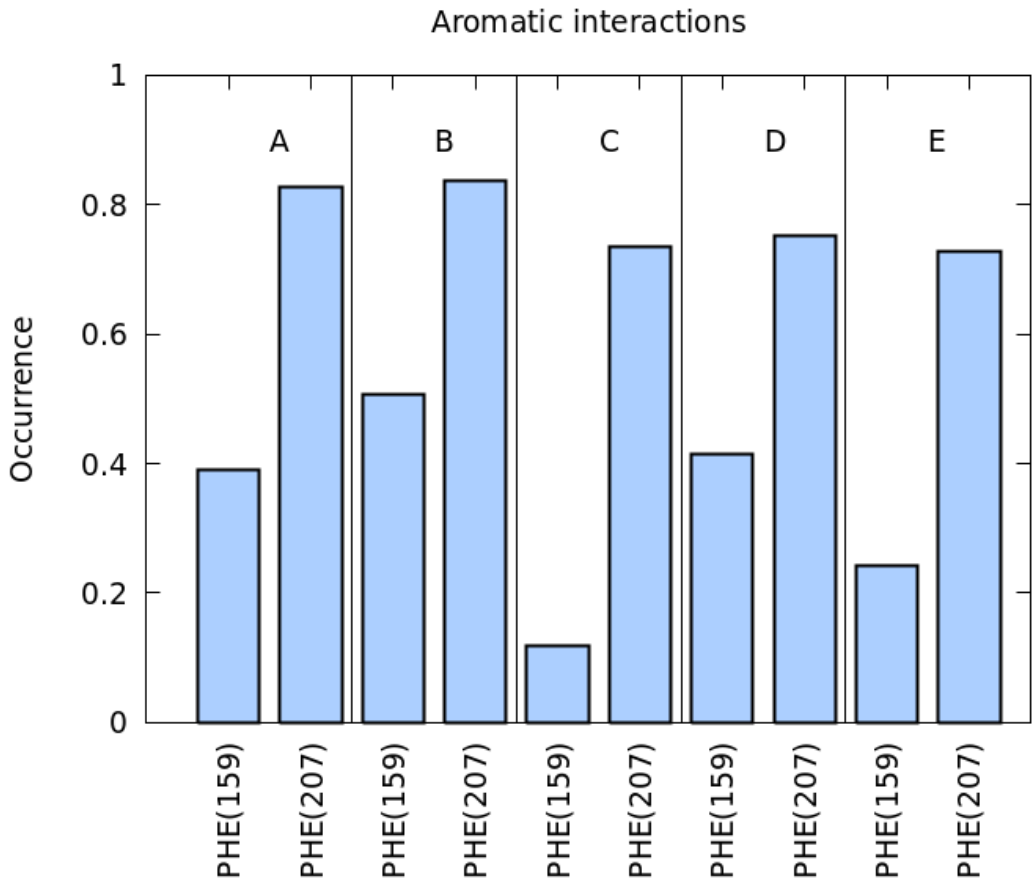

Figure 4.9: The RMSD of the protein backbone as a function of time during the production simulation.

4.3.4 The Binding Mode

Looking more specifically at the arrangement of the binding pocket and the orientation of the GABA ligand we can see that our system is in alignment the bound pose suggested in the cryo-EM data[26] in its original position with the carboxylate group on the ARG-65 side of the pocket and the amine moiety on the opposite side interacting with GLU-157 (Figure 4.8). Other features of the binding pocket that are important for ligand binding were also investigated including the positioning of Loop C. By measuring the distance between the tip of Loop C where the THR-220 residue is located and the centre of mass of the binding pocket we can see that the distance is mostly stable at around 6 Å for binding pockets B, C, D and E, as shown in Figure 4.10.A. However, there are some fluctuations that move this distance to around 8 Å for binding pocket A and transiently for pocket B and C. These movements although small indicate a degree of flexibility in the loop C structure that should otherwise be relatively fixed by interactions with the ligand. The instability seen in binding pocket A and B is likely a result of the reduced hydrogen bonding that we reported in Figure 4.7.A. Going beyond this, we can also measure the movement of the ligand within the binding pocket by recording the distance between the centre of mass of the ligand and the centre of mass of the key binding pocket residues (Figure 4.10).

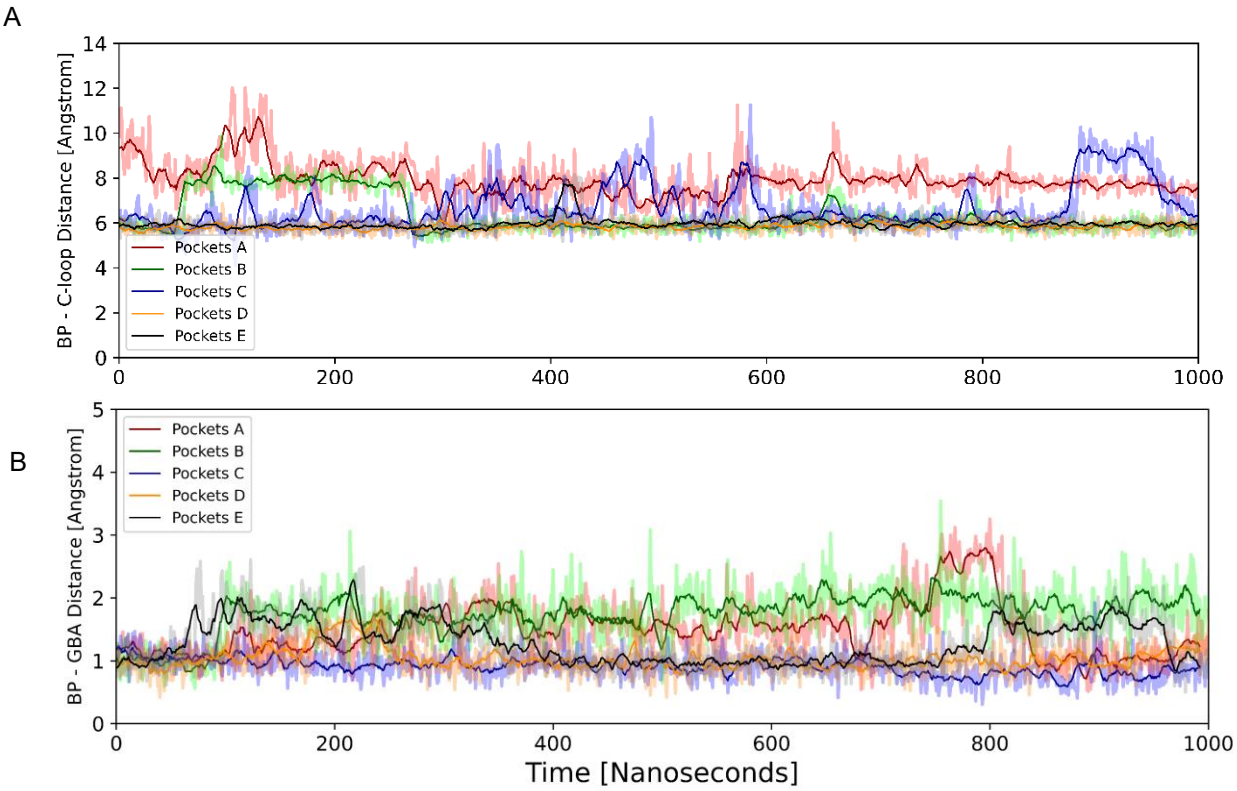

Figure 4.10: A, The distance between the tip of the C-loop centre of mass (THR-204) and the centre of mass of the binding pocket using the principal binding pocket residues. B, a measurement of the distance between the centre of mass of the ligand to the centre of mass of the binding pocket, each pocket plotted over the course of the production.

Doing this shows that whilst there is a degree of movement, the ligands in binding pocket C and D are mostly stable whereas the ligand in binding pocket A, B and E move up to 2-3 Å (Figure 4.10.B). Unlike the glycine bound system, there is no water in the pockets of the equilibrated structure with GABA bound. However, when investigating the pocket showing multiple elements of destabilisation, we observed a movement of water into pocket B, shown in Figure 4.11, shortly after the first spike in distance between the C-loop and the centre of the binding pocket.

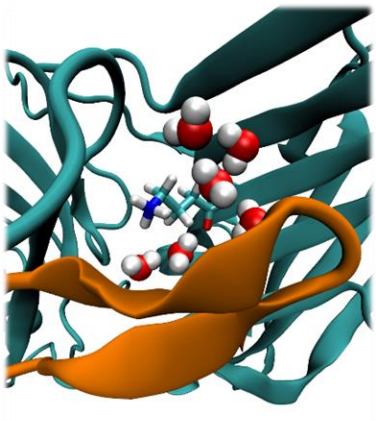

Figure 4.11: Depiction of binding pocket B, showing the positioning of the C-loop in orange, Ligand as well as the movement of water into the pocket.

The shift in the C-loop was sufficient to allow water molecules into the pocket and disrupt binding. The presence of water in this pocket aligns with the other results where as soon as water enters and the orientation of the ligand shifts, the total hydrogen bonds being formed between GABA and the GlyR drops to below 3 and never recovers back to the baseline level the other pockets share at around 5/6 hydrogen bonds at any given time.

The observation of this shift in water to the inside of the pocket explains what we see in the other results and allows us to identify this pocket as an unsuitable choice for further use given its instability and the likelihood of this to compromise any metadynamics simulations by providing a reduced bound/unbound energy barrier. Building on these results, measuring the binding pocket volume showed a similar pattern of destabilisation of pocket B when compared to the other pockets. We found that the stable well bound pockets demonstrated a binding pocket volume of around 145 $Å^3$ (Figure 4.12 on the next page), whereas the volume for pocket B fluctuated at much higher values. More specifically, the post equilibration state of the pocket did have a volume resembling stable binding but as soon as loop C moves and water enters the pocket, the volume drastically spikes to a peak of over 300 $Å^3$ and does not return to the baseline volume seen in more stable pockets.

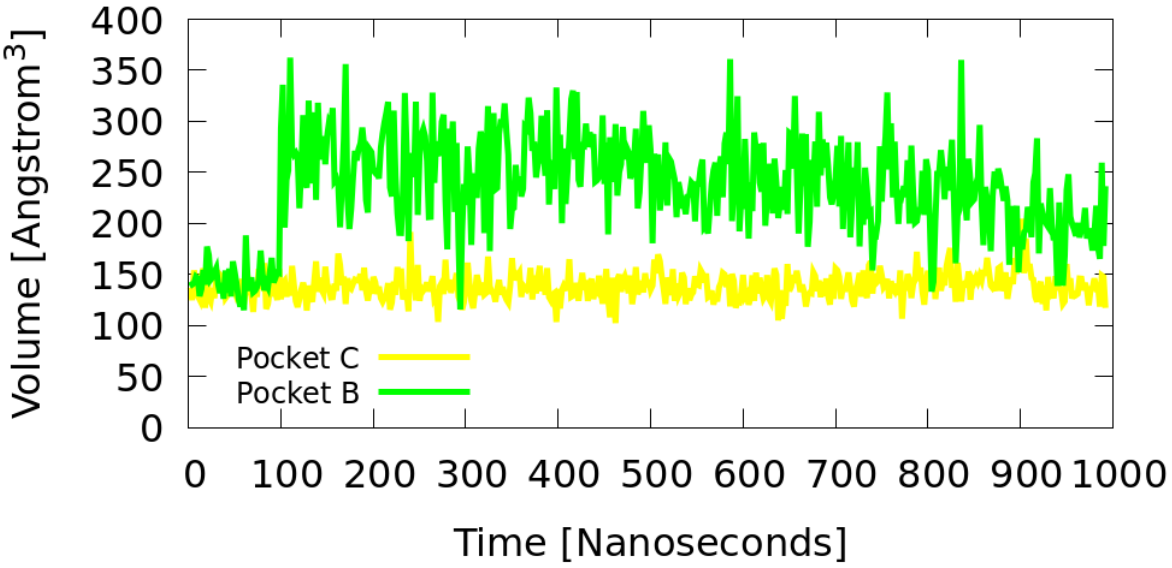

Figure 4.12: The volume of binding pockets B and C plotted over the course of the production, evaluated with the mdpocket tool[153].

4.3.5 Comparing The Glycine and GABA bound structures

Understanding how the structural changes brought about by the differing ligands leads to the functional changes observed is key to getting a clear picture of how full and partial agonism works. The main chemical moieties of GABA and glycine are the same, including an ammonium and carboxylate, meaning that the polar salt bridge across the pocket through the ligand from arginine to glutamic acid is maintained. The critical difference being that GABA is extended further, with a longer carbon chain, this is clear in Figure 4.13.A-B on the next page.

The structural differences are minimal in the whole structure, highlighted in figure 4.13.C, however a clearer picture can be seen of the binding pocket residues in figure 4.13.D where a shift in the arginine, threonine, phenylalanine and glutamic acid can be seen for the GABA structure. The similarity in structure makes sense when considering the key interactions for glycine binding, i.e. the salt bridge forming from the charged amino acids either side of the ligand being retained in both glycine and GABA bound structures. Without the ligand retaining these key similarities it's unlikely it'd bind as effectively as it does. Ultimately, the differences in the ligand leads to differences in the shape of the pocket when the ligands are bound.

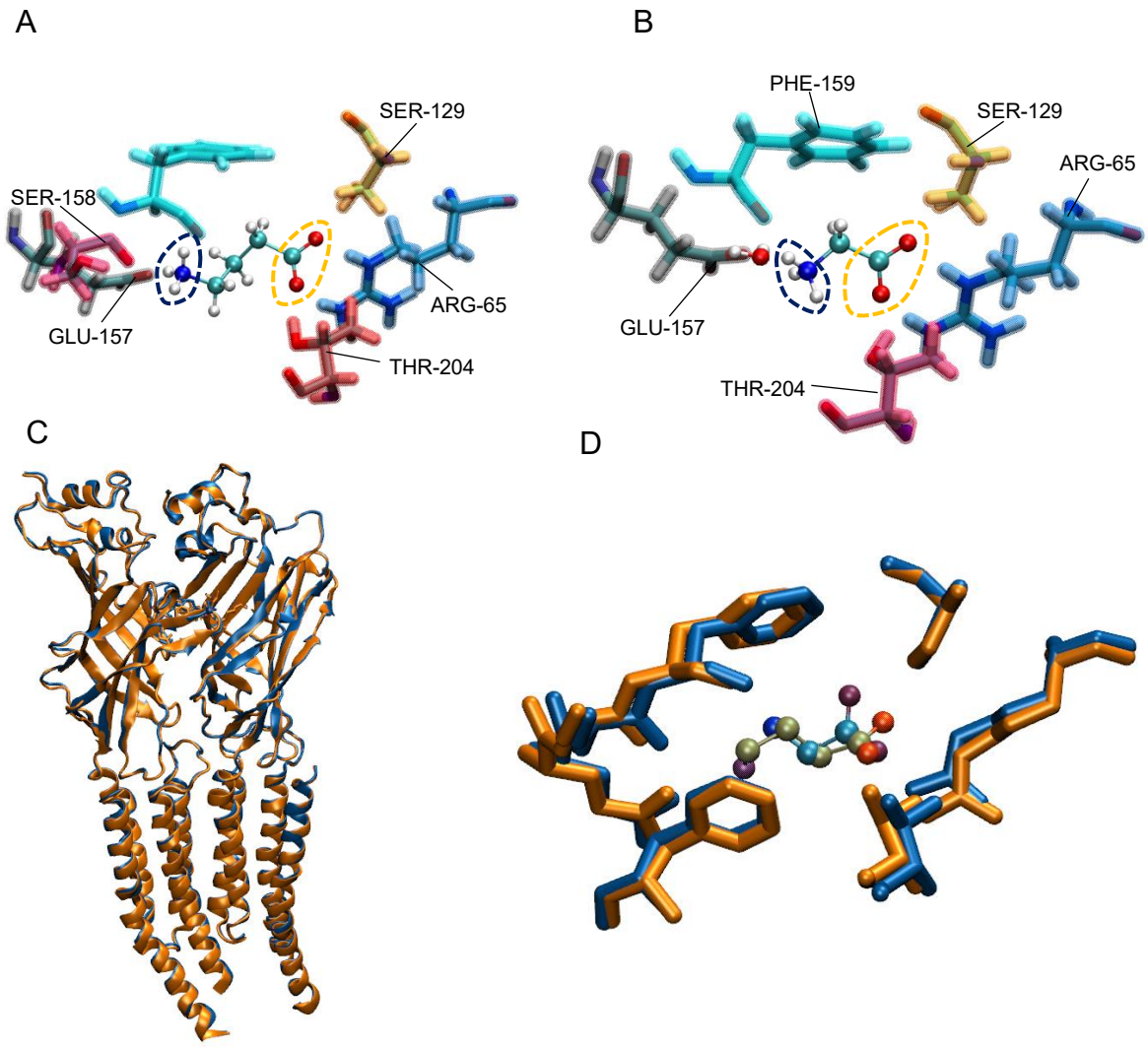

Figure 4.13: A snapshot of each pocket, demonstrating the layouts of each ligand, GABA in A and glycine in B. The ammonium moieties are highlighted by the dark blue dotted line and the carboxylate an orange dotted line. The water molecule can be seen within the glycine bound pocket on the right, between the ammonium moiety and the glutamic acid residue. The noticeable lack of differences in general structure can be seen in C, where the two subunits across a pocket are shown. The pocket itself and the shift in structure is more clear in D. The GABA bound structure is coloured orange, glycine blue, for the figures in C and D.

The general shape of the binding pocket with glycine bound and GABA bound is illustrated in figure 4.13, demonstrating how the pocket is deformed. The ammonium moiety of the GABA is placed further into the glutamic acid side of the pocket, displacing the water and changing the pocket shape. This effect is exemplified by measurement of volume that show the glycine bound pocket to be more constricted for the entirety of the production molecular dynamics simulations as is illustrated in figure 4.14, on the following page.

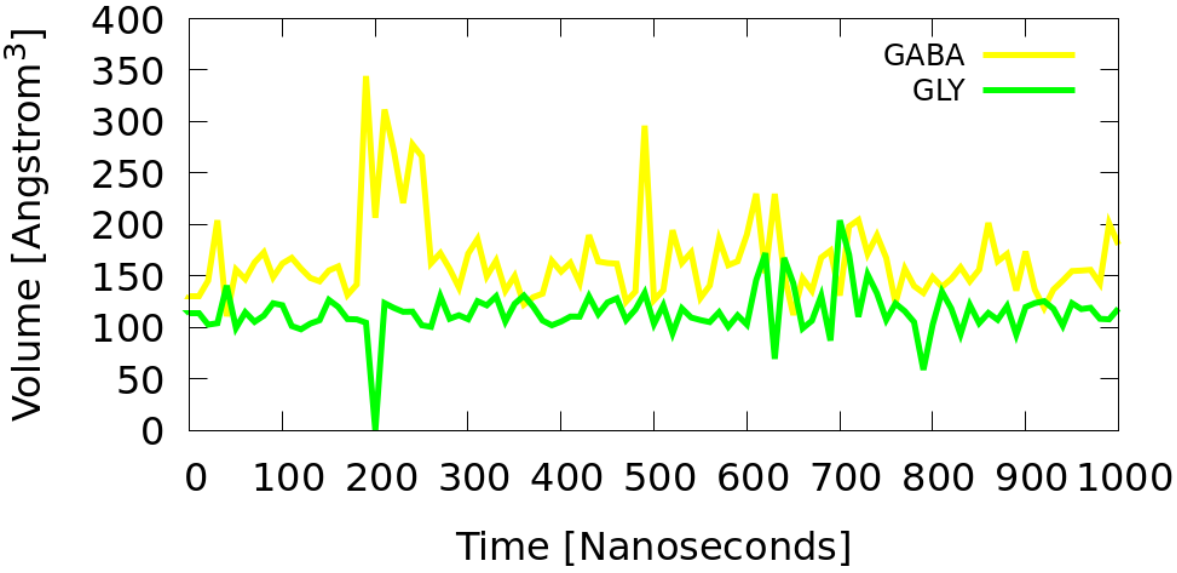

Figure 4.14: The volume measurements within the binding pocket for the GABA bound structure (yellow) and the glycine bound structure (Green). The volume was measured using the Mdpocket tool.[153]

Here in Figure 4.14, we see how the volume fluctuates at a higher value for the GABA bound structure, the larger ligand leading to a less stable, more expanded pocket. The effect on hydrogen bonding and cation-pi interactions seems minimal, the biggest difference comes from the less flexible nature of the interactions, with specific interactions taking priority in the GABA system whilst the glycine system appears more flexible.

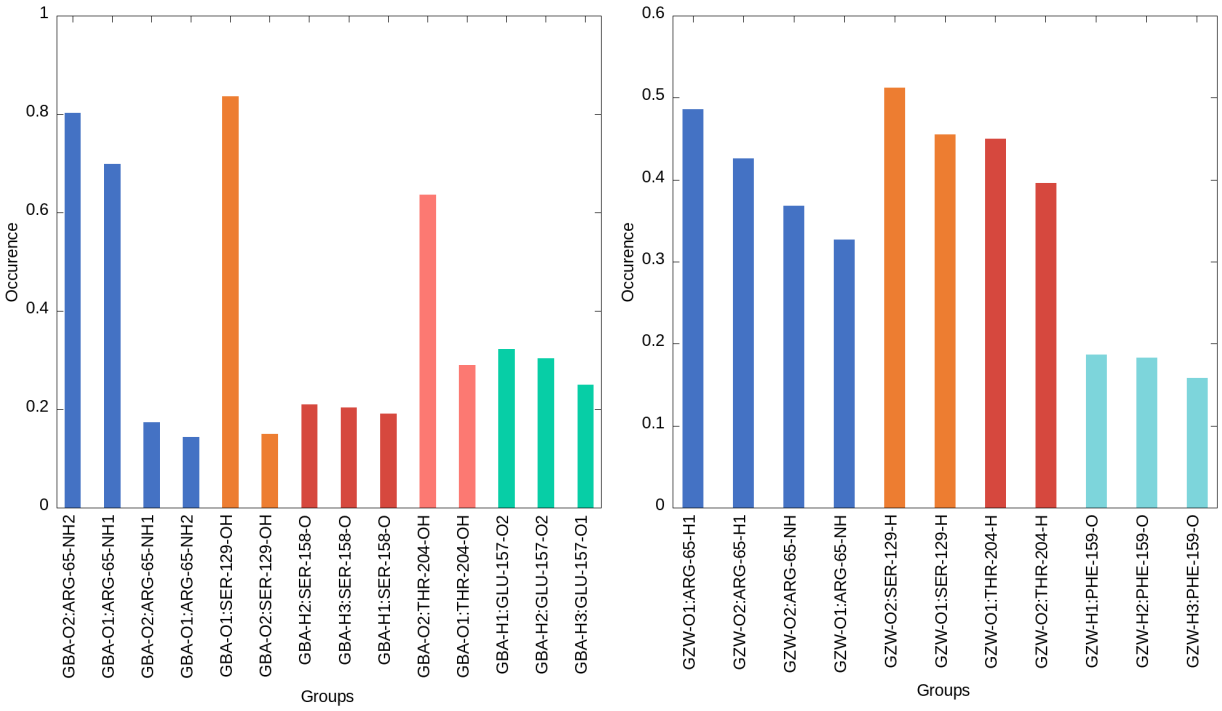

Figure 4.15: The hydrogen bond occurrence plots for GABA (Left) and glycine (right) shown alongside each other to highlight the relative differences. The types of hydrogen bonds are presented on the x axis and the occurrence over the course of the production is shown on the y axis.

4.3.6 Conclusion

The system discussed here has been shown to be a stable structure with 4 of the 5 binding pockets meeting our criteria. Binding pocket A, C, D and E all show the characteristic binding profile, stability and loop C conformation that would be expected given structural data suggestions and are all suitable for further use with enhanced sampling methods and for modelling a partial agonist bound to the GlyR. As well as this, we have been able to highlight some aromatic residues that interact with the ligand that have not been previously well documented, specifically PHE-207 and to a lesser extent PHE-159.

In the case of binding pocket B, due to the pocket starting equilibrated and in the same state as others, we hypothesise that the end state that the ligand reaches is likely due to the process of ejection starting. With the system confirmed as stable and representative of an expected GABA bound GlyR, we can move forward and use stable pocket D to system ligand binding/unbinding with enhanced sampling methods.

4.4 Funnel Metadynamics

4.4.1 Exploratory Metadynamics

To appropriately implement funnel metadynamics, the general path or area taken for the ligand to exit the pocket is required. To do this a series of exploratory metadynamics simulations were carried out to determine the area that would need to be within the funnel cone and the size required to avoid any interference of the boundary restraints confounding the binding path results. Gaussians were deposited every 1 ps along the trajectory with a height of 1.2 Kj/mol with a width of 0.2 Å, targeting one CV: the distance between the centre of mass of the ligand and the centre of mass of the key binding residues. Previous work on similar systems has determined this to be an effective method for funnel placement determination, and a valid approach to observing the slow movements involved in unbinding as, with a sufficient number of replicas used, the general direction and space used for unbinding can be highlighted as is shown in Figure 4.13. From this information, the funnel cone region can be placed so that none of the paths highlighted below would interact with the wall of the funnel whilst still engaged with the binding pocket residues or other key components. This is an important step because the key findings sought from the process includes the free energy of the system within any binding or pre-binding regions of conformational space; if the ligand has forces

exerted on it from the funnel walls in one of these sites, it will drastically alter the outcome and invalidate any result.

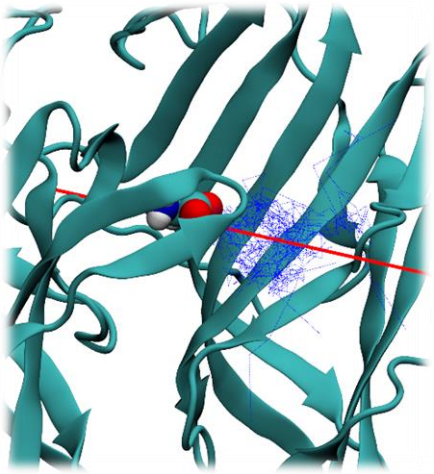

Figure 4.13: The exploratory metadynamics trajectories used to find a general direction for unbinding, the data here illustrates the many paths the ligand took overlayed with the structure of the protein. The dashed blue lines represent the trajectories of each of the 15 replicas used and the red solid line shows an average of these trajectories combined.

4.4.2 Simulation Details

The funnel metadynamics protocol followed utilised a tool called FMAP GUI[120] that involves both a pre-processing and post-processing phase before and after the simulation to reliably design and analyse the funnel metadynamics simulation. The first step of this protocol was to optimise the funnel potential placement onto the binding pocket and to define the parameters of both the funnel restraint properties and the metadynamics implementation itself. The axis of the funnel is set via two XYZ coordinates within the cell box. One point is chosen at the centre of the binding pocket, defining zero on the Z axis of the funnel. The second point chosen determines the general direction of approach the funnel takes; this point was picked to align the axis with the direction of most of the exploratory metadynamics ligand trajectories. Following this, the properties of the cone and cylinder were defined, again considering the exploratory simulation trajectories. The funnel is defined by the values outlined in the methods chapter and illustrated in Figure 4.14, with Zcc being the switching point for cylinder to change to funnel, α to determine the angle of the funnel cone and Rcyl to set the radius of the cylindrical region.

