UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Dynamics of Proteins: Applications in Complement and CRISPR-Cas9

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Bioengineering

by

Rohaine Veronica Hsu

December 2021

Dissertation Committee:
Dr. Giulia Palermo, Co-Chairperson
Dr. Hyle Park, Co-Chairperson
Dr. William Grover

The Dissertation of Rohaine Veronica Hsu is approved:

_____

_____

Committee Co-Chairperson

_____

Committee Co-Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This dissertation is thanks to the support of many people. I would first like to thank my advisor, Dr. Giulia Palermo. I am absolutely grateful for her continuous support. Thank you for your kindness, patience, and understanding. I would also like to thank my committee members, Dr. Hyle Park and Dr. William Grover, for their guidance and support throughout my time at UCR.

I would like to thank my lab members, both current and former. Thank you to those from the Palermo Lab, Dr. Yogesh Narkhede, Dr. Pablo Arantes, Dr. Lukasz Nierzwicki, Dr. Mohammed Ahsan, Dr. Souvik Sinha, Jacopo Manigrosso, Brandon Mitchell, Aakash Saha, and Amun Patel. Thank you also to the undergrads from the Palermo Lab.

I would especially like to acknowledge and thank my lab mates, the alumni from the Biomolecular Modeling and Design Lab (BioMoDeL). Thank you to Dr. Reed Harrison, Dr. Rohith Mohan, and Dr. Nehemiah Zewde. Thank you for being amazing mentors when I first joined BioMoDeL and teaching me so much. Your mentorship, guidance, support, and friendship have been the best part of my PhD. Thank you for always helping me, giving me advice, and being there for me anytime I needed you all. I am truly thankful I had the opportunity to meet you all. Thank you to the undergrads from BioMoDeL as well. Your advice and support were always appreciated.

Thank you to the funding sources that helped support my work, the department of bioengineering at UCR, and the National Institutes of Health. Thank you to the

bioengineering department graduate students, faculty, and staff for their support. I would also like to acknowledge Dr. Gloria Gonzalez-Rivera. Thank you for believing in me and motivating me to finish my PhD.

I would like to acknowledge and thank my late PhD advisor, Dr. Dimitrios Morikis. He was the kindest person I ever met. I would like to thank him for his guidance, support, and mentorship. He believed in me and gave me the opportunity to participate in research. I would not be the researcher I am today if it were not for him. He is the inspiration and motivation for my research. He was an amazing advisor. His brilliance and expertise will never be forgotten. He is truly missed.

Lastly, I would like to thank my family and friends for supporting me throughout my PhD. I especially would like to thank my parents and my brother for supporting me through my graduate career.

The text of this dissertation, in part, is a reprint of the material as it appears in:

1. Mitchell, B. P.; Hsu, R. V.; Medrano, M. A.; Zewde, N. T.; Narkhede, Y. B.; Palermo, G. Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9. *Front. Mol. Biosci.* **2020**, *7*, 39. https://doi.org/10.3389/fmolb.2020.00039.

The co-author Giulia Palermo directed and supervised the research which forms the basis for this dissertation. Other co-authors provided technical expertise.

DEDICATION

I dedicate this dissertation to my parents and my brother for their love and support.

This is in memory of my late PhD advisor, Dr. Dimitrios Morikis.

ABSTRACT OF THE DISSERTATION

Molecular Dynamics of Proteins: Applications in Complement and CRISPR-Cas9

by

Rohaine Veronica Hsu

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, December 2021
Dr. Giulia Palermo, Co-Chairperson
Dr. Hyle Park, Co-Chairperson

Molecular dynamics (MD) is a simulation technique that has been utilized to analyze biomolecular systems. Classical MD simulations and constant pH MD simulations have been applied to observe the behavior of components involved in two biomolecular systems of interest, the complement system and CRISPR-Cas9.

The complement system is a defense mechanism part of the innate immune system. Complement associated diseases, such as paroxysmal nocturnal hemoglobinuria and atypical hemolytic uremic syndrome, are treated with complement inhibitors. Monoclonal antibody, eculizumab, is used as treatment for these diseases and functions as an inhibitor of complement component 5 (C5). A next generation version of eculizumab has also been developed known as ravulizumab, resulting from mutations within the heavy chain of eculizumab. MD simulations elucidated key residues involved

in intermolecular interactions between complement inhibitors, eculizumab and ravulizumab, and C5.

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 is part of the bacterial adaptive immune system that functions as a genome editing tool. The HNH domain of CRISPR-Cas9 is involved in DNA cleavage. Through classical MD simulations, residues near the catalytic center of the HNH domain were observed. Mutations were applied to several residues within the HNH domain. The wild type structure was compared to different mutated structures to analyze the effect of the mutations. Distances between residues and RMSD were calculated. Constant pH MD simulations determined pKa values for histidine for the wild type and mutated structures. Taken together, our simulations clarified mechanisms and function of the complement and CRISPR-Cas systems, helping fundamental understanding and engineering.

TABLE OF CONTENTS

LIST OF FIGURES

LISTS OF TABLES

CHAPTER 1

Introduction

1.1 Molecular Dynamics

Molecular Dynamics (MD) is a simulation technique applied to observe the behavior of biomolecular systems.[1] With the initial positions of the atoms in a system, the forces acting on each atom can be calculated using a chosen force field.[1,2] Common force fields used for MD simulations include CHARMM[3], AMBER[4], and GROMOS.[5] A force field is a mathematical parameterized function **(Equation 1.1)**.[6–10]

$$E_{total} = \sum_{bonds} K_r \left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad \textbf{(1.1)}$$

Force fields consider bonded and non-bonded forces between atoms. Bonded forces can include angles and dihedrals. Non-bonded forces can include van der Waals interactions, expressed using the Lennard-Jones potential, and electrostatic interactions, expressed with Coulomb's law.[8,9,11,12] After the forces are determined, atoms move their positions based on Newton's laws of motion.[9] The time of the simulation is advanced by a time step in the femtosecond range[1], and the method is repeated throughout the simulation. With MD simulations, information about biomolecular systems can be obtained, such as

conformational changes. Observation of perturbations on biomolecular systems can also occur using MD simulations.[1] MD simulations were applied for two systems, the complement system and CRISPR-Cas9.

1.2 Complement System

The complement system is a defense mechanism that is part of the immune system.[13–15] It is composed of over thirty proteins. There are three pathways of activation for complement, alternative, classical, and lectin.[16–19] Complement functions to defend against infection. Complement is also associated with various diseases.[20–27] One disease is paroxysmal nocturnal hemoglobinuria (PNH).[28] Treatment of this disease relies on a complement inhibitor. Eculizumab, a monoclonal antibody, functions as a C5 inhibitor to prevent C5 from being cleaved by its convertases, and fragments C5a and C5b are not formed. Without C5b, the cascade of reactions forming the membrane attack complex does not occur.[29–32] To observe the interactions between eculizumab and C5, MD simulations were applied.

1.3 CRISPR-Cas9

Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 is a genome editing tool and a bacterial adaptive immune system.[33,34] CRISPR-Cas9 functions to defend against foreign DNA. In presence of a protospacer adjacent motif, DNA binds Cas9 through matching the target strand of DNA with the guide RNA to form an RNA:DNA hybrid, while the non-target strand is then displaced.[35] The HNH domain

is involved in cleavage of the target strand, and RuvC for cleavage for the non-target strand.[36–38] MD simulations were run to observe the behavior of components within CRISPR-Cas9, including residues near the catalytic center of the HNH domain.

1.4 References

(1) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/J.NEURON.2018.08.011.

(2) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations in Drug Discovery. *J. Comput. Mol. Des. 2011 261* **2011**, *26* (1), 15–26. https://doi.org/10.1007/S10822-011-9517-Y.

(3) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187–217. https://doi.org/10.1002/JCC.540040211.

(4) 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.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(5) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. The GROMOS Software for Biomolecular Simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26* (16), 1719–1751. https://doi.org/10.1002/JCC.20303.

(6) Palermo, G.; Ricci, C. G.; McCammon, J. A. The Invisible Dance of CRISPR-Cas9. *Phys. Today* **2019**, *72* (4), 30. https://doi.org/10.1063/PT.3.4182.

(7) Guvench, O.; MacKerell, A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. *Methods Mol. Biol.* **2008**, *443*, 63–88. https://doi.org/10.1007/978-1-59745-177-2_4.

(8) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. https://doi.org/10.1021/JA00124A002.

(9) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biology*. BioMed Central October 28, 2011, p 71. https://doi.org/10.1186/1741-7007-9-71.

(10) Salsbury, F. R. Molecular Dynamics Simulations of Protein Dynamics and Their Relevance to Drug Discovery. *Current Opinion in Pharmacology*. Elsevier December 1, 2010, pp 738–744. https://doi.org/10.1016/j.coph.2010.09.016.

(11) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nat. 1990 3476294* **1990**, *347* (6294), 631–639. https://doi.org/10.1038/347631a0.

(12) Lindahl, E. Molecular Dynamics Simulations. *Methods Mol. Biol.* **2015**, *1215*, 3–26. https://doi.org/10.1007/978-1-4939-1465-4_1.

(13) Walport, M. J. Complement. First of Two Parts. *N. Engl. J. Med.* **2001**, *344* (14), 1058–1066. https://doi.org/10.1056/NEJM200104053441406.

(14) Walport, M. J. Advances in Immunology: Complement (Second of Two Parts). *N. Engl. J. Med.* **2001**, *344* (15), 1140–114. https://doi.org/10.1056/NEJM200104123441506.

(15) Merle, N. S.; Church, S. E.; Fremeaux-Bacchi, V.; Roumenina, L. T. Complement System Part I - Molecular Mechanisms of Activation and Regulation. *Front. Immunol.* **2015**, *6* (JUN), 262. https://doi.org/10.3389/fimmu.2015.00262.

(16) Zewde, N. T.; Hsu, R. V.; Morikis, D.; Palermo, G. Systems Biology Modeling of the Complement System Under Immune Susceptible Pathogens. *Front. Phys.* **2021**, *9*, 201. https://doi.org/10.3389/fphy.2021.603704.

(17) Zewde, N.; Morikis, D. A Computational Model for the Evaluation of Complement System Regulation under Homeostasis, Disease, and Drug Intervention. *PLoS One* **2018**, *13* (6), e0198644. https://doi.org/10.1371/journal.pone.0198644.

(18) Dunkelberger, J. R.; Song, W. C. Complement and Its Role in Innate and Adaptive Immune Responses. *Cell Res. 2010 201* **2009**, *20* (1), 34–50. https://doi.org/10.1038/cr.2009.139.

(19) Nesargikar, P.; Spiller, B.; Chavez, R. The Complement System: History, Pathways, Cascade and Inhibitors. *Eur. J. Microbiol. Immunol.* **2012**, *2* (2), 103–111. https://doi.org/10.1556/EUJMI.2.2012.2.2.

(20) Ricklin, D.; Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *J. Immunol.* **2013**, *190* (8), 3831–3838. https://doi.org/10.4049/JIMMUNOL.1203487.

(21) Holers, V. M. The Spectrum of Complement Alternative Pathway-Mediated Diseases. *Immunol. Rev.* **2008**, *223* (1), 300–316. https://doi.org/10.1111/J.1600-065X.2008.00641.X.

(22) Morgan, B. P.; Harris, C. L. Complement, a Target for Therapy in Inflammatory and Degenerative Diseases. *Nat. Rev. Drug Discov. 2015 1412* **2015**, *14* (12), 857–877. https://doi.org/10.1038/nrd4657.

(23) Figueroa, J. E.; Densen, P. Infectious Diseases Associated with Complement Deficiencies. *Clin. Microbiol. Rev.* **1991**, *4* (3), 359–395. https://doi.org/10.1128/CMR.4.3.359.

(24) Ricklin, D.; Reis, E. S.; Lambris, J. D. Complement in Disease: A Defence System Turning Offensive. *Nat. Rev. Nephrol. 2016 127* **2016**, *12* (7), 383–401. https://doi.org/10.1038/nrneph.2016.70.

(25) Merle, N. S.; Noe, R.; Halbwachs-Mecarelli, L.; Fremeaux-Bacchi, V.; Roumenina, L. T. Complement System Part II: Role in Immunity. *Front. Immunol.* **2015**, *6* (MAY), 257. https://doi.org/10.3389/fimmu.2015.00257.

(26) Sarma, J. V.; Ward, P. A. The Complement System. *Cell Tissue Res.* **2011**, *343* (1), 227–235. https://doi.org/10.1007/s00441-010-1034-0.

(27) Ricklin, D.; Hajishengallis, G.; Yang, K.; Lambris, J. D. Complement: A Key System for Immune Surveillance and Homeostasis. *Nat. Immunol. 2010 119* **2010**, *11* (9), 785–797. https://doi.org/10.1038/ni.1923.

(28) Risitano, A. M. Paroxysmal Nocturnal Hemoglobinuria and the Complement System: Recent Insights and Novel Anticomplement Strategies. *Adv. Exp. Med. Biol.* **2013**, *734*, 155–172. https://doi.org/10.1007/978-1-4614-4118-2_10.

(29) Brodsky, R. A. Eculizumab: Another Breakthrough. *Blood* **2017**, *129* (8), 922–923. https://doi.org/10.1182/BLOOD-2017-01-760496.

(30) McKeage, K. Eculizumab: A Review of Its Use in Paroxysmal Nocturnal Haemoglobinuria. *Drugs* **2011**, *71* (17), 2327–2345. https://doi.org/10.2165/11208300-000000000-00000.

(31) Parker, C. J.; Kar, S.; Kirkpatrick, P. Eculizumab. *Nat. Rev. Drug Discov. 2007 67* **2007**, *6* (7), 515–516. https://doi.org/10.1038/nrd2369.

(32) Hillmen, P. The Role of Complement Inhibition in PNH. *Hematology* **2008**, *2008* (1), 116–123. https://doi.org/10.1182/ASHEDUCATION-2008.1.116.

(33) Doudna, J. A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science (80-.).* **2014**, *346* (6213). https://doi.org/10.1126/SCIENCE.1258096.

(34) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651–662. https://doi.org/10.1021/acscentsci.9b00020.

(35) Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key Role of the REC Lobe during CRISPR–Cas9 Activation by 'Sensing', 'Regulating', and 'Locking' the Catalytic HNH Domain. *Q. Rev. Biophys.* **2018**, *51*. https://doi.org/10.1017/S0033583518000070.

(36) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59* (5), 2394–2406. https://doi.org/10.1021/acs.jcim.8b00988.

(37) Zuo, Z.; Liu, J. Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Reports 2017 71* **2017**, *7* (1), 1–13. https://doi.org/10.1038/s41598-017-17578-6.

(38) Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V. J. T.; Hayatshahi, H. S.; Rajan, R.; Wang, Y. C.; Liu, J. Structural and Functional Insights into the Bona Fide Catalytic State of Streptococcus Pyogenes Cas9 HNH Nuclease Domain. *Elife* **2019**, *8*. https://doi.org/10.7554/ELIFE.46500.

CHAPTER 2

Elucidating Interactions Between C5 and Complement Inhibitors

2.1 Introduction

Eculizumab is a monoclonal antibody developed by Alexion Pharmaceuticals that functions as a C5 inhibitor. The antibody binds to C5 to prevent fragments C5a and C5b from being formed.[1–5] As an FDA approved drug, eculizumab is used as treatment for diseases including, paroxysmal nocturnal hemoglobinuria (PNH) and atypical hemolytic uremic syndrome.[1,6–14] Eculizumab has been considered one of the most expensive drugs worldwide, with a previous approximate cost of $500,000 per patient/year.[3,15] Patients receive treatment with eculizumab every two weeks intravenously.[6] A next generation version of eculizumab, known as ravulizumab, has been developed by Alexion Pharmaceuticals. Ravulizumab is based on its predecessor, where four mutations were applied to eculizumab, two mutations on the Fc region and two histidine mutations on the heavy chain of eculizumab.[16,17] Ravulizumab was modified from eculizumab to be pH dependent, where association with C5 occurs at pH 7.4 and dissociation at pH 6.[16–19] In comparison to eculizumab, patients receive treatment every eight weeks with ravulizumab.[9,11,20–23] For a subset of PNH patients, genetic polymorphisms have caused poor response to treatment with eculizumab. These genetic polymorphisms result in a mutation at residue arginine 885 within C5.[24–27] To observe the effect of the genetic polymorphisms, MD simulations were applied for both wild type C5 and mutated C5

when in complex with eculizumab, and in complex with ravulizumab. Key residues involved in intermolecular interactions between C5, and the antibodies were determined.

2.2 Methods

2.2.1 Structure Preparation

C5 in complex with eculizumab, obtained from the Protein Data Bank (PDB: 5I5K)[1,28], was used as a starting structure. The structure had C5 and the Fab region of the antibody. Missing residues were added using MODELLER.[29] To create C5 in complex with ravulizumab, two mutations in the heavy chain of eculizumab were applied to form the heavy chain of ravulizumab. Tyrosine 27 and serine 57 were mutated to histidine.[16] Mutations were applied using UCSF Chimera[30] **(Figure 2.1)**. For PNH patients with genetic polymorphisms, arginine 885 of C5 was mutated to either cysteine, histidine, or serine **(Figure 2.2)**.[24,25,31] Mutations at residues 885 were applied to complexes C5 with eculizumab and C5 with ravulizumab using UCSF Chimera. Eight structures in total were prepared **(Table 2.1)**.

Table 2.1. Summary of structures for simulation

C5	Antibody
Wild type	Eculizumab
ARG885CYS	Eculizumab
ARG885HIS	Eculizumab
ARG885SER	Eculizumab
Wild type	Ravulizumab
ARG885CYS	Ravulizumab
ARG885HIS	Ravulizumab
ARG885SER	Ravulizumab

PDB structures were modified for Amber[32] notation. Protonation states of histidine residues were modified based on PROPKA from PDB2PQR.[33] Cysteine residues involved in disulfide bonds were changed from CYS to CYX. Disulfide bond information was obtained from PDBsum.[34]

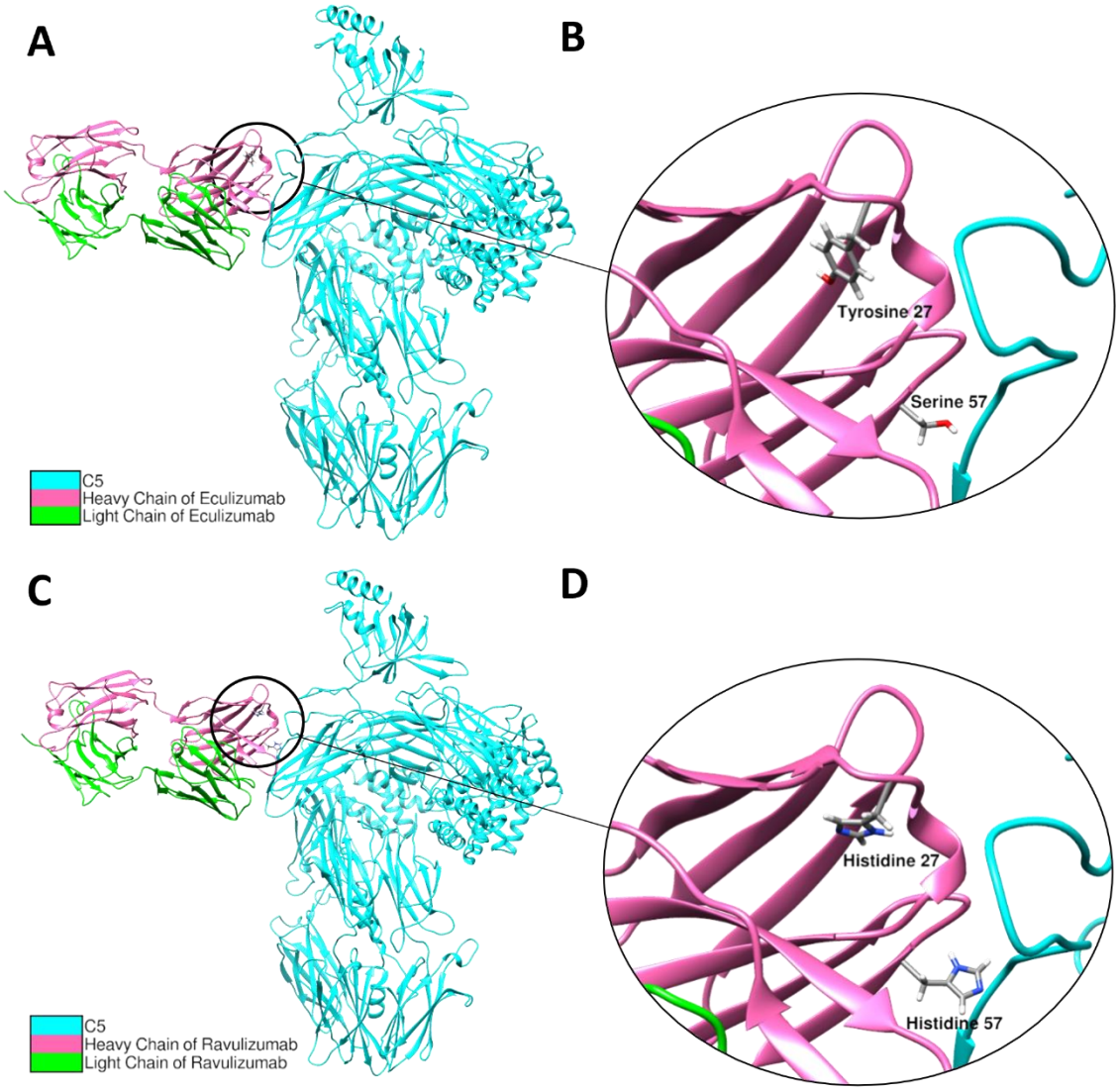

Figure 2.1. Structures of C5 in complex with eculizumab and with ravulizumab, shown in ribbon representation. C5 (cyan) with the heavy chain (pink) and light chain (green) of eculizumab **(A)**, and ravulizumab (C). Residues tyrosine 27 and serine 57 of the heavy chain of eculizumab **(B)** were mutated to histidine to form the heavy chain of ravulizumab **(D)**.

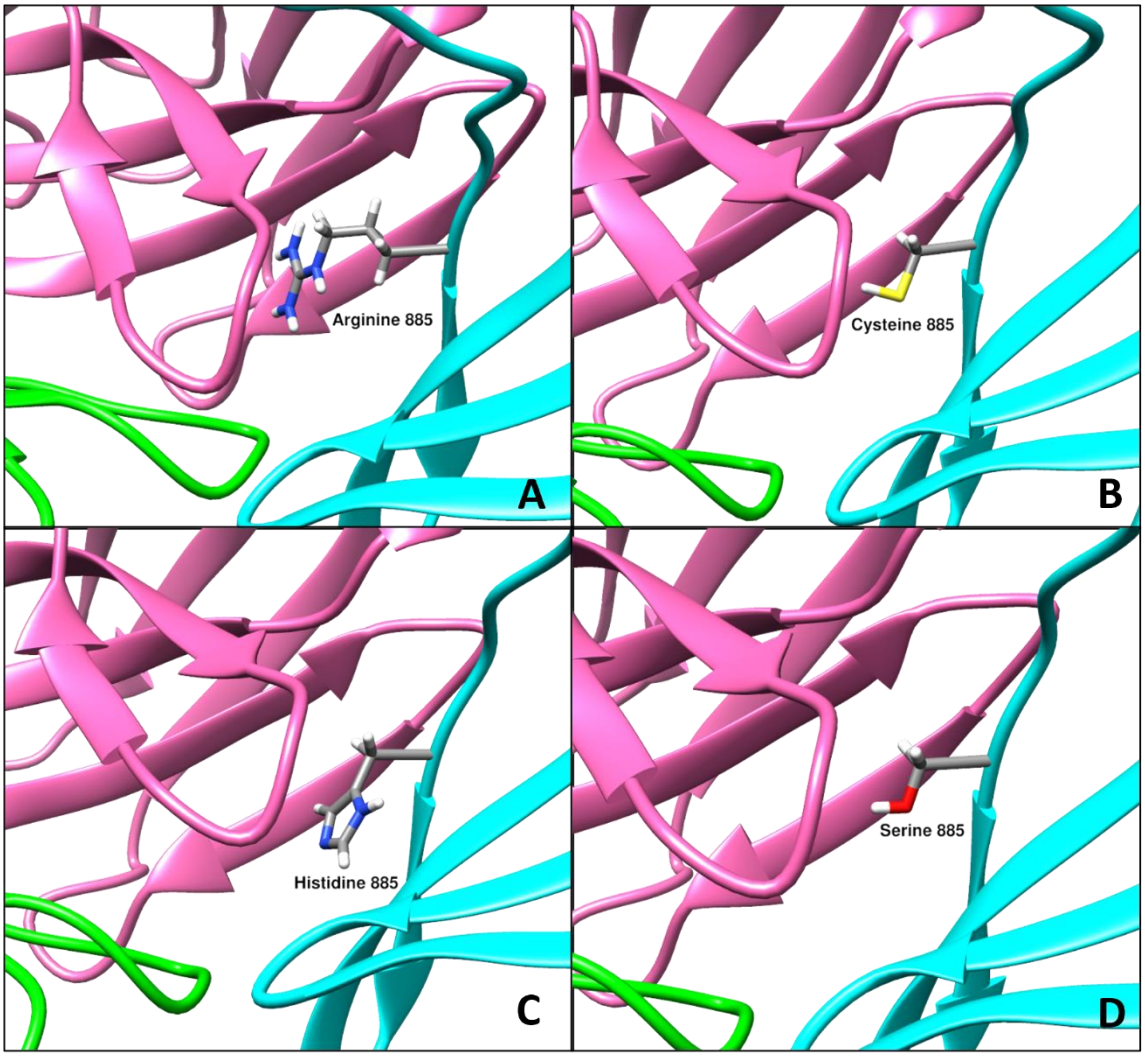

Figure 2.2. Residue 885 of C5. **(A)** Arginine 885 in the wild type C5. Arginine was mutated to cysteine **(B)**, histidine **(C)**, and serine **(D).**

2.2.2 Classical MD Simulations

MD simulations used the Amber ff12SB force field.[32] Using tleap[32], the charge of the system was neutralized with the addition of sodium ions. Each structure was solvated in a water box using TIP3P.[32] Topology and coordinate files of the solvated structures were generated. MD simulations were run using Amber20.[32] Minimization was performed in two steps. First, minimization was performed for 2000 cycles with

positional restraints on the protein with a force constant of 300 kcal/mol · $Å^2$. Restraints were then removed to minimize the entire system for 1000 minimization cycles. To constrain bond lengths involving hydrogen atoms, the SHAKE algorithm was applied.[32] Langevin dynamics was used for temperature control, and a collision frequency of 1/ps was utilized. A pressure relaxation time of 2ps was applied.[32] Heating was performed in four steps. The first two steps considered the NVT ensemble, and each step was run for 5ps. The temperature of the system was raised to 100K. The last two heating steps applied the NPT ensemble and were run for 100ps and 500ps. The temperature was raised by 100K at each step, raising the temperature from 100K to 300K. Positional restraints were applied during the first three steps of heating and then removed at the last step. For the first two heating steps, a force constant of 300 kcal/mol · $Å^2$ was applied. In the third heating step, the force constant was reduced to 25 kcal/mol · $Å^2$. Equilibration was performed for 10ns in the NPT ensemble. Positional restraints were not applied during equilibration. For each structure, production was run in triplicate for 1 microsecond with a simulation time step of 2fs. Production was run with the NVT ensemble and without positional restraints.

2.2.3 Classical MD Analysis

Intermolecular interactions were determined using CPPTRAJ.[35] Hydrogen bonds were considered based on the following parameters, a distance cutoff of 3Å and an angle cutoff of 135º. Salt bridges had a distance cutoff of 5Å. Intermolecular interactions were obtained for each production run. The average value for the interactions for each

simulation triplicate was calculated using MATLAB.[36] Occupancy maps for interactions occurring for 30% or more of the simulation trajectory were generated using pandas[37,38] and seaborn **(Figures 2.3-2.6, A.1-A.24)**.[39] Root mean square deviation (RMSD) of C5, the heavy chain, and the light chain for each structure was also determined with CPPTRAJ **(Figures A.25-A.30).**

2.2.4 Constant pH MD Simulations

Constant pH simulations[40] were performed for solvated structures of C5 in complex with ravulizumab to observe the pKa of histidine 27 and 57 in the heavy chain of ravulizumab. Protonation states of histidine 27 and histidine 57 were modified to HIP for amber notation for constant pH MD simulations. Topology and coordinate files were obtained through tleap. For each system, a cpin file was generated from cpinutil.py.[32] Minimization was performed for 5000 cycles with positional restraints on the protein. Heating was performed for 3ns, raising the temperature to 300K. Positional restraints were used during heating holding the protein fixed. For both minimization and heating, a force constant of 300 kcal/mol $\cdot$ Å^2 was used for the restrains. Equilibration was performed for 8ns. Restraints were removed during equilibration. Production was run for 20ns with a simulation time step of 2fs at pH values 1 to 14, and pH 7.4.

2.2.5 Constant pH MD Analysis

Fraction protonated, predicted pKa values, and populations of protonation states were calculated using cphstats.[32] Fraction deprotonated was calculated by subtracting the

fraction protonated from 1. Titration curves were produced by plotting the fraction deprotonated versus pH and fitted using the Hill equation[41] **(Equation 2.1)** on gnuplot.[42]

$$f_d = \frac{1}{1+10^{n(pK_a - pH)}} \quad \textbf{(2.1)}$$

The pKa value and the Hill coefficient for both histidine values were calculated from the Hill equation.

2.3 Results

2.3.1 Classical MD Simulations

Intermolecular interactions between C5 and the antibodies were analyzed. Hydrogen bonds and salt bridges occurring for 30% or more on average of the simulation trajectories were determined. Occupancy maps for hydrogen bonds and salt bridges involved in C5 in complex with eculizumab **(Figures 2.3-2.4)**, and ravulizumab **(Figures 2.5-2.6)** indicated the frequency the interactions occurred.

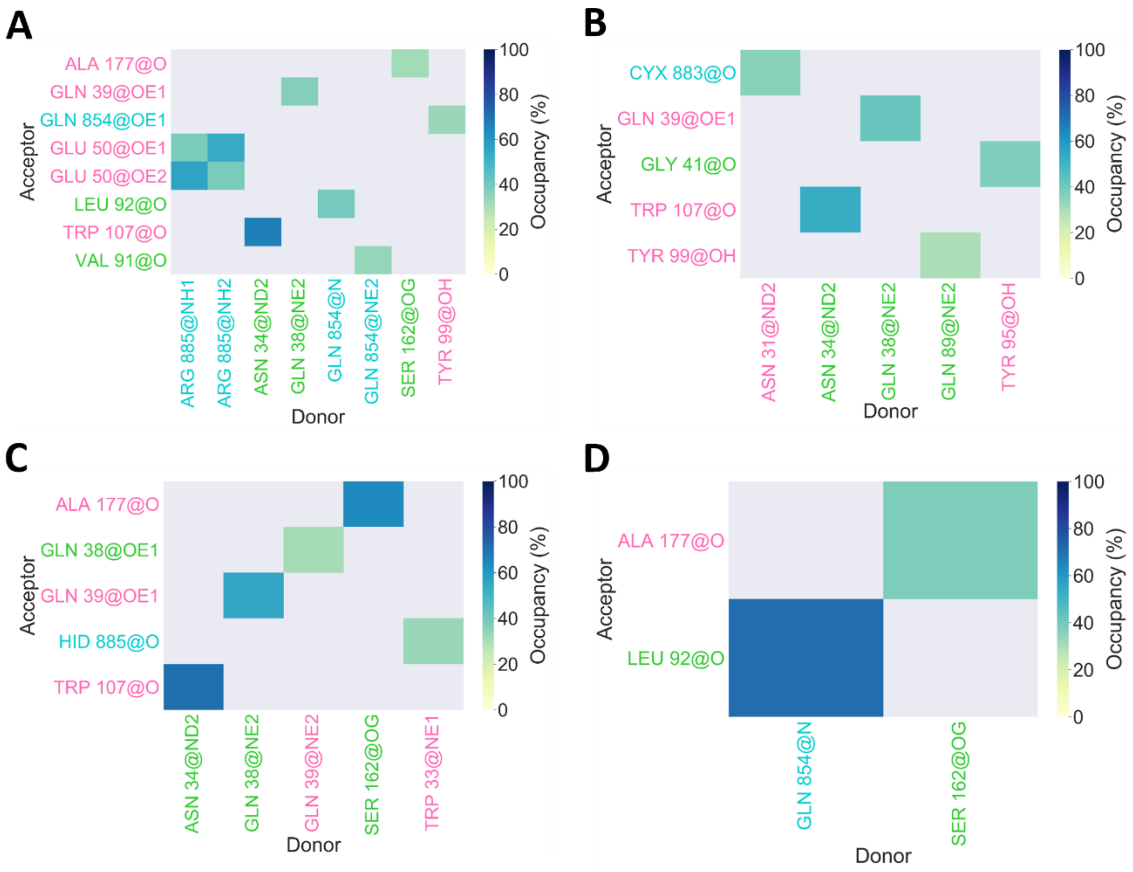

Figure 2.3. Occupancy maps for hydrogen bonds for C5 in complex with eculizumab that occurred for 30% or more on average of the simulation trajectory. **(A)** Wild type C5 in complex with eculizumab. C5 with residue 885 mutated to **(B)** Cysteine, **(C)** Histidine, and **(D)** Serine.

From averaging the intermolecular interactions from the triplicate runs, wild type C5 in complex with eculizumab observed ten hydrogen bonds occurring for 30% or more of the simulation trajectory. Mutated C5 in complex with eculizumab on average had less hydrogen bonds occurring throughout the simulation than the wild type C5. ARG885CYS and ARG885HIS C5 structures had five hydrogen bonds, whereas ARG885SER had two.

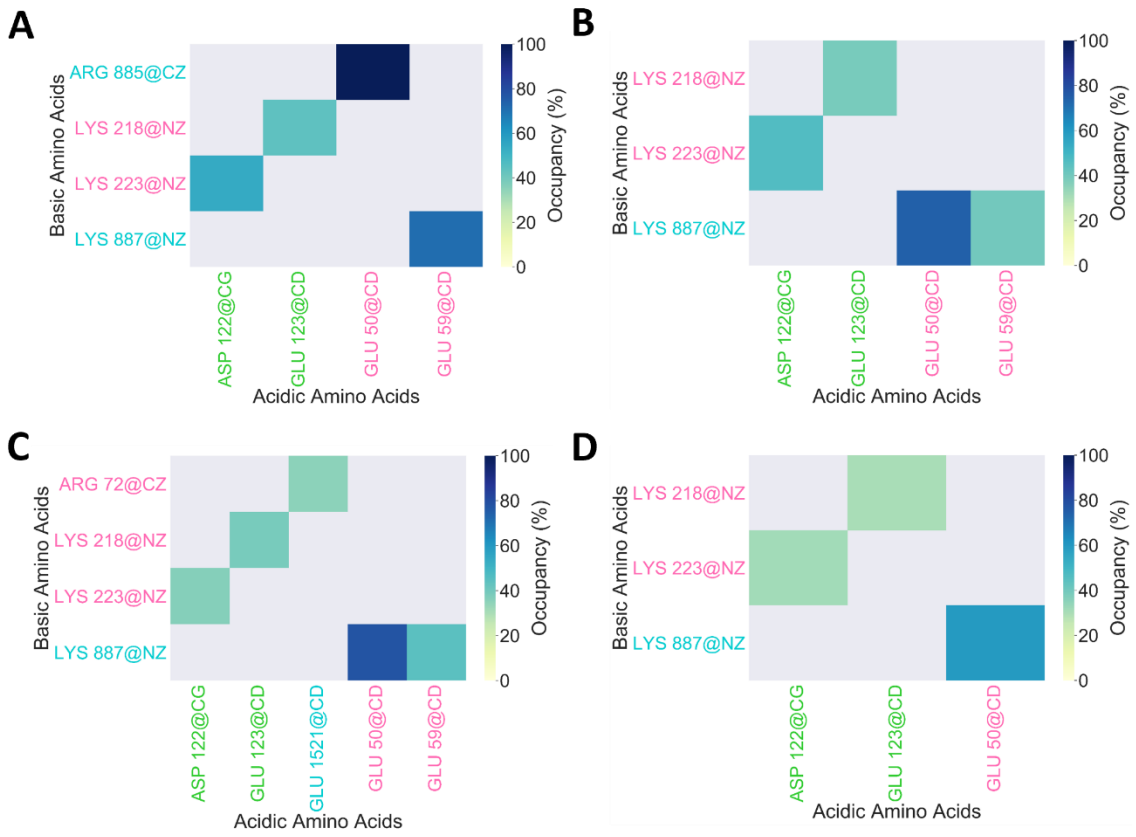

Figure 2.4. Occupancy maps for salt bridges for C5 in complex with eculizumab that occurred for 30% or more on average of the simulation trajectory. **(A)** Wild type C5 in complex with eculizumab. C5 with residue 885 mutated to **(B)** Cysteine, **(C)** Histidine, and **(D)** Serine.

The number of salt bridges ranged from three to five interactions for the structures. Mutated C5 with serine observed three salt bridges, wild type C5 and mutated C5 with cysteine observed four each. Mutated C5 with histidine observed five.

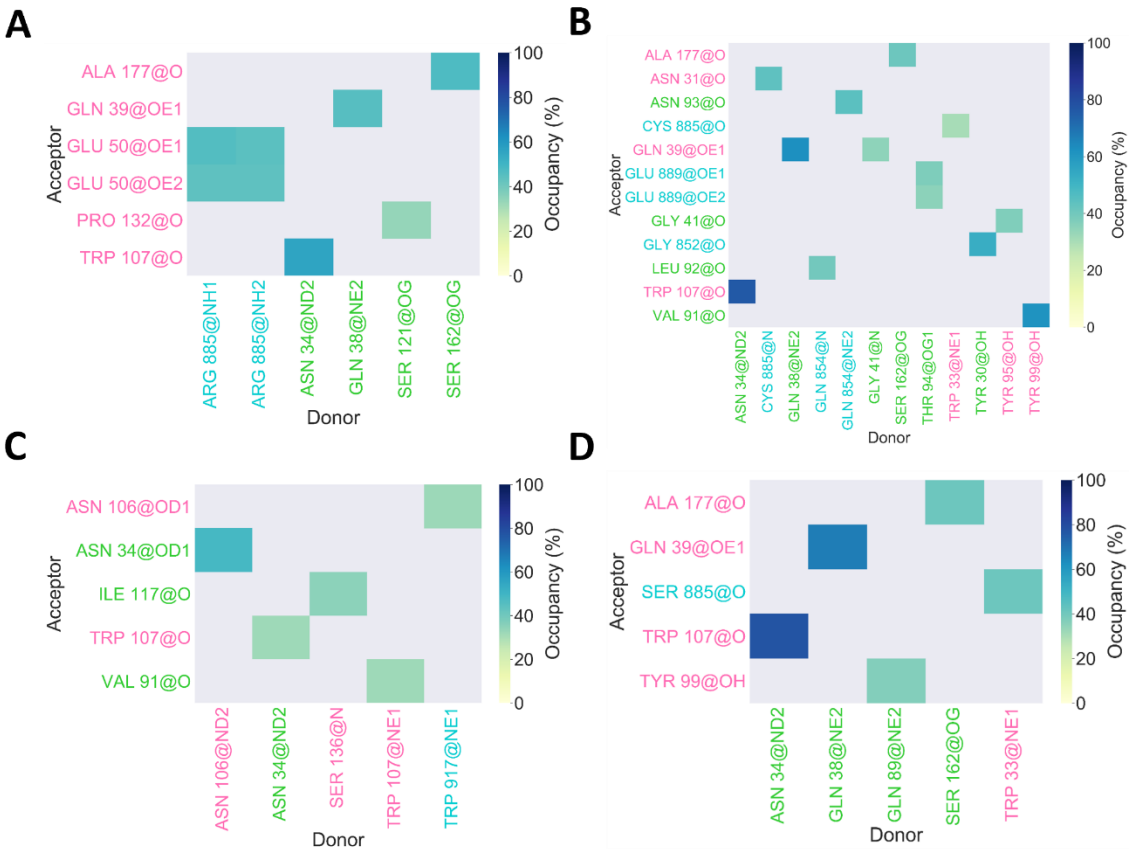

Figure 2.5. Occupancy maps for hydrogen bonds for C5 in complex with ravulizumab that occurred for 30% or more on average of the simulation trajectory. **(A)** Wild type C5 in complex with ravulizumab. C5 with residue 885 mutated to **(B)** Cysteine, **(C)** Histidine, and **(D)** Serine.

