

## UC Davis

### UC Davis Electronic Theses and Dissertations

#### Title

Applications in QM and MM: the ab initio nanoreactor solves mechanisms in green and biological chemistry, and simulations probe the role of sialic acids in viral mechanisms

#### Permalink

<https://escholarship.org/uc/item/57f86505>

#### Author

Oh, Lisa

#### Publication Date

2022

Peer reviewed|Thesis/dissertation

Applications in QM and MM: the *ab initio* nanoreactor solves mechanisms in green and biological chemistry, and simulations probe the role of sialic acids in viral mechanisms

By

LISA OH  
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

---

Lee-Ping Wang, Chair

---

Alexei Stuchebrukhov

---

Gang-yu Liu

Committee in Charge

2022

## Table of Contents

ABSTRACT .....	V
ACKNOWLEDGEMENTS .....	VI
CHAPTER 1: INTRODUCTION .....	1
GREEN CHEMISTRY .....	2
COMPUTATIONAL CHEMISTRY <i>AB INITIO</i> METHODS .....	3
SIALIC ACIDS IN VIRAL MECHANISMS .....	8
COMPUTATIONAL METHODS IN LIGAND-PROTEIN BINDING .....	12
CHAPTER 2: APPLICATION OF THE <i>AB INITIO</i> NANOREACTOR TO THE DEVELOPMENT OF A NEW PEROXIDE-DRIVEN KETONE OXIDATION REACTION THE WORK IN THIS CHAPTER IS BASED ON “APPLICATION OF THE <i>AB INITIO</i> NANOREACTOR TO THE DEVELOPMENT OF A NEW PEROXIDE-DRIVEN KETONE OXIDATION REACTION” WITH ANDREW L. OTSUKI, AIDIN R. BALO, MARK MASCAL, AND LEE-PING WANG, SUBMITTED TO ANGEWANDTE CHEMIE.....	15
INTRODUCTION .....	16
RESULTS AND DISCUSSION.....	19
<i>Initial mechanistic hypotheses based on 3-APA as an intermediate</i> .....	19
<i>Nanoreactor simulations involving 3-APA as an intermediate</i> .....	21
<i>Nanoreactor predictions involving acrylate as an intermediate</i> .....	24
<i>A role for radicals</i> .....	25
CONCLUSION .....	30
COMPUTATIONAL METHODS .....	31
ACKNOWLEDGEMENTS .....	34
CHAPTER 3: <i>O</i> -ACETYL MIGRATION WITHIN THE SIALIC ACID SIDE CHAIN: A MECHANISTIC STUDY BY THE <i>AB INITIO</i> NANOREACTOR THE WORK IN THIS CHAPTER IS BASED ON “ <i>O</i> -ACETYL MIGRATION WITHIN THE SIALIC ACID SIDE CHAIN: A MECHANISTIC STUDY BY THE <i>AB INITIO</i> NANOREACTOR” WITH YANG JI, WANQING LI, AJIT VARKI, XI CHEN, AND LEE-PING WANG, SUBMITTED TO BIOCHEMISTRY.....	35
ABSTRACT .....	36
BACKGROUND .....	36
METHODS .....	38
RESULTS AND DISCUSSION.....	43
CONCLUSION .....	52
ACKNOWLEDGEMENTS .....	53
CHAPTER 4: SARS-COV-2 AND MERS-COV SPIKE PROTEIN BINDING STUDIES SUPPORT STABLE MIMIC OF BOUND 9- <i>O</i> -ACETYLATED SIALIC ACIDS THE WORK IN THIS CHAPTER IS BASED ON “SARS-COV-2 AND MERS-COV SPIKE PROTEIN BINDING STUDIES SUPPORT STABLE MIMIC OF BOUND 9- <i>O</i> -ACETYLATED SIALIC ACIDS” WITH AJIT VARKI, XI CHEN, AND LEE-PING WANG, SUBMITTED TO NATURE COMMUNICATIONS.....	54
ABSTRACT .....	55
INTRODUCTION .....	55
RESULTS AND DISCUSSION.....	59

<i>Method Validation with Sias-MERS-CoV S binding</i> .....	59
<i>Discovery and analysis of Sias-SARS-CoV-2 S binding</i> .....	61
<b>CONCLUSION</b> .....	66
<b>METHODS</b> .....	67
<i>Choice of protein structures, sialic acids, and binding poses</i> .....	67
<i>Molecular dynamics simulations and MM-PBSA setup and procedure</i> .....	69
<i>Alchemical free energy simulations setup and procedure</i> .....	71
<i>Sialic acid parametrization procedure</i> .....	75
<i>Acknowledgements</i> .....	77
<b>CHAPTER 5: SUMMARY AND FUTURE WORK</b> .....	<b>78</b>
<b>REFERENCES</b> .....	<b>81</b>
<b>APPENDIX 1 FOR APPLICATION OF THE AB INITIO NANOREACTOR TO THE DEVELOPMENT OF A NEW PEROXIDE-DRIVEN KETONE OXIDATION REACTION</b> .....	<b>96</b>
<b>DETAILS OF THE EPR EXPERIMENTS</b> .....	96
<i>Supporting Figure 2.1. X-band EPR spectra of LA experiment and control reactions</i> .....	97
<b>COMPUTATIONAL METHODS</b> .....	97
<i>Nanoreactor Settings</i> .....	97
<i>Supporting Table 2.1. Sample simulation conditions for the ab initio nanoreactor</i> .....	102
<b>3-D RENDERINGS OF STRUCTURES IN MANUSCRIPT REACTION SCHEMES:</b> .....	102
<i>Supporting Figure 2.2: Details of reaction pathway LA<sup>-</sup> → HP-LA shown in Scheme 2.2a</i> .....	103
<i>Supporting Figure 2.3: Details of reaction pathway HP-LA → 3-APA<sup>-</sup> shown in Scheme 2.2a</i> .....	104
<i>Supporting Figure 2.4: Details of reaction pathway 3-APA<sup>-</sup> → 1 shown in Scheme 2.2b, 2d and 3</i> .....	105
<i>Supporting Figure 2.5: Details of reaction pathway 1 → 2 shown in Scheme 2.2b</i> .....	106
<i>Supporting Figure 2.6: Details of reaction pathway 2 → 3 shown in Scheme 2.2b</i> .....	107
<i>Supporting Figure 2.7: Details of reaction pathway 3 → 3-HPPA<sup>2-</sup> shown in Scheme 2.2b</i> .....	108
<i>Supporting Figure 2.8: Details of reaction pathway 3-APA<sup>-</sup> → 4 shown in Scheme 2.2c</i> .....	109
<i>Supporting Figure 2.9: Details of reaction pathway 4 → 5 shown in Scheme 2.2c</i> .....	110
<i>Supporting Figure 2.10: Details of reaction pathway 5 → 3-HPPA<sup>-</sup> shown in Scheme 2.2c</i> .....	111
<i>Supporting Figure 2.11: Details of reaction pathway 1 → 3 shown in Scheme 2.2d</i> .....	112
<i>Supporting Figure 2.12: Details of reaction pathway 3-APA<sup>-</sup> → 3-HPPA<sup>-</sup> shown in Scheme 2.3</i> .....	113
<i>Supporting Figure 2.13: Details of reaction pathway 3-APA<sup>-</sup> → 3-HPA<sup>-</sup> shown in Scheme 2.3</i> .....	114
<i>Supporting Figure 2.14: Details of reaction pathway 1 → 3-HPA<sup>-</sup> shown in Scheme 2.3</i> .....	115
<i>Supporting Figure 2.15: Details of reaction pathway 3-APA<sup>-</sup> → 6 shown in Scheme 2.3</i> .....	116
<i>Supporting Figure 2.16: Details of reaction pathway 6 → 3-HPA<sup>-</sup> shown in Scheme 2.3</i> .....	117
<i>Supporting Figure 2.17: Details of reaction pathway HP-LA → 7 shown in Scheme 2.4</i> .....	118
<i>Supporting Figure 2.18: Details of reaction pathway HP-LA' → 7 shown in Scheme 2.4</i> .....	119
<i>Supporting Figure 2.19: Details of reaction pathway HP-LA → 9 shown in Scheme 2.5</i> .....	120
<i>Supporting Figure 2.20: Details of reaction pathway HP-LA → 9 shown in Scheme 2.5</i> .....	121
<i>Supporting Figure 2.21: Details of reaction pathway 9 → 10 shown in Scheme 2.5</i> .....	122
<i>Supporting Figure 2.22: Details of reaction pathway HP-LA → 10 shown in Scheme 2.5, calculated using B3LYP</i> .....	123
<i>Supporting Figure 2.23: Details of reaction pathway 10 → 3-HPPA<sup>2-</sup> shown in Scheme 2.5</i> .....	124
<b>3-D RENDERINGS OF STRUCTURES RELATED TO MANUSCRIPT REACTION SEARCH</b> .....	125
<i>Supporting Figure 2.24: Details of HO- addition to 7 to form 3-HPA<sup>-</sup>, step 1/2</i> .....	125

<i>Supporting Figure 2.25: Details of HO<sup>-</sup> addition to 7 to form 3-HPA<sup>-</sup>, step 2/2</i> .....	126
<i>Supporting Figure 2.26: Elimination to Acrylate via 5-membered-ring TS with 2 K<sup>+</sup> and 6 H<sub>2</sub>O</i> .....	127
<i>Supporting Figure 2.27: Proton Transfer Between Oxo- and Peroxo- Groups of HP-LA</i> .....	128
<i>Supporting Figure 2.28: Proton Transfer from C2H to 4-Peroxo to Form “Enolate” Intermediate</i> .....	129
<i>Supporting Figure 2.29: Proton Transfer from C2H to 4-Oxo to Form “Enolate” Intermediate</i> .....	130
<i>Supporting Figure 2.30: Dissociation of “Enolate” Intermediate to Acrylate</i> .....	131

**APPENDIX 2 FOR O-ACETYL MIGRATION WITHIN THE SIALIC ACID SIDE CHAIN: A MECHANISTIC STUDY BY THE AB INITIO NANOREACTOR ..... 132**

FIGURE S3.1. REACTION FREE ENERGY AND BARRIER FOR NEU5,9AC <sub>2</sub> O-ACETYL MIGRATION CYCLIZATION STEP. ....	132
FIGURE S3.2. NANOREACTOR REACTION FREE ENERGIES AND ACTIVATION BARRIERS FOR THE DE-O-ACETYLATION AND DEPROTONATION OF THE C8-HYDROXYL GROUP IN NEU5,9AC <sub>2</sub> .....	133
FIGURE S3.3. FULL PATHWAY OF REACTION AND ACTIVATION FREE ENERGIES FOR BASE-CATALYZED O-ACETYL MIGRATION OF NEU5,9AC <sub>2</sub> , NEU5,8AC <sub>2</sub> AND NEU5,7AC <sub>2</sub> . ....	134
FIGURE S3.4. TORSIONAL ELECTRONIC ENERGY PROFILE OF GLYCEROL ROTATIONS BETWEEN 8 AND 9 IN THE O-ACETYL MIGRATION FROM C7-OH TO C8-OH.....	135
FIGURE S3.5. UMBRELLA SAMPLING OF EACH CYCLIZATION/O-ACETYL MIGRATION STEP BETWEEN NEU5,9AC <sub>2</sub> , NEU5,8AC <sub>2</sub> AND NEU5,7AC <sub>2</sub> . ....	136

**APPENDIX 3 FOR SARS-COV-2 AND MERS-COV SPIKE PROTEIN BINDING STUDIES SUPPORT STABLE MIMIC OF BOUND 9-O-ACETYLATED SIALIC ACIDS ..... 137**

<b>FIGURE S4.1:</b> MM-PBSA ENERGIES AND DECOMPOSITION ANALYSIS FOR NEU5,9AC <sub>2</sub> , NEU5AC9NAC AND NEU5AC IN MERS-CoV S PROTEIN.....	137
<b>FIGURE S4.2:</b> MM-PBSA ENERGIES AND DECOMPOSITION ANALYSIS FOR NEU5,9AC <sub>2</sub> , NEU5AC9NAC, NEU5AC, NEU5,9AC <sub>2</sub> A2-3GALBPNP, NEU5AC9NACA2-3GALBPNP AND NEU5ACA2-3GALBPNP IN ALL BINDING POSES (A-D) OF SARS-CoV-2 S PROTEIN .....	138
<b>FIGURE S4.3:</b> THERMODYNAMIC CYCLE TO ESTIMATE BINDING FREE ENERGIES.....	140
<b>FIGURE S4.4:</b> REPRESENTATIVE BINDING FREE ENERGY DIFFERENCE OF NEU5AC — NEU5,9AC <sub>2</sub> IN THE SARS-CoV-2 S PROTEIN WITH MBAR AND TI ENERGIES AND SIMULATION OVERLAP MATRICES.....	141

## Abstract

The development of environmentally friendly synthetic methodologies is a central and urgent goal of modern chemical science and industry. Advancing mechanistic understanding is crucial to accelerating the development of new reactions involving biobased platform molecules which can be applied to sustainable process design. The *ab initio* nanoreactor is a method for automated generation of mechanistic trajectories, in which high temperatures, high pressures, and external forces are applied to assemblies of reactants to advance discovery using first-principles MD simulations. I used this *ab initio* nanoreactor method to solve mechanistic mysteries of green and biological chemistry, which has proved pivotal in a synergistic theoretical and experimental studies. In another application, reaction design and accurate predictions of ligand-protein/protein-protein binding energies is critical in pharmaceutical drug development, and can drive efficient production from predicted reagents, minimize waste, and elucidate the core organizations of biology. Many disease-causing viruses target sialic acids (Sias) on the surface of host cells. Some viruses bind preferentially to sialic acids with *O*-acetyl modification at the hydroxyl group of C7, C8, or C9 on the glycerol-like side chain. Binding studies of proteins to sialosides containing *O*-acetylated sialic acids are crucial in understanding the related diseases, but experimentally difficult due to the lability of the ester group. As such, *N*-acetyl sialic acids have been proposed as stable mimics. I have studied the instability of the ester group across modified Sias, and the structural and biological similarities of these Sias in ligand-protein binding with MERS-CoV S and SARS-CoV-2 S proteins. I have found *N*-acetyl and *O*-acetyl Sias interchangeable, suggesting an experimentally reasonable mimic to probe viral mechanisms.

## Acknowledgements

I would like to dedicate this work to my family: my parents, Wayne and Mary Gong, who have supported me through the years, my uncle Professor William Gong and Uncle Steve Robb for encouraging me to pursue my Ph.D., my son Titus Oh for focusing my projects, and especially my husband Caleb Oh for inspiring and helping me finish my degree.

I am incredibly grateful to Lee-Ping Wang, my Ph.D. advisor, for all his advice, motivation, and feedback throughout the years. I appreciate not only your wisdom and experience, but also your kindness.

I am also thankful for all current and former members in the Wang group. I especially want to thank Yudong Qiu for helping me get started as a grad student and all your helpful tips, Nanhao Chen for all your helpful advice in any project, and Hyesu Jang, for continuing to inspire my research even after leaving lab, always being willing to share meals together, and for being a true friend.

I would like to also thank Ruby Reed, thank you for your mentorship and advice, I wouldn't be here without support.

Chapter 2 is from our submitted work "Application of the *ab initio* Nanoreactor to the Development of a New Peroxide-Driven Ketone Oxidation Reaction"; I am grateful to Andrew Otsuki, Aidin Balo, Mark Mascal, and Lee-Ping Wang, my co-authors in this collaboration.

Chapter 3 is from my submitted work "*O*-Acetyl migration within the sialic acid side chain: a mechanistic study by the *ab initio* nanoreactor"; I am grateful to my co-authors in this collaboration, Yang Ji, Wanqing Li, Ajit Varki, Xi Chen, and Lee-Ping Wang.

Additional to co-authors in Chapter 3, I am also grateful for Aniruddha Sasmal, Saurabh Srivastava, Audra A. Hargett, Brian R. Wasik, Hai Yu, Sandra Diaz, Biswa Choudhury, Colin R.

Parrish, and Darón I. Freedberg, for our collaboration in this migration project experimentally<sup>1</sup>. I also acknowledge Wanqing Li, Marcos D. Battistel, Hannah Reeves, Hai Yu, Xi Chen, Lee-Ping Wang, and Darón I. Freedberg for the exciting collaboration probing modified sialic acid structural similarities with NMR<sup>2</sup>.

Chapter 4 is from my submitted work “SARS-CoV-2 and MERS-CoV spike protein binding studies support stable mimic of bound 9-*O*-acetylated sialic acids”; I am grateful to my co-authors in this collaboration, Ajit Varki, Xi Chen, and Lee-Ping Wang for your helpful feedback and advice in this project.

I acknowledge financial support from the United States National Institutes of Health grant no. R01AI130684 for work in Chapters 3 and 4.

Above all, I thank God for leading me to this point and for all He has provided.



## **Chapter 1: Introduction**

## Green Chemistry

The development of environmentally friendly synthetic methodologies is a central and urgent goal of modern chemical science and industry.<sup>3-6</sup> Major thrusts of this goal involve replacing petrochemical feedstocks with renewable resources such as biomass and closer adherence to the principles of green chemistry.<sup>7-9</sup> A review by Gallezot, et al in 2012 points towards 5 driving forces for substituting fossil feedstocks with renewable carbon: increasing public confidence in the chemical industry by optimizing sustainable biomass-based chemicals production, industries in developed countries promoting the use and relevance of renewable resources in innovation, increasing bio-based products market with bulk chemicals, developing a specialized portfolio of bio-based products not currently represented in hydrocarbon-based manufacturing, and less legislative constraints.<sup>9</sup> Relevant bio-mass-derived building block chemicals for industrial applications have been screened based on a two-part process to determine the complexity of synthetic pathways and relevance of the building blocks and their respective derivatives.<sup>10</sup> Twelve building block chemicals were identified from over 300 by this process, some of which includes malic acids, aspartic acid, glutamic acid, levulinic acid, glycerol, and xylitol.<sup>10</sup>

Advancing mechanistic understanding is crucial to accelerating the development of new reactions involving biobased platform molecules which can be applied to sustainable process design.<sup>11</sup> For this reason, the pursuit of mechanistic knowledge in organic synthesis, using a synergistic combination of experimental and theoretical efforts, is expected to play an important role in bringing about the sustainable, carbon cycle-neutral chemical economy of the near future.

## Computational Chemistry *ab initio* Methods

Historically, experiment precedes computational studies, where computational methods were guided and validated by experimental results. Theoretical simulations yield detailed insight into mechanisms that is difficult to experimentally observe, if not impossible at experimental detection limits. Increasingly, theoretical simulations are viewed as a source for novel mechanistic hypotheses, not limited by precedent, chemical intuition, or extensive human input.

Time constrains this computer-aided molecular design. Improved accuracy of a computational method generally increases computational time, and the level of detail required to study a problem guides the choice of computational approach. Additionally, computational cost often scales non-linearly with system size, rendering some questions unsolvable with current methods and computer resources. The challenge in reactive computer-aided molecular design is computing interatomic forces without prior knowledge of chemical bonding, where these forces are used in resolving mechanistic details.<sup>12</sup> The electronic Schrodinger equation is often employed to compute these forces, using approaches such as semiempirical methods and density functional theory (DFT), to name a few.<sup>13,14</sup> A few holy grails of computational organic chemistry and biochemistry includes accurate predictions of ligand-protein/protein-protein binding energies, which elucidates the core organizations of biology and aids pharmaceutical drug development, and reaction design, which can drive efficient production from predicted reagents and reduced waste.<sup>15</sup>

Computational approaches to elucidating reaction pathways have recently been reported by several groups including ours.<sup>16-24</sup> A reaction pathway involves a single reactant/product combination,

connected by a single transition state (TS), where this TS is confirmed to connect the reactant/product by intrinsic reaction coordinate (IRC) computations. Given the electronic detail (first principles, or *ab initio*) required to describe a reaction path, and the high dimensionality ( $\sim 3N$ , where  $N$  is the number of atoms in a system) of reaction potential energy surface (PES) to search, these calculations are quite expensive, and prohibitive to exhaustively explore.

In general, there are 4 categories of approaches towards automated reaction path exploration by elementary steps.<sup>25</sup> In one category, reaction networks are based on encoded transformations such as from databases or chemical heuristics, where chemical heuristics have long since guided mechanistic searches. The quality of predicted reaction pathways depends on the quality of existing reactions that are used in this rules-based approach, and the similarity between the chemical system being explored and knowledge of similar existing reactions. Additionally, reaction barriers may be implicit, based on previous results, rather than explicit *ab initio* calculations. This could speed up reaction exploration at the cost of accuracy. In another, multiple transition state (TS) guesses are generated, such as by intuition, and subsequently optimized to single reactant/product species connecting the transition state. High quality potential energy surfaces (PES) at quantum mechanical level of detail required to describe TS's is massively dimensional. As such, TS approaches need to be reduced to look at smaller PES's, limiting the scope of reaction search. To mitigate the expensive of complete PES search, Maeda and coworkers developed an artificial force-induced reaction (AFIR) method, overcoming reaction barriers and inducing reactions by applying an artificial external force to push reactants together.<sup>26</sup> Methods such as AFIR generate TS and respective reaction paths, without necessarily knowing reaction pathways *a priori*. In a 3<sup>rd</sup> category, methods uncover potential reaction pathways and relevant

intermediates, and subsequently optimize TS's. In one example, generation of potential intermediates can be done through connectivity graphs generated from molecular dynamics trajectories, where differences in the connectivity graphs indicate bond forming/breakings steps for TS analysis.<sup>19,27</sup> Efficiency in this approach depends on the efficiency of TS optimization algorithms and potential reaction pathway generation. Given the expensive nature of quantum chemistry required to optimize TS of previous categories, another category explores reaction coordinates by combining MD with coordinate driving (CD), such as by Li et al.<sup>28</sup> While this method explores conformational isomers efficiency, the scope was limited to organic molecules, in part due to the expensive nature of MD.

Strategies employed have been highly diverse, including the use of semiempirical or fully *ab initio* molecular dynamics (MD) simulations to discover new mechanisms independent of pre-existing information about the reaction or putative intermediates.<sup>29</sup> We previously described the development of the *ab initio* nanoreactor approach to automated generation of mechanistic trajectories, in which high temperatures, high pressures, and external forces are applied to assemblies of reactants to advance discovery using first-principles MD simulations.<sup>17,19</sup> For example, such computations have revealed possible new pathways for amino acid synthesis from small inorganic compounds proposed to exist on the early Earth, and served as a proof of concept for dynamics-based computational reaction discovery.<sup>17</sup>

The nanoreactor is a specially modified AIMD simulation with the primary goal of inducing reaction events on a short simulation time scale, achieved by means of a time-oscillating boundary potential that generates high-energy molecular collisions. The initial conditions consist of one key

reactant molecule surrounded by other molecules, such as reacting reagents and a few solvent molecules, randomly placed within a relatively small sphere using Packmol software,<sup>30</sup> followed by energy minimization. Reaction events are induced by a time-dependent restraint potential in equation (1), where  $m$  is the atomic mass,  $R_1$  and  $R_2$  are radius parameters, and  $k_1$  and  $k_2$  are force constants. The restraint is a flat-bottomed harmonic potential with time-dependent parameters:

$$V(r,t) = \frac{m}{2} k(t) \rho(t)^2; \quad \rho = \begin{cases} r - R(t) & \text{if } r > R(t) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$k(t), R(t) = \begin{cases} k_1, R_1 & \text{if } t \bmod (t_1 + t_2) < t_1 \\ k_2, R_2 & \text{otherwise} \end{cases}$$

where the values of the radius  $R$  and spring constant  $k$  oscillate in a square wave pattern, and are equal to  $(R_1, k_1)$  for period  $t_1$  and  $(R_2, k_2)$  for period  $t_2$ . Generally,  $R_1 > R_2$  and during the period  $t_1$  the molecules are free to move within a spherical volume of radius  $R_1$  with no restraint force. When the radius decreases from  $R_1$  to  $R_2$ , atoms beyond  $R_2$  are accelerated inward, with the mass-dependent force ensuring near-uniform acceleration of atoms within a molecule. It is important to include this mass-weighted additional force to simultaneously accelerate all the atoms of a molecule inward. Throughout simulations, the key reactant is restrained to the center of the sphere. Reaction events are initiated from high velocity collisions of reactants, which generally occur within the smaller spherical volume of radius  $R_2$ . In general, increasing  $T$ ,  $R_1$ ,  $k_2$ , and  $t_1$  and decreasing  $R_2$  increases reactivity, but care must be to avoid unreasonably high-energy (explosive) simulations.

We varied the simulation setup and nanoreactor parameters to yield chemically reasonable reactivity within 2 ps of simulation time for each of our mechanistic studies. This involved

choosing which reactants and how many of each to model, generating multiple coordinates of randomized molecular placement within the initial sphere, and systematically varying nanoreactor simulation parameters. Simulations were limited to ~100 atoms for a few tens of picoseconds, given the high computational cost of *ab initio* molecular dynamics (AIMD) simulations. The implicit solvent was used in our simulations to relatively stabilize our charged intermediate species, compared to gas-phase calculations, while still generating relevant reaction pathway information, and without significantly increasing computational time, which would be the case in adding additional explicit water molecules. Simulations utilized the TeraChem quantum chemistry software<sup>31-37</sup> and were run on servers with an Intel Xeon CPU / Nvidia GPU (either E5-2630 v3 CPU / GTX 980 Ti GPU or E5-2637 v4 CPU / GTX 1080 Ti GPU). Use of TeraChem allowed these simulations to run on GPU's, which increases simulation calculation speeds in time compared to CPU's.

The reaction events observed in the nanoreactor trajectory underwent an energy refinement procedure to produce optimized transition states and corresponding reactant/product structures. Often, the refinement calculations used the  $\omega$ B97X-D3/TZVP/PCM level of theory,<sup>38-40</sup> which is more accurate (and more costly) compared to what was used for the nanoreactor simulations. The efficiency of TS optimization heavily depends on initial TS guesses, methods used in the optimization procedure, as well as software and hardware. Initial transition state guesses were optimized using a variety of methods, such as single and double-ended string methods (freezing and growing string),<sup>41-43</sup> and the nudged elastic band (NEB) method<sup>44,45</sup> implemented in a development version of the geomeTRIC optimization software<sup>46</sup> that calls the TeraChem software<sup>47,48</sup> for energy and gradient evaluations. The transition state optimizations were performed

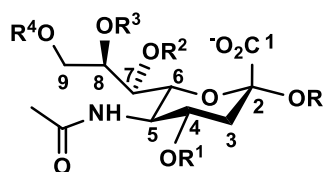
in Q-Chem<sup>49</sup> and TeraChem softwares.<sup>47,48</sup> A frequency calculation was used to confirm the presence of a single imaginary mode in the optimized transition state, and an intrinsic reaction coordinate (IRC) calculation was then performed to connect the optimized transition state with reactant to product structures.

I used this *ab initio* nanoreactor method across multiple projects (Chapters 2 and 3) and have presented this work at the following: West Coast Theoretical Chemistry Symposium in Stanford, CA in 2018; UC Davis 3rd Year Seminar in Davis, CA in 2019; and CALESS elevator pitch competition in 2021. The nanoreactor was pivotal in a synergistic theoretical and experimental study of a new ketone oxidation reaction that proceeds under mild conditions involving carbon–carbon bond cleavage to an alkyl hydroperoxide–carboxylic acid pair. The mechanism was derived using levulinic acid, an important biomass-derived platform molecule, then shown to be general for other ketones (Chapter 2). The *ab initio* nanoreactor played a key role in excluding intuitive mechanistic proposals and pointing instead to a radical O–O cleavage reaction that generates superoxide, a species whose presence in the reaction was ultimately confirmed using EPR spectroscopy. The initial analysis was performed on levulinic acid, an important biomass-derived platform molecule, but the method can be generalized to other ketones. This study exemplifies how combined theoretical and experimental effort can drive rational reaction design to give highly useful new products, in this particular case a precursor to biobased acrylic acid.

## **Sialic Acids in Viral Mechanisms**



Sialic acids (Sias) are part of a large family of over 50 derivatives of the 9-carbon sugar neuraminic acid and 2-keto-3-deoxy-nononic acid (Kdn)<sup>50,51</sup>. Among these forms, *N*-acetylneuraminic acid (Neu5Ac) is most common. Many disease-causing viruses target Sias, which commonly coat many cell surfaces, and is involved in many pathological and physiological processes.<sup>52-54</sup> Many disease-causing viruses target Sias with specific *O*-acetyl (OAc) modifications, where *O*-acetylation is the most common modification, and can be found at the C-4, 7, 8 and 9 positions (Figure 1.1)<sup>50,51</sup>.



**Figure 1.1.** *N*-Acetylneuraminic acid (Neu5Ac where  $R^1 = R^2 = R^3 = R^4 = H$ ). Its derived sialic acid forms in  $\alpha$ -linked sialosides. *R* may be modified to *O*- and *N*-acetyl.

*O*-acetyl groups are often found on gangliosides, sialylated glycosphingolipids found in all vertebrate cells and in the brain, where the GM3 ganglioside can inhibit tumor cell growth.<sup>55,56</sup> In another example, betacoronavirus OC43 and HKUI are known to bind to 9-*O*-acetylated Sias,<sup>57,58</sup> the most common Sia *O*-acetyl modification found in nature, and is among the three most abundant Sias found in mammals: 9-*O*-Acetyl-Neu5Ac (Neu5,9Ac<sub>2</sub>), Neu5Ac, and *N*-glycolylneuraminic acid (Neu5Gc).<sup>59</sup> Additionally, Neu5,9Ac<sub>2</sub> Sias are involved in regulating tissue morphogenesis,<sup>60</sup> binding of influenza C virus hemagglutinin and can mask the recognition of influenza A virus hemagglutinin,<sup>51</sup> and inhibiting tumor cell growth.<sup>51</sup> Sias are also involved in regulating immune responses to bacterial polysaccharides,<sup>61,62</sup> sialidase recognition, cellular apoptosis, and tumor

immunology.<sup>63-65</sup> The biological implications of *O*-acetylation in Sias is extensive, and have been reviewed across multiple sources.<sup>66-68</sup>

While it is important to understand the mechanisms of *O*Ac-Sia binding that cause the diseases, many are unknown due to experimental difficulties from *O*-acetyl ester instability. *O*-Acetylated Sias are sensitive to pH, temperature, and esterases.<sup>1,2,69,70</sup> The *O*-acetyl group in 9-*O*-acetylated sialoglycans on glycan microarrays is prone to cleavage.<sup>69</sup> Additionally, the *O*-acetyl group is prone to spontaneously migrate across the exocyclic glycerol-like Sia side-chain; *O*-acetyl on C-4, 7 and 8 can spontaneously migrate to C-9. This makes it very difficult to study glycan-protein binding using microarrays and characterization of *O*-acetylation on the other positions that exclude *O*-acetylation on C-9, as the expected *O*-acetyl group on the glycan may be cleaved or have migrated.<sup>59,71,72</sup> This spontaneous migration of *O*-acetyl group also influences many diseases that depend on Sias, such as infectious disease, immunology, oncology, cardiovascular disease, etc. For example, 9-*O*-acetylation is required for influenza C and D virus hemagglutinin binding, but prevents binding of influenza A and B virus hemagglutinins.<sup>73-77</sup> While 9-*O*-acetyled Sias block Siglec-9 binding, 7-*O*-acetylation Sias blocks binding to a greater extent.<sup>78</sup> The reaction mechanisms of *O*-acetyl dissociation and *O*-acetyl migration among the C-4, 7, 8, and 9 positions remained an open question until recently.<sup>1</sup>

To avoid spurious data resulting from migration and/or cleavage of the 9-*O*-acetyl group we considered replacing the *O*-acetyl group in these gangliosides by an *N*-Acetyl group. Chemically and biologically stable 9-acetamido-9-deoxy-*N*-acetylenuraminic acid (Neu5Ac9NAc)-containing sialoglycans have been proposed as stable mimics of Neu5,9Ac<sub>2</sub>-containing sialosides. Similarities

and differences of Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> were shown by glycan microarray studies, mammalian cell incorporation and surface expression studies, as well as sialidase cleavage activities<sup>2</sup>.

Two questions must be addressed when comparing the similarities between experimentally stable Neu5Ac9NAc and naturally occurring Neu5,9Ac<sub>2</sub>, in addition to the non-acetylated Neu5Ac sialic acid. First, how similar are the structures between modified sialic acid-containing glycans? We studied structural similarity computationally by performing conformational analysis and computing subsequent NMR <sup>3</sup>J couplings.<sup>2</sup> These <sup>3</sup>J couplings were compared against experiment, and confirmed conformational similarities between Neu5Ac9NAc and Neu5,9Ac<sub>2</sub>.<sup>2</sup> Secondly, how chemically similar in function are these glycans in biological conditions, such as in protein-ligand binding? To address this, a relevant system must be chosen. I chose a viral spike protein known to preferentially bind to Neu5Ac, with a crystal structure of Neu5Ac in the binding pocket, as well as a novel spike protein belonging to a virus that started the COVID-19 pandemic during my studies.

Beta coronavirus Middle East respiratory syndrome coronavirus (MERS-CoV) is known to co-bind to Neu5Ac in addition to its DPP4 primary receptor in a two-step binding mechanism, and depletion of Neu5Ac was found to inhibit MERS-CoV entry into human airway cells.<sup>79</sup> Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the beta coronavirus responsible for the COVID-19 pandemic, is structurally highly similar to SARS-CoV-1 (73% sequence identity between spike proteins), yet is remarkably more infective.<sup>58</sup> Both SARS-CoV-1 and SARS-CoV-2 share the same primary human cellular receptor angiotensin-converting enzyme-2 (ACE-2), but

this commonality in binding cannot explain the difference in infectivity. One possible mechanism for how SARS-CoV-2 achieves its high infectivity is by binding to Sias. Neu5Ac, Neu5,9Ac<sub>2</sub>, and Neu5Ac9NAc. Thus, I computed and analyzed the binding of Sias to MERS-CoV S and SARS-CoV-2 S proteins across multiple binding poses, to quantify the binding similarities across Sias, especially with Neu5,9Ac<sub>2</sub>, and Neu5Ac9NAc- containing glycans, elucidating the quality of the Neu5Ac9NAc analog in biological applications.