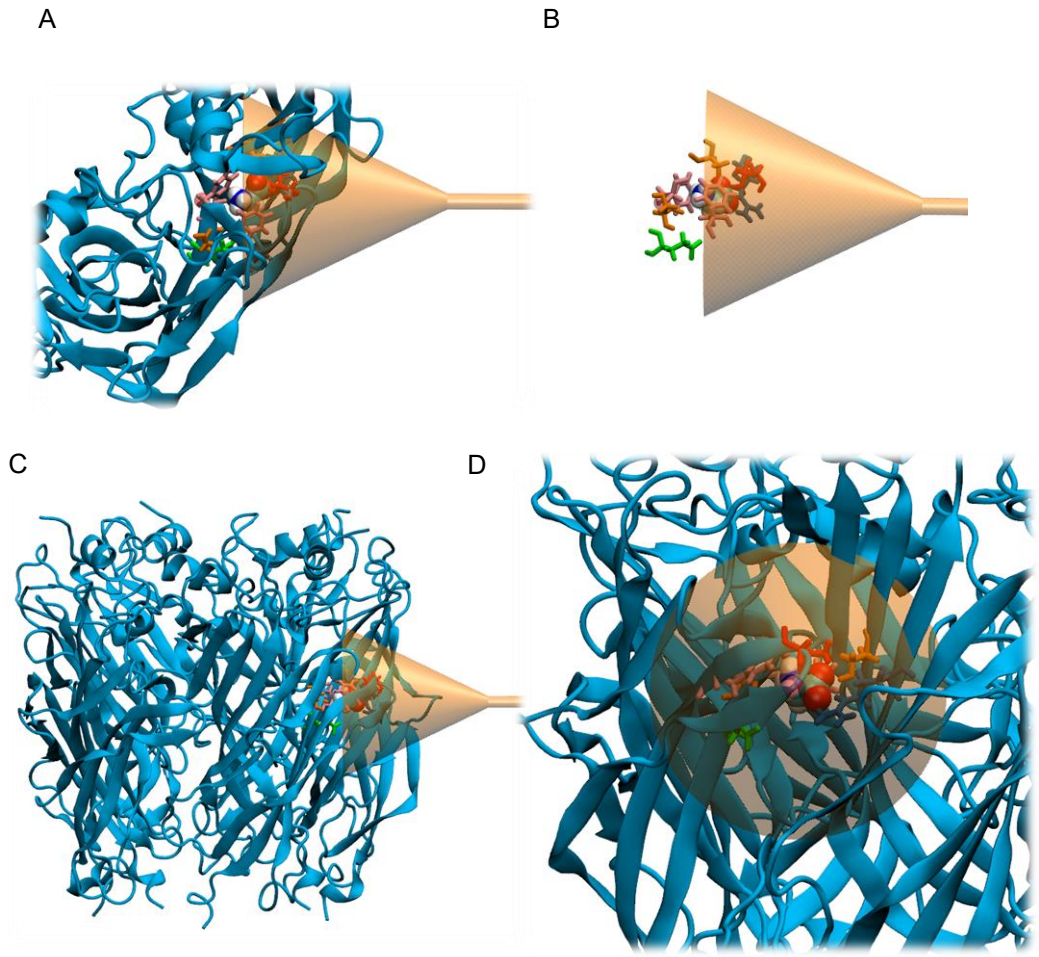

Figure 4.14: The 6PM5 ECD structure illustrated alongside the funnel representation. The view of the pocket and funnel from the top view is shown in A, the side on view in C and the view running along the funnel projected axis out of the pocket in D. Visualising of the pocket orientation is aided by B showing the principal residues alone with the funnel, colour coded for arginine (blue), phenylalanine (pink), threonine (red) and serine (orange) and glutamic acid (green).

By following the protocol set out in Limongelli 2020[120], using previous literature as a guide[46], carrying out test runs of the funnel metadynamics and also taking into account the exploratory metadynamics, we determined the suitable values of these parameters so as to avoid any interference from the funnel restraint. Zcc was set at 25 Å, α was chosen to be 0.45 rad, and Rcyl was set to 1 Å. To ensure full exploration of the pocket but limit any possible movements deeper into the protein, a harmonic restraining wall was set up at –5 Å. To measure the ligand binding free energy, the difference between the well in the binding site needs to be compared to a site in the solvent that is completely unimpeded by ligand-protein interactions. To guarantee that a region fitting this criterion will be explored by the ligand, the cylindrical region was designed to be 10 Å in length, projecting out into the solvent to an area where there would be no interactions with the protein. The cylinder length was also limited with a harmonic restraint on the upper bounds of the z axis at 35 Å. This whole process was carried out while keeping the trajectories of the exploratory metadynamics within the cone area, resulting in a

funnel shape that would allow the ligand to explore the most energetically favourable whole unbinding/binding path without touching the funnel restraint wall. The final funnel restraint shape and placement is illustrated figure 4.14, with the 6PM5 structure shown. Here the funnel can be seen placed ajar from the pocket, with the cone directed at the centre of the pocket between the two subunits that form it. Once the funnel was positioned, well-tempered metadynamics was carried out using the position of the ligand centre of mass along the funnel axis (fps.lp) and the distance from it (fps.ld) as collective variables. These CVs were selected based on the findings of the exploratory metadynamics and previous simulations that have studied zwitterionic ligand bound systems in this way, where these two axes as CVs can be applied to permit full exploration of the funnel space. The metadynamics parameters used were similar to the exploratory metadynamics, with gaussians deposited every 1 ps with an initial height of 1.2 kJ/mol, a width of 0.2 Å. The temperature was set at 310 K, the bias factor chosen was 15.

4.4.3 Comparing Full and Partial Agonism

In this section, the free energy landscapes from the system of the full agonist bound (Glycine) and partial agonist bound (GABA) receptor will be presented, along with other data collected during the funnel metadynamics simulation. We identify the primary binding mode and show the possible pre-binding poses that occur as well as provide a detailed description of the path the ligands take out of the pocket. More specifically, we demonstrate the stability of the system over the course of the funnel metadynamics simulation, then explore the free energy surfaces derived and the detailed mechanics involved in the ligand moving in and out of the pocket.

4.4.4 Stability + Flexibility

Before discussing the free energy landscapes, it is important to clarify the checks carried out to ensure the results presented here are reliable. There are two important aspects in particular, the stability of the structure and the convergence of the metadynamics simulation. For the stability, a common choice of metric is the measurement of the protein backbone RMSD over the course of the simulation and to ensure that there are no large drifts or movements that could distort the structure as a result of applying bias too aggressively.

The first step in ensuring that the simulations have been run adequately was to check the RMSD in a similar way that was done with the MD production of the systems. The RMSD of the protein backbone was measured across the trajectory for both systems. Whilst some drift

and instability would be expected, the important thing was that the RMSD during the metadynamics simulations was stable and similar to what was seen in the MD production. Any large differences could have indicated that metadynamics was inducing structural changes not intended or predicted. The RMSD of the whole protein shown in Figure 4.15A uses the first frame of funnel metadynamics for reference, for both the partial agonist (6PLX) system and full agonist (6PM5) system, and mirrors what was seen in the production with little difference.

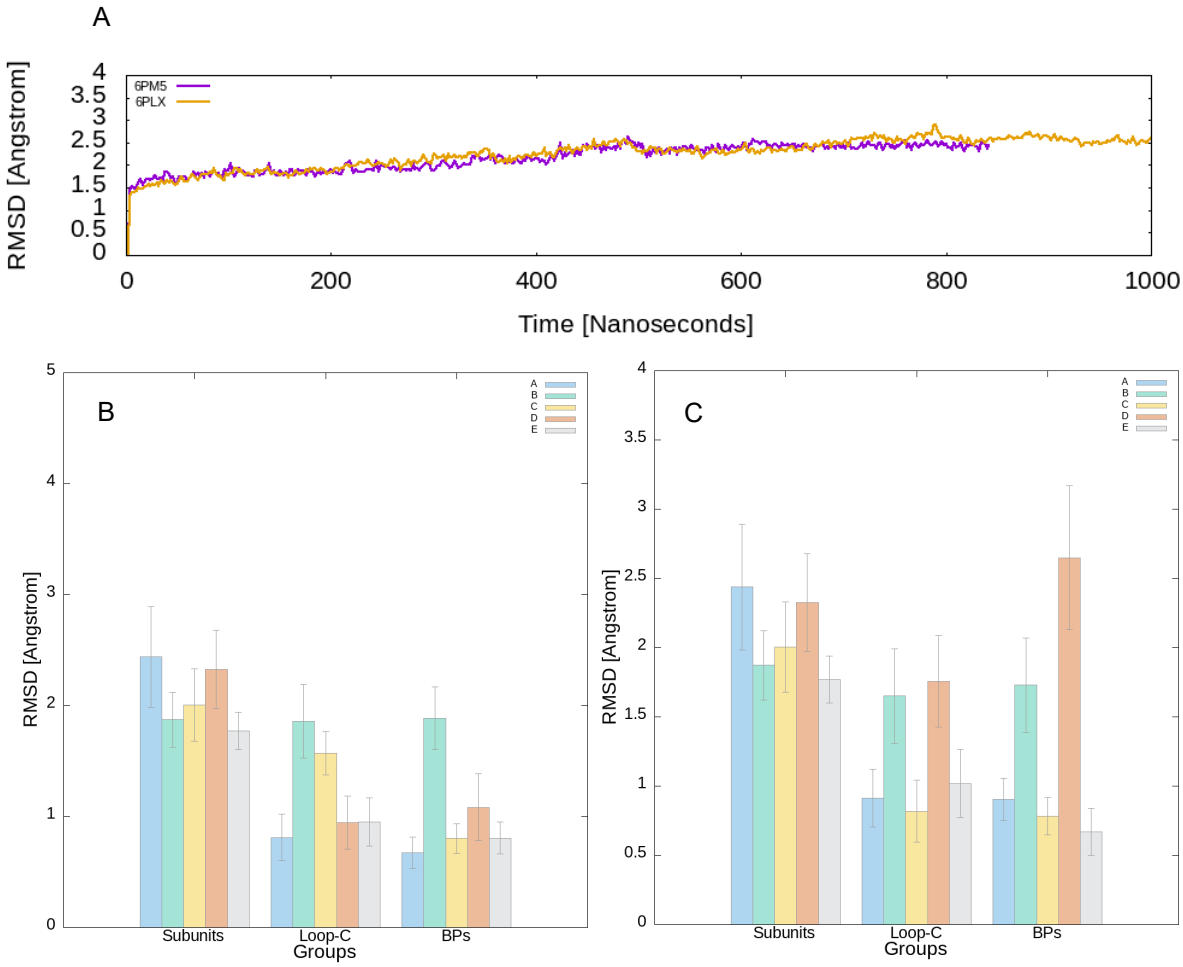

Figure 4.15: A, The RMSD of the protein backbone as a function of time during the whole funnel metadynamics simulation with both the 6PM5 and 6PLX models plotted. The average RMSD values for the important substructures identified, where plot B represents the 6PM5 system values and C shows the values measured for the 6PLX model.

Looking at the sub-structures within the systems can give us a more detailed check of the structural stability. Throughout the funnel metadynamics, we expected to see a greater amount of fluctuation and movement of atoms from the starting frame in the pockets targeted by the metadynamics as the movement of the ligands requires protein residue movement to be facilitated. However, what can be seen in Figure 4.15.B-C is that most of the average RMSD

values stays very close to what was seen in the production, particularly for the glycine bound structure. In the GABA system BP-D was used for funnel metadynamics and does show some changes in the RMSD of the Binding pockets and loop C. The increase seen in the GABA system RMSD values and very little change in BP C for the glycine bound system is likely related to the size of GABA, disrupting the position of residues during ejection much more, this then translates to a marked increase in the plots here.

To monitor individual parts of the structure, we can measure the RMSF of the protein backbone to see how much each residue moves in relation to the first frame of the simulation. Comparing the partial agonist and full agonist system can help identify key regions of interest that may be involved in different ways, but also any excessive movements that may point to instability in the systems. In Figure 4.16, we can see that, for many residues, there is little deviation between the systems, which is what we would expect for most residues not involved in the binding site. Small differences between the systems are apparent in the regions of residue 40 to 50 and 95 to 110, however these differences are around or below half an Å. The main conclusion from this RMSF plot is that the general profile of RMSF is consistent and any differences that are visible are too small to be impactful and are also outside of the binding pocket.

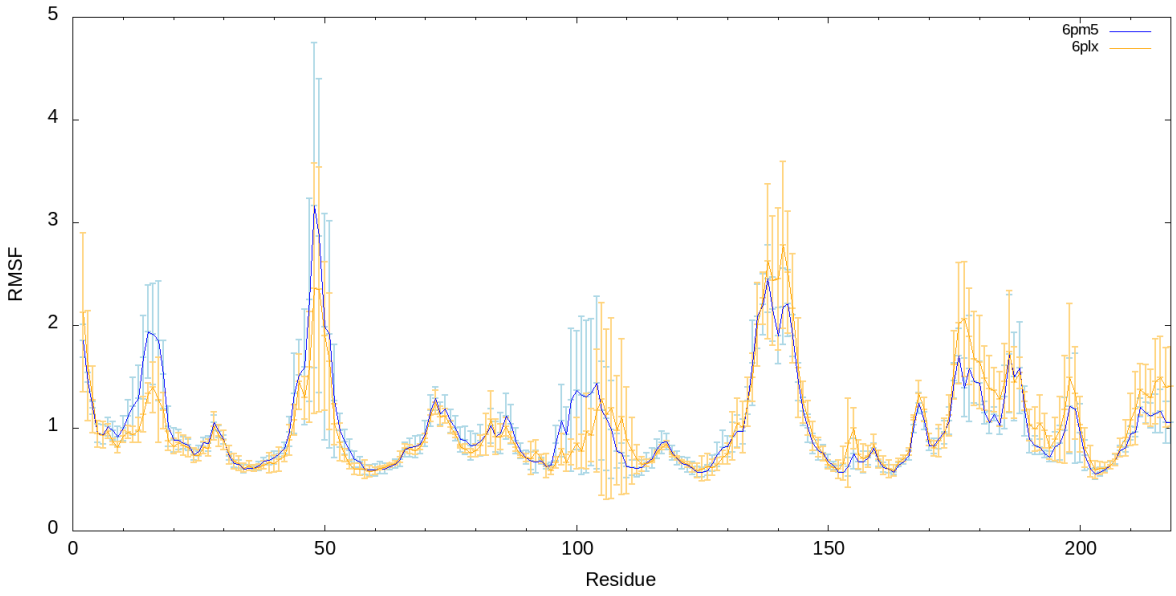

Figure 4.16: The RMSF values for the backbone atoms of residues in the 6PM5 system (blue) and the 6PLX system (orange), averaged across the subunits. The data collected here is taken from the funnel metadynamics simulation.

As for convergence there are two important aspects to consider, the movement of the ligand crossing between the bound/unbound states and the bound free energy reaching a constant. Here we show plots of the binding free energy difference between the bound and unbound

states over the last 100 ns of the funnel metadynamics along with the CV fps.lp plotted over time in figure 4.17.A-B. The plot binding free energy difference demonstrates in both systems how the value reaches a converged point, fluctuating around a specific end point over the last section of the simulation (Figure 4.17.C-D). Alongside this is a plot of the CV fps.lp across the funnel metadynamics where the bound and unbound states have been visited many times (Figure 4.17.A-B).

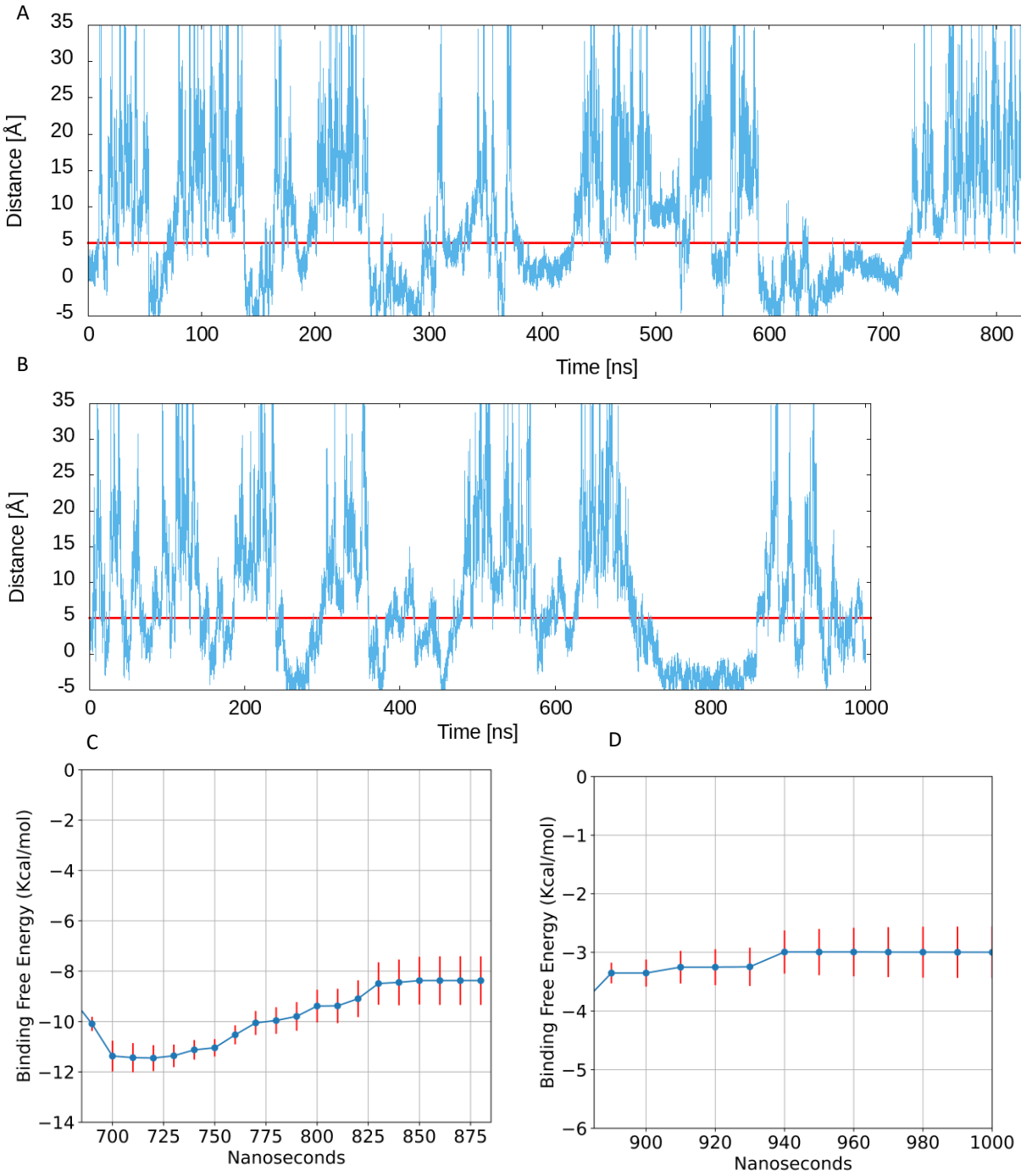

Figure 4.17: A, the CV fps.lp plotted along the funnel metadynamics trajectory of the 6PM5 model. B, the CV fps.lp plotted along the trajectory for the 6PLX model. The red line represents the point at which the ligand can be considered either unbound or bound. The binding free energy difference plots covering the last few hundred nanoseconds of the simulation with standard error plotted in red, 6PM5 in C and 6PLX in D.

4.4.5 The Free Energy Surface

The free energy surfaces shown in Figure 4.18 illustrate the key differences between full agonist and partial agonist binding.

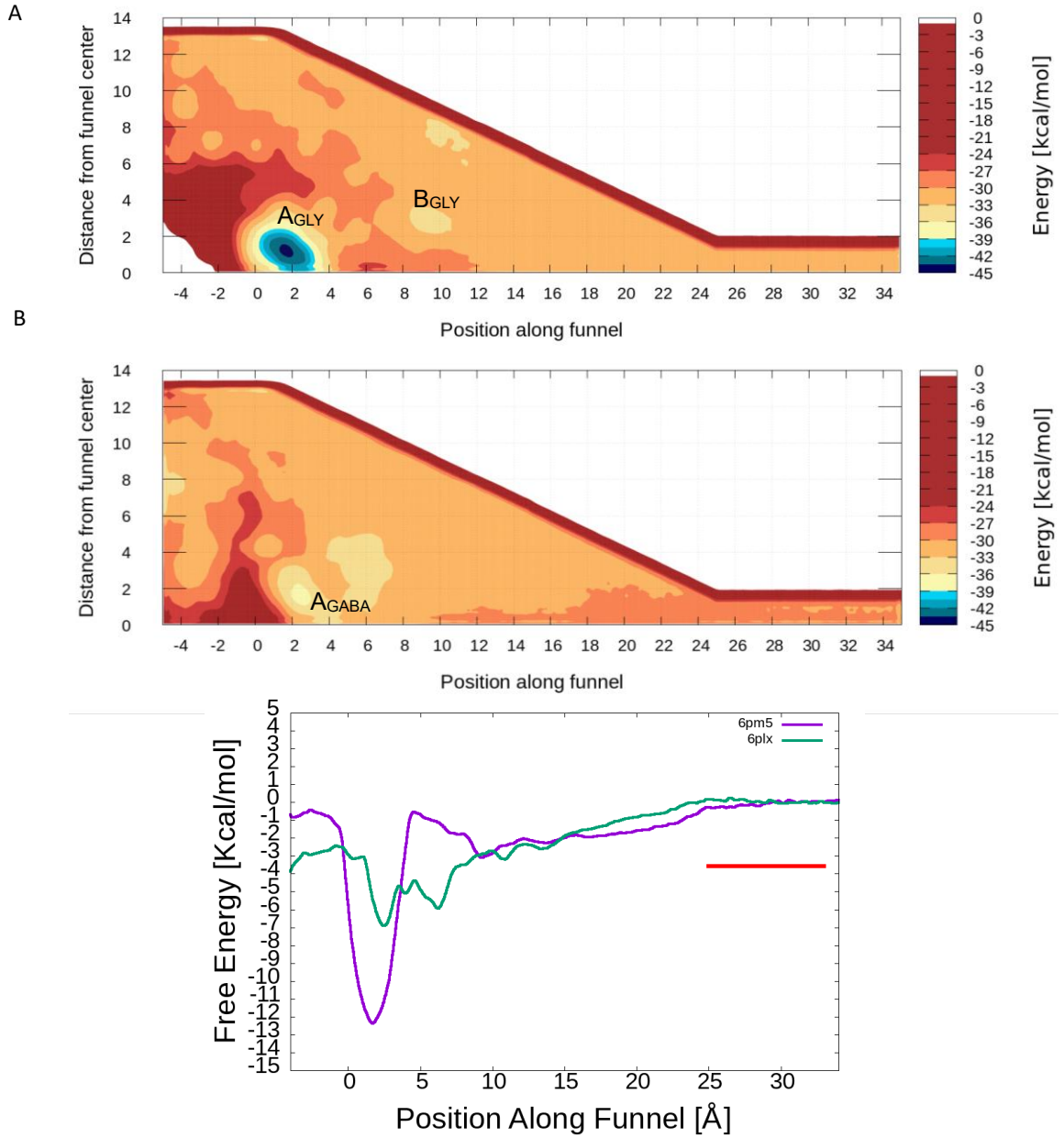

Figure 4.18: A, The free energy surface of the glycine bound 6PM5 system as a function of the CVs. B, The free energy surface of the GABA bound 6PLX system as a function of the CVs. C, The free energy profile projected along the funnel axis for both the 6PLX and 6PM5 models. The red bar illustrates where the cylindrical correction is applied. (needs rechecking as it seems a bit large)

Here the differences are clear within the free energy maps as a function of the CVs, with contour lines representing 2 Kcal/mol differences. In the case of GABA we can see much

lower energy barriers between the bound and unbound region, the bound region indicated by A_{GABA}. Glycine, however, binds with one strong well at A_{GLY} and a potential pre-binding pose at well B_{GLY}. In the 2-dimensional free energy plot, in Figure 4.18.C, of the CV along the Z axis of the funnel, this shows how the free energy surface changes as the ligand progresses further from the pocket. In this plot the energy barriers between bound and unbound states can be clearly defined, whilst GABA has much weaker energy barriers. In the glycine system there is a relatively stronger prebinding region where the entrance and exit from the pocket is hindered by interactions forming an additional well outside the Loop-C, the path which will be discussed in the next section of this chapter. From this data the ligand binding energy can be calculated between the bound well and the unbound region (Z > 25) finding the full agonist system to be -8.43 ± 0.93 Kcal/mol after the correction is applied and -3.01 ± 1.33 Kcal/mol in the case of the partial agonist. These initial findings were expected as GABA is a much less potent agonist of GlyRs with an apparent affinity (EC50) between 8 and 120 mM compared to that of glycine between 37 and 360 μM[167].To compensate for the interference of the of funnel restraint on the cylindrical section of the free energy surface a correction has been applied that amounts to -3.57 Kcal/mol.

4.4.6 Reweighting the Free Energy Surface

The choice of CVs does limit the visualisation of the free energy surface as the distance from the funnel axis compiles all sections of the cone radiating from the centre into one free energy surface. To gather further structural information, this free energy surface can be remapped as a function of spatial coordinates to get a clearer view of the binding pocket and path. This is done using the reweighting algorithm discussed in Chapter 2.

The reweighting procedure was applied to these two systems to produce the maps in Figure 4.19, which provide a more thorough picture of how the ligand explored conformational space. The reweighting process remaps the free energy surface produced with the fps.lp and fps.ld onto cartesian coordinates of the ligands position in in space p.x p.y p.z, allowing visualisation of all space explored by the ligand rather than the reduced dimensions seen in Figure 4.18. Here we show GABA binding in the 6PLX system on the left and glycine binding in the 6PM5 system on the right. The top two plots show a cross section vertically through the funnel space along the axis and the lower plots show the funnel from a horizontal plane cross-section. In the case of glycine binding we see that the reweighting gives a clearer view of the pre-binding region outside the pocket that was partially obscured by asymmetry in the free energy surface plot of fps.lp and fps.ld. This shallow well lines up with what we see when plotting the free energy of fps.lp alone in Figure 4.18.C.

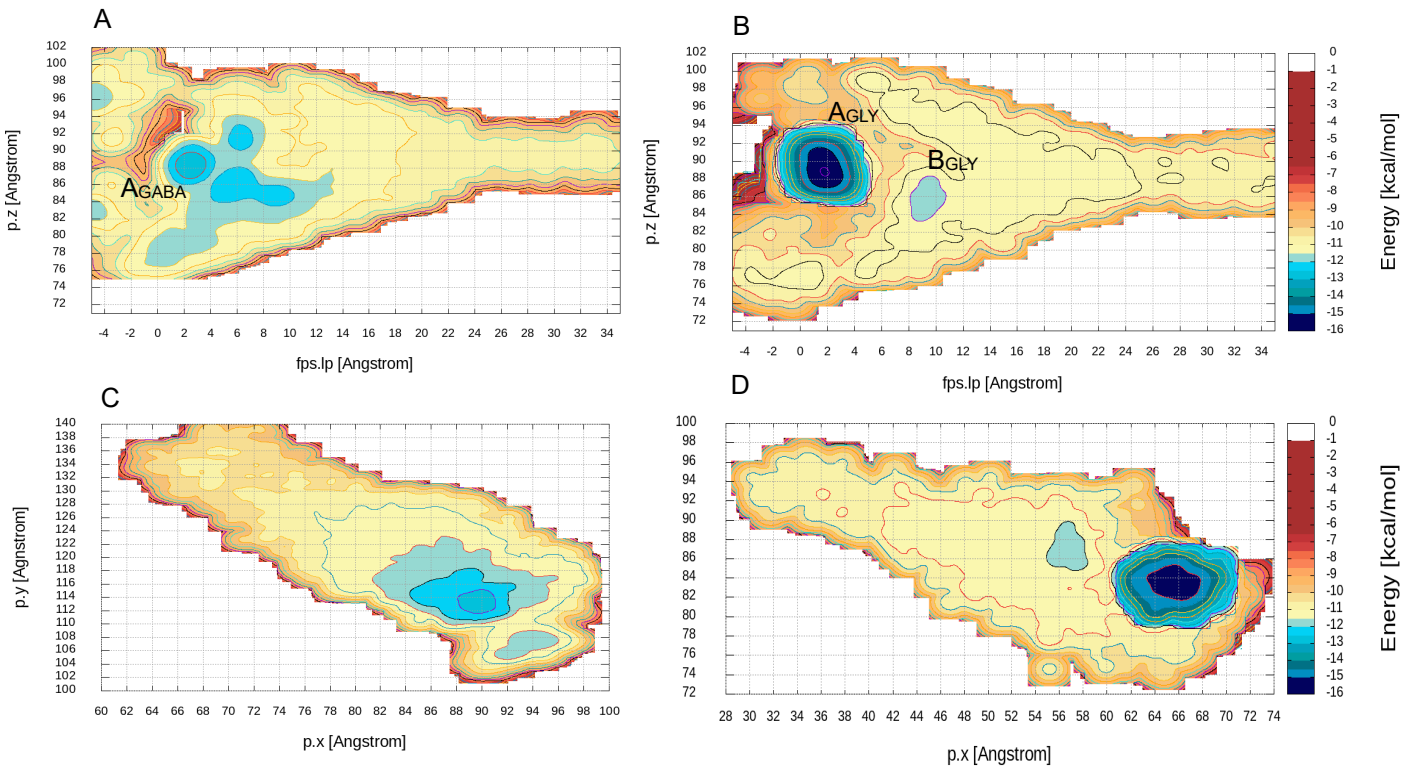

Figure 4.19: The reweighted free energy maps showing the funnel metadynamics reweighted from fps.lp and fps.ld onto the ligand coordinates in cartesian space p.x p.y and p.z. A, the reweighted funnel of the 6PLX system with fps.lp CV on the X axis and p.z on the Y, the same data is plotted for the 6PM5 system in B, both providing a vertical cross-section of the funnels. C and D are the plots of the reweighted p.x and p.y values, of the 6PLX and 6PM5 system respectively. C and D provide a horizontal cross-section of the funnel free energy surface.