For C5 in complex with ravulizumab, the mutated C5 with cysteine observed more hydrogen bonds compared to the wild type C5. Mutated C5 with histidine and serine observed five hydrogen bonds each.

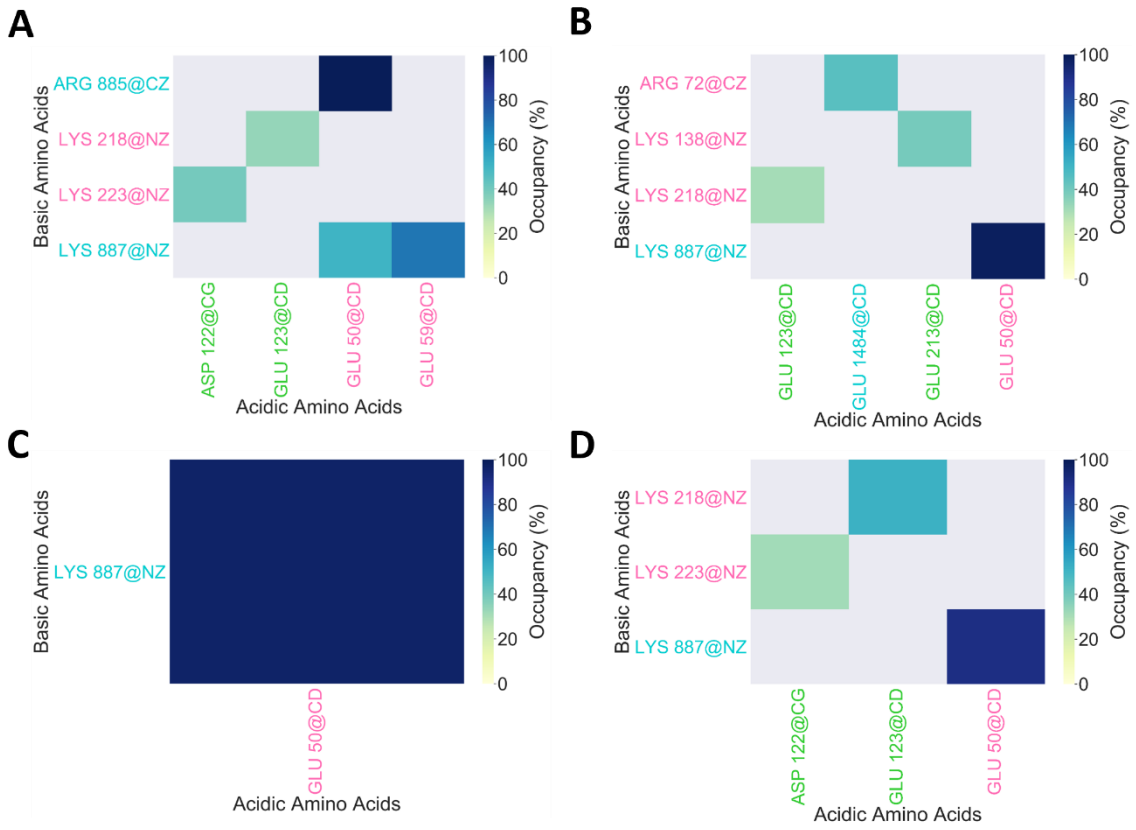

Figure 2.6. Occupancy maps for salt bridges for C5 in complex with ravulizumab that occurred for 30% or more on average of the simulation trajectory. **(A)** Wild type C5 in complex with ravulizumab. C5 with residue 885 mutated to **(B)** Cysteine, **(C)** Histidine, and **(D)** Serine.

Wild type C5 observed five salt bridges. Mutated C5 with cysteine had four salt bridges, and mutated C5 with serine had three. Mutated C5 with histidine on average observed one salt bridge occurring for 30% or more of the simulation trajectory.

2.3.2 Constant pH MD Simulations

Titration curves for HIS27 and HIS57 were generated for C5 with ravulizumab structures. From the titration curves, pKa values were determined. HIS 27 observed a lower pKa compared to HIS57 in all the structures. For wild type C5 **(Figure 2.7),** HIS 27 had a pKa of 5.04, whereas HIS57 has a pKa of 6.55. The hill coefficients were 0.69 and 1.19 for HIS27 and HIS57, respectively.

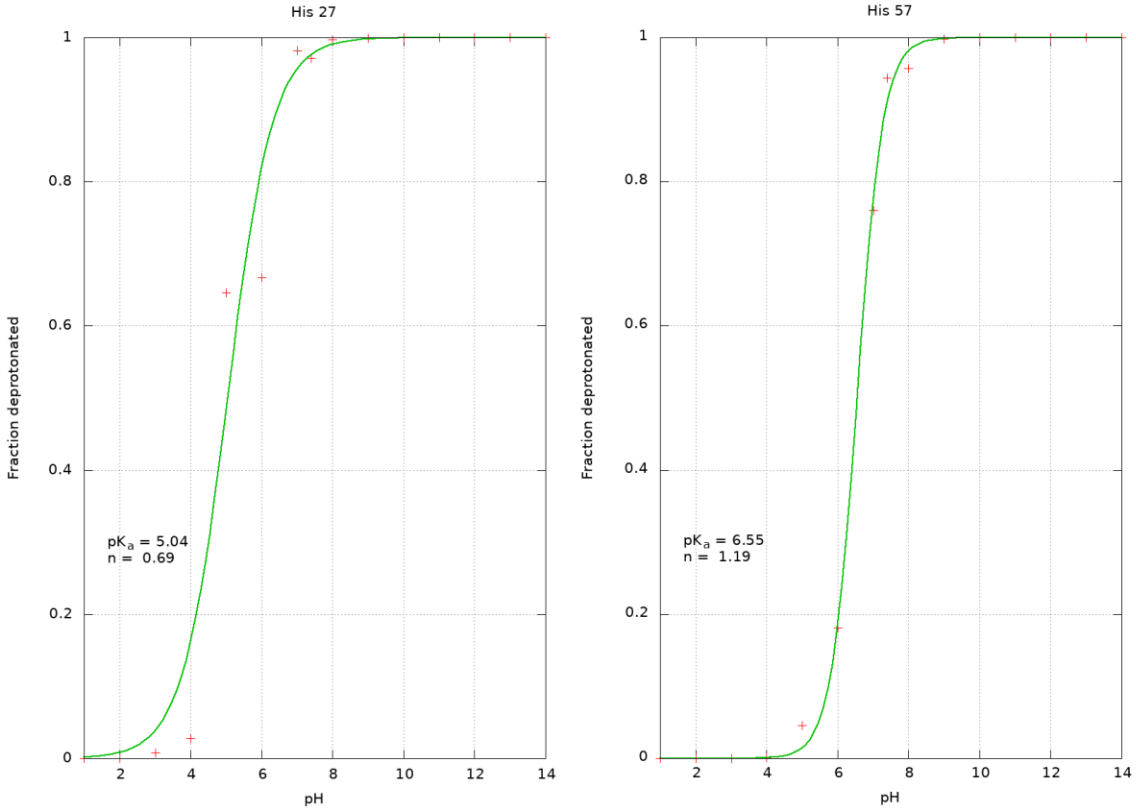

Figure 2.7. Titration curves of histidine 27 and histidine 57 for wild type C5 in complex with ravulizumab.

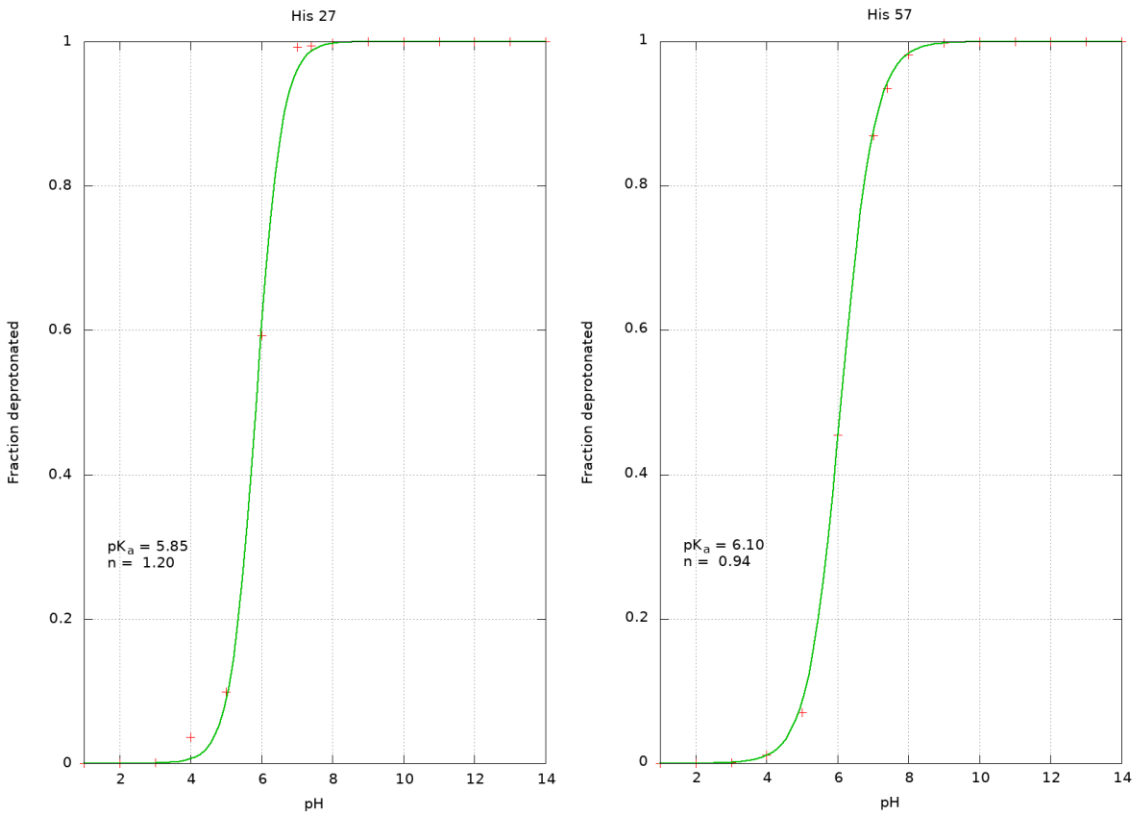

Figure 2.8. Titration curves of histidine 27 and histidine 57 for C5 in complex with ravulizumab with arginine 885 mutated to cysteine.

For mutated C5, ARG885CYS, the pKa value for HIS27 at 5.85 was higher than the wild type C5 case with a pKa of 5.04 **(Figure 2.8)**. However, for HIS57, the pKa value 6.10 was lower than the pKa from the wild type C5. For mutated C5, ARG885HIS, the pKa value for HIS 27 was calculated to be 5.93. This value was higher than the wild type pKa. Similarly, for HIS57, the calculated pKa of 6.77 was also higher than the wild type C5 case **(Figure 2.9)**. For mutated C5, ARG885SER, similar behavior as ARG885CYS occurred. The pKa value for HIS27 was higher than the wild type C5, but HIS57 observed a lower pKa **(Figure 2.10)**.

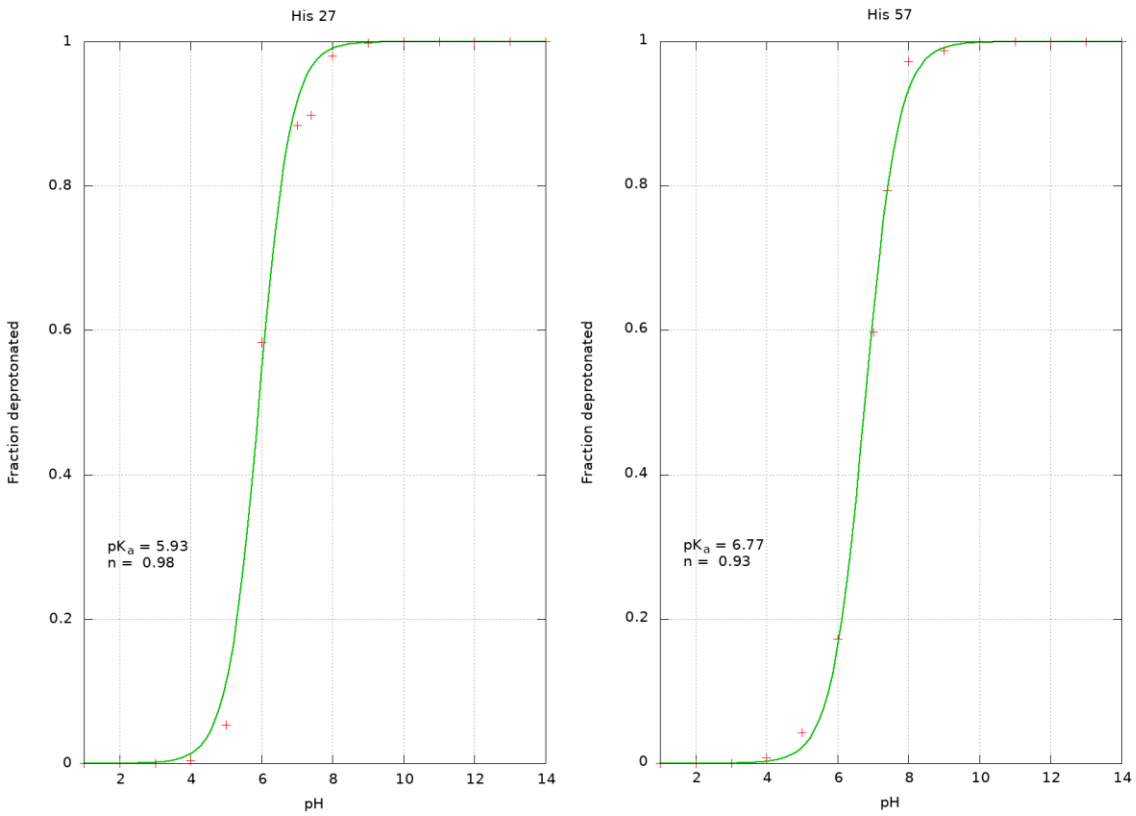

Figure 2.9. Titration curves of histidine 27 and histidine 57 for C5 in complex with ravulizumab with arginine 885 mutated to histidine.

The lowest pKa value for HIS27 was 5.04 in the wild type C5 structure. The highest pKa value for HIS27 was 5.93 for mutated C5 with histidine. The pKa value for HIS57 was the lowest for the mutated C5 structure with cysteine at 6.10, whereas the highest pKa value occurred for the mutated C5 structure with histidine at 6.77.

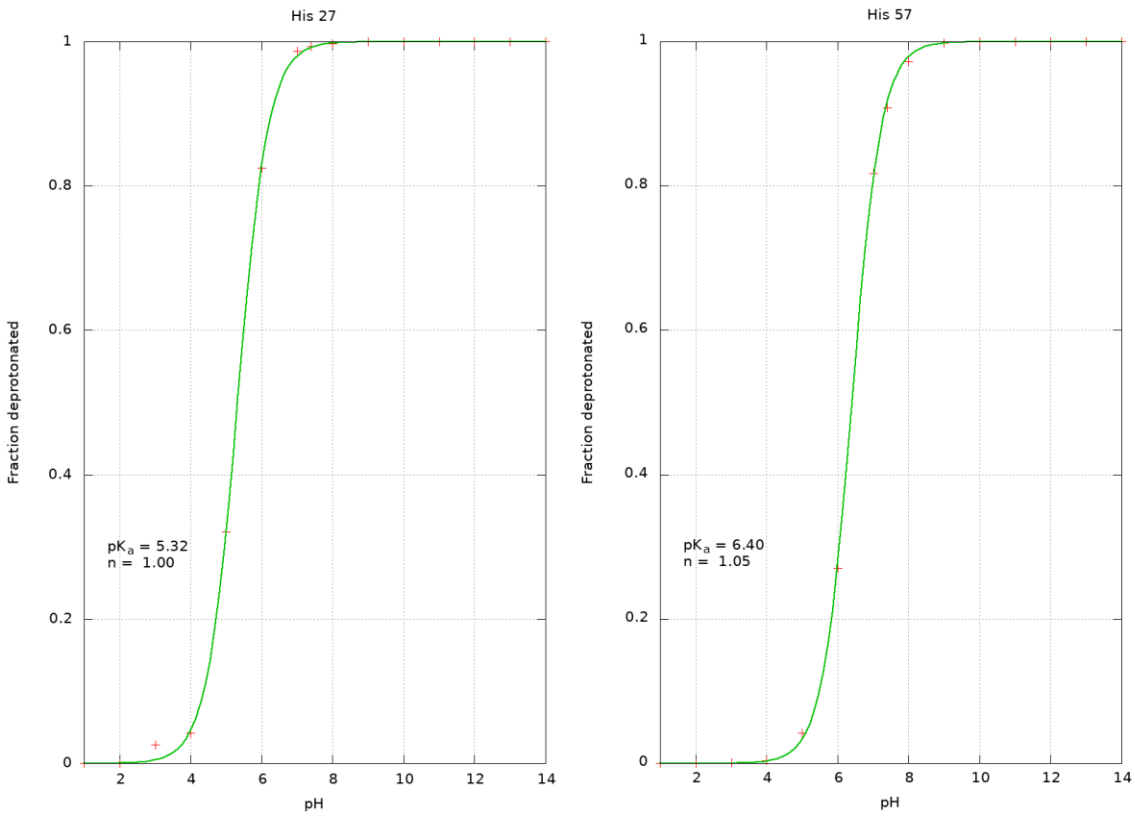

Figure 2.10. Titration curves of histidine 27 and histidine 57 for C5 in complex with ravulizumab with arginine 885 mutated to serine

2.4 Discussion

From the simulations, the average values for intermolecular interactions for each triplicate run was determined. Comparison between the wild type C5 structures and the mutated structures was observed when in complex of eculizumab and ravulizumab. For hydrogen bonds, wild type C5 had the most interactions when in complex with eculizumab. Residues GLN 854 and ARG 885 were involved in seven hydrogen bonds with eculizumab. Five of the interactions occurred with the heavy chain, residues GLU 50 and TYR 99, and two with the light chain, residues LEU 92 and VAL 91. For

eculizumab, three hydrogen bonds occurred between the heavy and light chains. For mutated C5 with cysteine, one hydrogen bond was between C5 and the heavy chain of eculizumab, CYX 883 of C5 and ASN 31 of the heavy chain. Four other hydrogen bonds occurred between the heavy and light chains. Similarly, mutated C5 with histidine also had five hydrogen bonds. One interaction between C5 and the heavy chain of eculizumab, HID 885 of C5 and TRP 33 of the heavy chain. The remaining four interactions were between the heavy and light chains of eculizumab. For mutated C5 with serine, two hydrogen bonds were determined. One interaction between residue GLN 854 of C5 with LEU 92 of the light chain of eculizumab. The other interaction was between the chains of eculizumab.

For salt bridges, four interactions occurred for the wild type structure. Two interactions between C5 and the heavy chain, ARG 885 and GLU 50, and LYS 887 and GLU 59. Two salt bridges between the heavy and light chains of eculizumab. Mutated C5 with cysteine also observed four salt bridges. Two interactions between the chains of eculizumab, and two interactions occurred between C5 and the heavy chain, LYS887 with GLU50, and LYS 887 with GLU 59. Mutated C5 with histidine had five salt bridges. Three interactions were between C5 and the heavy chain of the antibody, LYS 887 and GLU 50, LYS 887 and GLU 59, and GLU1521 and ARG 72. For mutated C5 with serine, three salt bridges occurred, with one interaction between C5 and the heavy chain, LYS 887 and GLU 50.

For C5 in complex with ravulizumab, the wild type structure had eight hydrogen bonds, four involving ARG 885 of C5 interacting with GLU 50 of the heavy chain. With

mutated C5 with cysteine, thirteen hydrogen bonds were observed. Seven of the interactions involved C5. Five of the interactions were between C5, residues GLY 852, GLN 854, and GLU 889, and the light chain of ravulizumab, residues TYR 30, ASN 93, LEU 92, and THR 94. Two interactions were between C5, residue CYS 885, and the heavy chain of ravulizumab, residues ASN 31 and TRP 33. Mutated C5 with histidine observed five hydrogen bonds. One interaction involved a residue of C5, TRP 917 interacting with ASN 106 of the heavy chain. Similarly, for mutated C5 with serine, five hydrogen bonds occurred with one interaction between C5 and the heavy chain of ravulizumab, SER 885 of C5 and TRP 33 of the heavy chain.

For the wild type structure, five salt bridges occurred. Three interactions involved C5 and the heavy chain of ravulizumab, ARG 885 and GLU 50, LYS 887 and GLU 59, and LYS 887 with GLU 50. Four salt bridges were observed for mutated C5 with cysteine. Two interactions involved C5 and the heavy chain, LYS 887 and GLU 50, GLU 1484 and ARG 72. One salt bridge interaction was observed for mutated C5 with histidine. An interaction between LYS 887 of C5 and GLU 50 of the heavy chain of ravulizumab. Mutated C5 with serine had three slat bridge interactions. One interaction occurred between LYS 887 of C5 and GLU 50 of the heavy chain of ravulizumab.

For constant pH MD simulations, the pKa values for the mutated structures was higher for HIS27 than the wild type C5 structure. However, for HIS57, mutated C5 structure ARG885HIS had a higher pKa value than the wild type structure, whereas ARG885CYS and ARG885SER both observed a lower pKa value.

2.5 Conclusion

Constant pH MD simulations determined the pKa values of HIS27 and HIS57 of the heavy chain of ravulizumab for the wild type and mutated C5 structures. The pKa of HIS27 ranged from 5.04 to 5.93, the lowest value from the wild type structure, and the highest value from the mutated structure with histidine. HIS57 pKa values ranged from 6.10 to 6.77, with the lowest pKa from mutated C5 structure with cysteine, and the highest from the mutated structure with histidine. Constant pH MD simulations indicated a higher pKa for HIS57 compared to HIS27. MD simulations provided insight to the effect of mutations against the wild type structures of C5 in complex with eculizumab and ravulizumab.

Classical MD simulations indicated key residues involved intermolecular interactions between C5 and the antibodies **(Tables 2.2-2.5)**. Residue 885 interacted in hydrogen bonds with eculizumab and ravulizumab for wild type C5 and while mutated. Wild type C5 (ARG885) participated in hydrogen bonds for both eculizumab and ravulizumab. When ARG885 was mutated histidine, the residue interacted in a hydrogen bond with eculizumab. When ARG885 was mutated cysteine and serine, the residue interacted in hydrogen bonds with ravulizumab. Residues GLN 854, ASN 31, TRP 33, GLU 50, and LEU 92 were involved in hydrogen bonds for both C5 in complex with eculizumab and C5 in complex with ravulizumab. For salt bridges, ARG 885, LYS 887, GLU 50, GLU 59, and ARG 72 occurred for both C5 in complex with eculizumab and C5 in complex with ravulizumab. These residues could provide a potential starting

framework for future drug development as they were involved in intermolecular interactions for both antibodies when binding to C5.

Table 2.2. Summary of residues involved in hydrogen bonds between C5 and eculizumab. ARG 885 is from the wild type C5. HID 885 is from the mutated C5 with histidine.

Structure component	Residue
C5	GLN 854
	CYX 883
	ARG 885
	HID 885
Heavy Chain	ASN 31
	TRP 33
	GLU 50
	TYR 99
Light Chain	VAL 91
	LEU 92

Table 2.3. Summary of residues involved in salt bridges between C5 and eculizumab.

Structure component	Residue
C5	ARG 885
	LYS 887
	GLU 1521
Heavy Chain	GLU 50
	GLU 59
	ARG 72

Table 2.4. Summary of residues involved in hydrogen bonds between C5 and ravulizumab. ARG 885 is from the wild type C5. CYS 885 is from the mutated C5 with cysteine. SER 885 is from the mutated C5 with serine.

Structure component	Residue
C5	GLY 852
	GLN 854
	ARG 885
	CYS 885
	SER 885
	GLU 889
	TRP 917
Heavy Chain	ASN 31
	TRP 33
	GLU 50
	ASN 106
Light Chain	TYR 30
	LEU 92
	ASN 93
	THR 94

Table 2.5. Summary of residues involved in salt bridges between C5 and ravulizumab.

Structure component	Residue
C5	ARG 885
	LYS 887
	GLU 1484
Heavy Chain	GLU 50
	GLU 59
	ARG 72

2.6 References

(1) Schatz-Jakobsen, J. A.; Zhang, Y.; Johnson, K.; Neill, A.; Sheridan, D.; Andersen, G. R. Structural Basis for Eculizumab-Mediated Inhibition of the Complement Terminal Pathway. *J. Immunol.* **2016**, *197* (1), 337–344. https://doi.org/10.4049/jimmunol.1600280.

(2) Parker, C. J.; Kar, S.; Kirkpatrick, P. Eculizumab. *Nat. Rev. Drug Discov. 2007 67* **2007**, *6* (7), 515–516. https://doi.org/10.1038/nrd2369.

(3) Ricklin, D.; Mastellos, D. C.; Reis, E. S.; Lambris, J. D. The Renaissance of Complement Therapeutics. *Nat. Rev. Nephrol. 2017 141* **2017**, *14* (1), 26–47. https://doi.org/10.1038/nrneph.2017.156.

(4) Hillmen, P. The Role of Complement Inhibition in PNH. *Hematology* **2008**, *2008* (1), 116–123. https://doi.org/10.1182/ASHEDUCATION-2008.1.116.

(5) Merle, N. S.; Church, S. E.; Fremeaux-Bacchi, V.; Roumenina, L. T. Complement System Part I - Molecular Mechanisms of Activation and Regulation. *Front. Immunol.* **2015**, *6* (JUN), 262. https://doi.org/10.3389/fimmu.2015.00262.

(6) McKeage, K. Eculizumab: A Review of Its Use in Paroxysmal Nocturnal Haemoglobinuria. *Drugs* **2011**, *71* (17), 2327–2345. https://doi.org/10.2165/11208300-000000000-00000.

(7) Keating, G. M. Eculizumab: A Review of Its Use in Atypical Haemolytic Uraemic Syndrome. *Drugs* **2013**, *73* (18), 2053–2066. https://doi.org/10.1007/s40265-013-0147-7.

(8) Rother, R. P.; Rollins, S. A.; Mojcik, C. F.; Brodsky, R. A.; Bell, L. Discovery and Development of the Complement Inhibitor Eculizumab for the Treatment of Paroxysmal Nocturnal Hemoglobinuria. *Nat. Biotechnol. 2007 2511* **2007**, *25* (11), 1256–1264. https://doi.org/10.1038/nbt1344.

(9) Zelek, W. M.; Xie, L.; Morgan, B. P.; Harris, C. L. Compendium of Current Complement Therapeutics. *Mol. Immunol.* **2019**, *114*, 341–352. https://doi.org/10.1016/J.MOLIMM.2019.07.030.

(10) Ricklin, D.; Reis, E. S.; Lambris, J. D. Complement in Disease: A Defence System Turning Offensive. *Nat. Rev. Nephrol. 2016 127* **2016**, *12* (7), 383–401. https://doi.org/10.1038/nrneph.2016.70.

(11) Mastellos, D. C.; Ricklin, D.; Lambris, J. D. Clinical Promise of Next-Generation Complement Therapeutics. *Nat. Rev. Drug Discov. 2019 189* **2019**, *18* (9), 707–729. https://doi.org/10.1038/s41573-019-0031-6.

(12) Harris, C. L. Expanding Horizons in Complement Drug Discovery: Challenges and Emerging Strategies. *Semin. Immunopathol.* **2018**, *40* (1), 125–140. https://doi.org/10.1007/s00281-017-0655-8.

(13) Wong, E. K. S.; Kavanagh, D. Diseases of Complement Dysregulation—an Overview. *Semin. Immunopathol. 2018 401* **2018**, *40* (1), 49–64. https://doi.org/10.1007/S00281-017-0663-8.

(14) Risitano, A. M. Paroxysmal Nocturnal Hemoglobinuria and the Complement System: Recent Insights and Novel Anticomplement Strategies. *Adv. Exp. Med. Biol.* **2013**, *734*, 155–172. https://doi.org/10.1007/978-1-4614-4118-2_10.

(15) Jore, M. M.; Johnson, S.; Sheppard, D.; Barber, N. M.; Li, Y. I.; Nunn, M. A.; Elmlund, H.; Lea, S. M. Structural Basis for Therapeutic Inhibition of Complement C5. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 378–386. https://doi.org/10.1038/nsmb.3196.

(16) Sheridan, D.; Yu, Z.-X.; Zhang, Y.; Patel, R.; Lasaro, M. A.; Bouchard, K.; Andrien, B.; Marozsan, A.; Wang, Y.; Tamburini, P. Design and Preclinical Characterization of ALXN1210: A Novel Anti-C5 Antibody with Extended Duration of Action. *PLoS One* **2018**. https://doi.org/10.1371/journal.pone.0195909.

(17) Stern, R. M.; Connell, N. T. Ravulizumab: A Novel C5 Inhibitor for the Treatment of Paroxysmal Nocturnal Hemoglobinuria. *Ther. Adv. Hematol.* **2019**, *10*, 2040620719874728. https://doi.org/10.1177/2040620719874728.

(18) Fattizzo, B.; Kulasekararaj, A. G. Second-Generation C5 Inhibitors for Paroxysmal Nocturnal Hemoglobinuria. *BioDrugs* **2020**, *34* (2), 149–158. https://doi.org/10.1007/s40259-019-00401-1.

(19) Lee, J. W.; Kulasekararaj, A. G. Ravulizumab for the Treatment of Paroxysmal Nocturnal Hemoglobinuria. *Expert Opin. Biol. Ther.* **2020**, *20* (3), 227–237. https://doi.org/10.1080/14712598.2020.1725468.

(20) McKeage, K. Ravulizumab: First Global Approval. *Drugs* **2019**, *79* (3), 347–352. https://doi.org/10.1007/s40265-019-01068-2.

(21) Syed, Y. Y. Ravulizumab: A Review in Atypical Haemolytic Uraemic Syndrome. *Drugs* **2021**, *81* (5), 587–594. https://doi.org/10.1007/s40265-021-01481-6.

(22) Lee, J. W.; de Fontbrune, F. S.; Lee, L. W. L.; Pessoa, V.; Gualandro, S.; Füreder, W.; Ptushkin, V.; Rottinghaus, S. T.; Volles, L.; Shafner, L.; Aguzzi, R.; Pradhan, R.; Schrezenmeier, H.; Hill, A. Ravulizumab (ALXN1210) vs Eculizumab in Adult Patients with PNH Naive to Complement Inhibitors: The 301 Study. *Blood*

2019, *133* (6), 530–539. https://doi.org/10.1182/BLOOD-2018-09-876136.

(23) Kulasekararaj, A. G.; Hill, A.; Rottinghaus, S. T.; Langemeijer, S.; Wells, R.; Gonzalez-Fernandez, F. A.; Gaya, A.; Lee, J. W.; Gutierrez, E. O.; Piatek, C. I.; Szer, J.; Risitano, A.; Nakao, S.; Bachman, E.; Shafner, L.; Damokosh, A. I.; Ortiz, S.; Röth, A.; Peffault de Latour, R. Ravulizumab (ALXN1210) vs Eculizumab in C5-Inhibitor–Experienced Adult Patients with PNH: The 302 Study. *Blood* **2019**, *133* (6), 540–549. https://doi.org/10.1182/BLOOD-2018-09-876805.

(24) Nishimura, J.; Yamamoto, M.; Hayashi, S.; Ohyashiki, K.; Ando, K.; Brodsky, A. L.; Noji, H.; Kitamura, K.; Eto, T.; Takahashi, T.; Masuko, M.; Matsumoto, T.; Wano, Y.; Shichishima, T.; Shibayama, H.; Hase, M.; Li, L.; Johnson, K.; Lazarowski, A.; Tamburini, P.; Inazawa, J.; Kinoshita, T.; Kanakura, Y. Genetic Variants in C5 and Poor Response to Eculizumab. *N. Engl. J. Med.* **2014**, *370* (7), 632–639. https://doi.org/10.1056/nejmoa1311084.

(25) Ueda, Y.; Takamori, H.; Jang, J.-H.; Ganzel, C.; Langemeijer, S.; Osato, M.; Muus, P.; Lee, J.-W.; Nishimura, J.-I.; Kanakura, Y. Current Status and Optimal Management of Eculizumab Poor-Responders Due to C5 Polymorphisms. *Blood* **2018**, *132* (Supplement 1), 2323–2323. https://doi.org/10.1182/blood-2018-99-111230.

(26) Razzak, M. Mutations in C5 Explain Eculizumab Resistance. *Nat. Rev. Nephrol. 2014 104* **2014**, *10* (4), 182–182. https://doi.org/10.1038/nrneph.2014.30.

(27) Nishimura, J.; Yamamoto, M.; Hayashi, S.; Ohyashiki, K.; Ando, K.; Noji, H.; Kitamura, K.; Eto, T.; Ando, T.; Masuko, M.; Shibayama, H.; Hase, M.; Lan, L.; Tamburini, P.; Inazawa, J.; Kinoshita, T.; Kanakura, Y. A Rare Genetic Polymorphism in C5 Confers Poor Response to the Anti-C5 Monoclonal Antibody Eculizumab by Nine Japanese Patients with PNH. *Blood* **2012**, *120* (21), 3197–3197. https://doi.org/10.1182/BLOOD.V120.21.3197.3197.

(28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research*. Oxford University Press January 1, 2000, pp 235–242. https://doi.org/10.1093/nar/28.1.235.

(29) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815. https://doi.org/10.1006/JMBI.1993.1626.

(30) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

https://doi.org/10.1002/jcc.20084.

(31) Langemeijer, S.; Nishimura, J.-I.; Weston-Davies, W.; Nunn, M. A.; Kanakura, Y.; Mackie, I. J.; Muus, P. C5 Polymorphism in a Dutch Patient with Paroxysmal Nocturnal Hemoglobinuria (PNH) and No Asian Ancestry, Resistant to Eculizumab, but in Vitro Sensitive to Coversin. *Blood* **2015**, *126* (23), 1209–1209. https://doi.org/10.1182/blood.v126.23.1209.1209.

(32) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; and Kollman, P. A. Amber 2021. University of California, San Francisco 2021.

(33) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G. W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27* (1), 112–128. https://doi.org/10.1002/PRO.3280.

(34) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural Summaries of PDB Entries. *Protein Sci.* **2018**, *27* (1), 129–134. https://doi.org/10.1002/PRO.3289.

(35) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(36) MATLAB 2019b. The MathWorks, Inc.: Natick, Massachusetts, United States.

(37) The pandas development team. Pandas-Dev/Pandas: Pandas. Zenodo March 12, 2019. https://doi.org/10.5281/ZENODO.3509134.

(38) Mckinney, W. Data Structures for Statistical Computing in Python. **2010**.

(39) Waskom, M. L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6* (60), 3021. https://doi.org/10.21105/JOSS.03021.

(40) Mongan, J.; Case, D. A.; McCammon, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048. https://doi.org/10.1002/JCC.20139.

(41) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant PH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10* (3), 1341–1352. https://doi.org/10.1021/CT401042B.

(42) Williams, T.; Kelley, C.; Lang, R.; Kotz, D.; Campbell, J.; Elber, G.; Woo, A. Gnuplot. 2012.

CHAPTER 3

Spontaneous Embedding of DNA Mismatches within the RNA:DNA Hybrid of CRISPR-Cas9

3.1 Introduction

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 is the core of a transformative genome editing technology that is innovating life science with cutting-edge impact in basic and applied biosciences.[1,2] This technology is based on a protein/nucleic acid complex, composed of the endonuclease Cas9, which associates with guide RNAs to recognize and cleave complementary DNA sequences **(Figure 3.1)**.[3] The Cas9 protein performs a site-specific recognition of the DNA, by binding a short sequence of 2-5 nucleotides, known as a Protospacer-Adjacent Motif (PAM), located within the DNA.[4] Upon PAM binding, the DNA base pairs guide the RNA with one strand (i.e., the so-called target strand, TS) to form an 20 base-paired RNA:DNA hybrid structure, while the other DNA non-target strand (NTS) is displaced and subsequently accommodated in the protein.