### **Computational Methods in Ligand-Protein Binding**

Accurate predictions of ligand-protein/protein-protein binding energies is an imperative. In the pharmaceutical industry, there is an urgent need to understand disease and viral mechanisms, to efficiently design drugs to improve lives of many people. Indeed, accurate predictions of ligand-protein/protein-protein binding energies, needed to describe drug-target interactions, are included in a list of holy grails of computational organic chemistry and biochemistry.<sup>15</sup> These involve two broadly classified computer-aided drug discover/design methods: ligand-based and structure-based.<sup>80</sup> Accurate predictions can greatly accelerate drug discovery by screening candidates against drug profile criteria, such as through structure-based approaches. This reduces synthetic targets, time, and cost. Improvements in computer hardware, methods, and algorithms have made computational approaches more accessible, with molecular dynamics (MD) simulations key in bridging electronic-structure and molecular mechanics methods.<sup>81</sup>

Modern free energy simulation methods are estimated to be accurate to within 1-2 kcal/mol for well-behaved<sup>82</sup> protein/ligand systems. Ligand-protein complexes can be modelled with explicit

water molecules, using classical force-fields in molecular dynamics and Monte Carlo algorithms, where currently the most accurate approaches are also the most rigorous.<sup>83</sup> SIRE-OpenMM molecular dynamics (SOMD)<sup>84,85</sup> is a rigorous alchemical simulations method. In alchemical ligand-protein binding free energy calculations, chemical groups are changed through unphysical states to estimate binding free energies using methods such thermodynamic integration (TI), Multistate Bennet acceptance ratio (MBAR), free energy perturbation (FEP), and others.<sup>86-89</sup> For example, Neu5,9Ac<sub>2</sub> was transformed to Neu5Ac9NAc by slowly turning off the interactions of Neu5,9Ac<sub>2</sub> and slowly turning them on Neu5Ac9NAc, over a set of discrete steps, as described the SI of Loeffler et al.<sup>90</sup> Based on a recent SAMPL6 challenge that evaluates binding free energy prediction methods,<sup>91</sup> we chose SOMD as the alchemical method to compute relative binding free energies of our modified Sia-containing glycans binding to the S proteins of MERS-CoV and SARS-CoV-2.

In SOMD, the unphysical states across the process of chemical modification is discretized into separate  $\lambda$  windows. We chose 13-21 linearly-spaced  $\lambda$  windows, depending on the transformation, based on a paper using SOMD and other alchemical free energy methods in an assessment of binding affinities,<sup>92</sup> and following the scheme outlined in the SI of Loeffler et al.<sup>90</sup> In my simulation cases, increasing  $\lambda$  to 21 did not immediately result in improved simulation results. Rather, a lower  $\lambda$  with multiple replicates would likely yield a more rigorous relative binding free energy value with an error estimate, and allow for discarding of simulation data without drastically reducing sampling size. However, multiple replicate alchemical simulations is indeed quite expensive and potentially time-prohibitive. The relative free energy of binding was computed as the free energy difference of the alchemical transformation of the ligand bound to the protein and the ligand in

solution, i.e. water, using the last 4 of 5 ns simulations.<sup>89,92</sup> Reliability of results was determined based on similarity between MBAR and TI, and from overlap matrices indicating sufficient thermodynamic overlap between simulation windows.

Herein describes studies of Neu5Ac-containing sialoglycans, Neu5,9Ac<sub>2</sub>-containing sialoglycans, and chemically and biologically stable Neu5Ac9NAc-containing sialoglycan counterparts. *O*-Acetyl instability and resultant migration across the glycerol-like side-chain was studied, with a full mechanism detailed in Chapter 3. Structural similarities between modified Sias have been quantified by NMR J-coupling values, known to quantify flexibility across glycosidic linkages,<sup>2</sup> and herein, biological similarities have been analyzed in ligand-protein binding studies in Chapter 4. Additionally, I have presented this work at the following: 22<sup>nd</sup> R. Bryan Miller Symposium in Davis, CA in 2022 and ACS Spring 2022 Meeting in San Diego, CA in 2022. As a culmination of these studies, we find that Neu5Ac9NAc-containing sialoglycans are structurally similar to Neu5,9Ac<sub>2</sub>-containing sialoglycans when studying sialoglycan conformations, and similar in function in Sias-protein binding with MERS-CoV S and SARS-CoV-2 S proteins.

## **Chapter 2: Application of the *ab initio* Nanoreactor to the Development of a New Peroxide-Driven Ketone Oxidation Reaction**

The work in this chapter is based on “Application of the *ab initio* Nanoreactor to the Development of a New Peroxide-Driven Ketone Oxidation Reaction” with Andrew L. Otsuki, Aidin R. Balo, Mark Mascall, and Lee-Ping Wang, submitted to *Angewandte Chemie*.

## Abstract

The development of new, synthetically useful reactions in organic chemistry can be effectively supported by detailed mechanistic insights from quantum chemistry and molecular dynamics simulations. Here we describe a synergistic theoretical and experimental study of a new ketone oxidation reaction that proceeds under mild conditions involving carbon–carbon bond cleavage to an alkyl hydroperoxide–carboxylic acid pair. The mechanism was derived using levulinic acid, an important biomass-derived platform molecule, then shown to be general for other ketones. The *ab initio* nanoreactor, a recently described simulation method for predicting new reaction pathways, played a key role in excluding intuitive mechanistic proposals and pointing instead to a radical cleavage involving superoxide, a species whose presence was confirmed using EPR spectroscopy. This study exemplifies how combining the forces of advanced theory and experiment can drive rational reaction design to give highly useful new products.

## Introduction

The development of environmentally friendly synthetic methodologies is a central and urgent goal of modern chemical science and industry.<sup>3–6</sup> Major thrusts of this goal involve replacing petrochemical feedstocks with renewable resources and closer adherence to the principles of green chemistry.<sup>7–9</sup> Advancing mechanistic understanding is crucial to accelerating the development of new reactions involving biobased platform molecules which can be applied to sustainable process design.<sup>11</sup> For this reason, the pursuit of mechanistic knowledge in organic synthesis, using a

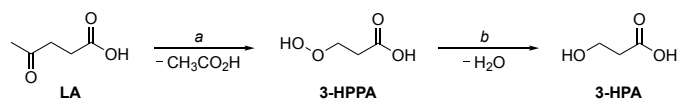


synergistic combination of experimental and theoretical efforts, is expected to play an important role in bringing about the sustainable, greener chemical economy of the near future.

Experimental knowledge constrains the space of mechanistic possibilities and serves as the ultimate proving ground of all hypotheses, whereas theoretical modeling is valuable for testing hypotheses by providing evidence for intermediates beyond experimental detection limits. For most of the history of theoretical chemistry, mechanistic deduction has largely followed this paradigm, originating from combinations of empirical data, precedent and, more peripherally, chemical intuition. On the other hand, theoretical simulations are increasingly being viewed as a source of *de novo* mechanistic hypotheses to complement and expand mechanistic knowledge. Computational approaches for automated discovery of reaction pathways, which are a major conceptual advance from traditional hypothesis-based approaches, have recently been reported by several groups including ours.<sup>16–25</sup> The strategies employed for automated exploration of reaction space are highly diverse, including the use of semiempirical or fully *ab initio* molecular dynamics (MD) simulations, which can predict mechanisms independent of pre-existing information about the reaction or heuristic reactivity rules.<sup>29</sup> We previously described the development of the *ab initio* nanoreactor approach to reaction discovery, in which high temperatures, high pressures, and external forces are applied to assemblies of reactants in first-principles MD simulations.<sup>17,19</sup> For example, such computations have revealed possible new pathways for amino acid synthesis from small inorganic compounds proposed to exist on the early Earth, and served as a proof of concept for dynamics-based computational exploration of chemical reactivity space.<sup>17</sup>

Herein, we show how the nanoreactor was used, in conjunction with experimental validation, to solve a mechanistic problem of strong interest due to the unprecedented nature of the reaction and its relationship to sustainable chemistry. We describe a systematic exploration of key steps in the production of 3-hydroperoxypropanoic acid (3-HPPA) from biobased levulinic acid (LA) by oxidation under alkaline conditions (Scheme 2.1). This reaction was first reported by Mascall *et al.* in a communication on the production of 3-hydroxypropanoic acid (3-HPA), itself a monomer of significant interest, but more importantly, a precursor (by dehydration) to acrylic acid, a mainstream commodity chemical.<sup>93</sup> Preliminary mechanistic proposals were advanced to account for the production of 3-HPPA from LA based on intuitive pathways. Subsequent attempts to model potential energy surfaces for these hypotheses and a range of alternatives using *ab initio* methods led however to unrealistic transition state (TS) energies. We therefore turned to the *ab initio* nanoreactor in an attempt to refocus the inquiry which, together with experimental confirmation of nanoreactor predictions, ultimately returned a supportable hypothesis.

Looking ahead, the reaction mechanism will be shown to feature O–O and C–C bond homolysis with an important driving force provided by the thermodynamic stability of superoxide ( $O_2^{\bullet-}$ ), whose presence in the reaction mixture was verified using electron paramagnetic resonance (EPR) spectroscopy. The generality of the proposed mechanism was further supported by experiments subjecting other ketones to the same reaction conditions, thus establishing a new class of oxidation reactions for organic synthesis.



**Scheme 2.1.** Transformation of LA into 3-HPA via 3-HPPA. Conditions: *a)* 30% H<sub>2</sub>O<sub>2</sub>, KOH, 0 °C to rt, 6 h; *b)* H<sub>2</sub>, Pd/C, MeOH, 2 h.<sup>19</sup>

This paper is organized in chronological order of the study to highlight the interplay between theory and experiment. We first describe the experiment and the initial mechanistic guesswork, followed by the development of new insights gained from nanoreactor simulations, and ultimate experimental confirmation of the theoretical predictions.

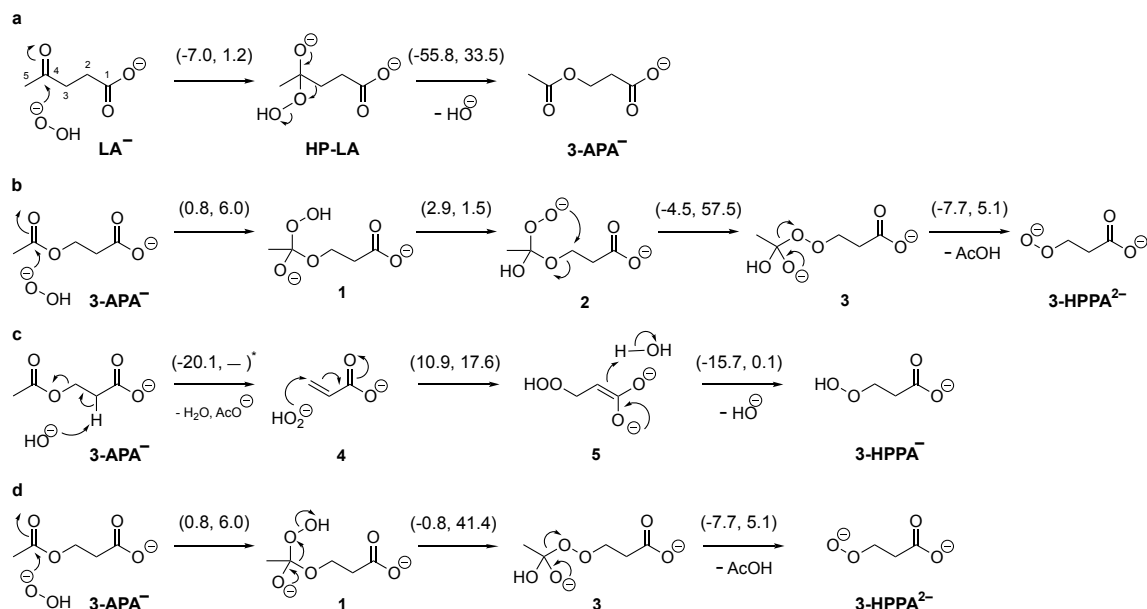
## Results and Discussion

### Initial mechanistic hypotheses based on 3-APA as an intermediate

It was previously shown that the addition of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) to a solution of LA under acidic conditions resulted in a standard Baeyer-Villiger oxidation involving methyl group migration where succinic acid is the major product.<sup>94</sup> However, as shown in Scheme 2.1, conducting the same reaction under basic conditions led to 3-HPPA as the major product in high yield.<sup>93</sup> Initially, it seemed reasonable to attribute the selectivity to a pH-switchable Baeyer-Villiger rearrangement that leads to either succinic acid or 3-acetoxypropanoic acid (3-APA) in acidic or basic conditions, respectively. The first step would involve nucleophilic addition of hydroperoxide anion (HO<sub>2</sub><sup>-</sup>) at C<sub>4</sub> of the levulinate anion to give HP-LA, followed by migration of C<sub>3</sub> and displacement of hydroxide to first give 3-APA (Scheme 2.2a).

Optimized structures, reaction energies ( $\Delta E$ ) and activation energies ( $E_a$ ) with harmonic free energy corrections were computed using the  $\omega$ B97X-V/def2-TZVPD level of theory and SWIG-PCM polarizable continuum model for water as described in the Methods section. Scheme 2a shows that the addition of  $\text{HO}_2^-$  to LA is predicted to proceed rapidly whereas the  $\text{C}_3$  shift to oxygen and synchronous displacement of hydroxide has an activation free energy  $\Delta G^\ddagger = 33.5$  kcal/mol. Although this was somewhat high for a reaction that takes place at 0 °C, 3-APA was considered the most obvious intermediate *en route* to the 3-HPPA reaction product.

Scheme 2.2b-d shows three chemically intuitive mechanistic proposals for the conversion of 3-APA to 3-HPPA, two of which (Scheme 2.2b and Scheme 2.2c) were proposed (without support) in the original communication.<sup>93</sup> Scheme 2.2b involves addition of  $\text{HO}_2^-$  to 3-APA followed by a proton transfer and rearrangement, cleaving to form 3-HPPA and acetic acid. However,  $\Delta G^\ddagger$  was calculated to be 57.5 kcal/mol for the rearrangement step, which was clearly prohibitive. Scheme 2.2c proposes an elimination-addition process, suggesting that acrylate could serve as a secondary intermediate. The elimination step appeared to be accessible at the B3LYP/6-31G\* level of theory ( $E_a < 10$  kcal/mol) but, despite efforts, TS structures could not be found using higher-level methods, so we were compelled to abandon this approach. Scheme 2.2d proposes a novel alkoxy group migration related to the Baeyer-Villiger reaction (second step); while the  $\Delta G^\ddagger$  of 41.4 kcal/mol is somewhat lower than the rearrangement in Scheme 2.2b, it was also considered an unlikely alternative. The inconclusiveness of these preliminary computational results was the impetus for carrying out *ab initio* nanoreactor simulations, which ultimately resulted in significant changes to our hypotheses.



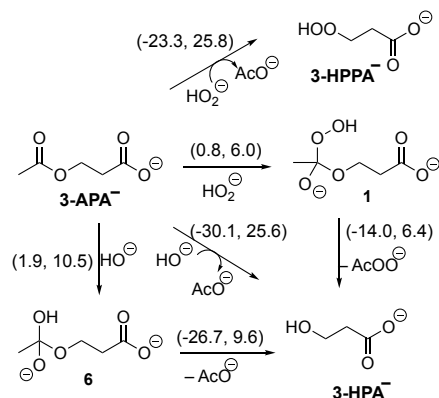
**Scheme 2.2.** Initial mechanistic proposals for oxidation of levulinate anion ( $\text{LA}^-$ ) to form 3-APA followed by transformation to 3-HPPA. **a.** Baeyer-Villiger oxidation of  $\text{LA}^-$  under basic conditions to give 3-APA via hydroperoxy adduct HP-LA. **b.** Cyclo-rearrangement pathway from 3-APA to 3-HPPA. **c.** Elimination-addition pathway from 3-APA to 3-HPPA. **d.** Alkoxy migration pathway from 3-APA to 3-HPPA. ( $\Delta G$ ,  $\Delta G^\ddagger$ ) values are in kcal/mol and calculated with  $\omega\text{B97X-V/def2-TZVPD/SWIG-PCM}$  level of theory, electronic energies given in the Supporting Information Figures 2.2-2.11. \*The TS optimization in the first step of **c** failed to converge.

### Nanoreactor simulations involving 3-APA as an intermediate

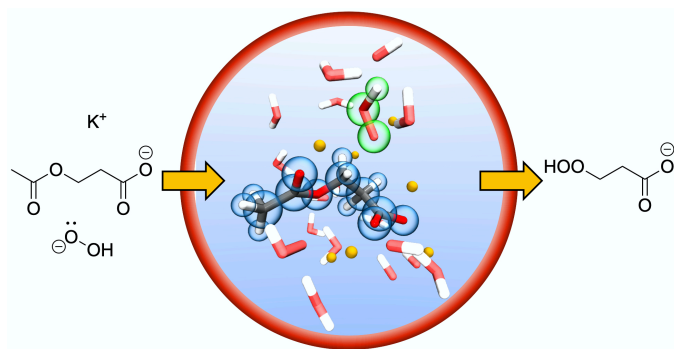
A total of 60 nanoreactor simulations were initially carried out, and reactions were observed over 270 ps of cumulative simulation time. The setup of a typical simulation is described in the Methods section. The simulated molecules were chosen to mirror the reaction conditions,

assuming that 3-APA was an intermediate in the production of 3-HPPA: 1 deprotonated 3-APA, 3 HO<sup>-</sup>, 3 HO<sub>2</sub><sup>-</sup>, 12 H<sub>2</sub>O, and 7 K<sup>+</sup> (or Na<sup>+</sup> in some runs). Among the reaction pathways observed in the nanoreactor, two involved S<sub>N</sub>2 substitution of acetate by HO<sup>-</sup> and HO<sub>2</sub><sup>-</sup> to form 3-HPA<sup>-</sup> and 3-HPPA<sup>-</sup>, respectively, with comparable activation free energies (Scheme 2.3) that were lower than in the mechanistic proposals of Scheme 2.2. Figure 2.1 is a 3-D rendering of one such substitution observed in the nanoreactor. However, among the other reaction pathways that the nanoreactor found, hydroperoxolysis of 3-APA via **1** led to 3-HPA with a much lower barrier than either substitution reaction. A hydrolysis reaction was also found via **6**, leading to 3-HPA with higher activation free energy than hydroperoxolysis, but still lower than alkyl substitution. These results cast doubt on the feasibility of 3-APA as an intermediate in the LA → 3-HPPA reaction.

In an attempt to corroborate this theoretical insight, we conducted follow-up experiments where 3-APA was subjected to the original reaction conditions that convert LA to 3-HPPA.<sup>93,94</sup> Multiple trials were conducted with variations of reagent addition. The only product observed was 3-HPA, as determined by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. These experiments supported the nanoreactor prediction that 3-APA reacts with hydroperoxide or hydroxide to form 3-HPA, and not 3-HPPA. 3-APA could therefore be ruled out as an intermediate.



**Scheme 2.3.** Reaction pathways found in the nanoreactor starting from the putative 3-APA intermediate, described in clockwise order as: nucleophilic substitution to form 3-HPPA<sup>-</sup>, hydroperoxolysis via **1** to form 3-HPA<sup>-</sup>, nucleophilic substitution to form 3-HPA<sup>-</sup>, and hydrolysis via **6** to form 3-HPA<sup>-</sup>. ( $\Delta G$ ,  $\Delta G^\ddagger$ ) values are in kcal/mol and calculated with  $\omega$ B97X-V/def2-TZVPD/SWIG-PCM level of theory, and electronic energies **given** in the Supporting Information Figure 4,12-16. Note path from **3-APA** to **1** is the same as Scheme 2.2b, step 1.



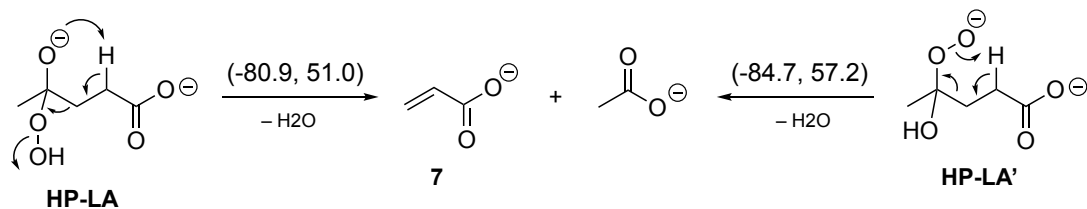
**Figure 2.1.** 3-D rendering of a trajectory frame of the nanoreactor simulations, showing the nucleophilic attack of HO<sub>2</sub><sup>-</sup> (highlighted green) on 3-APA (highlighted blue). Other species present are water molecules, K<sup>+</sup>, and HO<sup>-</sup>. C, gray; O, red; H, white; K, gold.

## Nanoreactor predictions involving acrylate as an intermediate

We thus initiated a search for alternative pathways to produce 3-HPPA that did not involve 3-APA. Nanoreactor simulations starting directly from LA predicted a high-energy C<sub>3</sub>–C<sub>4</sub> fragmentation of LA that resulted in acetaldehyde and acrylate. Also frequently observed was the nucleophilic addition of HO<sub>2</sub><sup>-</sup> at C<sub>4</sub> to form HP-LA, consistent with Scheme 2.2a. Although the elimination-addition hypothesis tested in Scheme 2.2c was problematic, these new results suggested an alternative path from HP-LA to acrylate. Acrylate easily admits conjugate addition of HO<sup>-</sup> or HO<sub>2</sub><sup>-</sup> at C<sub>3</sub> to give 3-HPA and 3-HPPA, respectively (Scheme 2.2c and Supporting Information Figure 2.24-2.25). These latter steps are overall exothermic, and HO<sub>2</sub><sup>-</sup> addition is preferred energetically to HO<sup>-</sup> addition at C<sub>4</sub>. Further experimental studies also confirmed that 3-HPPA is the major product when acrylate is subjected to LA reaction conditions (Supporting Information Experimental Methods).

Although experiments now pointed to acrylate as a plausible intermediate, the modeled activation barriers based on proposed mechanisms in Scheme 2.4 were too high. An exhaustive search was carried out by calculating the minimum energy paths of O–O and C–C cleavage starting from various protonation states of HP-LA (Supporting Information Figure 2.26-2.30). The lowest activation free energies found are shown in Scheme 2.4, but both are in excess of 50 kcal/mol and deemed prohibitive. Following this setback, we expanded the search to open-shell models. Theoretical and experimental studies involving a radical pathway are described in the next section.





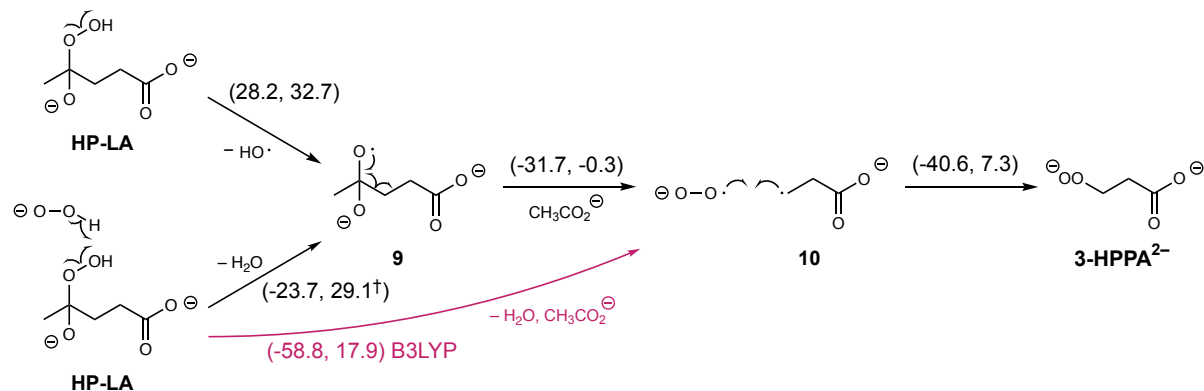
**Scheme 2.4.** Possible reaction pathways to acrylate starting from 4-hydroperoxy-levulinate adduct HP-LA in different protonation states. ( $\Delta G$ ,  $\Delta G^\ddagger$ ) values are in kcal/mol and calculated with  $\omega$ B97X-V/def2-TZVPD/SWIG-PCM level of theory, and electronic energies given in the Supporting Information Figure 17-18.

### A role for radicals

Studies involving the role of radicals in the decomposition of  $\text{H}_2\text{O}_2$  in alkaline solution date back several decades. A number of mechanisms have been advanced, and various experimental conditions have been shown to influence reaction rates, such as the presence of metal catalysts, temperature, or surface modification of the reaction vessel.<sup>95-98</sup> One proposed mechanism for the reaction  $2 \text{H}_2\text{O}_2 \rightarrow 2 \text{H}_2\text{O} + \text{O}_2$  contains a rate-determining step where hydroperoxyl radical ( $\text{HO}_2^\bullet$ ) and hydroxyl radical ( $\text{HO}^\bullet$ ) intermediates are formed,<sup>96</sup> with an optimal rate at pH 12 regardless of the presence or absence of metals.<sup>96,97</sup> A separate study of  $\text{H}_2\text{O}_2$  decomposition on an iron oxide surface invokes the following equilibrium:  $\text{HO}_2^\bullet \rightleftharpoons \text{O}_2^{\bullet-} + \text{H}^+$ ,<sup>98,99</sup> which favors  $\text{O}_2^{\bullet-}$  at high pH ( $\text{pK}_a$  of  $\text{HO}_2^\bullet = 4.8$ ).<sup>98,99</sup> Aqueous solutions of KOH and  $\text{H}_2\text{O}_2$  have been shown experimentally to generate  $\text{O}_2^{\bullet-}$  *in situ*,<sup>100</sup> providing further indication that radicals may play a part in the observed reactivity. Given the literature precedent, we investigated whether radicals were involved in the O–O plus C<sub>3</sub>–C<sub>4</sub> fragmentation pathway predicted by the nanoreactor.

As a first step, the reaction was tested for the presence of radicals. The following aqueous solutions were analyzed using X-band EPR spectroscopy with 0.2 mW of microwave power and a modulation amplitude of 1.0 G: (1) H<sub>2</sub>O<sub>2</sub>, (2) H<sub>2</sub>O<sub>2</sub> + LA, (3) KOH + LA, (4) H<sub>2</sub>O<sub>2</sub> + KOH, and (5) H<sub>2</sub>O<sub>2</sub> + KOH + LA (Supporting Information Experimental Methods). The first three solutions were found to be EPR silent, as expected. The EPR spectra measured for (4) and (5) were consistent with the presence of O<sub>2</sub><sup>-•</sup> (Supporting Figure 2.1), with axial EPR signals and g values matching reported values of O<sub>2</sub><sup>-•</sup>.<sup>101</sup> We observed a small shift in the g<sub>||</sub> value between the EPR spectra of the control (4) and reaction mixture (5), but could not assign this minor shift to the presence of any additional species in (5). We thus concluded that O<sub>2</sub><sup>-•</sup> was a likely participant in the reaction pathway.

Our computational study of radical pathways started by reoptimizing the fragmentation in Scheme 2.4 using an unrestricted wavefunction, leading to significantly altered chemistry (Scheme 2.5 upper pathway). O–O cleavage proceeded via homolysis, leading to HO• and a radical dianion that undergoes homolytic C–C cleavage to yield 3-propanoate radical anion and acetate. If O<sub>2</sub><sup>-•</sup> is present, radical coupling could generate 3-HPPA. While this bond dissociation has a significantly lower ΔG<sup>‡</sup> (32.7 kcal/mol) compared to those in Scheme 2.4, it has the undesirable feature where HO• is generated while O<sub>2</sub><sup>-•</sup> is consumed. EPR experiments did not show evidence for HO•, and since it is known to react rapidly and indiscriminately with organic species, its generation could preclude the selective formation of the 3-HPPA product.



**Scheme 2.5. Radical pathways from HP-LA to form 3-HPPA.** Upper left: Reaction initiated by homolytic cleavage of O–O bond in HP-LA. Lower left:  $\text{HO}_2^-$  anion is added serving as a hydrogen atom donor. ( $\Delta G$ ,  $\Delta G^\ddagger$ ) values are in kcal/mol and calculated with  $\omega\text{B97X-V/def2-TZVPD/SWIG-PCM}$  level of theory, and electronic energies given in the Supporting Information Figure 2.19-2.23. The pathway in purple was computed using the B3LYP functional as an approximate lower bound on the barrier height. †Computed using CDFT-CI single point calculations at the minimum energy crossing point.

A mechanism therefore had to be formulated where the involvement of discrete  $\text{HO}^\bullet$  was excluded.  $\text{HO}_2^-$  was added to the calculation, which can donate a hydrogen atom in a concerted fashion with O–O homolysis to yield water and  $\text{O}_2^\bullet$  (Scheme 2.5 lower pathway). Although this reaction still involves the generation of two radical species, it was considered that the relative thermodynamic stability of  $\text{O}_2^\bullet$  and its increased O–O bond order relative to  $\text{HO}_2^-$  could reduce the reaction energy and activation barrier. This reaction was unusual in that we could not find an optimized TS at our chosen level of theory ( $\omega\text{B97X-V/def2-TZVPD}$ ), because the ground state along the reaction path includes a discontinuous crossing from a closed-shell potential surface to a broken-symmetry

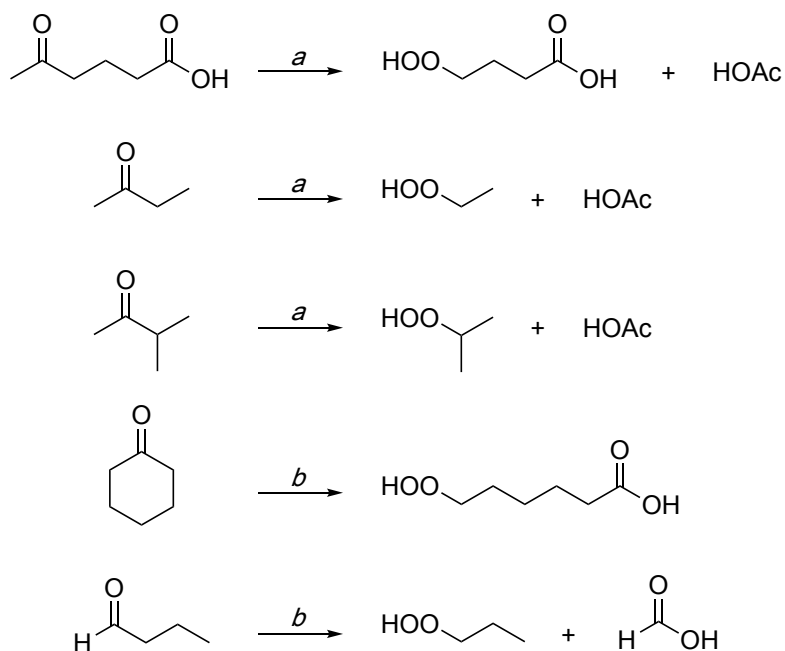
diradical surface. We thus optimized the minimum energy crossing point of the closed-shell and diradical states, and used constrained DFT-configuration interaction (CDFT-CI)<sup>102-105</sup> calculations to estimate the barrier taking multireference character into account. The CDFT-CI ground state at the crossing point geometry is a nearly equal mixture of the closed shell and broken-symmetry states.

Because it was not possible to optimize the TS in CDFT-CI, the CDFT-CI single-point energy difference of 29.1 kcal/mol between the crossing point and ground state was taken as an upper bound to the activation energy. To provide a lower bound, we looked for a TS at the B3LYP-D3(BJ)/def2-TZVPD level of theory. Unlike the case with  $\omega$ B97X-V, the B3LYP-D3 ground electronic state smoothly progresses from closed-shell to broken-symmetry character, allowing for the TS to be optimized (purple path in Scheme 2.5). Given the tendency for global hybrid functionals such as B3LYP to systematically underestimate barrier heights,<sup>106</sup> we take the computed activation energy of 21.4 kcal/mol as a lower bound, allowing us to estimate the TS energy as being within the 21.4–29.1 kcal/mol range, *i.e.* within the reach of the reaction conditions. Scheme 2.5 (lower pathway) thus provides a complete reaction path from LA to 3-HPPA where radical intermediates play a key role, as one equivalent of  $O_2^{\bullet}$  radical is produced and consumed. We also note that  $O_2^{\bullet}$  is produced independently of LA under the reaction conditions, which could increase the rate of the radical recombination of  $O_2^{\bullet}$  and **10** to form 3-HPPA<sup>2</sup>.

We tested the generality of this reaction, termed peroxide-induced radical ketone cleavage (PIRC), with similar substrates under the same reaction conditions (Scheme 2.6 and Supporting

Information Experimental Methods). 5-oxohexanoic acid, a homolog of LA, reacted to give 4-hydroperoxybutanoic acid, consistent with PIRC. Methyl ethyl ketone and methyl isopropyl ketone produced hydroperoxides from cleavage to more stable radical intermediates (ethyl and isopropyl, respectively). These results suggest that PIRC may be applied generally to the production of alkyl hydroperoxides from ketones.

Finally, since it is known that transition metal catalysts can significantly enhance rates in peroxide-based oxidations, we repeated the LA reaction in the presence of iron(III) acetate (Supporting Information Experimental Methods). The process immediately evolved gas more vigorously than in the absence of catalyst. The reaction rate was greater and the yield of 3-HPPA remained high, with trace succinic acid by-product and no 3-HPA. Likewise, cyclohexanone in the presence of iron(III) cleaved to 6-hydroperoxybutanoic acid (Scheme 2.6). An interesting aspect of PIRC is the reversal of selectivity compared to the analogous Baeyer-Villiger oxidation. Selectivity was reversed in LA and butanal (Scheme 2.6), which gives the C–C cleavage product, compared to the Baeyer-Villiger H-migration to give the corresponding acid. The role of oxygen-centered radicals in iron-catalyzed reactions involving peroxide has been reviewed,<sup>107–109</sup> and this catalytic mechanism is a promising avenue for further experimental and theoretical study.



**Scheme 2.6.** Oxidation of other ketones with basic hydrogen peroxide. Reagents and conditions, a. 30% H<sub>2</sub>O<sub>2</sub>, NaOH, 0 °C to rt, 6 h ; b. 30% H<sub>2</sub>O<sub>2</sub>, NaOH, FeCl<sub>3</sub>, 0 °C to rt, 1 h.

