

CLASSICAL FORCE FIELD SIMULATIONS OF  
BIOLOGICAL PROCESSES AND QUANTUM CHEMICAL  
COMPUTATIONS OF HOMOGENEOUS CATALYSTS

Thesis by  
Fan Liu

In Partial Fulfillment of the Requirements for  
the Degree of Doctor of Philosophy

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font, centered within a light orange rectangular background.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2016  
(Defended: May 26, 2016)

© 2016

Fan Liu

ORCID: 0000-0001-5650-2809

All rights reserved

## ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. Bill Goddard for teaching me how to solve problems and for his inspiration.

I would like to thank my thesis committee: Prof. Rudy Marcus, Prof. Harry Gray and Prof. Dennis Dougherty, for their kindness and support.

I would like to thank my teachers: Ms. Guo Gui, Mr. Libo Zhan, Mr. Changjiang Wan, Prof. Shiu-chun Wong, Prof. Jin Au Kong, Prof. Jian Wu, Prof. Tao Wu, Prof. Guanhua Chen, Prof. Jianshu Cao, Prof. Kieron Burke, Prof. Filipp Furche.

I would like to thank my family for their love.

## ABSTRACT

Computational chemistry methods and tools have enabled studies of biological processes and chemical reactions to get insights from detailed atomic structures and reaction mechanisms. In this thesis, two biological problems are attacked by the classical force fields simulations and two homogeneous catalysis problems are studied by quantum chemical calculations. In all four problems, new insights have been revealed by the computational results.

Chapter 1 briefly reviews the computational chemistry theories and methods developed and popularized in the past few decades. Chapter 2 addresses the protein-protein interaction problem in the onset of meningitis where *E. coli* OmpA interacts with Fc $\gamma$ RI  $\alpha$ -chain (Fc $\gamma$ RIa) to invade macrophages. Computationally predicted three-dimensional structure of the OmpA-Fc $\gamma$ RIa complex showed the role of three N-glycans in Fc $\gamma$ RIa in the interaction. Chapter 3 studies the molecular origin of the bitter aftertaste of a kind of natural sweetener called steviol glycosides. By examining the predicted binding complexes between the human bitter taste receptors 2R4 and 2R14 which could be activated by steviol glycosides, a general activation model is proposed to explain the structure-function relationship and to predict new natural sweeteners with less bitterness. Chapter 4 investigated the reaction mechanisms of methane to methanol conversion by a biomimetic tricopper cluster compound. An unusual exchange-stabilized multiradical state is found to be responsible for the hydrogen abstraction reactivity and a methyl radical rebound mechanism is proposed for methane oxidation. Calculations also show interesting spin crossing during the reaction cycle with high spin state forbidden for methyl rebound. Chapter 5 examines the reaction mechanisms in olefin hydrosilylation by the Pt-based Karstedt's catalyst. An unexpected rate-determining step of agostic bond dissociation is found in between the elementary reaction steps proposed previously. The regioselectivity of the products are studied. An alternative reaction cycle which is kinetically unflavored is proposed. Oxygen stability is studied.

## PUBLISHED CONTENT AND CONTRIBUTIONS

Krishnan, Subramanian and Liu, Fan and Abrol, Ravinder and Hodges, Jacqueline and Goddard, William A., III and Prasadarao, Nemani V. (2015). "The Interaction of N-glycans in Fc-gamma Receptor I Alpha Chain with Escherichia coli K1 Outer Membrane Protein A for Entry into Macrophages -- Experimental and Computational Analysis". In: *The Journal of Biological Chemistry* 289(45):30937-49. doi: 10.1074/jbc.M114.599407.

Fan Liu predicted and analyzed the structures, prepared the data, and participated in the writing of the manuscript.

## TABLE OF CONTENTS

Acknowledgements.....	iv
Abstract .....	v
Published Content and Contributions.....	vi
Table of Contents.....	vii
List of Figures.....	viii
List of Tables.....	x
Chapter I: Introduction .....	1
1.1 Classical Force Field Simulation .....	1
1.2 Quantum Chemistry Simulation .....	2
Chapter II: The Interaction of N-glycans in Fc-gamma Receptor I Alpha Chain with Escherichia coli K1 Outer Membrane Protein A for Entry into Macrophages.....	5
Abstract.....	5
2.1 Introduction .....	6
2.2 Simulation Details .....	8
2.3 Results and Discussion.....	17
2.4 Conclusions .....	24
Chapter III: The interaction of natural sweetener stevia glycosides with human bitter taste receptors TAS2R4 and TAS2R14.....	26
Abstract.....	26
3.1 Introduction .....	27
3.2 Simulation Details .....	28
3.3 Results and Discussion.....	31
3.4 Conclusions .....	42
Chapter IV: Reaction mechanisms of methane oxidation by the biomimetic tricopper cluster complexes.....	44
Abstract.....	44
4.1 Introduction .....	45
4.2 Simulation Details .....	50
4.3 Results and Discussion.....	51
4.4 Conclusions .....	60
Chapter V: Mechanism for the Karstedt catalyzed hydrosilylation reaction ...	61
Abstract.....	61
5.1 Introduction .....	62
5.2 Simulation Details .....	65
5.3 Results and Discussion.....	66
5.4 Conclusions .....	70
Bibliography .....	72