The formation of a well-matched RNA:DNA hybrid is a fundamental step of the CRISPR-Cas9 function.[5] Indeed, upon formation of the RNA:DNA hybrid, the catalytic HNH domain can change conformation from an inactive state (in which the catalysis is hampered, **Figure 3.1A**)[6,7] to a catalytically active conformation, which approaches the cleavage site on the TS (**Figure 3.1B**).[8] In spite of this fundamental requirement, the presence of DNA mismatches at specific positions of the RNA:DNA hybrid still enables

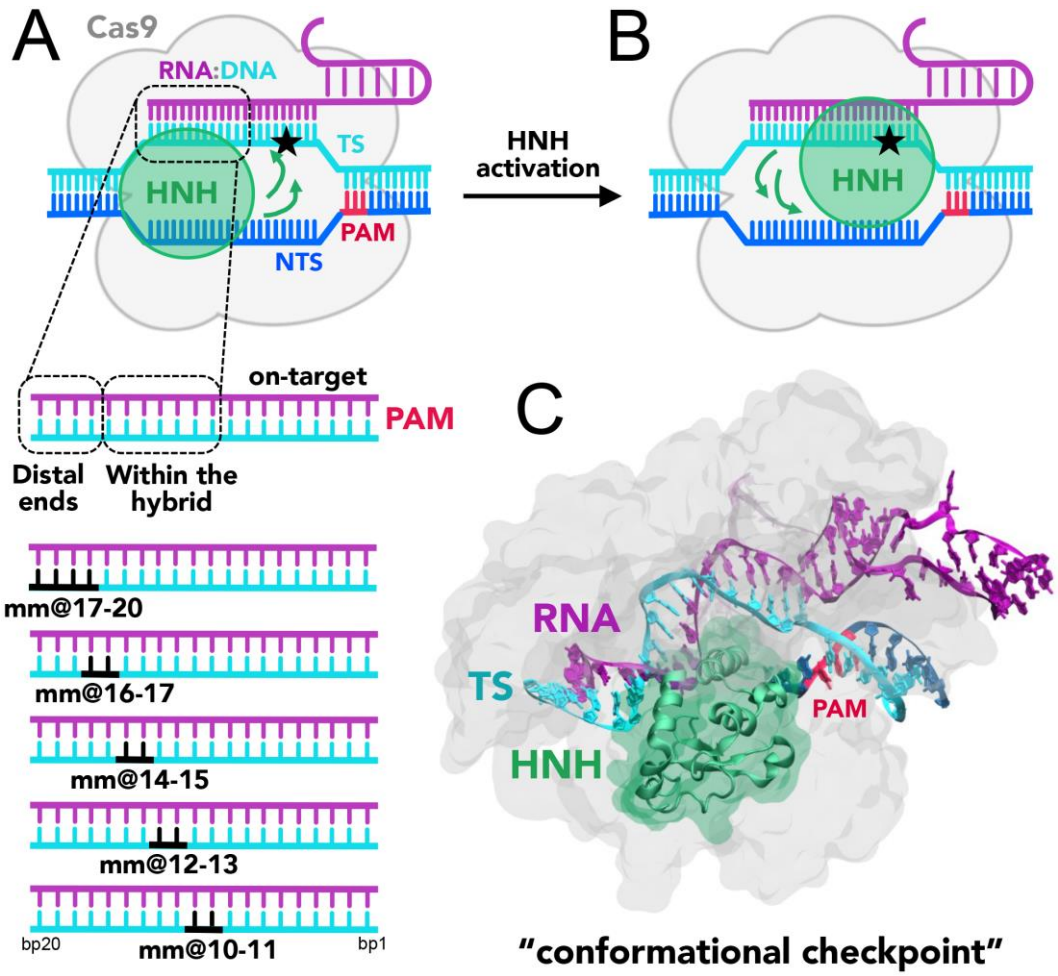

Figure 3.1 (A-B) Cartoon of the endonuclease Cas9 (grey) in complex with the nucleic acids. The DNA target strand (TS, cyan) base pairs the guide RNA (magenta), forming an RNA:DNA hybrid, while the DNA non-target strand (NTS, blue), which also includes the PAM recognition region (red), is displaced. Two conformational states of the catalytic HNH domain (green) are shown. In **(A)**, HNH assumes an inactive "conformational checkpoint" state, which requires a conformational transition (indicated using green arrows) to reach the activated state **(B)**, where it approaches the cleavage site on the TS (indicated using a star). On the bottom panel, a close-up view of the RNA:DNA hybrid highlights the regions at PAM distal ends and within the heteroduplex. In this work, base pair mismatches "mm" have been introduced at positions 17 to 20, 16-17, 14-15, 12-13 and 10-11 (shown in black). **(C)** X-ray structure of CRISPR-Cas9 identifying the "conformational checkpoint" state.[9] The protein is shown in molecular surface, with the HNH domain in green. The nucleic acids are shown as ribbons, color-coded as in the cartoon in panel **(A)**.

the partial activation of the HNH domain.[9,10] This leads to the off-target cleavages, which limit the applicability of CRISPR-Cas9, resulting in mutations at sites in the genome other than the desired target site. Several biophysical studies have investigated the effect of base pair mismatches within the RNA:DNA hybrid on the conformational dynamics of CRISPR-Cas9.[11–14] Single molecule and kinetics studies have revealed that the presence of 4 base pair mismatches at PAM distal ends can trap the catalytic HNH domain in an inactive conformation also referred to as "conformational checkpoint" (**Figure 3.1**, shown as a cartoon in panel A and as a 3D structure in panel B).[12] As a consequence, the cleavage of the TS gets hampered owing to lack of conformational changes that bring HNH in immediate vicinity to the cleavage site. Inversely, up to 3 base pair mismatches at PAM distal ends still allow the repositioning of HNH, thereby resulting in off-target cleavages. These studies indicate the occurrence of off-target cleavage is linked to the conformational states of HNH. In a recent computational study, we employed molecular dynamics (MD) simulations to investigate the factors affecting the HNH conformational dynamics prior to activation.[15] Our study employed the Gaussian accelerated MD (GaMD) method,[16] to broadly explore the conformational space of CRISPR-Cas9 in complex with an on-target DNA and in the presence of base pair mismatches. These simulations have revealed that the presence of 4 base pair mismatches at PAM distal sites (i.e., at positions 17–20 of the RNA:DNA hybrid) induced an extended opening of the RNA:DNA hybrid, with formation of conserved interactions between the TS and the HNH domain. This effectively decreased the conformational mobility of the HNH domain. Contrariwise, up to 3 base pair mismatches (at positions 18–20) display a lower

conformational effect on the RNA:DNA hybrid, and do not affect the conformational dynamics of HNH. These simulations thereby provided a theoretical rationale for the experimental evidence describing the molecular interactions that "lock" HNH in the presence of 4 base pair mismatches at PAM distal ends.[11,12,14]

However, mechanistic investigations of how DNA mismatches located upstream of the RNA:DNA heteroduplex affect the conformational dynamics of the hybrid structure and the HNH "conformational checkpoint" are absent. Knowledge of the conformational changes arising from base pair mismatches in the middle of the RNA:DNA hybrid are important to gain a deeper understanding of the molecular determinants of off-target binding, which consequently may offer insights for improving the specificity of CRISPR-Cas9. Moreover, understanding how base pair mismatches affect the RNA:DNA structure is important to characterize the dynamics of the heteroduplex itself. This is a key point considering the importance of RNA:DNA hybrids in a variety of biological processes, such as transcription, formation of Okazaki's fragments and R-loop structures, as well as in eukaryotic chromosomes.[17–22]

In this research report, we extend our recent investigations to 4 additional model systems, which include base pair mismatches upstream of the RNA:DNA hybrid **(Figure 3.1)**. Analysis of the results has been performed in comparison with our recently published data,[15] thereby evaluating similarities and differences with base pair mismatches at PAM distal ends and with an on-target DNA. We show that while base pair mismatches at PAM distal sites induce an opening of the RNA:DNA hybrid, at upstream positions they are incorporated within the heteroduplex, with minor effect on

the protein-nucleic acid interactions. Additionally, mismatches at PAM distal sites limit the mobility of HNH in the "conformational checkpoint" state and consequently affect its activation toward DNA cleavage. Conversely, mismatched pairs within the heteroduplex do not affect the dynamics of HNH, which can freely change conformation as needed to perform DNA cleavages.

3.2 Materials and Methods

Structural models have been based on the X-ray structure of the *Streptococcus pyogenes* CRISPR-Cas9 complex (4UN3.pd, 2.58 Å resolution),[6] which captures the inactivated state of the HNH domain (i.e., "conformational checkpoint").[12] MD simulations have been performed applying a well-established protocol for protein/nucleic acid complexes, which employs the Amber ff12SB force field, including the ff99bsc0[23] corrections for DNA and the ff99bsc0+χOL3[24,25] corrections for RNA. To broadly explore the conformational space of CRISPR-Cas9, we employed a recent accelerated MD (aMD) simulations method.[16] Specifically, we applied a Gaussian aMD (GaMD) method, which adds a harmonic boost potential to smoothen the potential energy surface, thereby decreasing energy barriers and accelerating transitions between the low-energy states (a complete description of the method is reported as a **Supplementary Material Appendix B.1**). The method has extended the use of aMD to large biomolecular systems, with applications of this method to G-protein coupled receptors,[26,27] the Mu opioid receptor,[28,29] T-cell receptors[30] and CRISPR-Cas9.[15,31,32]

3.3 Results and Discussion

To understand the effect of DNA mismatch pairs within the RNA:DNA hybrid on the conformational dynamics of CRISPR-Cas9 and on the HNH domain, we carried out molecular simulations. These investigations have been carried out in analogy to our recent study, which has investigated the effect of mismatch pairs at PAM distal ends.[15] In detail, molecular simulations have been performed on the X-ray structure of CRISPR-Cas9 capturing a "conformational checkpoint" state of the HNH domain (i.e., 4UN3.pdb),[6] thereby enabling us to understand if and how base pair mismatches could affect the dynamics of HNH prior its activation. A GaMD method has been employed,[16] adding a boost potential to the simulation that accelerates transitions between low-energy states (see section 3.2 Material and Methods). The method has been shown to enhance a broad sampling of the conformational space in large biomolecular systems,[26–30] including CRISPR-Cas9 as apo form and in complex with nucleic acids,[31,32] or bound to off-target DNAs.[15] Recently, GaMD has shown to capture long time scale motions in agreement with NMR relaxation experiments, showing that the method can efficiently capture the dynamics of large protein/nucleic acid complexes.[33] A set of model systems have been built; introducing couples of base pair mismatches "mm" within the hybrid complex at positions 10 to 17 (i.e. mm@10-11, @12-13, @14-15 and @16-17, **Figure 3.1A**, bottom panel). The dynamics of these systems have been compared with the simulations of CRISPR-Cas9 binding to an on-target DNA and including 1 to 4 mismatches at PAM distal sites (i.e. mm@17-20, @18-20, @19-20 and @20), which we have recently published.[15] For each system, ~1 μs of conformational sampling has been performed (see

section 3.2 Material and Methods), as in our previous study and by employing the same simulations conditions, thereby enabling proper comparison.

3.3.1 Dynamics of the RNA:DNA hybrid in the presence of DNA mismatches.

MD simulations of CRISPR-Cas9 bound to a fully matched RNA:DNA hybrid (i.e., on-target system) have revealed a stable Watson-Crick base pairing (**Figure 3.2A**, left panel), both at PAM distal ends and within the heteroduplex. Notably, transient openings at the end of a DNA duplex, or base flipping are not unusual over long timescales in MD simulations, as shown by several research groups.[34–38] However, in the simulations of the on-target CRISPR-Cas9 system, the RNA:DNA hybrid maintains the Watson-Crick base pairing, stabilized by the protein framework, as observed in several conventional and GaMD simulations of this system.[31,39] Contrariwise, in the presence of base pair mismatches at PAM distal ends (i.e. at positions 16 to 20), we previously observed the opening of the RNA:DNA hybrid (central panel).[15] Here, when we introduce DNA mismatches at the upstream positions (i.e. @10-11, @12-13, @14-15), we detect that the RNA:DNA hybrid preserves its overall shape (right panel), similarly to what observed in the on-target system. In order to estimate the conformational changes of the RNA:DNA hybrid, we analyzed in all simulated systems, the minor groove width from PAM distal ends up to the middle of the RNA:DNA hybrid **(Figure 3.2B)**. As a result, we observe that the presence of base pair mismatches at PAM distal ends (i.e., mm@17 to 20) induced an increase of the minor groove width at positions 18-20, which corresponds to the hybrid opening. Notably, the hybrid opening is also observed when including

mismatches at positions 16 and 17. This indicates that, perturbations at position 17 (as in the mm@17-20 and mm@16-17 systems) lead to major distortions in the heteroduplex. Conversely, when introducing mismatches at positions 10-11, 12-13 and 14-15, the minor groove width of the RNA:DNA hybrid preserves the conformation of the on-target system.

To understand the effects of the base pair mismatches on the Watson-Crick base pairing, we have used a key geometrical descriptor of the base pair complementarity. We have selected the Propeller Twist parameter **(Figure 3.2C)**, which describe the rotation of couples of base pairs with respect to each other. Based on our previous study, this parameter enables us to properly characterize alterations in the base pairing along the RNA:DNA hybrid.[15] **Figure 3.2C** reports the distribution of the Propeller Twist angle along the dynamics for each base pair from PAM distal ends up to the middle of the RNA:DNA hybrid. (i.e., from base pair bp20 to bp9). This analysis shows that the presence of base pair mismatches at positions 16 to 20 induces the remarkable loss of base pairing at PAM distal ends, as shown in the mm@20, m@19-20, mm@18-20, mm@17-20 and in the mm@16-17 systems ("major distortion" in **Figure 3.2C**). On the contrary, the geometrical requirements for the base pairing reveal "minor distortion" for mismatches within the RNA:DNA hybrid (i.e. mm@10-11, mm@12-13 and mm@14-15). Notably, this local distortion is due to the loss of base pair interactions (mainly H-bonds), which is typical between DNA mismatched pairs. However, the analysis of the minor groove width **(Figure 3.2B)** shows that the hybrid preserves its overall shape when base pair mismatches are introduced in the middle of the structure. Hence, a combined

analysis of the minor groove width and the base pair complementarity reveal that the presence of base pair mismatches within the hybrid does not influence the overall shape of the RNA:DNA hybrid, and that base pair mismatches result embedded within the heteroduplex structure.

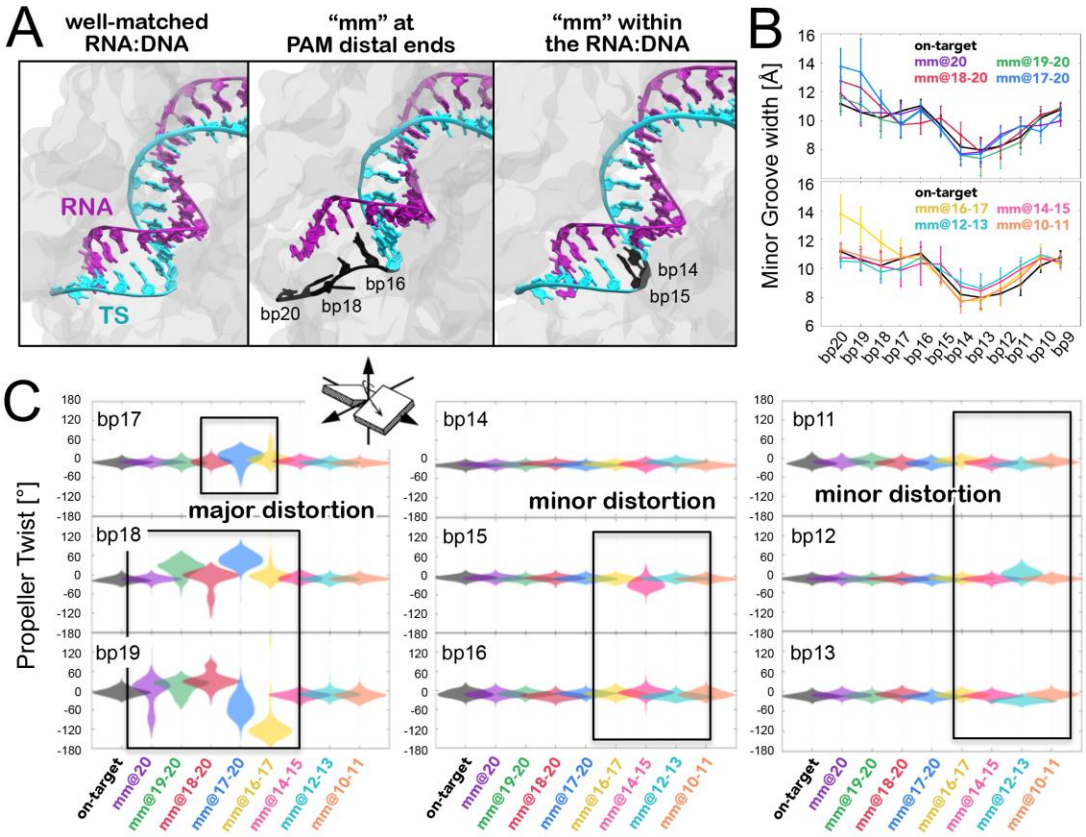

Figure 3.2. (A) Conformations adopted by the RNA:DNA hybrid, in the presence of an on-target DNA (left), including base pair mismatches "mm" at PAM distal ends (center) and within the heteroduplex (right). **(B)** Minor groove width measured at different levels of the RNA:DNA hybrid (i.e. from base pair bp20 to bp9) in the systems including "mm" at PAM distal ends (top panel) and within the heteroduplex (bottom panel). Data are compared with the on-target system. **(C)** Each graph reports the probability distribution (as violin plot) of the *Propeller Twist* angle for each base pair (bp) from PAM distal ends (bp19 to bp17) up to the middle of the RNA:DNA hybrid (bp16 to bp13), computed along the dynamics of each simulated system (reported on the x-axis). Regions of major and minor distortions are highlighted using boxes.

3.3.2 Mobility of the HNH domain in the presence of DNA mismatches.

Our previous study has revealed that in the presence of 4 base pair mismatches at PAM distal ends, the DNA TS establishes conserved interactions with the HNH domain.[15] These interactions restrict the mobility of HNH and affect its conformational activation toward DNA cleavage, while also contributing to the widening of the RNA:DNA hybrid. Here, in order to assess the conformational mobility of HNH in the presence of base pair mismatches within the RNA:DNA hybrid, we performed Principal Component Analysis (PCA). This analysis enabled to capture the essential degrees of freedom of the HNH domain (see section 3.2 Material and Methods). PCA has been carried out in comparison with the on-target system and with the system including 4 base pair mismatches at PAM distal ends (i.e. mm@17-20). **Figure 3.3A** reports the dynamics of the HNH domain along its first mode of motion (i.e., Principal Component 1, PC1), shown using arrows to indicate the direction and relative amplitude of the motions. The top panel shows a comparison between the system binding an on-target DNA and in the presence of 4 base pair mismatches at positions 17-20. In the mm@17-20 system, we observe that the unwound TS approaches the arrows corresponding to the HNH principal motion. A close-up view displays the interactions established by the DNA and the residues of the HNH domain. Notably, these interactions are stable along the dynamics, as discussed in our previous paper. The bottom panel reports the PCA analysis for the simulated systems including base pair mismatches within the RNA:DNA hybrid. We observe that for base pair mismatches at positions 16-17, the TS displays a similar unwinding of the mm@17-20 system, with conserved interactions established with the

HNH domain (close-up view). Indeed, the interaction between the nucleobases at position 17 and R904 is conserved in the two systems. This indicates that local distortions due to mismatched nucleobases at position 17, which is in close proximity to the HNH α-helices, can critically affect the dynamics of HNH. We note that the interaction established at position 17 involves the DNA backbone (rather than the nucleobases), which suggests that this interaction is not specific, but rather could be established also in the presence of different mismatched nucleobases. This hypothesis, however, warrants further investigations, which are currently ongoing in our lab as a follow-up of this study. On the contrary, base pair mismatches @10-11, @12-13 and @14-15 do not result in the approach of the TS to the HNH domain, resembling what observed the dynamics of the on-target system (top panel).

In order to characterize the conformational space sampled by the HNH domain, we plotted the first versus the second principal components (PC1 vs PC2, **Figure 3.3B**). This analysis revealed that in the mm@17-20 system, HNH explores a narrower conformational space with respect to the remaining systems, indicating a diminished mobility. A narrow conformational space is also observed for the mm@16-17 system. As discussed above, in these two systems, the TS tightly interacts with the HNH domain, thereby limiting its conformational dynamics. In the systems including base pair mismatches within the RNA:DNA hybrid, the HNH domain assumes a wider conformational space, similar to what observed in the on-target system. This indicates that the dynamics of HNH is not significantly affected by base pair mismatches in the middle of the RNA:DNA hybrid.

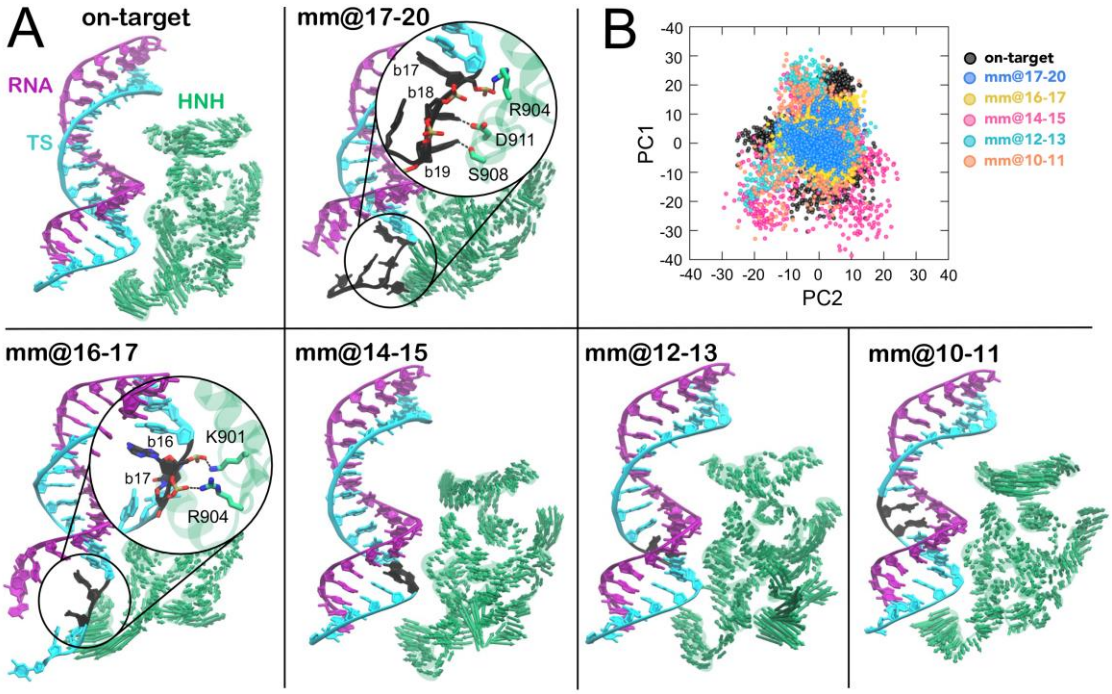

Figure 3.3. (A) "Essential dynamics",[40] derived from the first principal component (PC1), of the HNH domains in CRISPR-Cas9, binding an on-target DNA and base pair mismatches "mm" at positions 17-20 (top panel), 16-17, 14-15, 12-13 and 10-11 (bottom panel). PC1 is plotted on the three-dimensional structure of HNH (green) using arrows of sizes proportional to the amplitude of motions. The RNA:DNA hybrid is also shown. For the mm@17-20 and mm@16-17 systems, a close-up view shows the interaction between the unwound non-target strand and the HNH domain. **(B)** Projections of the first and second principal motions (PC1 vs PC2) for the HNH domain in the simulated systems (listed in the legend).

To further characterize the mobility of the systems and to understand the relation between the dynamics of the nucleic acids and the HNH domain, we performed cross-correlation (CC_{ij}) analysis. This analysis enabled us capturing coupled motions between the protein Cα atoms and the TS phosphate atoms (see section 3.2 Material and Methods). **Figure 3.4A** reports the CC_{ij} matrices computed between the residues of the HNH α-helices that locate in proximity of the hybrid, and the TS bases from position b20 (PAM distal ends) to position b9 (within the hybrid). Positive correlations ($CC_{ij} \geq 0$,

magenta) indicate highly coupled motions in the same direction, whereas anti-correlated motions display negative correlations ($CC_{ij} \leq 0$, green). A cartoon of the system, highlighting the regions used to compute the cross-correlations is shown in **Figure 3.4B**. For the sake of the clarity, the HNH α-helices in proximity of the hybrid are indicated in red (residues 890-900, Helix–A), yellow (901-910, Helix–B) and orange (911-920, Helix–C).

As a result of this analysis, in the presence of mismatches at PAM distal ends (i.e., in the mm@17-20 system) and at positions 16-17 (mm@16-17 system), Helix–A and Helix–B are highly correlated with the TS bases from position 18 to 14 (as highlighted using a box in **Figure 3.4A**). This indicates that the dynamics of the HNH and of the TS are mutually affected by each other, when in the presence of mismatched pairs at PAM distal ends. Moreover, we note that in the presence of mismatches at PAM distal ends, the DNA TS mainly interacts with Helix–B (**Figure 3.3A**, and also shown by Ricci and coauthors),[15] thereby affecting its conformational dynamics. Inversely, in the systems displaying base pair mismatches at upstream positions (mm@14 to 10), as well as in the on-target system, a weakening of the correlated motions can be seen. In these systems, there are no interactions being established between the TS and the HNH domain, signified by the diminished correlations between them. Overall, the cross-correlation analyses confirm that the presence of base pair mismatches at PAM distal ends affects the dynamics of HNH, while mismatches at upstream positions do not exert a relevant effect.

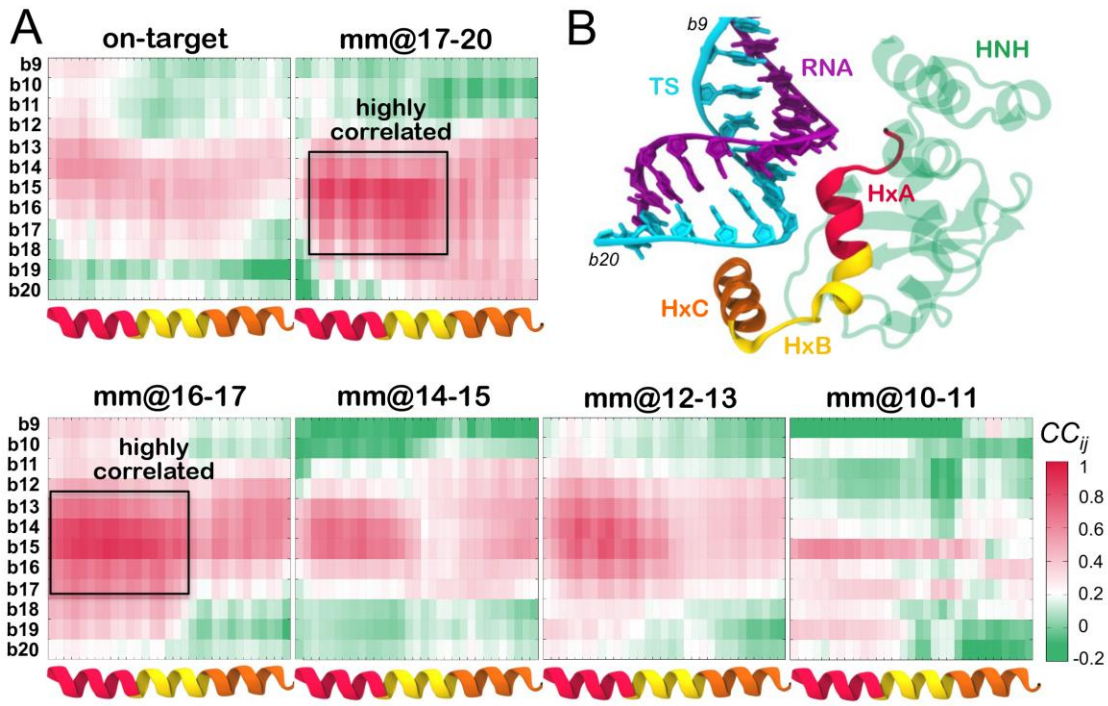

Figure 3.4. (A) Cross-Correlation (CC_{ij}) matrices computed between the HNH α-helices that locate in proximity of the hybrid (x-axis), and the DNA TS from position b20 to position b9 (y-axis). The CC_{ij} coefficients are computed between the protein Cα and the DNA phosphate atoms. Data are reported for CRISPR-Cas9 binding an on-target DNA and including base pair mismatches "mm" at positions 17-20 (top panel), as well as with "mm" at positions 16-17, 14-15, 12-13 and 10-11 (bottom panel). Positive correlations ($CC_{ij} \geq 0$) are shown in magenta, whereas anti-correlated motions display negative correlations ($CC_{ij} \leq 0$) are shown in green (legend on the bottom right). Two boxes indicate highly coupled motions in the mm@17-20 and mm@16-17 systems. **(B)** Cartoon of the system, displaying the regions used to compute the CC_{ij} matrix. The HNH domain is shown as cartoon (green), with the α-helices HxA (residues 890-900, red), HxB (residues 901-910, yellow) and HxC (residues 911-920, orange) in different colors. The RNA (violet) and the DNA TS (cyan) are shown as ribbons.

3.4 Conclusions

Here, molecular simulations have been used to characterize the conformational dynamics of CRISPR-Cas9 in the presence of base pair mismatches within the RNA:DNA hybrid. The simulations have shown that the presence of base pair mismatches at PAM distal ends of the RNA:DNA hybrid (i.e., positions 20 to 17) induce an opening of the heteroduplex.[15] As a result, newly formed interactions between the DNA TS and the catalytic HNH domain have been shown to "trap" HNH in an inactive "conformational checkpoint" state, hampering its activation for cleavage. On the contrary, base pair mismatches at upstream positions (i.e., within the RNA:DNA hybrid, at positions 14 to 10) are incorporated within the heteroduplex, with minor effect on the protein-nucleic acid interactions. Indeed, the presence of DNA mismatches within the hybrid does not affect the mobility of HNH, which is similar to that of the on-target system **(Figure 3.3)**. This suggests that mismatched base pairs within the RNA:DNA hybrid do not interfere with the process of HNH activation **(Figure 3.1A)**, where HNH changes in configuration from its "conformational checkpoint" state to an activated form are prone to cleave the DNA TS **(Figure 3.1A-B)**. Notably, these results agree with existing experimental studies and offer a rationale to the observed outcomes. Indeed, the presence of DNA mismatches at PAM distal ends has been experimentally shown to trap HNH in a "conformational checkpoint" state, likely due to interactions established with the DNA TS, as previously suggested.[11–13,41] However, mismatches in the middle of the hybrid are much more tolerated than at PAM distal ends, and lead to DNA cleavages. In light of this fact, our results indicate that mismatches at upstream positions (i.e., positions

14 to 10) still allow to preserve the overall structure of the RNA:DNA, without affecting the conformational dynamics of the catalytic HNH domain. As such, HNH can freely change conformation as needed to perform DNA cleavages **(Figure 3.1A-B)**. Overall, this brief research report constitutes a step forward in understanding the effect of DNA mismatches within the RNA:DNA hybrid in CRISPR-Cas9, offering insightful information on off-target effects. This work also forms the basis for further investigation, to characterize the effect of DNA mismatches along the entire RNA:DNA hybrid and therefore report an atomic-level understanding also for DNA mismatches at PAM-proximal sites (i.e. positions 1 to 9). These studies are currently ongoing in our laboratory, as inspired from the current work, taking also into account different conformations of the HNH **(Figure 3.1A)** domain and diverse mismatched nucleobases. Finally, we note that understanding how mismatched pairs affect the heteroduplex structure is per se important to understand the function of RNA:DNAs, which are critical in a variety of biological processes.[17–22]

In summary, this study provides an atomic-level understanding of the dynamic effects of the binding of DNA base pair mismatches within the RNA:DNA hybrid in CRISPR-Cas9. As a take-home message, the presence of mismatched pairs at distinctive locations of the RNA:DNA hybrid produces different conformational effects, which affect the protein counterpart. Specifically, mismatched pairs at PAM distal ends interfere with the activation of the catalytic HNH domain, while mismatches fully embedded in the RNA:DNA do not affect the HNH dynamics and enable its activation to cleave the DNA. This provides a reasonable explanation on why off-target sequences holding mismatches

at PAM distal ends are less likely to produce DNA cleavages in CRISPR-Cas9, than mismatched pairs within the heteroduplex, as experimentally observed.[11–13,41] These findings contribute in understanding the mechanistic basis of off-target effects in CRISPR-Cas9 and encourage novel experimental studies aimed at designing more specific variants of the system that prevent the onset of off-target effects.

3.5 References

(1) Doudna, J. A.; Charpentier, E. Genome Editing. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science (80-.).* **2014**, *346* (6213), 1258096. https://doi.org/10.1126/science.1258096.

(2) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *1576*, 1262–1278.

(3) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **2012**, *337*, 816–821.

(4) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9. *Nature* **2014**, *507* (7490), 62–67. https://doi.org/10.1038/Nature13011.

(5) Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A. Conformational Control of DNA Target Cleavage by CRISPR-Cas9. *Nature* **2015**, *527*, 110–113. https://doi.org/10.1038/nature15544.

(6) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. *Nature* **2014**, *513* (7519), 569–573. https://doi.org/10.1038/nature13579.

(7) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156* (5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001.

(8) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A. Structures of a CRISPR-Cas9 R-Loop Complex Primed for DNA Cleavage. *Science (80-.).* **2016**, *351* (6275), 867–871. https://doi.org/10.1126/science.aad8282.

(9) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells. *Nat. Biotechnol.* **2013**, *31* (9), 822–826. https://doi.org/10.1038/nbt.2623.

(10) Hsu, P. D.; Scott, D. A.; Weinstein, J. A.; Ran, F. A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E. J.; Wu, X.; Shalem, O.; Cradick, T. J.; Marraffini, L. A.; Bao, G.; Zhang, F. DNA Targeting Specificity of RNA-Guided Cas9 Nucleases. *Nat. Biotechnol.* **2013**, *31* (9), 827–832. https://doi.org/10.1038/nbt.2647.

(11) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced Proofreading Governs CRISPR–Cas9 Targeting Accuracy. *Nature* **2017**, *550* (7676), 407–410. https://doi.org/10.1038/nature24268.

(12) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. *Sci. Adv.* **2017**, *3*, eaao002. https://doi.org/10.1126/sciadv.aao0027.

(13) Singh, D.; Sternberg, S. H.; Fei, J.; Doudna, J. A.; Ha, T. Real-Time Observation of DNA Recognition and Rejection by the RNA-Guided Endonuclease Cas9. *Nat Commun* **2016**, *7*, 12778. https://doi.org/10.1038/ncomms12778.

(14) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep* **2018**, *22* (2), 372–382. https://doi.org/10.1016/j.celrep.2017.12.048.

(15) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651–662. https://doi.org/10.1021/acscentsci.9b00020.

(16) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theor. Comput.* **2015**, *11*, 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

(17) Nadel, J.; Athanasiadou, R.; Lemetre, C.; Wijetunga, N. A.; Ó Broin, P.; Sato, H.; Zhang, Z.; Jeddeloh, J.; Montagna, C.; Golden, A.; Seoighe, C.; Greally, J. M. RNA:DNA Hybrids in the Human Genome Have Distinctive Nucleotide Characteristics, Chromatin Composition, and Transcriptional Relationships. *Epigenetics and Chromatin* **2015**, *8* (1), 1–19. https://doi.org/10.1186/s13072-015-0040-6.

(18) Rich, A. Discovery of the Hybrid Helix and the First DNA-RNA Hybridization. *J. Biol. Chem.* **2006**, *281* (12), 7693–7696. https://doi.org/10.1074/JBC.X600003200.

(19) Shaw, N. N.; Arya, D. P. Recognition of the Unique Structure of DNA:RNA Hybrids. *Biochimie* **2008**, *90* (7), 1026–1039. https://doi.org/10.1016/J.BIOCHI.2008.04.011.

(20) Terrazas, M.; Genna, V.; Portella, G.; Villegas, N.; Sánchez, D.; Arnan, C.; Pulido-Quetglas, C.; Johnson, R.; Guigó, R.; Brun-Heath, I.; Aviñó, A.; Eritja, R.; Orozco, M. The Origins and the Biological Consequences of the Pur/Pyr

DNA·RNA Asymmetry. *Chem* **2019**, *5* (6), 1619–1631. https://doi.org/10.1016/J.CHEMPR.2019.04.002.

(21) Cheatham, T. E.; Kollman, P. A. Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes. *J. Am. Chem. Soc.* **1997**, *119* (21), 4805–4825. https://doi.org/10.1021/JA963641W.

(22) Palermo, G. Dissecting Structure and Function of DNA·RNA Hybrids. *Chem* **2019**, *5* (6), 1364–1366. https://doi.org/10.1016/J.CHEMPR.2019.05.015.

(23) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817–3829. https://doi.org/10.1529/biophysj.106.097782.

(24) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E. 3rd; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theor. Comput.* **2010**, *6*, 3836–3849. https://doi.org/10.1021/ct100481h.

(25) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(26) Miao, Y.; McCammon, J. A. Mechanism of the G-Protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3036–3041. https://doi.org/10.1073/pnas.1800756115.

(27) Miao, Y.; McCammon, J. A. Graded Activation and Free Energy Landscapes of a Muscarinic G Protein-Coupled Receptor. *Proc Natl Acad Sci U S A* **2016**, *113* (43), 12162–12167. https://doi.org/10.1073/pnas.1614538113.

(28) Liao, J. M.; Wang, Y. T. In Silico Studies of Conformational Dynamics of Mu Opioid Receptor Performed Using Gaussian Accelerated Molecular Dynamics. *J Biomol Struct Dyn* **2018**, 1–12. https://doi.org/10.1080/07391102.2017.1422025.

(29) Wang, Y.-T.; Chan, Y.-H. Understanding the Molecular Basis of Agonist/Antagonist Mechanism of Human Mu Opioid Receptor through Gaussian Accelerated Molecular Dynamics Method. *Sci. Rep.* **2017**, *7* (1), 7828. https://doi.org/10.1038/s41598-017-08224-2.

(30) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.;

McAffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W.; Dong, S.; Gee, M. H.; Jude, K. M.; Davis, M. M.; Groves, J. T.; Goddard, W. A.; Heath, J. R.; Evavold, B. D.; Vale, R. D.; Garcia, K. C. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **2018**, *174* (3), 672-687.e27. https://doi.org/10.1016/J.CELL.2018.06.017.

(31) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci.* **2017**, *114* (28), 7260–7265. https://doi.org/10.1073/pnas.1707645114.

(32) Palermo, G. Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59* (5), 2394–2406. https://doi.org/10.1021/acs.jcim.8b00988.

(33) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR–Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, jacs.9b10521. https://doi.org/10.1021/jacs.9b10521.

(34) Mura, C.; McCammon, J. A. Molecular Dynamics of a KB DNA Element: Base Flipping via Cross-Strand Intercalative Stacking in a Microsecond-Scale Simulation. *Nucleic Acids Res.* **2008**, *36* (15), 4941–4955. https://doi.org/10.1093/NAR/GKN473.

(35) Ricci, C. G.; de Andrade, A. S. C.; Mottin, M.; Netz, P. A. Molecular Dynamics of DNA: Comparison of Force Fields and Terminal Nucleotide Definitions. *J. Phys. Chem. B* **2010**, *114* (30), 9882–9893. https://doi.org/10.1021/jp1035663.

(36) Ma, Z.; Palermo, G.; Adhireksan, Z.; Murray, B. S.; von Erlach, T.; Dyson, P. J.; Rothlisberger, U.; Davey, C. A. An Organometallic Compound Displays a Unique One-Stranded Intercalation Mode That Is DNA Topology-Dependent. *Angew Chem Int Ed* **2016**, *128*, 7441–7444. https://doi.org/10.1002/anie.201602145.

(37) Alberto Pérez, †,‡; F. Javier Luque, § and; Modesto Orozco*, †,‡,∥. Dynamics of B-DNA on the Microsecond Time Scale. **2007**. https://doi.org/10.1021/JA0753546.

(38) Pérez, A.; Lankas, F.; Luque, F. J.; Orozco, M. Towards a Molecular Dynamics Consensus View of B-DNA Flexibility. *Nucleic Acids Res.* **2008**, *36* (7), 2379–2394. https://doi.org/10.1093/NAR/GKN082.

(39) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-Target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* **2016**, *2* (10), 756–763.

https://doi.org/10.1021/acscentsci.6b00218.

(40) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Bioinforma.* **1993**, *17* (4), 412–425. https://doi.org/10.1002/PROT.340170408.

(41) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* **2018**, *22* (2), 372–382. https://doi.org/10.1016/j.celrep.2017.12.048.

CHAPTER 4

Investigation of Residues within the HNH Domain of CRISPR-Cas9

4.1 Introduction

CRISPR (Clustered regularly interspaced short palindromic repeats)-Cas9 functions as a genome editing tool.[1–10] Guide RNAs are utilized by Cas9 for recognition and cleavage of DNA sequences. Site specific recognition is dependent on binding a protospacer adjacent motif (PAM), as PAM is located within the DNA.[11,12] One DNA strand will base pair match with the RNA to form an RNA:DNA hybrid. This strand is known as the target strand. The other DNA strand, known as the non-target strand, is displaced. The HNH domain of CRISPR-Cas9 is involved in cleavage of the DNA target strand.[1,11–17] For the active structure of CRISPR-Cas9, catalytic residues within the HNH domain included D839, H840, and N863, with a magnesium ion located at the catalytic center of the HNH domain.[18–20]

Classical MD and constant pH MD simulations were applied to observe the behavior of residues near the catalytic center of the HNH domain. Residues of interest included D839, D861, K862, N863, and K866. Distances were measured between the residues of interest and the magnesium ion. Constant pH MD was run to observe the pKa value of the H840 residue for wild type and mutant structures.

4.2 Methods

4.2.1 Structure Preparation

Wild type CRISPR-Cas9 was used a starting structure. Mutations were applied using UCSF Chimera.[21] Residues of interest, V838, D839, D861, K862, N863, and K866, were each individually mutated to alanine resulting in six mutated PDB structures. For constant pH MD simulations, the protonation state of HIS840 was changed to HIP in the wild type structure and mutated structures.