## Conclusion

The peroxide-induced radical ketone cleavage (PIRC) mechanism that is presented at the conclusion of this study is a result of the active development of insights using a synergistic combination of theory, dynamic simulation, synthetic organic chemistry, and EPR spectroscopy. The *ab initio* nanoreactor simulations played a central role in this work by suggesting new mechanistic hypotheses that directed the course of the scientific inquiry. This study is an example of the increasing integration of theory and experiment in chemistry research, standing in contrast to the more classic paradigm in which theory follows experiment by providing data to support experimentally derived mechanistic hypotheses.

The development of new synthetic methods for the conversion of biobased feedstocks into commercially useful chemicals can thus be supported by using computational methods to discover new modes of reactivity. In this work, we have described a method by which ketones, among the most common of functional motifs in organic chemistry, can be oxidatively cleaved to carboxylic acids and alkyl hydroperoxides, the latter being useful oxidants for a variety of applications.<sup>110-112</sup> As noted, the selectivity of this reaction is reversed compared to Baeyer-Villiger oxidation, since the direction of bond cleavage is determined by radical stability rather than migratory aptitude. We also highlighted the usefulness of *ab initio* nanoreactor simulations to predict reaction mechanisms, both in terms of excluding nonproductive lines of inquiry and providing insights that, in our case, ultimately led to a reasonable mechanistic hypothesis for a previously undescribed reaction. We are optimistic that this account of theoretical and experimental synergy will encourage innovative applications of the nanoreactor to the further discovery of new chemistries.

## **Computational Methods**

The computational studies are divided into two categories: nanoreactor simulations and reaction path refinement. The nanoreactor is a specialized *ab initio* molecular dynamics (AIMD) environment for discovering mechanistic hypotheses, whereas reaction path refinement refers to calculations that start from a given mechanistic hypothesis and aim to find the minimum energy path and estimate its thermodynamic and kinetic properties (*i.e.* reaction energy and activation barriers). Selected nanoreactor outputs and manually generated mechanistic guesses were inputs for in-depth reaction path refinement.

Nanoreactor simulations were carried out using the B3LYP density functional approximation with a DFT-D3 dispersion correction, 3-21G Gaussian basis set, and SWIG-PCM polarizable continuum model of the aqueous reaction medium ( $\epsilon=78.4$ ), denoted as B3LYP-D3/3-21G/PCM. Simulations utilized the TeraChem quantum chemistry software<sup>31-34</sup> and were run on servers with an Intel Xeon CPU / Nvidia GPU (either E5-2630 v3 CPU / GTX 980 Ti GPU or E5-2637 v4 CPU / GTX 1080 Ti GPU). The number of atoms in a typical nanoreactor system is on the order of 50-200, including reactants, solvent, and ions. Due to the high cost of AIMD, these simulations are limited to a few tens of picoseconds, and special simulation conditions are needed to broadly explore system reactivity within this short time frame.

Initial coordinates were generated by randomly placing reactants into a sphere using Packmol software.<sup>30</sup> Atomic coordinates were energy-minimized prior to starting the nanoreactor. Simulations used a time step of 0.5–1.0 fs with a Langevin thermostat set to 1000–3000 K, and a friction coefficient of 3.3–6.7 ps<sup>-1</sup>. The increased simulation temperature provides sufficient kinetic energy to rapidly dissociate intermolecular interactions without spontaneously breaking covalent bonds on its own. The reaction events are induced by a time-dependent restraint potential in equation (1), where  $m$  is the atomic mass,  $R_1$  and  $R_2$  are radius parameters, and  $k_1$  and  $k_2$  are force constants. The potential

$$V(r,t) = \frac{m}{2}k(t)\rho(t)^2; \quad \rho = \begin{cases} r - R(t) & \text{if } r > R(t) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$k(t), R(t) = \begin{cases} k_1, R_1 & \text{if } t \bmod (t_1 + t_2) < t_1 \\ k_2, R_2 & \text{otherwise} \end{cases}$$



oscillates in a square wave pattern with high and low periods  $t_1$  and  $t_2$ , where molecules are free to move in a sphere with radius  $R_1$  and  $R_2$ , respectively. When the radius decreases from  $R_1$  to  $R_2$ , atoms beyond  $R_1$  are accelerated inward, with the mass-dependent force ensuring near-uniform acceleration of atoms within a molecule. Reaction events occur following high velocity collisions of reactants (see Supporting Information Computational Methods for detailed parameter choice rationale and Table 1 for parameter selection).

Each reaction event, consisting of a subset of atom indices and trajectory frames, was automatically identified and extracted from the nanoreactor simulation for reaction path refinement to find the minimum energy path. Individual calculations include nudged elastic band (NEB),<sup>113,114</sup> growing string (GS),<sup>42,43</sup> and freezing string (FS)<sup>41</sup> for reaction path optimization; TS optimization; vibrational analysis to confirm the TS structure has only one imaginary vibrational mode; and intrinsic reaction coordinate (IRC) calculations<sup>115</sup> to confirm that following the imaginary mode "downhill" from the TS structure leads to the expected reactant and product.

Refinement calculations were carried out using the  $\omega$ B97X-V density functional approximation,<sup>116</sup> def2-TZVPD Gaussian basis set,<sup>117</sup> and SWIG-PCM implicit solvent model<sup>118</sup> with the dielectric constant set for water ( $\epsilon=78.4$ ), denoted as  $\omega$ B97X-V/def2-TZVPD/PCM, as implemented in the Q-Chem 5.0 quantum chemistry package.<sup>119</sup> Comprehensive benchmark studies showed that the  $\omega$ B97X-V/def2-QZVP level of theory gives mean absolute errors of 2.6 and 1.9 kcal/mol for reaction energies and barrier heights, respectively, relative to gold standard calculations.<sup>106</sup> Basis set incompleteness effects are assumed to be minor as studies of conformational relative energies show that the def2-TZVPD vs def2-QZVP basis sets agree to within 0.2 kcal/mol.<sup>120,121</sup> Therefore,

the estimated range of error in computed energetics in this study is  $\pm 2$ -3 kcal/mol. A development version of the geomeTRIC optimization software<sup>122</sup> was used to perform NEB calculations and optimizations of minimum energy crossing points between singlet and diradical electronic states,<sup>123</sup> which were calculated using a two-stage procedure in Q-Chem: constrained DFT<sup>102,103</sup> was used to generate the initial guess, followed by full orbital relaxation using the orbital tracking / maximum overlap method<sup>124</sup> during the self-consistent field (SCF) iterations. Free energies were estimated using the harmonic approximation based on results of vibrational analysis.

### **Acknowledgements**

LO and LPW acknowledge support from NIH/NIAID R01 AI130684-02. ARB acknowledges a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and support from NIH/NIGMS R35 GM126961-01. AO and MM acknowledge support from XSEDE TG-CHE150052.

**Chapter 3: *O*-Acetyl migration within the sialic acid side chain: a mechanistic study by the *ab initio* nanoreactor**

The work in this chapter is based on “*O*-Acetyl migration within the sialic acid side chain: a mechanistic study by the *ab initio* nanoreactor” with Yang Ji, Wanqing Li, Ajit Varki, Xi Chen, and Lee-Ping Wang, submitted to *Biochemistry*.

## Abstract

Many disease-causing viruses target sialic acids on the surface of host cells. Some viruses bind preferentially to sialic acids with *O*-acetyl modification at the hydroxyl group of C7, C8, or C9 on the glycerol-like side chain. Binding studies of proteins to sialosides containing *O*-acetylated sialic acids are crucial in understanding the related diseases, but experimentally difficult due to the lability of the ester group. We recently showed that *O*-acetyl migration among hydroxyl groups of C7, C8, and C9 in sialic acids occurs in all directions in a pH-dependent manner. In the current study, we elucidate a full mechanistic pathway for the migration of *O*-acetyl between C7, C8, and C9. We used the *ab initio* nanoreactor to explore potential reaction pathways, as well as density functional theory,  $pK_a$  calculations, and umbrella sampling to investigate elementary steps of interest. We found that when a base is present, migration is easy in any direction and involves three key steps: deprotonation of the hydroxyl group, cyclization between the two carbons, and the migration of the *O*-acetyl group. This dynamic equilibrium may play a defensive role against pathogens that evolve to gain entry to the cell by binding selectively to one acetylation state.

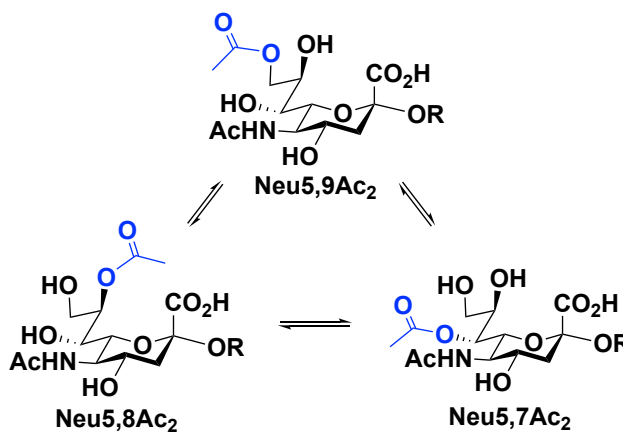
## Background

The sialic acid family includes more than 50 forms that are derived from the 9-carbon backbone monosaccharides neuraminic acid and 2-keto-3-deoxy-nononic acid (Kdn). Among these forms, *N*-acetylneuraminic acid (Neu5Ac) is most common.<sup>68</sup> Many disease-causing viruses target sialic acids (Sias) with specific *O*-acetyl (OAc) modifications. In one example, betacoronavirus OC43 and HKUI are known to bind to 9-*O*-acetylated Sias,<sup>57,58</sup> the most common Sia *O*-acetyl

modification found in nature. In our recent article, we noted the preferential binding of SARS-CoV-2 S protein to an immobilized acetyl group at the C9-OH by using an *N*-acetyl group analog Neu5Ac9NAc which is resistant to migration over its *O*-acetyl counterpart Neu5,9Ac<sub>2</sub>.<sup>125</sup> More recently, multivalent 9-*O*-acetylated-sialic acid glycoclusters were reported as potent inhibitors for SARS-CoV-2 infection.<sup>126</sup> While it is important to understand the mechanisms of OAc-Sia binding that cause the diseases, many are unknown due to experimental difficulties from *O*-acetyl ester instability. *O*-Acetylated Sias are sensitive to pH, temperature, and esterases.<sup>1,2,69</sup> Additionally, the *O*-acetyl group is prone to spontaneously migrate across the exocyclic glycerol-like Sia side-chain.

This makes it very difficult to study glycan-protein binding using microarrays, as the expected *O*-acetyl group on the glycan may be cleaved or have migrated.

In our recent study, we showed a pH-dependent Sia *O*-acetyl migration in all directions among the C9, C8, and C7 positions, to form Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub>, and Neu5,7Ac<sub>2</sub>, respectively (Fig. 3.1).<sup>1</sup>



**Figure 3.1.** *O*-acetyl migration is reversible between Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub>, and Neu5,7Ac<sub>2</sub>.

In this study, we probe the molecular details of this migration using density functional theory (DFT). There are many ways to computationally investigate mechanisms that involve finding critical points on the molecular potential energy surface, namely energy-minimized reactant, product, and intuitively-driven intermediate structures, and the transition states connecting them via minimum energy pathways. While intuition-guided approaches to optimize transition states

may connect the target reactant/product combination, success is highly dependent on user input and intuition. To overcome this, several approaches have been developed for automated reaction discovery.<sup>16–25</sup> Some methods include the use of semiempirical or fully *ab initio* molecular dynamics (MD) simulations, which can predict mechanisms independent of pre-existing information about the reaction or heuristic reactivity rules.<sup>29</sup> The nanoreactor is an *ab initio* molecular dynamics (AIMD)-based method specialized to induce a large number of reactions in a small amount of simulation time by increasing reactivity through elevated temperature and spherically symmetric compressions that oscillate in time.<sup>17,19,127</sup> The nanoreactor trajectories contain many reaction events that are automatically refined to individual mechanistic steps, where a single transition state connects each reactant/product structure pair.

We used the nanoreactor to generate mechanistic guesses, followed by a combination of automated and manual energy refinement to obtain minimum energy paths of *O*-acetyl group migration.<sup>17,19</sup> We then tested the feasibility of our proposed mechanism under experimental conditions by computing  $pK_a$  values of ionizable groups on the Sia glycerol chain, and free energy profiles of key bond dissociation steps using umbrella sampling simulations. We propose that an initial deprotonation step enables the *O*-acetyl group to form a cyclic intermediate, followed by migration with low activation barriers between the C9-, C8-, and C7-hydroxyl groups of Neu5Ac.

## Methods

The nanoreactor is a specially modified AIMD simulation with the primary goal of inducing reaction events on a short simulation time scale, achieved by means of a time-oscillating boundary

potential that generates high-energy molecular collisions. The initial conditions consisted of one Sia molecule (either Neu5,7Ac<sub>2</sub> or Neu5,9Ac<sub>2</sub>) surrounded by twelve water molecules randomly placed within a 5.8 Å and 5.4 Å sphere, respectively, using Packmol software,<sup>30</sup> followed by energy minimization. The level of theory for the nanoreactor simulations was B3LYP with a DFT-D3 dispersion correction, 3-21G Gaussian basis set, and SWIG-PCM polarizable continuum model of the aqueous reaction medium ( $\epsilon=78.4$ ), denoted as B3LYP-D3/3-21G/PCM. Simulations utilized the TeraChem quantum chemistry software<sup>31-37</sup> and were run on servers with an Intel Xeon CPU / Nvidia GPU (either E5-2630 v3 CPU / GTX 980 Ti GPU or E5-2637 v4 CPU / GTX 1080 Ti GPU). The computational cost of AIMD simulations is high, which limited the simulations to a few tens of picoseconds, despite our use of the relatively inexpensive 3-21G basis set.

Reaction events are induced by a time-dependent restraint potential in equation (1), where  $m$  is the atomic mass,  $R_1$  and  $R_2$  are radius parameters, and  $k_1$  and  $k_2$  are force constants. The restraint is a flat-bottomed harmonic potential with time-dependent parameters:

$$V(r,t) = \frac{m}{2} k(t) \rho(t)^2; \quad \rho = \begin{cases} r - R(t) & \text{if } r > R(t) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$k(t), R(t) = \begin{cases} k_1, R_1 & \text{if } t \bmod (t_1 + t_2) < t_1 \\ k_2, R_2 & \text{otherwise} \end{cases}$$

where the values of the radius  $R$  and spring constant  $k$  oscillate in a square wave pattern, and are equal to  $(R_1, k_1)$  for period  $t_1$  and  $(R_2, k_2)$  for period  $t_2$ . Generally,  $R_1 > R_2$  and during the period  $t_1$  the molecules are free to move within a spherical volume of radius  $R_1$  with no restraint force. When the radius decreases from  $R_1$  to  $R_2$ , atoms beyond  $R_2$  are accelerated inward, with the mass-

dependent force ensuring near-uniform acceleration of atoms within a molecule. Reaction events are initiated from high velocity collisions of reactants, which generally occur within the smaller spherical volume of radius  $R_2$ .

Simulations used a time step of 0.5 fs with a Langevin thermostat set to 1000 – 2000 K, and a friction coefficient of  $3.3 \text{ ps}^{-1}$ . Other simulation parameters were varied as we searched for the optimal combinations that yielded chemically reasonable reactivity within 2 ps of simulation time. At 2000 K, decarboxylation was observed to occur spontaneously, whereas temperatures between 1000–1500 K were optimal for relevant reaction exploration. The initial boundary radius ( $R_1$ ) and second boundary radius ( $R_2$ ) were chosen to be  $\sim 1.5\text{--}2\times$  and  $\sim 0.8\times$  that of the smallest initial radius used in Packmol. Reactivity can be further tuned by adjustment of the  $R_2$  parameter, where smaller values of  $R_2$  result in higher-energy collisions and greater frequency of generated reactivity, although the reaction pathways found in this way tended to have higher activation energies after refinement. For example, some of our simulations used a combination of the following parameters:  $T = 1500 \text{ K}$ ,  $R_1 = 12.0 \text{ \AA}$ ,  $k_1 = 1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2} \text{ amu}^{-1}$ ,  $t_1 = 1000 \text{ fs}$ ,  $R_2 = 4.4 \text{ \AA}$ ,  $k_2 = 0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2} \text{ amu}^{-1}$ , and  $t_2 = 500 \text{ fs}$ .

The reaction events observed in the nanoreactor trajectory underwent an energy refinement procedure to produce optimized transition states and corresponding reactant/product structures. The refinement calculations used the  $\omega\text{B97X-D3/TZVP/PCM}$  level of theory,<sup>38–40</sup> which is more accurate (and more costly) compared to what was used for the nanoreactor simulations. Initial transition state guesses were optimized using the nudged elastic band (NEB) method<sup>44,45</sup> implemented in a development version of the geomeTRIC optimization software<sup>46</sup> that calls the



TeraChem software<sup>47,48</sup> for energy and gradient evaluations. The transition state optimizations were performed in Q-Chem.<sup>49</sup> A frequency calculation was used to confirm the presence of a single imaginary mode in the optimized transition state, and an intrinsic reaction coordinate (IRC) calculation was then performed to connect the optimized transition state with reactant to product structures.

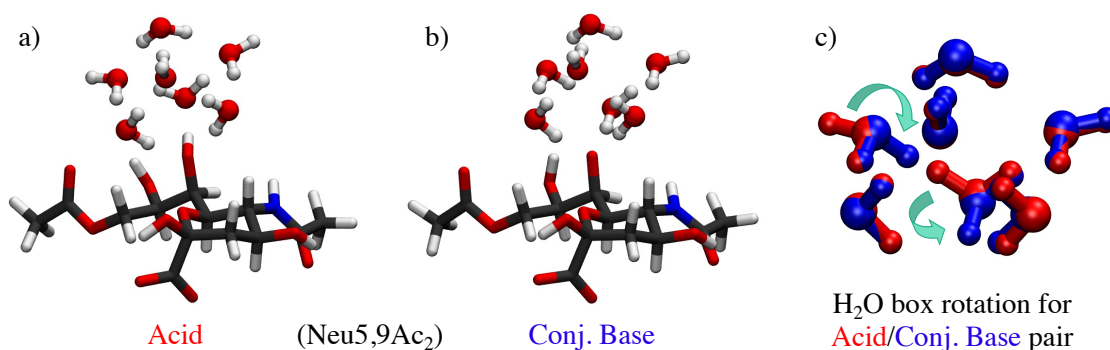
Torsional energy profiles were computed for a few minor intermediate steps, for the purpose of matching each geometric endpoint between elementary steps. More specifically, for a few glycerol rotation steps that were unable to optimize to single transition states, torsional energy profiles were computed by scanning relevant dihedrals by  $\sim 3^\circ$  per step using a development version of the geomeTRIC optimization software<sup>46</sup> at the  $\omega$ B97X-D3/TZVP level of theory. This was done to maintain maximal consistency between each intermediate within the reaction pathway.

The  $pK_a$  values for two conformers of each glycerol ionizable group on Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> or Neu5,7Ac<sub>2</sub>, resulting in twelve acid/conjugate base pairs, were computed according to the following equation.<sup>128,129</sup>

$$pK_a = \frac{\Delta\Delta G}{2.303 kT} = \frac{\Delta G_{base+H^+} - \Delta G_{acid}}{2.303 kT} \quad (2)$$

Eleven acid/conjugate base pairs were used to compute  $pK_a$  values for Neu5Ac, starting from the above structures and replacing the acetyl group with hydroxyl. In the  $pK_a$  calculations, the Sia ionizable group was modeled by adding seven explicit water molecules, which forms a cubic hydrogen bonded configuration together with the ionizable group. In the conjugate base structure,

three water molecules are positioned in a H-bond donor orientation, whereas for the acid, one of the H-bonding protons on water was rotated away, thereby maintaining a consistent number of H-bonds and maximal geometric consistency per acid/conjugate base pair.<sup>128</sup> The resulting structure was optimized at the  $\omega$ B97X-D3/TZVP/PCM level of theory. Frequency analyses for vibrational contributions to free energies were calculated at the same level of theory at 300 K, while electronic energy of the optimized structure was calculated using the larger def2-TZVP-f basis set. In our  $pK_a$  calculations, we used the free energy value of 11.803 eV<sup>130</sup> for a proton at  $pH = 0$ . Figures 3.2a and 3.2b show the optimized acid and conjugate base structures, and Figure 3.2c is an overlay of the water molecule structures to show the similarity in the H-bonding patterns.



**Figure 3.2.** Diagram of Neu5,9Ac<sub>2</sub> surrounded by a water box of seven waters. Water-box surrounding the ionizable C7-OH of Neu5,9Ac<sub>2</sub>, maintaining three H-bonds, in the acid (a) and the conjugate base (b) forms. (c) Illustration describing the proton bond rotations to maintain consistent H-bonding between the acid (red) and the conjugate base (blue) forms, highlighted in green arrows. Coloring key: H: white; C: gray; O: red; N: blue.

Free energy profiles for cyclized bond-dissociation steps were computed from umbrella sampling simulations and using the plumed<sup>131</sup> plugin in Terachem at B3LYP/6-31G\*/PCM level of theory with a harmonic spring constant of 0.50 hartree/bohr<sup>2</sup>, and Langevin thermostat set to 298.15 K. Eight explicit water molecules were placed around the oxide and near the dissociating bond to improve the description of solvation. Simulations of 5 ps in length were run with bond distances set from 2.60 bohr to 4.50 bohr (1.38 Å – 2.38 Å) in increments of 0.10 bohr (0.05 Å), where larger distances resulted in unrealistic multi-bond cleavage within the Sia. Simulations within these distances were excluded when unusual reactivity and multi-bond cleavage occurred over the desired bond dissociation. Free energy profiles were generated from the last 4 ps of simulation data, using the Grossfield implementation of WHAM, version 2.0.10,<sup>132</sup> with 17-20 windows (depending on relevant simulations), from 2.6025 bohr to 4.5975 bohr at 298.15 K, and with Monte Carlo bootstrap error analysis with 2000 iterations.<sup>132,133</sup>

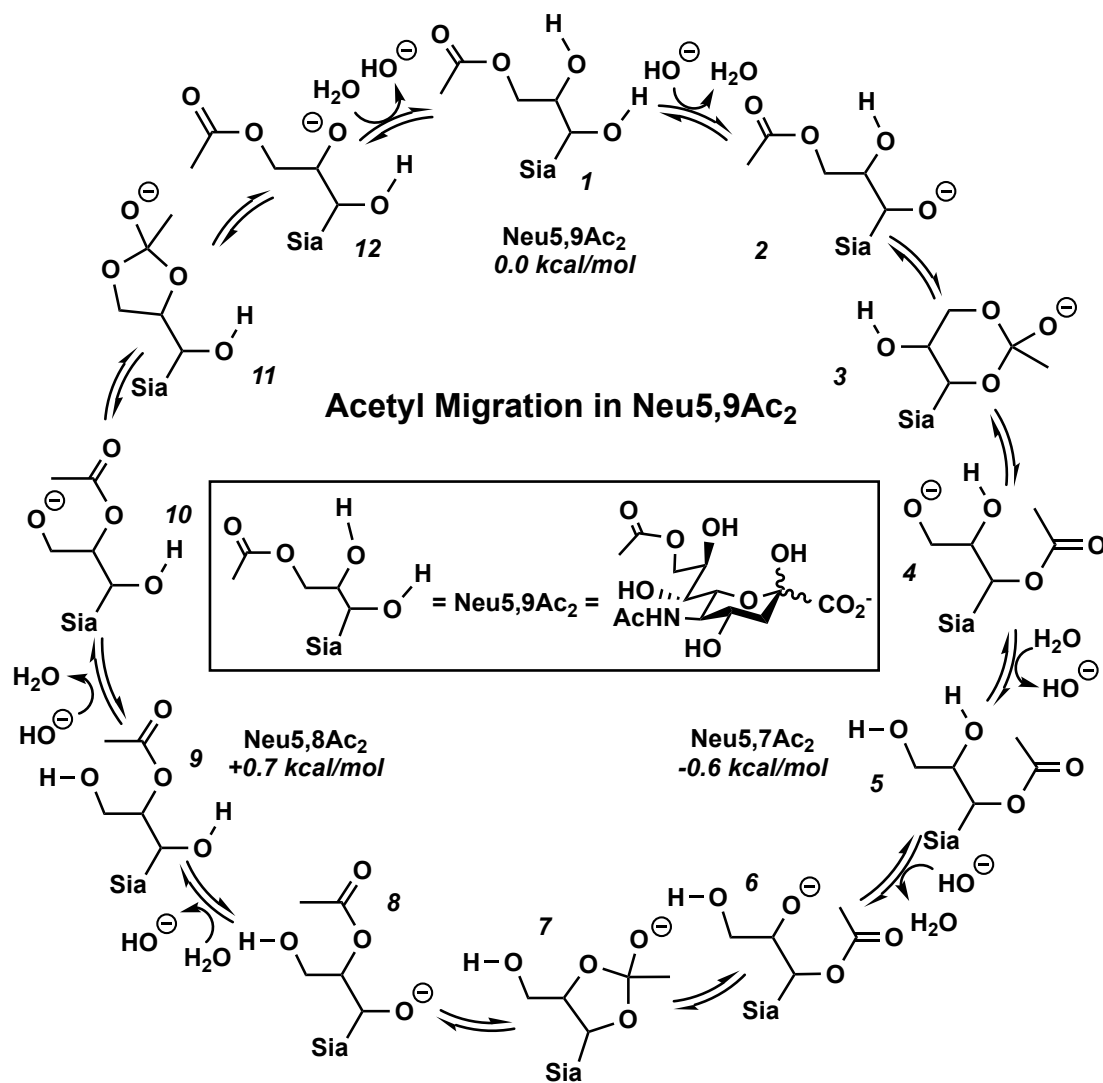
## Results and Discussion

We initially looked for the migration from Neu5,9Ac<sub>2</sub> to Neu5,7Ac<sub>2</sub> through an intramolecular cyclization step, but found this to be energetically prohibitive with barriers of ~50 kcal/mol (Fig. S3.1). Following this, we carried out nanoreactor simulations to explore *O*-acetyl reactivity using Neu5,9Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub> as key reactants, and observed many proton transfer events between water and the glycerol hydroxyl groups (C7-OH, C8-OH, or C9-OH), de-*O*-acetylation, and HO<sup>-</sup> addition to the *O*-acetyl group. After energy refinement, we found that most reactions that started from deprotonated species had low free energy barriers of less than 15 kcal/mol, (Fig. S3.2). This is consistent with experimentally observed trends of increasing *O*-acetyl migration with higher

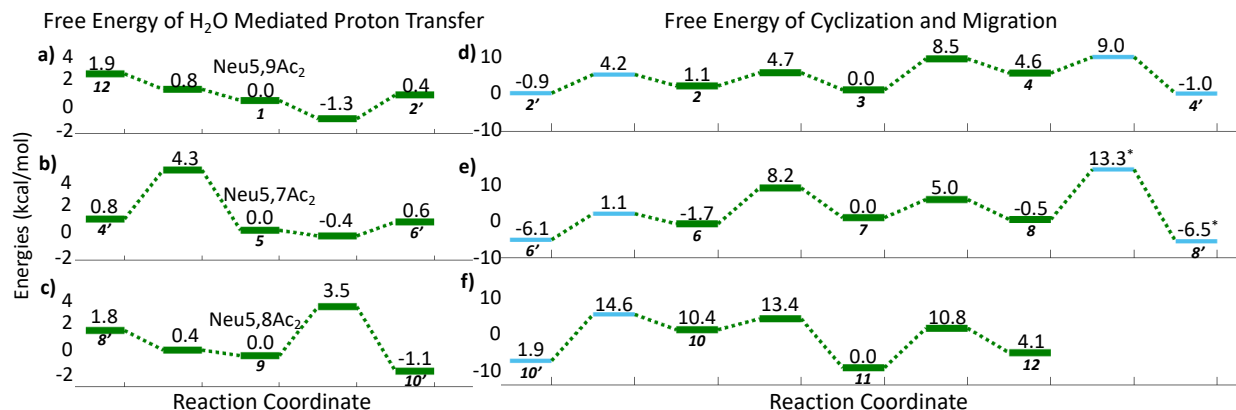
pH.<sup>134</sup> Based on this reactivity, we explored *O*-acetyl migration steps involving deprotonation by HO<sup>-</sup>.

Our calculations indicate the migration of the acetyl group on Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> or Neu5,7Ac<sub>2</sub> likely proceeds via a mechanism involving deprotonation, cyclization, followed by ring opening, subsequently resulting in migration (Fig. 3.3). When base is present, deprotonation of a glycerol hydroxyl group is facile, with reaction barriers under 5 kcal/mol. This is hypothesized to be the first step in the acetyl migration pathway (Fig. 3.4a,b,c). After deprotonation of Neu5,9Ac<sub>2</sub> C7OH (**1**), the resultant oxide (**2**) is a nucleophile that attacks the carbonyl carbon of the acetyl group and forms a cyclic intermediate with the carbonyl carbon adopting a tetrahedral geometry (**3**, **7**, **11**) (Fig. 3.4d,e,f). This cyclic intermediate then ring-opens to result in an oxide at the previously *O*-acetyl-substituted carbon and the migrated acetyl group. Within this process, the oxygen atoms directly bonded to C7, C8 and C9 remain in place, and only the C(=O)CH<sub>3</sub> atoms migrate. Re-protonation by water results in Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> or Neu5,7Ac<sub>2</sub> (**1**, **9**, **5**). Further migrations repeat the previous steps of deprotonation, cyclization, migration, and re-protonation. In addition to these chemical steps, some conformational differences were observed within an intermediate (e.g. the product structure of 1 → 2 and the reactant structure of 2 → 3) which required further calculations to determine the free energy barrier of the conformational changes. Figure 3.2 indicates a single highest barrier for a given step (deprotonation, cyclization, or migration) in the direction of Neu5,9Ac<sub>2</sub> to Neu5,7Ac<sub>2</sub> to Neu5,8Ac<sub>2</sub> (see Fig. S3.3 for full pathway), including any conformational changes. For example, in order to cyclize **2** to **3**, three rotations of glycerol O-H groups and C-C torsions are required to prepare the acetyl group for nucleophilic attack. Our

calculations require the presence of a  $\text{HO}^-$  anion, consistent with experiments that *O*-acetyl migration has a tendency to occur at higher pH values.



**Figure 3.3.** Mechanism of base-catalyzed *O*-acetyl migration between Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub>, and Neu5,7Ac<sub>2</sub>. Relative free energies calculated at  $\omega$ B97X-D3/TZVP level of theory as described in the Computational Details.



**Figure 3.4.** Reaction and activation free energies for base-catalyzed *O*-acetyl migration of Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub>. Small numbers in bold font indicate structures depicted in Figure 3.3. (a) through (c) show deprotonation of C(7/8)-OH of Neu5,9Ac<sub>2</sub>, C(9/7)-OH of Neu5,8Ac<sub>2</sub> and C(8/9)-OH of Neu5,7Ac<sub>2</sub> respectively, with free energies relative to the sialic acid and HO<sup>-</sup>. (d) through (f) show cyclization and ring-opening reactions for *O*-acetyl migration in C9-C7, C7-C8 and C8-C9 respectively, with free energies relative to the cyclic intermediate. Light blue lines (e.g., 2 ↔ 2') indicate the free energies of conformational changes between the endpoints of the deprotonation and cyclization steps (for example, deprotonation of **1** results in **2'**, followed by conformational changes to **2**, then cyclization to **3**). \*Energy barriers of conformational change steps were estimated by electronic energies from torsion scans (Fig. S3.4). Free energies are calculated at ωB97X-D3/TZVP level of theory as described in the Computational Details; negative activation free energies are an artifact of the quasi-harmonic approximation.

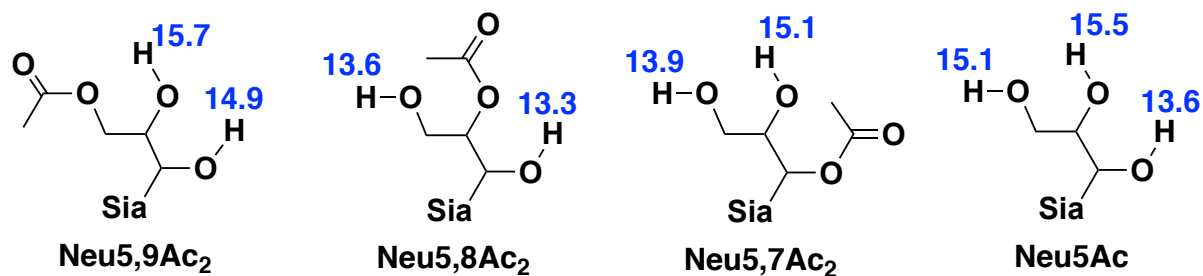
While the base-catalyzed migration between Neu5,9Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub> via the 6-membered cyclic intermediate (**3**) has the lowest free energy barriers (< 9 kcal/mol for each step), the transition state energies for the migrations between Neu5,9Ac<sub>2</sub> and Neu5,8Ac<sub>2</sub>, or Neu5,7Ac<sub>2</sub> and Neu5,8Ac<sub>2</sub>, are

also low ( $< 15$  kcal/mol). The largest barriers in the migration route from Neu5,9Ac<sub>2</sub> to Neu5,7Ac<sub>2</sub> to Neu5,8Ac<sub>2</sub> route is 12.7 kcal/mol for glycerol rotations of **10** (or, **10'** after deprotonation of **9**, to **10**) to geometrically pre-organize for the cyclization step in the *O*-acetyl migration from C8-OH to C9-OH (**11**). The reverse direction maximum barrier is 14.2 kcal/mol, for the **8** to **7** glycerol rotations needed for the cyclization step to migrate the *O*-acetyl from C8-OH to C7-OH. Excepting these two glycerol rotation barriers, the maximum barrier in this migration cycle is  $\sim 11$  kcal/mol in either direction. Additionally, relative free energy differences of Neu5,8Ac<sub>2</sub> (+0.7 kcal/mol vs. Neu5,9Ac<sub>2</sub>) and Neu5,7Ac<sub>2</sub> (-0.6 kcal/mol vs. Neu5,9Ac<sub>2</sub>), calculated for structures **1**, **5**, and **9**, are small and within the margin of error of our calculations. While Neu5,7Ac<sub>2</sub> is slightly lower in free energy than Neu5,9Ac<sub>2</sub>, this may in part be due to the formation of an intramolecular H-bond between the amide proton of the *N*-acetyl on C5 with the carbonyl of *O*-acetyl on C7, with a distance of 2.0 Å. For the other Sias, the *N*-acetyl proton is closest to C7-OH with a distance of  $\sim 2.5$  Å. Because the barriers are low in each step of the base-catalyzed migration, and the stabilities of Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub> are similar, we think the acetyl group can migrate readily between positions (C7/8/9) at an increased pH.