When looking at the trajectory and the positioning of this shallow well, we find an area past the C-loop where the glycine ligand is still flanked by protein residues. As shown in Figure 4.20, the ligand is located in a cavity devoid of water molecules but no longer interacting with the vast majority of the principal binding residues.

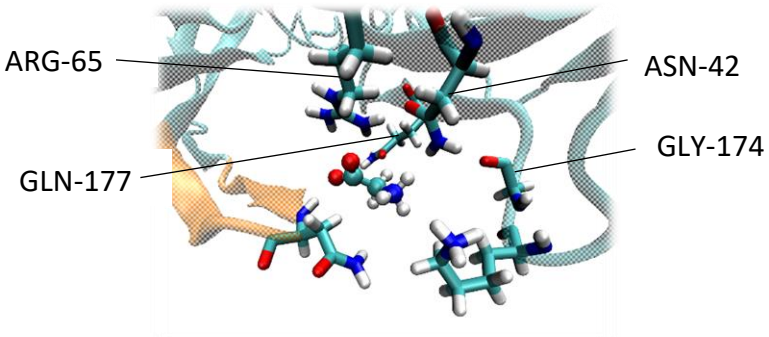

Figure 4.20: A snapshot taken from the funnel metadynamics trajectory showing a potential pre-binding pose where the ligand is flanked by protein residues ARG-65, ASN-42, GLY-174 and GLN-177. The C-loop is highlighted with orange as part of the illustrated protein structure.

This well is revisited multiple times throughout the funnel metadynamics simulation when the ligand transits from a fully solvated state into the pocket to bind. Here the ligand forms multiple transient interactions between its carboxyl moiety and ASN-42 and ARG-65, whilst the ammonium moiety interacts with GLY-174 and GLN-177. These interactions are relatively inconsistent with the glycine moving and rotating frequently but strong enough to hold the ligand in this position temporarily before moving in and out of the pocket, potentially long enough for the carboxyl moiety to form interactions with ARG-65 and move into the pocket and bind. The reweighted GABA plots show multiple regions outside the pocket where the GABA ligand interacts, potentially marking regions of semi-stable prebinding; however, the area covered is generally located across much of the protein surface, suggesting many potential arrangements with no interactions strong enough to point to a definitive well for a single pre-binding pose. This could be a matter of poor convergence, whilst the difference in ligand binding free energy suggests convergence, the CVs do not show a pattern of unhindered diffusion to the extent that would be expected of a converged system and so running the simulation longer could yield more accurate results.

4.4.7 Glycine Unbinding

To monitor the path of the ligand leaving the pocket, and to understand the series of events that leads to ejection of the ligand, we visualised the pocket over time alongside a set of measurements. The hydrogen bonds are the predominant interaction that mediate ligand-protein binding and so the bonds and types over time were plotted, as well as a measurement of ligand movement away from the pocket in the form of the distance between the centre of mass of glycine and the centre of mass of the principal residues.

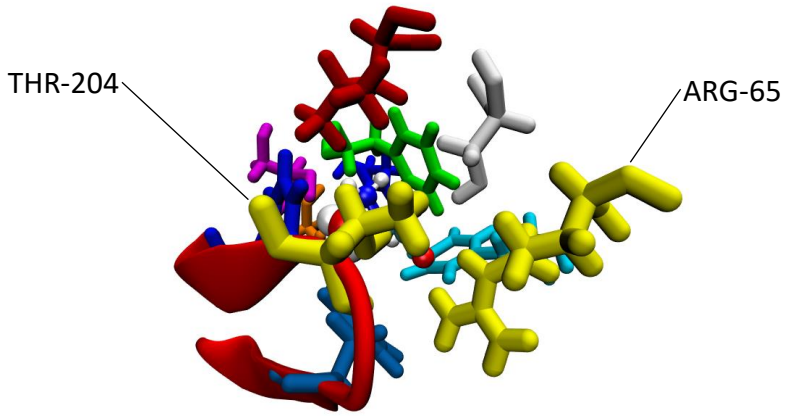

Figure 4.21: The binding pocket is illustrated here with the C-loop in red and the two residues used to judge opening and closing of the C-loop highlighted in yellow, THR-204 and ARG-65.

To judge the opening of the C-loop, the distance was measured between the COM (Centre of mass) of the THR-204 residue and the ARG-65 residue. The ARG-65 and THR-204 are highlighted in Figure 4.21 in yellow, along with the C-loop in red on which THR-204 is located. Here the proximity of the two residues can be seen, as well as why any C-loop opening would require a change in the distance between them.

For the glycine unbinding, we can look at the unbinding trajectories themselves along with the free energy plots to gain insight into the binding path taken by the ligands and the specific sequence of events involved. Looking first at the pattern of hydrogen bonding over time through the funnel metadynamics simulation we can see the consistent interactions between the ligand GZW carboxylate moiety oxygen groups and various hydrogen atoms in ARG-65, THR-204, SER-129 and PHE-159. The figure showing these details is displayed on the next page, followed by a full breakdown of the process.

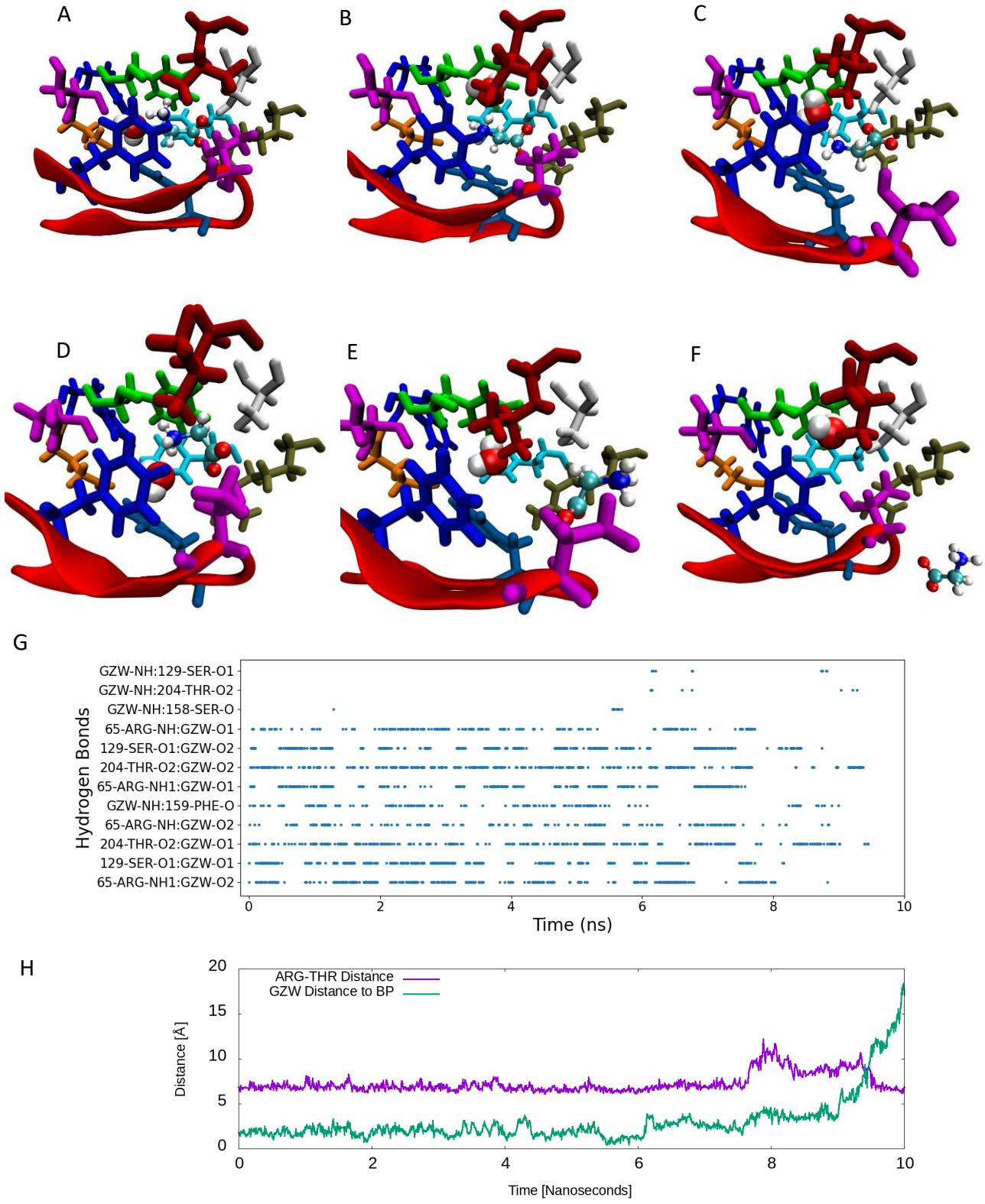

Figure 4.22: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white.. G, shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H, is a plot of distances both for the C-loop opening metric in blue and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for these analyses.

There is a notable lack of interactions between the ammonium moiety and the binding pocket residues; this is due to the water bridge mediated between the ligand and the SER-158 and GLU-157. We see the stable pocket in Figure 4.22.A at 0 ns of the funnel metadynamics: here the ligand is positioned between the water molecule on the GLU-157 side of the pocket where the water bridge is formed with the ammonium group. The ligand is oriented with the carboxylate group close to the ARG-65 and THR-204 of the C-loop. The hydrogen bonding at 0 ns in Figure 4.22.A is illustrated in Figure 4.22.G and shows the expected interactions seen in Cryo-EM and MD work[16,26,137] and has also been illustrated throughout the production previously. At this point, we can see that the C-loop is stable, in a closed position and the ligand moves very little (Figure 4.22.H). The first biggest change in the pocket is seen at 7 ns, Figure 4.22.B, where the water molecule shifts from its position between SER-158 and GLU-157 into a position between PHE-159, LEU-117 and SER-129.

The movement of this water coincides with a change in the ligand hydrogen bonding, specifically, the free rotation of oxygen groups on the ligand stops and the hydrogen bonds stabilise with SER-129, ARG-65 and THR-204. The next change is seen at 7.9 ns where the water molecule moves again: here the water movement occurs simultaneously with a shift in glycine movement away from the centre of the pocket and the first large shift in the C-loop positioning. This is clear in Figure 4.22.G as well as the molecular representation Figure 4.22.C, where we see that THR-204 moves away from ARG-65, placing glycine in a position in between the two residues. The hydrogen bond profile mirrors this change where other interactions are reduced and the main bonds formed are between glycine and ARG-65. It is important to note that after this point the glycine ligand is on a trajectory out of the pocket where the GZW to BP distance increases gradually.

The next major change is at 8.5 ns in Figure 4.22.D: the water molecule shifts lower in the pocket, switching with the ammonium moiety of the glycine ligand. The shift in positioning of glycine ammonium allows PHE-159 hydrogen bonding to resume. We also see a rotation in THR-204 on the C-loop that allows consistent THR-OH:GZW-O2 hydrogen bonding that becomes the main interaction between glycine and the principal binding pocket residues at this point in the ejection. From this point, the movement of glycine out of the pocket is much faster, illustrated in Figure 4.22.H, likely due to the reduced interactions the ligand is making with the pocket. Following this at 9.3 ns in Figure 4.22.E, we see another rotation in THR-204, back to its original position coupled with a rotation of the glycine molecule further out the pocket. This is coupled with the breaking of the GZW-NH:PHE-159 hydrogen bonds and a larger movement of glycine out of the pocket. Finally in Figure 4.21.F the ligand ejects the pocket with no interactions with the pocket residues remaining, the last of which being the

THR-204-OH:GZW-O interactions to break. The movement of the ligand out of the pocket is coupled with the closing of the C-loop back to its original position as seen in Figure Figure 4.22.A.

4.4.8 GABA Unbinding

In the GABA system, the resting state of binding can be seen in Figure 4.23.A on the following page, at 0 ns where the GABA molecule is positioned so that ammonium moiety is in the position where water typically would be in the glycine bound pocket between the GLU-157 and SER-158. In this context, the resting state is defined as a stable binding state where the ligand is not in transit out or in of the pocket. The Carboxylate moiety is on the opposite side of the pocket between the ARG-65 and THR-204 residues. This positioning and the proximity of these two charged moieties is reflected also in the hydrogen bonding profile in Figure 4.23.G where the key interactions taking place are either between the oxygen atoms of the ligand and the ARG-65, SER-129 and THR-204 residues.

The lack of water inside the pocket region means that direct contacts are possible between SER-158, GLU-157 and the GABA molecule. There is water close to the ligand, but not within the pocket or close enough to form water bridges: this water is at the base of the pocket below the GABA ammonium moiety. The first big shift in the GABA unbinding is seen at 3.6 ns, as shown in Figure 4.23.B, where the water below the ammonium moves up and into the pocket, pushing around the top of the ligand and displacing the GABA away from the PHE-159, SER-129 and GLU-157. We can see the impact of this in the hydrogen bonding plot where the SER-129 interaction stops along with interactions to the PHE-159. As the water moves in, we also see the first displacement of GABA further away from the centre of the binding pocket and the C-loop begin to open in Figure 4.22.G.

Following this, the next shift in positioning occurs at 3.9 ns in Figure 4.23.C. Here the water further displaces the ligand towards the entrance of the pocket where ARG-65 and THR-204 are located. There is a much more dramatic change in the hydrogen bonding in Figure 4.23.G, where we can see a complete cutoff of interactions between the ammonium moiety of the ligand and SER-158 and a weakening to eventual removal of the interaction with PHE-159. The carboxyl moiety interactions also flip on the ARG-65 from the rotation of the oxygen atoms of the ligand. We also see a sheer drop in the interactions between the ligand and the THR-204 residue that do not consistently return after this point. From this point onwards we see a gradual climb in both the opening of the C-loop and the movement of GABA out of the pocket.

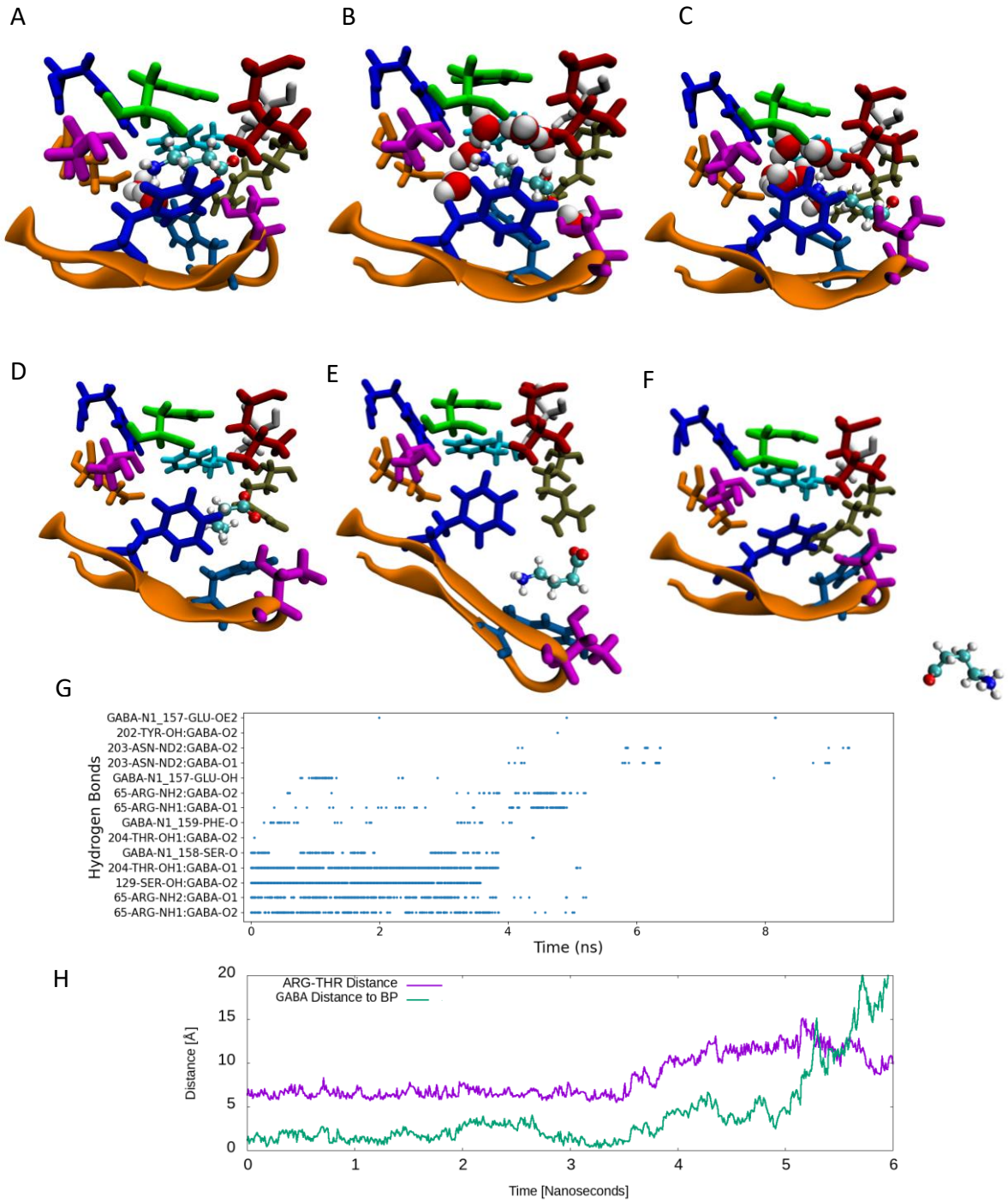

Figure 4.23: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white. G shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H is a plot of distances both for the C-loop opening metric in purple and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for this set of analyses.

At 4.5 ns, as shown in Figure 4.23.D, there is a rotation of the ARG-65 which the GABA carboxyl moiety is interacting with and the GABA molecule is pulled with this rotation further out of the pocket whilst maintaining the two ARG-65:GABA hydrogen bonds. This rotation continues, along with a small amount of interaction of the Carboxyl group of the ligand and THR-204, pulling GABA further out of the pocket to the 5.2 ns point, in Figure 4.23.E. At this point the C-loop is maximally open and GABA continues to move on an outward trajectory until completely removed from the pocket as in Figure 4.23.F, where the C-loop then closes in again at 6 ns.

4.5 Discussion

In this section we have described the results from modelling and simulating a glycine bound and GABA bound GlyR in conventional molecular dynamics as well as with enhanced sampling in funnel metadynamics. The results provide us an overview of how the ligands interact with the receptor binding pocket at rest and over the course of a trajectory involving a series of binding/unbinding events. This next section of the thesis will bring together each element of the results to discuss the implications for understanding the binding of partial and full agonists as well as what we can understand about full/partial agonism by comparing the two.

4.5.1 Full Agonism in the GlyR

From the results demonstrated so far, the glycine system presented here exhibits a typical binding behaviour by interacting with the known binding pocket residues ARG-65, THR-204, SER-129, PHE-159. This binding profile is supported by other work where cryo-EM data[16,26,33,137] looked at glycine within GlyR pockets. The dynamics however have not been characterised in as much detail as presented here, particularly in the context of movement out of the pocket. In standard molecular dynamics and in the early portion of our enhanced sampling, we see that the carboxyl and ammonium moiety regularly rotate in the pocket, resulting in a relatively even distribution of hydrogen bond interactions between the ligand and the principal binding residues (Figure 3.13). The funnel metadynamics simulation carried out takes current understanding of glycine binding a step further, demonstrating the dynamics over time from a bound to an unbound state. This is illustrated in the previous section in Figure 4.22.A-F where the snapshots of the trajectory provide a sequence of events leading to ejection of the ligand. At the end of the ejection, we see a temporary stable pre-binding of

glycine in a region also identified from the free energy calculations. Here the ligand is positioned just past the tip of the C-loop THR-204 residue but instead of being free to move in the bulk water, it is flanked by residues GLN-177, ARG-65, GLY-174 and ASN-42. The hydrogen bonding with these residues is sufficient to hold the ligand in this position before and after binding. The minimum identified on the free energy surface, coupled with the hydrogen bonding profile and positioning of the ligand suggests this to be a prebinding pose the ligand takes before and after ejection. This is an important element of the binding path that has not been reported before and serves as a potential mechanism for glycine movement in and out of the pocket, holding the ligand in a position where the critical interactions with THR-204 and ARG-65 can be mediated or disrupted to transport glycine in/out of the pocket.

One area of GlyR binding that has not been covered in much detail is the role of water in the binding pocket. Water is not resolved in the Cryo-EM structural images of these GlyRs and therefore is not typically factored into how glycine binds; however, some previous computational work has found water to play a role in the binding pocket from molecular dynamics simulations[67,137]. These simulations of the GlyR show a favourable site that water occupies when glycine is bound, mediating a water bridge between the ligand and the binding pocket GLU-157 residue. Our system mirrors this, with water entering the pocket early in equilibration and remaining in the same fixed position between the ligand and the GLU-157 and SER-158 residues of the binding site. The trajectory results support the idea that water is a key component of glycine binding, mediating interactions that would not be possible given the positioning[137] of the ligand taken from Cryo-EM data. Along the trajectory glycine takes to unbind, we also see an important role for this water molecule. The first step along the unbinding trajectory involves the shifting of this water molecule in the pocket, breaking interactions and displacing the ligand away from its resting bound state. The next two large changes in binding pocket topography also include the movement of this water that becomes positioned in between the ligand and key binding pocket residues ultimately leading to the ligand pushing further out the pocket and interacting with THR-204 alone, rotating and ejecting. Observations made both in the production and during the ejection of the ligand supports previous work and suggests a larger importance of this binding pocket water molecule that is often overlooked when discussing how glycine is bound within GlyRs.

4.5.2 Partial Agonism in the GlyR

The GABA system shown in this work shows a binding profile that matches well with the Cryo-EM data it was derived from, involving hydrogen bond interactions with ARG-65, THR-204, GLU-157, SER-158, SER-129[26]. The dynamics, however, has not been studied in the GlyR with a partial agonist like GABA. The hydrogen bonding profile suggests a relatively fixed position for the Carboxyl moiety, where there is very little flipping of the oxygen-hydrogen pairs between the ligand and the ARG-65 and THR-204 residues. The ammonium moiety however is flexible between the GLU-157 and SER-158 residue, showing an even proportion of hydrogen bonds spread across the three hydrogen atoms. Like the glycine bound system, the most novel information gathered about the binding comes from the funnel metadynamics simulation and the trajectory observed of the ligand leaving the pocket. Whilst there is no water initially in the pocket, the first major change we see in the binding site occurs as water moves in from below the GABA molecule, displacing it towards the entrance. The movement of water into the pocket and shift in the position of GABA leads to a drop off of the key interactions between the ligand and the pocket, weakening the binding and allowing the progressive movement of the ligand towards the entrance of the pocket and eventual ejection. The shift in the position of GABA is necessary for the formation of the strong interaction of the carboxyl moiety with ARG-65 and subsequent rotation of GABA out of the pocket. What is seen in this trajectory is a potential mechanism for the transit of GABA out of the pocket, with the identification of the key residues that play a role, as well as the way water is involved in disrupting the interactions that keep GABA stably bound. The work here serves to provide a greater understanding for how GABA interacts with the pocket and how each residue in the pocket is involved in keeping the ligand positioned to mediate its effects. The finding that water moves into the pocket prior to ejection also provides an invaluable insight into how the larger topography of the pocket and surrounding residues plays a role in the binding affinity, rather than just the specific principal binding residues that hold GABA in place.

4.5.3 Comparing Full and Partial Agonism of GlyRs

GABA and glycine are similar molecules with vastly different binding affinities to the GlyR. The results presented here provide an opportunity for comparing and understanding what makes these two ligands behave so differently. The clearest difference seen here is the interactions within the pocket. Glycine primarily forms hydrogen bonds with ARG-65, SER-129, THR-204 and PHE-159, and has a water bridge that interacts with SER-158 and GLU-157. GABA does not form consistent hydrogen bonds with PHE-159 but does directly make hydrogen bonds with SER-158 and GLU-157. This distinction is due to the longer length of GABA, projecting

its ammonium moiety into the pocket space where water is typically located with glycine binding. The size of the GABA molecule also leads to another key difference that can be seen in Figure 4.24, where the of the binding pocket is notably larger than with glycine bound. The importance of binding pocket volume has already been discussed in relation to the whole protein system where the stability of binding seemed to be linked to the stable volume of the pocket.

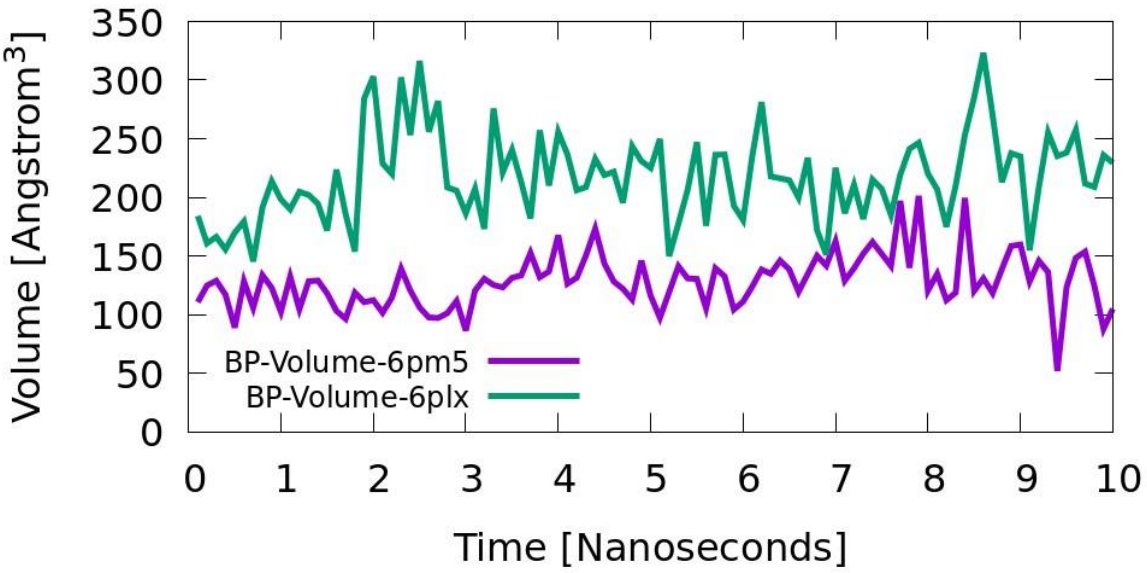

Figure 4.24: The volume of the binding pockets shown over the course of the ejection trajectory (The first 10 ns of FmetaD) with the GABA bound system in green and the glycine bound system in purple. These measurements were made using the funnel metadynamics trajectory analysed with the mdpocket tool[153].

When considering the implications of this for the binding affinity of GABA versus glycine the disparity in pocket volume could be a key factor in the difference between the two ligands. The work that produced the cryo-EM data that these systems were produced from also showed full agonists to possess a more compact binding site[26]. The results from the funnel metadynamics simulation provides valuable insight into how this difference in pocket volume may be responsible for the difference in affinity. In both GABA and glycine ejection trajectories, water movement around the pocket plays an important role in blocking interactions between the ligand and the binding pocket residues. In the case of glycine, a single water molecule shifts around the ligand as it moves to the exit of the pocket, however in the GABA trajectory, multiple water molecules shift into the pocket drastically and rapidly disrupting the key interactions required for stable binding. The glycine bound pocket in these simulations is almost entirely protected from the bulk water, with the only space available for water migration into the pocket being that of the space between SER-158 and GLU-157 that is always occupied and forms a water bridge between the ligand and the protein. These results suggest that the reduced volume of the full agonist glycine bound pocket limits the presence of water to only a single molecule that itself mediates protein-ligand interactions, whereas in the partial agonist GABA bound system, multiple water molecules can move into the pocket and disrupt interactions and lead to rapid ejection of the ligand. This points to a potential mechanism through which the reduced affinity and partial agonism of GABA can be explained even with the similar binding profile it possesses.