4.2.2 Classical MD Simulations

Amber ff12SB, ff99bsc0, and ff99 bsc0+χOL3 force fields were used.[22–24] Topology and coordinate files were generated for each structure using tleap.[25] Minimization was first performed at 2000 cycles with positional and distance restraints. Distance restraints were applied for residues near the catalytic center of the HNH domain. Positional restraints had a force constant of 300 kcal/mol · $Å^2$. The positional restraints were then removed for 1000 minimization cycles, while the distance restraints were kept. The system temperature was raised to 300K through multiple heating steps. Two heating steps in the NVT ensemble were run for 5ps each, bringing the temperature to 100K. Positional and distance restraints were applied. The temperature was then raised from 100K to 300K through two additional heating steps in the NPT ensemble for 100ps and 500ps. Positional restraints with a force constant of 25 kcal/mol · $Å^2$ were applied during heating for 100K to 200K, and then removed during 200K to 300K. Distance restraints remained through the heating steps. Equilibration was run in the NPT ensemble for 10ns

with distance restrains. Production was run in triplicate for each structure at one microsecond each, using the NVT ensemble and a simulation time step of 2fs. Restraints were not applied during production. The SHAKE algorithm was used to constrain bond lengths with hydrogen atoms.[25] For temperature control, Langevin dynamics were applied using a collision frequency of 1/ps. A 2ps pressure relaxation time was used. Simulations were run using Amber20.[25]

4.2.3 Classical MD Analysis

CPPTRAJ[26] was applied to calculate distances between residues near the catalytic center. Distances between ASP839:CG and Mg 1507, ASP861:CG and Mg 1507, LYS862:NZ and Mg 1507, ASN863:OD1 and Mg 1507, LYS866:NZ and Mg 1507, and DT1478:P and Mg 1507 were measured for each production run. RMSD was also determined using CPPTRAJ for residues near the catalytic center of the HNH domain.

4.2.4 Constant pH MD Simulations

For each solvated system, tleap created topology and coordinate files. Using cpinutil.py, a cpin file was generated.[25] Minimization was run for 5000 cycles using positional restraints with a force constant of 300 kcal/mol $\cdot$ Å^2. Heating was run for 3ns, where the temperature of the system was raised to 300K. During heating, positional restraints were applied with a force constant of 300 kcal/mol $\cdot$ Å^2. Equilibration was run for 8ns with distance restraints and a dihedral restraint. Production was run with and

without restraints on the dihedral of Histidine 840. Constant pH MD simulations were run for 20ns for pH values 1 to 14 with a simulation time step of 2fs.

4.2.5 Constant pH MD Analysis

Cphstats[25] was applied to determine the fraction protonated and populations of protonation states. Fraction deprotonated values were calculated from fraction protonated. Titration curves were generated from plotting fraction deprotonated vs pH values and fitting the data against the Hill equation **(Equation 4.1)** with gnuplot.[27,28]

$$f_d = \frac{1}{1+10^{n(pK_a - pH)}} \qquad \textbf{(4.1)}$$

The Hill equation calculations the pKa and value and Hill coefficient for the histidine residue.

4.3 Results

4.3.1 Classical MD Simulations

Distances between magnesium and ASP839, ASP861, LYS862, ASN863, LYS866, DT1478 were measured **(Figures 4.1-4.6)**. The distance between ASP839 and magnesium showed similar behavior across all runs for each trajectory. The mutated trajectories showed ASP839 at a distance of approximately 3Å from magnesium. For the K866A trajectory, runs 2 and 3 showed instances where ASP839 moved closer to magnesium and back to a similar distance of 3Å as before. For ASP861 and magnesium, the distances showed similar behavior for each simulation run. LYS862 appeared to fluctuate throughout the simulation trajectory, as the distances ranged from 5 to

approximately 15Å from magnesium. ASN863 was in close proximity to the magnesium and maintained the distance throughout the simulation trajectory. For the wild type and mutated structures, ASN863 stayed approximately 2Å from magnesium. LYS866 showed fluctuations throughout the simulation trajectory. The distances during each simulation run varied, ranging from 5 to 15Å. The distance between the phosphate ion of DT1478 and magnesium was consistent throughout the simulation trajectory in the wild type and mutated structures. RMSD of the residue near the catalytic center was also calculated **(Figure 4.7)**. For the wild type trajectory, RMSD for each run showed similar convergence. For V838A, K862A, and N863A trajectories, RMSD at each run was similar. RMSD differed between runs for D839A, D861A, and K866A trajectories.

4.3.2 Constant pH MD Simulations

Titration curves for HIS840 were generated for the wild type **(Figure 4.8)** and mutated structures **(Figures C.1-C.6)**. The pKa and Hill coefficient were calculated for each system. Production was run with and without restraints on the dihedral of HIS840. The restraints resulted in differences in pKa values and hill coefficients compared to the production run without restraints. For the wild type structure, the hill coefficient was larger for production with restraints compared to without restraints. The pKa value for the mutated structures differed from the wild type C5. For the mutated structures V838A, K862A, and N863A, the pKa of HIS840 was lower than the wild type, whereas D839A, D861A, and K866A structures had pKa values higher than the wild type structure.

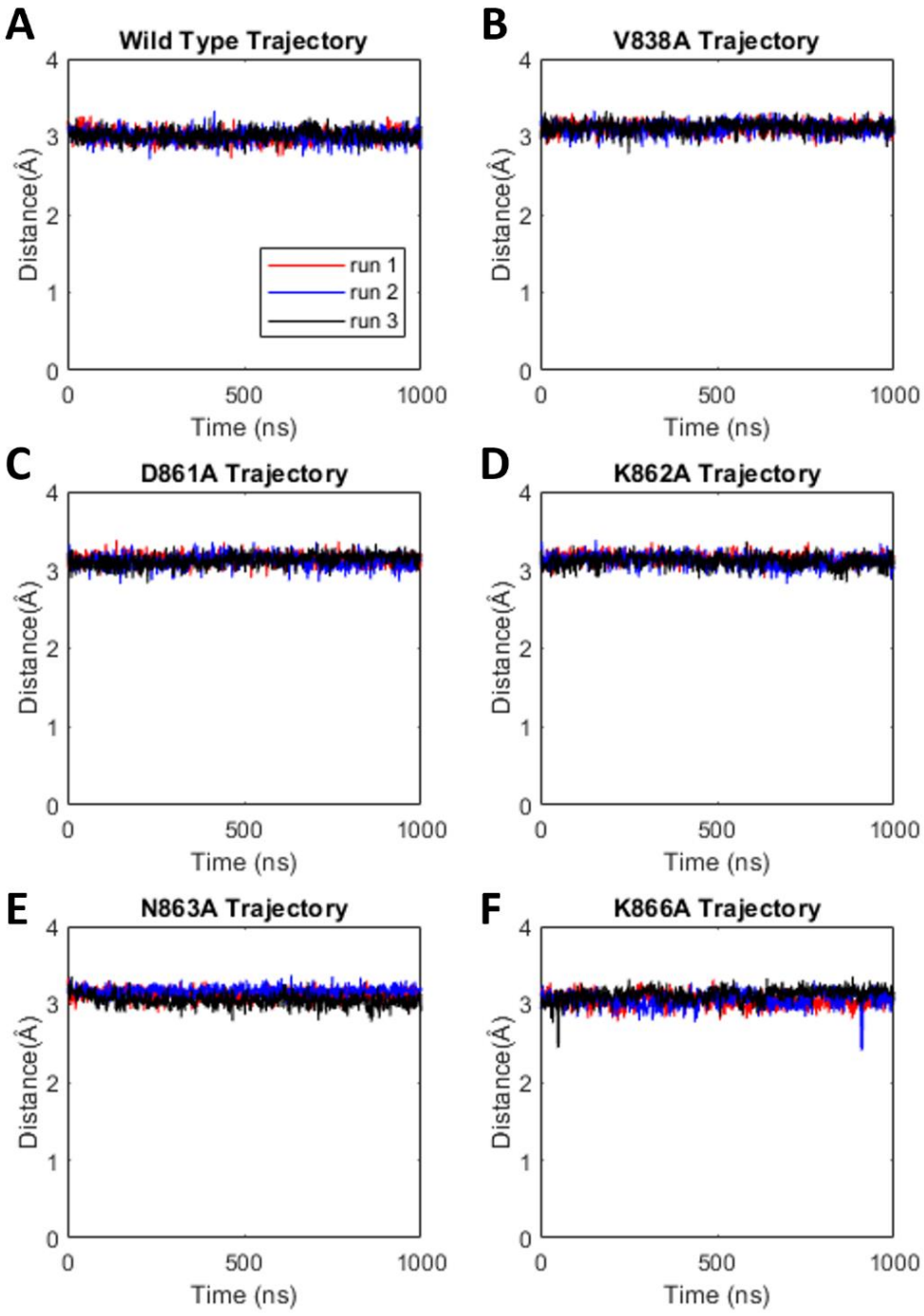

Figure 4.1. Distance between D839 and Mg 1507. Distances during each simulation run in the wild type **(A)**, V838A **(B)**, D861A **(C)**, K862A **(D)**, N863A **(E),** and K866A **(F)** trajectories.

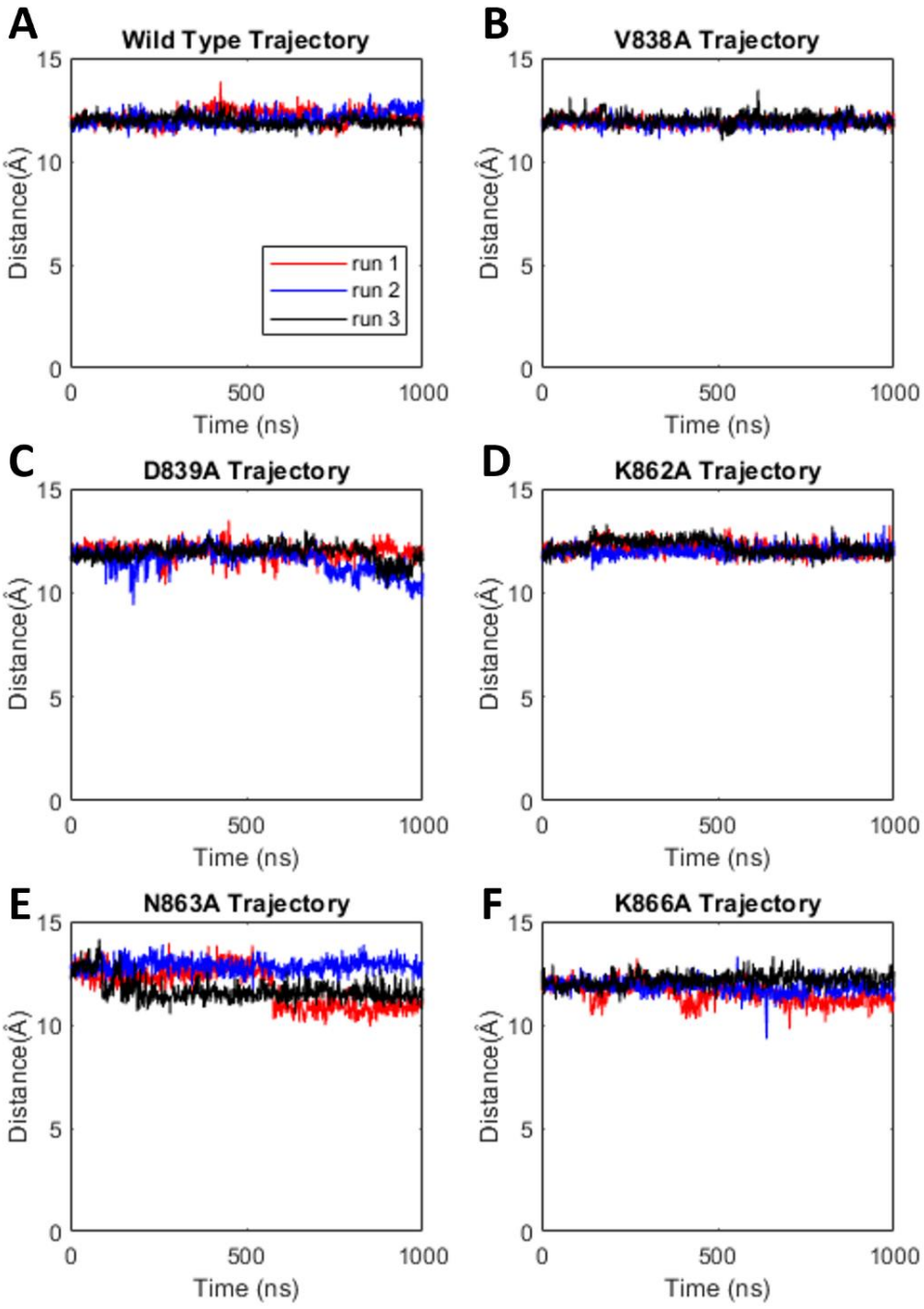

Figure 4.2. Distance between D861 and Mg 1507. Distance was measured during each simulation for the wild type **(A)**, V838A **(B)**, D839A **(C)**, K862A **(D)**, N863A **(E),** and K866A **(F)** trajectories.

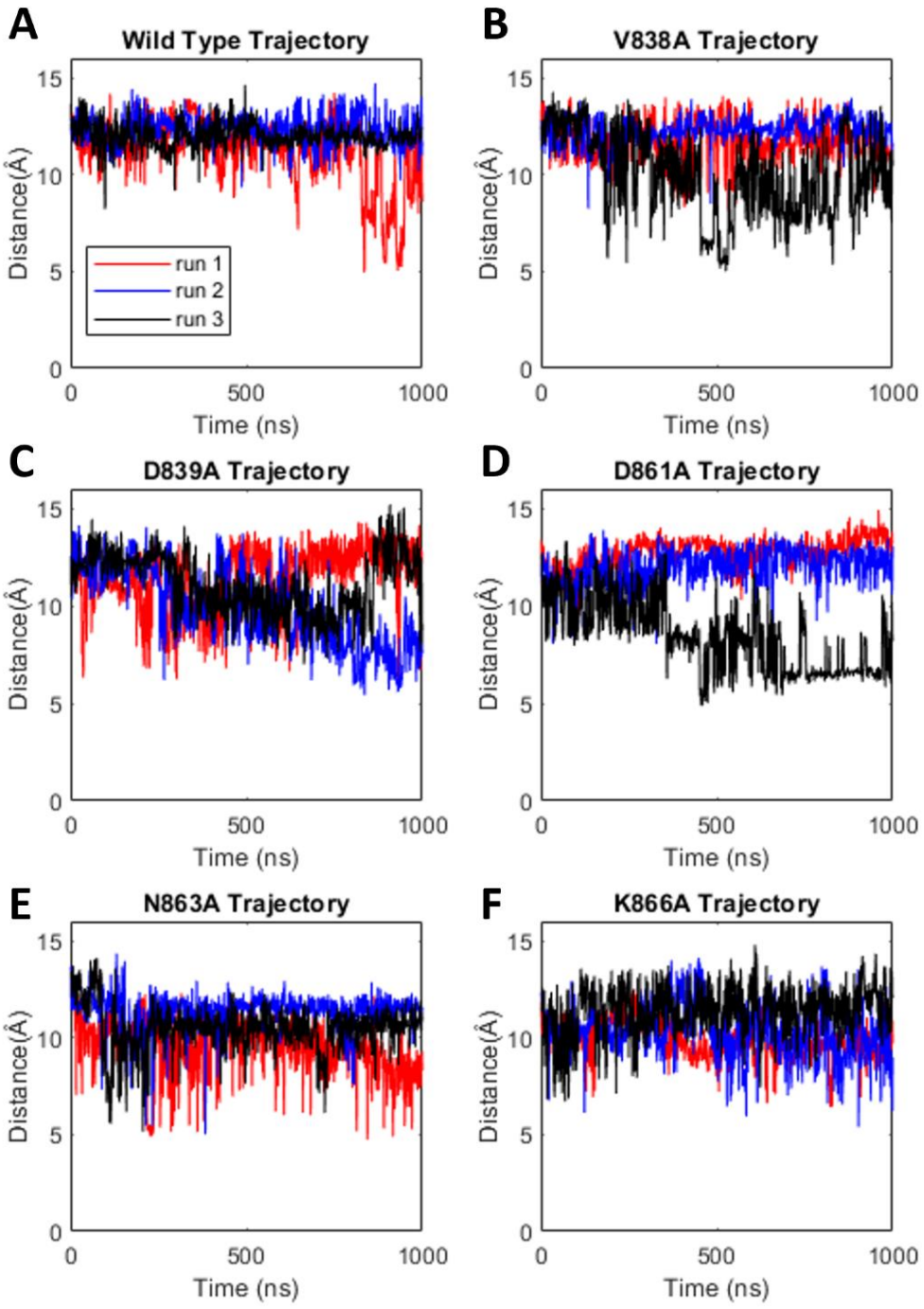

Figure 4.3. Distance between K862 and Mg 1507. Distance between lysine 862 and magnesium was measured for the wild type **(A)**, V838A **(B)**, D839A **(C)**, D861A **(D)**, N863A **(E),** and K866A **(F)** trajectories.

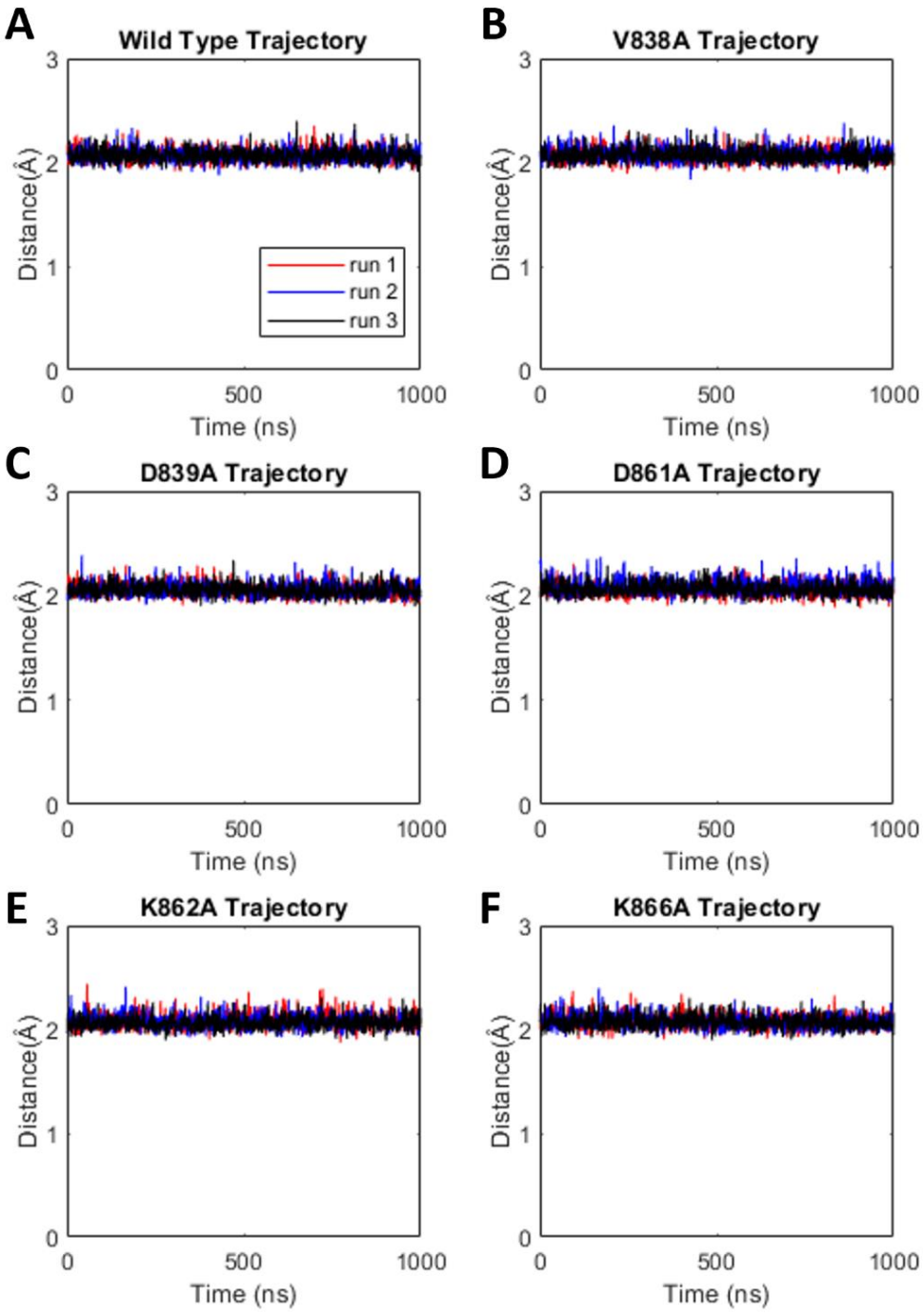

Figure 4.4. Distance between N863 and Mg 1507. Distance between asparagine 863 and magnesium for the wild type **(A)**, V838A **(B)**, D839A **(C)**, D861A **(D)**, K862A **(E),** and K866A **(F)** trajectories.

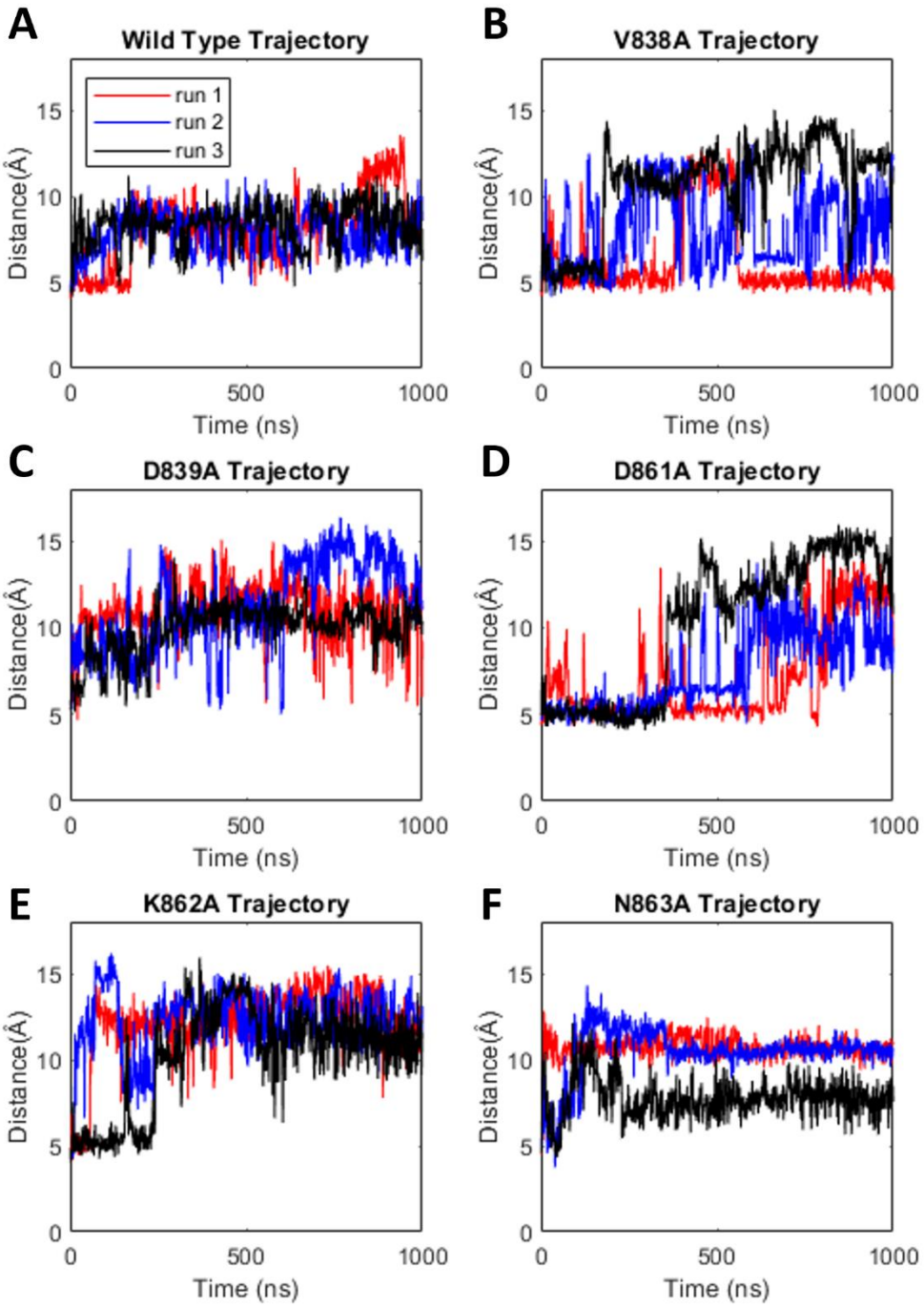

Figure 4.5. Distance between K866 and Mg 1507. Distance between lysine 866 and magnesium at the catalytic center was measured for wild type **(A)**, V838A **(B)**, D839A **(C)**, D861A **(D)**, K862A **(E),** and N863A **(F)** trajectories.

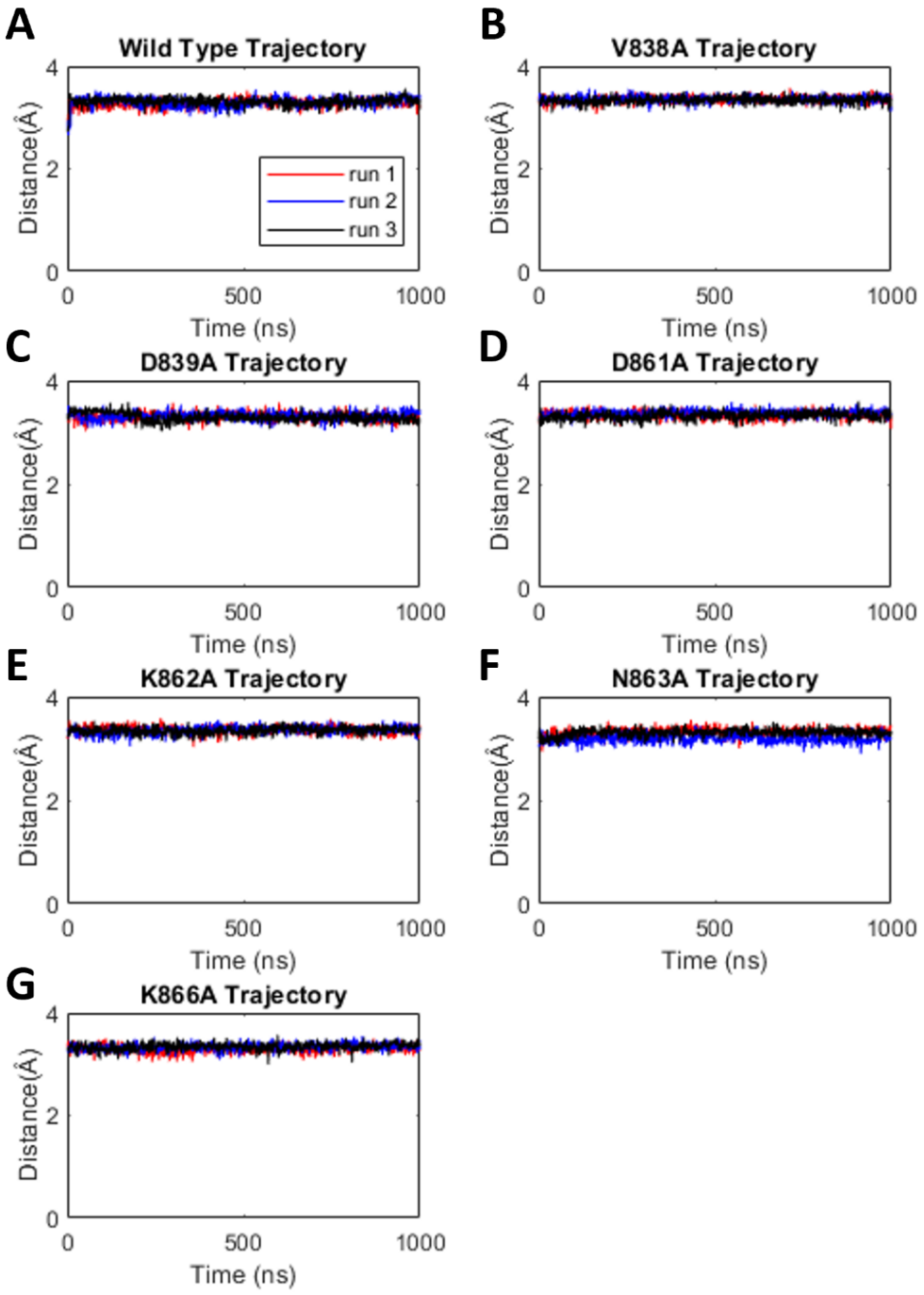

Figure 4.6. Distance between DT1478 and Mg 1507. Distances were calculated for each run for the wild type **(A)**, and V838A **(B)**, D839A **(C)**, D861A **(D)**, K862A **(E)**, N863A **(F)**, and K866A **(G)** simulation trajectories.

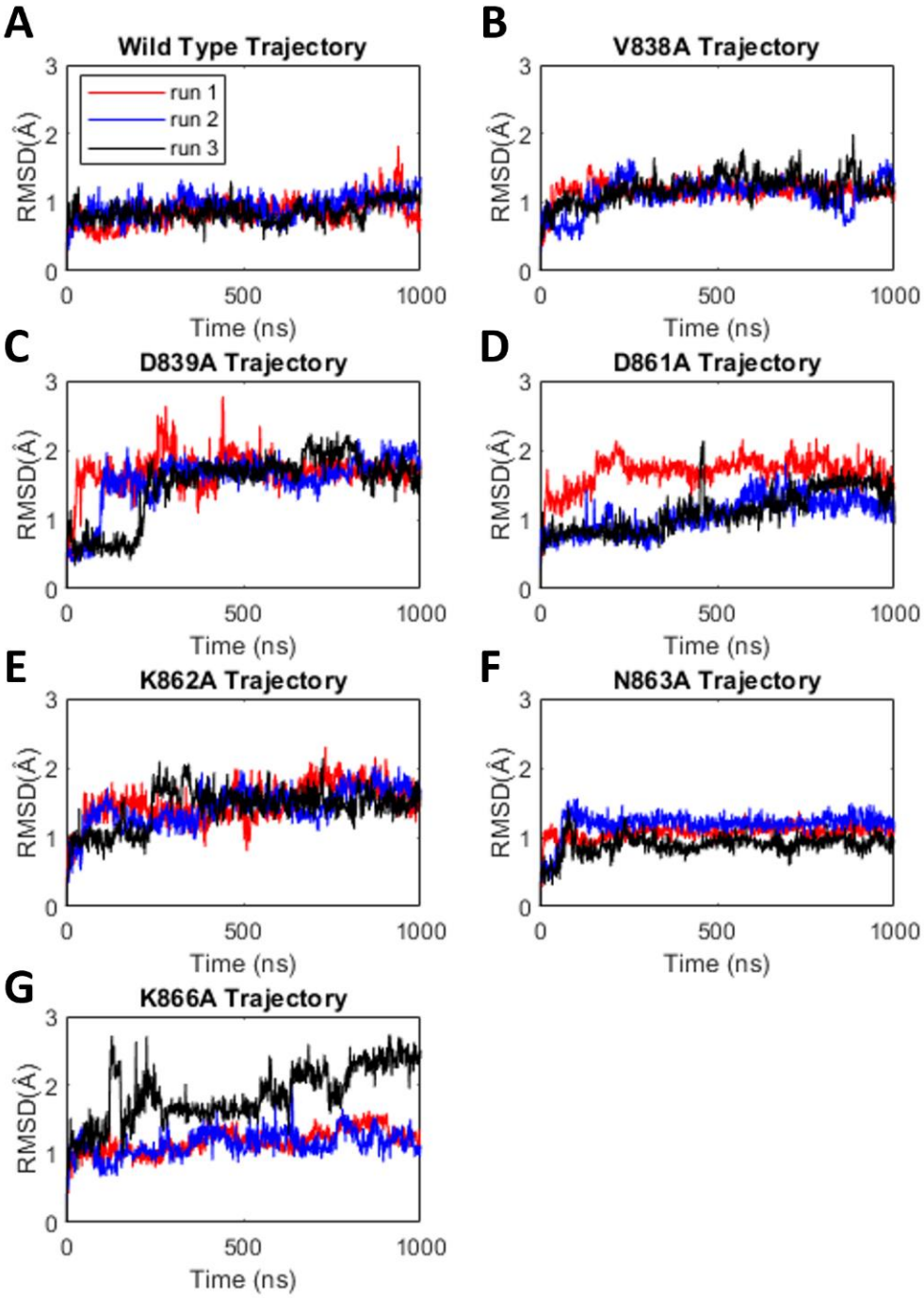

Figure 4.7. RMSD of residues near the catalytic center of the HNH domain. RMSD was calculated for each simulation run for the wild type **(A)**, and mutated structures V838A **(B)**, D839A **(C)**, D861A **(D)**, K862A **(E)**, N863A **(F)**, and K866A **(G)**.

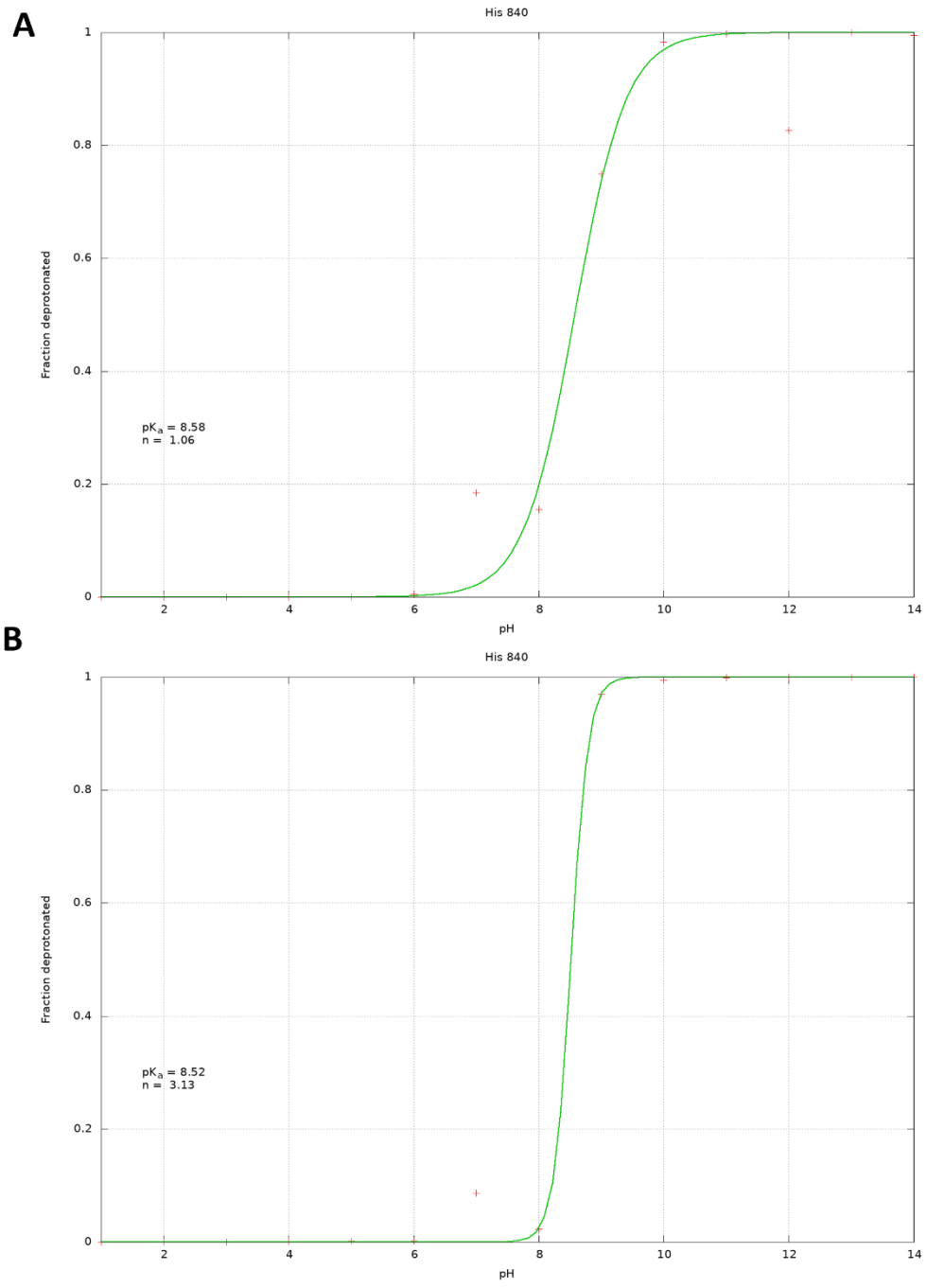

Figure 4.8. Titration curves for HIS840 in the wild type structure. Without restraints **(A)** and with restraints **(B)** applied on the dihedral during production.

4.4 Discussion

Distances between residues in the HNH domain were calculated. The differences in the mutations compared to the wild type was observed. The distances indicated the proximity of the residue of interest to the magnesium ion in the HNH domain. Calculated distances for each of the runs showed that D839 was within an approximate 3Å distance from the magnesium throughout the simulation trajectory. In the wild type simulations runs and the mutated structures, the distance between D839 and magnesium was maintained. The distance between D861 and the magnesium ion was similar throughout the simulation trajectories at approximately 12Å. The distance between K862 and magnesium fluctuated, ranging from approximately 5 to 15Å. The proximity of N863 and magnesium was consistent throughout the simulations of the wild type and mutated structures. The distance between N863 and magnesium was approximately 2Å. For K866, the distance from the magnesium ion fluctuated throughout the simulations. Distances fluctuated between approximately 5 to 15Å. The distance between the phosphate ion of DT1478 and magnesium was close in proximity throughout the simulations. RMSD was also calculated for the simulation trajectories. For V838A, K862A, and N863A trajectories, RMSD values were similar in the triplicate runs. RMSD values fluctuated for simulations runs for D839A, D861A, and K866A trajectories. Constant pH MD simulations showed different pKa values of HIS840 based on the mutation. For structures V838A, K862A, and N863A, the pKa was lower compared to the wild type structure. The pKa was calculated to be higher in the mutated structures for D839A, D861A, and K866A.

4.5 Conclusion

MD simulations allowed observation of behavior of residues near the catalytic center of the HNH domain. Mutations were applied residues of interest to determine the effect mutations have on the structures. Classical MD simulations results showed distances and RMSD. The distances between residues D839, D861, K862, N863, K866, and magnesium at the HNH domain were calculated. For residues D839, D861, and N863, the distances from magnesium were consistent between wild type and mutated structures. Distances for K862 and K866 to magnesium fluctuated throughout the simulation trajectories.

Constant pH MD simulations determined pKa values for HIS840 for the wild type structure and mutated structures. The mutated structures affected the pKa values as the value was lower than the wild type for V838A, K862A, and N863A, whereas the pKa values increased for D839A, D861A, and K866A.

4.6 References

(1) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651–662. https://doi.org/10.1021/acscentsci.9b00020.

(2) Zuo, Z.; Liu, J. Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Reports 2017 71* **2017**, *7* (1), 1–13. https://doi.org/10.1038/s41598-017-17578-6.

(3) Sternberg, S. H.; Lafrance, B.; Kaplan, M.; Doudna, J. A. Conformational Control of DNA Target Cleavage by CRISPR–Cas9. *Nat. 2015 5277576* **2015**, *527* (7576), 110–113. https://doi.org/10.1038/nature15544.

(4) Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key Role of the REC Lobe during CRISPR-Cas9 Activation by "Sensing", "Regulating", and "locking" the Catalytic HNH Domain. *Q. Rev. Biophys.* **2018**, *51*. https://doi.org/10.1017/S0033583518000070.

(5) Wu, X.; Kriz, A. J.; Sharp, P. A. Target Specificity of the CRISPR-Cas9 System. *Quant. Biol. 2014 22* **2014**, *2* (2), 59–70. https://doi.org/10.1007/S40484-014-0030-X.

(6) Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc. 2013 811* **2013**, *8* (11), 2281–2308. https://doi.org/10.1038/nprot.2013.143.

(7) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157* (6), 1262–1278. https://doi.org/10.1016/J.CELL.2014.05.010.