## LIST OF FIGURES

<i>Number</i>	<i>Page</i>
2.1 Ensemble of 15 OmpA conformations for docking .....	9
2.2 Crystal structure of FcγR15 and its N-glycosylation sites .....	11
2.3 Protein-protein complex structure analysis in 10-digit code.....	13
2.4 Top poses for OmpA-FcγR1 complex.....	16
2.5 N-linked glycans 1, 4 and 5 of FcγR1 in binding OmpA.....	20
2.6 Glycan 1 forms a hydrogen bond with D149 of OmpA.....	21
2.7 Glycan 4 forms indirect interaction with OmpA.....	22
2.8 Rotating the torsional angles of glycan 4 forms direct interaction .....	23
3.1 Predicted Transmembrane helical bundle structure of TAS2R4.....	32
3.2 Predicted Transmembrane helical bundle structure of TAS2R14.....	33
3.3 Predicted structure of 2R4 bound with its agonist parthenolide .....	34
3.4 Predicted structure of 2R14 bound with its agonist aristolochic acid. .	34
3.5 The general framework of steviol glycosides.....	35
3.6 Predicted structure of 2R4 bound with rubusoside. ....	36
3.7 Predicted structure of 2R14 bound with rubusoside. ....	37
3.8 Histogram of distribution of 94 molecules in Cargill priority list 1 ....	38
3.9 Histogram of distribution of 151 molecules in Cargill priority list 2 ..	39
3.10 Rule #1 for selecting low bitter targets.....	41
3.11 Rule #2 for selecting low bitter targets.....	42
4.1 The structures of pMMO, its active center and the designed ligand....	47
4.2 Productive cycling of the tricopper catalyst .....	49
4.3 Dioxygen activation of the Cu(I)Cu(I)Cu(I) complex.....	53
4.4 Transformation from Cu(I)Cu(I)Cu(II)(superoxo) to the active state..	54
4.5 The four natural orbitals from CASSCF calculation .....	55
4.6 Free energy profile of the methane activation .....	57
4.7 Enthalpy profile of the hydrogen peroxide activation .....	59

5.1 Redox cycles for olefin hydrosilylation.....	64
5.2 The free energy profile of the Pt(II) reductive elimination route.....	67
5.3 The free energy profile of several competing reaction routes.....	69
5.4 The structures of O <sub>2</sub> bound pre-catalyst and intermediates.....	70

## LIST OF TABLES

<i>Number</i>	<i>Page</i>
2.1 The output files of search for the matching poses.....	14
2.2 Interactions in the protein-protein interface of the best pose.....	17



*Chapter 1*

## INTRODUCTION

Computational chemistry has advanced in the past decades to reach a stage that it acquired explanatory, descriptive, and even predictive power to chemical and biological problems, due to the development of new computational methods, simulation software, and computer technology. Various computational chemistry methods have been developed for physical problems at different time and length scales. Classical force fields have been developed for small organic molecules and later for macromolecules like proteins, allowing all-atom molecular dynamics simulations of protein folding events in explicit water solvent. Quantum chemistry methods, from wave-function based methods to density functional theory, have enabled routine energy calculations and geometry optimizations for molecules, solid, and interfaces up to 100 atoms. These methods have been used to understand protein dynamics and reaction mechanisms, and the hybrid of these methods has been developed to understand enzyme reactions.

This chapter introduces the essence and the efficacy of the theories and applications of classical force field simulations and quantum chemistry computations.

**1. 1 Classical Force Fields Simulations**

Classical force fields simulations are based on the fundamental idea that the energy of a molecule could be represented as a function of its atomic coordinates. The exact energy surface on which the atoms move is the Born-Oppenheimer ground-state energy obtained from averaging the much-faster electronic motions. For macromolecules like proteins, it is not yet efficient to calculate such surfaces directly from quantum chemistry methods with high accuracy, thus approximations are made to use a set of simple classical functions to represent the energy, leading to practical “force field” methods to calculate the energy of systems of thousands to trillions of atoms and forces on individual atoms through molecular mechanics and dynamics<sup>1</sup>.