4.5.4 GABA resistance to dieldrin receptor (RDL)

Previous work carried out in the Molteni group utilised similar approaches to those used here, the methods presented taking strong influence from prior work done on the RDL receptor binding with GABA. Therefore, it's important to compare the system bound with GABA here to the RDL receptor results to identify important distinctions and also ensure that this work aligns with similar systems. The work carried out previously utilised a structure of the insect RDL receptor, a pLGIC described as prototypical for the family of cys-loop receptors that has been studied experimentally with GABA ligand application.

The topology of the pockets is very similar between the structures, this is highlighted in Figure 4.25.A where the C-loop and positioning of the pocket residues around the ligand can be seen for both the α-1 glycine receptor system in orange and the RDL receptor in blue. The key similarities involve the interactions that form with the binding pocket residues, primarily the salt bridge between the charged glutamic acid and arginine residues at either side of the pocket flanking the GABA ligand at its also charged carboxylate and ammonium moieties. From the structure alone, it would be expected that they would share a similar affinity and consequently produce similar ligand binding free energies.

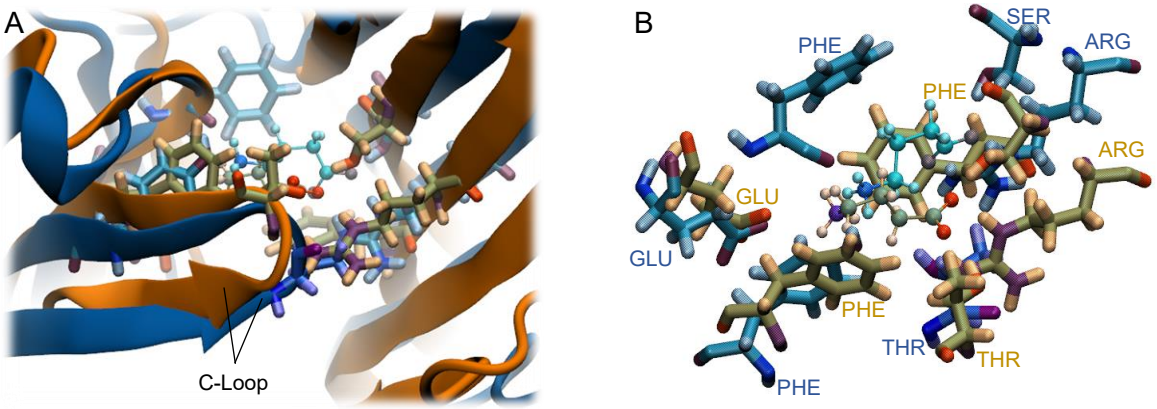

Figure 4.25: Illustration of the the RDL receptor (blue) and the α-1 glycine receptor (orange) structures superimposed. In A, the protein structures are represented as a cartoon illustration along with the key binding pocket residues of each system. In B a clearer representation of the binding pockets along with labels of important residues is shown with the GABA ligands in the centre.

The methods involved a similar protocol, applying a funnel to the protein-ligand system and using the same CVs across the funnel to bias the ligand to explore conformational space around the pocket. The RDL receptor results yielded a binding free energy for well A at 3.44 ± 0.69 Kcal/mol, after the funnel corrections. This matches up similarly to the results gained in the α-1 glycine receptor here which had a binding free energy of -3.01 ± 1.33 Kcal/mol. These results are within a reasonable range and are an expected result, this is primarily due to the composition of both pockets being very similar.

There is however a larger difference in the area of conformational space outside the binding pocket, labelled with B in figure 4.26.A. The reason for this basin to be present in the RDL receptor and not in the α-1 glycine receptor is due to the absence of the key residue responsible for the pre-binding pose, an arginine residue outside the binding pocket, on the complementary subunit. In the RDL work, it was suggested that the charge arginine residue positioned into the solvent was responsible in their simulations for capturing GABA and shuttling the ligand into the pocket. The α-1 glycine receptor has no charged molecule in this position and therefore the basin does not appear, supporting both the validity of this work and the RDL simulations, associating this pre-binding to the residue suggested to be key.

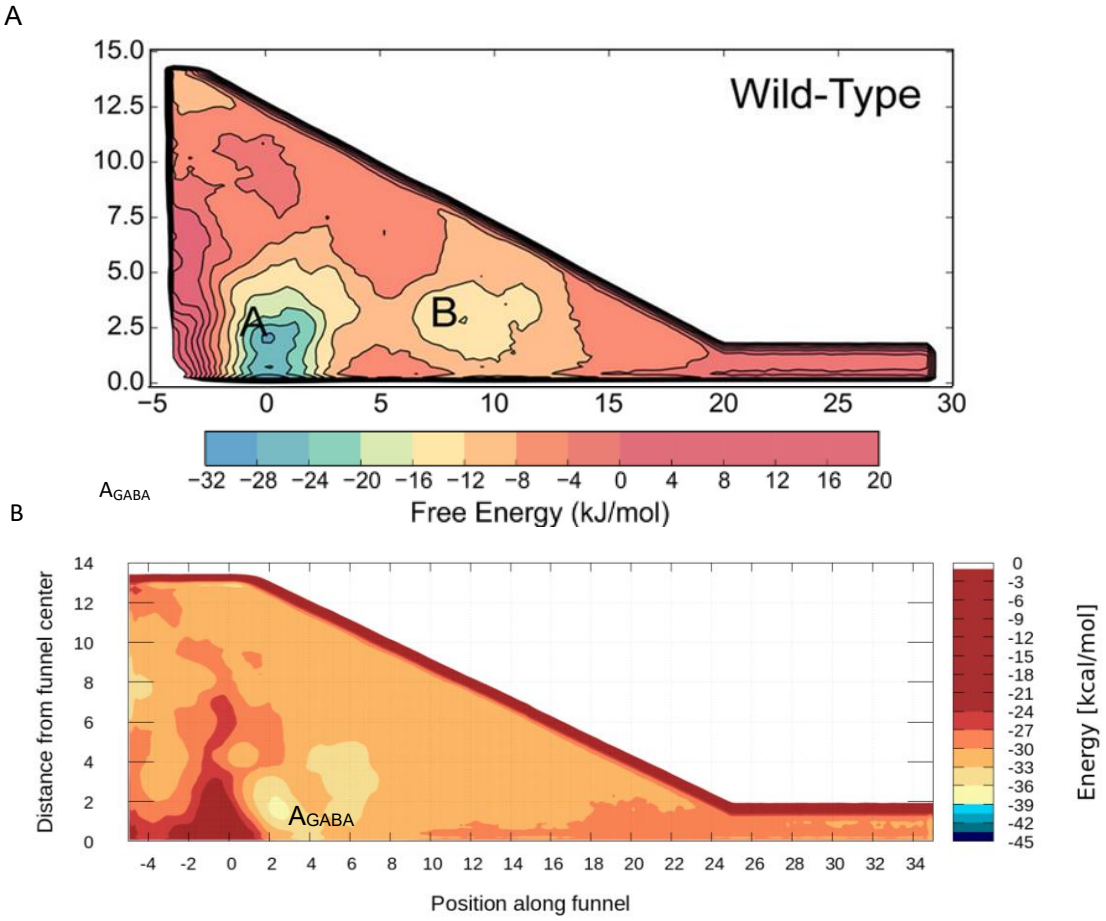

Figure 4.26: The free energy landscapes produced from this work alongside the results gained in the Comitani et al paper[86]. A shows the past work with the RDL receptor with the funnel axis fps.lp on the x axis and the fps.ld collective variable on the y axis. There are two visible wells labelled A (the orthosteric binding site) and B indicating a pre-binding pose. B shows the results from this work previously discussed for GABA binding to the α-1 glycine receptor.

4.5.5 Conclusion

Overall, the data collected here provides an in depth look at the protein-ligand interactions involved in both partial and full agonism of GlyRs, giving a detailed profile of the key residues that interact in both ligands' stable bound poses. The results highlight a potential mechanism for GABA and glycine to transit out of the binding pocket as well as suggesting a potential pre-binding pose that exists outside the immediate binding site for glycine. The pre-binding pose can serve as a starting point for future work to investigate the role of residues around the pocket that may be impacting binding affinity in a way that is unclear from static Cryo-EM structures and experimental data alone.

The work presented in this chapter also demonstrates the importance of taking into consideration the role of water within the pocket that has otherwise been overlooked, both for full and partial agonists. The glycine bound system supports the idea that water in the pocket is a key element in mediating ligand-protein interactions from the strong tendency for a single water molecule to enter the binding pocket. For the partial agonist system, water shifting into the pocket as a potential result of increased pocket volume may be a mechanism and explanation for the reduced affinity GABA has for the receptor, whilst still possessing many of the same hydrogen bond interactions that the potent full agonist glycine has. On this basis, further work could explore other ligands and other pLGICs that share similar binding pocket topography as the role of water in ligand affinity may be a common phenomenon based on the structural arrangement of the binding site.

Chapter 5

The N46K Mutation in the Glycine Receptor and its Effects on Glycine Binding

5.1 Introduction

Inhibitory neurotransmission in humans can be disrupted by mutations in GlyRs, leading to neurological disorders like hyperekplexia. In humans in fact, mutations in α-1 GlyR gene (GLRA1) result in hyperekplexia that is characterised by massive tremors and exaggerated startle response. The specific mechanism of the disorder is not completely understood although a range of mouse systems have been used to analyse the pathology; in general, it is suggested that specific mutations change important residues of the GlyRs, reducing their functionality and therefore disturbing inhibitory glycine-mediated neurotransmission, altering synaptic activity, and leading to the disease phenotype observed[52].

The mutants that have been found for glycine receptors are of key interest not only in resolving the pathology of hyperekplexia but can also be used to understand better the mechanisms through which pLGICs themselves function. There are a variety of mutants that have been identified with a range of severity levels and impacts on channel function. The placement of the mutation within the receptor also varies; the GLRA1 mutants that have been associated with hyperekplexia are often located near or in the M2 pore-lining domain of the α-1 GlyR, the suggested mechanism for these mutations being an alteration of the allosteric signalling pathway that couples binding to gating[51]. Other mutants across the receptor lead to similar channel disfunction but with varying mechanisms of action. The mutant Y128C, for example, is found in the extracellular domain and causes spontaneously open channels in the absence of an agonist. The suggested mechanism of action for the Y128C mutant is that it results in disrupted packing and rearrangement of the α1 subunit M1 and M2 helices, therefore altering the pore structure[51]. Alternatively mutant Q266I is located within the transmembrane domain, and results in reduced sensitivity of the channel to glycine and reduced maximal currents. The location of Q266I places it in the TM2 (second transmembrane domain) segment lining the

pore, it also results in smaller water cavities within the TMD changing how the GlyR is impacted by glycine binding[168].

Whilst studying these mutations is a valuable approach to understanding the hyperekplexia pathology, the data collected also inform us as to the role of various residues in the general functioning of pLGICs. The severity of the impact of the mutations gives a strong indication of the importance of a given residue in receptor functionality. These mutants also offer a starting point for investigations by providing evidence for a residue having a crucial functional role that can be focused on. Mutants are regularly explored in this way with computational methods to approach complex biological systems like neurotransmission, particularly with pLGICs[169]. By developing a system based upon a known mutant structure or by carrying out mutagenesis of an important residue to investigate the effects of its disruption, a wealth of data can be collected on the resulting disturbance and then related back to understanding how the wild type functions. An example of this approach being used recently when studying pLGICs is work carried out on mutants in the TM1 and TM2 sections of glycine receptors. The work used molecular dynamics to explore the conformational changes and transitional movements between closed and desensitised states, focussing on mutations V260M, S241L and R252S. The study of these mutants with molecular dynamics demonstrated a reduced capacity for the receptor to transition towards the closed state, a necessary step for functionality of GlyRs after they have been bound, open and then desensitised. By utilising molecular dynamics, it was found that the mutations led to disruption of TM1-2 contacts and changing the orientation of the sub-structures.

The molecular mechanisms by which these mutations mediate functional changes in the protein vary based on the type of residue substituted in and where they are localised within the structure. The systems developed here were designed for investigating the binding and unbinding of GlyRs and so a mutant useful for this objective likely should be in the ECD and preferably close to or in the binding pocket. Previous work carried out in our group[46] on pLGICs has utilised mutant residues in the binding pocket to investigate the mechanisms of binding/unbinding. The work utilised molecular dynamics and funnel metadynamics, mutated two residues in the binding pocket to observe the changes in the binding process. The system used was the insect resistance to dieldrin (RDL) γ-aminobutyric acid (GABA)-gated pLGIC, and the mutant residues were the ARG-111, GLU-204 (ARG-65 and GLU-157 are the equivalents of in our α1 GlyR systems). These mutants were chosen based on previous electrophysiology mutagenesis experiments demonstrating nonfunctional channels as well as their own MD simulations demonstrating the critical importance of these residues. Funnel metadynamics was carried out and a comparative analysis of each free energy landscape and

binding mechanism studied, the results providing a more in depth understanding of how these mutants have such a potent impact on channel functionality.

In the case of our GlyR systems, a similar method was applied, including choosing a mutant localised around the pocket. An important factor that was considered when choosing a mutant to system was whether it would be impacting affinity or efficacy. If a mutant alters efficacy, the residue changed likely has a role in the conformational cascade that links binding to gating or the gating function itself. These conformational changes and the gating itself happen on a much longer timescale[157] than the methods used here are capable to simulate. The goal of this chapter is to investigate altering binding affinity by choosing a mutant that has experimental work demonstrating reduced binding affinity rather than efficacy. It is also important that the impact of the mutant is subtle enough to disrupt binding partially and not completely stop binding as there are less insights to gather about the path and binding mechanism from simulations of a completely broken pocket than one that is still mostly functional. The N46K mutant mentioned in Chapter 1 is what was ultimately chosen as it fits all these criteria; specifically, a shift in concentration-response curves has been demonstrated without changes in the relative maximum response that indicates N46K is a mutant which impacts binding affinity rather than efficacy (Figure 1.10).

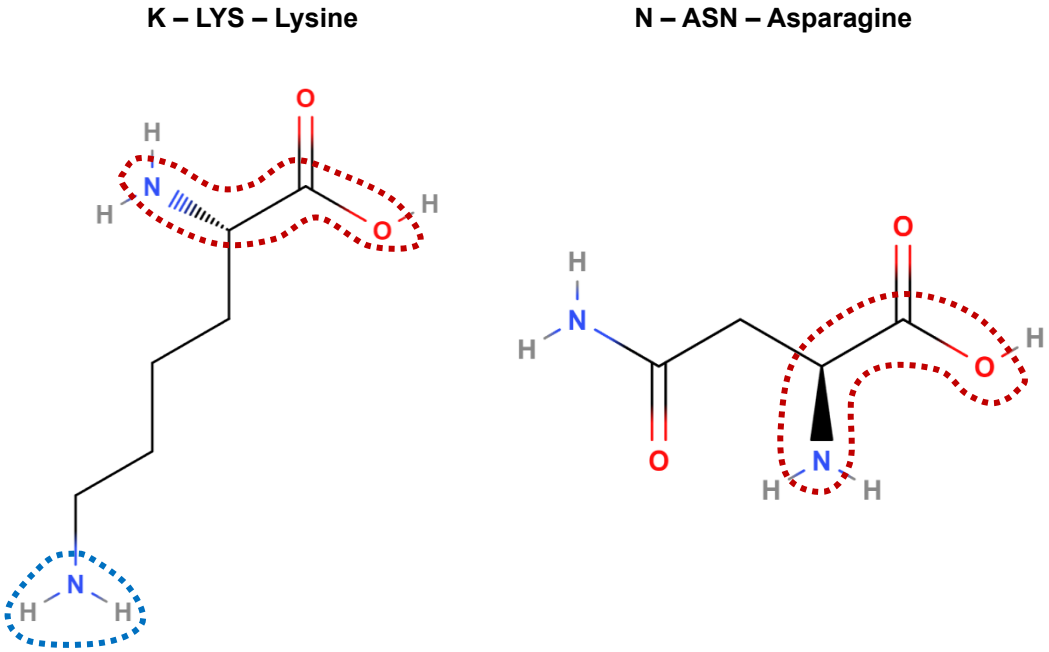

Figure 5.1: Lysine and Asparagine molecular structure on the left and right respectively. The structures shown are the complete molecules outside of the protein structure. The backbone atoms are highlighted by a red dotted line and the amine group that becomes positively charged at physiological pH is outlined with a blue dotted line.

The N46K mutant involves a substitution of the Asparagine residue 46 for a Lysine residue. The Lysine is a longer molecule with an amine group that becomes a positively charged ammonium in physiological conditions as illustrated in Figure 5.1. The positioning of this residue in relation to the rest of the glycine receptor is made clearer in Figure 5.2, with the relevant substructures labelled for both the wild type on the left and mutant variant on the right.

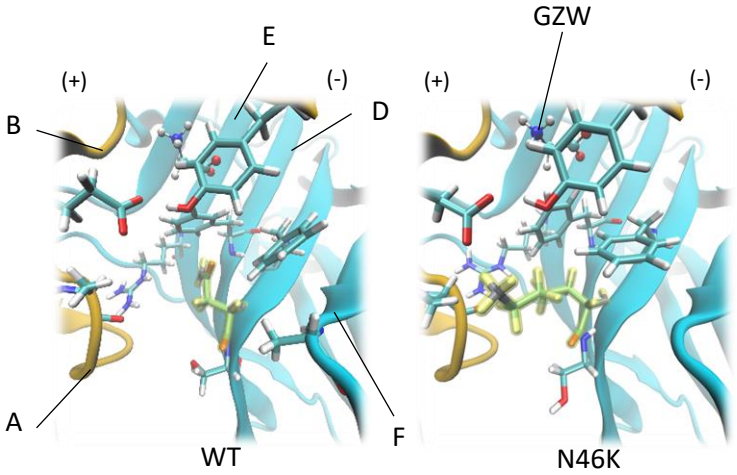

Figure 5.2: Snapshots from during the production showing the location of the mutation, changing the residue from Asparagine to Lysine. The binding pocket loops A, B, D and F are also labelled with the principal (-) and complementary (+) subunits coloured cyan and orange respectively.

The work carried out on the N46K mutant suggested that the role of ASN-46 could be to interact with residues on the adjacent subunit or residues important for binding, therefore stabilising ligand binding[44]. This mutant specifically, whilst not a mutation found in human hyperekplexia patients but causes hyperekplexia in the Nmf11 mutant mice, still serves an important role in understanding the GlyR and hyperekplexia. Mutations of the α-1 GlyR gene is the predominant cause of hyperekplexia in humans but the specific ways in which mutants in regions like the subunit-subunit interface that the N46K mutant impacts is still unclear[44,170]. Inducing the N46K mutant in a system like those shown here is a vital step in understanding the role of these residues in hyperekplexia and understanding how these residues contribute to properties of the binding pocket.

The aim of this chapter is to replicate this mutant in an MD system and test these ideas, identifying at the molecular level how the ASN-46 impacts ligand binding and see what changes when the Asparagine residue is substituted for a Lysine. The data collected also provides a foundation for understanding the mechanical processes that underly hyperekplexia, an important step in developing efficacious treatments. The next few sections of this chapter

will explain how this mutant system was set up and then will discuss the results found from both unbiased molecular dynamics and funnel metadynamics. The objective of the work presented in this chapter is to explore the binding/unbinding of GlyRs by using a mutant to disrupt the process and compare the outcomes when the N46 residue function is disrupted.

5.2 modelling the N46K mutant

The N46 residue is located at the subunit interface, below the binding pocket in the ECD. It is positioned between loops D and F whilst also being in close proximity to loop A on the adjacent subunit, important sub-structures involved in agonist binding[171]. It does not interact directly with the ligand but instead has been suggested to play a role through its interactions with binding pocket residues and other amino acids at the subunit interface[44]. The Figure below shows the positioning of the ASN-46 in the wild type α-1 GlyR from the 6PM5 system produced in Chapter 3 as well as the mutant system with the LYS-46 in place on the right made from the 6PM5 system in this chapter.

5.2.1 Mutation Procedure

To system the impacts of this mutation, the raw PDB data needed to be modified prior to the setup of the structure to incorporate the change. To ensure that comparisons could be made from a wild type as similar as possible, the same 6PM5 PDB structure was used as a starting point to carry out the mutation. The way this was done was through the removal of all atoms in residue 46 of each subunit, excluding atoms of the backbone as to not disturb the structure of the loops and GlyR as a whole. Once these atoms were removed, the residue name was modified from ASN to LYS which could then be reformatted with the pdb4amber tool and tleap within the AmberTools package[146], to fill in the missing LYS atoms. Because this mutant requires editing of the residues before the generation of parameter and topology files, the whole process of system creation was required to be carried out, including minimisation, heating, equilibration, and production.

5.2.2 Simulation Details

The 6PM5-N46K mutant system was developed with identical parameters and procedure to that of the 6PM5 system in Chapter 3 apart from this initial alteration of residue 46. The first 218 residues of the ECD were isolated, leaving five residues for restraints to be applied onto at the tail of the structure. Hydrogen atoms were added using the H++ tool[150] at a neutral pH and the structure was solvated and formatted for use in NAMD 2.14[96] with the amber forcefield

ff14sb[90] and GAFF[165]. The solvation and water system were the same used in Chapter 3 for the 6PM5 system, using a 15 Å buffer, 0.15 M of Na^+, Cl^- ions. Disulphide bonds were checked in the same way and the final structure produced contained 123,339 atoms including the water, protein, ions and zwitterionic glycine ligand.

The minimisation was carried out in the same way as Chapter 3 where restraints were released over a 4 step process and the equilibration/production used the same NPT parameters with the Langevin thermostat and barostat. A 2 fs timestep was used again, the SHAKE algorithm was utilised to constrain hydrogen bonds, partial mesh ewald used for electrostatic interactions with the same 10 Å cutoff. Heating also followed the same procedure, increasing the temperature at increments of 25 K until 310 K was reached. Equilibration also followed the same sequence of releasing harmonic restraints down from 25 Kcal/mol/$Å^2$ until only 1 Kcal/mol/$Å^2$ remained on the last five residues of the ECD.

5.3 The Mutant N46K GlyR system Results

To quantify the stability of the system and to identify any artifacts of the preparation, the same routine of profiling RMSD of the whole protein and various substructures was used. Due to the use of the 6PM5 structure once again, we expected very similar results and across the production. In addition to profiling the interactions in the pocket, we also looked at how the positioning and interactions of residues around the N46/K46 differed between the wild type and mutant system.

5.3.1 Structure and Stability

The whole protein RMSD shows similar results to that of the 6PM5 system, with the RMSD plateauing slightly lower than the wild type at around 1.75 Å in Figure 5.3. The lack of any spikes or any substantial increase in RMSD suggests that the preparation of the system was successful; however, to ensure this was the case, the RMSD values for specific substructures was also measured for comparison.

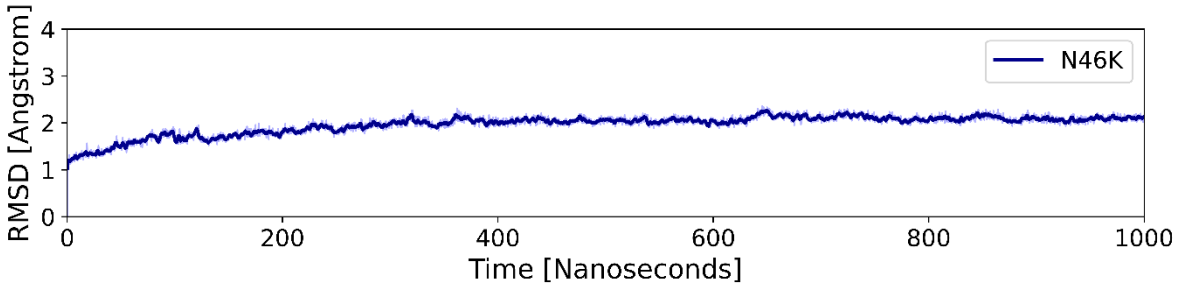

Figure 5.3: The RMSD of the protein backbone as a function of time during the production molecular dynamics simulation.

When looking at the sub-structures in Figure 5.4 we see that there is little difference between the 5 subunits. The C-loop structures are also consistent across the pockets which would be expected for a structure with 5 pockets all bound with a ligand.

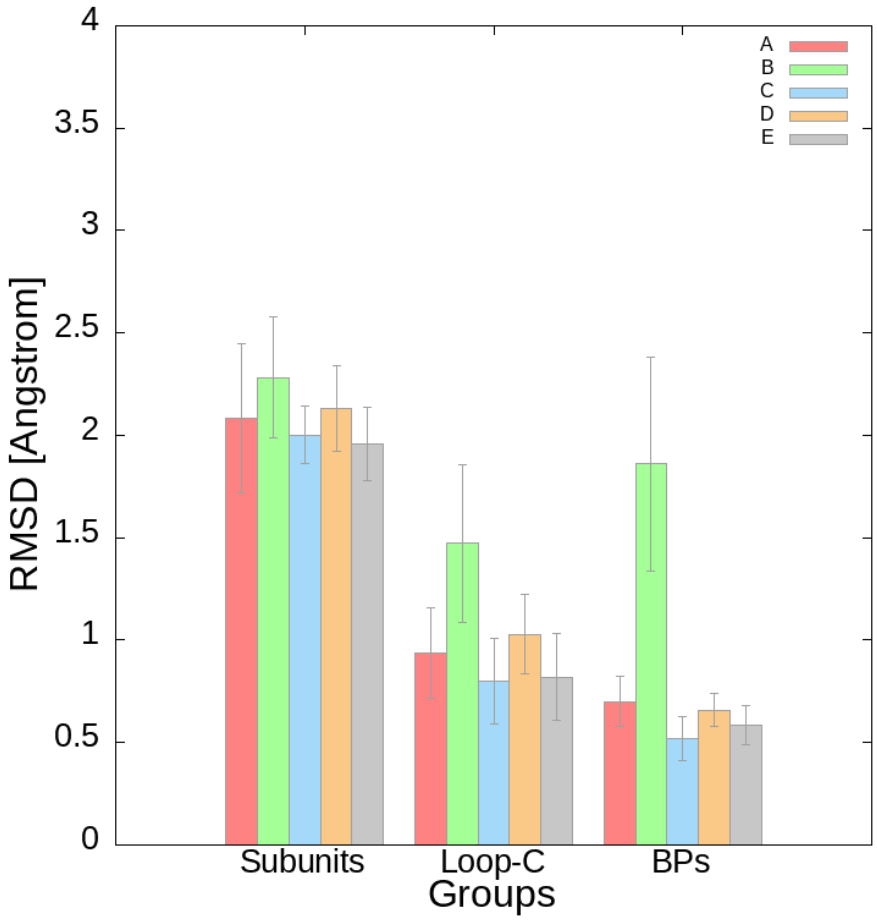

Figure 5.4: The average RMSD values for each substructure across the five pockets. The error bars here represent the standard deviation measured over the course of the production.

The binding pocket RMSD values are also mostly uniform with an exception for binding pocket B which could suggest some instability if the key binding pocket residues are moving more relative to the other pockets but any relevance of this elevated RMSD in pocket B would be

more apparent after profiling the interactions in the pocket and the time resolved RMSD for each substructure.