(8) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (28), 7260–7265. https://doi.org/10.1073/pnas.1707645114.

(9) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139* (45), 16028–16031. https://doi.org/10.1021/jacs.7b05313.

(10) Doudna, J. A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science (80-.).* **2014**, *346* (6213). https://doi.org/10.1126/SCIENCE.1258096.

(11) Mitchell, B. P.; Hsu, R. V.; Medrano, M. A.; Zewde, N. T.; Narkhede, Y. B.; Palermo, G. Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9. *Front. Mol. Biosci.* **2020**, *7*, 39. https://doi.org/10.3389/fmolb.2020.00039.

(12) Nierzwicki, Ł.; Arantes, P. R.; Saha, A.; Palermo, G. Establishing the Allosteric Mechanism in CRISPR-Cas9. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11* (3), e1503. https://doi.org/10.1002/WCMS.1503.

(13) Palermo, G.; Ricci, C. G.; McCammon, J. A. The Invisible Dance of CRISPR-Cas9. *Phys. Today* **2019**, *72* (4), 30. https://doi.org/10.1063/PT.3.4182.

(14) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59* (5), 2394–2406. https://doi.org/10.1021/acs.jcim.8b00988.

(15) Jiang, F.; Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46*, 505–529. https://doi.org/10.1146/ANNUREV-BIOPHYS-062215-010822.

(16) Huai, C.; Li, G.; Yao, R.; Zhang, Y.; Cao, M.; Kong, L.; Jia, C.; Yuan, H.; Chen, H.; Lu, D.; Huang, Q. Structural Insights into DNA Cleavage Activation of CRISPR-Cas9 System. *Nat. Commun. 2017 81* **2017**, *8* (1), 1–9. https://doi.org/10.1038/s41467-017-01496-2.

(17) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. *Sci. Adv.* **2017**, *3* (8). https://doi.org/10.1126/sciadv.aao0027.

(18) Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V. J. T.; Hayatshahi, H. S.; Rajan, R.; Wang, Y. C.; Liu, J. Structural and Functional Insights into the Bona Fide Catalytic State of Streptococcus Pyogenes Cas9 HNH Nuclease Domain. *Elife* **2019**, *8*. https://doi.org/10.7554/eLife.46500.

(19) Tang, H.; Yuan, H.; Du, W.; Li, G.; Xue, D.; Huang, Q. Active-Site Models of Streptococcus Pyogenes Cas9 in DNA Cleavage State. *Front. Mol. Biosci.* **2021**, *8*, 235. https://doi.org/10.3389/fmolb.2021.653262.

(20) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156* (5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001.

(21) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for

Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/JCC.20084.

(22) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829. https://doi.org/10.1529/BIOPHYSJ.106.097782.

(23) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836–3849. https://doi.org/10.1021/ct100481h.

(24) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(25) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; and Kollman, P. A. Amber 2021. University of California, San Francisco 2021.

(26) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(27) Williams, T.; Kelley, C.; Lang, R.; Kotz, D.; Campbell, J.; Elber, G.; Woo, A. Gnuplot. 2012.

(28) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant PH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10* (3), 1341–1352. https://doi.org/10.1021/CT401042B.

CHAPTER 5

Conclusion

5.1 Summary

MD simulations provided insight to the behavior of residues involved in biomolecular systems. Through classical and constant pH MD simulations, information about residues involved in biomolecular systems were determined. For the complement system, wild type C5 and mutated C5 structures were observed when binding to complement inhibitors eculizumab and ravulizumab. MD simulations determined key residues involved in binding. Intermolecular interactions indicated residues involved in hydrogen bonds and salt bridges throughout the simulation trajectories. Classical MD simulations allowed observation of perturbations. By simulating wild type C5 and mutated C5 structures, information about how mutations affected the systems was obtained. Residues important for binding when eculizumab was in complex with C5, versus when ravulizumab was in complex with C5 were determined. Through constant pH MD simulations, pKa values were determined for the histidine residues form the heavy chain of ravulizumab.

Classical MD simulations on the CRISPR-Cas9 structure observed the behavior of residues near the catalytic center of the HNH domain. The effect of mutations was analyzed through RMSD and distance calculations. By mutating individual residues of interest to alanine, and comparing the structures to the wild type structure, the effect of mutations was observed. Constant pH MD simulations for both the wild type and the

mutated structures indicated how the pKa is affected by the mutated residues near the catalytic center of the HNH domain. Overall, MD simulations provided information at the molecular level for residues of interest allowing analysis of interactions between residues.

Appendix A

Supporting Information for Chapter 2

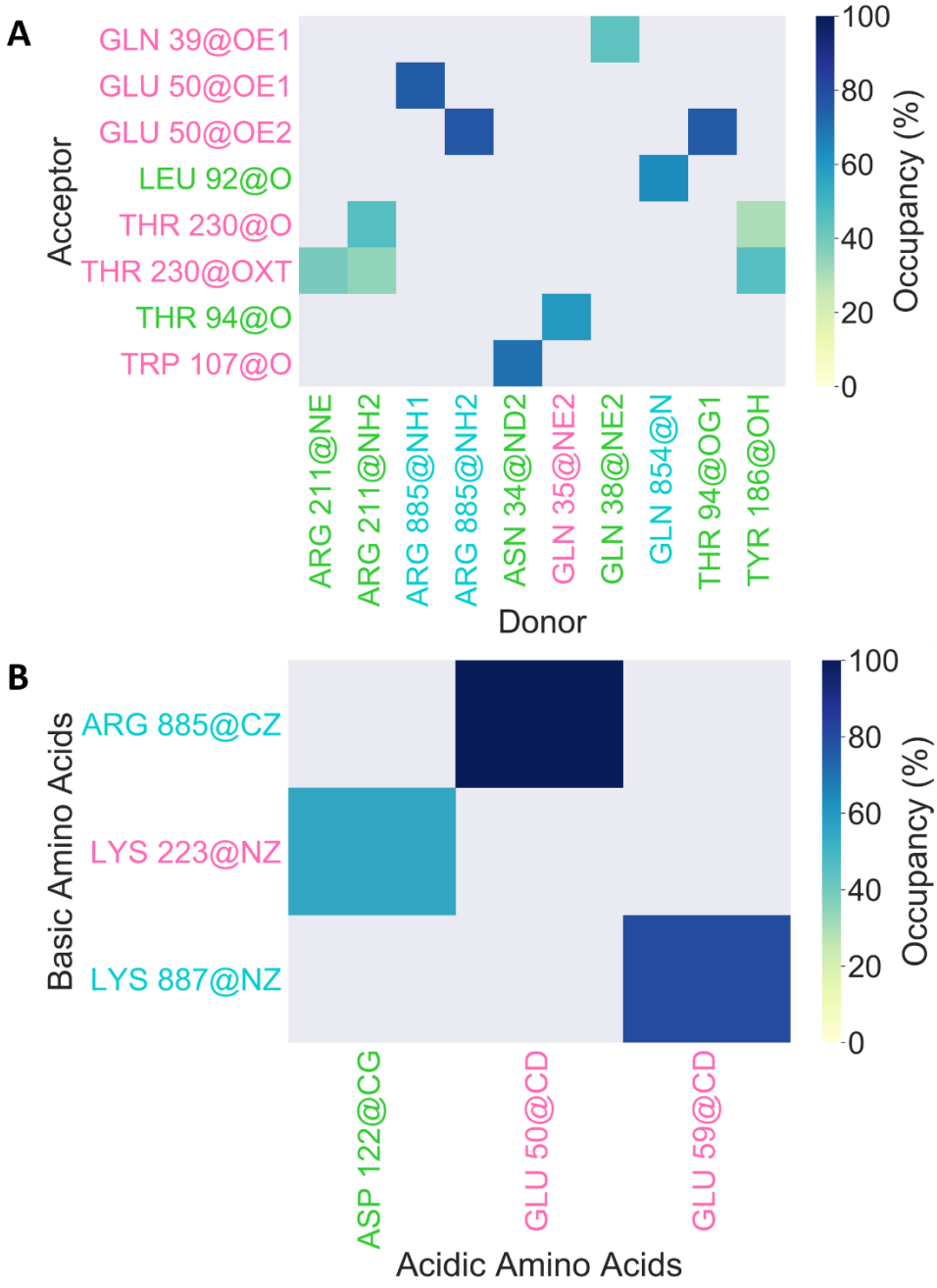

Figure A.1. Occupancy maps for production run 1 for wild type C5 in complex with eculizumab. **(A)** Hydrogen bonds and **(B)** salt bridges that occurred for 30% or more of the simulation trajectory.

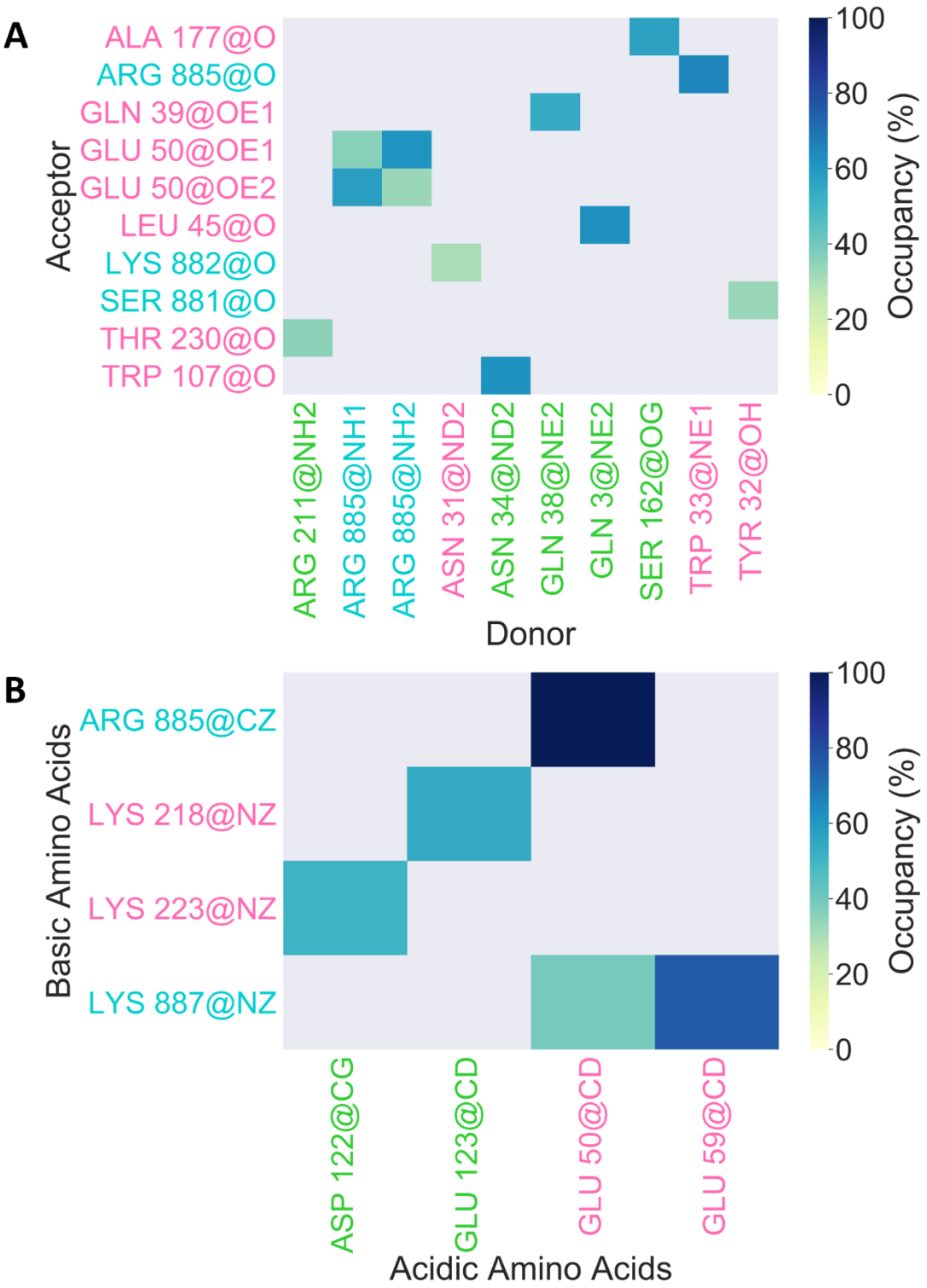

Figure A.2. Occupancy maps for production run 2 for wild type C5 in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

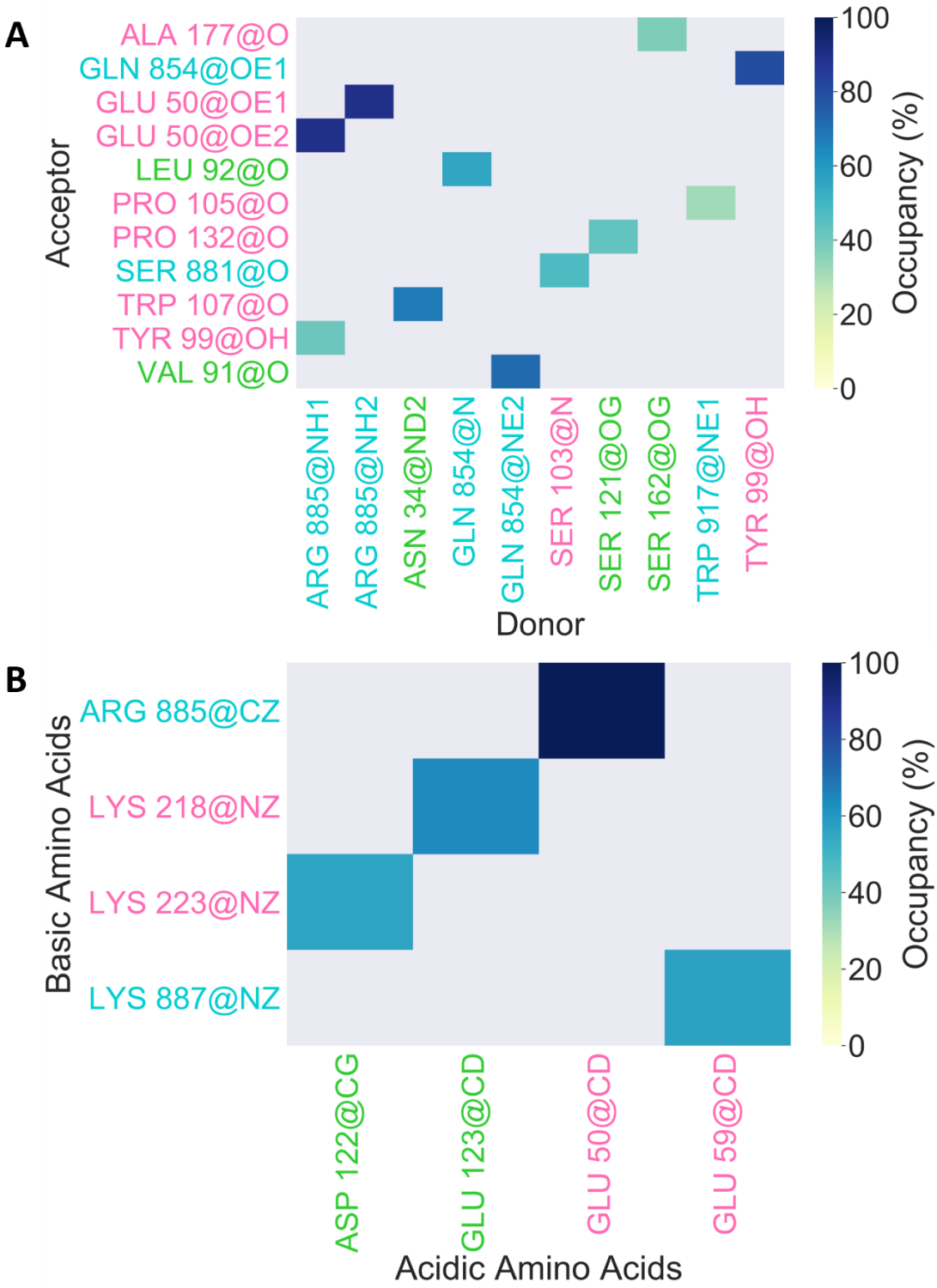

Figure A.3. Occupancy maps for production run 3 for wild type C5 in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

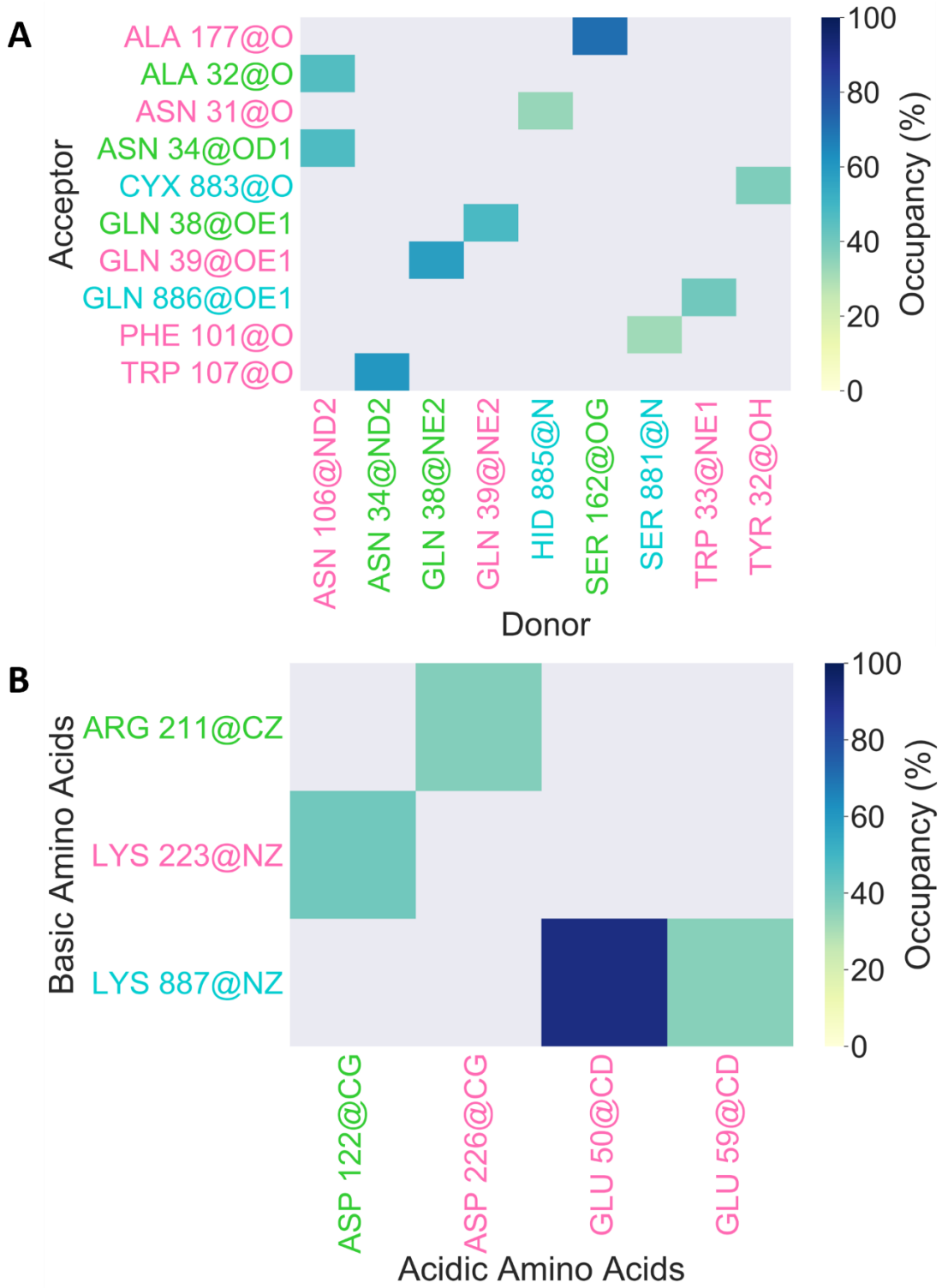

Figure A.4. Occupancy maps for production run 1 for C5 mutated with histidine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

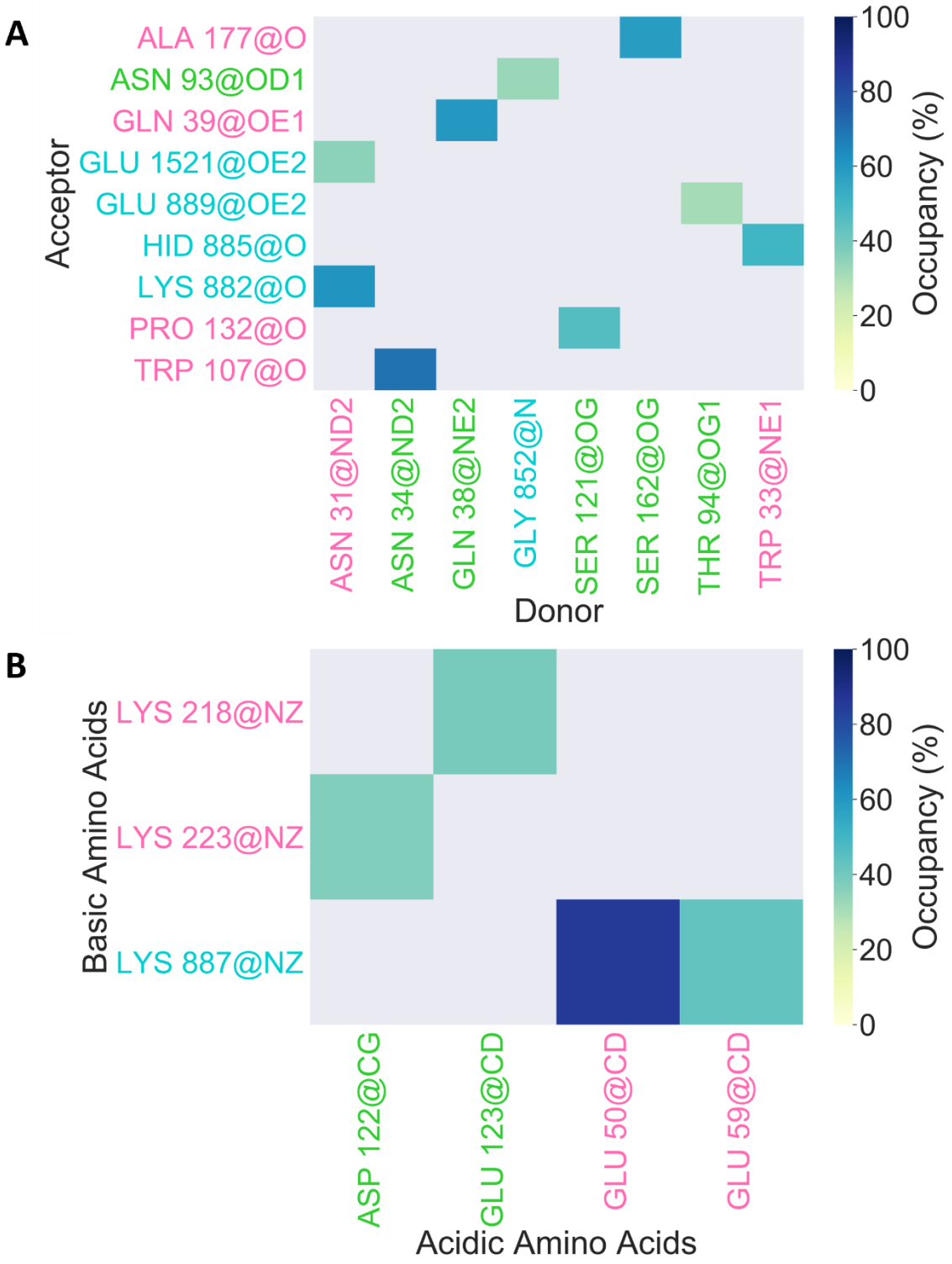

Figure A.5. Occupancy maps for production run 2 for C5 mutated with histidine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

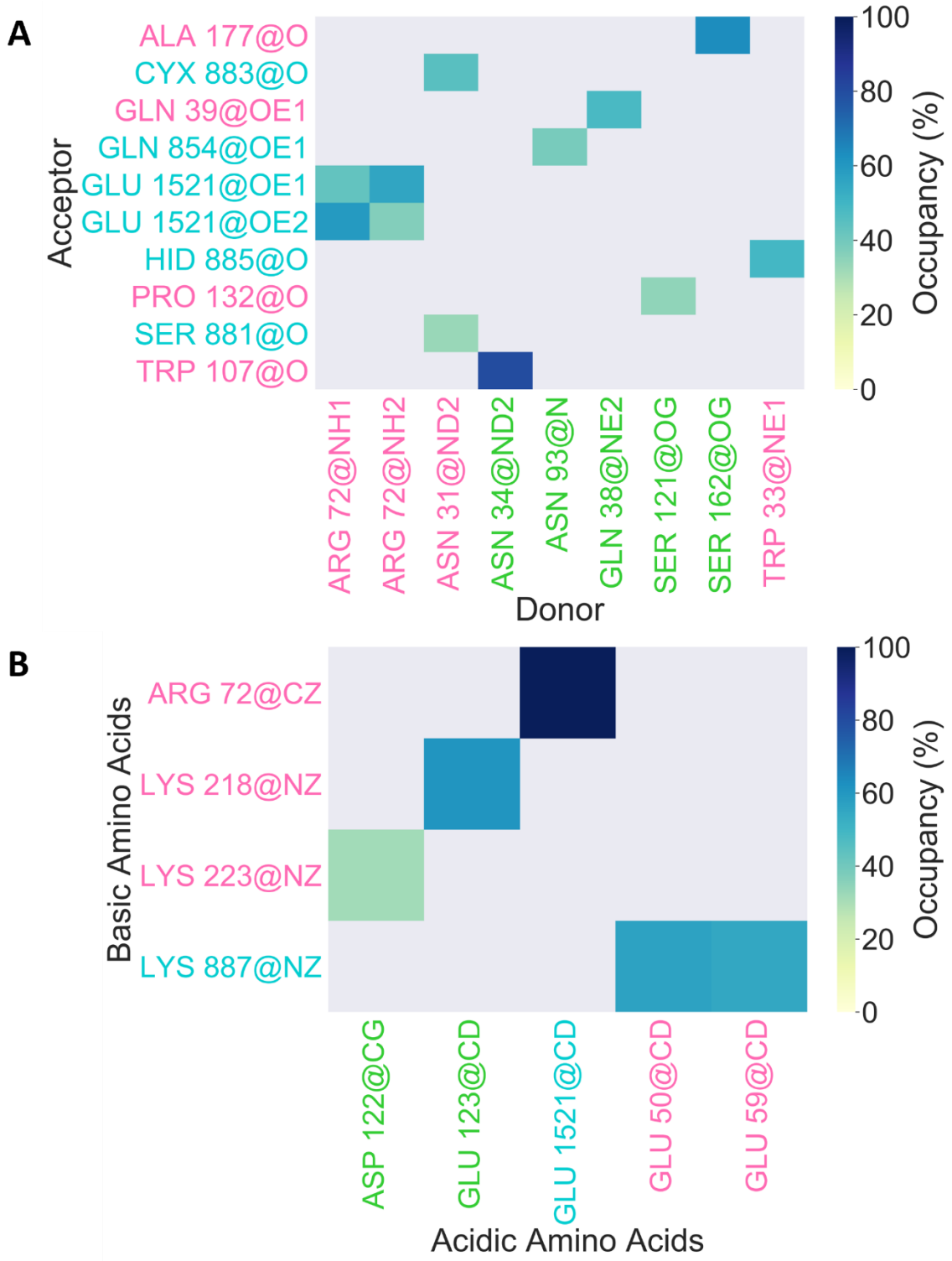

Figure A.6. Occupancy maps for production run 3 for C5 mutated with histidine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

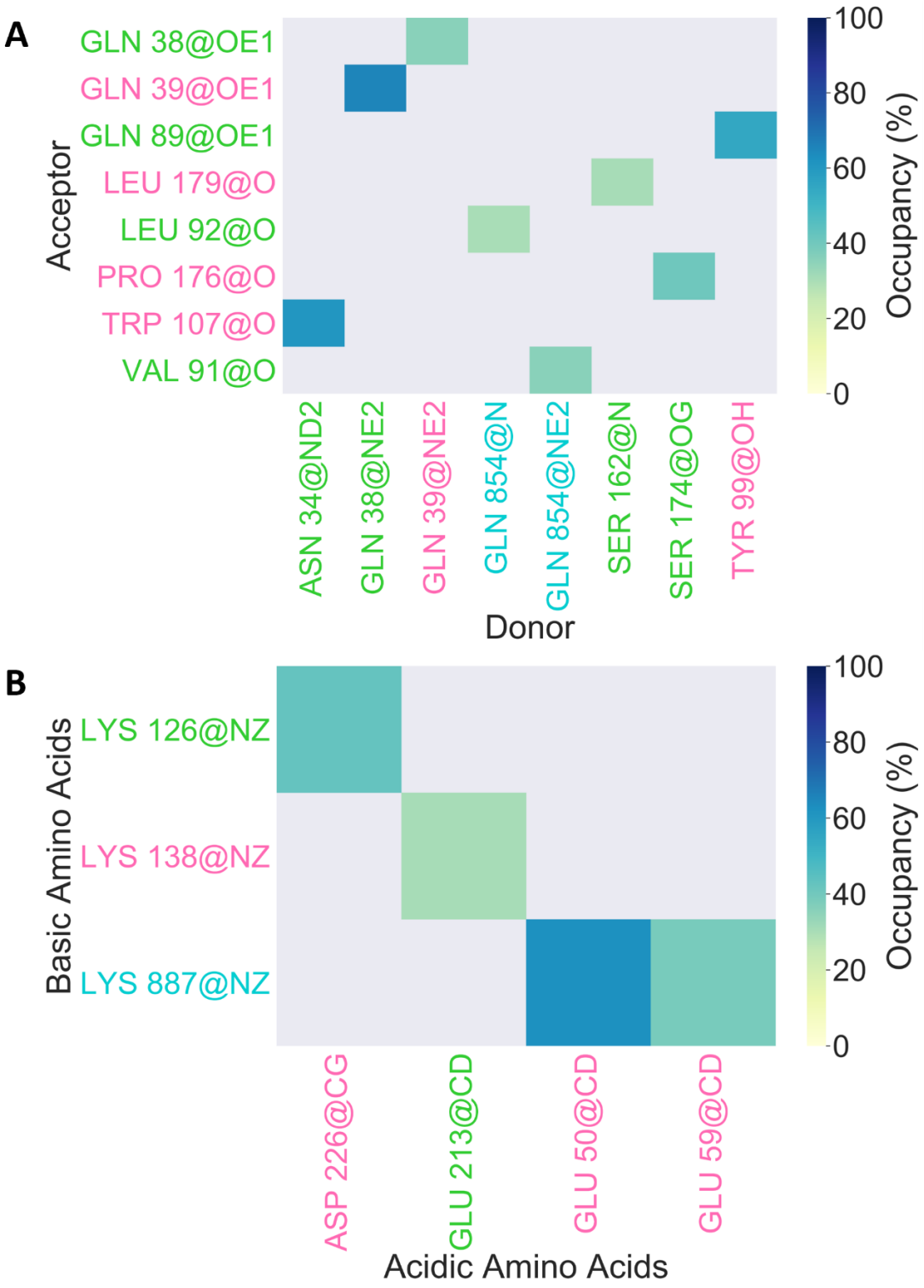

Figure A.7. Occupancy maps for production run 1 for C5 mutated with cysteine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

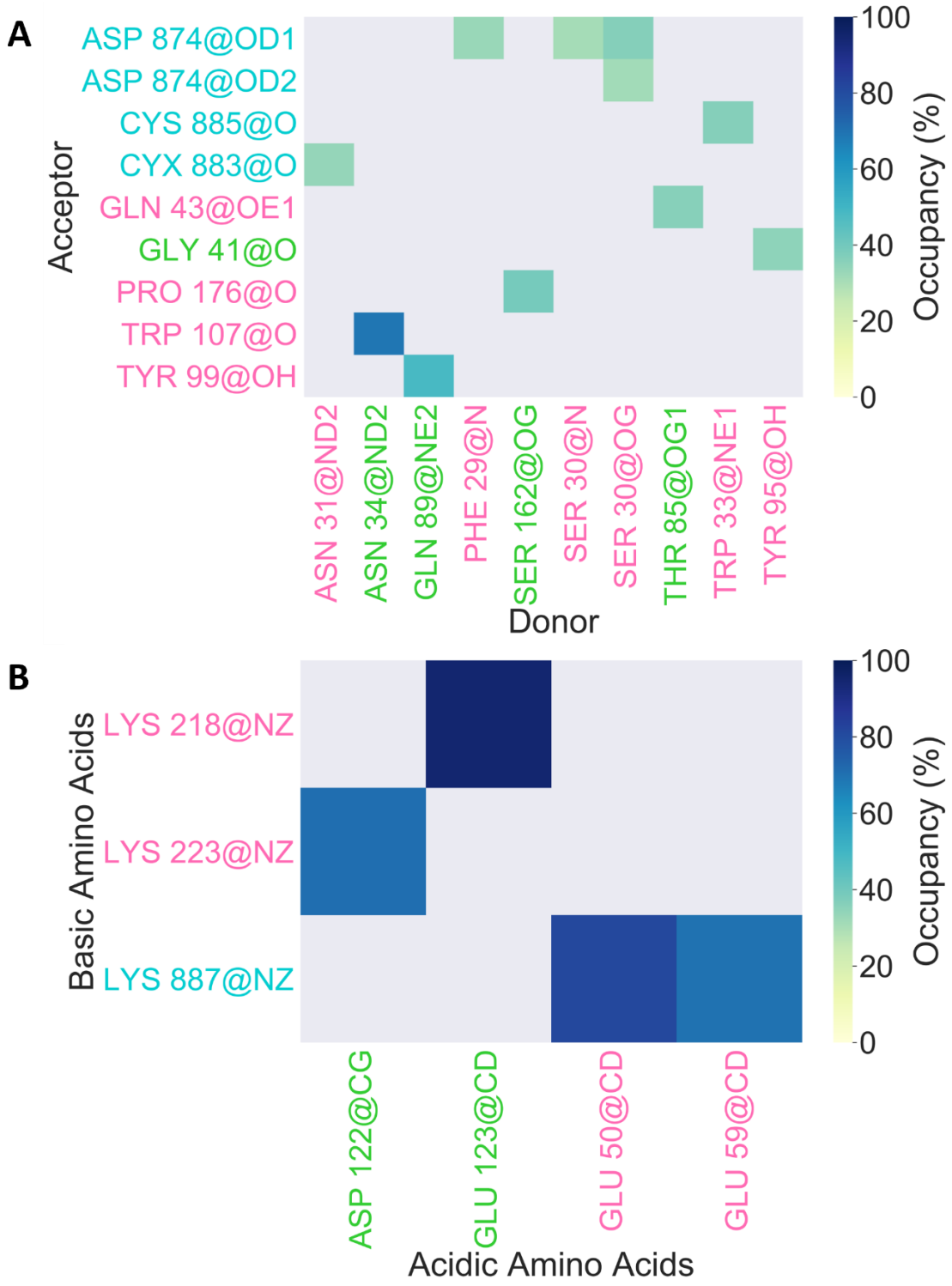

Figure A.8. Occupancy maps for production run 2 for C5 mutated with cysteine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

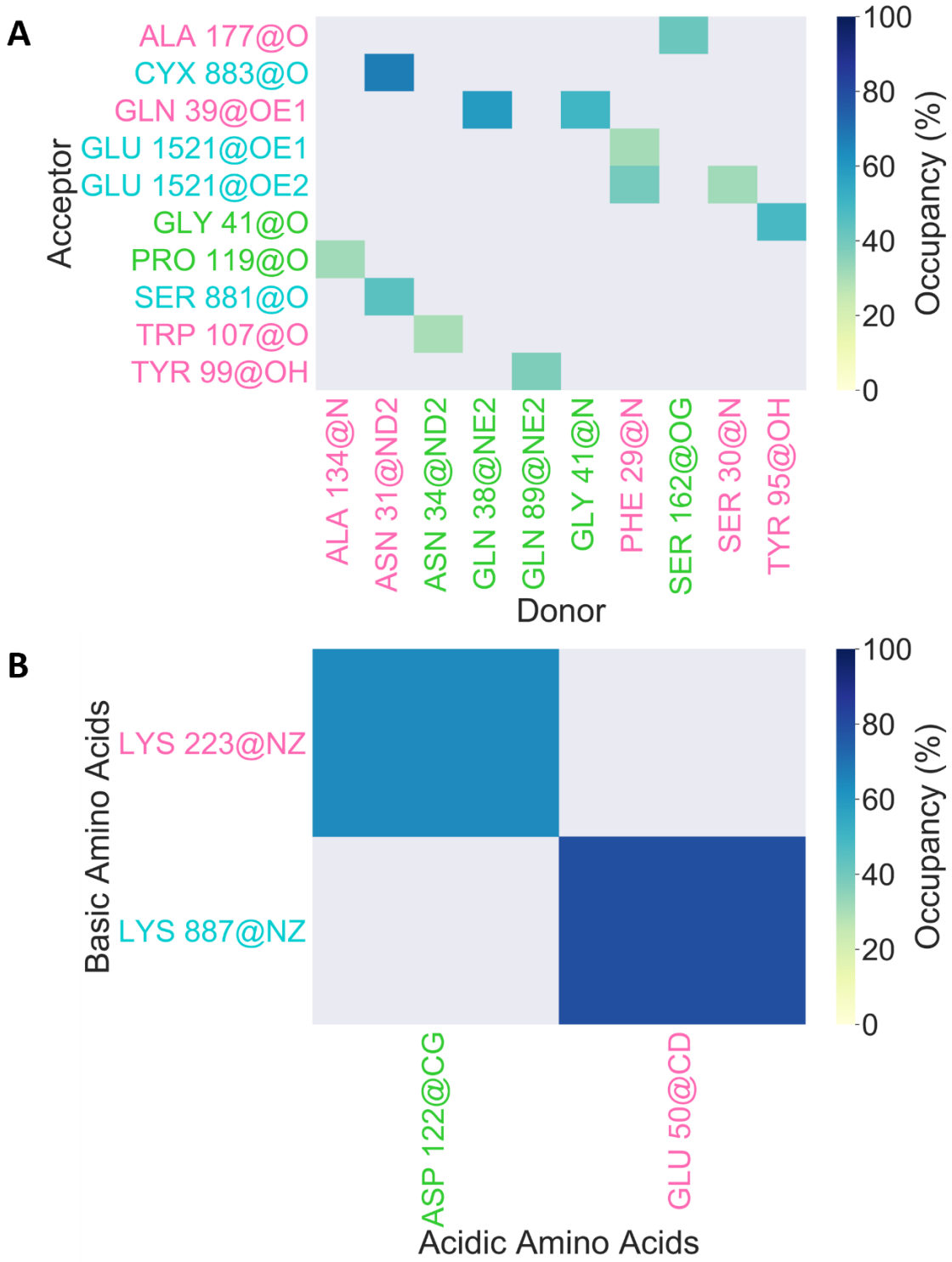

Figure A.9. Occupancy maps for production run 3 for C5 mutated with cysteine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

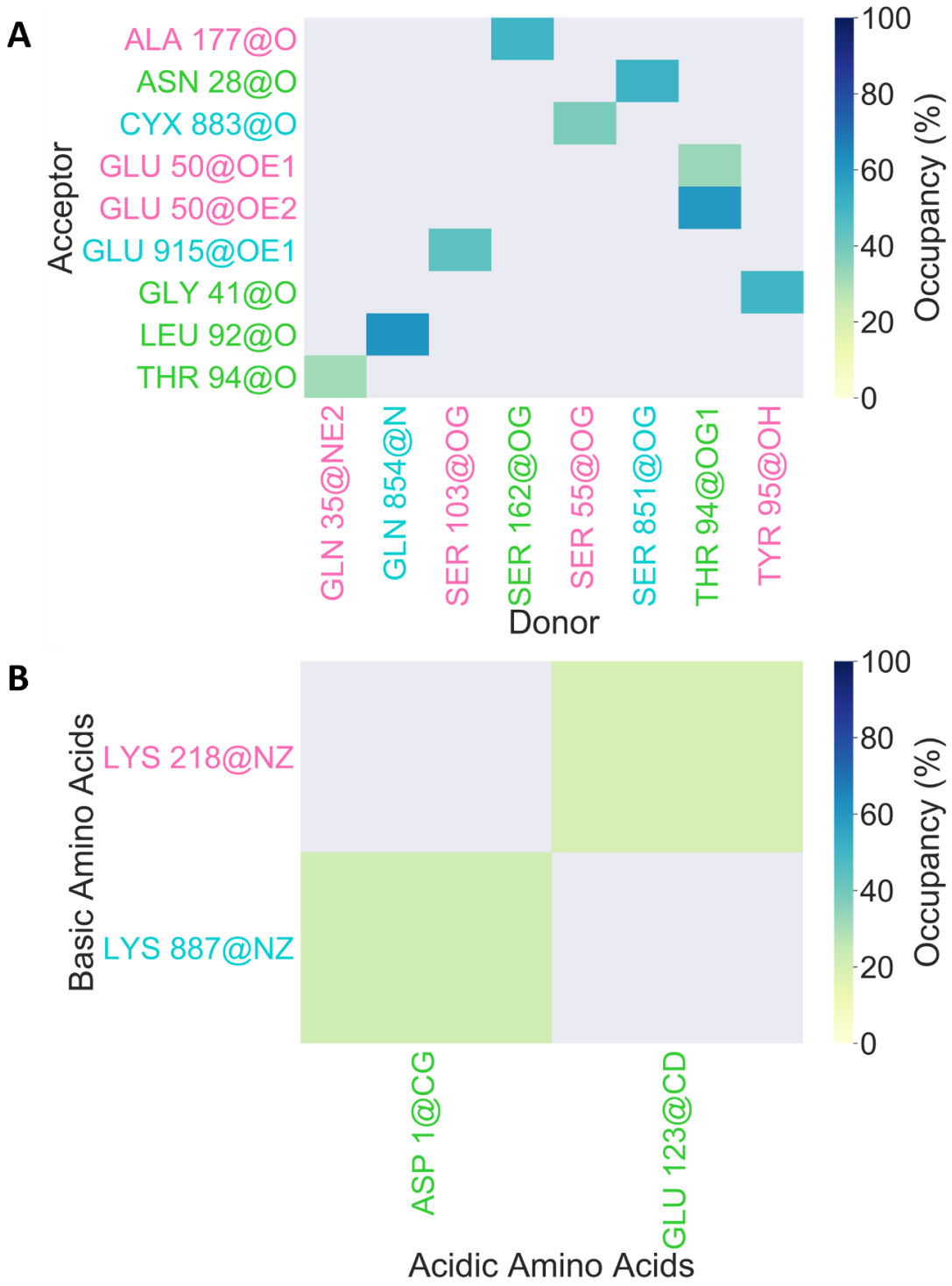

Figure A.10. Occupancy maps for production run 1 for C5 mutated with serine in complex with eculizumab. **(A)** Hydrogen bonds that occurred for 30% or more of the simulation trajectory. **(B)** Salt bridges bonds that occurred for 20% or more of the simulation trajectory.