This migration competes with de-*O*-acetylation, which is exergonic by  $-20.5$  kcal/mol with a highest point on the reaction pathway  $\sim 18$  kcal/mol higher than the starting state. Given that the free energy barriers of *O*-acetyl migration in either direction from Neu5,9Ac<sub>2</sub> are  $< 12$  kcal/mol, the de-*O*-acetylation pathway therefore has a higher barrier, indicating that *O*-acetyl migration is kinetically more favorable. Thus, the de-*O*-acetylated Neu5Ac is the thermodynamic product, while the *O*-acetylated isomers are kinetic products.



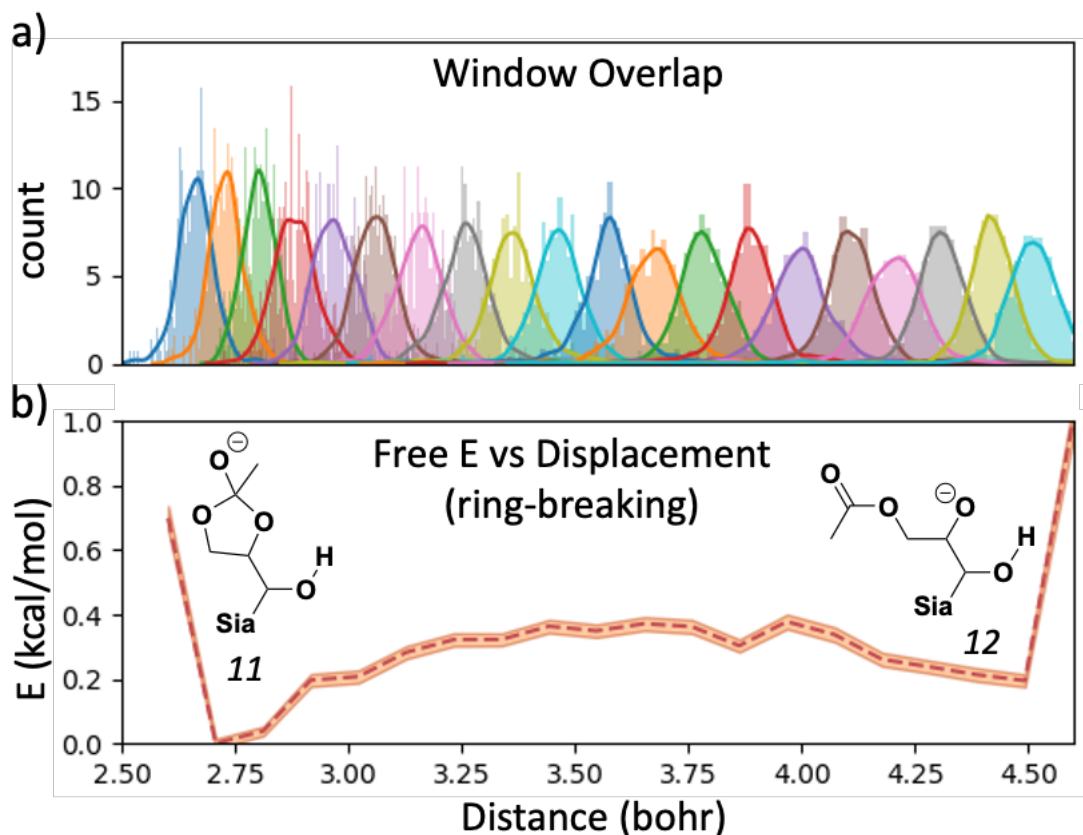
In the cyclization and ring-opening calculations, we did not use explicit solvent, which is needed to describe hydrogen bonding with solvent molecules. Given the mechanism's dependence on initial deprotonation, and that these initial deprotonation steps have low barriers, we calculated  $pK_a$  values in the presence of explicit water to determine which hydroxyl may be deprotonated first. This could suggest a potential migration direction between Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub> (see Methods for details). By taking the average of each ionizable group across the respective two Sias (ex. C9-OH results averaged from Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub>), the  $pK_a$  values of C9-OH and C7-OH are more similar at 13.8 and 14.1, compared to the slightly higher  $pK_a$  of C8-OH at 15.4 (Fig. 3.5). We performed additional  $pK_a$  calculations on the unmodified Neu5Ac for reference and found that the  $pK_a$  of C7-OH to be 13.6, lower than the  $pK_a$ s for C8-OH and C9-OH at 15.5 and 15.1, respectively. From our calculations on Sias C9/8/7-OH, as distance from the *O*-acetyl group increased,  $pK_a$  values tend to decrease by about 1. Even so, given the small differences between the  $pK_a$  values, this study further suggests that deprotonation of any hydroxyl group is comparatively easy between the Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub>, and that all deprotonated forms are likely present in an equilibrium mixture.



**Figure 3.5.** Computed  $pK_a$  values are highlighted in blue next to each ionizable group of Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub>, Neu5,7Ac<sub>2</sub>, and Neu5Ac; values accurate within a 0.9 standard deviation, the largest standard deviation for the listed values.  $pK_a$  values were calculated with  $\omega$ B97X-

D3/def2-TZVP-f level of theory for single point energy contributions and  $\omega$ B97X-D3/TZVP for vibrational contributions to free energies, as described in the Computational Details.

The conformational flexibility of the ring and glycerol chain across Sias was evident in our calculations of energy minima and transition states. To improve the sampling of particular reaction steps, we carried out umbrella sampling along the C-O distance coordinate for the cyclization and de-*O*-acetylation steps. Umbrella sampling may give a clearer picture of which ring is easier to break, and to which direction the *O*-acetyl group is more likely to migrate. Figure 3.6 shows the free energy of ring-opening from **11** to **12** in the *O*-acetyl migration from C8-OH to C9-OH, which involves a free energy barrier of < 1.0 kcal/mol, and a window overlap figure from our simulations. The figure indicates sufficient thermodynamic overlap between each window where the cyclized bond distance increases until the ring breaks, ensuring that the resultant free energy results are reliable. Each of the other ring-breaking steps has similar results, with free energy barriers of < 1.0 kcal/mol and similar window overlaps (Fig. 3.5 and Fig. S3.5). Umbrella sampling results with explicit water yielded lower activation free energies compared to those obtained from vibrational analysis without explicit water. This further supports the low-barriered, facile *O*-acetyl migration between the hydroxyl groups at C7, C8, and C9 of Sia.



**Figure 3.6.** Example umbrella sampling of the cyclization/*O*-acetyl migration step between C9/C8-hydroxyl groups. (a) Overlap of windows sampling over bond length. (b) Free energy (red dashed line) of ring-break for *O*-acetyl migration from C8-OH to C9-OH with estimated error shaded in orange. Umbrella sampling carried out at B3LYP/6-31G\* level of theory as described in the Computational Details. See Fig. S3.5 for all six steps.

To compare with experiment, we calculated free energies from population ratios in our previous paper<sup>1</sup>. From the average population ratios (between SM1 and SM2) that reached an equilibrium at *pH* 8.0 and 37 °C, the free energies of Neu5,7Ac<sub>2</sub> and Neu5,8Ac<sub>2</sub> with respect to Neu5,9Ac<sub>2</sub> were +2.06 and +1.25 kcal/mol. Our values from a single structure of Neu5,7Ac<sub>2</sub> and Neu5,8Ac<sub>2</sub> (Fig. 3.3) were -0.6 and +0.7 kcal/mol, with respect to Neu5,9Ac<sub>2</sub>. The free energies for Neu5,8Ac<sub>2</sub>

match within the margin of error, but the computed value for Neu5,7Ac<sub>2</sub> was slightly lower than expected. Our computational results are not precise enough to reproduce the experimentally measured population ratios of isomers; this could partly be due to the approximate level of theory used, and/or our use of the quasi-harmonic approximation to evaluate these free energy differences. While it may be possible to obtain more accurate free energy estimations using more advanced sampling methods,<sup>135–137</sup> it does not affect our main conclusion regarding the proposed migration mechanism, in which migration in any direction is facile provided that base is present.

## Conclusion

Our calculations indicate that when a base is present, *O*-acetyl migration occurs readily among the hydroxyl groups at C9, C7, and C8 of the Sia via deprotonation, migration, and cyclization steps. This conclusion is based on computed *pK<sub>a</sub>* values, umbrella sampling, and DFT calculations on the migration steps and the competing de-*O*-acetylation reaction. When explicit water molecules were included in the calculations, their presence affected relative energies of intermediates but did not change the main conclusion that base-catalyzed acetyl migration is facile. *O*-Acetyl migration may allow host cells to evade viral binding. In our recent article, we noted the preferential binding of SARS-CoV-2 S protein to immobilized acetyl group at the C9-OH by using an *N*-acetyl group analog Neu5Ac9NAc which is resistant to migration over its *O*-acetyl counterpart Neu5,9Ac<sub>2</sub>.<sup>125</sup> Further work is necessary to probe the role of *O*-acetyl migration in viral recognition and pathogenesis.

## **Acknowledgements**

We are grateful for the insightful discussions with Darón I. Freedberg and Audra A. Hargett at the Laboratory of Bacterial Polysaccharides, Food and Drug Administration (FDA). This work was supported by the United States National Institutes of Health grant no. R01AI130684.

**Chapter 4: SARS-CoV-2 and MERS-CoV spike protein binding studies support stable mimic of bound 9-*O*-acetylated sialic acids**

The work in this chapter is based on “SARS-CoV-2 and MERS-CoV spike protein binding studies support stable mimic of bound 9-*O*-acetylated sialic acids” with Ajit Varki, Xi Chen, and Lee-Ping Wang, submitted to Nature Communications.

## Abstract

Many disease-causing viruses target sialic acids (Sias), a class of nine-carbon sugars known to coat the surface of many cells including those in the lungs. Human beta coronaviridae, known for causing respiratory tract diseases, often bind Sias, some preferentially bind to those with 9-*O*-Ac modification. Currently, co-binding of SARS-CoV-2, a beta coronavirus responsible for the COVID-19 pandemic, to human Sias has been reported and its preference towards  $\alpha$ 2-3-linked Neu5Ac shown. Nevertheless, *O*-acetylated Sias-protein binding studies are difficult, due to the ester lability. We studied the binding free energy differences between Neu5,9Ac2 $\alpha$ 2-3Gal $\beta$ pNP and its more stable 9-NAc mimics binding to SARS-CoV-2 spike protein using molecular dynamics and alchemical free energy simulations. We identified multiple Sias-binding pockets, including two novel sites, with similar binding affinities to those of MERS-CoV, a known co-binder of sialic acid. In our binding poses, 9-NAc and 9-OAc Sias bind similarly, suggesting an experimentally reasonable mimic to probe viral mechanisms.

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the beta coronavirus responsible for the COVID-19 pandemic, is structurally highly similar to SARS-CoV-1 (73% sequence identity between spike proteins), yet is remarkably more infective.<sup>58</sup> Both SARS-CoV-1 and SARS-CoV-2 share the same primary human cellular receptor angiotensin-converting enzyme-2 (ACE-2), but this commonality in binding cannot explain the difference in infectivity. One possible mechanism for how SARS-CoV-2 achieves its high infectivity is by binding to sialic acids (Sias),

which coat many cell surfaces, especially in the lungs, and are targeted by many disease-causing viruses. Sias are part of a large family of over 50 derivatives of the 9-carbon sugar neuraminic acid, where N-acetylneuraminic acid (Neu5Ac) is most common.<sup>68</sup> Beta coronavirus Middle East respiratory syndrome coronavirus (MERS-CoV) is known to co-bind to Neu5Ac in addition to its DPP4 primary receptor in a two-step binding mechanism, and depletion of Neu5Ac was found to inhibit MERS-CoV entry into human airway cells.<sup>79</sup> Additionally, MERS-CoV binds preferentially to  $\alpha$ 2-3-linked Sias over  $\alpha$ 2-6-linked ones.<sup>79</sup> In another example, beta coronavirus OC43 and HKUI are known to bind to 9-*O*-acetylated (9-OAc) Sias,<sup>57,58</sup> where *O*-acetylation is one of the most common Sia modifications found in nature.

While SARS-CoV-1 is not known for its binding to Sias, recent studies have shown Sia binding by the SARS-CoV-2 spike (S) protein, and suggest that sialylated glycans can facilitate viral entry.<sup>138,139</sup> In a lateral flow test, SARS-Cov-2 S protein bound to both Neu5Ac and  $\alpha$ 2-3/ $\alpha$ 2-6-linked Sias, where each Sia is each attached to gold nanoparticles (auNP).<sup>140</sup> While stronger binding under experimental conditions was observed with Neu5Ac-auNP over  $\alpha$ 2-3/ $\alpha$ 2-6-linked Sias-auNPs, the paper did not preclude involvement of  $\alpha$ 2-3/ $\alpha$ 2-6-linked Sias in SARS-CoV-2 S protein binding.<sup>140</sup> Binding studies using sialylated glycans indicate low affinities to the SARS-CoV-2 S protein ( $\sim$  -10 kcal/mol)<sup>125</sup> and specifically along the ACE-2 receptor binding domain (RBD), with monosialylated ganglioside glycan affinities of 100-200  $\mu$ M and multisialylated glycan affinities approaching 900  $\mu$ M, using catch-and-release ESI-MS (CaR-ESI-MS) ( $\sim$  -20 kcal/mol).<sup>138</sup> A few potential sialic acid-binding domains have been proposed additional to the ACE-2 RBD, especially along the flat region of the SARS-CoV-2 S protein N-terminal domain (NTD), but specific sites have not been confirmed experimentally.<sup>141-148</sup> The S protein is made up

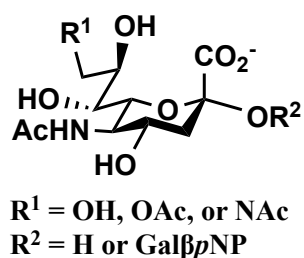


of two subunits, S1 for cell recognition, containing both NTD and ACE-2 RBD , and S2, responsible for viral cell membrane fusion.<sup>149</sup> Given the density of sialic acids along cell surfaces, and role of S1 in cell recognition, Sias could increase viral binding affinity by acting as an intermediate target or co-binder. We attempt to provide a quick overview of current work on Sias-SARS-CoV-2 S protein binding, where a more thorough study may be found in the recent review by Sun.<sup>150</sup>

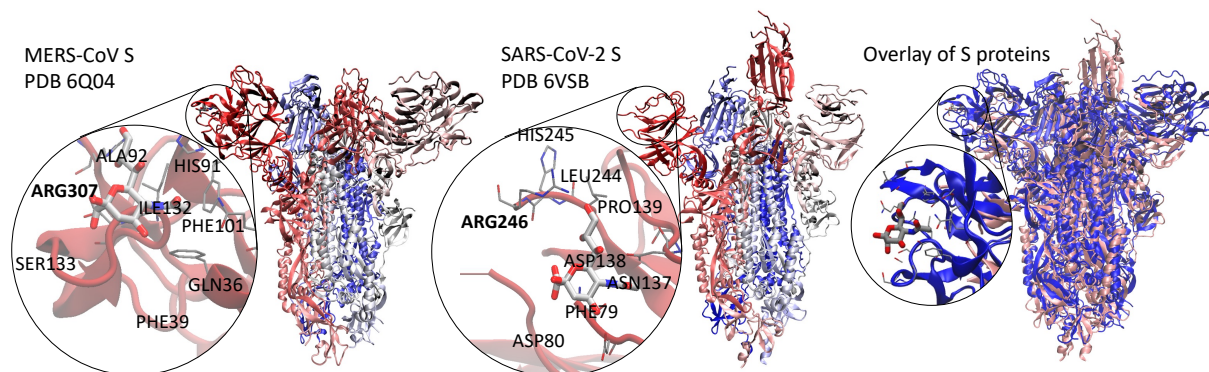
Experimental studies that test S protein binding to Neu5Ac and *O*-acetylated Sias using glycan microarrays are informative towards understanding the disease and preferential binding.<sup>125</sup> Even so, experimental testing is difficult with *O*-acetylated Sias due to the instability of the ester with respect to migration and cleavage, which depends on pH, temperature, and presence of esterases.<sup>1,2,69</sup> *N*-Acetylated (NAc) Sias have been proposed to be a stable synthetic mimic, as they are chemically and structurally similar, as seen in experimental and computational NMR studies<sup>2</sup>. Determining the similarity in spike protein binding of 9-NAc to 9-OAc Sias would be valuable for when performing binding array studies, and understanding the binding sites of sialic acids to SARS-CoV-2 spike protein is important when considering potential druggable sites.

To this end, we computed the binding free energies of modified Neu5Ac monomers and oligosaccharides to SARS-CoV-2 S protein using molecular dynamics (MD) simulations (Fig. 4.1). Our simulations revealed new possible Sia binding sites ranging along the S1 unit, along the NTD and RBD, with some approaching the S2 domain. Each binding site contains a salt bridge connecting a conserved arginine residue to the carboxylate group of Sia, a known motif in Sia-lectin binding;<sup>151</sup> while the binding to individual sites is predicted to be weak ( $\sim -15$  kcal/mol),

multiple binding to cell surface Sias could strengthen binding overall.<sup>58,152</sup> In addition, we calculated relative binding free energies of ligands that differ in the chemical modification at C9 (9-OH, 9-OAc, and 9-NAc), showing that the synthetic 9-NAc are excellent structural mimic of their naturally occurring 9-OAc counterparts. These insights into Sia-S protein binding could lead to the design of therapeutics that inhibit the binding of S protein to Sias on the cell surface, thereby limiting SARS-CoV-2 transmission.



**Figure 4.1:** Structure of N-Acetylneuraminic acid (Neu5Ac), the simulated ligand in this study. Three substitutions at  $R^1$  were simulated corresponding to the unsubstituted Sia and its 9-OAc and 9-NAc forms. Two substitutions at  $R^2$  were simulated, corresponding to the Sia monomer or an  $\alpha$ 2-3-sialoside containing the terminal 2 sugars of the GM3 ganglioside, commonly found on cell membranes, followed by *p*NP, a molecule used in quantifying Sias binding and cleavage in sialidase microarray studies.



**Figure 4.2:** Initial binding pose of Neu5Ac in SARS-CoV-2 spike protein (a) used as a starting point for MD simulations, based on cryo-EM structure of MERS-CoV spike protein with Neu5Ac bound (b). (c) The initial pose was modeled based on structural overlay of pink SARS-CoV-2 S (PDB ID: 6VSB) with blue MERS-CoV S in complex with Neu5Ac (PDB ID: 6Q04), using MultiSeq<sup>153</sup> an extension of the Multiple Alignment tool available in VMD.<sup>154</sup>

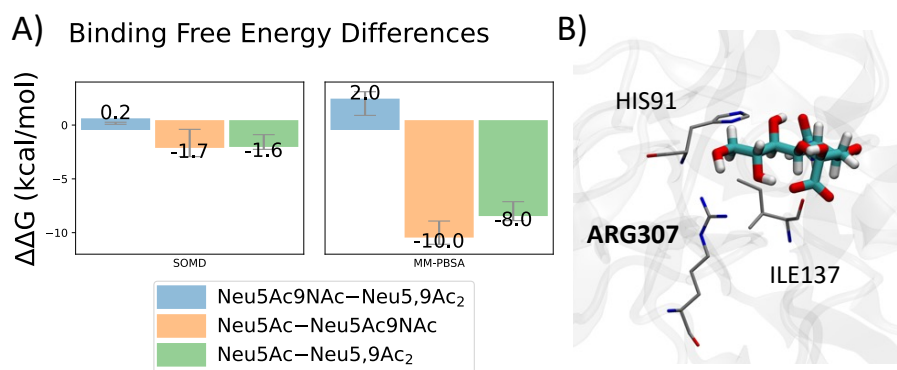
## Results and Discussion

### Method Validation with Sias-MERS-CoV S binding

Modern free energy simulation methods are estimated to be accurate to within 1-2 kcal/mol for well-behaved<sup>82</sup> protein/ligand systems, but this system presents additional challenges due to the exceptional flexibility of both the SARS-CoV-2 S protein and the ligands, the latter of which we have studied computationally<sup>69</sup> and using NMR.<sup>2</sup> Based on a recent SAMPL6 challenge that evaluates binding free energy prediction methods,<sup>91</sup> we used two methods in this study – an inexpensive implicit solvent approach known as molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)<sup>155</sup> to estimate absolute binding free energies, and a more rigorous approach based on alchemical intermediates known as SIRE-OpenMM molecular dynamics (SOMD)<sup>84,85</sup> to compute relative binding free energies of ligands that differ only in the chemical modification at the Sia C9.

Given the exploratory nature of Sias-SARS-CoV-2 S binding, we first validated our methods with Sias bound to the RBD in MERS-CoV S, known to preferentially bind to Neu5Ac-containing Sias

over Neu5,9Ac<sub>2</sub>- and Neu5Ac9NAc-containing Sias. Both MM-PBSA binding energy and SOMD alchemical free energy differences show that Neu5Ac binds stronger than either modified variant, consistent with experimental findings (Fig. 4.3). In the case of SOMD, Neu5Ac binds stronger than Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> by 2.7-3.0 kcal/mol. Both Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> bind very similarly, with a nominal difference of 0.2 kcal/mol, within the margin of error. While the magnitudes of the binding free energy differences are larger in the MM-PBSA results, they follow the same trend as our SOMD results, where Neu5Ac binds stronger than Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> (by 8-10 kcal/mol), with the modified Sias binding very similarly. Given the shallow nature of the binding pocket, we did not perform binding free energy simulations with the larger sialosides, as the modified Neu5Ac test case is sufficient for validating our methods. As expected from shallow binding and experimental reports for weak binding affinity<sup>156</sup>, MM-PBSA binding energy results indicated overall weak binding (-20 kcal/mol for Neu5Ac). Based on energy decomposition analysis, one of the key residues contributing to this difference in binding is ARG307, which binds stronger to Neu5Ac (Fig. 4.3).



**Figure 4.3:** A: SOMD and MM-PBSA binding free energy differences for Neu5Ac, Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> bound in MERS-CoV S protein. Neu5Ac binds stronger than either modified

variants, where the Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> result in nominal binding energy differences. B: Binding poses of Neu5Ac, Neu5Ac9NAc and Neu5,9Ac<sub>2</sub> in MERS-CoV S protein. Key contributing residues, based on energy decomposition analysis, are annotated with top contributor bolded. See Fig. S4.1 for energy decomposition results.

Given that the MERS-CoV simulation predictions largely aligned with the experiment, we proceeded toward computational discovery of Sia binding sites in SARS-CoV 2 and estimation of binding free energies using the same computational approach.

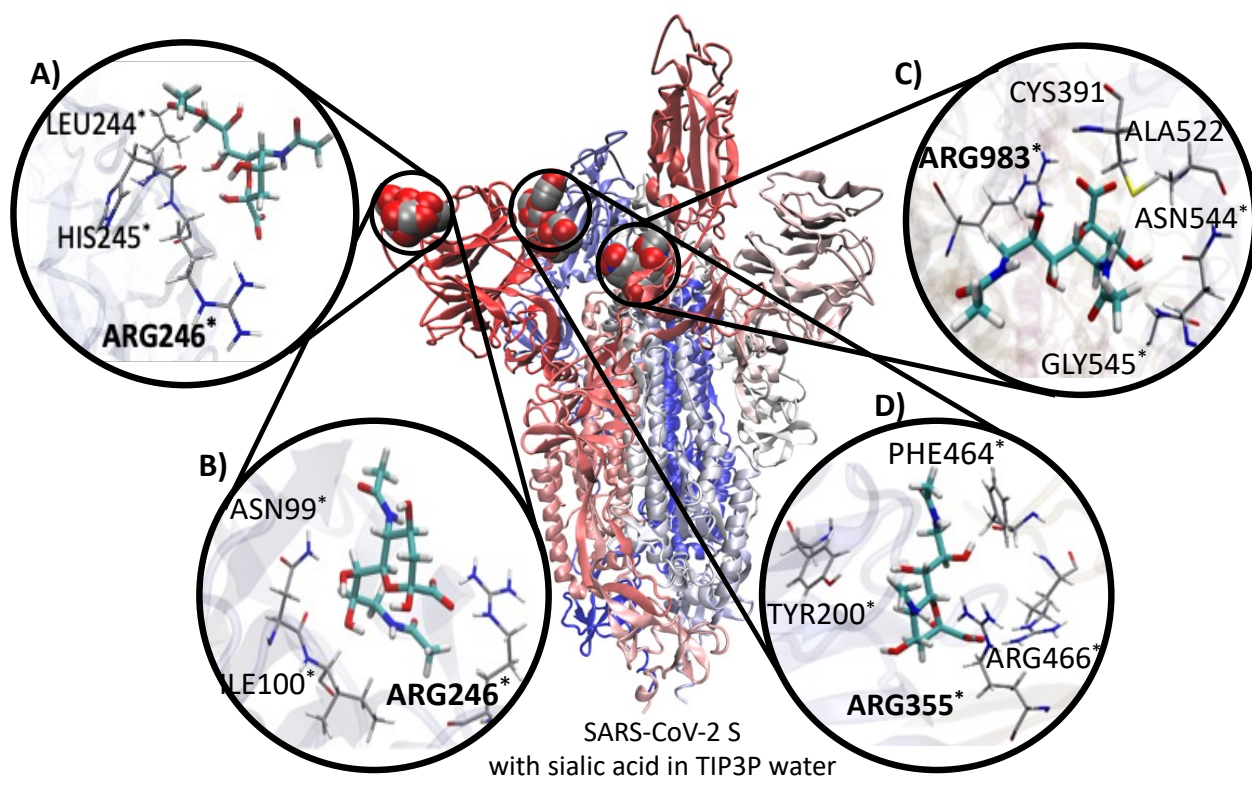
### **Discovery and analysis of Sias-SARS-CoV-2 S binding**

We performed MD simulations of Neu5Ac, Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc bound in SARS-CoV-2 S protein, starting with a docked pose based on the Neu5Ac-MERS-CoV S complex (see Methods for details). After initially setting up docked ligands in the crystal structure of SARS-CoV-2 S protein from cryo-EM (PDB ID: 6VSB), we found a large RMSD in protein structure, indicating a conformational change of the S protein. During the final equilibration phase with no restraints, each Sia (Neu5Ac, Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc) along the NTD of SARS-CoV-2 S protein became unbound (Fig. 4.2). Interestingly, Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc were observed to return to the original RBD before unbinding again. All 3 Sias subsequently sampled temporary binding events to many regions of the S protein. From the simulation trajectories, we initially selected six poses for further investigation, four of which are in Figure 4.2.

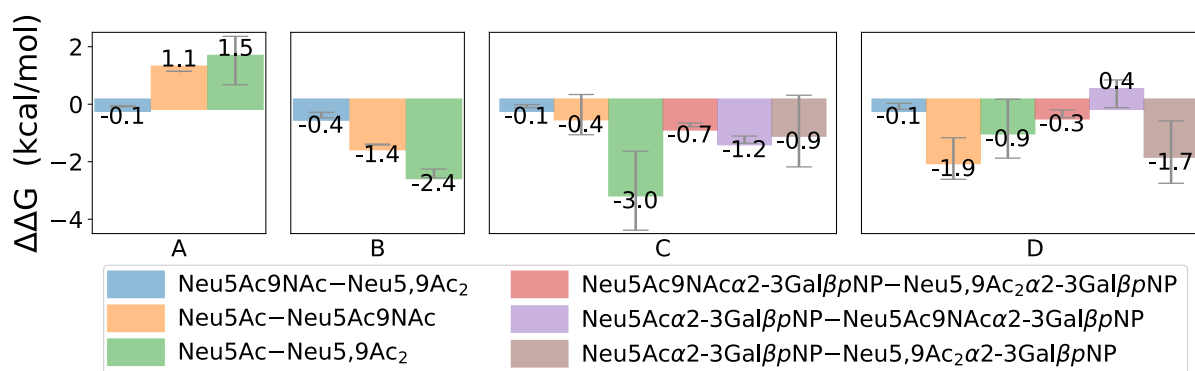
All of the proposed sites from MD involve arginine forming a salt bridge with the Sia's carboxylate, a known motif in Sia-lectin binding.<sup>151</sup> In other Sias-S protein observations without an interacting arginine, the Sias quickly dissociated quickly. While we observed many other binding sites in the MD simulations, such as along the flat top of the NTD within the S1 domain (quick dissociation observed), along the S2 domain and between the S1 and S2 domains (deeper within the S protein trimer), we only considered regions with longer association times and that may be accessible by Sias.

Including the initial pose based on MERS-CoV, we chose 4 binding poses to investigate in closer detail (Fig. 4.4). Here, we describe how each pose was found. Structures of Neu5Ac, Neu5,9Ac<sub>2</sub>, Neu5Ac9NAc were docked into the initial binding pocket along the NTD front (initial pose based on MERS-CoV) (Fig. 4.4A). Given the shallow pocket, each sialic acid unbound and sampled different regions along the S protein. Neu5Ac and Neu5,9Ac<sub>2</sub> sampled the region behind this initial pose (Fig. 4.4B). This pose is close to pose A, potentially allowing for more accessible Sias-binding domains along the S protein NTD. Neu5Ac9NAc sampled a deeper binding pose between chains A (NTD) and C (ACE-2 RBD) in the initial simulations (Fig. 4.4C). In initial MD simulations, Neu5Ac sampled the region between the ACE-2 RBD and S2 domain, and between chains A and B (Fig. 4.4D). This region is accessible for gangliosides to attach to the spike protein and direct or inhibit the down to up state of the ACE-2 RBD. Indeed, a previous study indicated that residues along this region may function as a binding pose for glycosaminoglycans including heparan sulfate,<sup>157-159</sup> highlighting the importance of this region in cellular recognition. Both of these deeper binding domains (Fig. 4.4C, 4.4D) overlap with key regions in stabilizing the S protein and affecting its ability to transition the 3 ACE-2 RBD between “down” and “up” states.<sup>160</sup>

Sias binding in these regions may shift the S protein flexibility and ability to transition between states, altering ACE-2 binding. Notably, and to the best of our knowledge, these two Sia-binding poses (Fig. 4.4C, 4.4D) have not been previously proposed for Sia binding. One possible explanation is that a number of simulations have been performed using a single NTD or subunit rather than the full S protein trimer, which is needed to describe Sias binding between multiple trimer units (such as between chains A & B in pose D).



### SOMD Binding Free Energy Differences in 4 Poses



**Figure 4.4:** Sias binding poses with the SARS-CoV-2 S protein and SOMD binding free energy differences. Upper panel: A) binding pose based on MERS-CoV, Neu5,9Ac<sub>2</sub> highlighted; B and C) binding poses based on initial MD with Neu5Ac and Neu5Ac9NAc highlighted; and D) binding poses based on initial MD with Neu5Ac9NAc, Neu5Ac9NAc highlighted. Nearby protein residues are annotated, with top contributor bolded, based on MM-PBSA energy decomposition analysis. \*Residues conserved across omicron (PDB ID: 7TB4), kappa (PDB ID: 7VXB), delta (PDB ID: 7W92), gamma (PDB ID: 7M8K), original (PDB ID: 6VSB) variants of SARS-CoV-2. Lower panel: SOMD binding free energy differences for Sias bound in SARS-CoV-2 S protein. Results for transformations between Neu5Ac, Neu5,9Ac<sub>2</sub>, and Neu5Ac9NAc shown for all poses (A-D). Results for transformations between Neu5Acα<sub>2</sub>-3GalβpNP, Neu5,9Ac<sub>2</sub>α<sub>2</sub>-3GalβpNP, and Neu5Ac9NAcα<sub>2</sub>-3GalβpNP sialosides shown for poses C and D. Error bars are plotted from standard error of means across 4 simulations, when available. Pose D results in stronger binding to Neu5Acα<sub>2</sub>-3GalβpNP, Neu5,9Ac<sub>2</sub>α<sub>2</sub>-3GalβpNP, and Neu5Ac9NAcα<sub>2</sub>-3GalβpNP over Neu5Ac, Neu5,9Ac<sub>2</sub>, and Neu5Ac9NAc. See Fig. S4.2 for MM-PBSA energies and energy decomposition results.



In all initial MD simulations using the cryo-EM structure, Sias-unbinding events occurred rapidly, given the high flexibility of the SARS-CoV-2 S protein and the resolution of the cryo-EM structure. This flexibility was indicated by relatively fast increase in RMSD at the start of simulations that used the cryo-EM structure. As such, we ran simulations with the SARS-CoV-2 S protein in water to equilibrate the S protein, ideally to a more stable form. We extracted several SARS-CoV-2 S protein structures from the equilibrated trajectories to use for docking of the described 4 poses for production runs, with results in Figure 4.4.

Figure 4.4 show that in each binding pose, Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc bind very similarly, with differences and errors much less than 1 kcal/mol. Each pose contains a conserved arginine motif, which are conserved across most SARS-CoV-2 variants (omicron (7TB4), kappa (7VXB), delta (7W92), gamma (7M8K), original (6VSB)). Binding residues conserved across these variants are annotated in Figure 4.4. Neu5Ac binds weaker than either modified Sia in pose A, with slightly weaker binding to LEU244, but each Sia binding strongest to ARG246. In pose B, ARG246 is again the tightest binding residue to the Sias, and Neu5Ac binds more strongly than either Neu5,9Ac<sub>2</sub> or Neu5Ac9NAc, likely due to its size fitting better in the shallow pocket. In pose C, ARG983 was the tightest binding residue, and ARG355 and ARG466 in pose D. Neu5Ac binds slightly stronger than either modified variant, especially to these key arginine residues. Due to their size, Neu5Ac $\alpha$ 2-3Gal $\beta$ pNP, Neu5,9Ac<sub>2</sub> $\alpha$ 2-3Gal $\beta$ pNP, and Neu5Ac9NAc $\alpha$ 2-3Gal $\beta$ pNP in pose C were slightly displaced from the deep ARG983, resulting in the top binding residue as LYS529. While both Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc bound weaker than the Neu5Ac, these differences were not large (~1 kcal/mol). In binding pose D, while the Neu5Ac $\alpha$ 2-3Gal $\beta$ pNP binds stronger than Neu5,9Ac<sub>2</sub> $\alpha$ 2-3Gal $\beta$ pNP, the standard error of means is relatively large. Larger error

bars calculated for binding poses C and D are due to the range of flexibility observed in the S protein, Sias, and the position and orientation of Sia in the binding pocket.