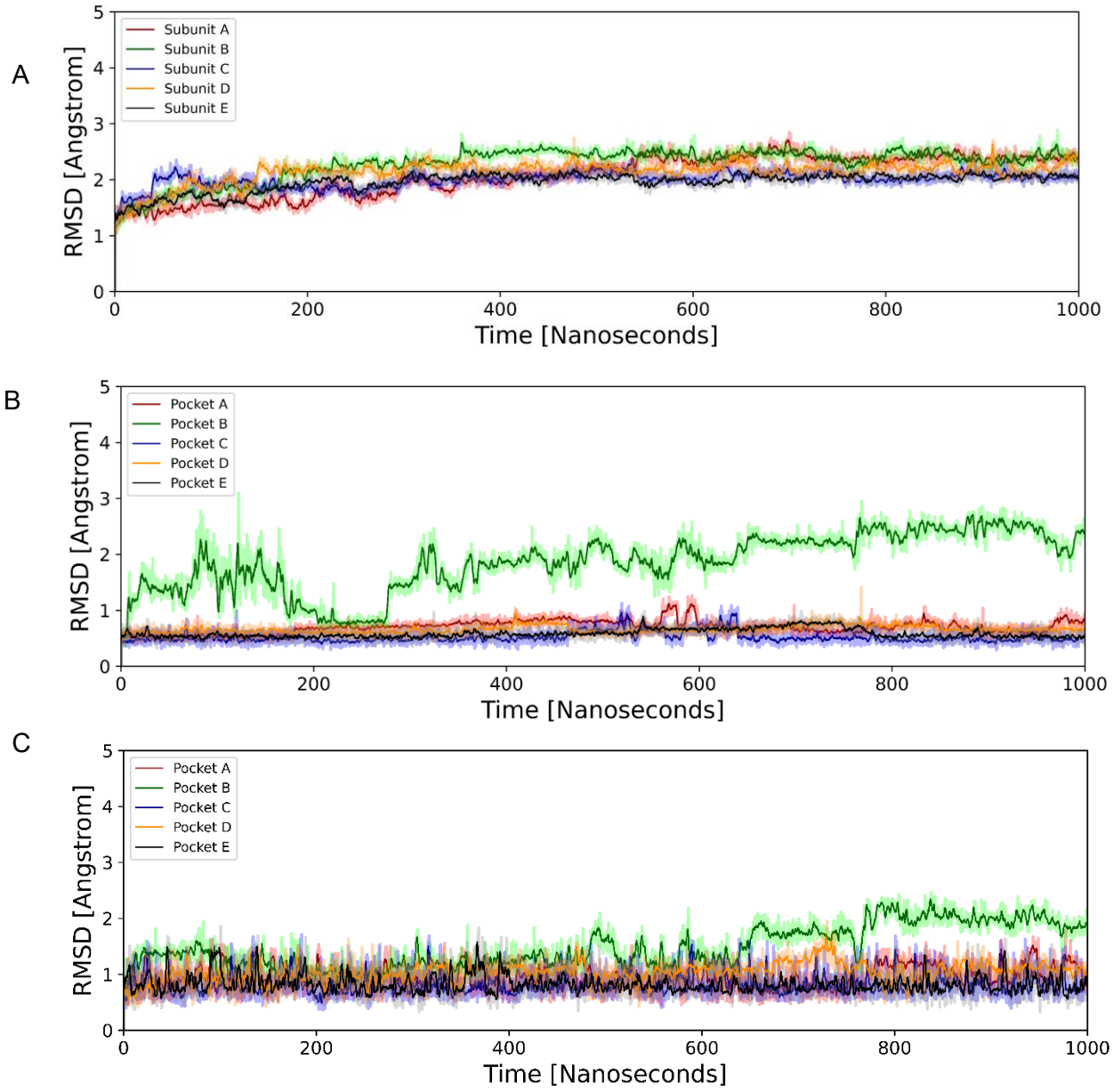

Figure 5.5: The RMSD values plotted over time for the key substructures. A, Subunits. B, pockets. C, C-loops.

Looking at these structures RMSD values over time provides the opportunity to identify spikes and specific time periods of increased RMSD. The RMSD for the subunits over time shows a similar consistency to the average RMSDs where we see little deviation between the structures over time, shown in Figure 5.5.A. The RMSD value for the BPs over time however highlights further instability seen in the averages plot, where pocket B shows up to 5 times higher values, indicating the pocket differs much more in motion compared to pocket A, C, D

and E (Figure 5.5.B). For the RMSD values taken for the C-loop structures in Figure 5.5.C we see the same conformity across the 5 pockets with the exception of pocket B. Here we see much larger movements in pocket B C-loop that is not highlighted as well in the averages plot, adding to the idea that pocket B has lost the stability the other pockets appear to share. We can conclude from these measurements that whilst the system is mostly stable, binding pocket B is showing signs that the pocket is not completely stable, an important piece of information to consider when choosing a pocket to explore the binding events for.

5.3.2 The Binding Mode

To further understand the binding profile of this system as well as check more thoroughly the stability of the binding interactions, the ligand and C-loop position were measured. The same procedure was carried out for these measurements as in Chapter 3 and 4.

5.3.3 Hydrogen Bonding

The hydrogen bonding profile mimics that of the 6PM5 wild type system, where the total number of hydrogen bonds levels out at around 4 overall (Figure 5.5). There is however another exception for pocket B where we see a dropping off of interactions and eventually complete lack of hydrogen bonds with only a few consistent interactions over the course of the production. The drastic reduction of hydrogen bond interactions of the ligand in pocket B, combined with the RMSD measurements, points to instability substantial enough to disrupt binding.

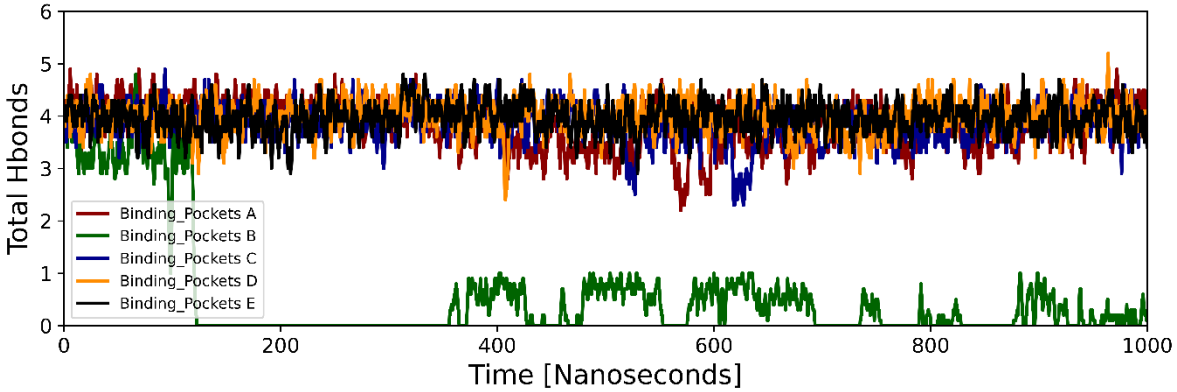

Figure 5.6: The total number of hydrogens for each pocket over the course of the production, measured between the ligands and binding pocket residues of each respective pocket A-E.

The hydrogen bonding profile of a stable pocket, such as E used in this example, provides us with better insights into which binding pocket residues are involved in keeping the ligand stable

and bound. In Figure 5.7.B the types of interactions for pocket E are plotted along with the proportion of time they were measured as present. Here we see an almost identical plot and bonding profile to that seen in Chapter 3 Figure 3.17 for the wild type. Looking at the snapshot taken from the simulation (Figure 5.7.A) we can see the positioning of binding residues and the ligand. The carboxylate moiety of the ligand interacts with the ARG-65, THR-204 and SER-129 residues on the entrance side of the pocket whereas the ammonium moiety only interacts with PHE-153 and being in position to form a water bridge to GLN-157 and SER-158 in the same way as the wild type 6PM5 wild type system. We also see again a flexibility of the ammonium and carboxylate moieties to rotate, meaning there is no preference for any one oxygen/hydrogen to interact with any one binding residue atom. These results were expected as the N46K mutation does not directly alter any of the binding pocket residues themselves and so it would be reasonable to assume that the impacts of the mutation would be clearer in the dynamic binding/unbinding events of funnel metadynamics rather than over the relatively short period of MD.

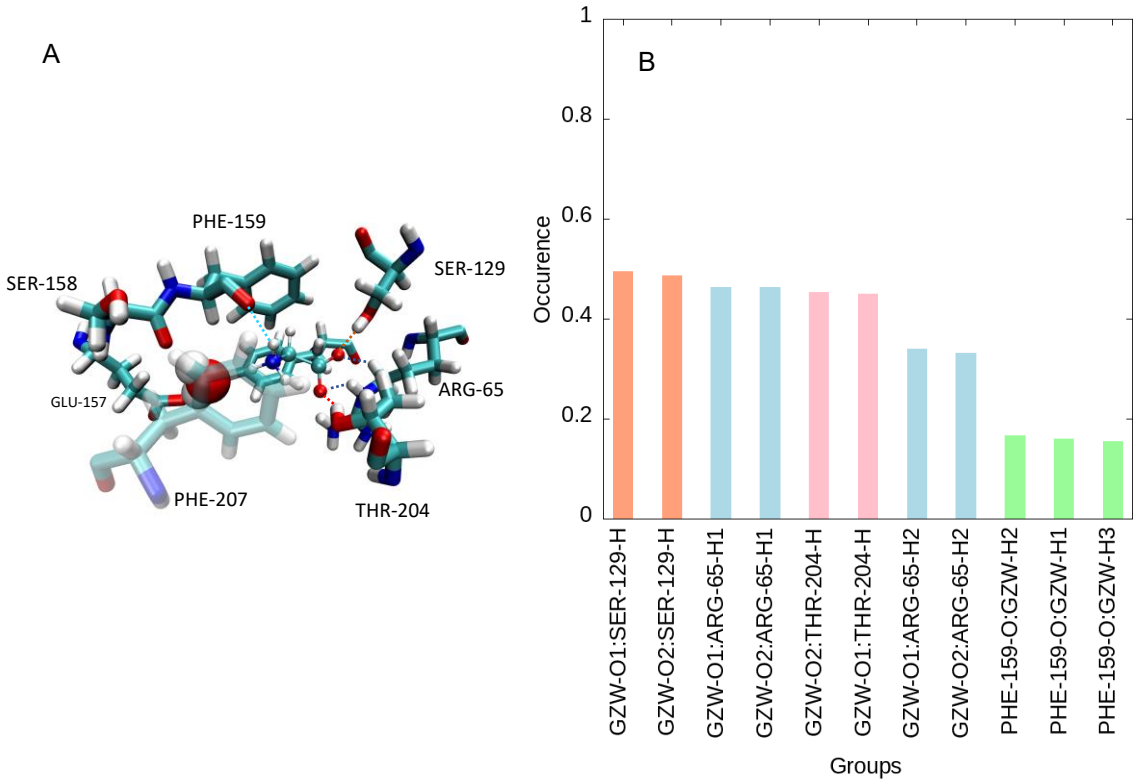

Figure 5.7: A, the hydrogen bonding over the course of the production including the types of interactions formed along with the proportion of the simulation they were present for. B, a labelled illustration of the binding pocket taken from a frame of the simulation.

5.3.4 Cation-π Interactions

Further examination of the pocket interactions to include cation-π interactions in Figure 5.8 shows a similar binding profile to that of the wild type system in Figure 3.18. The same two cation-π interactions are present but with a shift in the proportions for the PHE-207 interactions where the pockets seem to have a larger proportion of the MD with the interaction present. This could be indicative of a repositioning of the ligand closer to the outer wall of the pocket making this interaction more favourable compared to the wild type system. There is also a notable lack of cation-π interactions in pocket B, likely due to the early destabilisation of the pocket resulting in ligand B never forming the full profile of interactions required to keep it bound.

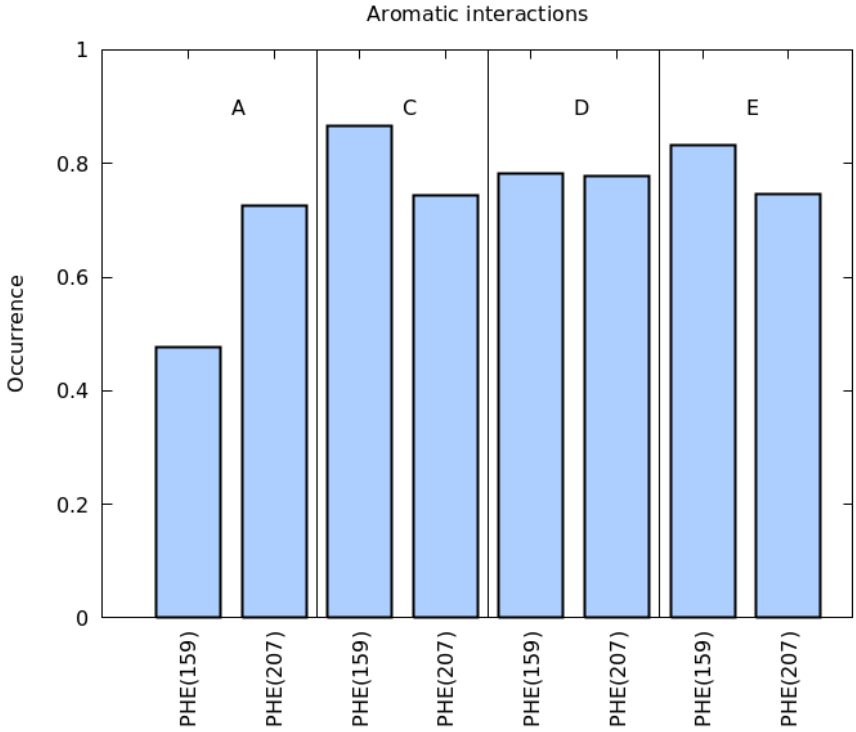

Figure 5.8: The cation-π interactions recorded for each binding pocket, plotted as a proportion of the simulation they were detected as present for. Binding pocket B showed no cation-pi interactions formed and therefore was left out of the plot.

The last set of measurements taken regarding the binding mode concern the positioning of both the ligand and the C-loop structure. By investigating these properties, we can see how the C-loop moves in relation to the pocket and whether it is in an open/closed state as well as understand how much the ligands are moving in the pocket, if at all. The movement of the C-loop was very limited for most of the pockets staying fixed at around 6 Å, but the binding pocket B C-loop appears to open and fluctuate regularly over the MD simulation, this finding fitting well with previous measurements highlighting its instability. The C-loop opening in Figure 5.9.A also aligns with the movement of the ligand in binding pocket B in Figure 5.9.B where an

ejection can be seen after around 125 ns. The movement of the ligand in binding pocket B from around 1.75 Å where the other ligands are almost fixed, increases until the ligand is free to move away in the bulk solvent. Other than the movement of ligand B, the only other slight fluctuation in distance between the BP centre of mass and the ligand centre of mass is in binding pocket C which is not as drastic and only very temporary.

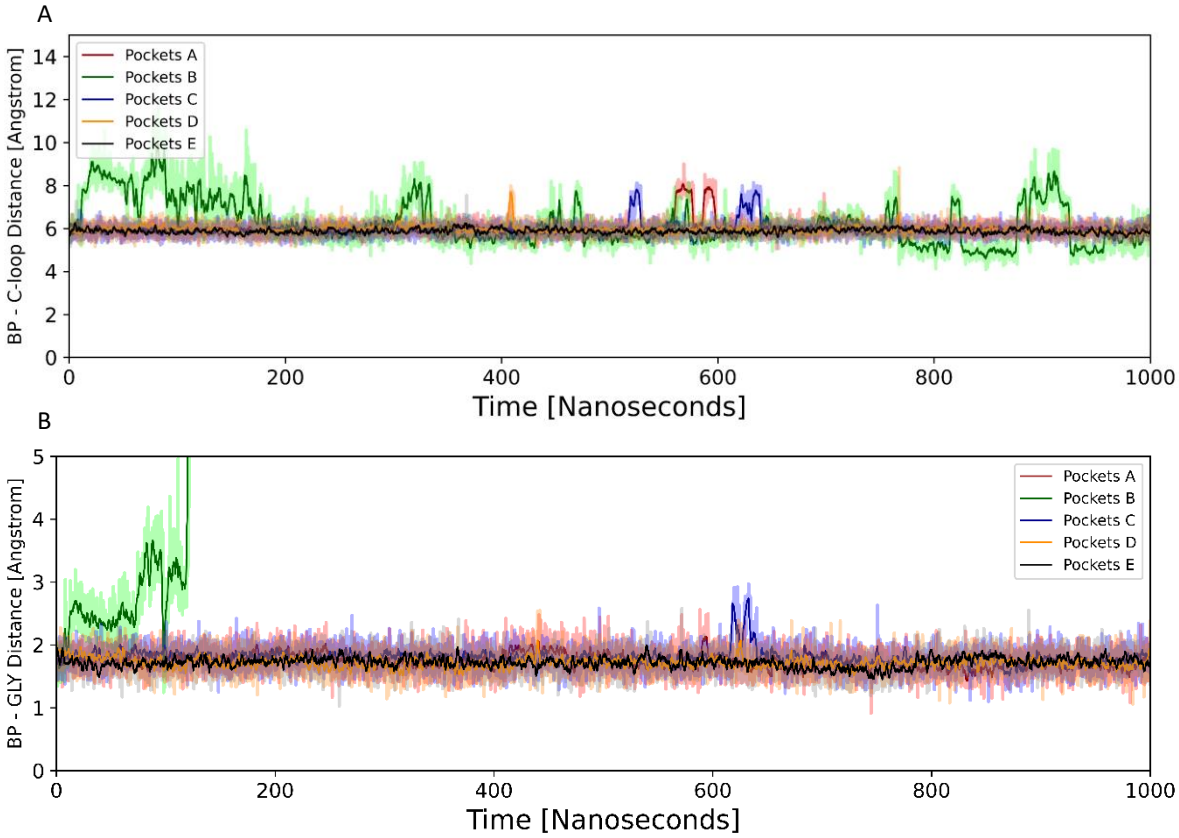

Figure 5.9: A, The distance measured between the centre of mass of the tip of the C-loop (THR-204) and the centre of mass of the binding pocket residues over time. B, The distance measured between the centre of mass of the ligand and the centre of mass of the binding pocket residues over time.

5.3.5 Comparing residue 46

The primary aim of this chapter is to explore the impacts of the N46K mutation and therefore it is important to quantify and investigate how the structure of the protein changes between systems and in particular what happens at residue 46 when the ASN-46 is mutated into LYS-46. Looking at the wild type system, we can measure the interactions that take place over the course of the production between residues 46 and the protein surrounding it. ASN-46 is located below the binding pocket, as shown in Figure 5.10.A, and is positioned in a nexus of three asparagine residues on both subunits making up the pocket. These other two asparagine

residues form repeated hydrogen bonds with the ASN-46 residue, the strongest being that between ASN-46-NH and ASN-61-O onto loop D. The other interaction is slightly weaker and less consistent, forming between ASN-46-O and ASN-102-NH on loop A of the complementary subunit.

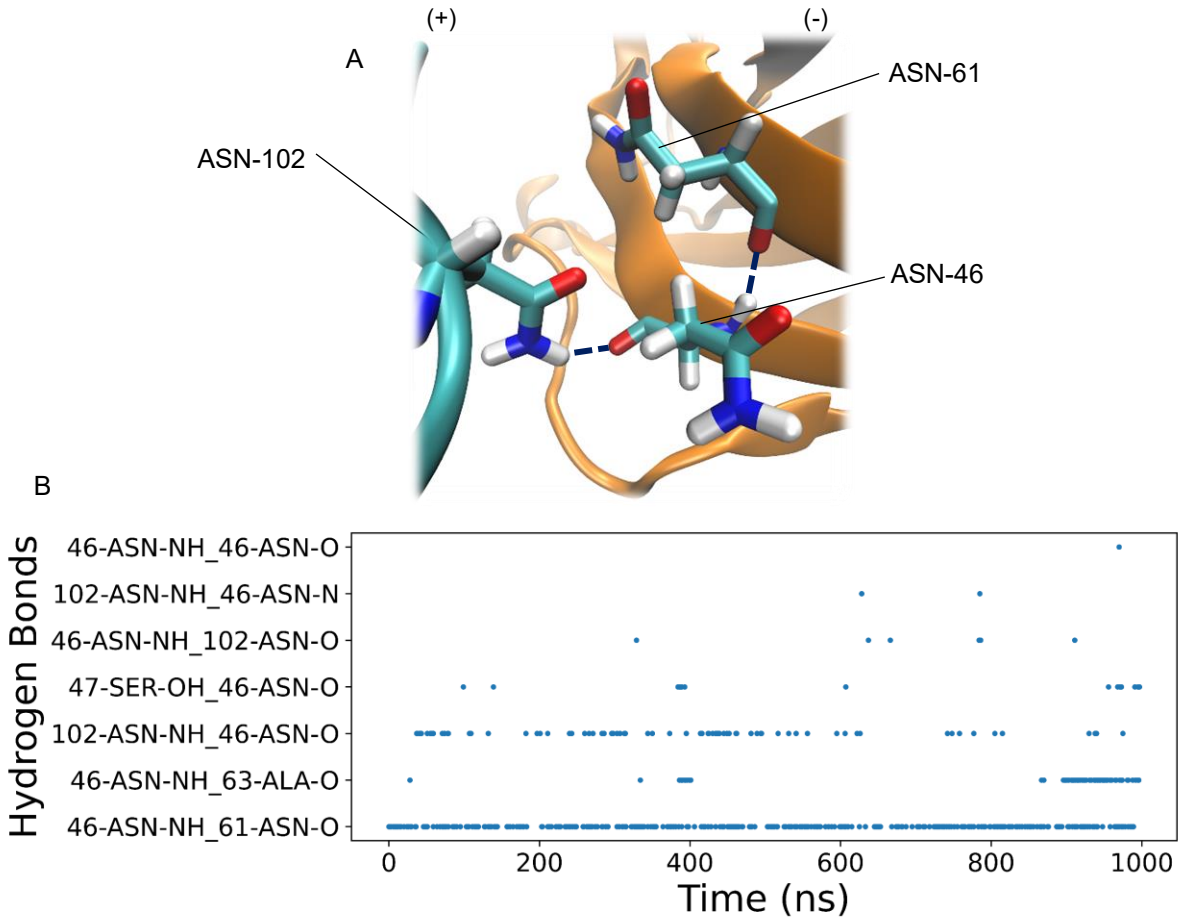

Figure 5.10: A, an illustration of residue 46 in between the subunits below the binding pocket, the interacting residues highlighted. B, the hydrogen bonding profile of the ASN-46 residue during the production of the wild type model.

These results can then be compared to that of the mutant system with LYS-46 in Figure 5.11 where the only consistent hydrogen bonding interaction is between the LYS-46 and ASN-61. The same interaction exists between LYS-46-NH and the ASN-61-O on loop D but the positioning of ASN-102 results in limited interactions with LYS-46.

There are some interactions both in the wild type system and mutant that are present linking residue 46 and the binding pocket residues; however, these only make up a small fraction of the MD, only existing for a few nanoseconds at a time. It can be concluded that the main impact, structurally, from replacing ASN-46 with LYS-46 is the interference with the ASN-102 hydrogen bond linking the subunits, potentially having broader implications to how binding/unbinding events proceed, the simulations carried out to explore this and their results will follow in the next section of this chapter.

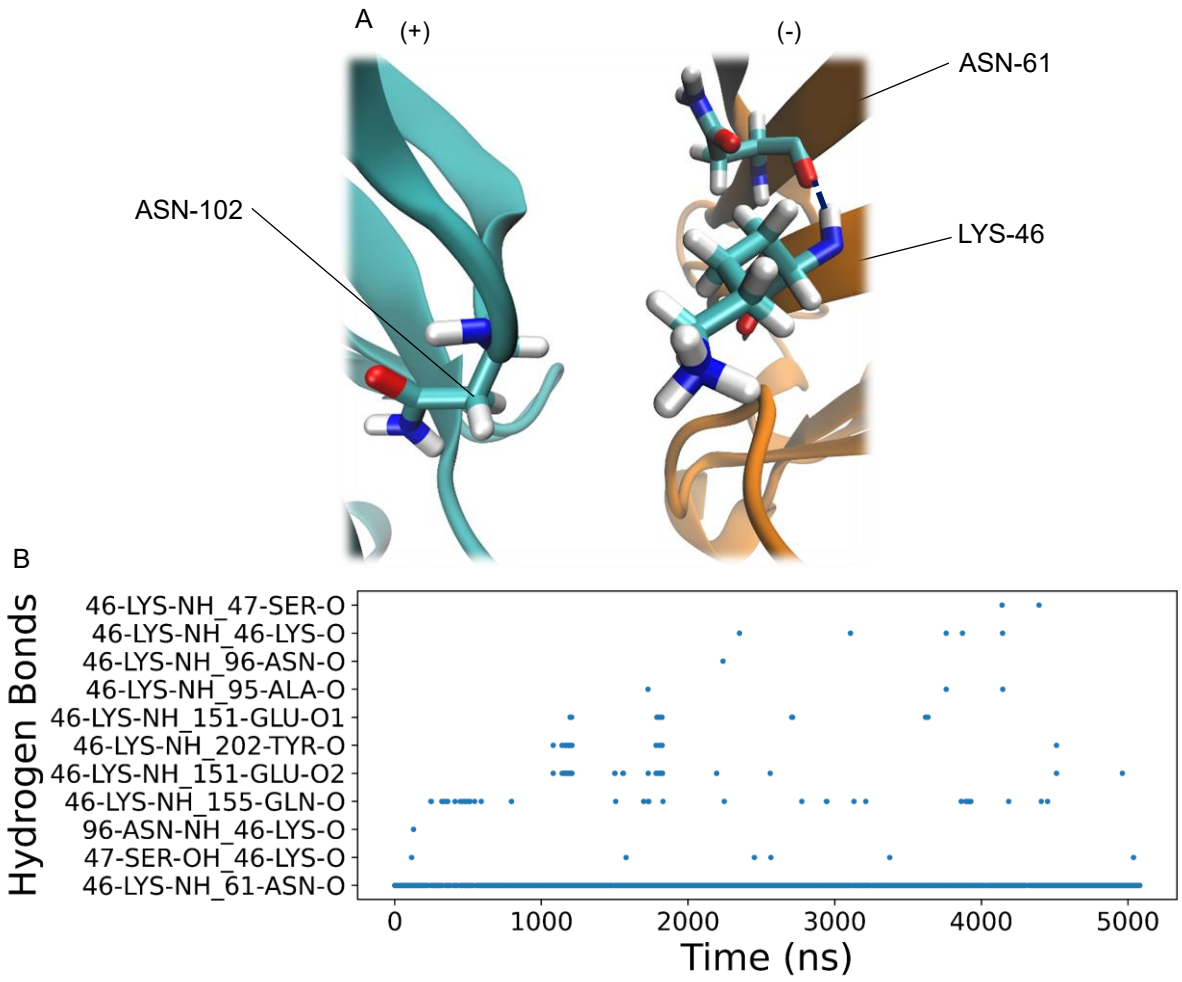

Figure 5.11: A, an illustration of residue 46 in between the subunits below the binding pocket, the interacting residues highlighted. B, the hydrogen bonding profile of the LYS-46 residue during the production of the wild type model.

5.3.5 Conclusions

The system described here has been shown to demonstrate suitable stability and an expected binding profile for a glycine bound receptor. Binding pocket A, C, D and E show stable RMSD values, closed C-loop structures the typical types of hydrogen bonds formed between their respective ligands and binding pocket residues. Here we see ,again, the water molecule located in a relatively fixed position in the binding pocket where water bridges form between the ligand and binding pocket, in line with what we observed in the 6PM5 wild type system. Binding pocket B does however show a large amount of instability, eventually devolving into an ejection that removes the ligand from the pocket. The results shown here however point to pocket A, C, D and E being suitable for moving on with enhanced sampling to characterise binding/unbinding events. As this is a mutant system, some comparative analysis was carried out regarding the residue changed, the main outcome of which showed that the N46K mutant results in less hydrogen bonding than the wild type, potentially having implications during binding/unbinding events that will be explored in the next section.

5.4 Funnel Metadynamics

The funnel metadynamics procedure used in this chapter utilised the same funnel setup as the previous chapter, using the trajectories taken from exploratory metadynamics as a guide to inform how glycine exits the binding pocket and applying appropriate parameters. To accurately compare the wild type and mutant systems it was important to ensure the funnel shape and positioning was identical to limit confounding effects of the methodology and since the system is almost identical the same parameters are appropriate.

5.4.1 Simulation Details

For the mutant system the same FMAP GUI tool was used to position the funnel, following the same protocol of defining an axis which had a trajectory following the ejection paths seen in the exploratory metadynamics. The same parameters as used in Chapter 4 were implemented here with the funnel axis origin set to roughly where the centre of mass of the binding pocket is positioned. The cone of the funnel was defined by an α angle of 0.45 rad, covering the axis region (z) from -5 Å to 25 Å. The cylindrical region of the funnel had the same radius R_{cyl} of 1 Å and a length of 10 Å. Harmonic restraining walls were applied in the same way at the end's

origin of the funnel cone and end of the funnel cylinder to prevent exploration of the ligand outside of the target area. The metadynamics was carried out with the same well-tempered metadynamics parameters, depositing gaussians ever 1 ps with an initial height of 1.2 Kj/mol, a target temperature of 310 K and a bias factor of 15. These simulations were carried out also in an NVT ensemble.