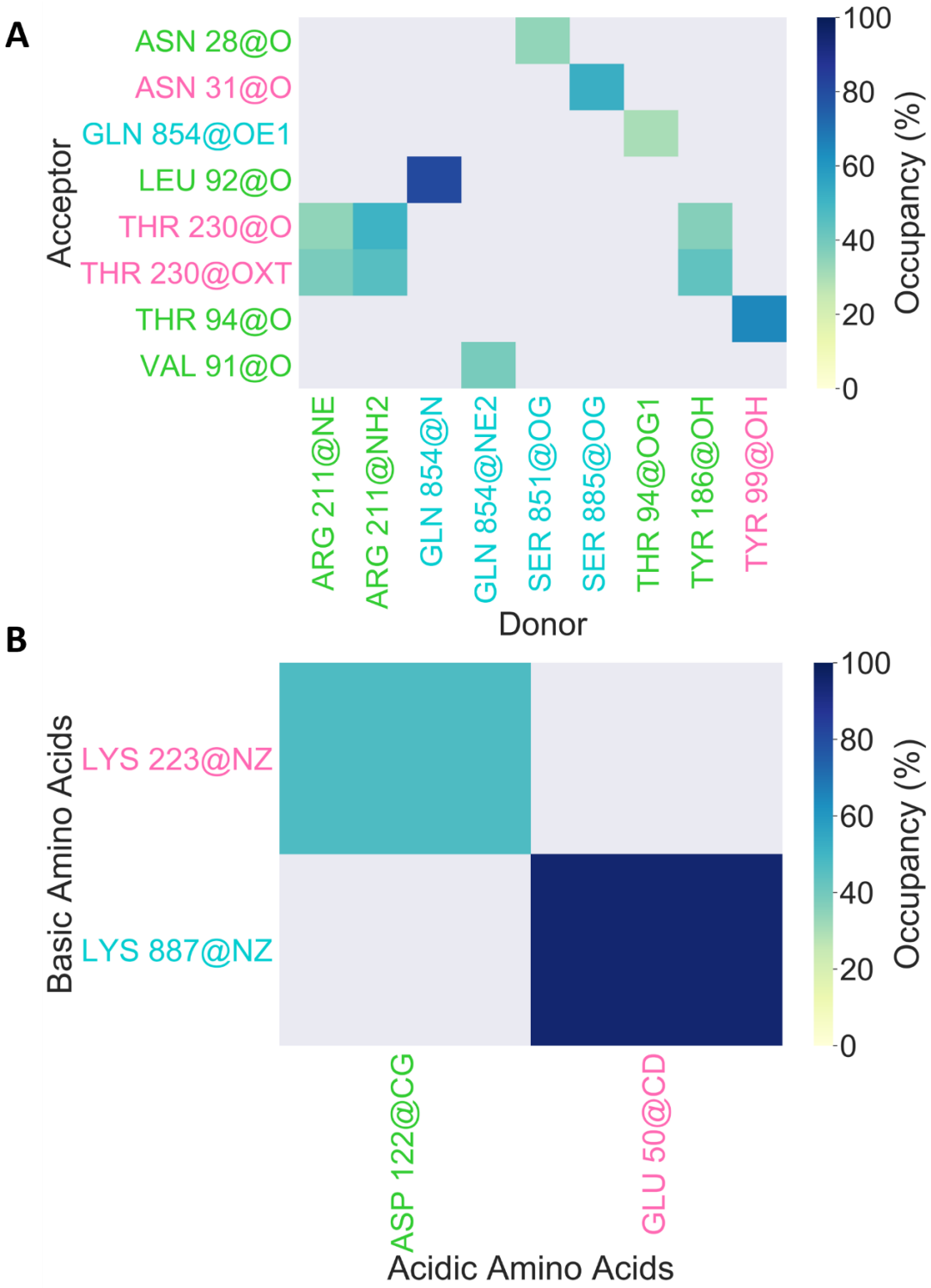

Figure A.11. Occupancy maps for production run 2 for C5 mutated with serine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

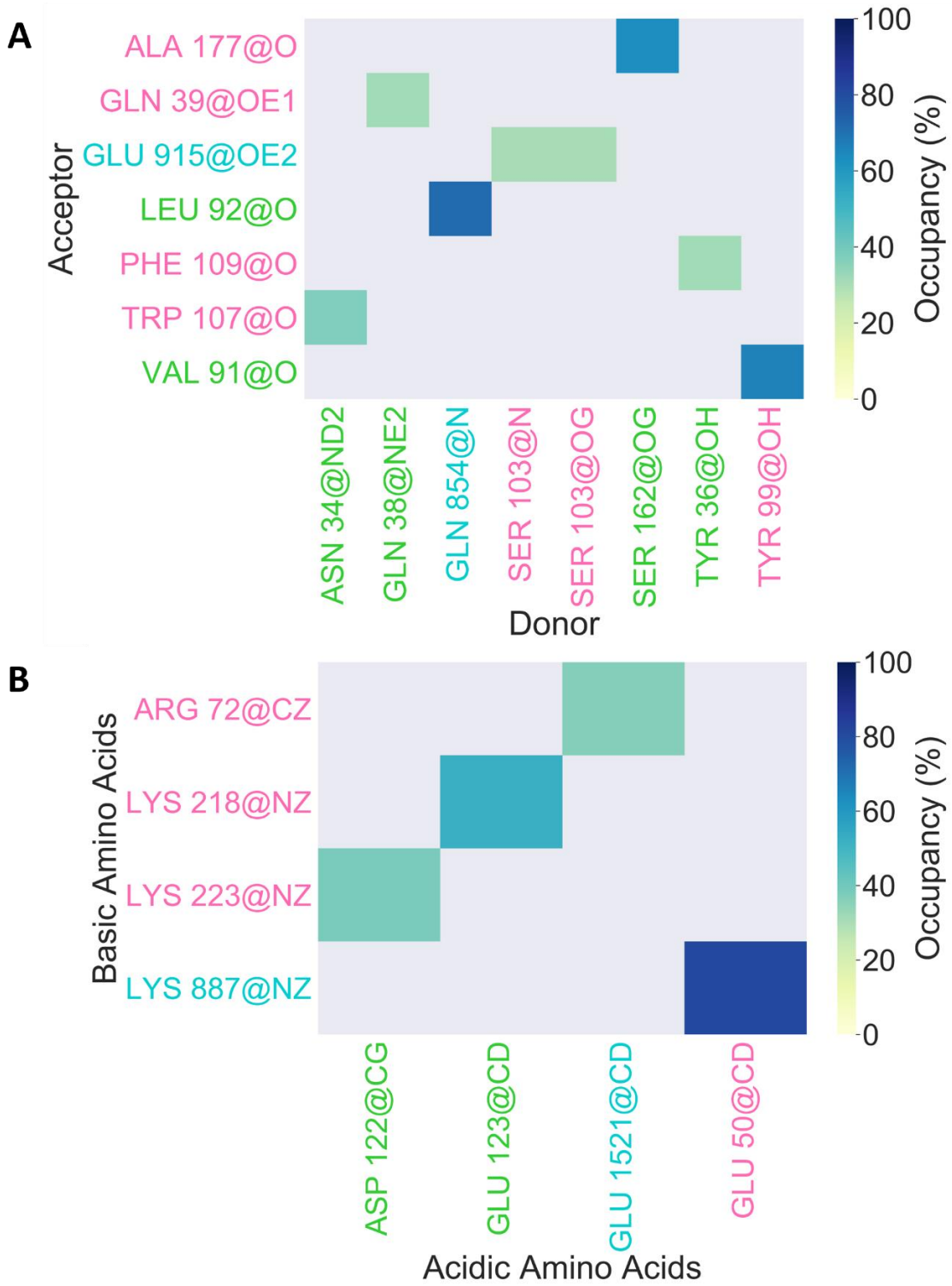

Figure A.12. Occupancy maps for production run 3 for C5 mutated with serine in complex with eculizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

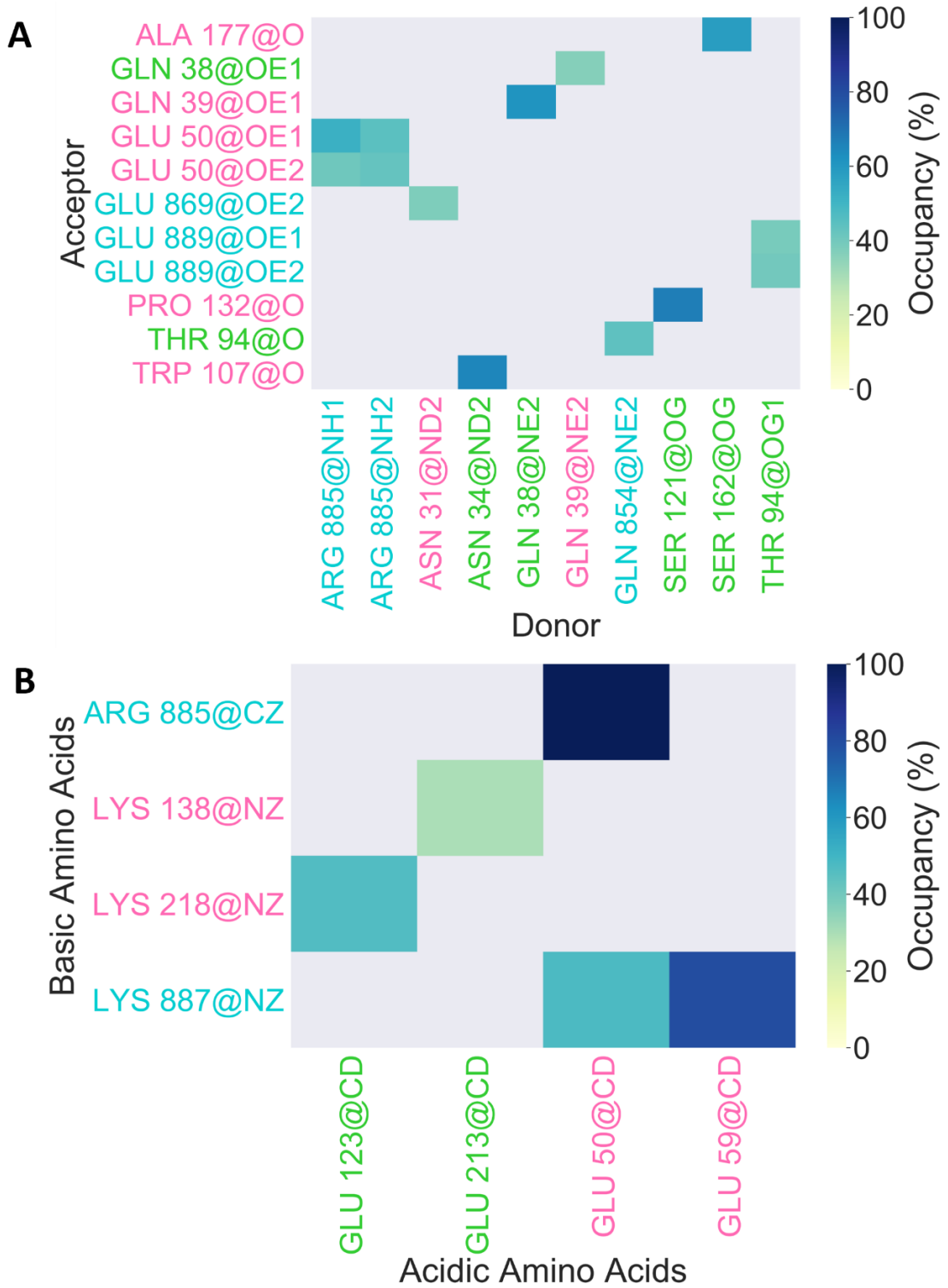

Figure A.13. Occupancy maps for production run 1 for wild type C5 in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

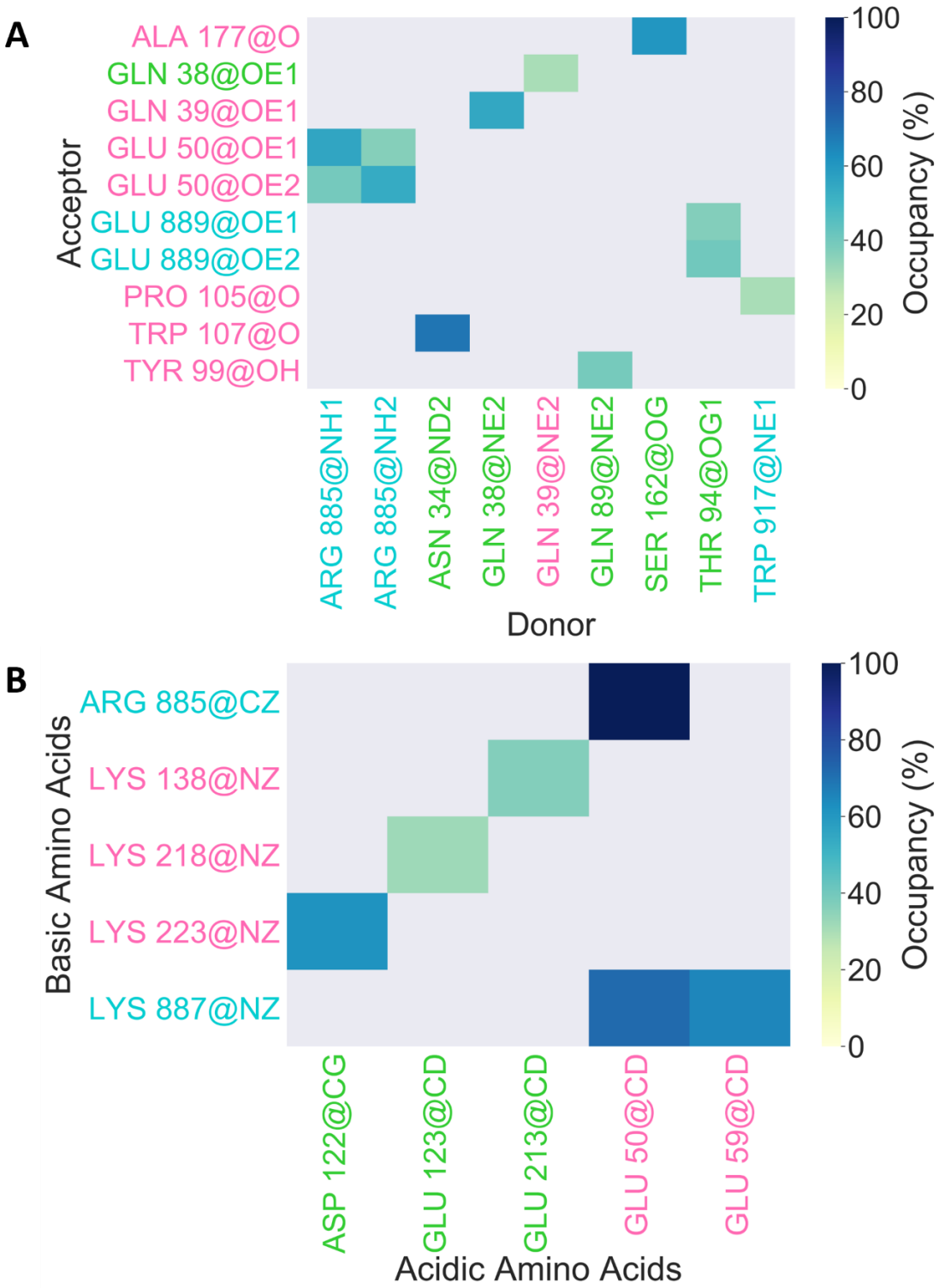

Figure A.14. Occupancy maps for production run 2 for wild type C5 in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

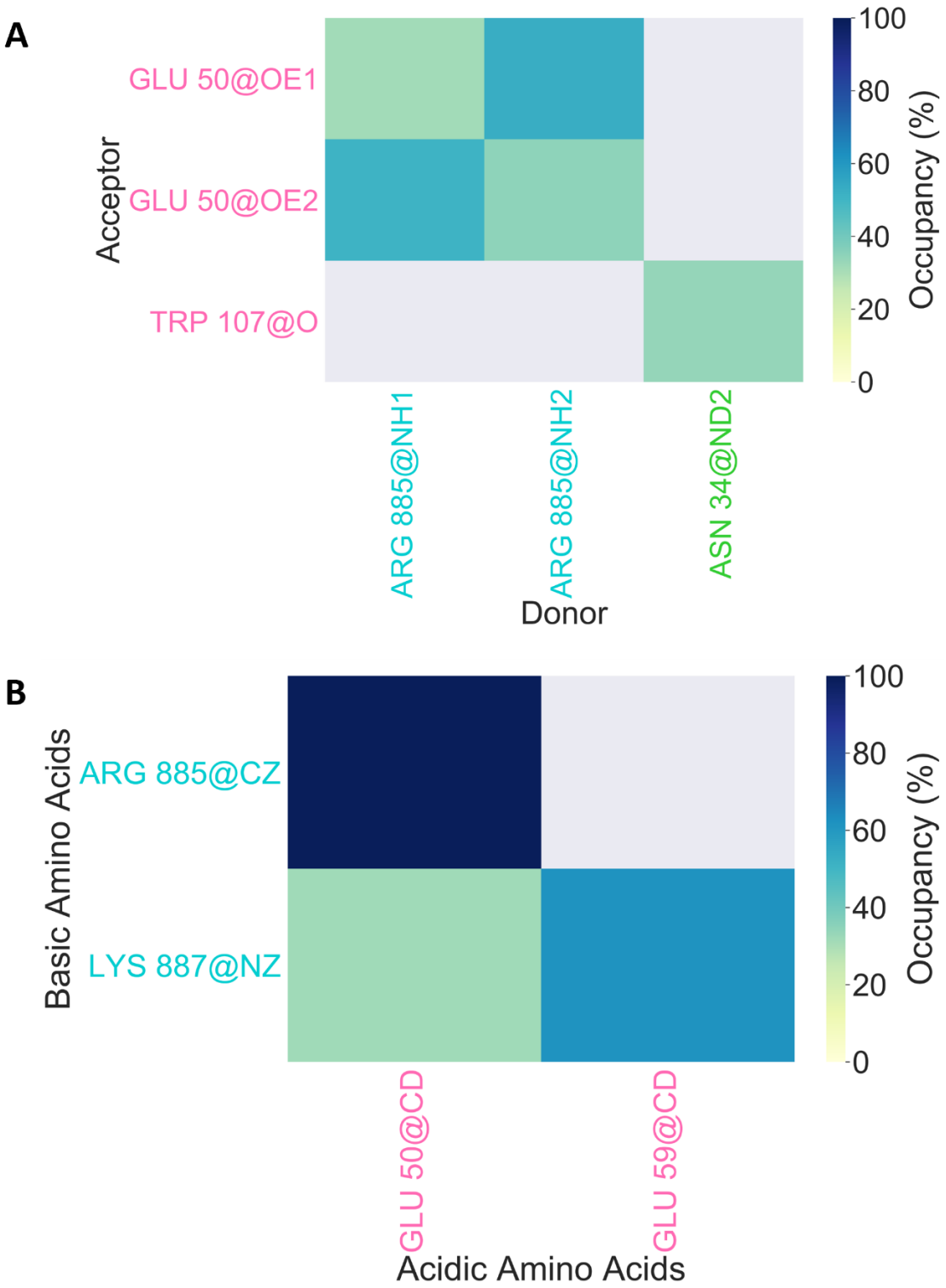

Figure A.15. Occupancy maps for production run 3 for wild type C5 in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

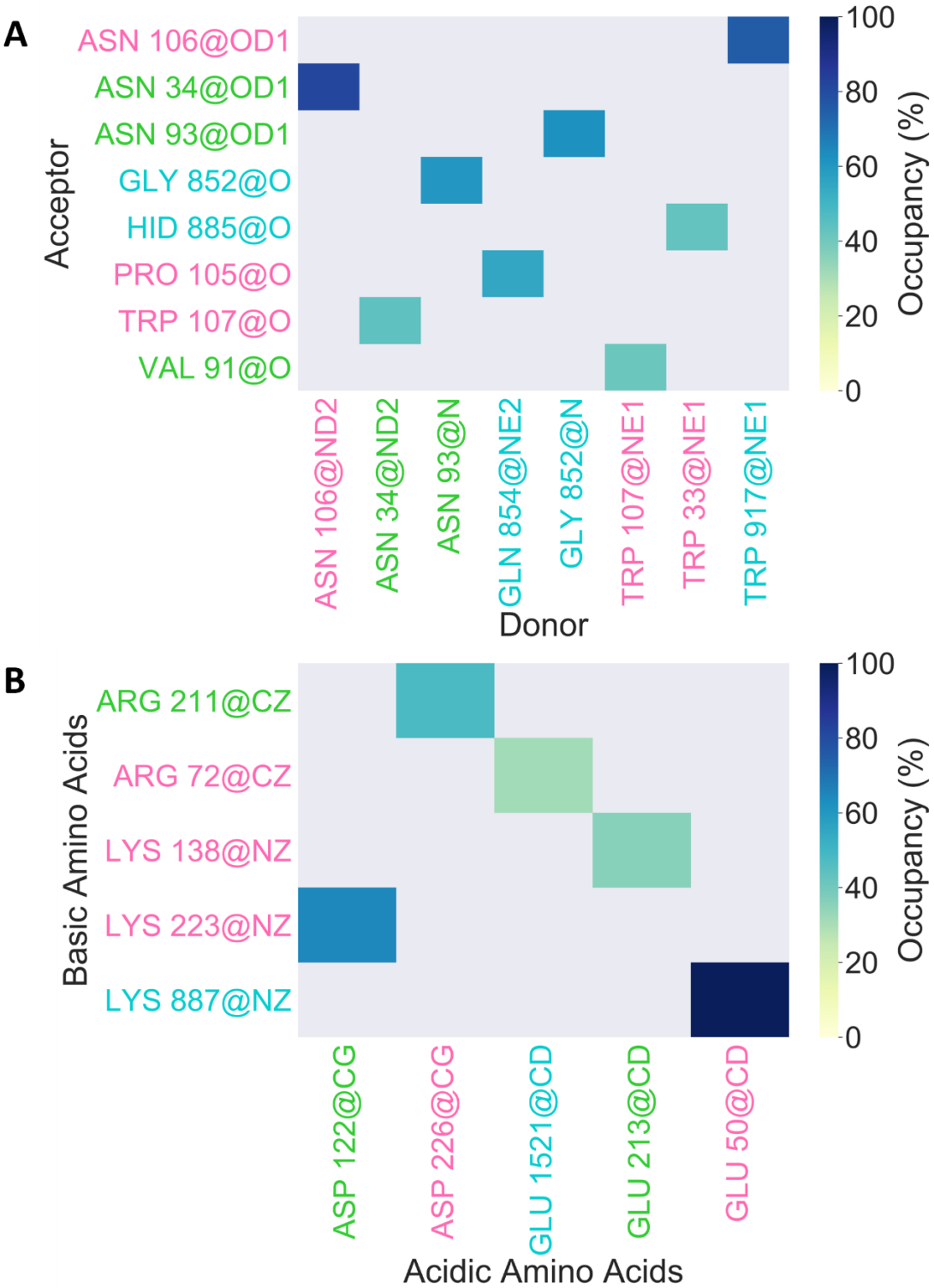

Figure A.16. Occupancy maps for production run 1 for C5 mutated with histidine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

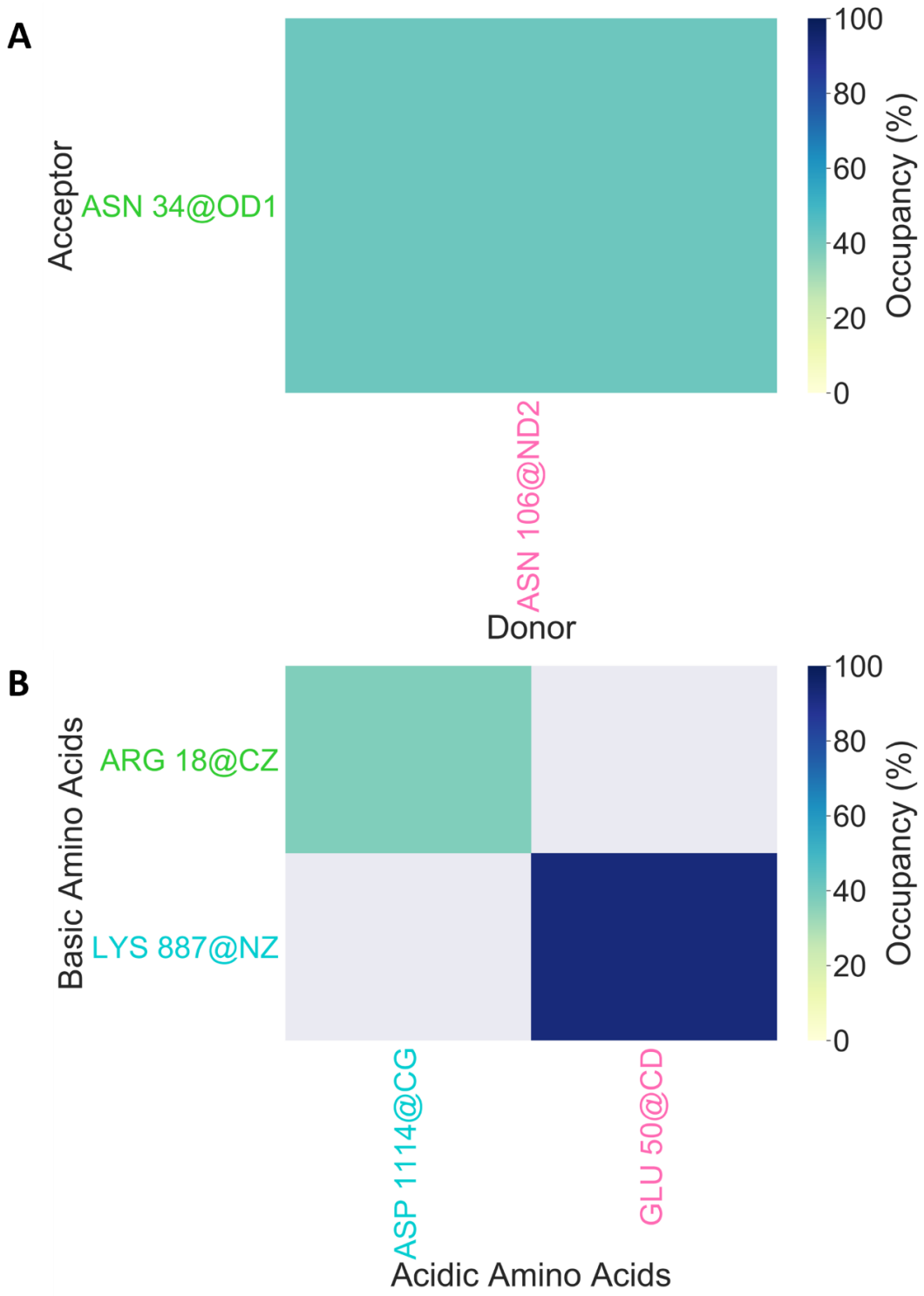

Figure A.17. Occupancy maps for production run 2 for C5 mutated with histidine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

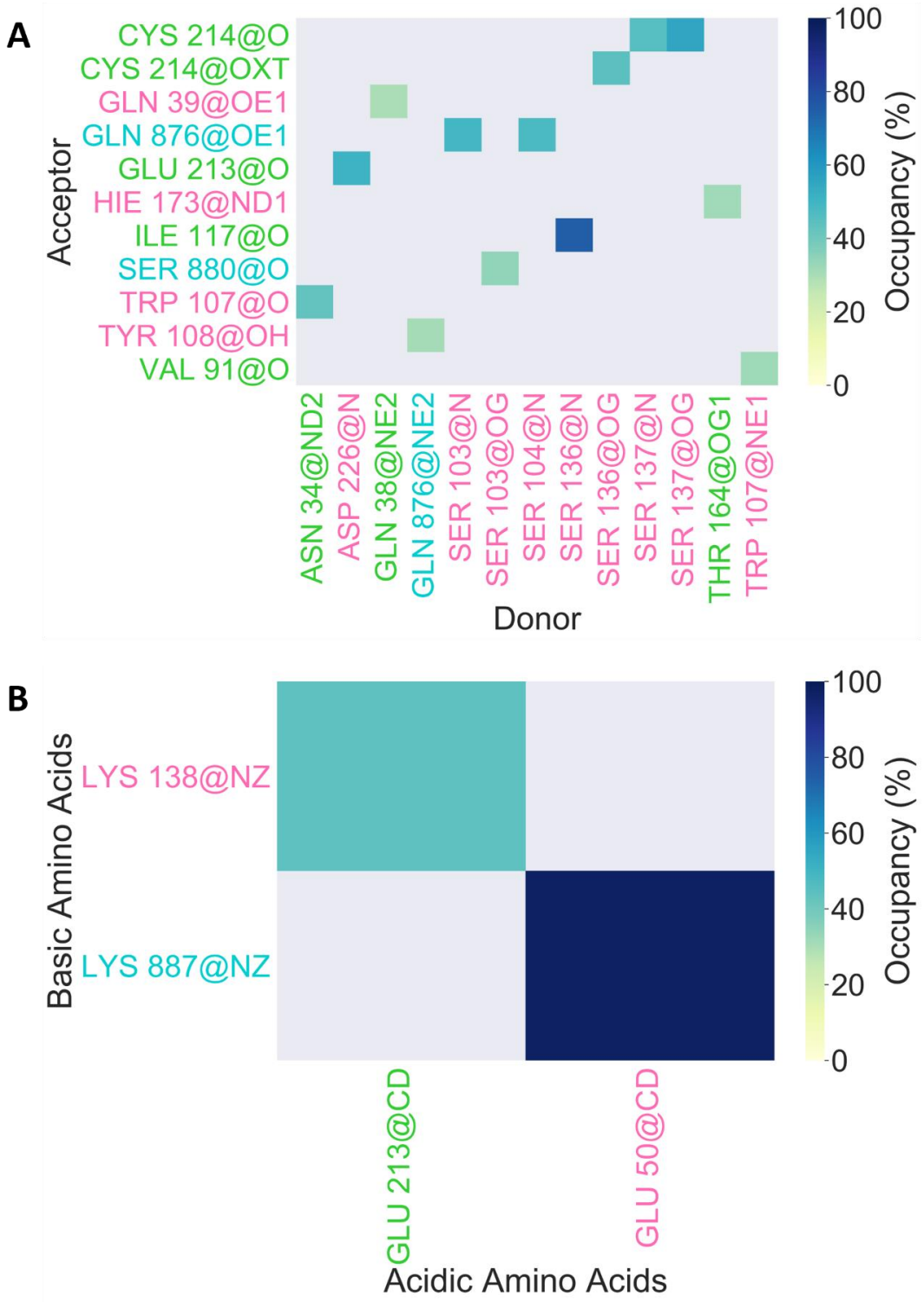

Figure A.18. Occupancy maps for production run 3 for C5 mutated with histidine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

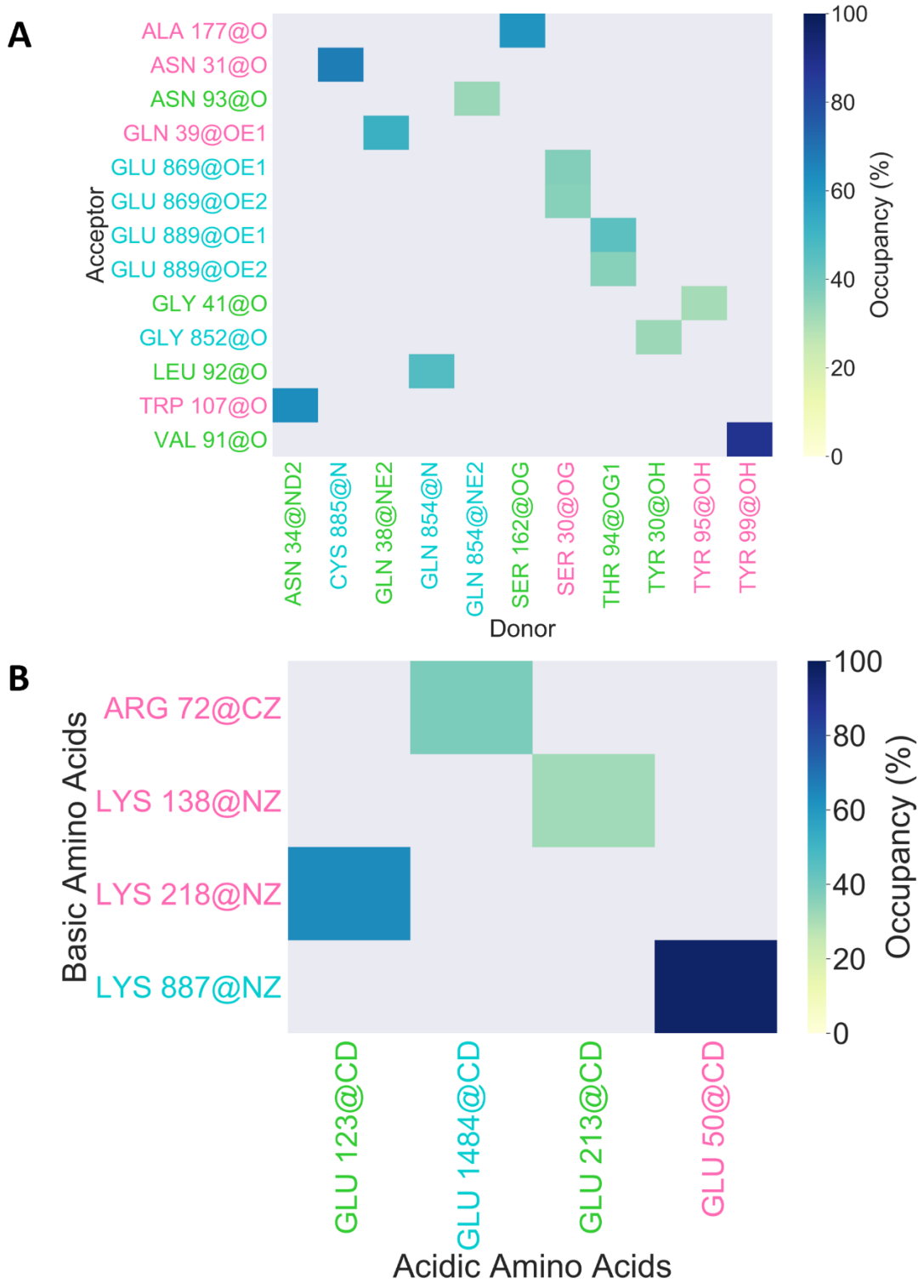

Figure A.19. Occupancy maps for production run 1 for C5 mutated with cysteine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