Based on MM-PBSA binding free energies, Sia binding in each pose of the SARS-CoV-2 S protein is weak (<25 kcal/mol), but is consistent with experimental results,<sup>125</sup> and is comparable to that of Sia-MERS-CoV spike protein binding (see Fig. S4.1 and S4.2 for MM-PBSA energies and decomposition analysis). Of all the binding poses, binding pose D resulted in tighter binding of  $\alpha$ 2-3-linked Sias over their monomer counterparts in its initial equilibrated pose used in SOMD. This highlights the importance of pose D in tighter SARS-CoV-2 S protein co-binding to cell-surface Sias, influencing cellular recognition and ACE-2 binding.

A number of other possible binding sites were observed in the MD simulations, but we did not investigate them further due to the very short dwell times in the trajectories. Weak binding was also observed with ARG158 and ARG237 along the NTD, but the binding pocket was shallow and dissociation quickly occurred during the equilibration of the Sias-S protein complex. Binding was also observed to ARG328 along the outside of the S below the ACE-2 RBD, but further simulations also showed the site as shallow and a weak binder.

## **Conclusion**

We have identified multiple weak sialic acid binding sites on the SARS-CoV-2 S protein, with two novel Sia-binding poses. Each novel pose, between the NTD and ACE-2 RBD (chains A and C) and between the ACE-2 RBD and S2 domain (chains A and B), is accessible to gangliosides for S

attachment, and overlaps with regions known in stabilizing the S protein. The binding in each pose is predicted to be weak, in overall agreement with glycan microarray experiments.<sup>125,138</sup> We validated our methods using MERS-CoV-S, confirming the experimental result that Neu5Ac binds stronger than that with either 9-OAc or 9-NAc modification, and binding energies comparable to Sias-SARS-CoV-2 S protein binding. Each Sias-binding pose in SARS-CoV-2 S protein contains an arginine residue that is conserved across SARS-CoV-2 variants. The multiple Sia-binding sites on SARS-CoV-2 S protein may lead to increased binding affinity to multiple Sias collocated on the cell surface, and the existence of multiple binding sites to the S protein may be validated experimentally. Neu5Ac in SARS-CoV-2 binding sites tend to have stronger binding than their 9-OAc or 9-NAc modifications, but the differences are within the expected margin of error. Binding free energies support 9-NAc Sias as close structural and chemical mimics of 9-OAc Sias in SARS-CoV-2 and MERS-CoV S proteins, given the small energy differences (all <1 kcal/mol with SOMD). This supports 9-NAc Sias as an experimentally stable mimic of 9-OAc counterparts to probe Sias-viral binding. Future experimental studies can validate our differential binding free energy results and understanding the binding of modified Sias can further elucidate the role of Sias in cellular recognition and the high transmissibility of SARS-CoV-2.

## **Methods**

### **Choice of protein structures, sialic acids, and binding poses**

The initial protein-ligand complex for molecular dynamics simulations was chosen based on available structures from the protein data bank, ideally a protein with a sialic acid bound in a

known receptor site. At the initiation of this project, there were no SARS-CoV-2 spike proteins bound to sialic acids or suspected sialic acid binding. As such, we searched for other human beta coronavirus spike protein structures available on the protein data bank with a bound sialic acid. While the cryo-EM SARS-CoV-2 spike protein (PDB ID: 6VSB)<sup>161</sup> has a 73% sequence identity with SARS-CoV,<sup>58</sup> there is no known sialic acid binding or bound complex with SARS-CoV. MERS-CoV spike protein (PDB ID: 6Q04)<sup>79</sup> has a 35% identity with SARS-CoV-2 spike protein, and does have an available structure with a bound sialic acid (Neu5Ac).

To generate an initial sialic acid binding pose with the SARS-CoV-2 spike protein, we performed sequence alignment with both spike proteins (6VSB and 6Q04) using MultiSeq,<sup>153</sup> an extension of the Multiple Alignment tool available in VMD,<sup>154</sup> a freely available structural graphics program for visualization and analysis. From the aligned structures and conserved binding residues, notably arginine, we placed the sialic acid from MERS-CoV spike protein directly into the overlaid SARS-CoV-2 spike protein N-terminal domain as our initial ligand-protein complex, each without sialylation of the spike proteins (Fig. 4.2).

For additional ligand-protein complexes in both spike proteins, we docked Sias (Fig. 4.1) into binding pockets using OEDocking.<sup>162-165</sup> Given other beta coronavirus virus (OC43 and HKUI)<sup>57,58</sup> preference to Neu5,9Ac<sub>2</sub> over Neu5Ac, and that *N*-acetylated sialic acids have shown to be chemically and structurally reasonable mimics to *O*-acetylated counterparts, we docked Neu5Ac, Neu5,9Ac<sub>2</sub>, and Neu5Ac9NAc into each spike protein. Given that MERS-CoV spike protein preferentially binds to  $\alpha$ 2,3 over  $\alpha$ 2,6-linked sialic acids, we docked  $\alpha$ 2,3-linked sialic acids (Neu5Ac $\alpha$ 2-3Gal $\beta$ pNP, Neu5,9Ac<sub>2</sub> $\alpha$ 2-3Gal $\beta$ pNP, and Neu5Ac9NAc $\alpha$ 2-3Gal $\beta$ pNP) into deeper

binding pockets of SARS-CoV-2 spike protein, where these new poses were observed from molecular dynamics simulations where a free sialic acid would associate to various residues along the spike trimer.

### **Molecular dynamics simulations and MM-PBSA setup and procedure**

Molecular dynamics simulations were performed using AMBER18 software suite, using *tleap* for setup and *pmemd.cuda* for dynamics on NVIDIA GTX 980 Ti or 1080 Ti hardware.<sup>166–170</sup> We used the ff14SB and GAFF protein force fields, GLYCAM06 carbohydrate forcefield, TIP3P water model. Protein or protein-ligand complexes were solved in a truncated octahedron box with 12.0 Å distance, resulting in around 120,000 water molecules, and Na<sup>+</sup> ions were added for a net neutral charge. Simulations were run with periodic boundary conditions, a 2.0 fs time step and Langevin thermostat set to 298.15 K and a collision frequency of 5.0 ps<sup>-1</sup>. The particle mesh Ewald method was used to treat long-range electrostatics with a real-space cutoff of 9.0 Å and the SHAKE algorithm to constrain all bonds involving hydrogen. 1000 minimization steps were performed with restraints on complex heavy atoms with a restraint weight of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>, followed by an additional 500 steps with no constraints to minimize the entire system. We subsequently heated our system linearly from 0.1 K to 298.15 K at constant pressure molecular dynamics, using a Berendsen barostat set to 1.0 atm, and a compressibility of 44.6 x 10<sup>-6</sup> bar<sup>-1</sup>, NMR restraints on the protein complex heavy atoms with a restraint weight of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>, and a timescale of 500 ps. Subsequent equilibration was performed at constant pressure over 5 steps, each at 100 ns. Restraint on protein-ligand complex heavy atoms decreased in order of 10.0, 1.0, 0.1, 0.01 and 0 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Production simulations were performed with the NVT ensemble.

Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach to binding energies was also computed from the initial molecular dynamics simulations and the equilibration runs for the alchemical free energy simulations.<sup>171-173</sup> Key interacting protein residues are identified by energy decomposition analysis (EDA). MM-PBSA can give comparative estimates on absolute binding free energies using an end-state free-energy approach, where the binding free energy is the difference between the sum of ligand-protein complex binding and its solvation, and the sum of the solvation energies of the free ligand and protein (Fig. S4.3). Implicit solvation in the thermodynamic cycle calculations reduces computational time compared to the explicit water molecules used for solvation in the above alchemical free energy approach. Difference in binding free energies were computed from binding free energies. Calculations were performed using MMPBSA.py<sup>155</sup> AMBER software, from molecular dynamics simulation trajectories. Trajectories (in triplicate) were sampled over 300 ns, depending on retention of Sias in the binding pocket. This was deemed sufficient time for relative binding free energies, as our total simulation lengths sampled in our MM-PBSA calculations are greater than the 1-10 ns range tested as sufficient in a paper studying MM-PBSA protein-ligand binding free energy convergences from crystal structures.<sup>171</sup> The importance of the initial structure is especially important in our case, given the flexibility of glycans and the SARS-CoV-2 S protein. When performing initial docking and simulations of Sias directly into the Cryo-EM SARS-CoV-2 S protein structure, Sias would quickly unbind. We then equilibrated the S protein, took select structures from the MD, and minimized each as receptor structures to dock Sias and run MD. This resulted in longer simulation lengths prior to Sias unbinding. Given the ligand and protein flexibilities, triplicate simulations were run to increase total simulation time while maintaining the binding pose.

## Alchemical free energy simulations setup and procedure

Alchemical free energy simulations were setup in a similar fashion, but employed the automated Sire FESetup<sup>174</sup> tool to prepare systems for alchemical free energy simulations using free and open source Sire and OpenMM (SOMD).<sup>85,175</sup> To prepare the system, we specified sets of ligands with docked poses to respective receptors, specific ligand GLYCAM parameters and forcefield parameters described above. Different from the above procedure, we used a rectangular box with a 12.0 Å distance, performed minimization over 300 steps, heated the system from 5.0 K to 298.15 K, and ran equilibration simulations decreasing constraints in order of 8.0, 6.0, 4.0, 2.0, and 0 kcal mol<sup>-1</sup> Å<sup>-2</sup>, where the first equilibration ran for 100 ps (and subsequent simulations ran for 100 ns, as in the above case). We again used FESetup to generate the alchemical free energy simulations perturbation input describing the transformation occurring (such as Neu5Ac to Neu5,9Ac<sub>2</sub>), dummy atoms to transform into additional atoms required in the transformation, and the respective simulation parameters for each atom pair or transforming pair between ligands. Soft-core potentials were used to describe electrostatic and van der Waals interactions according to the following equations:<sup>90</sup>

$$r_{LJ}=(2\sigma_{ij}\lambda+r_{ij}^2)^{1/2}$$

$$r_{Coul}=(\lambda+r_{ij}^2)^{1/2}$$

Alchemical free energy simulations were performed to compute the difference in binding free energies of multiple sialic acids to spike protein receptors using the SOMD method.<sup>85,175</sup> Ligands were transformed (ex: Neu5,9Ac<sub>2</sub> to Neu5Ac9NAc) over a number of artificial intermediate states

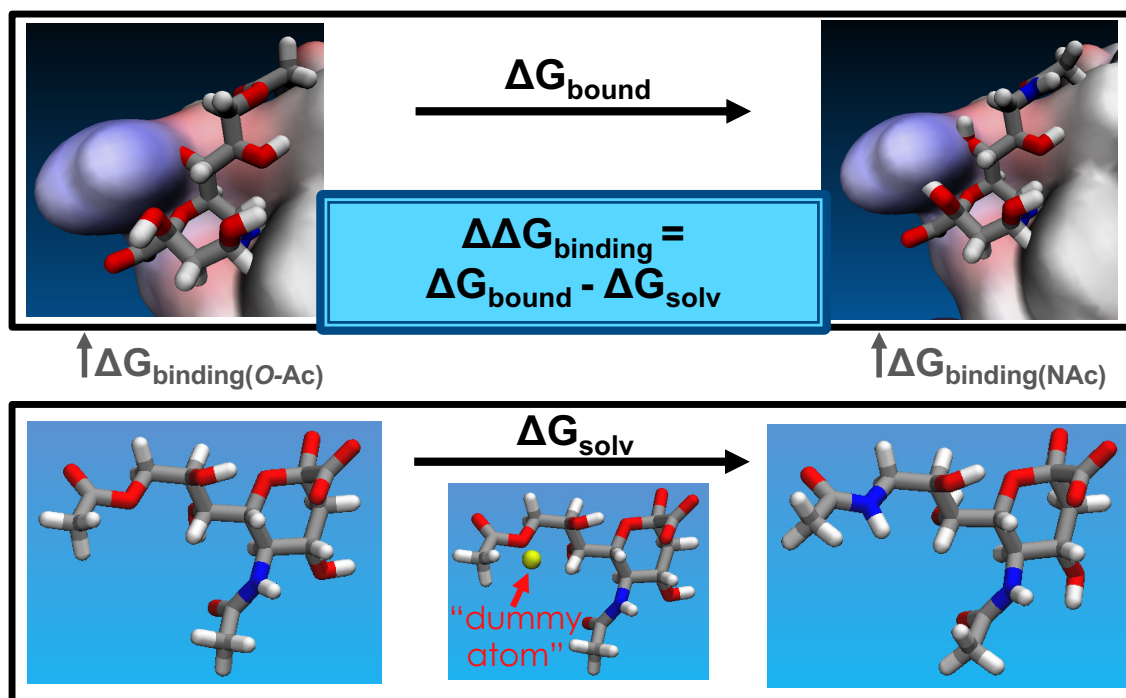
(windows) created to connect the thermodynamic ensembles of the two physical states of interest (Fig. 4.5). We chose 13-21 linearly-spaced  $\lambda$  windows, depending on the transformation, based on a paper using SOMD and other alchemical free energy methods in an assessment of binding affinities,<sup>92</sup> and following the scheme outlined in the SI of Loeffler et al.<sup>90</sup> In general, the first replicate for each transformation was ran with 13 windows, and the duplicate with 21 for greater certainty. Discretion must be used when choosing the  $\lambda$  value, as a value too small can result in insufficient thermodynamic overlap between simulations, whereas a value too large results in excessive simulation time. In my simulation cases, increasing  $\lambda$  to 21 did not immediately result in improved simulation results. Rather, a lower  $\lambda$  with multiple replicates would likely yield a more rigorous relative binding free energy value with an error estimate, and allow for discarding of simulation data without drastically reducing sampling size. However, multiple replicate alchemical simulations is indeed quite expensive and potentially time-prohibitive.

Additional to varying the  $\lambda$  value in simulations, the duration of simulations at each  $\lambda$  step must be chosen. Each simulation was minimized over 1000 steps prior to 5 ns production run in the NPT ensemble, where free energies were computed from the last 4 ns, based on alchemical free energy simulation best practices.<sup>89,92</sup> Quality of simulations did not necessarily improve with increased simulation time at each step. This is likely due to the flexibility of our Sia ligands and S proteins. Instead, confidence improved with additional replicates rather than doubling individual simulation times. Additionally, but not required, each transformation was computed in the reverse direction, and duplicate simulations were performed. For example, Neu5,9Ac<sub>2</sub> was transformed to Neu5Ac9NAc by slowly turning off the interactions of Neu5,9Ac<sub>2</sub> and slowly turning them on for Neu5Ac9NAc, over a set of discrete steps, as described the SI of Loeffler et al.<sup>90</sup>



We used *analyse\_frenerg* to calculate free energy differences at 298 K using multistate Bennett acceptance ratio (MBAR) and thermodynamic integration (TI) as a reference estimate.<sup>86-88</sup> The relative free energy of binding was then computed as the free energy difference of the alchemical transformation of the ligand bound to the protein and the ligand in solution, i.e. water.

Reliability of results was determined based on similarity between MBAR and TI, and from overlap matrices, where the first off-diagonal element is at least 0.03,<sup>89</sup> but ideally close to 0.10 to ensure enough thermodynamic overlap between simulation windows. To determine the reliability of alchemical free energy simulation results, we compared values from MBAR and TI, as well as analyzing overlap matrices (Fig. S4.4). Some test cases of longer simulations were ran, but the thermodynamic overlap did not significantly improve compared to more windows. The high degree of flexibility known for the SARS-CoV-2 spike protein may contribute to this observation. Thus, we increased the number of windows when necessary, and duplicate simulations were run for higher certainty and increased sampling. Solvated ligand simulations each took about 45 min, and the solvated ligand-protein complex calculations took about 1 day each, running on NVIDIA GTX 980 Ti or 1080 Ti hardware.



**Figure 4.5:** Visual representation of  $\Delta\Delta G_{\text{binding}}$  of Neu5Ac9NAc versus Neu5,9Ac<sub>2</sub> as the difference of two alchemical transformations in explicit solvent (boxed). The upper box is the transformation from Neu5,9Ac<sub>2</sub> to Neu5Ac9NAc bound in the protein and solvated in explicit water. The lower box is the transformation in explicit water. Additional atoms required in transformations are represented by “dummy atoms” that slowly turn on over the simulation steps.

Trajectory analyses were carried out using the *cpptraj*<sup>176</sup> simulation analysis package and free energy differences plots generated with *matplotlib* plotting software package in python. Molecular structures were visualized with VMD<sup>154</sup>.

Structures used in molecular dynamics simulations: 6VSB; 6Q04, each without sialylation

## Sialic acid parametrization procedure

In order to carry out simulations of diverse Sia modifications, we used a computational protocol developed in our group to generate molecular dynamics simulation parameters for modified carbohydrate residues that have excellent compatibility with the widely used GLYCAM06 force field model;<sup>2</sup> this was originally developed for the N-Ac Sias.

We parametrized our ligands to match GLYCAM format, following GLYCAM protocol and Neu5,9Ac<sub>2</sub> and Neu5Ac9NAc results from our previous publication.<sup>69,177</sup> The MD simulations with sialosides used a combination of the GLYCAM06 force field for the monosaccharide building blocks and GAFF<sup>178</sup> for the para-nitrophenol (*p*NP) aglycon, with customized parameters to describe the *p*NP in a consistent manner. The procedure for deriving the new simulation parameters was designed to be maximally consistent with the original GLYCAM procedure.

A set of atomic partial charges was derived for *p*NP. The procedure involves averaging over an ensemble of structures, sampled using a MD simulation of Gal $\beta$ *p*NP in TIP3P water. The system was equilibrated for 1 ns at 298.15 K, 1.0 atm (NPT), followed by a 100 ns production run at 298.15 K (NVT). One hundred structures of Gal $\beta$ *p*NP were saved at 2 ns intervals. Each of these structures was energy-minimized at the HF/6-31G\* level of theory with all exocyclic dihedral angles constrained to their MD- sampled values.

For each the 100 constrained energy-minimized structures, electrostatic potential (ESP) calculations were computed on a rectilinear grid of points using the Gaussian16 software<sup>179</sup> using

the Pop = chelpg option in the route file. This ESP data was used in a single-stage restrained ESP fitting (RESP) calculation with a restraint weight of 0.01 applied to all atoms using the resp program from AmberTools and with charges on the OMe group fixed to the original GLYCAM values. An arithmetic average over the 100 sets of fitted charges yielded the final set of charges in the model.

The GLYCAM and GAFF force fields lacked torsional parameters to describe the torsional energy profiles about the *p*NP functional group and Gal-*p*NP linkage; these parameters were derived by fitting to reproduce torsional profiles from DFT. In order to generate data for fitting the torsional profiles, constrained optimized geometries were obtained at the HF/6-31G\* level of theory with the selected torsional degree of freedom constrained to values on a 30-degree grid. We used the TeraChem quantum chemistry software for the energy minimizations and the *torsiondrive* software to scan over the dihedral angles recursively.<sup>31,32,46,47,180</sup> The optimized geometries were used for single-point energies and atomistic forces calculated at the  $\omega$ B97X-D3/6-31++G(2d,2p) level of theory.<sup>181</sup> We chose a different DFT functional from what was used in the original GLYCAM procedure because dispersion-corrected DFT functionals have become the standard for accurate conformational energy calculations over the past decade.<sup>38,179,181,182</sup> We also do not expect differences in the torsional parameterization procedure to cause incompatibilities with the rest of the GLYCAM force field, because the effect of parameter optimization on the potential energy surface is localized to the glycerol chain. The parameters were optimized by fitting to the quantum chemical energies using the ForceBalance optimization software.<sup>180,183</sup> Bond stretching and angle bending parameters for the Gal-*p*NP linkage were copied from analogous parameters available in GLYCAM06.

## **Acknowledgements**

We are grateful for the insightful discussions with Caleb Oh at the UC Davis Medical Center. This work was supported by the United States National Institutes of Health grant no. R01AI130684.

## **Chapter 5: Summary and Future Work**

In this dissertation, I have applied the *ab initio* nanoreactor method to solve mechanistic mysteries in green and biological chemistry (Chapters 2 and 3). In one case, the nanoreactor guided the mechanistic search and inspired elementary radical steps, and through a synergistic effort between theory and experiment, solved the mechanism to form 3-HPPA from LA, and unlocked a new class of reactions. Future work to support this mechanism could involve EPR experiments of trapped radicals on labelled starting materials, with results invaluable to tuning this reaction to alternative products and milder conditions. In another application, I used the nanoreactor to study the lability of the *O*-acetyl modification of Sias at the hydroxyl group of C7, C8, or C9 on the glycerol-like side chain, lending insights into the defensive role of Sias against pathogens that evolve to gain entry to the cell by binding selectively to one acetylation state. Further work is necessary to probe the role of *O*-acetyl migration in viral recognition and pathogenesis. Given the above precedence where this nanoreactor method has aided in mechanistic searches, next steps could include using the nanoreactor with larger starting materials and probing reactions involving two relatively large species. Given the high energy nature of the nanoreactor and its automated nature, it could be used to predict reaction mechanisms and rates in combustion chemistry. Additional applications would elucidate the nanoreactor's scope and parameter sensitivity to specific chemistries.

In Chapter 4, I used simulations to probe the binding of Sia to MERS-CoV S and SARS-CoV-2 S proteins and determined that *N*-acetyl and *O*-acetyl Sias interchangeable, suggesting an experimentally reasonable mimic to probe viral mechanisms. Future work needs to be done to probe modified Sias-protein binding in other Sia-binding proteins and validate differential binding trends between Sias. Additionally, repeating similar simulations in other viral systems may suggest trends of top-interacting protein residues to Sias additional to arginine. Finally, future steps could

involve ligand-protein binding simulations, where the protein is sialylated, which may affect Sia affinity, and the use of Sias with alternative linkages, variable sialoside length, and modifications at other locations (such as C4, C7, C8 of Neu5Ac).



## References

- (1) Ji, Y.; Sasmal, A.; Li, W.; Oh, L.; Srivastava, S.; Hargett, A. A.; Wasik, B. R.; Yu, H.; Diaz, S.; Choudhury, B.; Parrish, C. R.; Freedberg, D. I.; Wang, L.-P.; Varki, A.; Chen, X. Reversible O-Acetyl Migration within the Sialic Acid Side Chain and Its Influence on Protein Recognition. *ACS Chem. Biol.* **2021**, *16* (10), 1951–1960. <https://doi.org/10.1021/acscchembio.0c00998>.
- (2) Li, W.; Battistel, M. D.; Reeves, H.; Oh, L.; Yu, H.; Chen, X.; Wang, L.-P.; Freedberg, D. I. A Combined NMR, MD and DFT Conformational Analysis of 9-O-Acetyl Sialic Acid-Containing GM3 Ganglioside Glycan and Its 9-N-Acetyl Mimic. *Glycobiology* **2020**, *30* (10), 787–801. <https://doi.org/10.1093/glycob/cwaa040>.
- (3) Li, C.-J.; Trost, B. M. Green Chemistry for Chemical Synthesis. *Proc. Natl. Acad. Sci.* **2008**, *105* (36), 13197–13202. <https://doi.org/10.1073/pnas.0804348105>.
- (4) Cue, B. W.; Zhang, J. Green Process Chemistry in the Pharmaceutical Industry. *Green Chem. Lett. Rev.* **2009**, *2* (4), 193–211. <https://doi.org/10.1080/17518250903258150>.
- (5) Corma, A.; Iborra, S.; Velty, A. Chemical Routes for the Transformation of Biomass into Chemicals. *Chem. Rev.* **2007**, *107* (6), 2411–2502. <https://doi.org/10.1021/cr050989d>.
- (6) Mäki-Arvela, P.; Holmbom, B.; Salmi, T.; Murzin, D. Yu. Recent Progress in Synthesis of Fine and Specialty Chemicals from Wood and Other Biomass by Heterogeneous Catalytic Processes. *Catal. Rev.* **2007**, *49* (3), 197–340. <https://doi.org/10.1080/01614940701313127>.
- (7) Sheldon, R. A. Green and Sustainable Manufacture of Chemicals from Biomass: State of the Art. *Green Chem* **2014**, *16* (3), 950–963. <https://doi.org/10.1039/C3GC41935E>.
- (8) Donate, P. M. Green Synthesis from Biomass. *Chem. Biol. Technol. Agric.* **2014**, *1* (1), 4. <https://doi.org/10.1186/s40538-014-0004-2>.
- (9) Gallezot, P. Conversion of Biomass to Selected Chemical Products. *Chem Soc Rev* **2012**, *41* (4), 1538–1558. <https://doi.org/10.1039/C1CS15147A>.
- (10) Werpy, T.; Petersen, G. *Top Value Added Chemicals from Biomass: Volume I – Results of Screening for Potential Candidates from Sugars and Synthesis Gas*; Office of Scientific and Technical Information (OSTI), 2004. <https://doi.org/10.2172/15008859>.
- (11) Cheng, G. J.; Zhang, X.; Chung, L. W.; Xu, L.; Wu, Y. D. Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chemical Reactions. *J Am Chem Soc* **2015**, *137*, 1706–1725. <https://doi.org/10.1021/ja5112749>.
- (12) Martínez, T. J. Ab Initio Reactive Computer Aided Molecular Design. *Acc. Chem. Res.* **2017**, *50* (3), 652–656. <https://doi.org/10.1021/acs.accounts.7b00010>.
- (13) Stewart, J. J. P. Application of the PM6 Method to Modeling Proteins. *J. Mol. Model.* **2009**, *15* (7), 765–805. <https://doi.org/10.1007/s00894-008-0420-y>.
- (14) Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140* (18), 18A301. <https://doi.org/10.1063/1.4869598>.
- (15) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50* (3), 539–543. <https://doi.org/10.1021/acs.accounts.6b00532>.
- (16) Pietrucci, F.; Andreoni, W. Graph Theory Meets Ab Initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale. *Phys. Rev. Lett.* **2011**, *107* (8), 085504. <https://doi.org/10.1103/PhysRevLett.107.085504>.

- (17) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering Chemistry with an *Ab Initio* Nanoreactor. *Nat. Chem.* **2014**, *6* (12), 1044–1048. <https://doi.org/10.1038/nchem.2099>.
- (18) Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M. Heuristics-Guided Exploration of Reaction Mechanisms. *J. Chem. Theory Comput.* **2015**, *11* (12), 5712–5722. <https://doi.org/10.1021/acs.jctc.5b00866>.
- (19) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways. *J. Chem. Theory Comput.* **2016**, *12* (2), 638–649. <https://doi.org/10.1021/acs.jctc.5b00830>.
- (20) Lapkin, A. A.; Heer, P. K.; Jacob, P.-M.; Hutchby, M.; Cunningham, W.; Bull, S. D.; Davidson, M. G. Automation of Route Identification and Optimisation Based on Data-Mining and Chemical Intuition. *Faraday Discuss.* **2017**, *202* (0), 483–496. <https://doi.org/10.1039/C7FD00073A>.
- (21) Varela, J. A.; Vázquez, S. A.; Martínez-Núñez, E. An Automated Method to Find Reaction Mechanisms and Solve the Kinetics in Organometallic Catalysis. *Chem. Sci.* **2017**, *8* (5), 3843–3851. <https://doi.org/10.1039/C7SC00549K>.
- (22) Jafari, M.; Zimmerman, P. M. Uncovering Reaction Sequences on Surfaces through Graphical Methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (11), 7721–7729. <https://doi.org/10.1039/C8CP00044A>.
- (23) Plehiers, P. P.; Marin, G. B.; Stevens, C. V.; Van Geem, K. M. Automated Reaction Database and Reaction Network Analysis: Extraction of Reaction Templates Using Cheminformatics. *J. Cheminformatics* **2018**, *10* (1), 11. <https://doi.org/10.1186/s13321-018-0269-8>.
- (24) Unsleber, J. P.; Reiher, M. The Exploration of Chemical Reaction Networks. *Annu. Rev. Phys. Chem.* **2020**, *71* (1), 121–142. <https://doi.org/10.1146/annurev-physchem-071119-040123>.
- (25) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8* (2), e1354. <https://doi.org/10.1002/wcms.1354>.
- (26) Maeda, S.; Harabuchi, Y.; Takagi, M.; Saita, K.; Suzuki, K.; Ichino, T.; Sumiya, Y.; Sugiyama, K.; Ono, Y. Implementation and Performance of the Artificial Force Induced Reaction Method in the GRRM17 Program. *J. Comput. Chem.* **2018**, *39* (4), 233–251. <https://doi.org/10.1002/jcc.25106>.
- (27) Schmitz, G.; Yönder, Ö.; Schnieder, B.; Schmid, R.; Hättig, C. An Automated Workflow from Molecular Dynamic Simulation to Quantum Chemical Methods to Identify Elementary Reactions and Compute Reaction Constants. *J. Comput. Chem.* **2021**, *42* (32), 2264–2282. <https://doi.org/10.1002/jcc.26757>.
- (28) Yang, M.; Zou, J.; Wang, G.; Li, S. Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method. *J. Phys. Chem. A* **2017**, *121* (6), 1351–1361. <https://doi.org/10.1021/acs.jpca.6b12195>.
- (29) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003**, *90* (23), 238302. <https://doi.org/10.1103/PhysRevLett.90.238302>.
- (30) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164. <https://doi.org/10.1002/jcc.21224>.

- (31) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4* (2), 222–231. <https://doi.org/10.1021/ct700268q>.
- (32) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5* (4), 1004–1015. <https://doi.org/10.1021/ct800526s>.
- (33) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619–2628. <https://doi.org/10.1021/ct9003004>.
- (34) Liu, F.; Sanchez, D. M.; Kulik, H. J.; Martínez, T. J. Exploiting Graphical Processing Units to Enable Quantum Chemistry Calculation of Large Solvated Molecules with Conductor-like Polarizable Continuum Models. *Int. J. Quantum Chem.* **2019**, *119* (1), e25760. <https://doi.org/10.1002/qua.25760>.
- (35) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J Chem Soc Perkin Trans 2* **1993**, No. 5, 799–805. <https://doi.org/10.1039/P29930000799>.
- (36) Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015**, *11* (7), 3131–3144. <https://doi.org/10.1021/acs.jctc.5b00370>.
- (37) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010**, *133* (24), 244111. <https://doi.org/10.1063/1.3511297>.
- (38) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections. *J. Chem. Theory Comput.* **2013**, *9* (1), 263–272. <https://doi.org/10.1021/ct300715s>.
- (39) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. <https://doi.org/10.1063/1.3382344>.
- (40) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. <https://doi.org/10.1002/jcc.21759>.
- (41) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient Exploration of Reaction Paths via a Freezing String Method. *J. Chem. Phys.* **2011**, *135* (22), 224108. <https://doi.org/10.1063/1.3664901>.
- (42) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.* **2004**, *120* (17), 7877–7886. <https://doi.org/10.1063/1.1691018>.
- (43) Zimmerman, P. M. Growing String Method with Interpolation and Optimization in Internal Coordinates: Method and Examples. *J. Chem. Phys.* **2013**, *138* (18), 184102. <https://doi.org/10.1063/1.4804162>.
- (44) Jonsson, H.; Mills, G.; Jacobsen, K. W. CHAPTER 16 Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. 20.
- (45) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113* (22), 9901–9904. <https://doi.org/10.1063/1.1329672>.

- (46) Wang, L.-P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108. <https://doi.org/10.1063/1.4952956>.
- (47) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619–2628. <https://doi.org/10.1021/ct9003004>.
- (48) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9* (1), 213–221. <https://doi.org/10.1021/ct300321a>.
- (49) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscammann, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215. <https://doi.org/10.1080/00268976.2014.952696>.
- (50) Schauer, R. *Glycoconj. J.* **2000**, *17* (7/9), 485–499. <https://doi.org/10.1023/a:1011062223612>.
- (51) Klein, A.; Roussel, P. O-Acetylation of Sialic Acids. *Biochimie* **1998**, *80* (1), 49–57. [https://doi.org/10.1016/s0300-9084\(98\)80056-4](https://doi.org/10.1016/s0300-9084(98)80056-4).
- (52) Varki, N. M.; Varki, A. Diversity in Cell Surface Sialic Acid Presentations: Implications for Biology and Disease. *Lab. Invest.* **2007**, *87* (9), 851–857. <https://doi.org/10.1038/labinvest.3700656>.