The parameters that go into the simple classical energy function terms in a force field are optimized in agreement with various sets of experimental data and also with accurate energies of small molecules calculated from quantum chemistry methods. The common force field for proteins has the following potential energy function:

$$\begin{aligned}
 V(r) = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\
 & + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1] \\
 & + \sum_{nonbond\ pairs} \left[ \frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right]
 \end{aligned} \tag{1}$$

The first two terms shown in equation (1) represent bond length stretching and bond angle bending as harmonic springs. The third term is the torsion term which could account for both the dihedral torsions and the improper torsions. The last term describes electrostatics via Coulomb's law and the van der Waals interaction by a Lennard-Jones 6-12 potential. With different terms using different parameters which are transferable between molecules, one could calculate all static, dynamic, and thermodynamic properties of such molecular systems.

A large number of force fields have been developed for simulating proteins. The earlier force fields include the ECEPP potentials from Scheraga<sup>2</sup>, the consistent force field (CFF) from Lifson<sup>3</sup>, and the potential energy functions of organic molecules from Allinger<sup>4</sup>. The current standard force fields for simulations of proteins include the Amber force fields<sup>5</sup>, the CHARMM force fields<sup>6</sup>, and the OPLS force fields<sup>7</sup>.

## 1.2 Quantum Chemical Computations

Quantum chemical methods aim to solve the electronic Schrodinger equation and calculate the electronic energy and other properties for a system of given nuclei positions and total number of electrons. To get the exact wave-function solution of the electronic Schrodinger equation is an infeasible task, as the computational complexity grows exponentially with the

number of electrons. There exist two ways to circumvent this “exponential curse”: one is to work with approximated wave-functions, the other is to work with electron densities.

Wave-function based methods starts with a mean-field approximation called Hartree–Fock theory<sup>8</sup>, where a single Slater determinant is optimized variationally to approximate the exact wave-function, assuming each electron moves independently of all the other electrons except the Coulomb repulsion and the exchange interaction due to antisymmetrization. This method could calculate the molecular orbitals due to its independent particle approximation, but for the reason it does not include the correlations between electrons due to their instantaneous motions. To get a robust description of the electron correlation energy, more elaborate theoretical methods use Hartree-Fock results as a starting point to calculate correlated wave-functions. These post-Hartree-Fock methods include second-order Moller-Plesset perturbation theory (MP2)<sup>9</sup>, methods based on the coupled-cluster ansatz<sup>10</sup>, and multireference methods such as Complete Active Space Self-Consistent Field (CASSCF)<sup>11</sup>.

Instead of working with approximated wave-functions, density functional theory (DFT) works with electron density, which is a function of only three Cartesian coordinates as opposed to  $3N$  coordinates of  $N$  electrons<sup>12</sup>. The Hohenberg–Kohn theorem<sup>13</sup> establishes that there exists a universal mapping between the electron density and the electronic energy, in other words, that the total energy of the system is a functional of the electron density. Kohn and Sham<sup>14</sup> offered a recipe of calculating the ground state energy by the virtue of Hohenberg-Kohn theorem, as shown in equations (2)-(4):

$$E = T_s + U + V_{nuc} + E_{xc}[\rho] \quad (2)$$

$$\left( -\frac{1}{2}\nabla^2 + v_s(r) \right) \phi_j(r) = \epsilon_j \phi_j(r) \quad (3)$$

$$v_s(r) = v_{nuc}(r) + \int d^3r' \frac{\rho(r')}{|r - r'|} + v_{xc}[\rho](r) \quad (4)$$

A Kohn-Sham calculation uses a single Slater determinant function as an auxiliary wavefunction to calculate the non-interacting kinetic energy term  $T_s$  in equation (2), sweeping the kinetic part of the electron correlation energy in the  $E_{xc}$  which is the exchange-correlation energy term that makes equation (2) exact. The so-called Hartree term  $U$  is the classical Coulomb energy,  $V_{nuc}$  is the electron-nuclei attraction energy. The single Slater determinant auxiliary orbitals are optimized in equation (3), which is the famous Kohn-Sham equation, and the Kohn-Sham potential  $v_s$  is explicitly expressed in equation (4). Knowing the universal functional in Hohenberg–Kohn theorem is equivalent to knowing the form of functional  $E_{xc}$ .