5.5 Comparing Results of the Wild Type to the N46K mutant

In this section the results taken from the funnel metadynamics simulations of the N46K mutant will be illustrated and discussed along with a comparison between the mutant variant system and the wild type system first describe in Chapter 3. Firstly, the analysis of the funnel metadynamics simulation stability will be covered, along with measurements made to quantify convergence.

5.5.1 Funnel Metadynamics Stability

The stability of the structure during funnel metadynamics and the convergence of the simulation are both important aspects to be considered before any conclusions can be drawn upon the results. The same process described in Chapter 4 is repeated here to assess how stable the protein structure was during funnel metadynamics and identify any behaviour that could indicate destabilisation of the greater protein structure not being targeted directly by the metadynamics biasing scheme. The first checks carried out were measurements of the RMSD of the whole protein, along with averaging of the RMSD values for important substructures. The whole protein RMSD shows both systems plateau to a similar level

and remain stable over the course of the funnel metadynamics with only a small gradual increase (Figure 5.12.A).

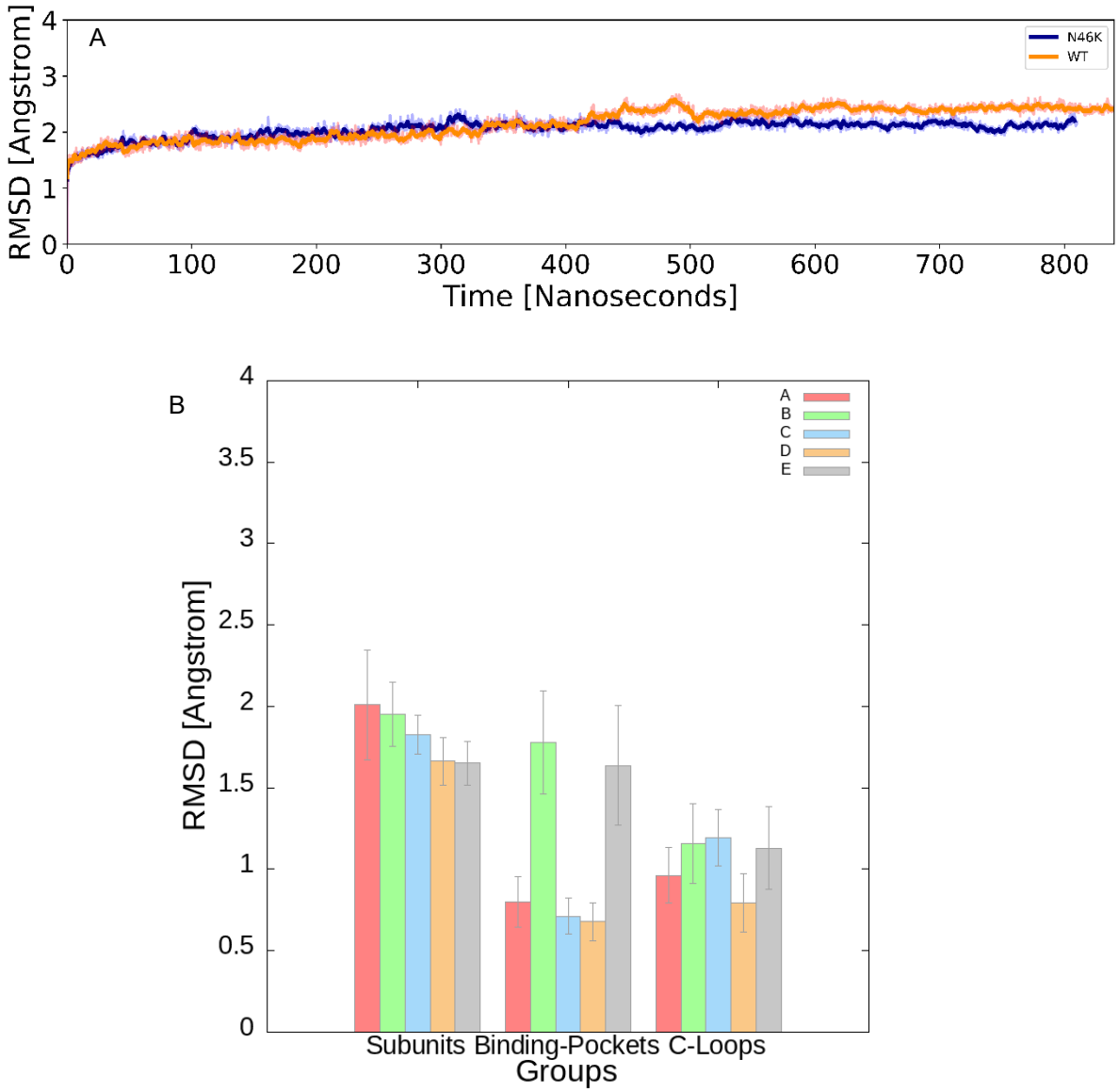

Figure 5.12: A, The RMSD of the protein backbone as a function of time during the whole funnel metadynamics simulation with both the 6PM5 and N46K models plotted. The average RMSD values for the important substructures identified where plot B, shows the values measured for the N46K model.

The average RMSD values show a similar conformity between the systems but highlights the areas more transient and specific movements that are not captured in the whole protein RMSD plot. There is an increase in the RMSD of the binding pockets for both pocket E and B in the mutant system (Figure 5.12.B). The increase in RMSD for binding pocket E is expected given that this is the pocket targeted in the mutant system, however binding pocket B showing increased RMSD values is indicative of instability which is likely a continuation of the destabilisation seen in the production. There is an increase in RMSD for the targeted binding pocket E that is not seen in the wild type system, suggesting reduced stability in that pocket during this process that was not observed in the production however this cannot be determined without the more in-depth measurements and observations of the binding pocket discussed later in this chapter.

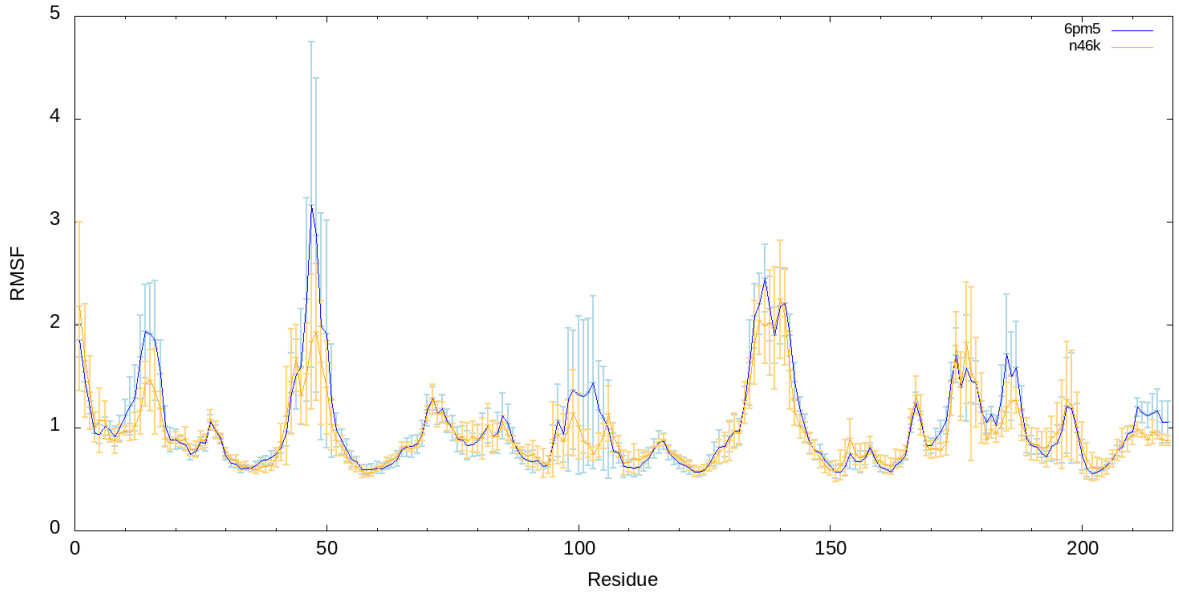

Figure 5.13: The RMSF values for the backbone atoms of residues in the 6PM5 system (Blue) and the N46K system (orange). The data here has been averaged across the 5 subunits to demonstrate general differences between the models, standard deviation has been plotted as error bars.

To understand better the differences between the two structures over the course of the funnel metadynamics, as well as check for any structure specific changes outside the directly binding related regions the RMSF was plotted for both the wild type and mutant system. The RMSF shows very few major differences between the structures which is what would be expected due to most residues not being actively biased by metadynamics. The main conclusion that can be drawn from this, along with the other RMSD measurements made, is that both systems are suitably stable and can ultimately be used effectively to characterise binding/unbinding events.

5.5.2 Funnel Metadynamics Convergence

When deciding that a metadynamics simulation has converged, the two important factors to consider are the movement of the ligand and the stabilisation of the binding free energy. Both have been measured here, in the same way as in Chapter 4, to determine the convergence is sufficient to verify any results found. The calculations were terminated and considered converged when multiple crossings of the ligand between bound/unbound states had been made and the free energy difference between the states plateaued. Looking at the CV plot in Figure 5.14.A we can see there are many crossing events observed and a smoothening of the movement of the ligand as the simulation comes to an end indicating less impeded movement by the flattening of the free energy surface from bias deposition.

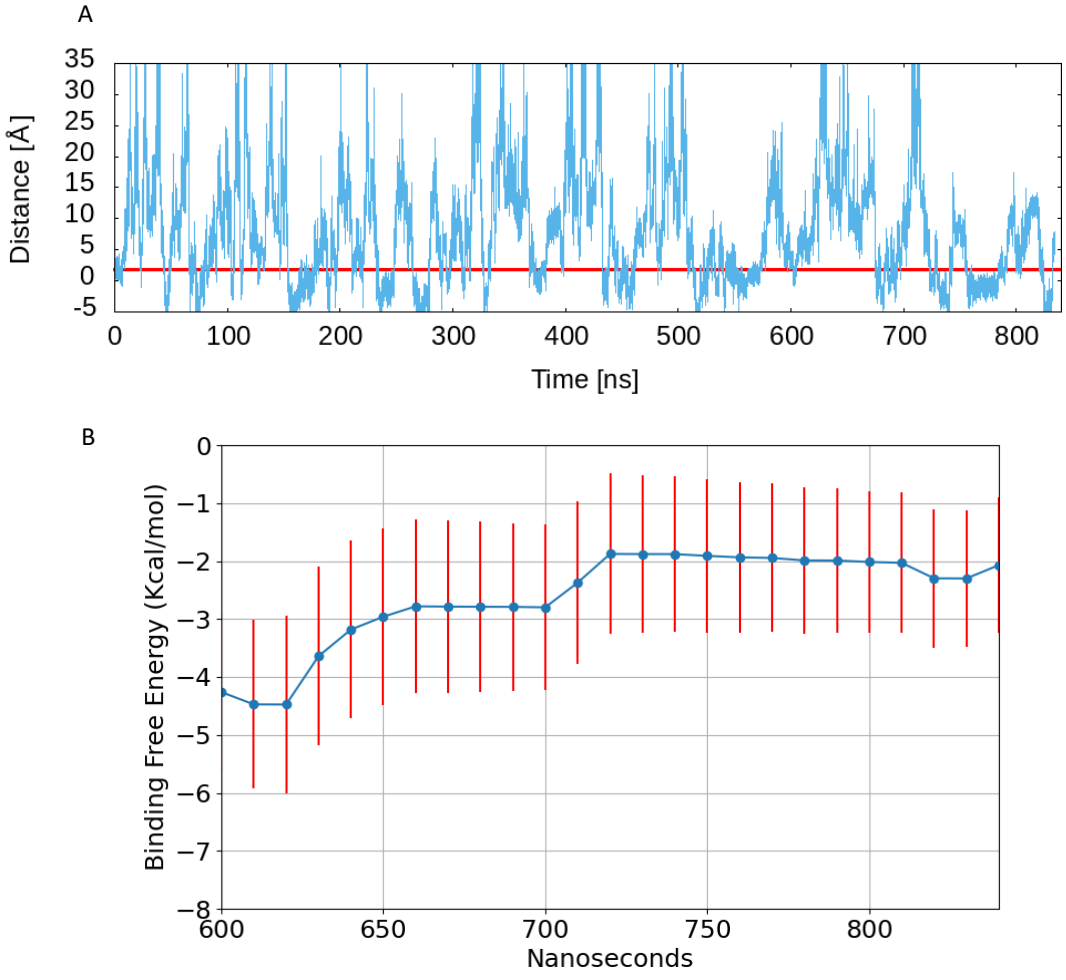

Figure 5.14: Plots taken from the funnel metadynamics to measure convergence. A, The CV fps.lp shows the movement along the z axis of the funnel in Å measured over the course of the simulation, the red line denotes the binding pocket position. B, The binding free energy over the course of the second part of the funnel metadynamics stabilising, standard error is plotted in red.

The measurement of the binding free energy over the last few hundred nanoseconds of the simulation, Figure 5.14.B, shows how the binding free energy plateaus over time, reaching a converged state.

5.5.3 Funnel Metadynamics Convergence

The free energy maps as a function of the two CVs and the profiles projected onto the funnel axis of both systems are shown in Figure 5.15. Major differences can be observed in the 2-dimensional plot, showing mainly that the mutant system (Figure 5.15.A) demonstrates much weaker glycine binding than the wild type (Figure 5.15.B). The same binding pocket well and pre-binding pose well can be seen in the wild type at **A-WT** and **B-WT** respectively however the mutant demonstrates only a single strong binding pocket well at **A-N46K.** This is further exemplified in the 2-dimensional free energy plot in Figure 5.15.C where in blue the WT can be seen to have a much more potent binding effect when compared to the plot of N46K. The result of these measurements finding the binding free energy of the N46K system to be -2.07 ± 1.16 Kcal/mol, an over 6 Kcal/mol drop from that of the wild type.

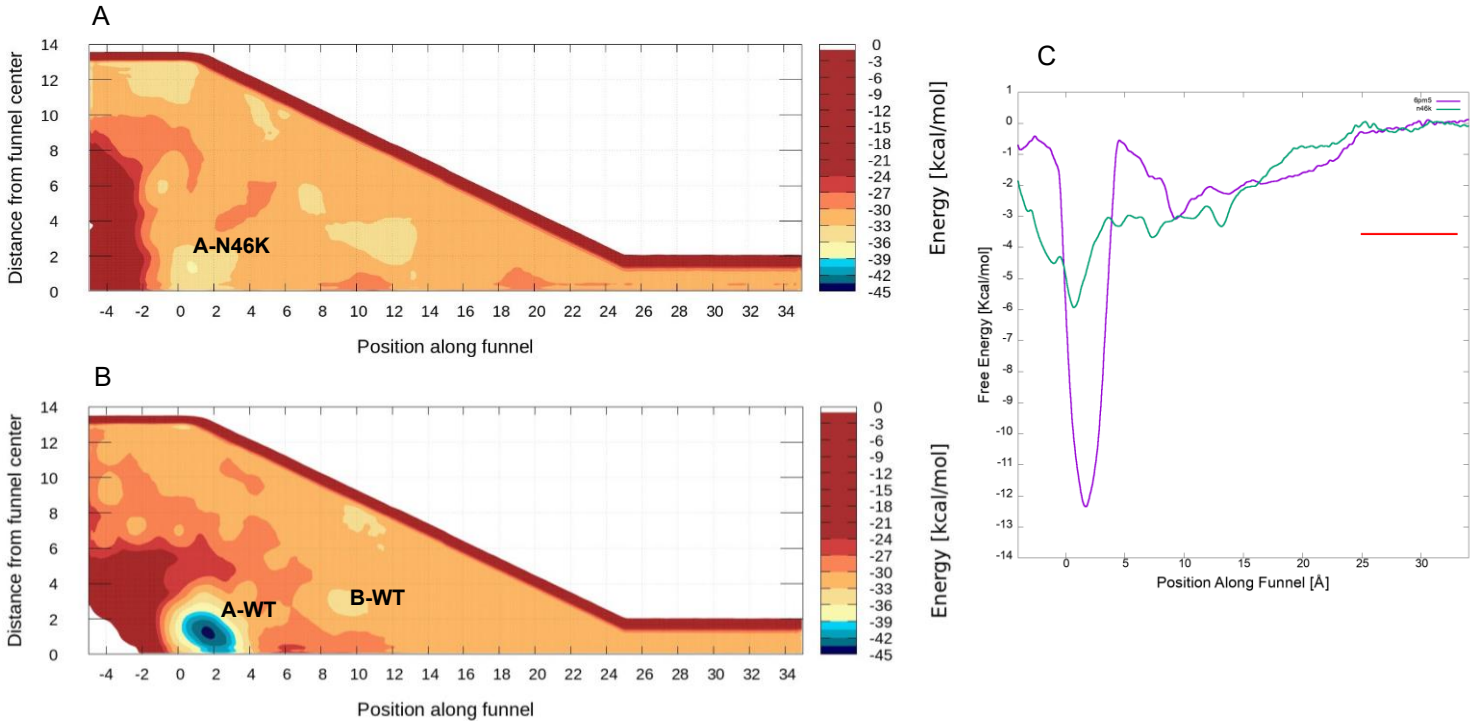

Figure 5.15: The free energy maps as a function of the CVs of the wild type, B, and mutant models,A. Projected free energy profiles as a function of the ligand position along the funnel axis, C, with the correction due to the cylindrical part of the funnel illustrated by a red bar.

Given the known consequences of this mutant impeding glycine receptor function, these results were to be expected, but provide further details at the molecular level which cannot be easily obtained from experiments.

5.5.3 Reweighting the Free Energy Surface

Again, using a select number of CVs limits the visualisation of the free energy surface to a narrow band of dimensions and so the reweighting procedure was applied again here as it was in Chapter 4. Expanding the observation of the free energy surface to the full three-dimensional coordinate space, using the ligand coordinates provides a clearer view the free energy across the conformational space in and around the pocket. The top two plots, Figure 5.16.A-B, are the reweighted plots for the N46K system and the wild type plots are shown below, Figure 5.16.C-D. Here the binding pocket well is replicated in both systems, along with the pre-binding pose for the wild type discussed in Chapter 4. The explored space outside of the pocket is shown in greater detail, highlighting a potential binding poses, as seen by the wells outside the pocket, like the wild type again but also much weaker than in the wild type, this binding pose is also much more solvated and open than that seen in the wild type.

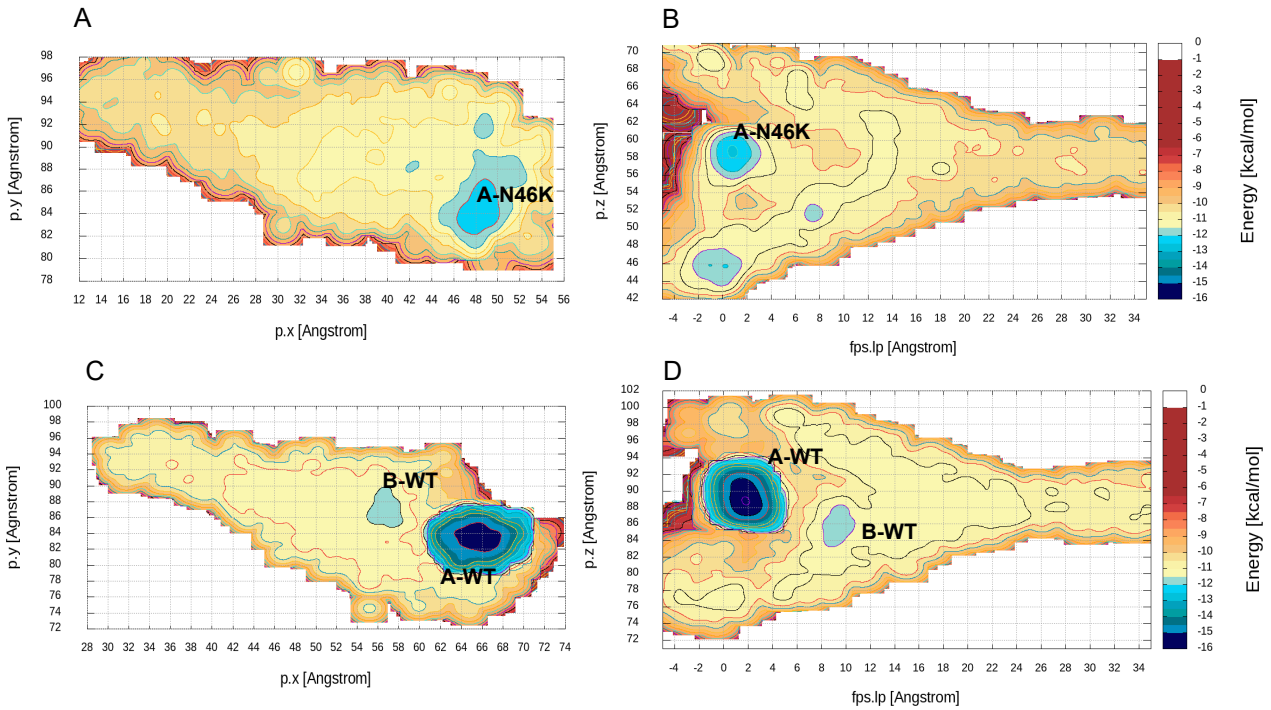

Figure 5.16: The reweighted free energy maps showing the funnel metadynamics reweighted from the fps.lp and fps.ld CVs onto the position of the ligand along the x (p.x), y (p.y) and z (p.z) coordinates. A and B show the reweighted free energy surfaces for the N46K mutant whilst C and D show the reweighted plots for the wild type model.

These results also confirm the idea that the mutant does indeed alter binding affinity, as a clear reduction in binding affinity is demonstrated here. The mechanics of how this mutant is leading to such a reduction in binding free energy requires further investigation and observations of the binding residues and the activity of the N46K mutant during this process which will be discussed further in the next section of this chapter.

5.5.4 Glycine Unbinding from the N46K GlyR Mutant

To understand more about the binding/unbinding process and the path the ligand takes, the movement of the ligand in the first ejection observed in funnel metadynamics can be utilised. The movements of the ligand were visualised as it ejects from the pocket using the same procedure as shown in Chapter 4. The metrics used to quantify the ligand distance from the centre of the pocket as well as the motions of the C-loop were also employed alongside the measurement of the hydrogen bonds between the ligand and the protein residues.

The unbinding trajectory alongside the free energy landscape produced from funnel metadynamics provides a clearer and more complete picture of the dynamics and the changes that mediate the unbinding of the ligand. At the beginning of the trajectory shown in Figure 5.17.A, over the next page, the ligand is positioned in the pocket in an almost identical way to the glycine wild type illustrated in Figure 4.22.A. The carboxylate moiety is positioned close to the entrance of the pocket interacting with the THR-204 residue on the C-loop, and the ARG-65 residue. There is a slight increase in the favourability in the mutant type for the ligand carboxylate to interact with SER-129 as it appears more frequently and for longer periods in the starting position of the mutant system. The ammonium moiety is positioned on the other side of the pocket, interacting with a water molecule that fits in between the ligand and the GLU-157 and SER-158 residues forming a water bridge. Whilst there is some rotating of the carboxylate moiety, the first largest change also happens at around 7ns depicted in Figure 5.17.B where we see a rotation of the PHE-63 residue at the back of the pocket, along with the starting destabilisation of the ligand. Here the ligand begins to move around in the pocket space towards the SER-129 residue. These movements in the pocket are coupled with a disruption of hydrogen bonding seen in Figure 5.17.G and the start of C-loop opening in Figure 5.17.H. At 7.2ns the C-loop opening and ligand movement towards the entrance of the pocket continue (Figure 5.17.H), along with further rearrangement of the positioning of PHE-63 and PHE-153 at the inner wall of the pocket (Figure 5.17.C).

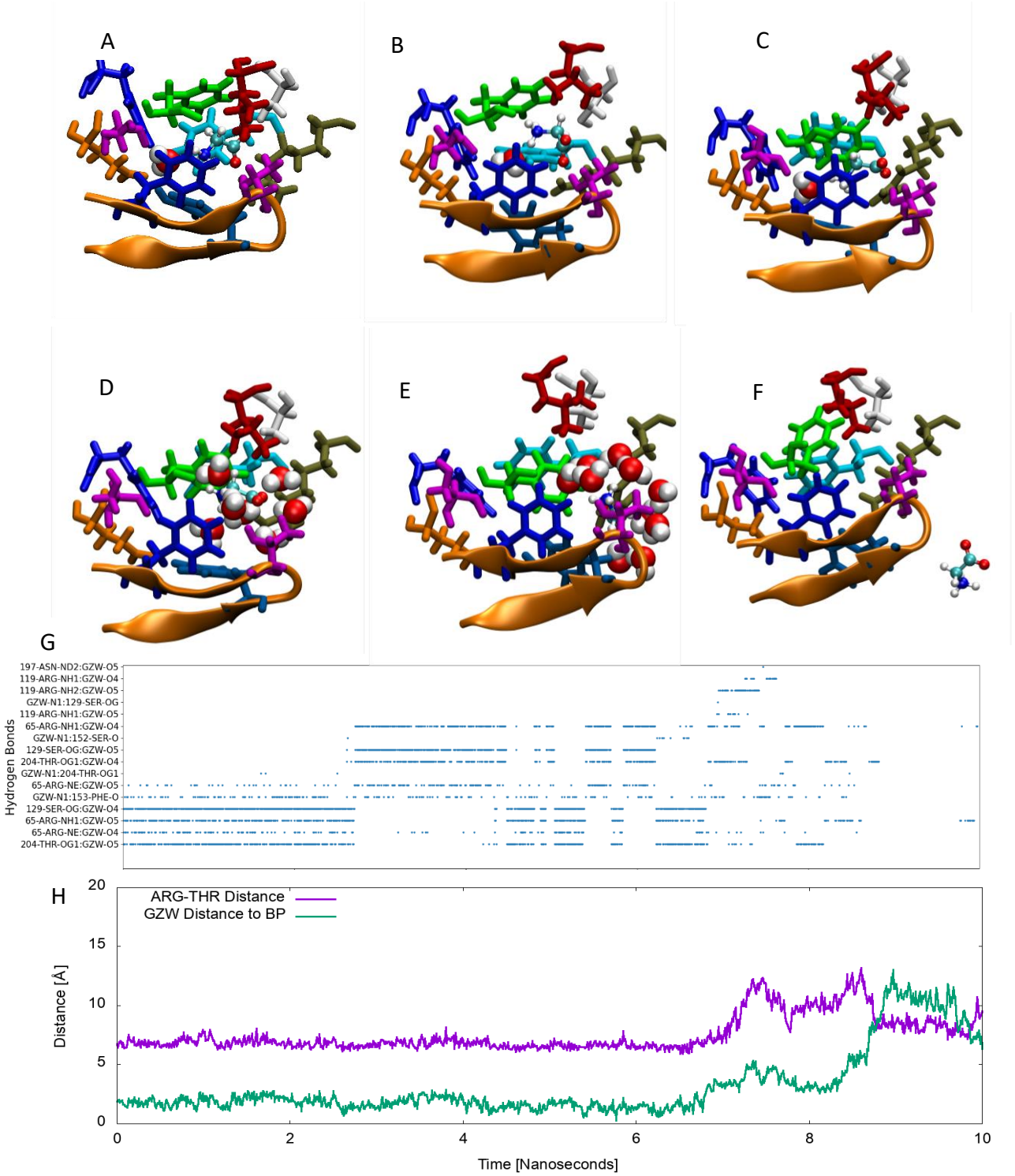

Figure 5.17: A-F show the binding pocket residues and C-loop structure over the course of the ejection trajectory. The residues shown are PHE-99:upper-blue, GLU-157:orange, SER-158:magenta-left, PHE-159:green, TYR-202:navy, THR-204:magenta-right, PHE-207:lower-blue, PHE-63:cyan, ARG-65:khaki, LEU-117:red, SER-129:white. G, shows the hydrogen bonding over the course of the ejection period with the type of hydrogen bond labelled on the Y axis. H, is a plot of distances both for the C-loop opening metric in purple and the distance between the COM of the ligand and the COM of the BP in green. The first unbinding event was used for these analyses.