- (53) Angata, T.; Varki, A. Chemical Diversity in the Sialic Acids and Related  $\alpha$ -Keto Acids: An Evolutionary Perspective. *Chem. Rev.* **2002**, *102* (2), 439–470. <https://doi.org/10.1021/cr000407m>.
- (54) Chen, X.; Varki, A. Advances in the Biology and Chemistry of Sialic Acids. *ACS Chem. Biol.* **2010**, *5* (2), 163–176. <https://doi.org/10.1021/cb900266r>.
- (55) Vyas, A. A.; Schnaar, R. L. Brain Gangliosides: Functional Ligands for Myelin Stability and the Control of Nerve Regeneration. *Biochimie* **2001**, *83* (7), 677–682. [https://doi.org/10.1016/s0300-9084\(01\)01308-6](https://doi.org/10.1016/s0300-9084(01)01308-6).
- (56) Zheng, C.; Terreni, M.; Sollogoub, M.; Zhang, Y. Ganglioside GM3 and Its Role in Cancer. *Curr. Med. Chem.* **2019**, *26* (16), 2933–2947. <https://doi.org/10.2174/0929867325666180129100619>.
- (57) Hulswit, R. J. G.; Lang, Y.; Bakkers, M. J. G.; Li, W.; Li, Z.; Schouten, A.; Ophorst, B.; van Kuppeveld, F. J. M.; Boons, G.-J.; Bosch, B.-J.; Huizinga, E. G.; de Groot, R. J. Human Coronaviruses OC43 and HKU1 Bind to 9-O-Acetylated Sialic Acids via a Conserved Receptor-Binding Site in Spike Protein Domain A. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (7), 2681–2690. <https://doi.org/10.1073/pnas.1809667116>.
- (58) Tortorici, M. A.; Walls, A. C.; Lang, Y.; Wang, C.; Li, Z.; Koerhuis, D.; Boons, G.-J.; Bosch, B.-J.; Rey, F. A.; de Groot, R. J.; Veesler, D. Structural Basis for Human Coronavirus Attachment to Sialic Acid Receptors. *Nat. Struct. Mol. Biol.* **2019**, *26* (6), 481–489. <https://doi.org/10.1038/s41594-019-0233-y>.
- (59) Schauer, R. Chemistry, Metabolism, and Biological Functions of Sialic Acids. In *Advances in Carbohydrate Chemistry and Biochemistry*; Elsevier, 1982; pp 131–234. [https://doi.org/10.1016/s0065-2318\(08\)60109-2](https://doi.org/10.1016/s0065-2318(08)60109-2).
- (60) Klotz, F. W.; Orlandi, P. A.; Reuter, G.; Cohen, S. J.; Haynes, J. D.; Schauer, R.; Howard, R. J.; Palese, P.; Miller, L. H. Binding of Plasmodium Falciparum 175-Kilodalton Erythrocyte Binding Antigen and Invasion of Murine Erythrocytes Requires N-Acetylneuraminic Acid but Not Its O-Acetylated Form. *Mol. Biochem. Parasitol.* **1992**, *51* (1), 49–54. [https://doi.org/10.1016/0166-6851\(92\)90199-t](https://doi.org/10.1016/0166-6851(92)90199-t).
- (61) Deszo, E. L.; Steenbergen, S. M.; Freedberg, D. I.; Vimr, E. R. *Escherichia Coli* K1 Polysialic Acid *O*-Acetyltransferase Gene, *neuO*, and the Mechanism of Capsule Form Variation Involving a Mobile Contingency Locus. *Proc. Natl. Acad. Sci.* **2005**, *102* (15), 5564–5569. <https://doi.org/10.1073/pnas.0407428102>.
- (62) Berti, F.; Ricco, R. D.; Rappuoli, R. Role of O-Acetylation in the Immunogenicity of Bacterial Polysaccharide Vaccines. *Molecules* **2018**, *23* (6), 1340. <https://doi.org/10.3390/molecules23061340>.
- (63) Varki, A.; Kornfeld, S. An Autosomal Dominant Gene Regulates the Extent of 9-O-Acetylation of Murine Erythrocyte Sialic Acids. A Probable Explanation for the Variation in Capacity to Activate the Human Alternate Complement Pathway. *J. Exp. Med.* **1980**, *152* (3), 532–544. <https://doi.org/10.1084/jem.152.3.532>.
- (64) Shi, W.-X.; Chammas, R.; Varki, N. M.; Powell, L.; Varki, A. Sialic Acid 9-O-Acetylation on Murine Erythrocyte Cells Affects Complement Activation, Binding to I-Type Lectins, and Tissue Homing. *J. Biol. Chem.* **1996**, *271* (49), 31526–31532. <https://doi.org/10.1074/jbc.271.49.31526>.

- (65) Sjöberg, E. R.; Powell, L. D.; Klein, A.; Varki, A. Natural Ligands of the B Cell Adhesion Molecule CD22 Beta Can Be Masked by 9-O-Acetylation of Sialic Acids. *J. Cell Biol.* **1994**, *126* (2), 549–562. <https://doi.org/10.1083/jcb.126.2.549>.
- (66) Schauer, R.; Kamerling, J. P. Exploration of the Sialic Acid World. In *Advances in Carbohydrate Chemistry and Biochemistry*; Elsevier, 2018; pp 1–213. <https://doi.org/10.1016/bs.accb.2018.09.001>.
- (67) Park. Post-Glycosylation Modification of Sialic Acid and Its Role in Virus Pathogenesis. *Vaccines* **2019**, *7* (4), 171. <https://doi.org/10.3390/vaccines7040171>.
- (68) Varki, A.; Schauer, R. Sialic Acids. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2009.
- (69) Khedri, Z.; Xiao, A.; Yu, H.; Landig, C. S.; Li, W.; Diaz, S.; Wasik, B. R.; Parrish, C. R.; Wang, L.-P.; Varki, A.; Chen, X. A Chemical Biology Solution to Problems with Studying Biologically Important but Unstable 9-O-Acetyl Sialic Acids. *ACS Chem. Biol.* **2017**, *12* (1), 214–224. <https://doi.org/10.1021/acscchembio.6b00928>.
- (70) Herrler, G.; Rott, R.; Klenk, H. D.; Müller, H. P.; Shukla, A. K.; Schauer, R. The Receptor-Destroying Enzyme of Influenza C Virus Is Neuraminidase-O-Acetyltransferase. *EMBO J.* **1985**, *4* (6), 1503–1506. <https://doi.org/10.1002/j.1460-2075.1985.tb03809.x>.
- (71) Schauer, R. [6] Characterization of Sialic Acids. In *Methods in Enzymology*; Elsevier, 1978; pp 64–89. [https://doi.org/10.1016/0076-6879\(78\)50008-6](https://doi.org/10.1016/0076-6879(78)50008-6).
- (72) KAMERLING, J. P.; SCHAUER, R.; SHUKLA, A. K.; STOLL, S.; HALBEEK, H.; VLIAGENTHART, J. Migration of O-Acetyl Groups in N,O-Acetylneuraminic Acids. *Eur. J. Biochem.* **1987**, *162* (3), 601–607. <https://doi.org/10.1111/j.1432-1033.1987.tb10681.x>.
- (73) Muñoz-Barroso, I.; García-Sastre, A.; Villar, E.; Manuguerra, J.-C.; Hannoun, C.; Cabezas, J. A. Increased Influenza A Virus Sialidase Activity with N-Acetyl-9-O-Acetylneuraminic Acid-Containing Substrates Resulting from Influenza C Virus O-Acetyltransferase Action. *Virus Res.* **1992**, *25* (1–2), 145–153. [https://doi.org/10.1016/0168-1702\(92\)90106-j](https://doi.org/10.1016/0168-1702(92)90106-j).
- (74) Muchmore, E. A.; Varki, A. Selective Inactivation of Influenza C Esterase: A Probe for Detecting 9- $\text{O}$ -Acetylated Sialic Acids. *Science* **1987**, *236* (4806), 1293–1295. <https://doi.org/10.1126/science.3589663>.
- (75) Song, H.; Qi, J.; Khedri, Z.; Diaz, S.; Yu, H.; Chen, X.; Varki, A.; Shi, Y.; Gao, G. F. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLOS Pathog.* **2016**, *12* (1), e1005411. <https://doi.org/10.1371/journal.ppat.1005411>.
- (76) Barnard, K. N.; Wasik, B. R.; LaClair, J. R.; Buchholz, D. W.; Weichert, W. S.; Alford-Lawrence, B. K.; Aguilar, H. C.; Parrish, C. R. Expression of 9- $\text{O}$ -Acetyl and 7,9- $\text{O}$ -Acetyl Modified Sialic Acid in Cells and Their Effects on Influenza Viruses. *mBio* **2019**, *10* (6). <https://doi.org/10.1128/mbio.02490-19>.
- (77) Liu, R.; Sreenivasan, C.; Yu, H.; Sheng, Z.; Newkirk, S. J.; An, W.; Smith, D. F.; Chen, X.; Wang, D.; Li, F. Influenza D Virus Diverges from Its Related Influenza C Virus in the Recognition of 9-O-Acetylated N-Acetyl- or N-Glycolyl-Neuraminic Acid-Containing Glycan Receptors. *Virology* **2020**, *545*, 16–23. <https://doi.org/10.1016/j.virol.2020.02.007>.

- (78) Weiman, S.; Dahesh, S.; Carlin, A. F.; Varki, A.; Nizet, V.; Lewis, A. L. Genetic and Biochemical Modulation of Sialic Acid O-Acetylation on Group B Streptococcus: Phenotypic and Functional Impact. *Glycobiology* **2009**, *19* (11), 1204–1213. <https://doi.org/10.1093/glycob/cwp111>.
- (79) Park, Y.-J.; Walls, A. C.; Wang, Z.; Sauer, M. M.; Li, W.; Tortorici, M. A.; Bosch, B.-J.; DiMaio, F.; Velesler, D. Structures of MERS-CoV Spike Glycoprotein in Complex with Sialoside Attachment Receptors. *Nat. Struct. Mol. Biol.* **2019**, *26* (12), 1151–1157. <https://doi.org/10.1038/s41594-019-0334-7>.
- (80) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334–395. <https://doi.org/10.1124/pr.112.007336>.
- (81) Bereau, T. Computational Compound Screening of Biomolecules and Soft Materials by Molecular Simulations. *Model. Simul. Mater. Sci. Eng.* **2021**, *29* (2), 023001. <https://doi.org/10.1088/1361-651X/abd042>.
- (82) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703. <https://doi.org/10.1021/ja512751q>.
- (83) Leelananda, S. P.; Lindert, S. Computational Methods in Drug Discovery. *Beilstein J. Org. Chem.* **2016**, *12* (1), 2694–2718. <https://doi.org/10.3762/bjoc.12.267>.
- (84) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>.
- (85) Woods, C.J.; Mey, A.S.; Calabro, G.; Julien, M. Sire Molecular Simulation Framework. [Http://Siremol.Org.2020.1.0](http://Siremol.Org.2020.1.0).
- (86) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12), 124105. <https://doi.org/10.1063/1.2978177>.
- (87) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300–313. <https://doi.org/10.1063/1.1749657>.
- (88) Merz, K. M.; Kollman, P. A. Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. *J. Am. Chem. Soc.* **1989**, *111* (15), 5649–5658. <https://doi.org/10.1021/ja00197a022>.
- (89) Mey, A. S. J. S.; Allen, B. K.; Macdonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2020**, *2* (1), 18378. <https://doi.org/10.33011/livecoms.2.1.18378>.
- (90) Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J. Reproducibility of Free Energy Calculations across Different Molecular

- Simulation Software Packages. *J. Chem. Theory Comput.* **2018**, *14* (11), 5567–5582. <https://doi.org/10.1021/acs.jctc.8b00544>.
- (91) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *J. Comput. Aided Mol. Des.* **2018**, *32* (10), 937–963. <https://doi.org/10.1007/s10822-018-0170-6>.
- (92) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. *Automated Assessment of Binding Affinity via Alchemical Free Energy Calculations*; preprint; 2020. <https://doi.org/10.26434/chemrxiv.11812053.v1>.
- (93) Wu, L.; Dutta, S.; Mascal, M. Efficient, Chemical-Catalytic Approach to the Production of 3-Hydroxypropanoic Acid by Oxidation of Biomass-Derived Levulinic Acid with Hydrogen Peroxide. *ChemSusChem* **2015**, *8*, 1167–1169. <https://doi.org/10.1002/cssc.201500025>.
- (94) Dutta, S.; Wu, L.; Mascal, M. Efficient, Metal-Free Production of Succinic Acid by Oxidation of Biomass-Derived Levulinic Acid with Hydrogen Peroxide. *Green Chem.* **2015**, *17*, 2335–2338. <https://doi.org/10.1039/c5gc00098j>.
- (95) Haas, T. W. Kinetics of the Uncatalyzed, Alkaline Decomposition of Hydrogen Peroxide. Retrospective Theses and Dissertations, Iowa State University, 1960.
- (96) Abel, E. Über Die Selbstzersetzung von Wasserstoffsperoxyd. *Monatshefte Für Chem. Verwandte Teile Anderer Wiss.* **1952**, *83* (2), 422–439. <https://doi.org/10.1007/BF00938566>.
- (97) Erdey, L.; Inczedy, I. Beiträge Zur Kinetik Der Alkalischen Zersetzung Des Wasserstoff Peroxyds. *Acta Chim. Acad. Sci. Hung.* **1955**, *7*, 94–115.
- (98) Lin, S.-S.; Gurol, M. D. Catalytic Decomposition of Hydrogen Peroxide on Iron Oxide: Kinetics, Mechanism, and Implications. *Environ. Sci. Technol.* **1998**, *32* (10), 1417–1423. <https://doi.org/10.1021/es970648k>.
- (99) Behar, D.; Czapski, G.; Rabani, J.; Dorfman, L. M.; Schwarz, H. A. Acid Dissociation Constant and Decay Kinetics of the Perhydroxyl Radical. *J. Phys. Chem.* **1970**, *74* (17), 3209–3213. <https://doi.org/10.1021/j100711a009>.
- (100) Stoin, U.; Shames, A. I.; Malka, I.; Bar, I.; Sasson, Y. In Situ Generation of Superoxide Anion Radical in Aqueous Medium under Ambient Conditions. *ChemPhysChem* **2013**, *14* (18), 4158–4164. <https://doi.org/10.1002/cphc.201300707>.
- (101) Samuel, E. L. G.; Marcano, D. C.; Berka, V.; Bitner, B. R.; Wu, G.; Potter, A.; Fabian, R. H.; Pautler, R. G.; Kent, T. A.; Tsai, A.-L.; Tour, J. M. Highly Efficient Conversion of Superoxide to Oxygen Using Hydrophilic Carbon Clusters. *Proc. Natl. Acad. Sci.* **2015**, *112* (8), 2343–2348. <https://doi.org/10.1073/pnas.1417047112>.
- (102) Dederichs, P. H.; Blügel, S.; Zeller, R.; Akai, H. Ground States of Constrained Systems: Application to Cerium Impurities. *Phys. Rev. Lett.* **1984**, *53* (26), 2512–2515. <https://doi.org/10.1103/PhysRevLett.53.2512>.
- (103) Wu, Q.; Van Voorhis, T. Direct Optimization Method to Study Constrained Systems within Density-Functional Theory. *Phys. Rev. A* **2005**, *72* (2), 024502. <https://doi.org/10.1103/PhysRevA.72.024502>.
- (104) Wu, Q.; Cheng, C.-L.; Van Voorhis, T. Configuration Interaction Based on Constrained Density Functional Theory: A Multireference Method. *J. Chem. Phys.* **2007**, *127* (16), 164119. <https://doi.org/10.1063/1.2800022>.



- (105) Wu, Q.; Kaduk, B.; Van Voorhis, T. Constrained Density Functional Theory Based Configuration Interaction Improves the Prediction of Reaction Barrier Heights. *J. Chem. Phys.* **2009**, *130* (3), 034109. <https://doi.org/10.1063/1.3059784>.
- (106) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19* (48), 32184–32215. <https://doi.org/10.1039/C7CP04913G>.
- (107) Yaremenko, I. A.; Vil', V. A.; Demchuk, D. V.; Terent'ev, A. O. Rearrangements of Organic Peroxides and Related Processes. *Beilstein J. Org. Chem.* **2016**, *12*, 1647–1748. <https://doi.org/10.3762/bjoc.12.162>.
- (108) Sivaguru, P.; Wang, Z.; Zaroni, G.; Bi, X. Cleavage of Carbon–Carbon Bonds by Radical Reactions. *Chem. Soc. Rev.* **2019**, *48* (9), 2615–2656. <https://doi.org/10.1039/C8CS00386F>.
- (109) Thompson, Q. E. Ozonolysis of Dihydropyran. Reactions of 4-Hydroperoxy-4-Methoxybutyl Formate. *J. Org. Chem.* **1962**, *27* (12), 4498–4502.
- (110) Berkessel, A. Peroxides, Inorganic Esters (RO–X, X = Hal, S, Se, Te, N). *Sci. Synth.* **2009**, *38*, 1. <https://doi.org/10.1055/sos-SD-038-00001>.
- (111) Scarso, A.; Strukul, G. Product Class 1: Alkyl and Cycloalkyl Hydroperoxides. *Sci. Synth.* **2009**, *38*, 9. <https://doi.org/10.1055/sos-SD-038-00002>.
- (112) Vassilikogiannakis, G.; Montagnon, T. Product Class 5: Alkyl and Cycloalkyl Peroxides. *Sci. Synth.* **2009**, *38*, 179. <https://doi.org/10.1055/sos-SD-038-00222>.
- (113) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113* (22), 9978–9985. <https://doi.org/10.1063/1.1323224>.
- (114) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113* (22), 9901–9904. <https://doi.org/10.1063/1.1329672>.
- (115) Fukui, K. Variational Principles in a Chemical Reaction. *Int. J. Quantum Chem.* **1981**, *20* (S15), 633–642. <https://doi.org/10.1002/qua.560200866>.
- (116) Mardirossian, N.; Head-Gordon, M.  $\omega$ B97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.* **2014**, *16* (21), 9904–9924. <https://doi.org/10.1039/C3CP54374A>.
- (117) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. <https://doi.org/10.1039/B508541A>.
- (118) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010**, *133* (24), 244111. <https://doi.org/10.1063/1.3511297>.
- (119) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; III, H. L. W.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.;

- Jr, R. A. D.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscammann, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; III, W. A. G.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; III, H. F. S.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215. <https://doi.org/10.1080/00268976.2014.952696>.
- (120) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12* (1), 444–454. <https://doi.org/10.1021/acs.jctc.5b01066>.
- (121) Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. *J. Chem. Theory Comput.* **2018**, *14* (3), 1254–1266. <https://doi.org/10.1021/acs.jctc.7b01074>.
- (122) Wang, L.-P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108. <https://doi.org/10.1063/1.4952956>.
- (123) Levine, B. G.; Coe, J. D.; Martínez, T. J. Optimizing Conical Intersections without Derivative Coupling Vectors: Application to Multistate Multireference Second-Order Perturbation Theory (MS-CASPT2). *J. Phys. Chem. B* **2008**, *112* (2), 405–413. <https://doi.org/10.1021/jp0761618>.
- (124) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). *J. Phys. Chem. A* **2008**, *112* (50), 13164–13171. <https://doi.org/10.1021/jp801738f>.
- (125) Dhar, C.; Sasmal, A.; Diaz, S.; Verhagen, A.; Yu, H.; Li, W.; Chen, X.; Varki, A. Are Sialic Acids Involved in COVID-19 Pathogenesis? *Glycobiology* **2021**, *31* (9), 1068–1071. <https://doi.org/10.1093/glycob/cwab063>.
- (126) Petitjean, S. J. L.; Chen, W.; Koehler, M.; Jimmidi, R.; Yang, J.; Mohammed, D.; Juniku, B.; Stanifer, M. L.; Boulant, S.; Vincent, S. P.; Alsteens, D. Multivalent 9-O-Acetylated-Sialic Acid Glycoclusters as Potent Inhibitors for SARS-CoV-2 Infection. *Nat. Commun.* **2022**, *13*, 2564. <https://doi.org/10.1038/s41467-022-30313-8>.

- (127) Das, T.; Ghule, S.; Vanka, K. Insights Into the Origin of Life: Did It Begin from HCN and H<sub>2</sub>O? *ACS Cent. Sci.* **2019**, *5* (9), 1532–1540. <https://doi.org/10.1021/acscentsci.9b00520>.
- (128) Thapa, B.; Schlegel, H. B. Improved PKa Prediction of Substituted Alcohols, Phenols, and Hydroperoxides in Aqueous Medium Using Density Functional Theory and a Cluster-Continuum Solvation Model. *J Phys Chem A* **2017**, *9*.
- (129) Rao, G.; Chen, N.; Marchiori, D. A.; Wang, L.-P.; Britt, R. D. Accumulation and Pulse Electron Paramagnetic Resonance Spectroscopic Investigation of the 4-Oxidobenzyl Radical Generated in the Radical S-Adenosyl-L-Methionine Enzyme HydG. *Biochemistry* **2022**, *61* (2), 107–116. <https://doi.org/10.1021/acs.biochem.1c00619>.
- (130) Wang, L.-P.; Wu, Q.; Van Voorhis, T. Acid–Base Mechanism for Ruthenium Water Oxidation Catalysts. *Inorg. Chem.* **2010**, *49* (10), 4543–4553. <https://doi.org/10.1021/ic100075k>.
- (131) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961–1972. <https://doi.org/10.1016/j.cpc.2009.05.011>.
- (132) Grossfield, A. WHAM: An Implementation of the Weighted Histogram Analysis Method, <Http://Membrane.Urnc.Rochester.Edu/Content/Wham/>, Version 2.0.10.
- (133) Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Monographs on statistics and applied probability; Chapman & Hall/CRC: New York, 1994.
- (134) Chokhawala, H. A.; Yu, H.; Chen, X. High-Throughput Substrate Specificity Studies of Sialidases by Using Chemoenzymatically Synthesized Sialoside Libraries. *ChemBioChem* **2007**, *8* (2), 194–201. <https://doi.org/10.1002/cbic.200600410>.
- (135) Sherwood, P.; Brooks, B. R.; Sansom, M. S. Multiscale Methods for Macromolecular Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18* (5), 630–640. <https://doi.org/10.1016/j.sbi.2008.07.003>.
- (136) Lee, J.; Miller, B. T.; Brooks, B. R. Computational Scheme for PH-Dependent Binding Free Energy Calculation with Explicit Solvent. *Protein Sci. Publ. Protein Soc.* **2016**, *25* (1), 231–243. <https://doi.org/10.1002/pro.2755>.
- (137) Jia, X.; Wang, M.; Shao, Y.; König, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J. Chem. Theory Comput.* **2016**, *12* (2), 499–511. <https://doi.org/10.1021/acs.jctc.5b00920>.
- (138) Nguyen, L.; McCord, K. A.; Bui, D. T.; Bouwman, K. M.; Kitova, E. N.; Elaish, M.; Kumawat, D.; Daskhan, G. C.; Tomris, I.; Han, L.; Chopra, P.; Yang, T.-J.; Willows, S. D.; Mason, A. L.; Mahal, L. K.; Lowary, T. L.; West, L. J.; Hsu, S.-T. D.; Hobman, T.; Tompkins, S. M.; Boons, G.-J.; de Vries, R. P.; Macauley, M. S.; Klassen, J. S. Sialic Acid-Containing Glycolipids Mediate Binding and Viral Entry of SARS-CoV-2. *Nat. Chem. Biol.* **2022**, *18* (1), 81–90. <https://doi.org/10.1038/s41589-021-00924-1>.
- (139) Nguyen, K.; Chakraborty, S.; Mansbach, R. A.; Korber, B.; Gnanakaran, S. Exploring the Role of Glycans in the Interaction of SARS-CoV-2 RBD and Human Receptor ACE2. *Viruses* **2021**, *13* (5), 927. <https://doi.org/10.3390/v13050927>.
- (140) Baker, A. N.; Richards, S.-J.; Guy, C. S.; Congdon, T. R.; Hasan, M.; Zwetsloot, A. J.; Gallo, A.; Lewandowski, J. R.; Stansfeld, P. J.; Straube, A.; Walker, M.; Chessa, S.; Pergolizzi, G.; Dedola, S.; Field, R. A.; Gibson, M. I. The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care

- Diagnostic Device. *ACS Cent. Sci.* **2020**, *6* (11), 2046–2052.  
<https://doi.org/10.1021/acscentsci.0c00855>.
- (141) Li, B.; Wang, L.; Ge, H.; Zhang, X.; Ren, P.; Guo, Y.; Chen, W.; Li, J.; Zhu, W.; Chen, W.; Zhu, L.; Bai, F. Identification of Potential Binding Sites of Sialic Acids on the RBD Domain of SARS-CoV-2 Spike Protein. *Front. Chem.* **2021**, *9*.
- (142) Wang, C.-W.; Lee, O. K.; Fischer, W. B.; Wang, C.-W.; Lee, O. K.; Fischer, W. B. Screening Coronavirus and Human Proteins for Sialic Acid Binding Sites Using a Docking Approach. *AIMS Biophys.* **2021**, *8* (3), 248–263.  
<https://doi.org/10.3934/biophy.2021019>.
- (143) Milanetti, E.; Miotto, M.; Di Rienzo, L.; Nagaraj, M.; Monti, M.; Golbek, T. W.; Gosti, G.; Roeters, S. J.; Weidner, T.; Otzen, D. E.; Ruocco, G. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Front. Mol. Biosci.* **2021**, *8*, 509.  
<https://doi.org/10.3389/fmolb.2021.690655>.
- (144) Fantini, J.; Di Scala, C.; Chahinian, H.; Yahi, N. Structural and Molecular Modelling Studies Reveal a New Mechanism of Action of Chloroquine and Hydroxychloroquine against SARS-CoV-2 Infection. *Int. J. Antimicrob. Agents* **2020**, *55* (5), 105960.  
<https://doi.org/10.1016/j.ijantimicag.2020.105960>.
- (145) Zamorano Cuervo, N.; Grandvaux, N. ACE2: Evidence of Role as Entry Receptor for SARS-CoV-2 and Implications in Comorbidities. *eLife* **2020**, *9*, e61390.  
<https://doi.org/10.7554/eLife.61390>.
- (146) Vandelli, A.; Monti, M.; Milanetti, E.; Armaos, A.; Rupert, J.; Zacco, E.; Bechara, E.; Delli Ponti, R.; Tartaglia, G. G. Structural Analysis of SARS-CoV-2 Genome and Predictions of the Human Interactome. *Nucleic Acids Res.* **2020**, *48* (20), 11270–11283.  
<https://doi.org/10.1093/nar/gkaa864>.
- (147) Seyran, M.; Takayama, K.; Uversky, V. N.; Lundstrom, K.; Palù, G.; Sherchan, S. P.; Attrish, D.; Rezaei, N.; Aljabali, A. A. A.; Ghosh, S.; Pizzol, D.; Chauhan, G.; Adadi, P.; Mohamed Abd El-Aziz, T.; Soares, A. G.; Kandimalla, R.; Tambuwala, M.; Hassan, Sk. S.; Azad, G. K.; Pal Choudhury, P.; Baetas-da-Cruz, W.; Serrano-Aroca, Á.; Brufsky, A. M.; Uhal, B. D. The Structural Basis of Accelerated Host Cell Entry by SARS-CoV-2†. *FEBS J.* **2021**, *288* (17), 5010–5020. <https://doi.org/10.1111/febs.15651>.
- (148) Awasthi, M.; Gulati, S.; Sarkar, D. P.; Tiwari, S.; Kateriya, S.; Ranjan, P.; Verma, S. K. The Sialoside-Binding Pocket of SARS-CoV-2 Spike Glycoprotein Structurally Resembles MERS-CoV. *Viruses* **2020**, *12* (9), 909. <https://doi.org/10.3390/v12090909>.
- (149) Huang, Y.; Yang, C.; Xu, X.; Xu, W.; Liu, S. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41* (9), 1141–1149. <https://doi.org/10.1038/s41401-020-0485-4>.
- (150) Sun, X.-L. The Role of Cell Surface Sialic Acids for SARS-CoV-2 Infection. *Glycobiology* **2021**, *31* (10), 1245–1253. <https://doi.org/10.1093/glycob/cwab032>.
- (151) Varki, A.; Angata, T. Siglecs--the Major Subfamily of I-Type Lectins. *Glycobiology* **2006**, *16* (1), 1R-27R. <https://doi.org/10.1093/glycob/cwj008>.
- (152) Marsh, L. Strong Ligand-Protein Interactions Derived from Diffuse Ligand Interactions with Loose Binding Sites. *BioMed Res. Int.* **2015**, *2015*, e746980.  
<https://doi.org/10.1155/2015/746980>.
- (153) Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: Unifying Sequence and Structure Data for Evolutionary Analysis. *BMC Bioinformatics* **2006**, *7* (1), 382.  
<https://doi.org/10.1186/1471-2105-7-382>.

- (154) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (155) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321. <https://doi.org/10.1021/ct300418h>.
- (156) Li, W.; Hulswit, R. J. G.; Widjaja, I.; Raj, V. S.; McBride, R.; Peng, W.; Widagdo, W.; Tortorici, M. A.; Dieren, B. van; Lang, Y.; Lent, J. W. M. van; Paulson, J. C.; Haan, C. A. M. de; Groot, R. J. de; Kuppeveld, F. J. M. van; Haagmans, B. L.; Bosch, B.-J. Identification of Sialic Acid-Binding Function for the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein. *Proc. Natl. Acad. Sci.* **2017**, *114* (40), E8508–E8517. <https://doi.org/10.1073/pnas.1712592114>.
- (157) Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A. Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb. Haemost.* **2020**, *120* (12), 1700–1715. <https://doi.org/10.1055/s-0040-1721319>.
- (158) Kim, S. H.; Kearns, F. L.; Rosenfeld, M. A.; Casalino, L.; Papanikolas, M. J.; Simmerling, C.; Amaro, R. E.; Freeman, R. *GlycoGrip*: Cell Surface-Inspired Universal Sensor for Betacoronaviruses. *ACS Cent. Sci.* **2022**, *8* (1), 22–42. <https://doi.org/10.1021/acscentsci.1c01080>.
- (159) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* **2020**, *183* (4), 1043–1057.e15. <https://doi.org/10.1016/j.cell.2020.09.033>.
- (160) Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S. M. C.; Kopp, M.; Li, D.; Parks, R.; Hsu, A. L.; Borgnia, M. J.; Haynes, B. F.; Acharya, P. Controlling the SARS-CoV-2 Spike Glycoprotein Conformation. *Nat. Struct. Mol. Biol.* **2020**, *27* (10), 925–933. <https://doi.org/10.1038/s41594-020-0479-4>.
- (161) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* **2020**, *367* (6483), 1260–1263. <https://doi.org/10.1126/science.abb2507>.
- (162) McGann, M. FRED and HYBRID Docking Performance on Standardized Datasets. *J. Comput. Aided Mol. Des.* **2012**, *26* (8), 897–906. <https://doi.org/10.1007/s10822-012-9584-8>.
- (163) McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51* (3), 578–596. <https://doi.org/10.1021/ci100436p>.
- (164) Kelley, B. P.; Brown, S. P.; Warren, G. L.; Muchmore, S. W. POSIT: Flexible Shape-Guided Docking For Pose Prediction. *J. Chem. Inf. Model.* **2015**, *55* (8), 1771–1780. <https://doi.org/10.1021/acs.jcim.5b00142>.

- (165) *OEDOCKING 3.3.1: OpenEye Scientific Software, Inc., Santa Fe, NM.*  
*Http://Www.Eyesopen.Com.*
- (166) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, 8 (5), 1542–1555.  
<https://doi.org/10.1021/ct200909j>.
- (167) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, 184 (2), 374–380. <https://doi.org/10.1016/j.cpc.2012.09.022>.
- (168) *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald | Journal of Chemical Theory and Computation.*  
<https://pubs.acs.org/doi/10.1021/ct400314y> (accessed 2022-02-16).
- (169) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *WIREs Comput. Mol. Sci.* **2013**, 3 (2), 198–210.  
<https://doi.org/10.1002/wcms.1121>.
- (170) D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), *AMBER 2018, University of California, San Francisco.*
- (171) Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T.-B. N.; Wang, H.; Ren, P.; Luo, R. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis: Calculating Protein-Ligand Binding Affinities. *J. Comput. Chem.* **2016**, 37 (27), 2436–2446. <https://doi.org/10.1002/jcc.24467>.
- (172) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, 33 (12), 889–897.  
<https://doi.org/10.1021/ar000033j>.
- (173) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of RNA Hairpin Loops and Helices. *J. Biomol. Struct. Dyn.* **1998**, 16 (3), 671–682. <https://doi.org/10.1080/07391102.1998.10508279>.
- (174) *FESetup: Automating Setup for Alchemical Free Energy Simulations | Journal of Chemical Information and Modeling.* <https://pubs.acs.org/doi/10.1021/acs.jcim.5b00368>.
- (175) Mey, A. S. J. S.; Jiménez, J. J.; Michel, J. Impact of Domain Knowledge on Blinded Predictions of Binding Energies by Alchemical Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 199–210. <https://doi.org/10.1007/s10822-017-0083-9>.
- (176) 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>.
- (177) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field.

- Carbohydrates. *J. Comput. Chem.* **2008**, 29 (4), 622–655.  
<https://doi.org/10.1002/jcc.20820>.
- (178) 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>.
- (179) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. 2016. *Gaussian 16, Revision C.01*. Wallingford CT, USA: Gaussian, Inc.
- (180) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, 121 (16), 4023–4039. <https://doi.org/10.1021/acs.jpcc.7b02320>.
- (181) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, 12 (1), 444–454. <https://doi.org/10.1021/acs.jctc.5b01066>.
- (182) Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. *J. Chem. Theory Comput.* **2018**, 14 (3), 1254–1266.  
<https://doi.org/10.1021/acs.jctc.7b01074>.
- (183) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, 5 (11), 1885–1891.  
<https://doi.org/10.1021/jz500737m>.

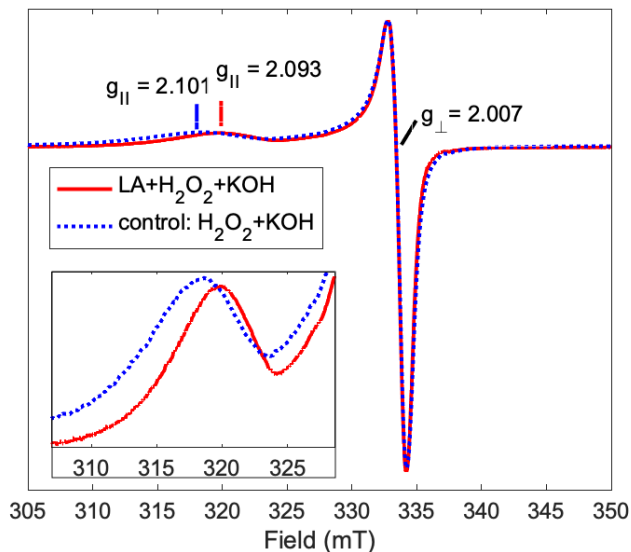
## **Appendix 1 for Application of the ab Initio Nanoreactor to the Development of a New Peroxide-Driven Ketone Oxidation Reaction**

### **Details of the EPR experiments**

Samples were prepared at the reaction concentrations with 30% aq. H<sub>2</sub>O<sub>2</sub> and 5.2 M aq. KOH and were rapidly freeze-quenched at multiple time points in the course of the reactions. X-band (9.37 GHz) CW EPR spectra were recorded at 90 K with 0.2 mW of microwave power and a modulation amplitude of 1.0 G on a Bruker EleXsys E500 spectrometer equipped with a superhigh Q resonator (ER4122SHQE). Simulations of the EPR spectra were performed using the EasySpin 5.2.23 toolbox within the Matlab 2018b software suite (The Mathworks Inc., Natick, MA).