There are two major schools of thoughts to approximate the exchange-correlation energy  $E_{xc}$ . One school uses exact conditions required by quantum mechanics to derive the parameters so that their approximate functionals satisfy these exact conditions. This school of thought, pioneered by Perdew and Burke, has yielded successful functionals like PBE<sup>15</sup> that is favored by physicists doing material calculations. The other school uses highly accurate results of specific systems, such as noble gas atoms, to fit the parameters in their approximated functionals. This school, lead by Becke, Parr and Yang, have generated functionals like B3LYP<sup>16,17</sup> that is popular among chemists for molecular calculations.

## Chapter 2

### THE INTERACTION OF N-GLYCANS IN FC-GAMMA RECEPTOR I ALPHA CHAIN WITH ESCHERICHIA COLI K1 OUTER MEMBRANE PROTEIN A FOR ENTRY INTO MACROPHAGES

#### **Abstract**

Neonatal meningitis, due to *Escherichia coli* K1, is a serious central nervous system disease. It has been established that macrophages serve as permissive niches for *E. coli* K1 to multiply in the host and for attaining a threshold level of bacterial load, which is a pre-requisite for the onset of the disease. It has been demonstrated experimentally that three N-glycans in Fc $\gamma$ RIa interact with OmpA of *E. coli* K1 for binding to and entering the macrophages. Our molecular dynamics and simulation studies predict that N-glycan 5 exhibits strong binding at the barrel site of OmpA formed by loops 3 and 4 while N-glycans 1 and 4 interact with the tip regions of the same loops. Molecular modeling data also suggest no role for the IgG binding site in the invasion process. In agreement, experimental mutations in IgG binding site had no effect on the *E. coli* K1 entry into macrophages in vitro or on the onset of meningitis in newborn mice. Together, this integration of experimental and computational studies reveals how the N-glycans in Fc $\gamma$ RIa interact with the OmpA of *E. coli* K1 for inducing the disease pathogenesis.

## 2.1 Introduction

In the event of bacterial invasion of host tissues, the intruder encounters an arsenal of host-defense mechanisms, which result in either resolution of the pathogen by the host or subversion of the defense by the pathogen. Cross talk between host defense components is critical for successful resolution of the infection. Macrophages are long-lived cells that play a critical role in engulfing pathogenic microorganisms and degrading them. They express a range of receptors that recognize bacteria, including TLRs, Fc- $\gamma$  receptors, complement receptors, scavenger receptors and mannose receptors<sup>18,19</sup>. Most microbial structures are recognized by more than one macrophage receptor and these receptors also interact with each other. Several pathogens subvert the anti-microbial mechanism by exploiting receptor interactions to create their own safe havens inside of which they survive. The bacterial pathogens upon entering the host are coated with complement proteins enabling them to be recognized by complement receptors on the macrophages<sup>20</sup>. Similarly, antibody coated bacterial recognition involves interaction with Fc-  $\gamma$  receptors, which subsequently elicits anti-microbial mechanisms. A large number of bacteria introduce microbial factors that govern macrophage function by type III or type IV secretions systems<sup>18</sup>. However, very few bacteria control the attack of macrophages at the receptor level. One example is *S. aureus*, which uses protein A to bind to the Fc region of IgG, thus avoiding recognition by Fc- $\gamma$  receptors<sup>21</sup>.

Several studies have shown that Fc $\gamma$ RI expression increases during septicemia and meningitis caused by a variety of bacteria. The Fc region of IgG recognizes Fc $\gamma$  receptors (Fc $\gamma$ R) enabling it to play an important role in linking the cellular and humoral immune response. Fc $\gamma$ R comprises a multigene family divided into 3 classes (Fc $\gamma$ RI, II, and III), which are defined by their affinity for IgG. Fc $\gamma$ RI is a transmembrane receptor that binds IgG with high affinity and induces the association of the  $\gamma$ -chain for signal transduction and triggering of effector responses such as macrophage phagocytosis. The ligation of Fc $\gamma$ RI with IgG also mediates antibody dependent cellular cytotoxicity induced transcription of cytokine genes and release of inflammatory mediators. It has been previously demonstrated that outer membrane protein A (OmpA) of *E. coli* K1, which causes neonatal meningitis, directly