The C-loop continues to open as the system reaches 7.4 ns, with the C-loop now open, water molecules rapidly move into the pocket space surrounding the ligand an further inhibiting

ligand-protein hydrogen bonds from forming consistently (Figure 5.17.D,G). From this point onwards the only consistent hydrogen bonds forming are between the carboxylate moiety of the ligand and the ARG-65 and THR-204 residues at the entrance to the pocket (Figure 5.17.G). Leading up to 8.5 ns (Figure 5.17.E), the ligand moves further away from the centre of the binding pocket (Figure 5.17.G), flanked by water molecules until the only two interactions remaining between the ligand and the pocket are with ARG-65 and THR-204. The C-loop begins to close again after this point and the ligand becomes fully solvated and free to move in the bulk water (Figure 5.17.F).

5.5.5 The N46K Residue and Unbinding

To consider the role that the N46K mutant had in the results of these simulations, the movement of the residue and its interactions can be visualised and quantified during the ejection trajectories discussed in section 5.5.3 and Chapter 4. For the wild type, when visualising the movements of the N46 Asparagine residue, there is a singular stable pose that is maintained throughout the unbinding process. Consistent hydrogen bonds form between ASN-46 and ASN-61, as in the production. However, the hydrogen bond between ASN-46 and ASN-102 seen intermittently in the production (Figure 5.10) is not present here.

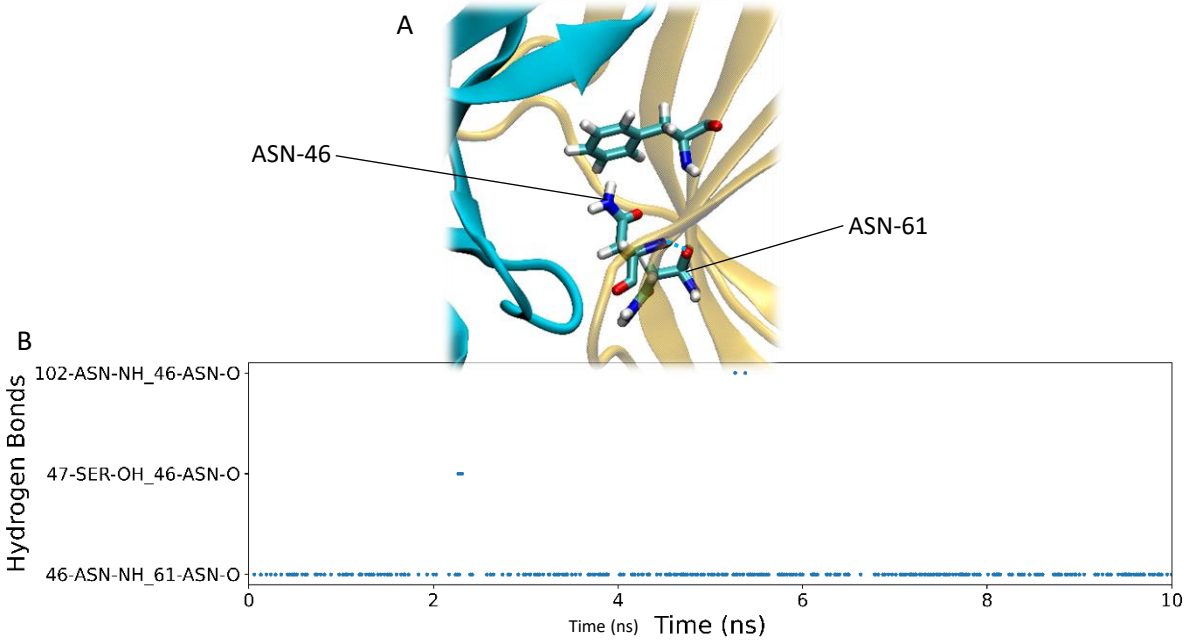

Figure 5.18: The wild type glycine receptor model, with the Asparagine targeted by the N46K mutation highlighted along with residues that it forms interactions with and the closest binding pocket residue for reference: PHE-63. A, during the ejection period in Figure 4.23. The hydrogen bonding profile for the N46 residue is plotted in B.

The N46K mutant system possesses a very different profile of hydrogen bonds and positioning. Whilst the interaction between ASN-46 and ASN-61 is reproduced here between

the LYS-46 and ASN-61 (Figure 5.19.A-B) across the loops of the subunit (-), there is a change later in the ejection trajectory that puts LYS-46 in a position to directly influence a critical binding pocket residue. This interaction is seen forming in the hydrogen bonding plot in Figure 5.19.C where at around 8.5 ns the LYS-46 ammonium moiety forms an interaction with the GLU-157 residue carboxylate moiety. This interaction is consistent and maintained for the rest of the ejection, the positioning change highlighted between Figure 5.19.A and 5.19.B showing the start and end points of the LYS-46 residue respectively.

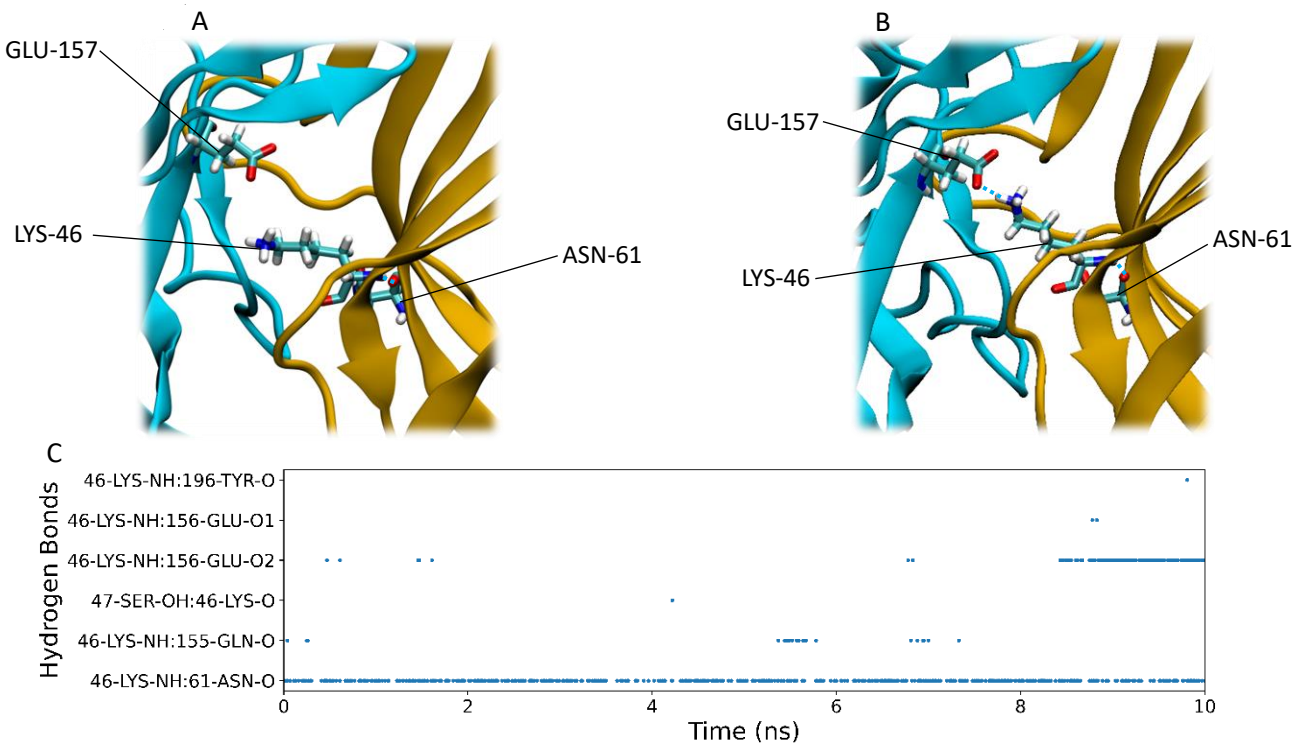

Figure 5.19: The mutant glycine receptor model, with the Asparagine mutated to Lysine by the N46K mutation highlighted along with residues that it forms interactions with in A. A second pose during the ejection process is also identified along with the key residue interaction highlighted in B. These results are taken from the ejection period in Figure 5.17. The hydrogen bonding profile for the K46 residue is plotted in C.

5.5 Discussion

This section will discuss the results described in this chapter, including data collected with the molecular dynamics simulations and the funnel metadynamics. The results presented so far have provided an in-depth view into the impact of the N46K at the atomistic level. The following sections bring together these results to explore the insights gained from simulating this mutant system and discuss implications and understanding gained about the mechanisms through which the GlyR functions by comparing the wild type and mutant systems.

5.6.1 The Mutant Glycine Receptor system

The results in this chapter have shown the binding of the mutant system to possess very similar properties to that of the wild type discussed in Chapter 4. The binding pocket interactions often cited from structural experiments and other simulations[16,26,33,137], involving ARG-65, THR-204, SER-129 and PHE-159 are all present in the mutant system along with the same positioning suggested from cryo-EM data. This was expected due to the cryo-EM data used to develop both the wild type and the mutant system are the same (PDB:6PM5[26]). No major differences were expected in the resting binding pocket either since residue 46 is not a principal binding pocket residue. Because there are no N46K mutant specific cryo-EM systems or other computational systems to use as a reference, determining whether the structure accurately systems the N46K mutant stems from observations made in the production and comparing it to the wild type system. The results from the production molecular dynamics simulations show very similar levels of stability in the dynamics, similar binding pocket hydrogen bonding and cation-π interactions. The similar properties of the binding pocket and structure of the wild type and mutant systems suggests the N46K mutant acts in a nuanced way during binding/unbinding, rather than causing any larger scale changes to the protein structure that would likely have been observed in the molecular dynamics. This is particularly of interest given the dramatic changes in the free energy landscapes, demonstrating how delicate the potent binding of the wild type is. During the production in the mutant system there is some destabilisation in binding pocket B, leading to an ejection of the ligand, highlighted in the average RMSD values collected for the sub-structure (Figure 5.3) and the plot of BP-GZW distance (Figure 5.8.B). This instability and the ejection of a ligand is not seen in the wild type system and could suggest instability in the system but the other pockets remaining stable and the other properties of the structure mirroring that of the wild type likely suggests it was an isolated ejection event. Given the timescale involved in this destabilisation and ejection, less than 200 ns, it is difficult to link this to the effects of the

mutant; however, it would be plausible given the findings from the funnel metadynamics and ejection trajectory.

5.6.2 Funnel metadynamics and the N46K Mutant

The funnel metadynamics simulation carried out with the mutant system demonstrated a very different trajectory to that of the wild type. The changes in the pocket documented in Figure 5.17.A-F provides a series of snapshots illustrating the movement of glycine within the pocket of the mutant system during an unbinding event. The process starts with destabilisation within the pocket and increasing opening of the C-loop. Whilst the trajectory starts with only one water molecule within the pocket, as the C-loop opens, there is a movement of water molecules into the space and around the ligand, rapidly leading to its transit and ejection out of the pocket. The entrance of water into the pocket quickly leads to a deterioration in the hydrogen bonds between the ligand and the protein allowing it to move further out towards the ARG-65 and THR-204 residues at the exit of the pocket. Like the other systems, the disruption of interactions with the pocket residues, whilst maintaining interactions with the ARG-65 and THR-204 appears, to be a critical last step before the ligand progresses out of the pocket space. What is presented here is a mechanism for the transit of glycine out of the pocket in the mutant system. The path taken by the ligand allows for the identification of the key residues involved in mediating movement out of the pocket in this system; moreover, the results further exemplify the importance of water within the binding pocket during binding/unbinding events. The water bridge is a defining element of the pocket from the start, the destabilisation of which occurs while hydrogen bonds with the pocket residues starts to break down. Whilst these results provide a valuable mechanism for which glycine unbinds in the mutated form of the GlyR, understanding how the ejection has been influenced by the modification of ASN-46 to LYS-46 requires observations and measurements made around the changed residue which is discussed in the next section.

5.6.3 The Mechanism of Action of The N46K Mutant

The binding pockets of the wild type and N46K mutant behave drastically differently during funnel metadynamics. Whilst the starting point is almost identical, the ligand in the wild type is completely insulated from the bulk water for the duration of the trajectory until it is completely outside the pocket, even stopping in a pre-binding pose where water is still unable to interact with it. In the mutant system, the ligand initially destabilises in the pocket in a similar way to the wild type but when the C-loop begins to open there is an influx of water leading to a rapid breakdown of hydrogen bonding and ultimately unbinding. The inclusion of water into the

pocket suggests a drastic change in stability of the pocket structure, in a similar way that the size of GABA and increased pocket volume likely led to water influx discussed in Chapter 4.

In the case of the mutant system, the LYS-46 residue replacing ASN-46 leads to a difference in interactions below the pocket and between the subunits. During the production molecular dynamics the wild type ASN-46 residue forms intermittent hydrogen bonds both between subunits (ASN-46:ASN-102) and within its subunit across the binding pocket loops. The mutant variant LYS-46 is unable to form the inter-subunit interaction during the production and is characterised by only one stable interaction with ASN-61, positioned in the same way as the wild type. During the unbinding trajectory the wild type hydrogen bonding profile and positioning is constant, mirroring the behaviour from the production. The mutant variant, whilst maintaining the same LYS-46:ASN-61 interaction, rotates and forms a consistent inter-subunit hydrogen bond with the GLU-157. GLU-157 is one of the critical binding pocket residues for the GlyR, containing a carboxylate moiety that forms part of the water bridge with the ligand and maintains binding pocket stability. The mutant LYS-46 residue in this system may therefore be causing the destabilising effect on the binding pocket by acting at this residue. Whilst the interaction was not present in the molecular dynamics production, its intermittent formation during unbinding and then stabilisation post-ejection could be deforming the pocket and leading to the reduced binding affinity reported in experimental work[44].

The positioning of the binding pocket residues and the resulting movement of the glutamic acid are illustrated in figure 5.20.A-B. Here we see how in the wild type, the glycine forms a water bridge to glutamic acid, in the mutant system the LYS-46 residue interacts with this glutamic acid and displaces it, disturbing the water bridge and therefore reducing the likelihood of stable binding. The frequency of this interaction is demonstrated by a plot of hydrogen bonds LYS-46 engages in through funnel metadynamics in figure 5.20.C showing it is a regular interaction following ejection for a large part of the simulation.

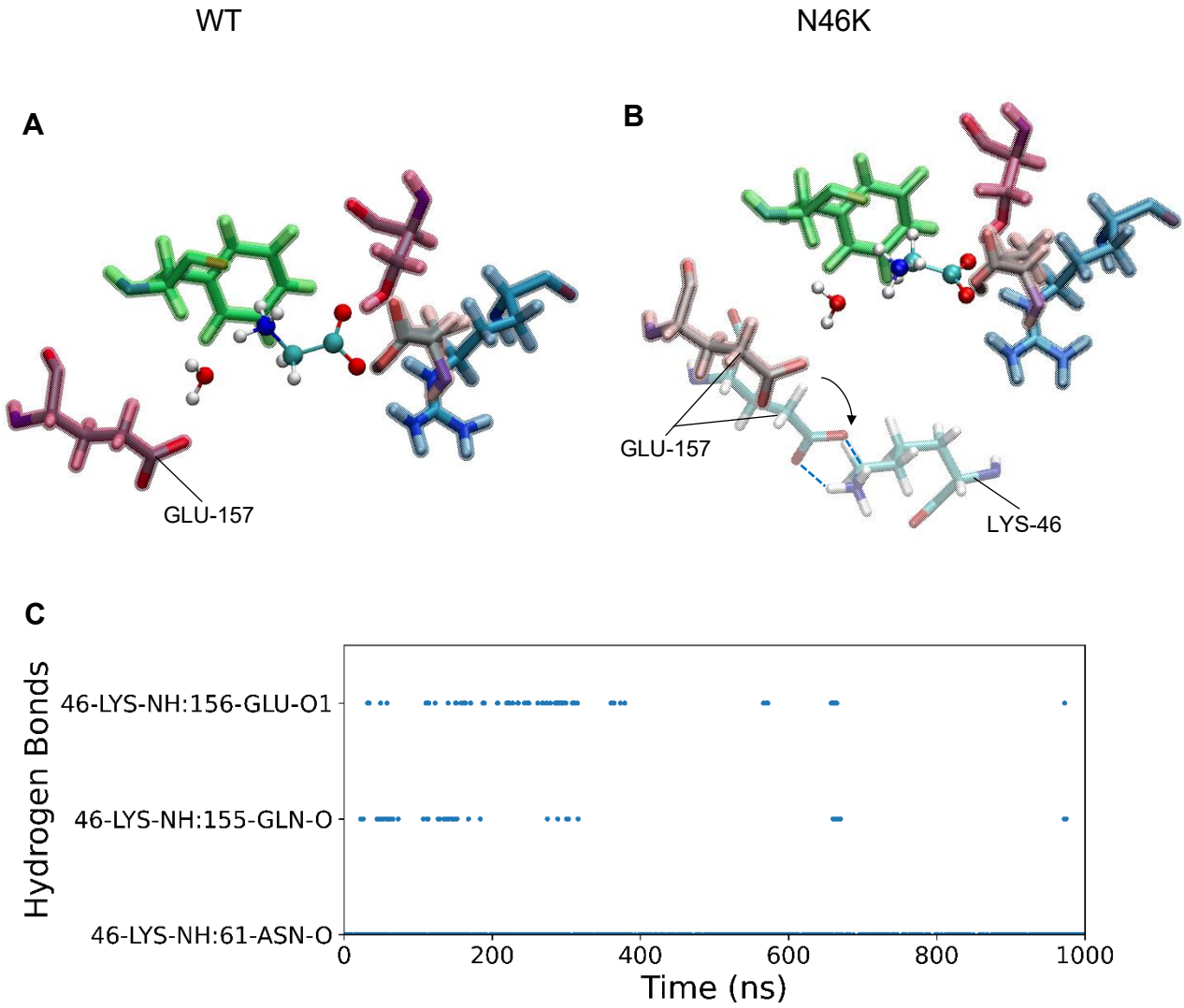

Figure 5.20: Snapshots of the binding pockets during funnel metadynamics with the wild type structure (A) and the mutant system (B). In B, the movement of the principal binding pocket residue glutamic acid is shown from the start of the simulation to the position it's in following during and following ejection, along with the mutant residue 46 (Lysine). The interaction this glutamic acid makes with Lysine-46 in the mutant system is demonstrated in C.

5.6.4 Conclusion

In conclusion, the data presented in this chapter provides detailed insights at the molecular level into the N46K mutant and the impact of substituting the Asparagine at residue 46 with Lysine. The work provides a detailed profile of the principal interactions between the protein and the ligand both in the wild type and mutant receptor. The results explain how the mutant variant leads to reduced affinity via its action on a key binding pocket residue and also puts forward a mechanism through which the glycine transits out of the pocket differently in the mutated structure compared to a wild type system. Understanding the mechanistic and dynamic properties of mutant residues is an important step in furthering understanding of

mutant disease phenotypes and how larger physiological effects are mediated. The results here provide a starting point for further experiments and research to investigate the role of the N46K mutant and other mutants that could be functioning in similar ways in other pLGICs, as the exact interactions and mechanical properties of residues, like the N46 residue, are unclear from cryo-EM data alone.

The work discussed in this chapter also further highlights the importance of considering the role of water in binding, the water bridge present in both the wild type and mutant system being a key component of the stable bound pose and unbinding trajectories. Moreover, the data collected supports the suggestions made in previous work[44] that there is an inter-subunit ASN-46:ASN-61 interaction involved in maintaining structural stability of the subunit interface and glycine binding site. The results here showing that the LYS-46 mutation eliminates this interaction, coupled with the destabilisation and alteration of the binding pocket during the ejection trajectory. This work sets the foundation for further investigation of mutants along the subunit interface for both the study of disease phenotypes and for the identification of residues like ASN-46 and ASN-61 that play an active role in mediating binding from outside the binding pocket itself.

Chapter 6

Conclusions

The focus of the work presented in this thesis has been a pentameric ligand-gated ion channel, the glycine receptor. Specifically, the ligand binding that inactivates this receptor has been the area of investigation, using molecular dynamics and enhanced sampling techniques to probe the dynamics of the binding pocket at rest and during unbinding events. The ultimate goal of this field of research is to elucidate the mechanisms of action of proteins, by studying the role of each component until a complete fundamental understanding is attained. The starting point for pentameric ligand gated ion channel activation is the binding of specific ligands, hence understanding how a ligand interacts with a protein binding pocket is an essential stepping stone to understanding the larger dynamics at play involved in receptor function. By utilising state-of-the-art computational techniques and the systems developed in this work, based on recently resolved structures, valuable insights have been gained that contribute to this understanding in several ways.

In Chapter 3, the ligand-protein interactions were studied with glycine bound receptors using molecular dynamics simulations. Two systems were developed to this end, the first including both the extracellular and the transmembrane domain based on an X-ray structure originally bound to strychnine (5CFB[15]), and the second including only the extracellular domain based on a cryo-EM structure bound to glycine (6PM5). Whilst the first system was not optimal for later work using enhanced sampling methods, both systems provided results that have implications for our understanding of glycine receptor ligand-protein interactions. The use of a system with a diverse range of stability levels for each of the five pockets provided an opportunity to identify important elements that when disrupted led to poor binding properties. In the case of the whole protein system 5CFB, with the original binding pockets optimal for strychnine but too large for glycine, we demonstrated how fluctuations in pocket volume can be a contributing factor to destabilisation of binding. Some pockets also allowed us to show how the loss of the binding pocket water molecule compromises the integrity of the glycine binding which is water-mediated, also leading to ejection. These results compared with 6PM5 cryo-EM data where consistent: stabilised binding was obtained in all pockets with secure water bridging and stable volume, optimized for the specific ligand, provides strong evidence for the important role of these binding pocket properties. This work also set up an effective system of the GlyR that would be used in later chapters for more advanced calculations and

in the projects of other group members currently researching different properties of the glycine receptors such as inter-subunit interactions and the consequences they may have on binding and subunit cooperativity. The original whole protein system also can serve as a foundation for future work to investigate lipid-protein interactions and how these may also play a role in stabilising the structure and in protein function.

The work in Chapter 4 took a specific look at the differences between partial and full agonism within GlyRs. The free energy landscape of receptors bound with both the full agonist (Glycine) and the partial agonist (GABA) were derived providing information about the differences in binding and affinity between the two by investigating the binding pocket wells and potential pre-binding poses. From the trajectories in the funnel metadynamics, the movement of the ligand was characterised along with the interactions that play a role in mediating unbinding/binding. Here a likely mechanism for the unbinding of both the full agonist and the partial agonist was identified through alongside a comparative analysis to understand the differences. The mechanisms proposed, incorporate findings regarding the key interactions and changes that occur through the transit of ligands out of the binding pocket. This demonstrates how full agonism relies on the presence of a single water molecule in the binding pocket and a process whereby water is excluded from the binding path almost entirely. The full agonist simulation also identified a novel pre-binding pose where glycine is flanked by protein residues outside the pocket prior to ejection into the bulk solvent. For partial agonism, it was identified that an increased volume along with water movement into the pocket during unbinding could explain why GABA has reduced binding affinity, regardless of having a similar binding profile at rest. The suggestion from this being that the increased length of the GABA ligand pushes on the pocket, deforming it and reducing its capacity to exclude water during unbinding. The inclusion of a water bridge in the pocket in the full agonism and not partial agonism also supports previous work suggesting the importance of water in the pocket and demonstrates how crucial the water bridge is for high affinity binding in pLGICs. The pre-binding pose identified in the full agonist system highlights GLN-177, GLY-174 and ASN-42 as having a greater importance than originally thought in the binding of the GlyR. Identification of this site and residues could explain in part how glycine has such a high affinity for the pocket. A pre-binding pose where the ligand can be captured and guided into the pocket from the bulk could be a factor in producing the high affinity glycine has for the GlyR. The pre-binding pose provides a foundation for further work that could investigate more precisely the role of these residues outside the binding pocket and highlights the important of residues that are not directly observed to interact with ligands from cryo-EM data collected alone. In general, the findings with both systems demonstrate the importance of considering both the role of water, and non-binding pocket residues for future research in the area of pLGICs.

The results in Chapter 5 utilised a mutant system that is implicated in hyperekplexia, the startle disease, a neurological condition that is characterised at a molecular level by GlyR dysfunction. The mutant is linked to reduced affinity of glycine in the receptors affected by the N46K mutant and while there have been suggestions as to how ASN-46 plays a role in binding pocket stability, no modelling had been carried out to understand the mechanism before now. This chapter compares the binding in an N46K mutant system with a wild type system with the intention of identifying important elements and dynamics that involved in the high affinity binding of glycine. A key result of this work was the observation of an inter-subunit hydrogen bond, from residue ASN-46 to ASN-61, in the wild type that was repeatedly formed throughout the production. This supports the suggestions made in prior experimental work that the interaction could be involved in maintaining binding pocket stability[44]. Whilst the recreation of the interaction in these simulations system provides support to this assertion, finding that the inter-subunit interaction from residue 46 is completely removed in the mutant system verifies the claim further. The mechanism through which the N46K mutant mediates its effect was however more complex than the removal of this interaction, a key result being that during unbinding a consistent hydrogen bond formed with the GLN-157 binding pocket residue. The data collected from the unbinding trajectory presents a sequential mechanism for the ligand ejection that mirrors the path seen for the partial agonist, involving movement of water into the pocket and interfering with the typical hydrogen bonding profile expected. The reduced ligand binding free energy combined with the observed trajectory and interaction of the mutant LYS-46 with GLN-157 provides a detailed potential mechanism of action for the N46K mutant, demonstrating how, when mutated, the inter-subunit hydrogen bond is replaced with a direct interaction with the binding pocket, altering its structure and likely leading to the instability and water inclusion during unbinding that was observed. These findings explain how the N46K mutant functions in GlyRs and highlight how important the residues outside of both the transduction pathway and binding pocket are in the functioning of the receptor, especially at regions such as the inter-subunit interface where small changes could lead to pocket destabilisation like that which was reported here.

The results presented in this thesis provides an excellent platform for future study, whether utilising the same systems or in the study of other pLGICs. Future studies would benefit from applying the insights gained here to better quantify the effects water molecules not captured by cryo-EM may be having during binding, whether in GlyRs specifically or in other systems based in the pLGIC family of receptors where similar phenomena may occur. The production of a series of replicas would also benefit the reliability of the results shown here. Whilst the ligand binding free energy comparison to the RDL work does support the results, additional

replicas would provide a stronger foundation for future work. The systems and data presented in this thesis also could be utilised moving forward for the application of more advanced analysis and specific techniques such as the use of kinetics calculations which could be carried out to quantify differences in binding affinity more accurately. Further work to understand better the differences between partial and full agonists could be carried out by use of a broader range of ligands, perhaps even comparing other similar receptors. Finding a consistent link between pocket volume, water bridge formation or even differences in how inter-subunit interactions behave could provide a more solid understanding of partial agonism in pLGICs, also using molecular dynamics and enhanced sampling. Moving on from the N46K mutant, the work here demonstrates the utility of investigating the mechanisms of mutant action with these methodologies. Exploring the impacts of more and different mutant variants would not only provide a mechanistic explanation for the mutant effects but also demonstrate how the wild type residue may be serving important functions as was demonstrated here. Along with this, an area that is explored by this work with the mutant is the role of inter-subunit interactions. Carrying out more work focusing on mutant variants of residues at this inter-subunit interface could yield greater insights into the way in which pockets interact with each other and influence receptor function as a whole. Overall, the various elements of this thesis have contributed to the understanding of GlyR binding and pLGICs in general.

Acknowledgements

The work presented here would not have been possible without the support and encouragement of a great many people, no one more supportive and caring than my partner Hannah, for which this Ph.D certainly wouldn't at all have been possible without.

For whom the most time and dedication came from my supervisor Carla Molteni, whose guidance and many hours of discussion has been invaluable.

I would also like to thank my secondary supervisor Lucia Sivilotti, for providing the advice, perspective, and time for discussion necessary to navigate the challenges of an interdisciplinary Ph.D.

I must also show my utmost gratitude to Alessandro Crnjar, who during his final year in the Molteni group dedicated invaluable time to teach me the basics of molecular dynamics that would serve as a foundation for all the skills learnt and work produced over the course of this Ph.D.

Many thanks also go to the many collaborators and colleagues who have provided many beneficial discussions both of a technical and conceptual nature regarding the work. It also goes without saying my gratitude to the many dedicated members of King's College London who have built an environment that fosters collaboration and growth. I have been lucky to be a part of such a dedicated and passionate community at KCL.