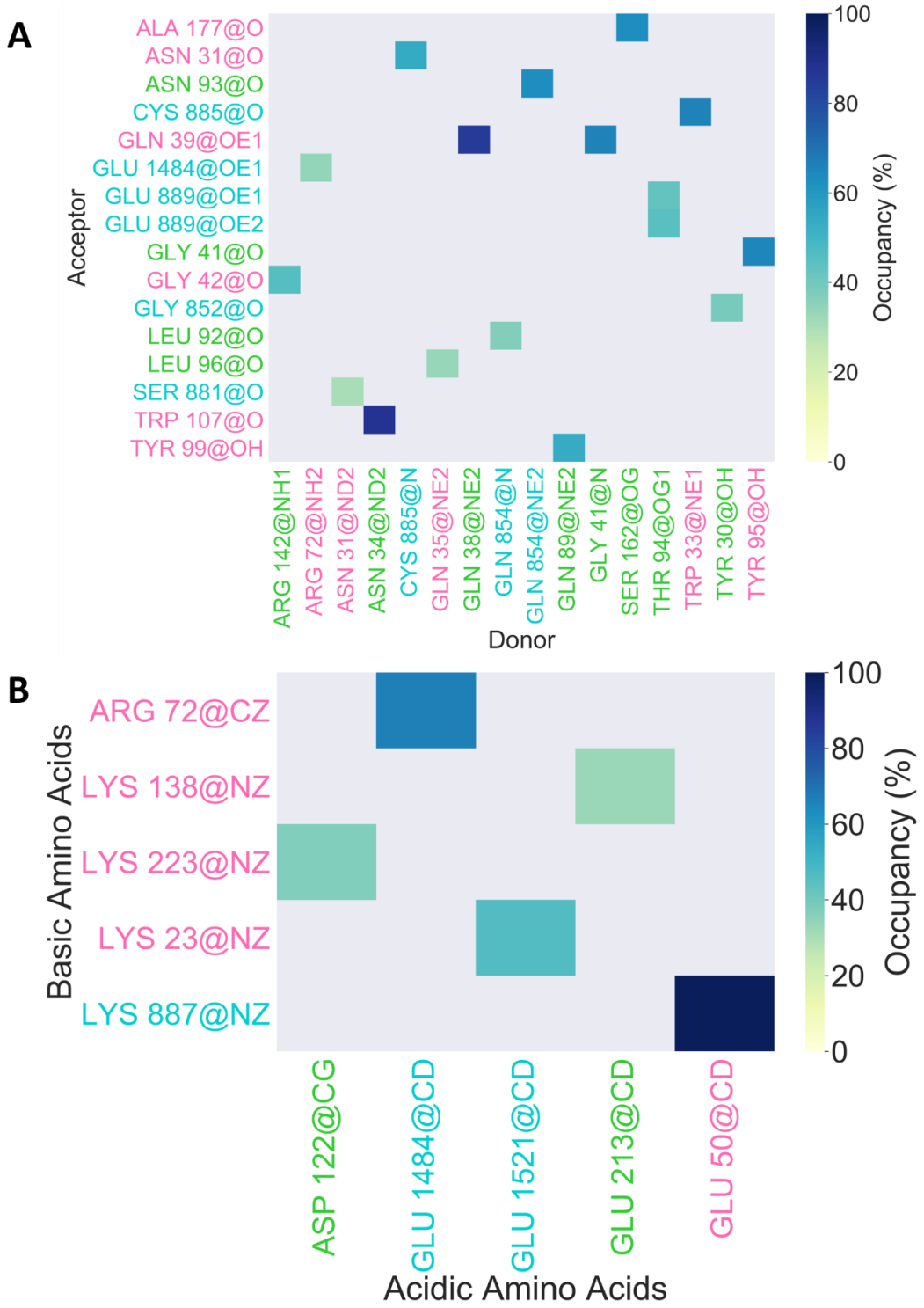

Figure A.20. Occupancy maps for production run 2 for C5 mutated with cysteine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

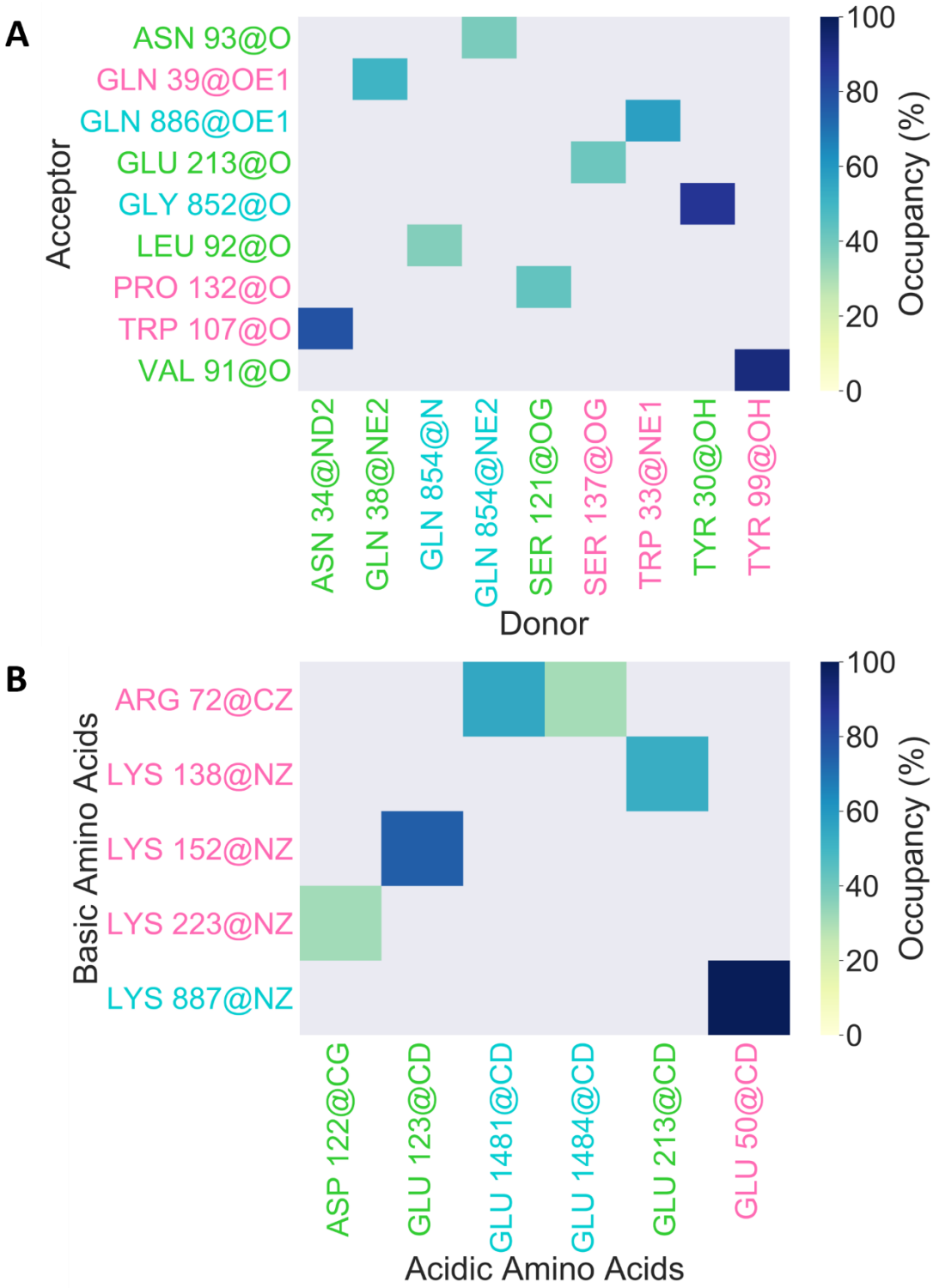

Figure A.21. Occupancy maps for production run 3 for C5 mutated with cysteine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

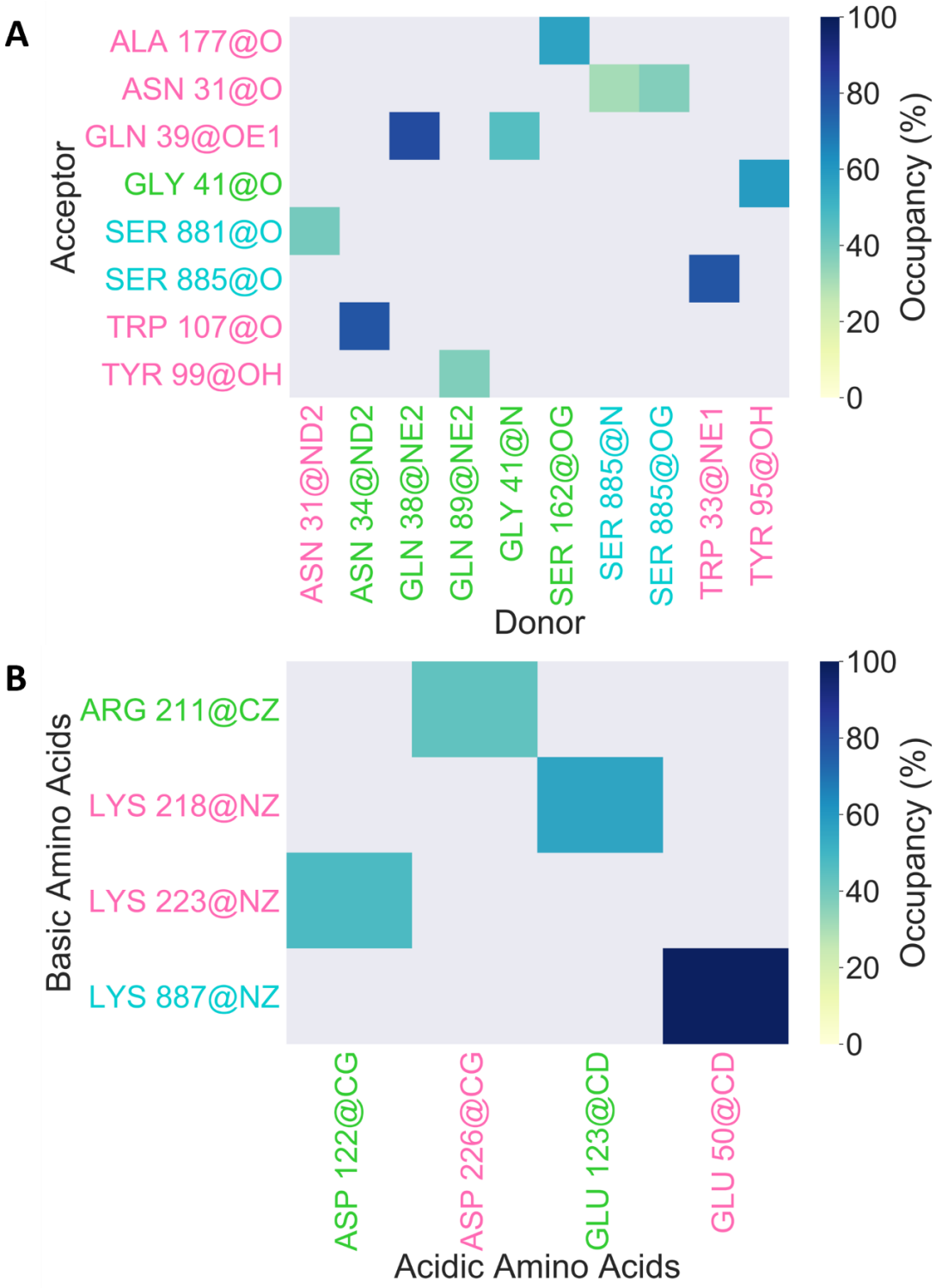

Figure A.22. Occupancy maps for production run 1 for C5 mutated with serine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

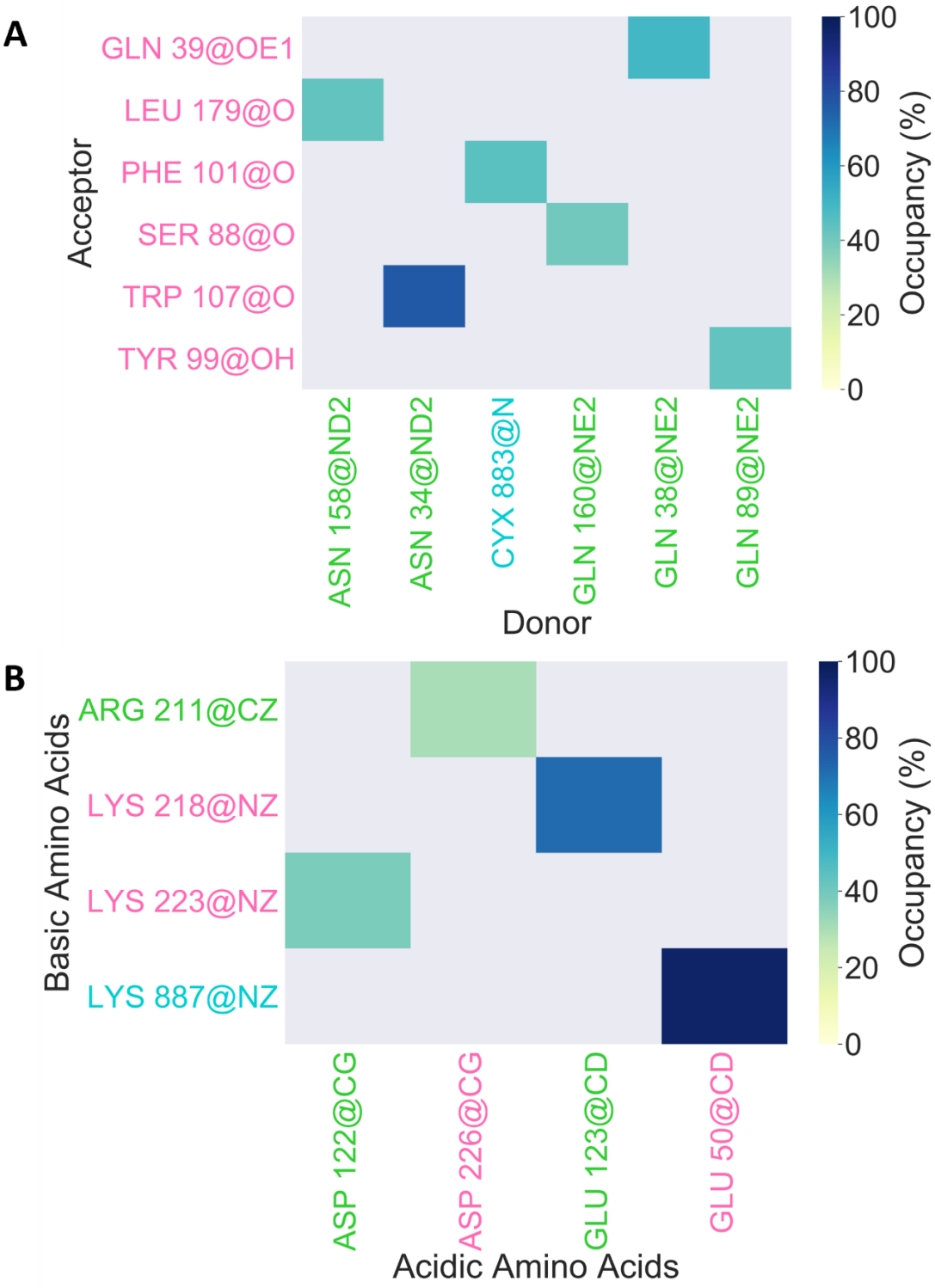

Figure A.23. Occupancy maps for production run 2 for C5 mutated with serine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

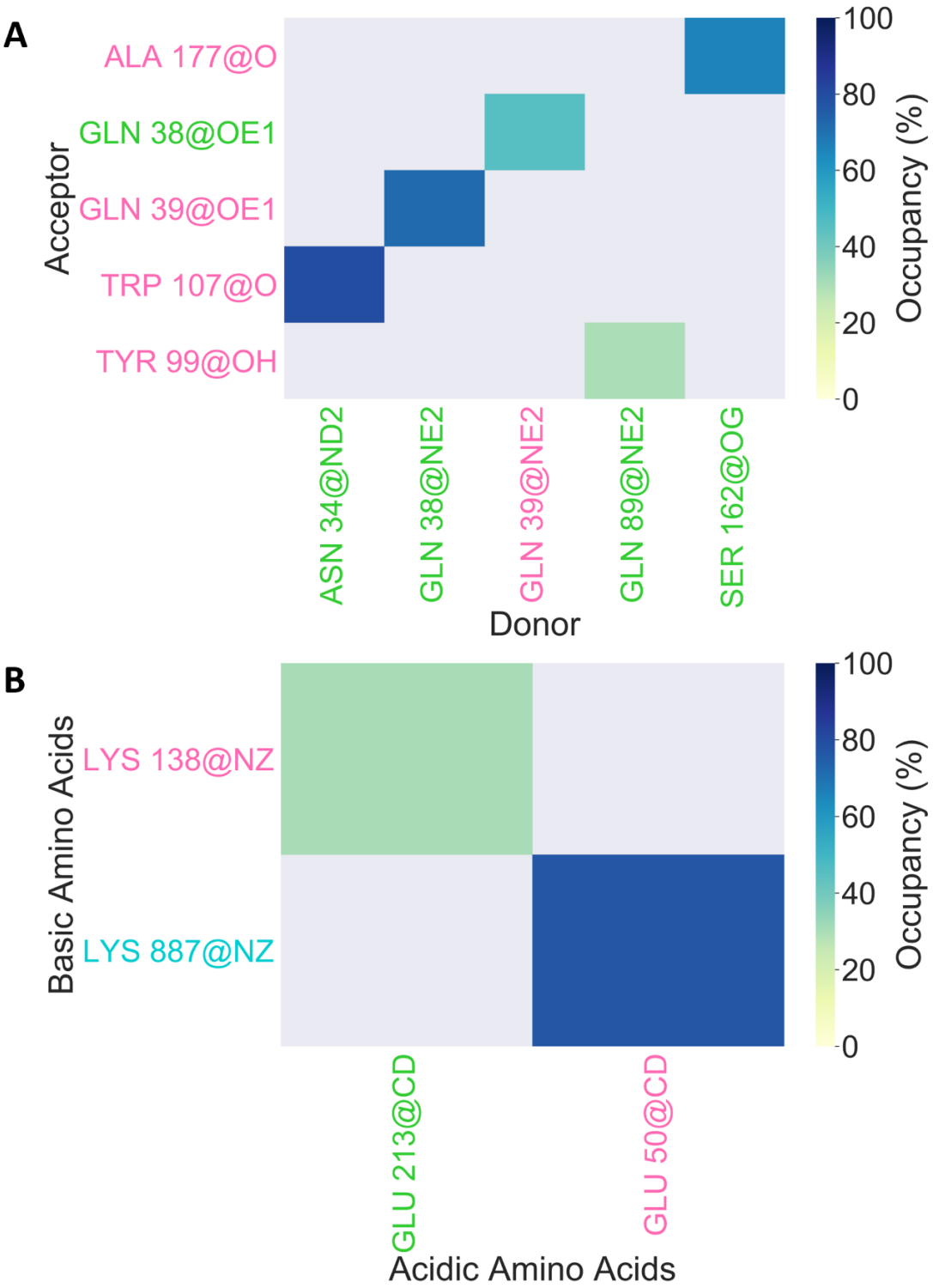

Figure A.24. Occupancy maps for production run 3 for C5 mutated with serine in complex with ravulizumab. (A) Hydrogen bonds and (B) salt bridges that occurred for 30% or more of the simulation trajectory.

For each simulation production run, RMSD was calculated for the C5 component, the heavy chain, and the light chain of eculizumab and ravulizumab. RMSD for each component of interest was measured for the wild type structure and the mutated C5 structures **(Figures A.25-A.30)**.

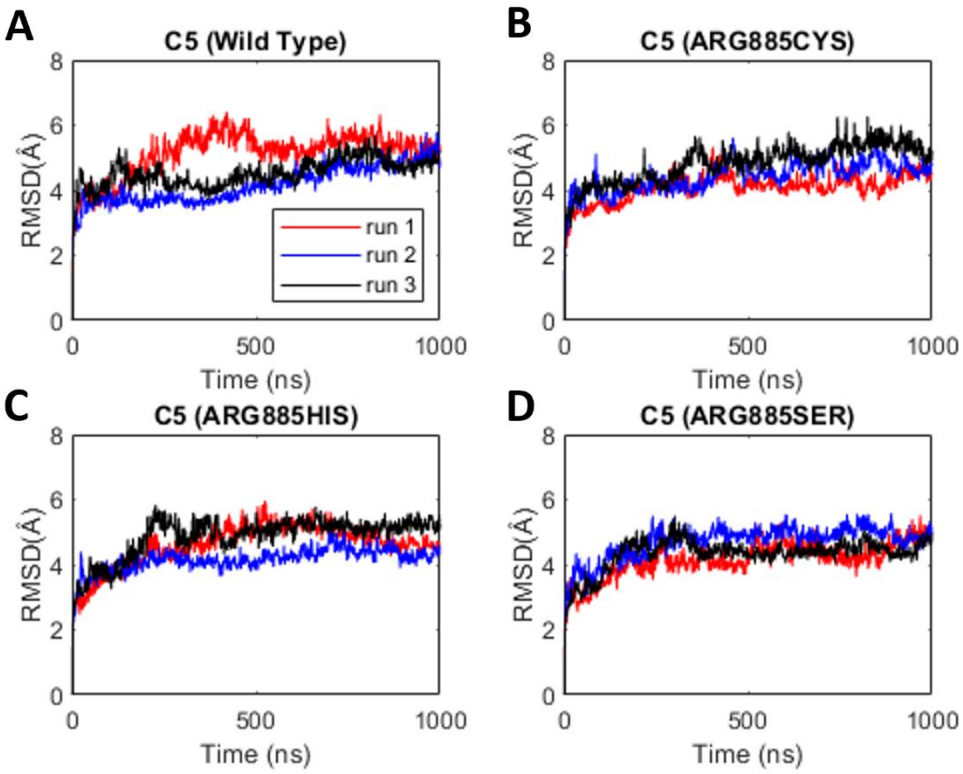

Figure A.25. RMSD of C5 when in complex with eculizumab. **(A)** Wild type C5 in complex with eculizumab. Mutated C5 structures, **(B)** Cysteine 885, **(C)** Histidine 885, and **(D)** Serine 885 of C5 in complex with eculizumab.

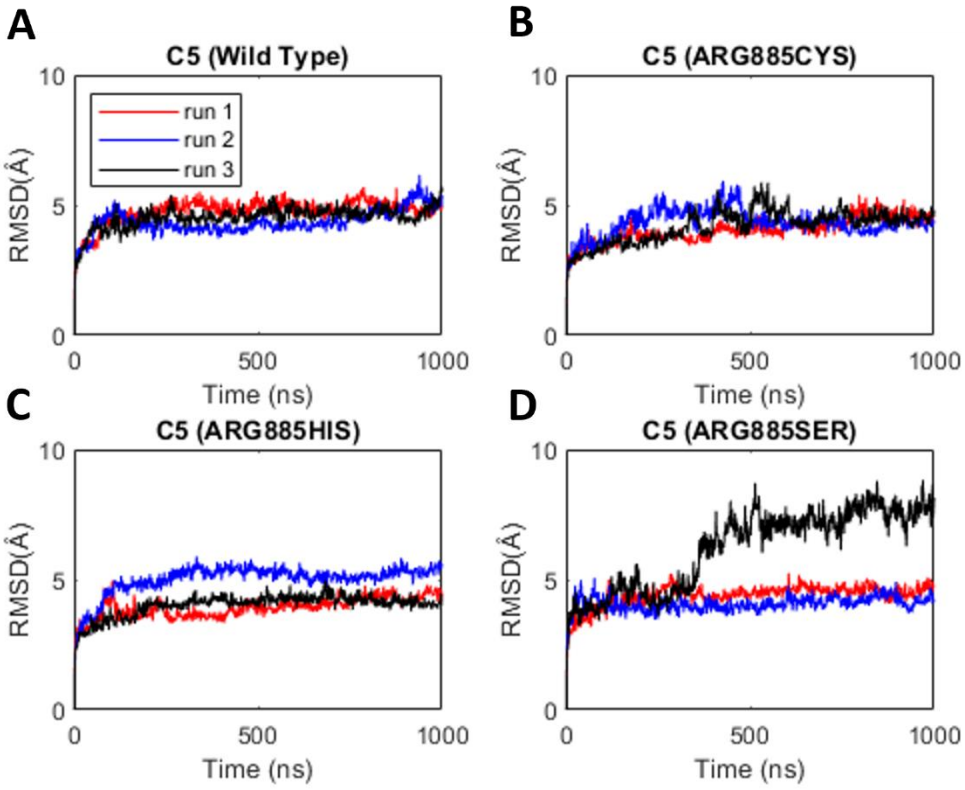

Figure A.26. RMSD of C5 when in complex with ravulizumab. **(A)** Wild type C5 in complex with ravulizumab. Mutated C5 structures, **(B)** Cysteine 885, **(C)** Histidine 885, and **(D)** Serine 885 of C5 in complex with ravulizumab.

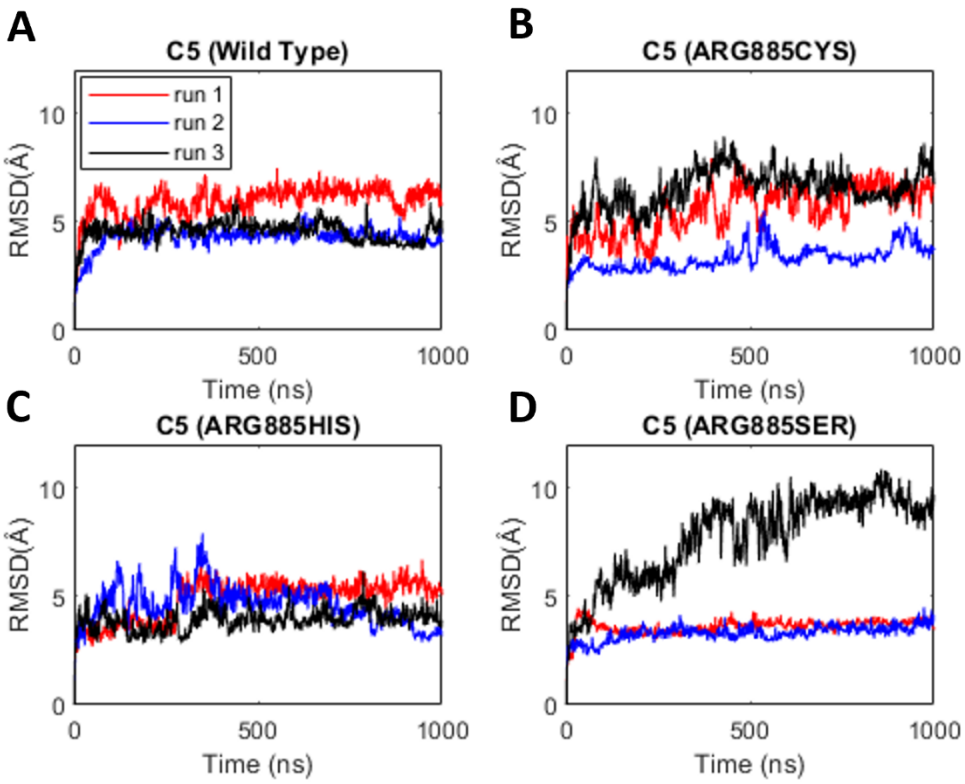

Figure A.27. RMSD of the heavy chain of eculizumab for C5 in complex with eculizumab. **(A)** Wild type C5 and mutated C5 structures, **(B)** Cysteine 885, **(C)** Histidine 885, and **(D)** Serine 885.

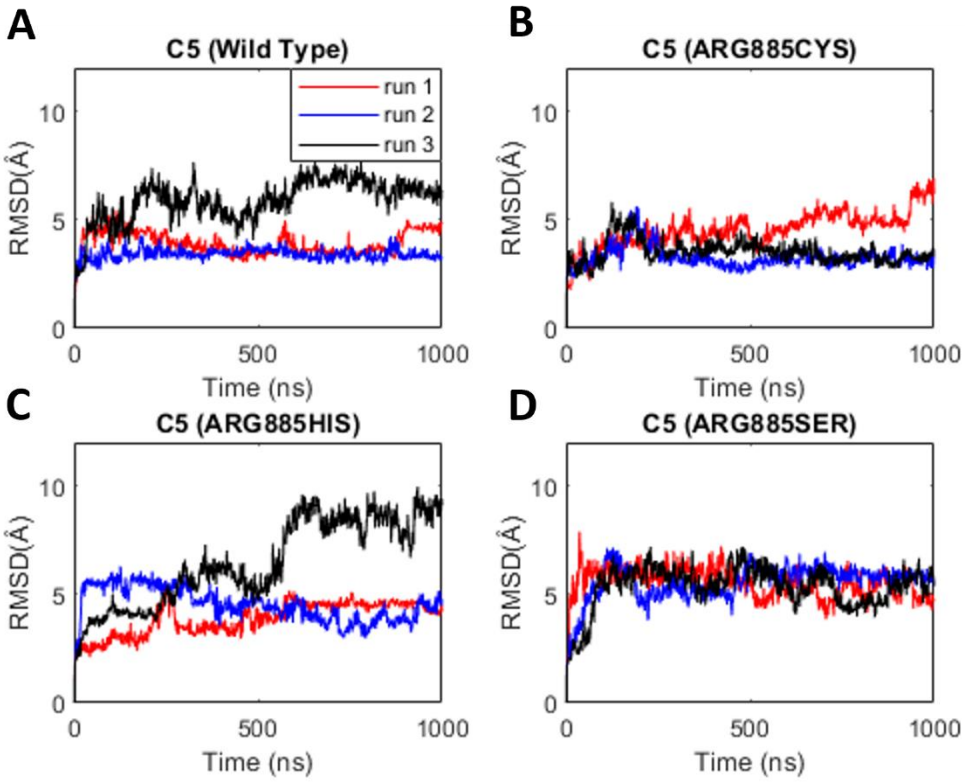

Figure A.28. RMSD of the heavy chain of ravulizumab for C5 in complex with ravulizumab. **(A)** Wild type C5 and mutated C5 structures, **(B)** Cysteine 885, **(C)** Histidine 885, and **(D)** Serine 885.

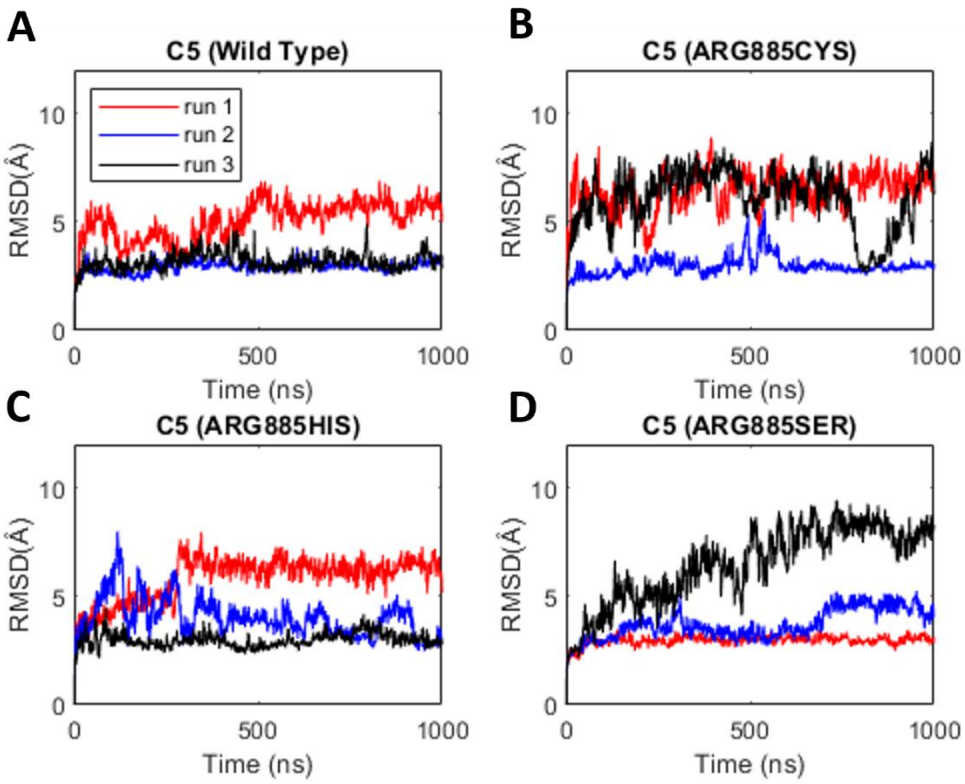

Figure A.29. RMSD of the light chain of eculizumab for C5 in complex with eculizumab. **(A)** Wild type C5 and mutated C5 structures, **(B)** Cysteine 885, **(C)** Histidine 885, and **(D)** Serine 885.

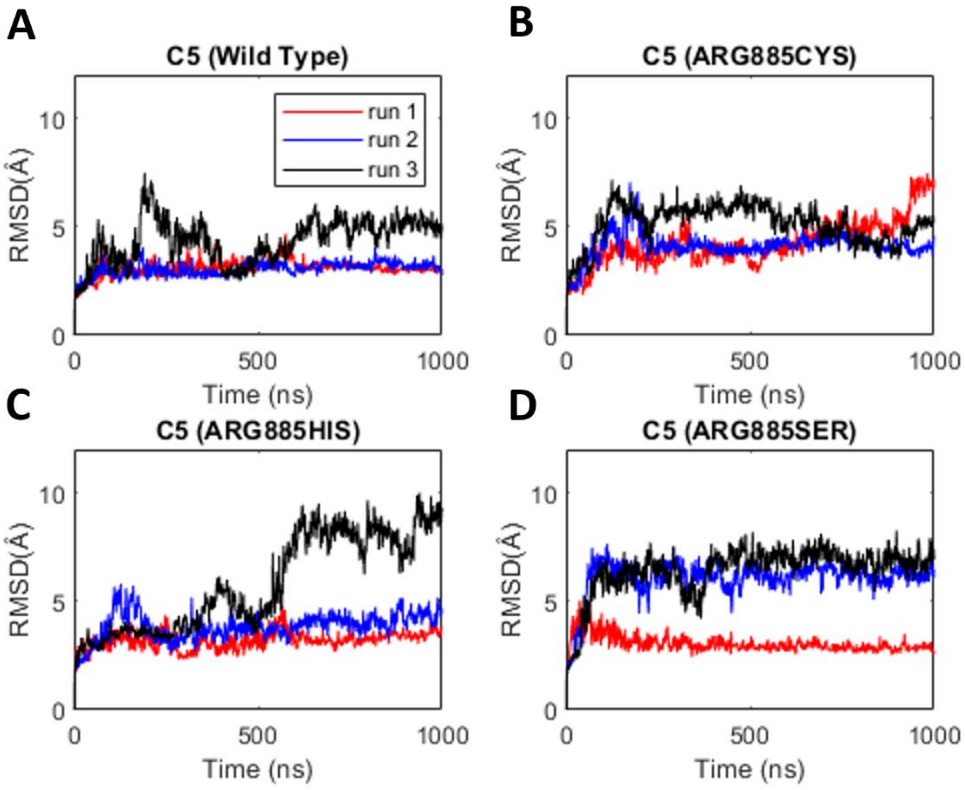

Figure A.30. RMSD of the light chain of ravulizumab for C5 in complex with ravulizumab. (A) Wild type C5 and mutated C5 structures, (B) Cysteine 885, (C) Histidine 885, and (D) Serine 885.

Appendix B

Supporting Information for Chapter 3

B.1 Supplementary Materials and Methods

B.1.1 Structural models

Molecular Dynamics (MD) simulations have been based on the X-ray structure of the *Streptococcus pyogenes* (*Sp*) CRISPR-Cas9 in complex with RNA and DNA (4UN3.pdb), solved at 2.58 Å resolution.[1] This structure identifies the inactivated state of the HNH domain (i.e., the "conformational checkpoint").[2] Four model systems have been built, including base pair mismatches "mm" within the RNA:DNA at different positions (i.e., mm@16-17, mm@14-15, mm@12-13, mm@10-11, as in **Figure 3.1A**). These structural models have been embedded in explicit waters, while Na+ ions were added to neutralize the total charge, leading to an orthorhombic periodic simulation cell of ~145 · 110 · 145 $Å^3$, containing a total of ~220,000 atoms. Notably, the simulation systems have been built similarly to our recent paper,[3] which investigated the dynamics of CRISPR-Cas9 in the presence of base pair mismatches at positions 17-20, 18-20, 19-20 and 20. The outcomes of these previously published systems have been compared with the results presented in this paper.

B.1.2 Molecular Dynamics (MD)

MD simulations have been performed in analogy to our previous paper. In detail, conventional MD simulations have been carried out to equilibrate the systems, prior to

Gaussian accelerated MD (GaMD) simulations.[4] MD simulations have been performed using a simulation protocol successfully employed for CRISPR-Cas9[3,5–8] and widely adopted for other RNA/DNA nucelases,[9] using of the Amber ff12SB force field, which includes the ff99bsc0[10] corrections for DNA and the ff99bsc0+$\chi OL3$[11,12] corrections for RNA. The Åqvist force field[13] has been employed for Mg^{2+} ions. An integration time step of 2 fs has been employed. All bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Temperature control (300 K) has been performed via Langevin dynamics,[14] with a collision frequency $\gamma = 1/ps$. Pressure control was accomplished by coupling the system to a Berendsen barostat,[15] at a reference pressure of 1 atm and with a relaxation time of 2 ps. The system has been subjected to energy minimization to relax water molecules and counter ions, keeping the protein, the RNA, DNA and Mg ions fixed with harmonic position restraints of 300 kcal/mol · $Å^2$. Then, the system has been heated up from 0 to 100 K in the canonical ensemble (NVT), by running two simulations of 5 ps each, imposing position restraints of 100 kcal/mol · $Å^2$ on the above-mentioned elements of the system. The temperature was further increased up to 200 K in ~100 ps of MD in the isothermal-isobaric ensemble (NPT), reducing the restraint to 25 kcal/mol $Å^2$. Subsequently, all restraints were released and the temperature of the system was raised up to 300 K in a single NPT simulation of 500 ps. After ~ 1.1 ns of equilibration, ~ 10 ns of NPT runs were carried out allowing the density of the system to stabilize around 1.01 g/cm^{-3}. Finally, the ~100 ns have been carried out in NVT ensemble. Simulations have been performed using the GPU version of AMBER pmemd 18.[16]

B.1.3 Gaussian accelerated Molecular Dynamics (GaMD)

Accelerated MD (aMD) is an enhanced sampling method that works by adding a non-negative boost potential to smoothen the system potential energy surface (PES), thus effectively decreasing the energy barriers and accelerating transitions between the low-energy states.[17] The method has been extensively employed to accelerate protein dynamics in a variety of biomolecules (see Markwick & McCammon as a review).[18] However, the use of aMD for large biomolecular systems, such as CRISPR-Cas9, can suffer from high statistical noise, which hampers the characterization of the correct statistical ensemble.[19–22] To overcome this limitation, we employed here a novel and more robust Gaussian aMD (or GaMD)[4] method, in which the boost potential follows Gaussian distribution. This allows smoothly reconstructing the original shape of the potential energy surface, through accurate reweighting using cumulant expansion to the 2nd order. This has expanded the use of aMD to large biological systems, with applications of this method to G-protein coupled receptors,[23,24] the Mu opioid receptor,[25,26] T-cell receptors[27] and CRISPR-Cas9.[3,6,8]

Considering a system with N atoms at positions $\vec{r} = \{\vec{r}_1 \ldots \vec{r}_N\}$, when the system potential $V(\vec{r})$ is lower than a threshold energy E **(Equation B.1)**, the energy surface is modified by adding a boost potential as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), \qquad V(\vec{r}) < E, \tag{B.1}$$

$$\Delta V(\vec{r}) = \frac{1}{2} k \left(E - V(\vec{r})\right)^2, \tag{B.2}$$

where *k* is the harmonic force constant **(Equation B.2)**. The two adjustable parameters *E* and *k* are automatically determined by applying the following three criteria. First, for any two arbitrary potential values $V_1(\vec{r})$ and $V_2(\vec{r})$ found on the original energy surface, if $V_1(\vec{r}) < V_2(\vec{r})$, ΔV should be a monotonic function that does not change the relative order of the biased potential values, i.e., $V_1^*(\vec{r}) < V_2^*(\vec{r})$. Secondly, if $V_1(\vec{r}) < V_2(\vec{r})$, the potential difference observed on the smoothened energy surface should be smaller than that of the original, i.e., $V_2^*(\vec{r}) - V_1^*(\vec{r}) < V_2(\vec{r}) - V_1(\vec{r})$. By combining the first two criteria and plugging in the formula of $V^*(\vec{r})$ and ΔV, we obtain:

$$V_{\max} \le E \le V_{\min} + \frac{1}{k}, \qquad \textbf{(B.3)}$$

where $V_{\min}$ and $V_{\max}$ are the system minimum and maximum potential energies **(Equation B.3)**. To ensure that **(Equation B.3)** is valid, *k* has to satisfy $k \le \frac{1}{V_{\max} - V_{\min}}$. By defining $k \equiv k_0 \frac{1}{V_{\max} - V_{\min}}$, then $0 < k_0 \le 1$. Thirdly, the standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure accurate reweighting using cumulant expansion to the second order: $\sigma_{\Delta V} = k\left(E - V_{avg}\right)\sigma_V \le \sigma_0$, where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV and σ_0 as a user-specified upper limit (e.g., 10 *k*BT) for accurate reweighting. When *E* is set to the lower bound, $E = V_{\max}$, according to **(Equation B.4)**, k_0 can be calculated as:

$$k_0 = \left(1.0, k_0^{'}\right) = \left(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{\max} - V_{\min}}{V_{\max} - V_{avg}}\right). \quad \textbf{(B.4)}$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{\min} + \frac{1}{k}, k_0$ is:

$$k_0 = k_0^{''}\left(1 - \frac{\sigma_0}{\sigma_V}\right) \cdot \frac{V_{\max} - V_{\min}}{V_{avg} - V_{\min}}, \quad \textbf{(B.5)}$$

if $k_0^{''}$ is calculated between 0 and 1 **(Equation B.5)**. Otherwise, k_0 is calculated using **(Equation B.4)**, instead of being set to 1 directly as described in the original paper. In Gaussian aMD, even with biasing potential, the same low-energy physical states are sampled, such enabling quantitative recovery of conformational distributions through reweighting, while unreweighted results can be used to sample low-energy physical state provide a useful semi-quantitative ranking of their probabilities. For our purposes, here we analyze unreweighted results, as in our previous paper on the off-target effects in CRISPR-Cas9,[3] such obtaining a broad exploration of the conformational dynamics.