Measurements were taken after 30 min at 90 K with 0.2 mW of microwave power and a modulation amplitude of 1.0 G. Both reactions give rise to an axial EPR signal with g values matching the reported values of O<sub>2</sub><sup>•</sup> (Supporting Figure 2.1). While the g<sub>II</sub> value varies slightly in the presence of LA, the chemical interpretation of this variation remains unclear.





Supporting Figure 2.1. X-band EPR spectra of LA experiment and control reactions.

## Computational Methods

### Nanoreactor Settings

There are multiple variables that are used in these nanoreactor simulations to induce reactivity, and careful choices must be made to produce relevant reactions. We varied simulation parameters and determined the optimal combinations that yielded chemically reasonable reactivity within 2 ps of simulation time for each of our mechanistic studies. Firstly, there is a dependence on where molecules are placed, and how many molecules are chosen. To avoid excess user input time in reactant placement, molecules are randomly placed within a sphere. Still, users need to decide how large the initial sphere should be. In our experience, the smaller the sphere for the number of reactants, the quicker reactions are observed in simulations. Given the approach of random reactant placement into the nanoreactor sphere, multiple sets of starting conditions, and evolution of

molecular coordinates in simulation time, molecular collisions occur across a wide set of conditions (distance between molecules, velocities, molecular orientations, etc). This may reduce the setup effort required in other reactive simulations that collide two or more reacting molecules, while achieving a range of observed reactions.

In addition to the importance of reactant placement into the nanoreactor sphere, the number of molecules matters. While simulation size is limited due to the expensive nature of AIMD, such that a full explicit solvent shell within the nanoreactor is impossible with current hardware and methods, sufficient solvent must be present to describe solvent-solute interactions, such as proton transfers in aqueous medium. If only one key reactant is used at the center of the sphere, then bimolecular reactions between replicates of the key reactant are inherently impossible. Special consideration must also be made when considering how many of other reagents, such as base, are used. If using around three co-reactants, perhaps this is sufficient to interact with the key reactant without significantly increasing simulation time. Ultimately, there is a balance of the number of reactants and solvents used and the simulation time. For example, using fewer solvent molecules may speed up simulations, resulting in the ability to run multiple additional trajectories, thereby increasing potential reaction observations.

A key nanoreactor variable is the choice of simulation temperature ( $T$ ), which effectively determines how much energy the reactants have. Sufficient temperature is needed to quickly dissociate intermolecular interactions, but not so much as to break C-C bonds by temperature alone. Simulating at room temperature is drastically too cold to observe any reactions in a month of simulation time. For exploring organic reactivity, simulations often used a time step of 0.5 fs

with a Langevin thermostat set to 1000 – 2000 K, and a friction coefficient of 3.3 ps<sup>-1</sup>. While choice of thermostat, time-step and friction coefficients may influence reactivity efficiency, it is unlikely to change the outcomes as much as temperature choice, based on our experience. In some cases, at 2000 K, decarboxylation was observed to occur spontaneously, and at 3000 K, organic molecules completely exploded!

The second set of key nanoreactor parameter choices are the boundary radii for the two time periods. In the first time period, the initial boundary radius ( $R_1$ ) must be large enough to allow for molecules to spread out. If this radius is too large, then the boundary molecules are accelerated inward with massive speeds, due to the additional potential applied to molecules in the second time period. This results in explosive reactions! Modifying  $R_1$  was more relevant than significantly decreasing the force constant used to accelerate molecules inward. I found initial boundary radius ( $R_1$ ) and second boundary radius ( $R_2$ ) were best chosen to be  $\sim 1.5\text{--}2\times$  and  $\sim 0.8\times$  that of the smallest initial radius used in Packmol. Reactivity can be further tuned by adjustment of the  $R_2$  parameter, where smaller values of  $R_2$  result in higher-energy collisions and greater frequency of generated reactivity, although the reaction pathways found in this way tended to have higher activation energies after refinement. If  $R_2$  is too small, then all the reactants are crammed into a small space and essentially pulse and fragment. If  $R_2$  is too large, then reactants are not accelerated inward with enough force to induce reactivity. Mass-weighted force constants  $k_1$  and  $k_2$  can also be tuned, in a similar fashion as described in the radii boundary conditions.

Finally, another nanoreactor parameter to tune is the time conditions – the duration for the first period of the nanoreactor where the molecules spread out, and the duration of the second and

highly reactive period. While reactions may occur during the first time period, I generally found it better for molecules to diffuse in the first period, then collide and react in the second period. The more cycles of this diffuse and react period, the more reaction events observed. So, I optimized the time periods to be 1000 fs and 500 fs for  $t_1$  and  $t_2$  to induce the most reactivity in the shortest time. As an example, some of our simulations used a combination of the following parameters:  $T = 1500$  K,  $R_1 = 12.0$  Å,  $k_1 = 1.0$  kcal mol<sup>-1</sup> Å<sup>-2</sup> amu<sup>-1</sup>,  $t_1 = 1000$  fs,  $R_2 = 4.4$  Å,  $k_2 = 0.5$  kcal mol<sup>-1</sup> Å<sup>-2</sup> amu<sup>-1</sup>, and  $t_2 = 500$  fs.

Additional to nanoreactor boundary condition choices, other factors influence the speed of calculations. The computational cost of *ab initio* molecular dynamics (AIMD) simulations is high, which limits simulations to a few tens of picoseconds for ~100 atoms when using a relatively inexpensive 3-21G basis set. Often, the level of theory for the nanoreactor simulations was B3LYP with a DFT-D3 dispersion correction, 3-21G Gaussian basis set, and SWIG-PCM polarizable continuum model of the aqueous reaction medium ( $\epsilon=78.4$ ), denoted as B3LYP-D3/3-21G/PCM. We considered B3LYP may outperform Hartree Fock (HF) for similar simulation times. The implicit solvent was used in our simulations to relatively stabilize our charged intermediate species, compared to gas-phase calculations, while still generating relevant reaction pathway information, and without significantly increasing computational time, which would be the case in adding additional explicit water molecules. Simulations utilized the TeraChem quantum chemistry software<sup>31-37</sup> and were run on servers with an Intel Xeon CPU / Nvidia GPU (either E5-2630 v3 CPU / GTX 980 Ti GPU or E5-2637 v4 CPU / GTX 1080 Ti GPU). Use of TeraChem allowed these simulations to run on GPU's, which increases simulation calculation speeds in time compared to CPU's.

Simulation parameters were varied as we searched for the optimal combinations that yielded chemically reasonable reactivity. The initial boundary radius ( $R_1$ ) and second boundary radius ( $R_2$ ) were chosen to be  $\sim 2x$  and  $\sim 0.7x$  that of the smallest initial radius used in Packmol. The temperature limits ranged from 1000 K to 3000 K. At the lower end of the temperature range no reactions were observed, whereas formation of  $H_2$  and  $CH_4$  from organic reactants was seen at the highest temperatures. Temperatures between 1500-2000 K tended to result in the generation of potentially relevant reaction pathways. Reactivity can be further tuned by adjustment of the  $R_2$  parameter, in that smaller values of  $R_2$  result in higher-energy collisions and greater frequency of generated reactivity, although the reaction pathways found in this way tended to have higher activation energies after refinement.

Sample Condition	T (K)	$\Delta t$ (fs)	$\gamma$ ( $ps^{-1}$ )	$R_1$ ( $\text{\AA}$ )	$k_1$ ( $kcal/mol/\text{\AA}^2$ )	$R_2$ ( $\text{\AA}$ )	$k_2$ ( $kcal/mol/\text{\AA}^2$ )	$t_1:t_2$ (steps)	$m_H$ (amu)
1	1500	0.5	6.7	10.0	1.0	4.0	0.03	1:9999	2.0
2	1500	0.5	6.7	8.0	1.0	4.0	0.01	2000:1000	2.0
3	2500	0.5	6.7	8.0	1.0	4.0	0.50	3000:1000	2.0
4	2500	1.0	6.7	8.0	1.0	4.0	0.50	1500:500	2.0
5	3000	0.5	6.7	8.0	1.0	4.0	0.50	3000:1000	1.0
6	1000	1.0	1.0	10.0	1.0	0.1	0.01	1:9999	2.0

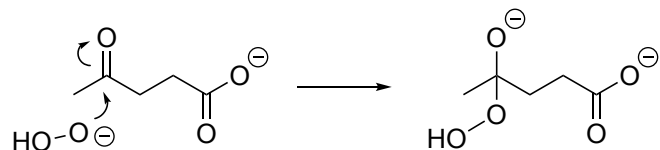
**Supporting Table 2.1.** Sample simulation conditions for the *ab initio* nanoreactor.

$\Delta t$  = time step,  $\gamma$  = friction coefficient,  $m_{\text{H}}$  = hydrogen atomic mass. Sphere radii  $R_1$ ,  $R_2$ , force constants  $k_1$ ,  $k_2$  and time intervals  $t_1$ ,  $t_2$  are described in the main text.

### **3-D Renderings of Structures in Manuscript Reaction Schemes:**

Supporting Figure 2.2: Details of reaction pathway  $LA^- \rightarrow HP-LA$  shown in Scheme 2.2a

**Mechanism:**



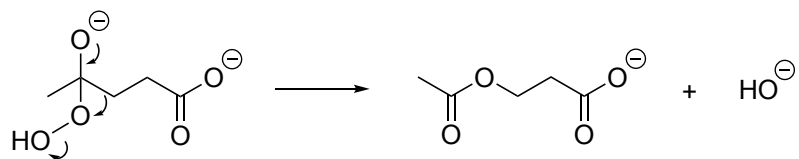
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 0.2$ , $\Delta G^\ddagger = 1.2$
Final		$\Delta E = -10.8$ , $\Delta G = -7.0$

Supporting Figure 2.3: Details of reaction pathway HP-LA  $\rightarrow$  3-APA<sup>-</sup> shown in Scheme 2.2a

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

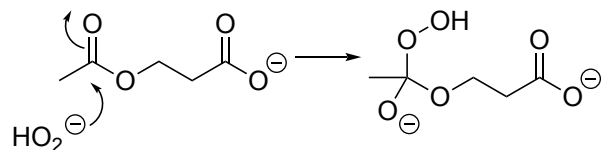
**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E, \Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 37.0, \Delta G^\ddagger = 33.5$
Final		$\Delta E = -51.4, \Delta G = -55.8$



Supporting Figure 2.4: Details of reaction pathway 3-APA<sup>-</sup> → 1 shown in Scheme 2.2b, 2d and 3

**Mechanism:**



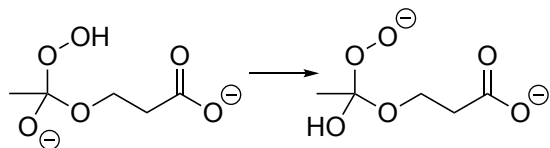
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 3.5$ , $\Delta G^\ddagger = 6.0$
Final		$\Delta E = -3.6$ , $\Delta G = +0.8$

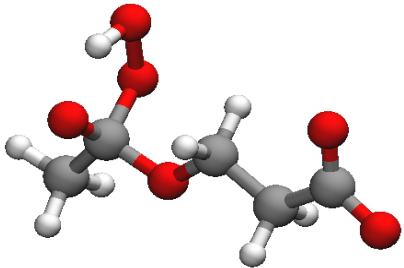
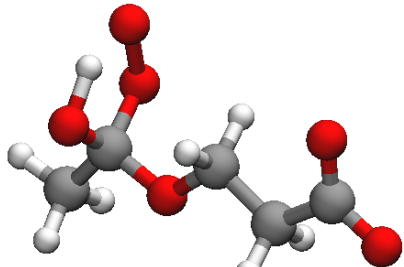
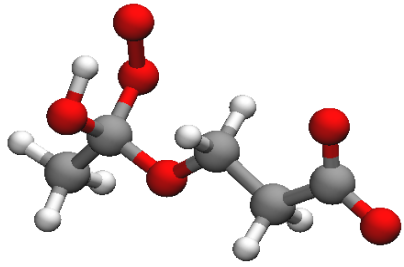
Supporting Figure 2.5: Details of reaction pathway 1 → 2 shown in Scheme 2.2b

**Mechanism:**



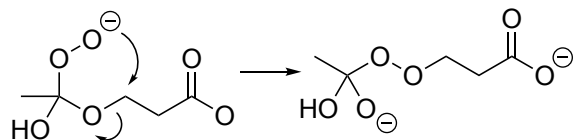
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 3.2$ , $\Delta G^\ddagger = 1.5$
Final		$\Delta E = 2.9$ , $\Delta G = 2.9$

Supporting Figure 2.6: Details of reaction pathway 2 → 3 shown in Scheme 2.2b

**Mechanism:**



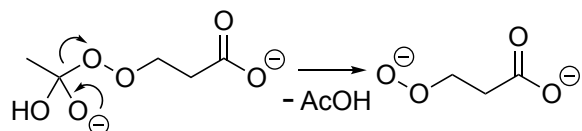
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 59.1$ , $\Delta G^\ddagger = 57.5$
Final		$\Delta E = -3.8$ , $\Delta G = -4.5$

Supporting Figure 2.7: Details of reaction pathway 3 → 3-HPPA<sup>2-</sup> shown in Scheme 2.2b

**Mechanism:**



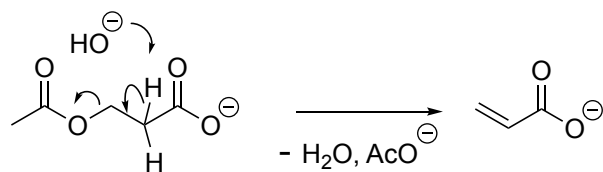
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 7.1$ , $\Delta G^\ddagger = 5.1$
Final		$\Delta E = -2.1$ , $\Delta G = -7.7$

Supporting Figure 2.8: Details of reaction pathway 3-APA<sup>-</sup> → 4 shown in Scheme 2.2c

**Mechanism:**



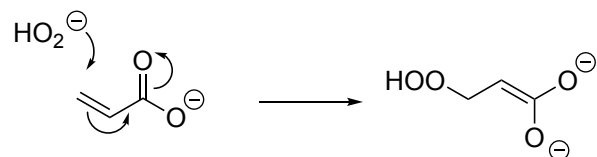
**Method / Basis / Solvent:** B3LYP / 6-31G\* / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed

Name	3D Structure	$\Delta E, \Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 6.8$
Final		$\Delta E = -33.1$ At wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ ): $\Delta E = -15.0, \Delta G = -20.1$

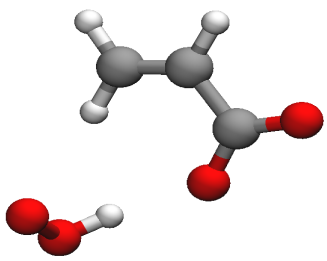
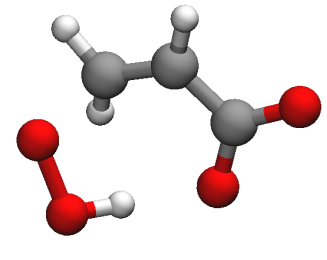
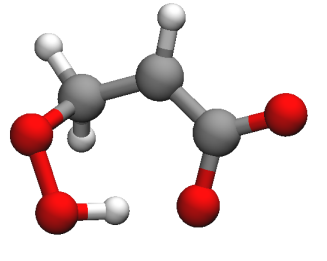
Supporting Figure 2.9: Details of reaction pathway 4 → 5 shown in Scheme 2.2c

**Mechanism:**



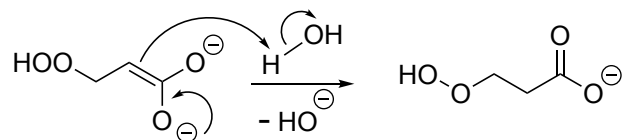
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 15.4$ , $\Delta G^\ddagger = 17.6$
Final		$\Delta E = 7.5$ , $\Delta G = 10.9$

Supporting Figure 2.10: Details of reaction pathway 5  $\rightarrow$  3-HPPA<sup>-</sup> shown in Scheme 2.2c

**Mechanism:**



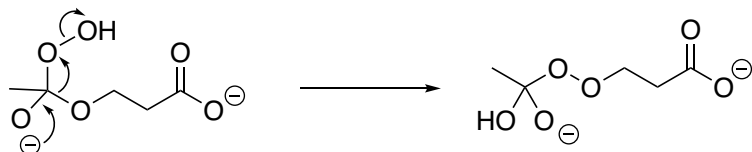
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E, \Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 1.3, \Delta G^\ddagger = 0.1$
Final		$\Delta E = -16.0, \Delta G = -15.7$

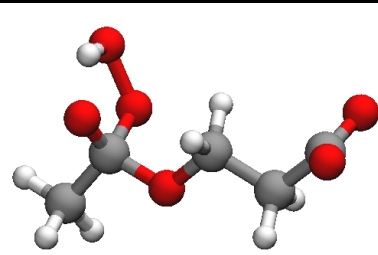
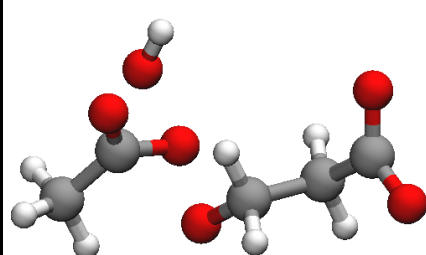
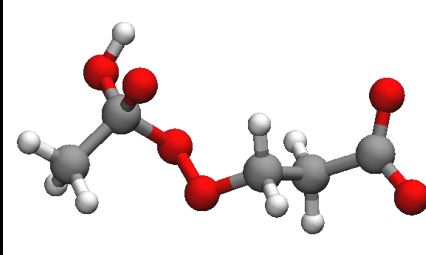
Supporting Figure 2.11: Details of reaction pathway 1 → 3 shown in Scheme 2.2d

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

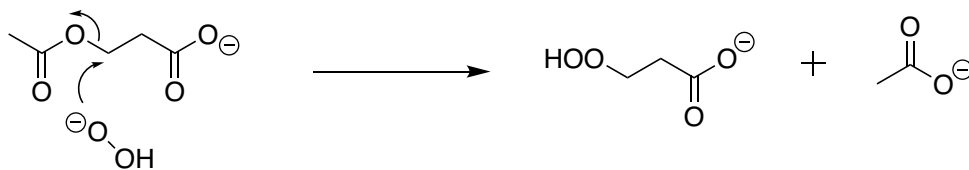
**Closed / Open shell:** Closed

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 45.5$ , $\Delta G^\ddagger = 41.4$
Final		$\Delta E = -0.1$ , $\Delta G = -0.8$



Supporting Figure 2.12: Details of reaction pathway 3-APA<sup>-</sup> → 3-HPPA<sup>-</sup> shown in Scheme 2.3

**Mechanism:**



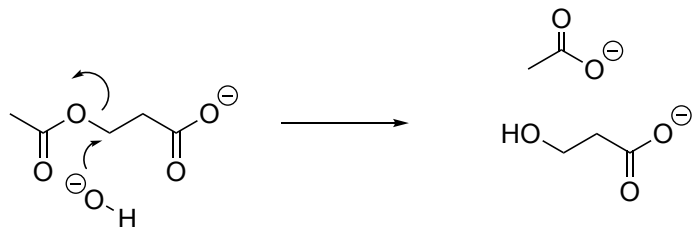
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 24.8$ , $\Delta G^\ddagger = 25.8$
Final		$\Delta E = -22.9$ , $\Delta G = -23.3$

Supporting Figure 2.13: Details of reaction pathway 3-APA<sup>-</sup> → 3-HPA<sup>-</sup> shown in Scheme 2.3

**Mechanism:**



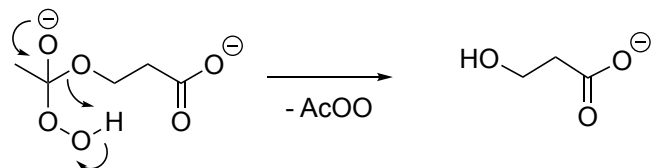
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E, \Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 25.5, \Delta G^\ddagger = 25.6$
Final		$\Delta E = -29.5, \Delta G = -30.1$

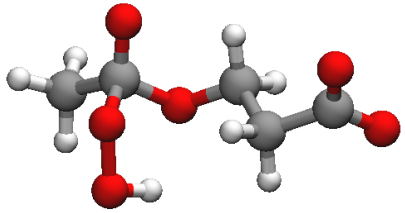
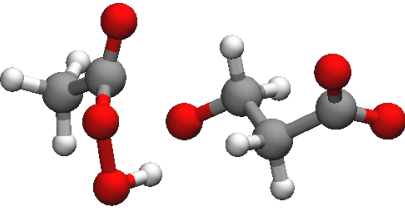
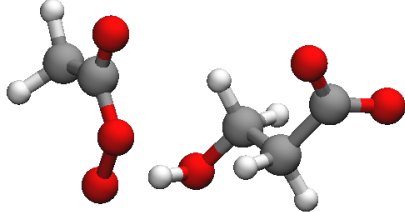
Supporting Figure 2.14: Details of reaction pathway 1 → 3-HPA<sup>-</sup> shown in Scheme 2.3

**Mechanism:**



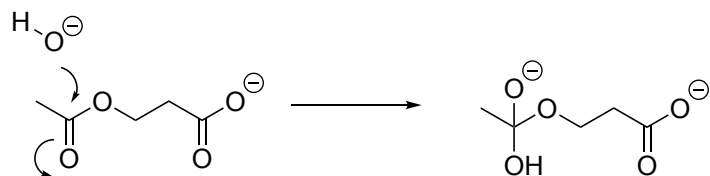
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 8.0$ , $\Delta G^\ddagger = 6.4$
Final		$\Delta E = -11.0$ , $\Delta G = -14.0$

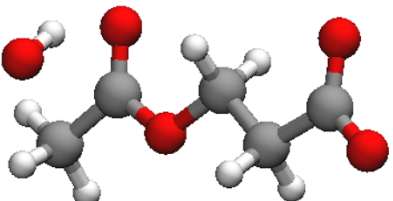
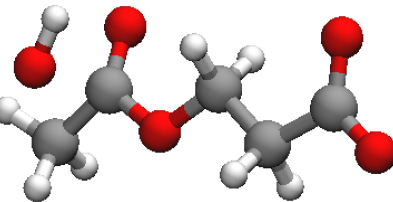
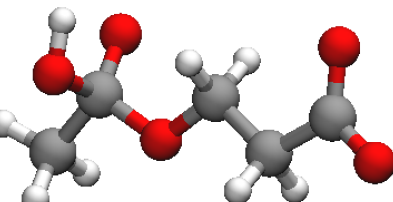
Supporting Figure 2.15: Details of reaction pathway 3-APA<sup>-</sup> → 6 shown in Scheme 2.3

**Mechanism:**



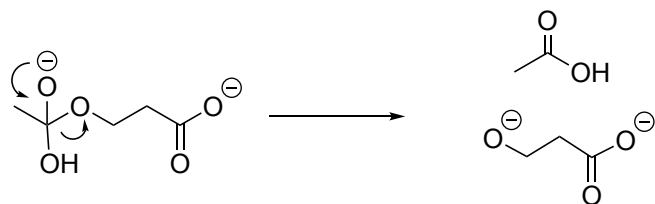
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 7.4$ , $\Delta G^\ddagger = 10.5$
Final		$\Delta E = -3.4$ , $\Delta G = 1.9$

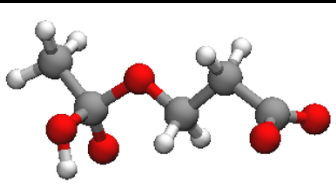
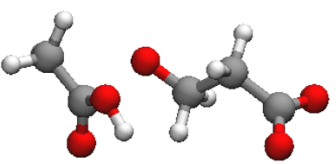
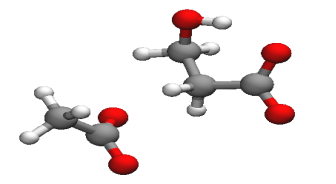
Supporting Figure 2.16: Details of reaction pathway 6 → 3-HPA<sup>-</sup> shown in Scheme 2.3

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

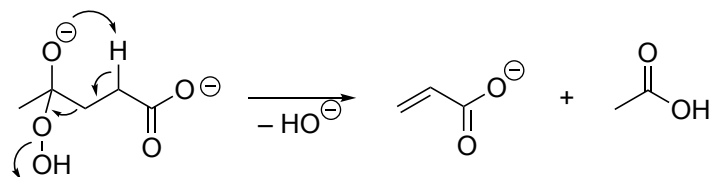
**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E, \Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 11.8, \Delta G^\ddagger = 9.6$
Final		$\Delta E = -22.4, \Delta G = -26.7$

\* Note: The final structure at the endpoint of the IRC calculation had a higher-energy protonation state, which we assumed would undergo PT with a negligible barrier. Therefore, final  $\Delta E$  and  $\Delta G$  values are calculated using the same product structure as Supporting Information Fig. 13.

Supporting Figure 2.17: Details of reaction pathway HP-LA  $\rightarrow$  7 shown in Scheme 2.4

**Mechanism:**



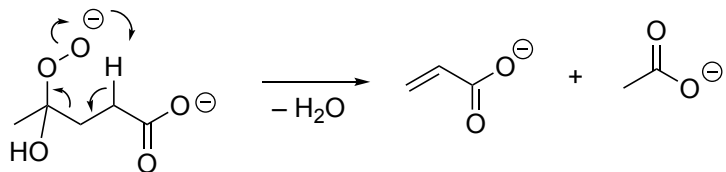
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 55.1$ , $\Delta G^\ddagger = 51.0$
Final		$\Delta E = -71.2$ , $\Delta G = -80.9$

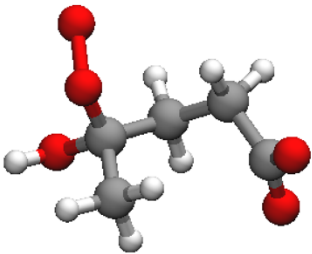
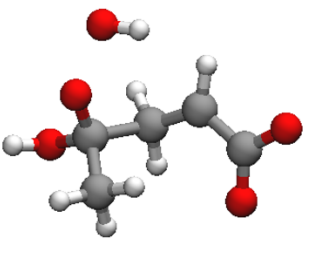
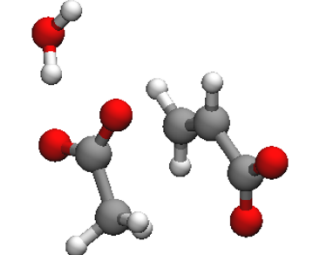
Supporting Figure 2.18: Details of reaction pathway HP-LA'  $\rightarrow$  7 shown in Scheme 2.4

**Mechanism:**



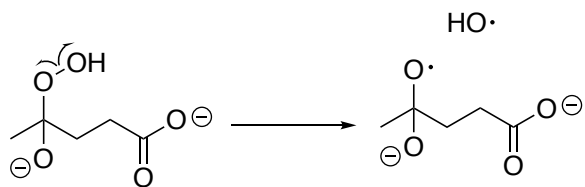
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 62.0$ , $\Delta G^\ddagger = 57.2$
Final		$\Delta E = -74.5$ , $\Delta G = -84.7$

Supporting Figure 2.19: Details of reaction pathway HP-LA  $\rightarrow$  9 shown in Scheme 2.5

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

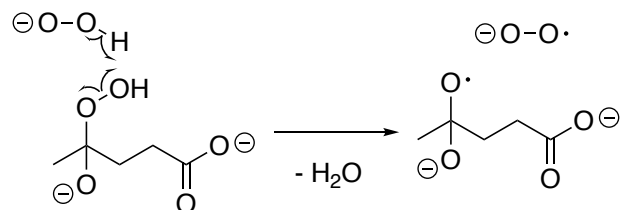
**Closed / Open shell:** Open shell TS & intermediates, Closed shell products

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 36.9$ , $\Delta G^\ddagger = 32.7$
Final		$\Delta E = 32.6$ , $\Delta G = 28.2$



Supporting Figure 2.20: Details of reaction pathway HP-LA  $\rightarrow$  9 shown in Scheme 2.5

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

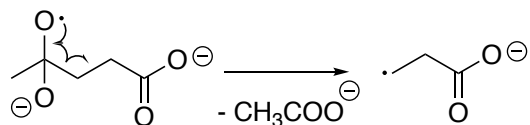
**Closed / Open shell:** Open shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
MECP		$\Delta E$ (MECP) = 29.1 <sup>†</sup>
Final		$\Delta E = -18.4$ , $\Delta G = -23.7$

<sup>†</sup> Computed using CDFT-CI single point calculations at the minimum energy crossing point.

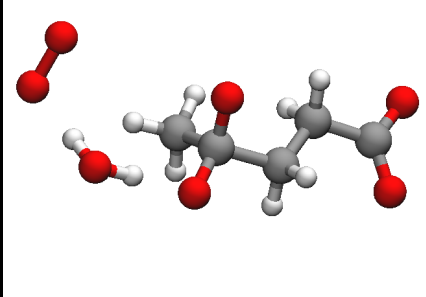
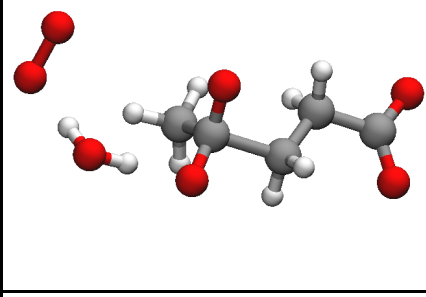
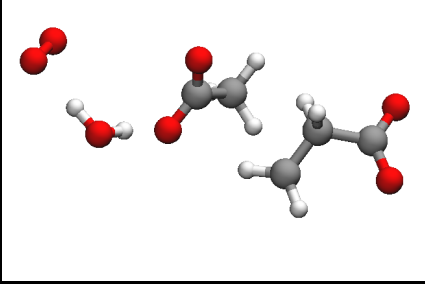
Supporting Figure 2.21: Details of reaction pathway 9 → 10 shown in Scheme 2.5

**Mechanism:**



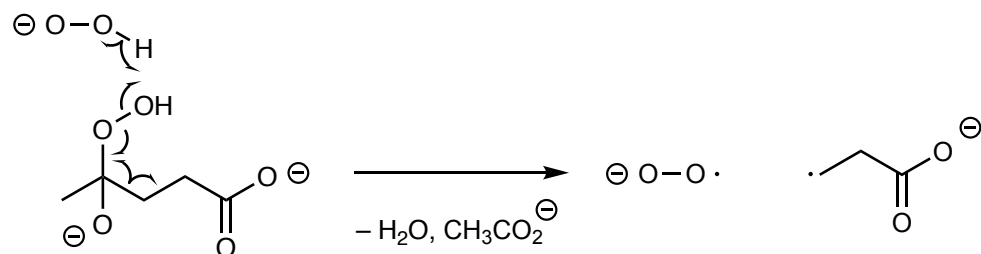
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Open shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 0.4$ , $\Delta G^\ddagger = -0.3$
Final		$\Delta E = -23.3$ , $\Delta G = -31.7$

Supporting Figure 2.22: Details of reaction pathway HP-LA  $\rightarrow$  10 shown in Scheme 2.5, calculated using B3LYP

**Mechanism:**



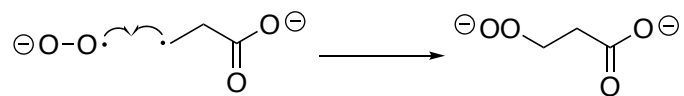
**Method / Basis / Solvent:** B3LYP-D3(BJ) / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Open shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 21.4$ , $\Delta G^\ddagger = 17.9$
Final		$\Delta E = -47.6$ , $\Delta G = -58.8$

Supporting Figure 2.23: Details of reaction pathway 10  $\rightarrow$  3-HPPA<sup>2-</sup> shown in Scheme 2.5

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

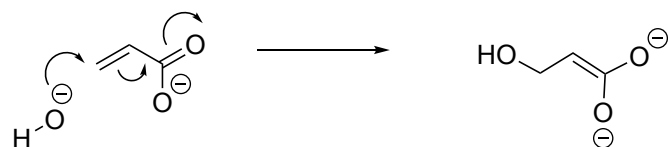
**Closed / Open shell:** Open shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 4.2$ , $\Delta G^\ddagger = 7.3$
Final		$\Delta E = -50.1$ , $\Delta G = -40.6$

### 3-D Renderings of Structures Related to Manuscript Reaction Search

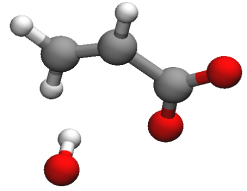
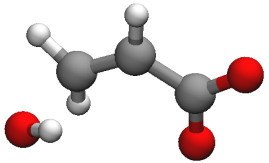
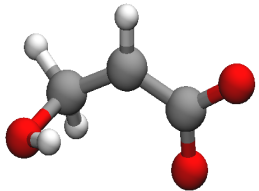
Supporting Figure 2.24: Details of HO<sup>-</sup> addition to 7 to form 3-HPA<sup>-</sup>, step 1/2

#### Mechanism:



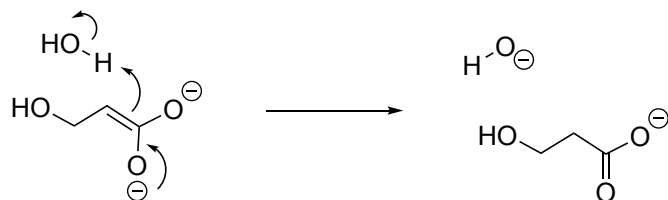
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 18.5$ , $\Delta G^\ddagger = 21.6$
Final		$\Delta E = 11.7$ , $\Delta G = 16.4$

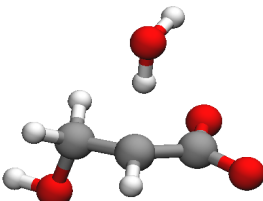
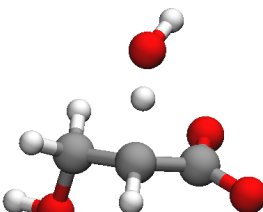
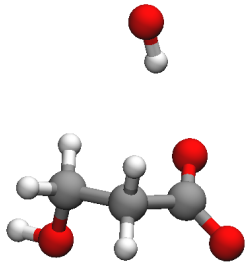
Supporting Figure 2.25: Details of HO<sup>-</sup> addition to 7 to form 3-HPA<sup>-</sup>, step 2/2