A special thanks also is required for both the staff and students at LIDo. Being a part of the LIDo community has been a privilege and the support provided by the programme manager Nadine Mogford and the KCL institutional Lead QueeLim Ch'ng during my Ph.D has been integral in my personal and professional development.

Lastly, this thesis is the culmination of over 10 years of personal aspirations and each step of the way family, friends, mentors and colleagues have contributed enormously to the person and researcher I have become.

Bibliography

1. Rizzolatti, G., Fabbri-Destro, M., Caruana, F. & Avanzini, P. System neuroscience: Past, present, and future. *CNS Neurosci Ther* **24**, 685–693 (2018).
2. Bear, M. F. *Neuroscience: Exploring the Brain*. *Neuroscience: exploring the brain* (Wolters Kluwer, Philadelphia, 2016).
3. Südhof, T. C. Molecular Neuroscience in the 21st Century: A Personal Perspective. *Neuron* **96**, 536–541 (2017).
4. Evans, W. H. & Martin, P. E. M. Gap junctions: Structure and function (review). *Mol Membr Biol* **19**, 121–136 (2002).
5. Kesters, D. *et al.* Structural basis of ligand recognition in 5-HT $_3$ receptors. *EMBO Rep* **14**, 49–56 (2013).
6. Roux, B. Ion channels and ion selectivity Essays in Biochemistry. *Essays Biochem* **61**, 201 (2017).
7. Raghavan, M., Fee, D. & Barkhaus, P. E. Generation and propagation of the action potential. *Handb Clin Neurol* **160**, 3–22 (2019).
8. Du, J., Lü, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature 2015 526:7572* **526**, 224–229 (2015).
9. Wu, Z. S., Cheng, H., Jiang, Y., Melcher, K. & Xu, H. E. Ion channels gated by acetylcholine and serotonin: structures, biology, and drug discovery. *Acta Pharmacologica Sinica 2015 36:8* **36**, 895–907 (2015).
10. Connolly, C. N. & Wafford, K. A. The Cys-loop superfamily of ligand-gated ion channels: the impact of receptor structure on function. *Biochem Soc Trans* **32**, 529–534 (2004).
11. Gielen, M. & Corringer, P.-J. The dual-gate model for pentameric ligand-gated ion channels activation and desensitization. *J Physiol* **596**, 1873–1902 (2018).
12. Absalom, N. L., Liao, V. W. & Chebib, M. Ligand-gated ion channels in genetic disorders and the question of efficacy. *Int J Biochem Cell Biol* **126**, (2020).
13. Sparling, B. A. & DiMauro, E. F. Progress in the discovery of small molecule modulators of the Cys-loop superfamily receptors. *Bioorg Med Chem Lett* **27**, 3207–3218 (2017).
14. Nys, M., Kesters, D. & Ulens, C. Structural insights into Cys-loop receptor function and ligand recognition. *Biochem Pharmacol* **86**, 1042–1053 (2013).
15. Huang, X., Chen, H., Michelsen, K., Schneider, S. & Shaffer, P. L. Crystal structure of human glycine receptor-α3 bound to antagonist strychnine. *Nature 2015 526:7572* **526**, 277–280 (2015).
16. Kumar, A. *et al.* Mechanisms of activation and desensitization of full-length glycine receptor in lipid nanodiscs. *Nature Communications 2020 11:1* **11**, 1–14 (2020).
17. Lara, C. O., Burgos, C. F., Moraga-Cid, G., Carrasco, M. A. & Yévenes, G. E. Pentameric Ligand-Gated Ion Channels as Pharmacological Targets Against Chronic Pain. *Front Pharmacol* **11**, (2020).
18. Matta, J. A., Gu, S., Davini, W. B. & Bredt, D. S. Nicotinic acetylcholine receptor redux: Discovery of accessories opens therapeutic vistas. *Science (1979)* **373**, (2021).
19. Albuquerque, E. X., Pereira, E. F. R., Alkondon, M. & Rogers, S. W. Mammalian Nicotinic Acetylcholine Receptors: From Structure to Function. *Physiol Rev* **89**, 73–120 (2009).

20. Amiri, S., Sansom, M. S. P. & Biggin, P. C. Molecular dynamics studies of AChBP with nicotine and carbamylcholine: the role of water in the binding pocket. *Protein Engineering Design and Selection* **20**, 353–359 (2007).
21. Balakrishna, P., George, S., Hatoum, H. & Mukherjee, S. Serotonin Pathway in Cancer. *Int J Mol Sci* **22**, 1268 (2021).
22. Langlhofer, G. & Villmann, C. The Intracellular Loop of the Glycine Receptor: It's not all about the Size. *Front Mol Neurosci* **9**, (2016).
23. Capogna, M. & Pearce, R. A. GABAA,slow: causes and consequences. *Trends Neurosci* **34**, 101–112 (2011).
24. Kim, J. J. & Hibbs, R. E. Direct Structural Insights into GABAA Receptor Pharmacology. *Trends Biochem Sci* **46**, 502–517 (2021).
25. Dutertre, S., Becker, C.-M. & Betz, H. Inhibitory Glycine Receptors: An Update. *Journal of Biological Chemistry* **287**, 40216–40223 (2012).
26. Yu, J. *et al.* Mechanism of gating and partial agonist action in the glycine receptor. *Cell* **184**, 957-968.e21 (2021).
27. Kirsch, J. Glycinergic transmission. *Cell Tissue Res* **326**, 535–540 (2006).
28. Imlach, W. L., Bhola, R. F., Mohammadi, S. A. & Christie, M. J. Glycinergic dysfunction in a subpopulation of dorsal horn interneurons in a rat model of neuropathic pain. *Sci Rep* **6**, (2016).
29. Burgos, C. F., Yévenes, G. E. & Aguayo, L. G. Structure and pharmacologic modulation of inhibitory glycine receptors. *Molecular Pharmacology* vol. 90 318–325 Preprint at https://doi.org/10.1124/mol.116.105726 (2016).
30. Iovino, L., Tremblay, M. E. & Civiero, L. Glutamate-induced excitotoxicity in Parkinson's disease: The role of glial cells. *J Pharmacol Sci* **144**, 151–164 (2020).
31. Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim Biophys Acta Gen Subj* **1862**, 324–334 (2018).
32. Betz, H. & Laube, B. Glycine receptors: recent insights into their structural organization and functional diversity. *J Neurochem* **97**, 1600–1610 (2006).
33. Zhu, H. Structure and Mechanism of Glycine Receptor Elucidated by Cryo-Electron Microscopy. *Front Pharmacol* **13**, (2022).
34. Yu, H., Bai, X.-C. & Wang, W. Characterization of the subunit composition and structure of adult human glycine receptors. *Neuron* **109**, 2707-2716.e6 (2021).
35. Lynagh, T. & Pless, S. A. Principles of agonist recognition in Cys-loop receptors. *Front Physiol* **5**, (2014).
36. Huang, X. *et al.* Crystal structures of human glycine receptor α3 bound to a novel class of analgesic potentiators. *Nat Struct Mol Biol* **24**, 108–113 (2017).
37. Perrin, C. L. & Nielson, J. B. "STRONG" HYDROGEN BONDS IN CHEMISTRY AND BIOLOGY. *Annu Rev Phys Chem* **48**, 511–544 (1997).
38. Nekoei, A.-R. & Vatanparast, M. π-Hydrogen bonding and aromaticity: a systematic interplay study. *Physical Chemistry Chemical Physics* **21**, 623–630 (2019).
39. Dougherty, D. A. Cation-π Interactions Involving Aromatic Amino Acids. *J Nutr* **137**, 1504S-1508S (2007).
40. Pless, S. A. *et al.* A cation-π interaction in the binding site of the glycine receptor is mediated by a phenylalanine residue. *Journal of Neuroscience* **28**, 10937–10942 (2008).
41. Lape, R., Colquhoun, D. & Sivilotti, L. G. On the nature of partial agonism in the nicotinic receptor superfamily. *Nature* **454**, 722 (2008).

42. Yevenes, G. E. & Zeilhofer, H. U. Allosteric modulation of glycine receptors. *Br J Pharmacol* **164**, 224–236 (2011).
43. MacDonald, M. L. *et al.* Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol* **2**, 329–337 (2006).
44. Wilkins, M. E., Caley, A., Gielen, M. C., Harvey, R. J. & Smart, T. G. Murine startle mutant *Nmf11* affects the structural stability of the glycine receptor and increases deactivation. *J Physiol* **594**, 3589–3607 (2016).
45. Dutta, S., Selvam, B., Das, A. & Shukla, D. Mechanistic origin of partial agonism of tetrahydrocannabinol for cannabinoid receptors. *Journal of Biological Chemistry* **298**, 101764 (2022).
46. Comitani, F., Limongelli, V. & Molteni, C. The Free Energy Landscape of GABA Binding to a Pentameric Ligand-Gated Ion Channel and Its Disruption by Mutations. *J Chem Theory Comput* **12**, 3398–3406 (2016).
47. Cooper, D. N. Human gene mutation in pathology and evolution. *J Inherit Metab Dis* **25**, 157–182 (2002).
48. Schaefer, N. *et al.* Disruption of a Structurally Important Extracellular Element in the Glycine Receptor Leads to Decreased Synaptic Integration and Signaling Resulting in Severe Startle Disease. *The Journal of Neuroscience* **37**, 7948–7961 (2017).
49. Mocatta, J., Mesoy, S. M., Dougherty, D. A. & Lummis, S. C. R. 5-HT $_3$ Receptor MX Helix Contributes to Receptor Function. *ACS Chem Neurosci* **13**, 2338–2345 (2022).
50. Bode, A. & Lynch, J. W. The impact of human hyperekplexia mutations on glycine receptor structure and function. *Mol Brain* **7**, 2 (2014).
51. Chung, S.-K. *et al.* Pathophysiological Mechanisms of Dominant and Recessive GLRA1 Mutations in Hyperekplexia. *Journal of Neuroscience* **30**, 9612–9620 (2010).
52. Schaefer, N., Langlhofer, G., Kluck, C. J. & Villmann, C. Glycine receptor mouse mutants: model systems for human hyperekplexia. *Br J Pharmacol* **170**, 933–952 (2013).
53. Maveyraud, L. & Mourey, L. Protein X-ray Crystallography and Drug Discovery. *Molecules* **25**, (2020).
54. Thangaratnarajah, C., Rheinberger, J. & Paulino, C. Cryo-EM studies of membrane proteins at 200 keV. *Curr Opin Struct Biol* **76**, 102440 (2022).
55. Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci* **26**, 32–39 (2017).
56. Read, R. J. *et al.* A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* **19**, 1395–1412 (2011).
57. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12 (2010).
58. Marabelli, A., Moroni, M., Lape, R. & Sivilotti, L. G. The kinetic properties of the α3 rat glycine receptor make it suitable for mediating fast synaptic inhibition. *J Physiol* **591**, 3289–3308 (2013).
59. Beato, M., Groot-Kormelink, P. J., Colquhoun, D. & Sivilotti, L. G. The Activation Mechanism of α1 Homomeric Glycine Receptors. *The Journal of Neuroscience* **24**, 895 (2004).
60. Jan, D. D. Saint, David-Watine, B., Korn, H. & Bregestovski, P. Activation of human α1 and α2 homomeric glycine receptors by taurine and GABA. *J Physiol* **535**, 741–755 (2001).
61. Mhashal, A. R., Yoluk, O. & Orellana, L. Exploring the Conformational Impact of Glycine Receptor TM1-2 Mutations Through Coarse-Grained Analysis and Atomistic Simulations. *Front Mol Biosci* **9**, (2022).

62. Speranskiy, K., Cascio, M. & Kurnikova, M. Homology modeling and molecular dynamics simulations of the glycine receptor ligand binding domain. *Proteins* **67**, 950–960 (2007).
63. Thompson, A. J., Lester, H. A. & Lummis, S. C. R. The structural basis of function in Cys-loop receptors. *Q Rev Biophys* **43**, 449–499 (2010).
64. Gonzalez-Gutierrez, G., Wang, Y., Cymes, G. D., Tajkhorshid, E. & Grosman, C. Chasing the open-state structure of pentameric ligand-gated ion channels. *Journal of General Physiology* **149**, 1119–1138 (2017).
65. Shi, S. *et al.* Illumination of a progressive allosteric mechanism mediating the glycine receptor activation. *Nat Commun* **14**, 795 (2023).
66. Crnjar, A. & Molteni, C. Cholesterol content in the membrane promotes key lipid-protein interactions in a pentameric serotonin-gated ion channel. *Biointerphases* **15**, (2020).
67. Dämgen, M. A. & Biggin, P. C. A Refined Open State of the Glycine Receptor Obtained via Molecular Dynamics Simulations. *Structure* **28**, 130-139.e2 (2020).
68. Gibbs, E. *et al.* Conformational transitions and allosteric modulation in a heteromeric glycine receptor. *Nat Commun* **14**, 1363 (2023).
69. Cerdan, A. H., Martin, N. E. ´ R. & Cecchini, M. An Ion-Permeable State of the Glycine Receptor Captured by Molecular Dynamics Theory An Ion-Permeable State of the Glycine Receptor Captured by Molecular Dynamics. *Structure/Folding and Design* **26**, 1555-1562.e4 (2018).
70. Kumar, A. *et al.* Mechanisms of activation and desensitization of full-length glycine receptor in lipid nanodiscs. *Nat Commun* **11**, 1–14 (2020).
71. Lindahl, E. Molecular Dynamics Simulations. *Methods in Molecular Biology* **1215**, 3–26 (2015).
72. Narayan, B., Yuan, Y., Fathizadeh, A., Elber, R. & Buchete, N. V. Long-time methods for molecular dynamics simulations: Markov State Models and Milestoning. *Prog Mol Biol Transl Sci* **170**, 215–237 (2020).
73. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129 (2018).
74. Limongelli, V., Bonomi, M. & Parrinello, M. Funnel metadynamics as accurate binding free-energy method. doi:10.1073/pnas.1303186110.
75. Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P. & Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. (2017) doi:10.1021/jacs.6b12950.
76. Ciccotti, G., Dellago, C., Ferrario, M., Hernández, E. R. & Tuckerman, M. E. Molecular simulations: past, present, and future (a Topical Issue in EPJB). *Eur Phys J B* **95**, 3 (2022).
77. Maginn, E. J. & Elliott, J. R. Historical Perspective and Current Outlook for Molecular Dynamics As a Chemical Engineering Tool. *Ind Eng Chem Res* **49**, 3059–3078 (2010).
78. Marco, E. & Gago, F. Overcoming the Inadequacies or Limitations of Experimental Structures as Drug Targets by Using Computational Modeling Tools and Molecular Dynamics Simulations. *ChemMedChem* **2**, 1388–1401 (2007).
79. Groenhof, G. Introduction to QM/MM Simulations. in 43–66 (2013). doi:10.1007/978-1-62703-017-5_3.
80. Jerzy Leszczynski, Chris Lorenz & Nikos L. Doltsinis. *Handbook of Computational Chemistry*. (Springer Netherlands, Dordrecht, 2012). doi:10.1007/978-94-007-0711-5.
81. Tang, Z., Chen, S.-H. & Chang, C. A. Transient States and Barriers from Molecular Simulations and the Milestoning Theory: Kinetics in Ligand–Protein Recognition and Compound Design. *J Chem Theory Comput* **16**, 1882–1895 (2020).

82. Miao, Y., Bhattarai, A. & Wang, J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J Chem Theory Comput* **16**, 5526–5547 (2020).
83. Söldner, C. A., Horn, A. H. C. & Sticht, H. Binding of histamine to the H1 receptor—a molecular dynamics study. *J Mol Model* **24**, 346 (2018).
84. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip Rev Comput Mol Sci* **1**, 826–843 (2011).
85. Capelli, R. *et al.* Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *J Chem Theory Comput* **15**, 3354–3361 (2019).
86. Comitani, F. *et al.* Insights into the binding of GABA to the insect RDL receptor from atomistic simulations: a comparison of models. *J Comput Aided Mol Des* **28**, 35–48 (2014).
87. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology 2002 9:9* **9**, 646–652 (2002).
88. Frenkel, D. & Smit, B. *Understanding Molecular Simulation*. *Understanding Molecular Simulation* (Elsevier, 2002). doi:10.1016/B978-0-12-267351-1.X5000-7.
89. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174 (2004).
90. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).
91. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J Phys Chem A* **105**, 9954–9960 (2001).
92. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).
93. Kadaoluwa Pathirannahalage, S. P. *et al.* Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J Chem Inf Model* **61**, 4521–4536 (2021).
94. Ponder, J. W. & Case, D. A. FORCE FIELDS FOR PROTEIN SIMULATIONS. (2003).
95. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. SOC* **117**, 5179–5197 (1995).
96. Bernardi, R. NAMD User's Guide. Preprint at https://www.ks.uiuc.edu/Research/namd/cvs/ug.pdf (2022).
97. Allen, M. P. & Tildesley, D. J. Computer simulation of liquids: Second edition. *Computer Simulation of Liquids: Second Edition* 1–626 (2017) doi:10.1093/oso/9780198803195.001.0001.
98. Stenberg, S. & Stenqvist, B. An Exact Ewald Summation Method in Theory and Practice. *J Phys Chem A* **124**, 3943–3946 (2020).
99. Wells, B. A. & Chaffee, A. L. Ewald Summation for Molecular Simulations. *J Chem Theory Comput* **11**, 3684–3695 (2015).
100. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).
101. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**, 17 (2012).
102. Frisch, M. J., Trucks, G. W., Schlegel, H. B. & et al. Gaussian 09. Preprint at (2009).
103. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* **97**, 10269–10280 (1993).

104. Bhuiyan, A., Needham, P. & Walker, R. C. Amber PME Molecular Dynamics Optimization. in *High Performance Parallelism Pearls* 73–89 (Elsevier, 2015). doi:10.1016/B978-0-12-803819-2.00019-7.
105. Bailey, A. G. & Lowe, C. P. MILCH SHAKE: An efficient method for constraint dynamics applied to alkanes. *J Comput Chem* **30**, 2485–2493 (2009).
106. Coffey, W., Kalmykov, Yu. P. & World Scientific (Firm). The Langevin equation : with applications to stochastic problems in physics, chemistry and electrical engineering. 827 (2012).
107. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **103**, 4613–4621 (1995).
108. Roy, K., Kar, S. & Das, R. N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* 1–479 (2015) doi:10.1016/C2014-0-00286-9.
109. Grewer, C. Investigation of the alpha(1)-glycine receptor channel-opening kinetics in the submillisecond time domain. *Biophys J* **77**, 727 (1999).
110. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99**, 12562–12566 (2002).
111. Liao, Q. Enhanced sampling and free energy calculations for protein simulations. *Prog Mol Biol Transl Sci* **170**, 177–213 (2020).
112. Laio, A. & Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **71**, 126601 (2008).
113. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* **100**, 020603 (2008).
114. Tiwary, P. & Parrinello, M. From metadynamics to dynamics. *Phys Rev Lett* **111**, (2013).
115. Wang, Y., Valsson, O., Tiwary, P., Parrinello, M. & Lindorff-Larsen, K. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *J Chem Phys* **149**, (2018).
116. Bonomi, M., Barducci, A. & Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J Comput Chem* **30**, 1615–1621 (2009).
117. Valsson, O., Tiwary, P. & Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *undefined* **67**, 159–184 (2016).
118. Du, X. *et al.* Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **17**, 144 (2016).
119. Wang, D. D., Zhu, M. & Yan, H. Computationally predicting binding affinity in protein–ligand complexes: free energy-based simulations and machine learning-based scoring functions. *Brief Bioinform* **22**, (2021).
120. Raniolo, S. & Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat Protoc* **15**, 2837–2866 (2020).
121. Söldner, C. A., Horn, A. H. C. & Sticht, H. A Metadynamics-Based Protocol for the Determination of GPCR-Ligand Binding Modes. *Int J Mol Sci* **20**, 1970 (2019).
122. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084–3095 (2013).
123. Gowers, R. J. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* 98–105 (2016) doi:10.25080/MAJORA-629E541A-00E.

124. Bitencourt-Ferreira, G., Veit-Acosta, M. & de Azevedo, W. F. Hydrogen Bonds in Protein-Ligand Complexes. *Methods in Molecular Biology* **2053**, 93–107 (2019).
125. Kumar, K. *et al.* Cation–π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem Sci* **9**, 2655–2665 (2018).
126. Platzer, G. *et al.* PI by NMR: Probing CH-π Interactions in Protein-Ligand Complexes by NMR Spectroscopy. *Angew Chem Int Ed Engl* **59**, 14861–14868 (2020).
127. Melis, C., Lummis, S. C. R. & Molteni, C. Molecular Dynamics Simulations of GABA Binding to the GABA C Receptor: The Role of Arg 104. doi:10.1529/biophysj.107.127589.
128. Alvarez, L. D. & Alves, N. R. C. Molecular determinants of tetrahydrocannabinol binding to the glycine receptor. *Proteins: Structure, Function, and Bioinformatics* **91**, 400–411 (2023).
129. PRECUPAS, A. & IONESCU, S. Exploring the interaction of 5,6- benzocoumarin-3-carboxylic acid with bovine serum albumin at the molecular level: A biophysical investigation using molecular dynamics. *Revue Roumaine de Chimie* **66**, 49–58 (2021).
130. Duan, L. L., Feng, G. Q. & Zhang, Q. G. Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand binding energy. *Sci Rep* **6**, 31488 (2016).
131. Cole, C. C. *et al.* Cation−π Interactions and Their Role in Assembling Collagen Triple Helices. *Biomacromolecules* **23**, 4645–4654 (2022).
132. Wu, Y., Lee, J. & Wang, Y. A Comparative Study of GROMACS and NAMD. (2007).
133. Henderson, R. *et al.* Outcome of the First Electron Microscopy Validation Task Force Meeting. *Structure* **20**, 205–214 (2012).
134. Read, R. J. *et al.* A New Generation of Crystallographic Validation Tools for the Protein Data Bank. *Structure* **19**, 1395–1412 (2011).
135. Huang, X., Chen, H., Michelsen, K., Schneider, S. & Shaffer, P. L. Crystal structure of human glycine receptor-α3 bound to antagonist strychnine. *Nature* **526**, 277–280 (2015).
136. Ivica, J. *et al.* The intracellular domain of homomeric glycine receptors modulates agonist efficacy. *J Biol Chem* **296**, (2021).
137. Yu, R. *et al.* Agonist and Antagonist Binding in Human Glycine Receptors. *Biochemistry* **53**, 6041–6051 (2014).
138. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).
139. Pless, S. A. *et al.* A Cation-π Interaction at a Phenylalanine Residue in the Glycine Receptor Binding Site Is Conserved for Different Agonists. *Mol Pharmacol* **79**, 742–748 (2011).
140. Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997–2004 (2014).
141. Vogel, N. *et al.* Mapping of Disulfide Bonds within the Amino-terminal Extracellular Domain of the Inhibitory Glycine Receptor. *J Biol Chem* **284**, 36128 (2009).
142. Rajendra, S. *et al.* The unique extracellular disulfide loop of the glycine receptor is a principal ligand binding element. *EMBO J* **14**, 2987 (1995).
143. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **40**, (2012).
144. Periole, X., Zeppelin, T. & Schiøtt, B. Dimer Interface of the Human Serotonin Transporter and Effect of the Membrane Composition. *Scientific Reports 2018 8:1* **8**, 1–15 (2018).
145. Hamedi, F. & Mohammad-Aghaie, D. Molecular Insight into the Mutual Interactions of Two Transmembrane Domains of Human Glycine Receptor (TM23-GlyR), with the Lipid Bilayers. *Physical Chemistry Research* **8**, 373–397 (2020).
146. P.A. Kollman. Amber 2021. Preprint at (2021).

147. Dickson, C. J. *et al.* Lipid14: The Amber Lipid Force Field. *undefined* **10**, 865–879 (2014).
148. Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
149. Gonzalez-Gutierrez, G., Wang, Y., Cymes, G. D., Tajkhorshid, E. & Grosman, C. Chasing the open-state structure of pentameric ligand-gated ion channels. *Journal of General Physiology* **149**, 1119–1138 (2017).
150. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537–W541 (2012).
151. Byrne, J. & Heidelberger, R. *From Molecules to Networks*. (Elsevier, 2014). doi:10.1016/C2011-0-07251-4.
152. Safar, F. *et al.* The Startle Disease Mutation E103K Impairs Activation of Human Homomeric α1 Glycine Receptors by Disrupting an Intersubunit Salt Bridge across the Agonist Binding Site. *Journal of Biological Chemistry* **292**, 5031–5042 (2017).
153. Schmidtke, P., Bidon-Chanal, A., Luque, F. J. & Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **27**, 3276–3285 (2011).
154. Poveda, J. A. *et al.* Lipid modulation of ion channels through specific binding sites. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838**, 1560–1567 (2014).
155. Thompson, M. J. & Baenziger, J. E. Structural basis for the modulation of pentameric ligand-gated ion channel function by lipids. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1862**, 183304 (2020).
156. Shoemaker, S. C. & Ando, N. X-rays in the Cryo-EM Era: Structural Biology's Dynamic Future. *Biochemistry* **57**, 277 (2018).
157. Grewer, C. Investigation of the alpha(1)-glycine receptor channel-opening kinetics in the submillisecond time domain. *Biophys J* **77**, 727 (1999).
158. Colquhoun, D. & Sakmann, B. Fast events in single-channel currents activated by acetylcholine and its analogues at the frog muscle end-plate. *J Physiol* **369**, 501 (1985).
159. Legendre, P. A Reluctant Gating Mode of Glycine Receptor Channels Determines the Time Course of Inhibitory Miniature Synaptic Events in Zebrafish Hindbrain Neurons. *Journal of Neuroscience* **18**, 2856–2870 (1998).
160. Grudzinska, J. *et al.* The β Subunit Determines the Ligand Binding Properties of Synaptic Glycine Receptors. *Neuron* **45**, 727–739 (2005).
161. Ganser, L. R. *et al.* Distinct phenotypes in zebrafish models of human startle disease. *Neurobiol Dis* **60**, 139–151 (2013).
162. Alexander, S. P. H. *et al.* THE CONCISE GUIDE TO PHARMACOLOGY 2021/22: Ion channels. *Br J Pharmacol* **178**, (2021).
163. Bussi, G. & Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2**, 200–212 (2020).
164. Hata, H., Phuoc Tran, D., Marzouk Sobeh, M. & Kitao, A. Binding free energy of protein/ligand complexes calculated using dissociation Parallel Cascade Selection Molecular Dynamics and Markov state model. *Biophys Physicobiol* **18**, bppb-v18.037 (2021).
165. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174 (2004).
166. Lummis, S. C. R., McGonigle, I., Ashby, J. A. & Dougherty, D. A. Two Amino Acid Residues Contribute to a Cation-π Binding Interaction in the Binding Site of an Insect GABA Receptor. *The Journal of Neuroscience* **31**, 12371–12376 (2011).

167. Fucile, S., de Saint Jan, D., David-Watine, B., Korn, H. & Bregestovski, P. Comparison of glycine and GABA actions on the zebrafish homomeric glycine receptor. *J Physiol* **517**, 369–383 (1999).
168. Borghese, C. M. *et al.* Characterization of Two Mutations, M287L and Q266I, in the α1 Glycine Receptor Subunit That Modify Sensitivity to Alcohols. *Journal of Pharmacology and Experimental Therapeutics* **340**, 304–316 (2012).
169. Crnjar, A., Comitani, F., Melis, C. & Molteni, C. Mutagenesis computer experiments in pentameric ligand-gated ion channels: the role of simulation tools with different resolution. *Interface Focus* **9**, 20180067 (2019).
170. Traka, M., Seburn, K. L. & Popko, B. Nmf11 is a novel ENU-induced mutation in the mouse glycine receptor alpha 1 subunit. *Mammalian Genome* **17**, 950–955 (2006).
171. Miller, P. S. & Smart, T. G. Binding, activation and modulation of Cys-loop receptors. *Trends Pharmacol Sci* **31**, 161–174 (2010).
172. Kmiecik, S. *et al.* Coarse-Grained Protein Models and Their Applications. *Chem Rev* **116**, 7898–7936 (2016).
173. Liu, X. & Wang, W. Asymmetric gating of a human hetero-pentameric glycine receptor. *Nat Commun* **14**, 6377 (2023).