Based on extensive testing, performed in our previous study on the CRISPR-Cas9 conformational dynamics,[3,6,8] the system threshold energy has be set to $E = V_{\max}$ for all Gaussian aMD simulations. The boost potential has been applied in a dual-boost scheme, in which two acceleration potentials are applied simultaneously to the system: (i) the torsional terms only and (ii) across the entire potential. A timestep of 2 fs has been used. Given an average system size of ~220K atoms, the maximum, minimum, average and standard deviation values of the system potential ($V_{\max}$, $V_{\min}$, V_{avg} and σ_V) has been obtained from an initial ~100 ns NVT simulation with no boost potential (see details

above). Each Gaussian aMD simulation proceeded with a ~50 ns run, in which the boost potential has been updated every 1.6 ns, thus reaching equilibrium values. Finally, ~1 μs of Gaussian aMD simulations have been carried out in the NVT ensemble for each system, in analogy to our previous paper,[3] to enable proper comparison.

B.1.4 Analysis of the RNA:DNA hybrid structure

Analysis of the RNA:DNA dynamics has been done over the Gaussian aMD production runs using the CURVES+ code.[28] As a measure of the base pair complementarity, we computed the *Propeller Twist* angle, which describes the rotation of couples of base pairs with respect to each other. Based on our previous study,[3] this parameter enables to properly characterize alterations in the base pairing along the RNA:DNA hybrid. The computed *Propeller Twist* angles have been plotted employing "violin plots" (or vioplot), which provide an overall view of the probability distribution. The minor groove has been measured between cubic spline curves running through the phosphorus atoms of the nucleic backbone and then reduced by 5.8 Å (2 x 2.9 Å) to discount the average radius of two adjacent phosphodiester backbones. Analysis of the minor groove width includes the calculation of the statistical error at each level of the RNA:DNA hybrid **(Figures 3.2 and B.2)**. The analysis of the results has been performed over the last ~800 ns of GaMD, as in our previous paper,[3] which is used as a comparison for the current paper. This choice has been motivated by the analysis of the RMSD of the RNA:DNA hybrid, which stabilizes after the first ~200 ns of GaMD **(Figure B.1)**. Hence, the last converged part (i.e., last ~800 ns) of the simulated runs has been object of

analysis. To father validate this choice, the conformational changes of the RNA:DNA hybrid structure have also been analyzed over the last ~400 ns of GaMD, reporting no significant difference with the analysis performed over the last ~800 ns **(Figure B.2)**.

B.1.5 Principal Component Analysis (PCA)

In PCA, the covariance matrix of the protein Cα atoms is calculated and diagonalized to obtain a new set of generalized coordinates (eigenvectors) to describe the system motions. Each eigenvector – also called Principal Component (PC) – is associated to an eigenvalue corresponding to the mean square fluctuation contained in the system's trajectory projected along that eigenvector. By sorting the eigenvectors according to their eigenvalues, the first few Principal Components (PCs) corresponds to the system's largest amplitude motion (variance), and the dynamics of the system along these PCs is referred as *"essential dynamics".*[29] Here, each conformation of the HNH domain sampled during the Gaussian aMD trajectories is projected into the collective coordinate space defined by the first two eigenvectors (PC1 and PC2), such allowing the characterization of the essential conformational sub-space sampled by Cas9 during Gaussian aMD. Importantly, each simulated system has been superposed onto the same reference structure and aligned, such allowing the projection into the same collective coordinate space. PCA has been performed using cpptraj of Amber18,[16] while the Normal Mode Wizard plugin of the Visual Molecular Dynamics[30] program has been used for the graphical rendering in **Figure 3.3**.

B.1.6 Cross-Correlation analysis

Cross-Correlation (CC_{ij}) analysis has been performed in order to identify the coupling of the motions between the residues of the HNH domain and of the DNA TS. The CC_{ij} coefficients have been computed between the Cα atoms of the HNH domain (*i*) and the TS phosphate atoms (*j*), as follows:

$$CC_{ij} = \frac{\left\langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \right\rangle}{\left(\left\langle \Delta\vec{r}_i(t)^2 \right\rangle \left\langle \Delta\vec{r}_j(t)^2 \right\rangle \right)^{\frac{1}{2}}} \quad \textbf{(B.6)}$$

where Δr_i and Δr_j are the fluctuation vectors of the atoms *i* and *j*, respectively **(Equation B.6)**. The angle brackets represent an average over the sampled time period. The value of CC_{ij} ranges from -1 to 1. Positive CC_{ij} values describe a correlated motion between atoms *i* and *j*, while negative CC_{ij} values describe anti-correlated motions. The CC_{ij} have been computed between the residues of the HNH domain that locate in proximity of the hybrid (i.e., residues 890-900, 901-910 and 911-920, which form three α-helices, **Figure 3.4**) and the TS bases from position b20 to b9, and have been plotted as a 2x2 matrix **(Figure 3.4)**.

B.2 Supplementary Figures

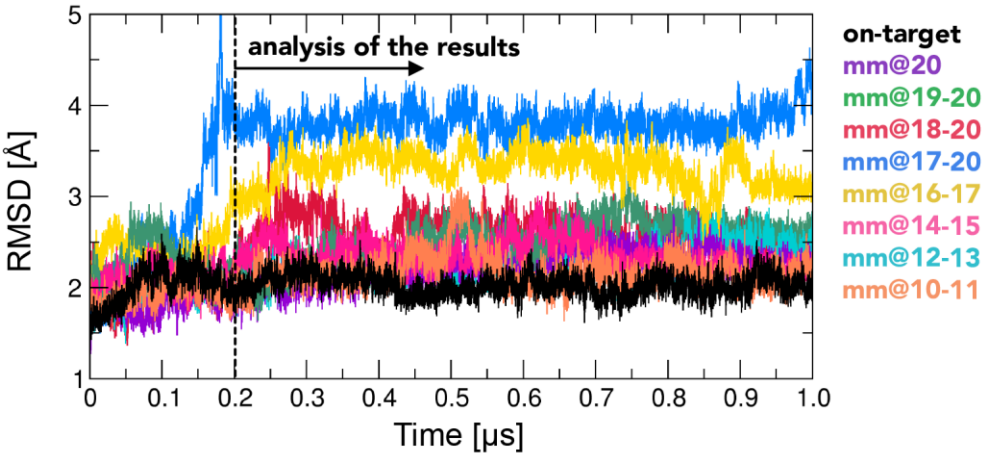

Figure B.1. Time evolution of the Root Mean Square Deviation (RMSD) of the RNA:DNA hybrid structure, along Gaussian accelerated MD (GaMD) of the CRISPR-Cas9 system including the on-target DNA (i.e., on-target system) and base pair mismatches (mm) at different positions of the hybrid (i.e., mm@20, mm@19−20, mm@18−20, mm@17−20, mm@16−17, mm@14−15, mm@12−13 and mm@10−11 systems). The RMSD of the RNA:DNA hybrid stabilizes after the first ~0.2 μs of GaMD. Hence, the last converged ~0.8 μs have been considered for analysis. Notably, the mm@17−20 and mm@16−17 systems display increased RMSD values, due to the fact that the RNA:DNA hybrid undergoes structural changes.

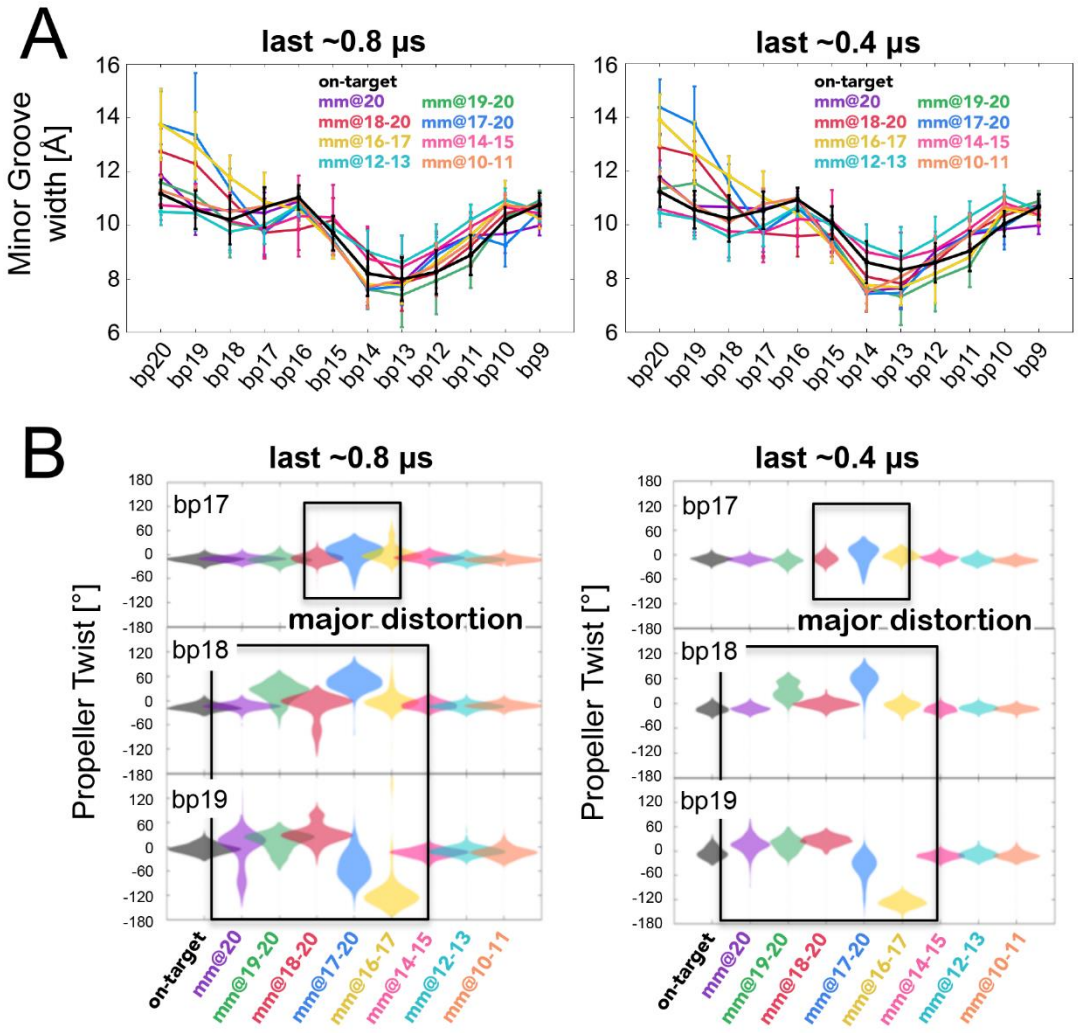

Figure B.2. Analysis of the conformations adopted by the RNA:DNA hybrid over the last ~0.8 µs of GaMD (left panel) and over the last ~0.4 µs of GaMD (right panel). **(A)** Minor groove width measured at different levels of the RNA:DNA hybrid (i.e., from base pair 20 to 9) of the CRISPR-Cas9 system including the on-target DNA (i.e., on-target system) and base pair mismatches (mm) at different positions of the hybrid (i.e., mm@20, mm@19−20, mm@18−20, mm@17−20, mm@16−17, mm@14−15, mm@12−13 and mm@10−11 systems). **(B)** Probability distribution (as violin plot) of the *Propeller Twist* angle for the base pairs (bp) at positions 17 (top graph), 18 (central graph) and 19 (bottom graph) of the RNA:DNA hybrid.

B.3 Supplementary References

(1) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. *Nature* **2014**, *513* (7519), 569–573. https://doi.org/10.1038/nature13579.

(2) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. *Sci. Adv.* **2017**, *3*, eaao002. https://doi.org/10.1126/sciadv.aao0027.

(3) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5* (4), 651–662. https://doi.org/10.1021/acscentsci.9b00020.

(4) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theor. Comput.* **2015**, *11*, 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

(5) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139* (45), 16028–16031. https://doi.org/10.1021/jacs.7b05313.

(6) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (28), 7260–7265. https://doi.org/10.1073/pnas.1707645114.

(7) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-Target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* **2016**, *2* (10), 756–763. https://doi.org/10.1021/acscentsci.6b00218.

(8) Palermo, G. Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59* (5), 2394–2406. https://doi.org/10.1021/acs.jcim.8b00988.

(9) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48* (2), 220–228. https://doi.org/10.1021/ar500314j.

(10) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids:

Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817–3829. https://doi.org/10.1529/biophysj.106.097782.

(11) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E. 3rd; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theor. Comput.* **2010**, *6*, 3836–3849. https://doi.org/10.1021/ct100481h.

(12) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(13) Aqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94* (21), 8021–8024. https://doi.org/10.1021/j100384a009.

(14) Turq, P.; Lantelme, F.; Friedman, H. L. Brownian Dynamics: Its Applications to Ionic Solutions. *J. Chem. Phys.* **1977**, *66*, 3039. https://doi.org/10.1063/1.434317.

(15) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684. https://doi.org/10.1063/1.448118.

(16) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; and Kollman, P. A. AMBER 2016. University of California, San Francisco 2016.

(17) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929. https://doi.org/10.1063/1.1755656.

(18) Wereszczynski, J.; McCammon, J. A. Nucleotide-Dependent Mechanism of Get3 as Elucidated from Free Energy Calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (20), 7759–7764. https://doi.org/10.1073/pnas.1117441109.

(19) Fajer, M.; Hamelberg, D.; McCammon, J. A. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J. Chem. Theory Comput.* **2008**, *4* (10), 1565–1569.

https://doi.org/10.1021/ct800250m.

(20) Jiang, W.; Thirman, J.; Jo, S.; Roux, B. Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis. *J. Phys. Chem. B* **2018**, *122* (41), 9435–9442. https://doi.org/10.1021/acs.jpcb.8b03277.

(21) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677–2689. https://doi.org/10.1021/ct500090q.

(22) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. Accelerated Molecular Dynamics Simulations of Protein Folding. *J. Comput. Chem.* **2015**, *36* (20), 1536–1549. https://doi.org/10.1002/JCC.23964.

(23) Miao, Y.; McCammon, J. A. Mechanism of the G-Protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (12), 3036–3041. https://doi.org/10.1073/pnas.1800756115.

(24) Miao, Y.; McCammon, J. A. Graded Activation and Free Energy Landscapes of a Muscarinic G-Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (43), 12162–12167. https://doi.org/10.1073/pnas.1614538113.

(25) Liao, J. M.; Wang, Y. T. In Silico Studies of Conformational Dynamics of Mu Opioid Receptor Performed Using Gaussian Accelerated Molecular Dynamics. *J Biomol Struct Dyn* **2018**, 1–12. https://doi.org/10.1080/07391102.2017.1422025.

(26) Wang, Y.-T.; Chan, Y.-H. Understanding the Molecular Basis of Agonist/Antagonist Mechanism of Human Mu Opioid Receptor through Gaussian Accelerated Molecular Dynamics Method. *Sci. Rep.* **2017**, *7* (1), 7828. https://doi.org/10.1038/s41598-017-08224-2.

(27) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; McAffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W.; Dong, S.; Gee, M. H.; Jude, K. M.; Davis, M. M.; Groves, J. T.; Goddard, W. A.; Heath, J. R.; Evavold, B. D.; Vale, R. D.; Garcia, K. C. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **2018**, *174* (3), 672-687.e27. https://doi.org/10.1016/J.CELL.2018.06.017.

(28) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Res.* **2009**, *37* (17), 5917–5929. https://doi.org/10.1093/NAR/GKP608.

(29) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of

Proteins. *Proteins Struct. Funct. Bioinforma.* **1993**, *17* (4), 412–425. https://doi.org/10.1002/PROT.340170408.

(30) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 27-28,33-38.

Appendix C

Supporting Information for Chapter 4

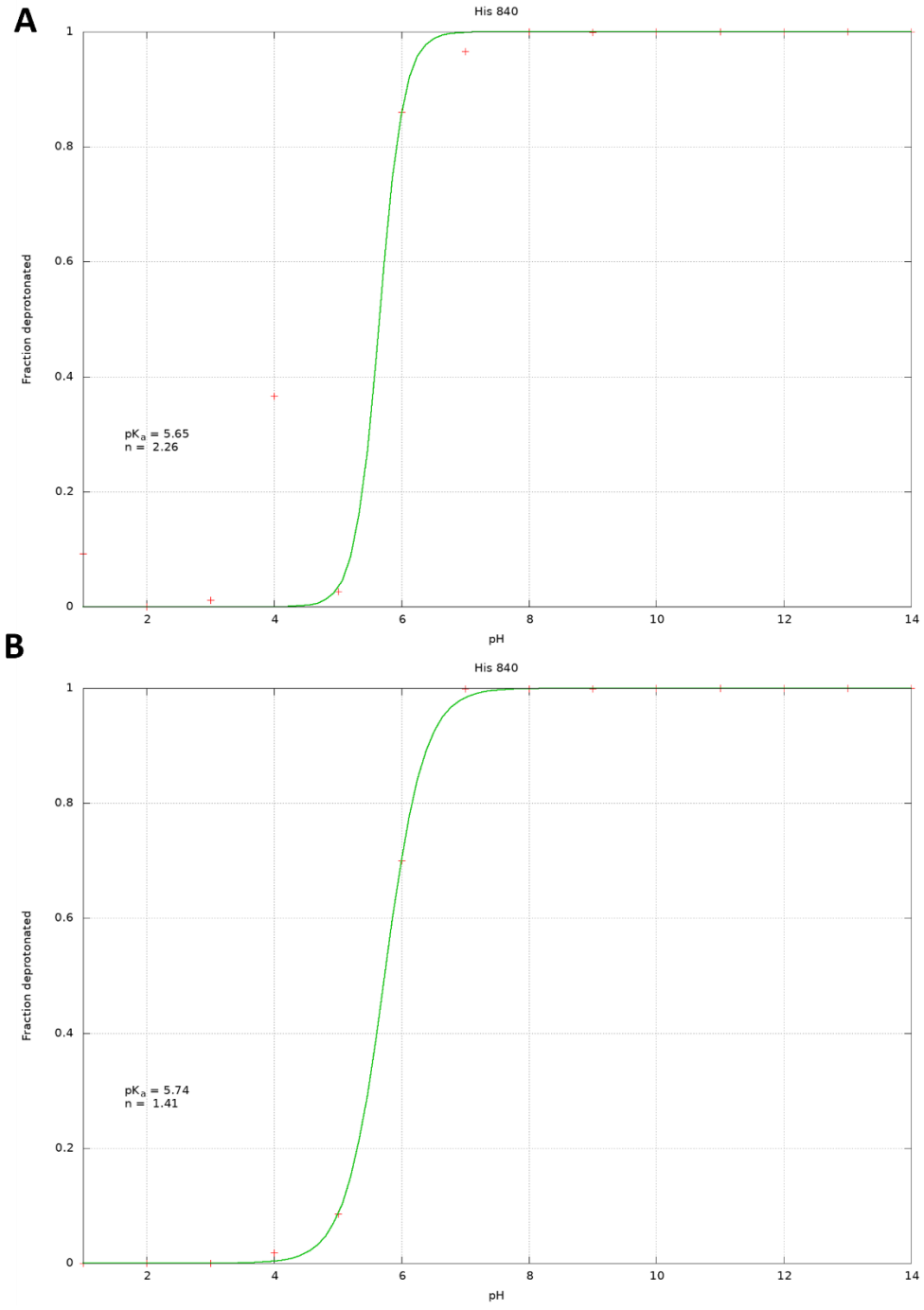

Figure C.1. Titration curves for HIS840 for the V838A simulation trajectory. Without restraints (A) and with restraints (B) applied during production.

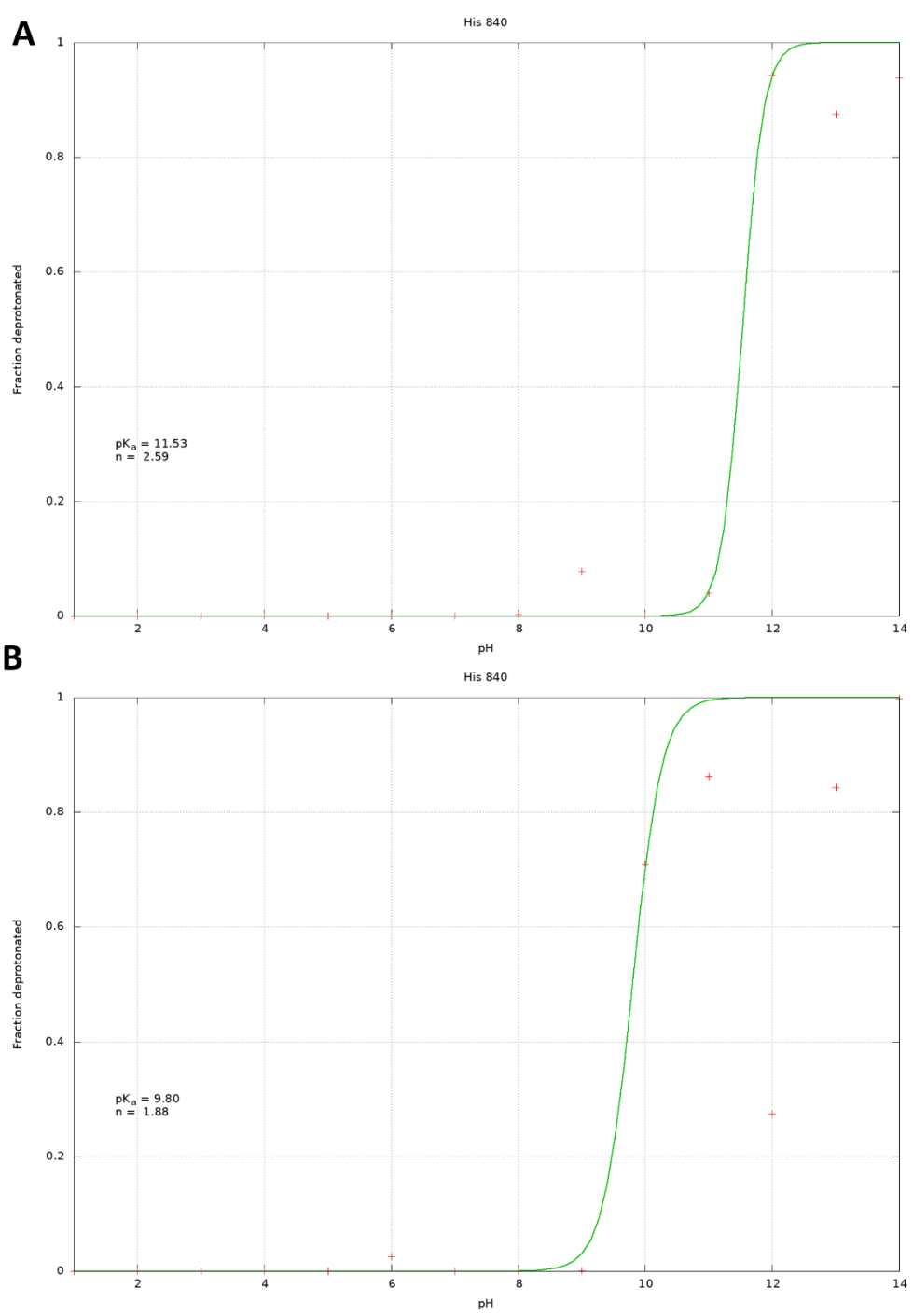

Figure C.2. Titration curves for HIS840 for the D839A simulation trajectory. Without restraints **(A)** and with restraints **(B)** applied during production.

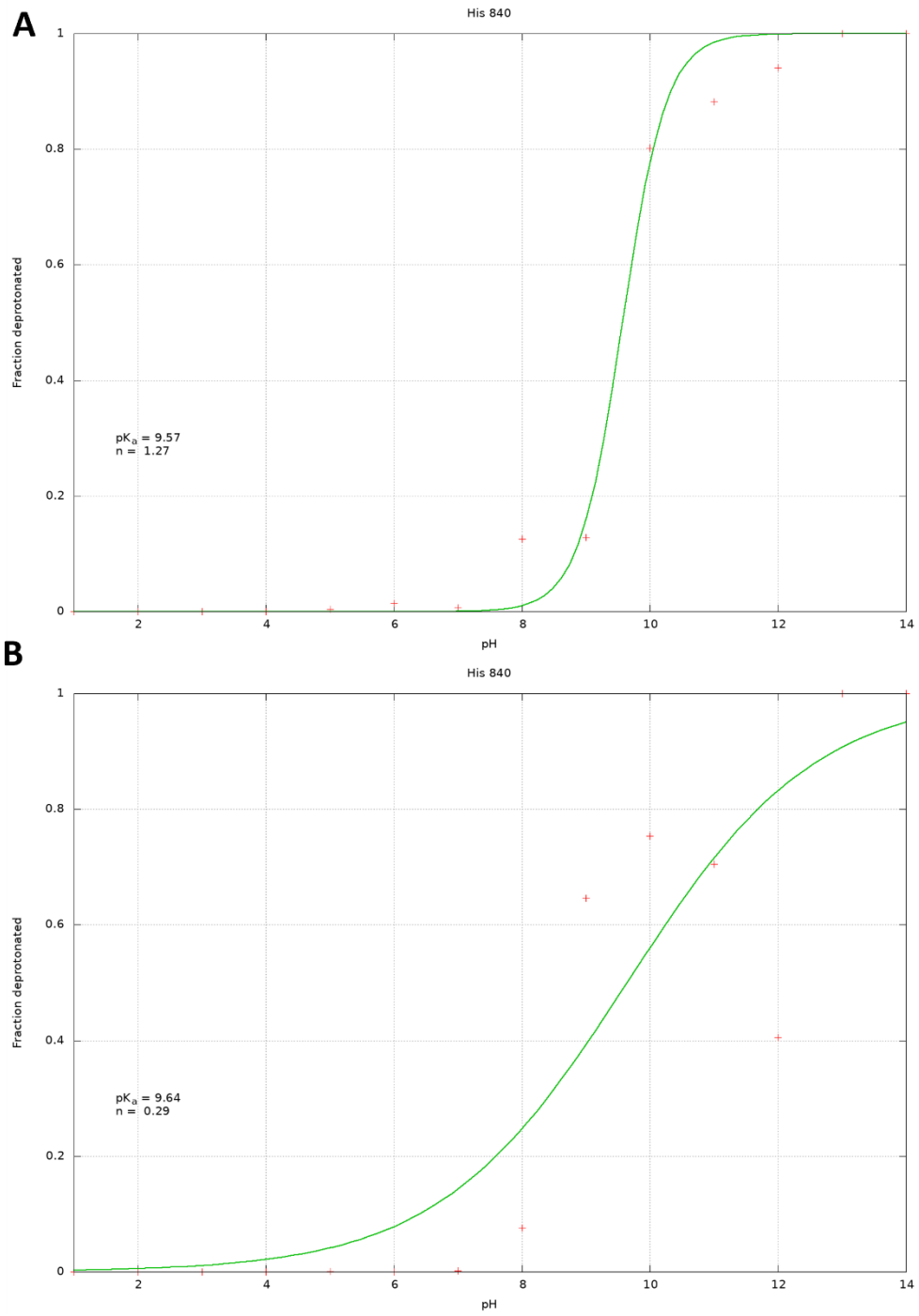

Figure C.3. Titration curves for HIS840 for the D861A simulation trajectory. Without restraints **(A)** and with restraints **(B)** applied during production.

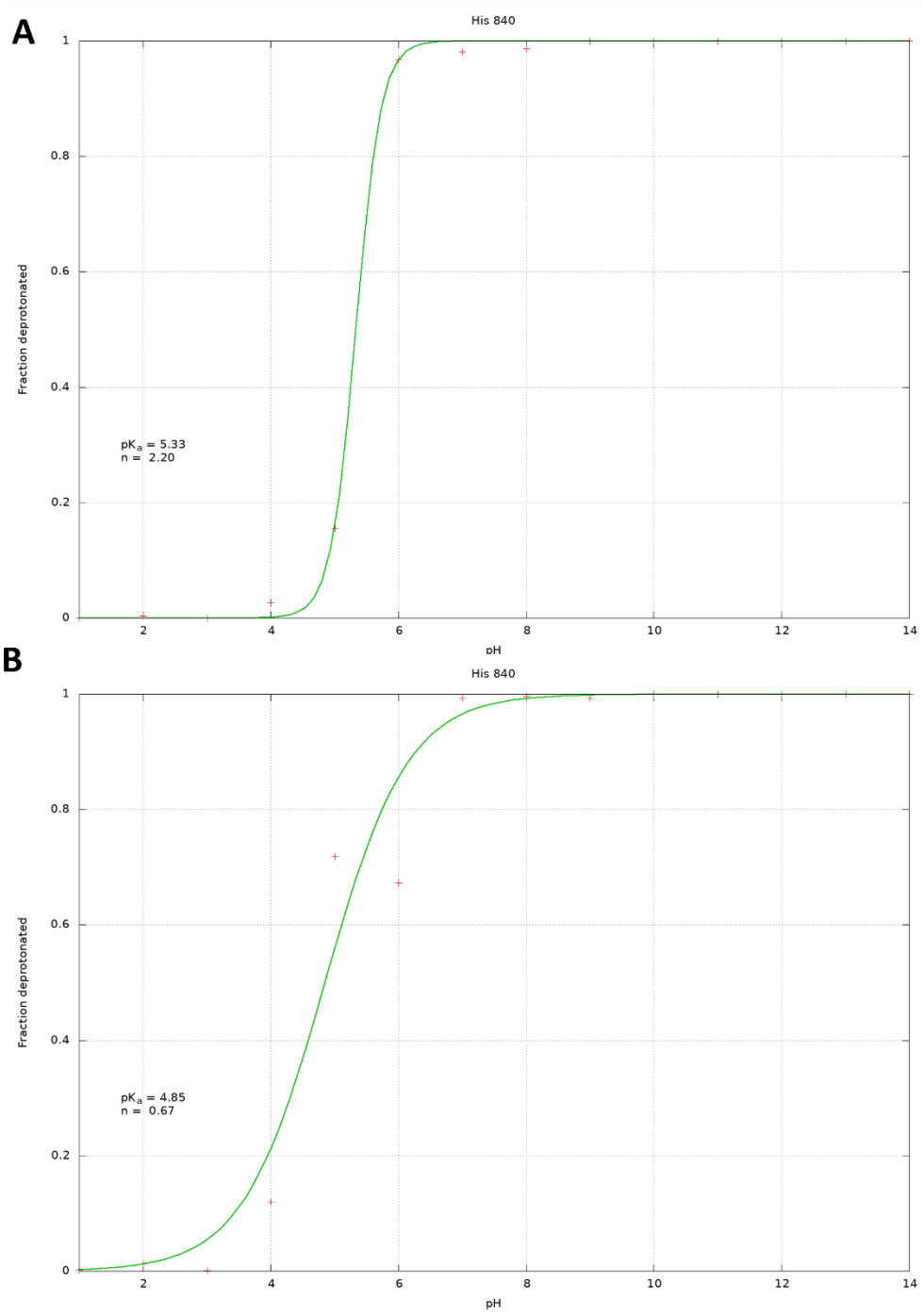

Figure C.4. Titration curves for HIS840 for the K862A simulation trajectory. Without restraints **(A)** and with restraints **(B)** applied during production.

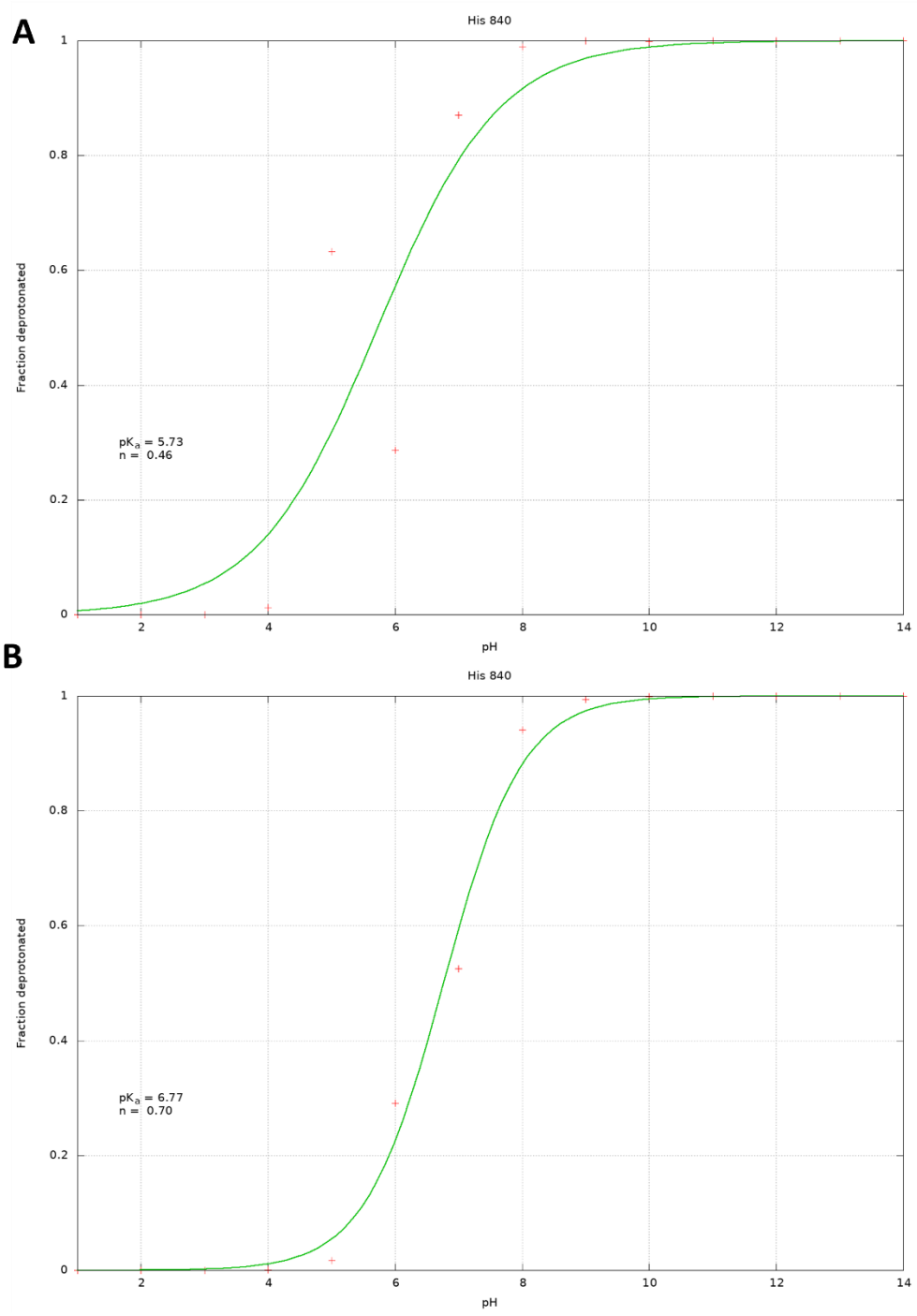

Figure C.5. Titration curves for HIS840 for the N863A simulation trajectory. Without restraints **(A)** and with restraints **(B)** applied during production.

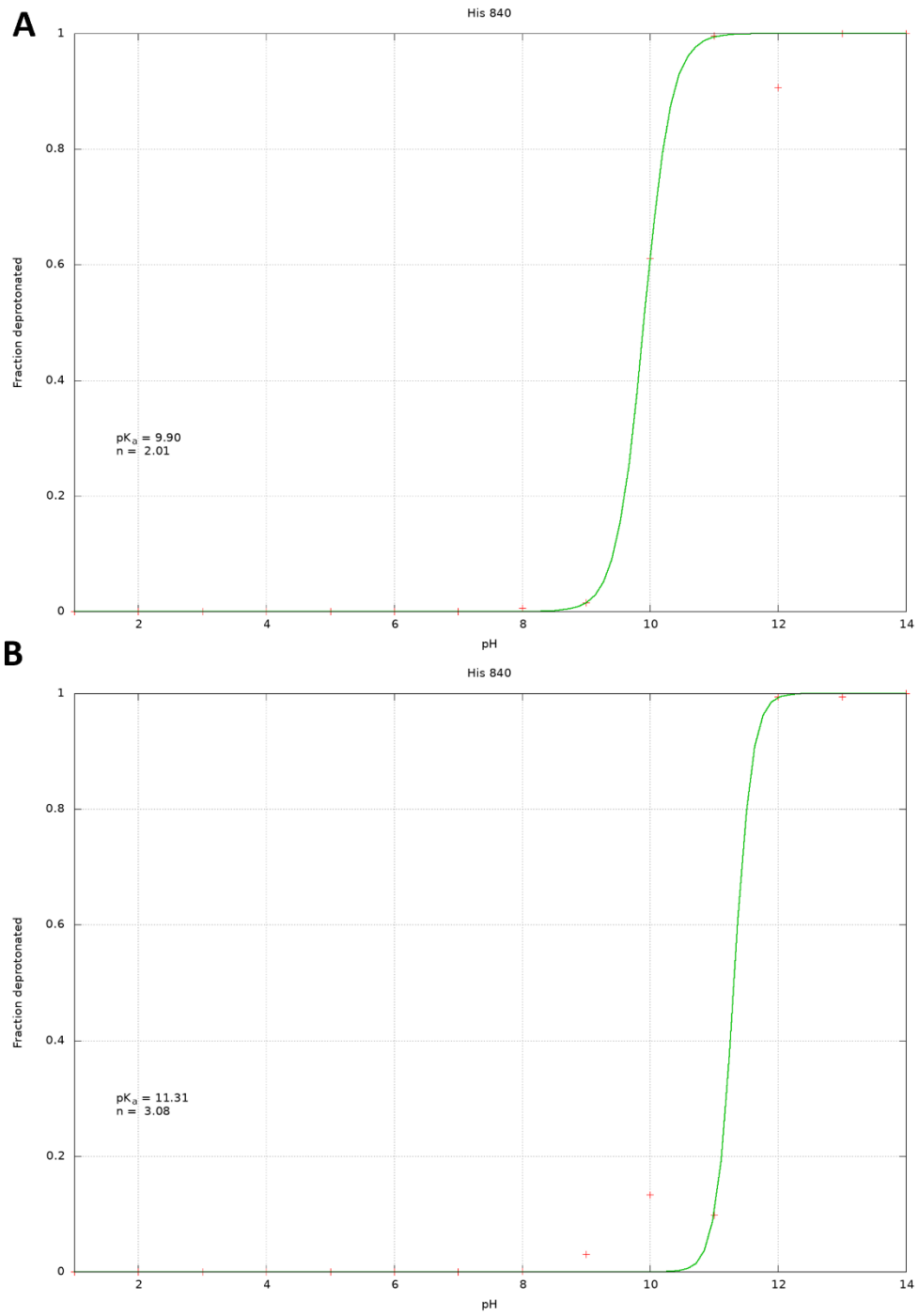

Figure C.6. Titration curves for HIS840 for the K866A simulation trajectory. Without restraints **(A)** and with restraints **(B)** applied during production.