**Mechanism:**



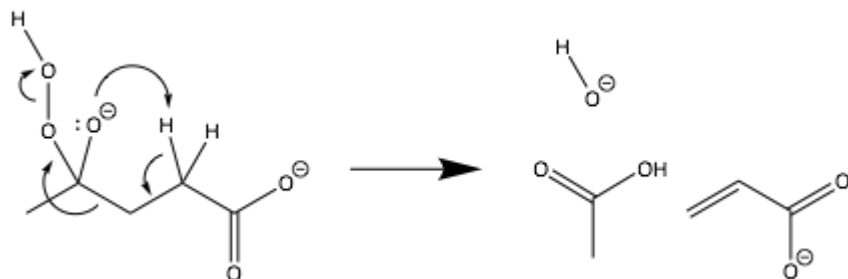
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ , $\Delta G$ (kcal/mol)
Initial		0.0
TS		$E_a = 0.8$ , $\Delta G^\ddagger = 0.9$
Final		$\Delta E = -18.0$ , $\Delta G = -17.7$

Supporting Figure 2.26: Elimination to Acrylate via 5-membered-ring TS with 2 K<sup>+</sup> and 6 H<sub>2</sub>O

**Mechanism:**



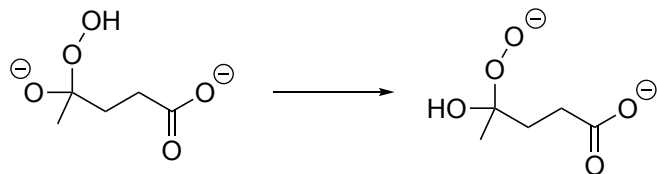
**Method / Basis / Solvent:** wB97X-V / TZVP / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ (kcal/mol)
Initial		0.0
TS		$E_a = 54.3$
Final		$\Delta E = -49.7$

Supporting Figure 2.27: Proton Transfer Between Oxo- and Peroxo- Groups of HP-LA

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

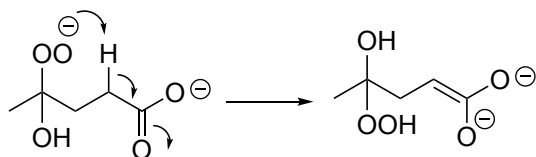
**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ (kcal/mol)
Initial		0.0
TS		$E_a = 2.0$
Final		$\Delta E = -0.3$



Supporting Figure 2.28: Proton Transfer from C2H to 4-Peroxo to Form “Enolate” Intermediate

**Mechanism:**



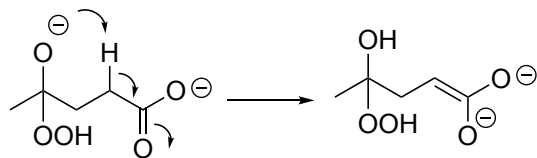
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ (kcal/mol)
Initial		0.0
TS		$E_a = 32.2$
Final		$\Delta E = 22.7$

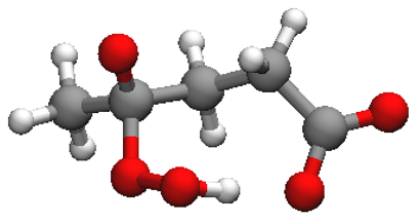
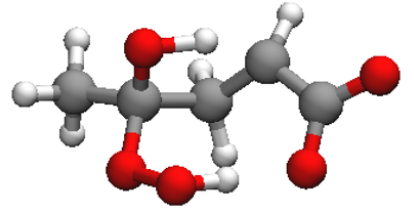
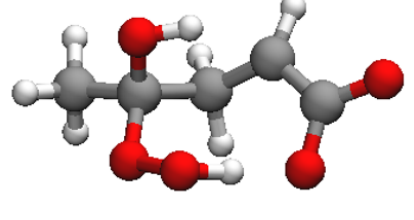
Supporting Figure 2.29: Proton Transfer from C2H to 4-Oxo to Form “Enolate” Intermediate

**Mechanism:**



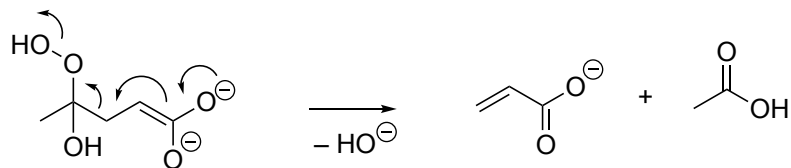
**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ (kcal/mol)
Initial		0.0
TS		$E_a = 23.8$
Final		$\Delta E = 21.8$

Supporting Figure 2.30: Dissociation of “Enolate” Intermediate to Acrylate

**Mechanism:**



**Method / Basis / Solvent:** wB97X-V / def2-TZVPD / SWIG-PCM ( $\epsilon=78.4$ )

**Closed / Open shell:** Closed shell

Name	3D Structure	$\Delta E$ (kcal/mol)
Initial		0.0
TS		$E_a = 36.1$
Final		$\Delta E = -63.6$

**Appendix 2 for *O*-Acetyl migration within the sialic acid side chain: a mechanistic study by the *ab initio* nanoreactor**

Figure S3.1. Reaction free energy and barrier for Neu5,9Ac<sub>2</sub> *O*-acetyl migration cyclization step to 6-membered intermediate in C9-C7 acetyl migration, without an initial deprotonation step.

Free energies calculated at ωB97X-D3/TZVP/PCM level of theory.

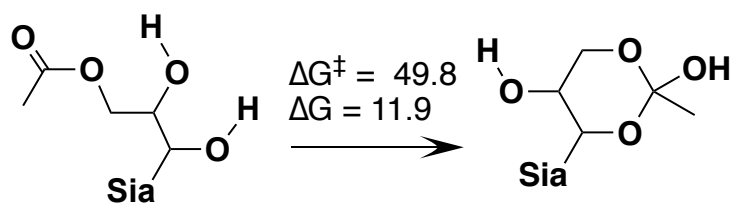


Figure S3.2. Nanoreactor reaction free energies and activation barriers for the de-*O*-acetylation and deprotonation of the C8-hydroxyl group in Neu5,9Ac<sub>2</sub>. (a) Hydrolysis of Neu5,9Ac<sub>2</sub> to form Neu5Ac in the absence of an initial deprotonation step, as an initial de-*O*-acetylation step prior to acetylation at C7 or C8 positions. (b) Deprotonation of Neu5,9Ac<sub>2</sub> at C8-OH by HO<sup>-</sup>, (c) Concerted protonation of C8-O<sup>-</sup> by H<sub>2</sub>O and OH<sup>-</sup> addition to the ester carbonyl, (d) De-*O*-acetylation of Neu5,9Ac<sub>2</sub>-intermediate to form Neu5Ac, Free energies calculated at ωB97X-D3/TZVP level of theory as described in the Methods section.

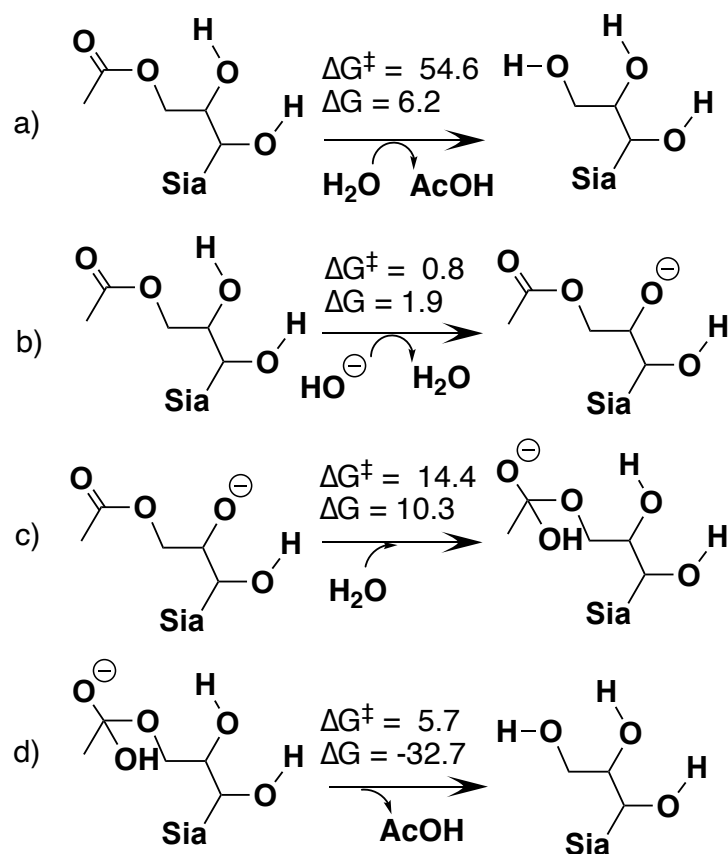


Figure S3.3. Full pathway of reaction and activation free energies for base-catalyzed *O*-acetyl migration of Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub> (green). Energies were estimated for a few glycerol rotation steps where transition states could not be found (orange) using electronic energies of torsion scans (Fig. S4). Free energies are calculated at ωB97X-D3/TZVP level of theory as described in the Computational Details.

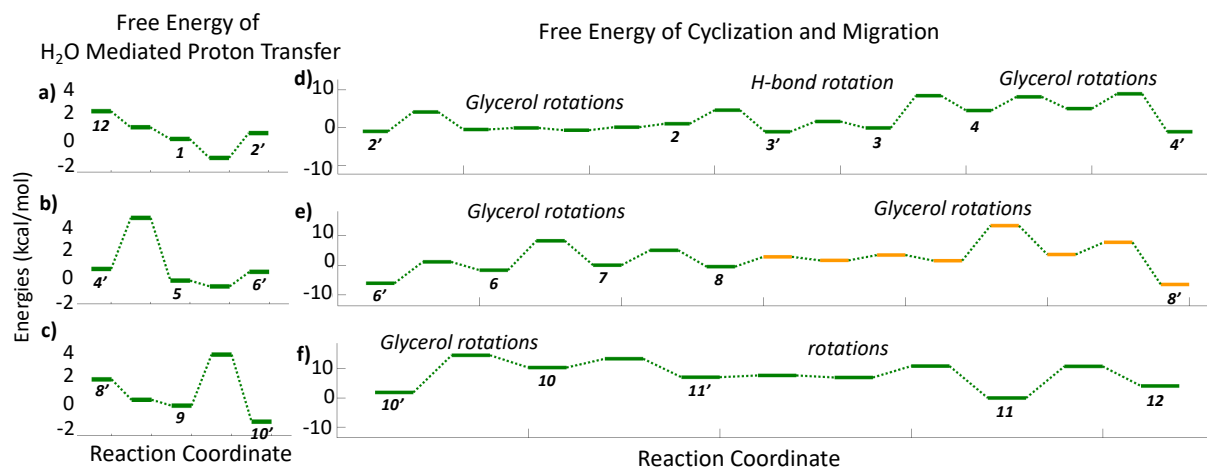


Figure S3.4. Torsional electronic energy profile of glycerol rotations between 8 and 9 in the *O*-acetyl migration from C7-OH to C8-OH. Energies in kcal/mol. This involved glycerol dihedral rotation about C8-C9, C8-OH proton rotation, *O*-acetyl dihedral rotation, followed by a final glycerol rotation about C7-C8. This approximates the barrier in this conformational change, as a single TS between each step with connecting endpoints was not found (the *O*-acetyl rotation was the most difficult). The maximum individual barriers in the forward and reverse directions are 11.9 kcal/mol and 14.2 kcal/mol. Torsion scans were carried out at  $\omega$ B97X-D3/TZVP level of theory as described in the Computational Details.

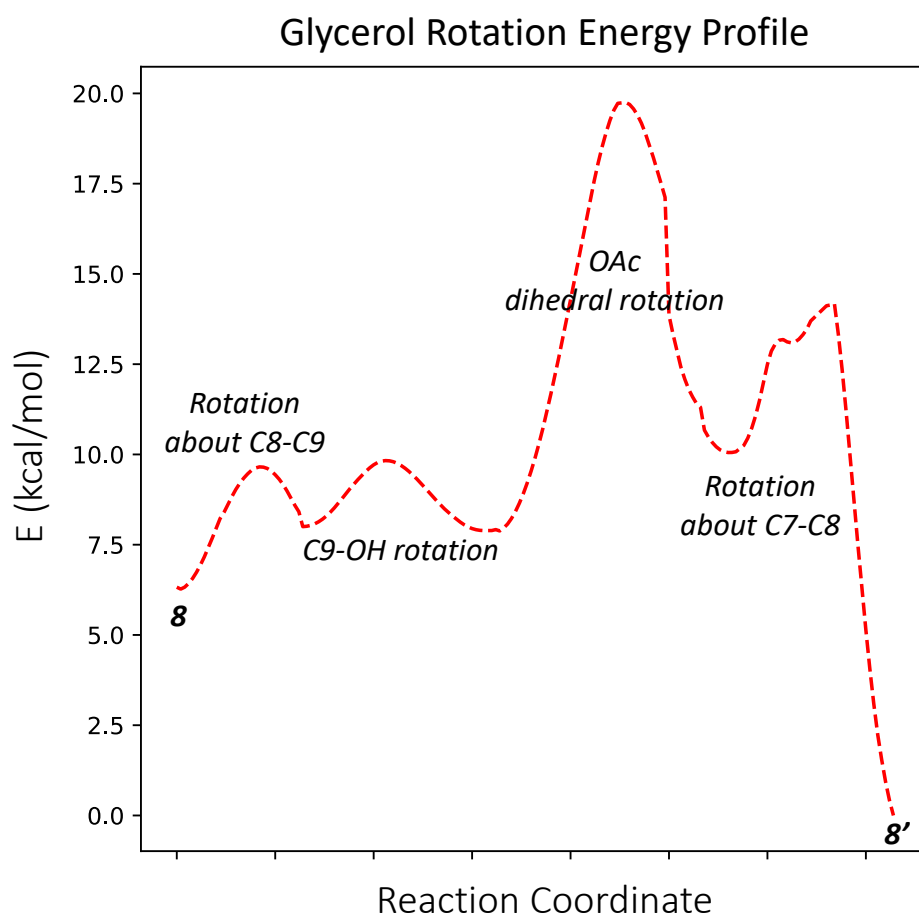
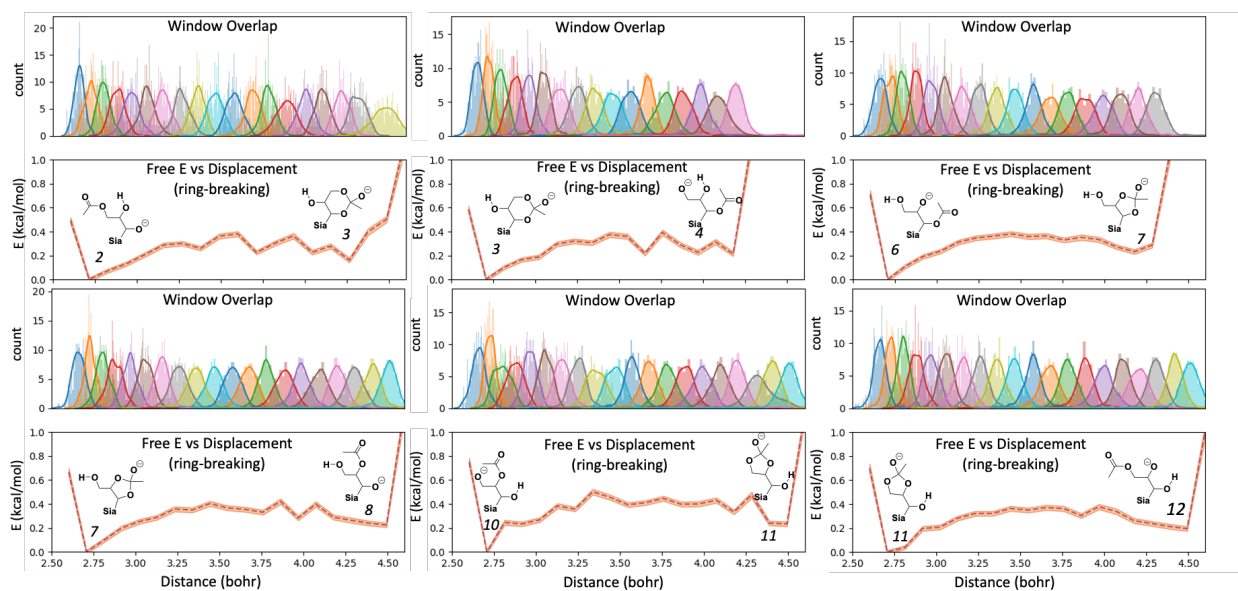


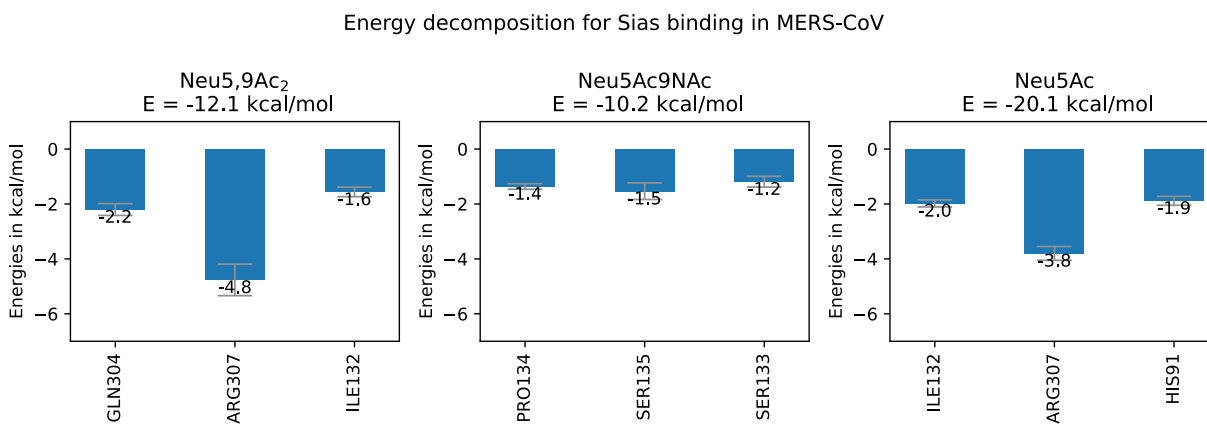
Figure S3.5. Umbrella sampling of each cyclization/*O*-acetyl migration step between Neu5,9Ac<sub>2</sub>, Neu5,8Ac<sub>2</sub> and Neu5,7Ac<sub>2</sub>. Overlap of windows sampling over bond length is plotted above free energy profiles in red-dashed lines with estimated error bars shaded in orange, where each barrier is < 1 kcal/mol. Umbrella sampling was carried out at B3LYP/6-31G\* level of theory as described in the Computational Details.



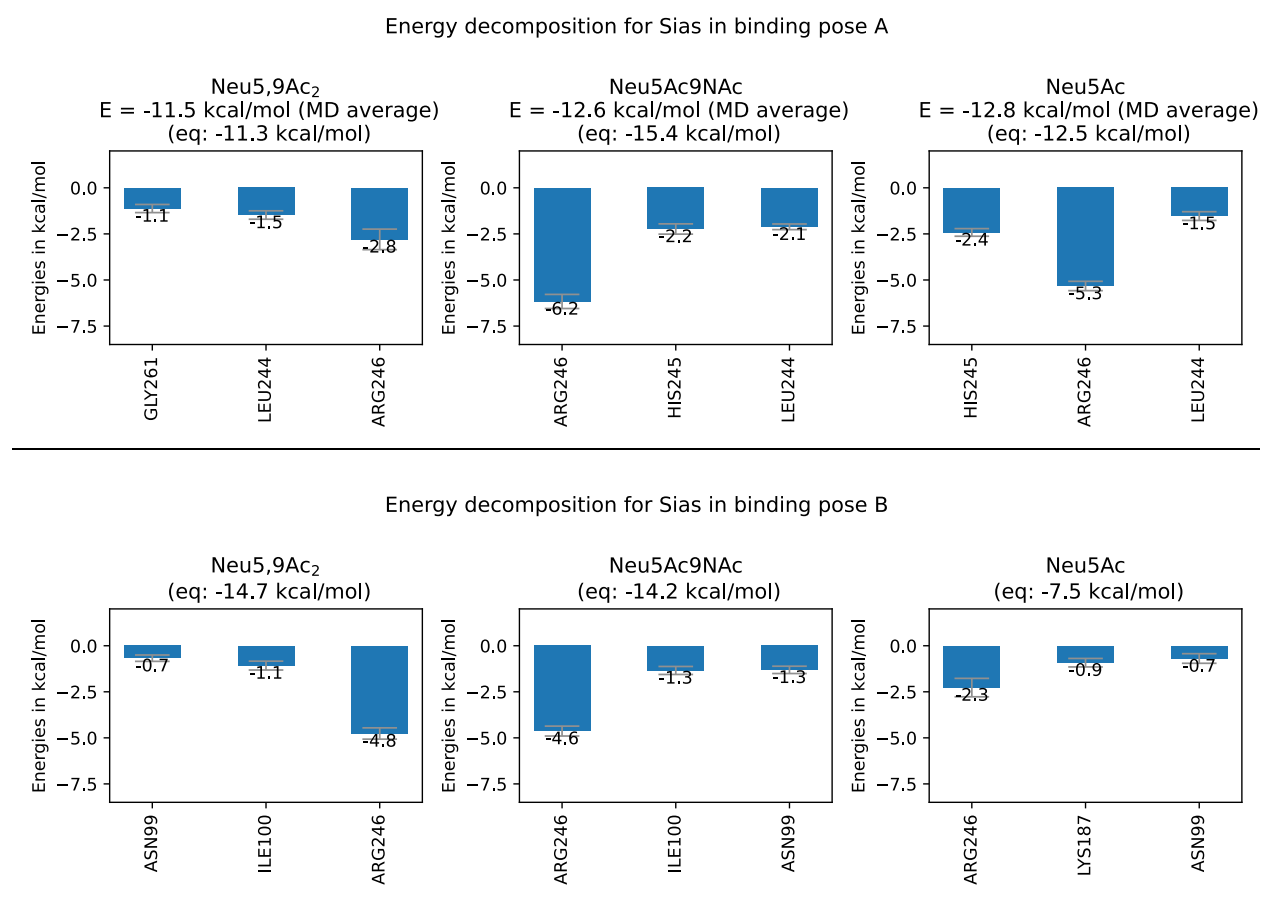


**Appendix 3 for SARS-CoV-2 and MERS-CoV spike protein binding studies support stable mimic of bound 9-O-acetylated sialic acids**

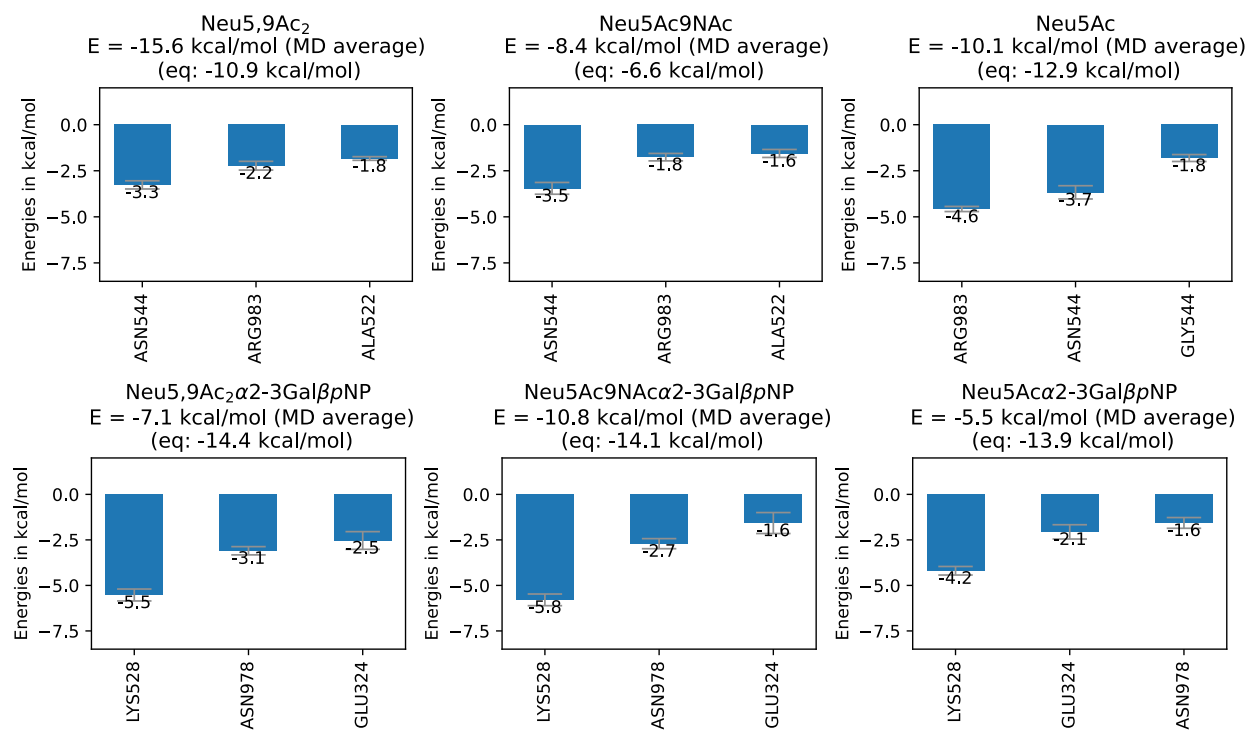
**Figure S4.1:** MM-PBSA energies and decomposition analysis for Neu5,9Ac<sub>2</sub>, Neu5Ac9NAc and Neu5Ac in MERS-CoV S protein, with the top 3 binding residues are highlighted.



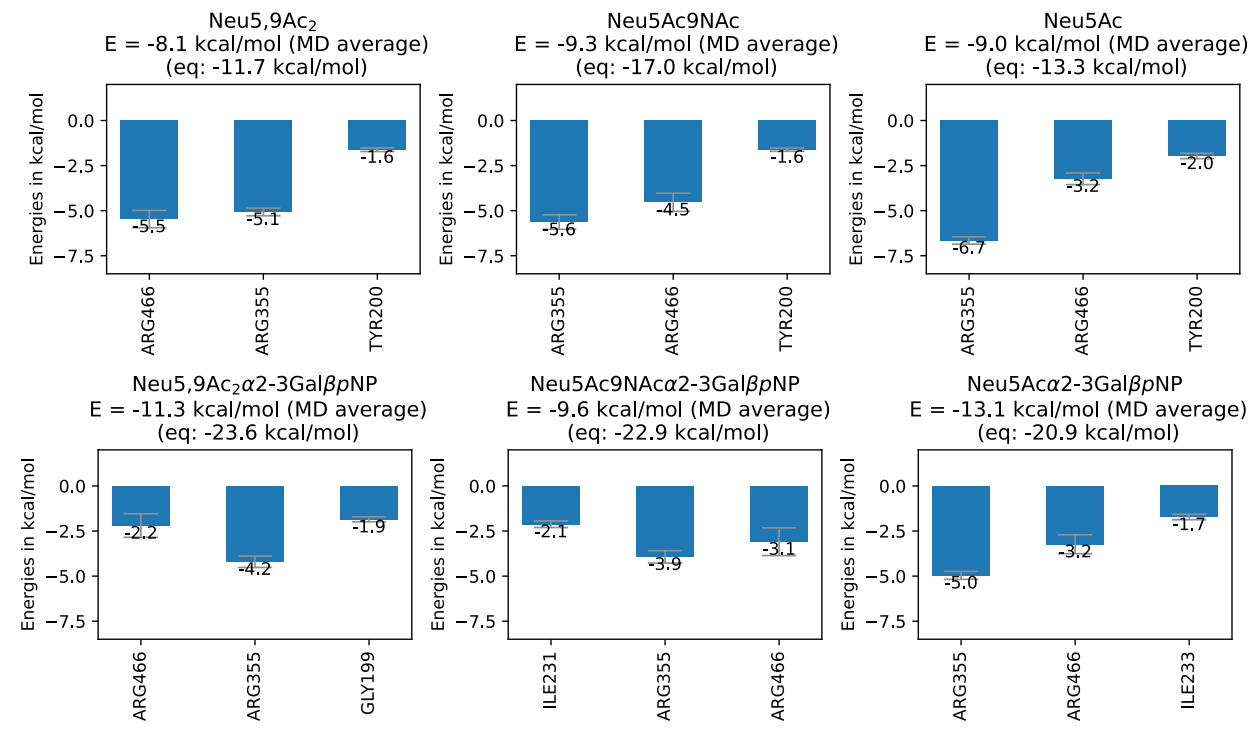
**Figure S4.2:** MM-PBSA energies and decomposition analysis for Neu5,9Ac<sub>2</sub>, Neu5Ac9NAc, Neu5Ac, Neu5,9Ac<sub>2</sub>α2-3GalβpNP, Neu5Ac9NAcα2-3GalβpNP and Neu5Acα2-3GalβpNP in all binding poses (A-D) of SARS-CoV-2 S protein, when available, with the top 3 binding residues are highlighted. Two binding energies are written, one averaged across MD simulations, and the second binding energy from the last equilibration step in the SOMD complex setup is for reference, where the ligands are close to docked locations. The exception is pose B with weak binding and only MM-PBSA energy reported for the SOMD equilibration step. Error bars are plotted from standard error of means.



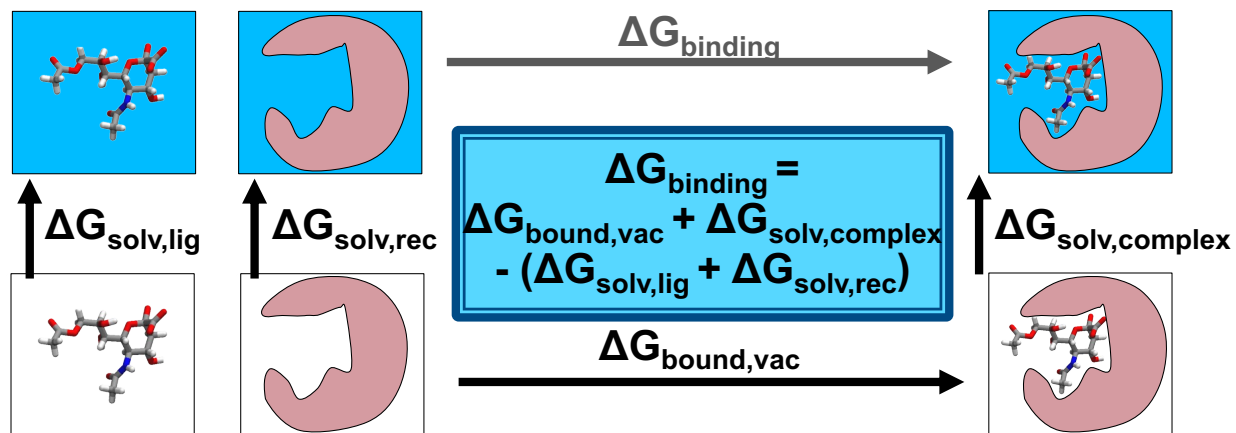
Energy decomposition for Sias in binding pose C



Energy decomposition for Sias in binding pose D



**Figure S4.3:** Thermodynamic cycle to estimate binding free energies using MM-PBSA, with Neu5,9Ac<sub>2</sub> in protein receptor as example.



**Figure S4.4:** Representative binding free energy difference of Neu5Ac — Neu5,9Ac2 in the SARS-CoV-2 S protein, where the transformation is listed as final – initial Sia. Left side plots display the MBAR and TI energies versus  $\lambda$ . Right side plots display the lower triangles for simulation overlap matrices, where the first off-diagonal quantifies the overlap of one simulation window with the next. Upper plots show analysis for the alchemical transformation of Neu5,9Ac2 to Neu5Ac when bound to the protein and in explicit solvent, and lower plots show results for this transformation in explicit solvent. The binding free energy difference of the Neu5,9Ac2 to Neu5Ac is taken as the difference in energies of the transformations in the bound and solvated systems. Multiple replicates are shown, when available. Plots excluded from analysis when MBAR and TI differ significantly, or in cases where the ligand unbinds during the simulations.

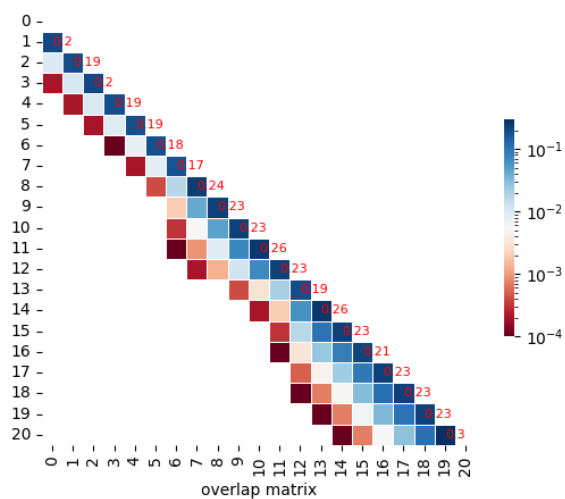
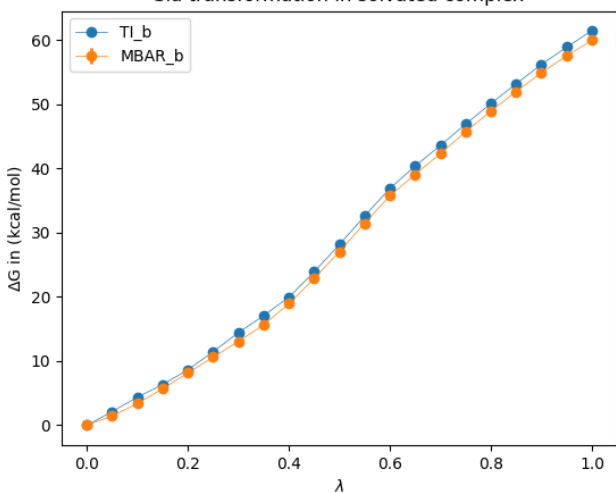
# Sias binding in SARS-CoV-2 S protein pose B

Neu5Ac–Neu5,9Ac<sub>2</sub>

$\Delta\Delta G$  (MBAR): -2.2 kcal/mol

$\Delta\Delta G$  (TI): -1.4 kcal/mol

Sia transformation in solvated complex



Sia transformation in solvent