interacts with Fc- $\gamma$  receptor I alpha chain (Fc $\gamma$ RIa) to bind to and enter macrophages. Indeed, depletion of macrophages or lack of Fc $\gamma$ RIa expression in macrophages in newborn mice renders the animals resistant to *E. coli* K1 induced meningitis. Despite the general requirement of Fc $\gamma$ RIa association with the  $\gamma$ -chain for the internalization of the receptor, the interaction of OmpA+ *E. coli* with Fc $\gamma$ RIa and the subsequent entry into macrophages does not require the  $\gamma$ -chain to facilitate *E. coli* K1 entry into macrophages, which is a novel mechanism. *E. coli* K1 interaction with macrophages in the absence of Fc $\gamma$ RIa induces the expression of complement receptor 3 (CR3), which elicits antimicrobial mechanisms to kill the intracellular bacteria<sup>22</sup>. In addition, macrophages generate biopterin and neopterin upon *E. coli* K1 infection to suppress the production of nitric oxide and superoxide, respectively<sup>23</sup>. We showed previously that mutation of three amino acids in loops 1 and 3 of the extracellular domains of OmpA prevented the bacterial survival in macrophages. Concomitantly, *E. coli* K1 containing a mutation in loop 1 could not cause meningitis in the newborn mouse model<sup>24</sup>. However, there has been no molecular level understanding of how OmpA interaction with Fc $\gamma$ RIa controls these cellular events.

OmpA has been shown to interact with GlcNAc1-4GlcNAc epitopes of host receptors<sup>25,26</sup>. In addition, our previous molecular modeling predictions of OmpA interaction with GlcNAc1-4GlcNAc epitopes demonstrated that this moiety can bind to OmpA at two sites, one at the tip of loops 1 and 2, and the second at the barrel site formed by loops 3 and 4<sup>22</sup>. We now report investigations on the role of N-glycans in Fc $\gamma$ RIa in *E. coli* K1 entry of macrophages and for the onset of meningitis in the newborn mouse model. The experimental studies show that N-glycosylation sites 1, 4, and 5 of Fc $\gamma$ RIa are critical for interacting with OmpA of *E. coli* K1, both for binding to and entry of macrophages. Adoptive transfer of Fc $\gamma$ RIa<sup>-/-</sup> macrophages transfected with N-glycosylation (NG) mutants of Fc $\gamma$ RIa into Fc $\gamma$ RIa<sup>-/-</sup> mice revealed that the presence of Fc $\gamma$ RIa with 1, 4, or 5 NG sites are important for the onset of meningitis. To determine a structural basis for the experimental results, we conducted computer simulations to predict the atomistic structure for the OmpA protein complexed with the glycosylated Fc $\gamma$ RIa, which identified how N-glycans contribute to the *E. coli* K1 interaction with macrophages.

## 2.2 Simulation Details

**Protein-protein docking protocol:** A hybrid docking strategy was adopted that first considered an ensemble of 15 protein conformations for OmpA, each of which was docked to Fc $\gamma$ RIa to obtain an ensemble of 810,000 OmpA-Fc $\gamma$ RIa protein-protein poses through exhaustive rotational and translational sampling of the relative position space that satisfies shape complementarity of the two given protein conformations. Then experimental observations and topological constraints were used to dramatically reduce the most likely poses to just a few that could be subjected to more detailed considerations including molecular dynamics. This strategy overcomes the limitations of finding scoring functions that can reliably rank the energies for protein-protein interactions and it reduces the issues regarding the conformational flexibility of glycosylated amino acids.

First, an ensemble of 15 structures of OmpA that consider the flexibility of its four loops was predicted (Figure 2.1). To construct this ensemble, (a) we selected as one candidate the PDB: 1BXW X-ray crystal structure of OmpA, which resolves more of the loop regions than the PDB: 1QJP X-ray crystal structure<sup>27,28</sup>. (b) In addition, 2.5 nanosecond molecular dynamics (MD) simulation of the protein in explicit membrane and water (including salt) was carried out, starting from a conformation previously generated from the 1BXW X-ray crystal structure equilibrated with MD at 300K. Here the protein was embedded in a periodically infinite box containing the 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) lipid system fully solvated with water and salt. We used the CHARMM26 forcefield method for the lipid and OmpA protein and the TIP3P water forcefield<sup>29-31</sup>. The VMD Membrane Plugin to build the membrane and the VMD

Solvate Plugin to solvate the protein at both the intracellular and extracellular regions was used<sup>32</sup>. All MD simulations used the NAMD program<sup>33</sup> with the NPT constant temperature-pressure ensemble and periodic boundary conditions at 300K and 1 atm. From this trajectory 2500 conformations (1 snapshot every 1 picosecond) have been selected, which we clustered into 13 diverse structures using an RMSD Voronoi criterion of 1.5 Å from which we selected the 13 family heads. (c) In addition, an ensemble of 10 candidate structures based on distance-geometry fits to NMR (PDB number: 1G90)<sup>34</sup> is available and their energies were minimized in vacuum using the DREIDING force field<sup>35</sup>. From this the one (number 4) with the lowest total energy was selected.

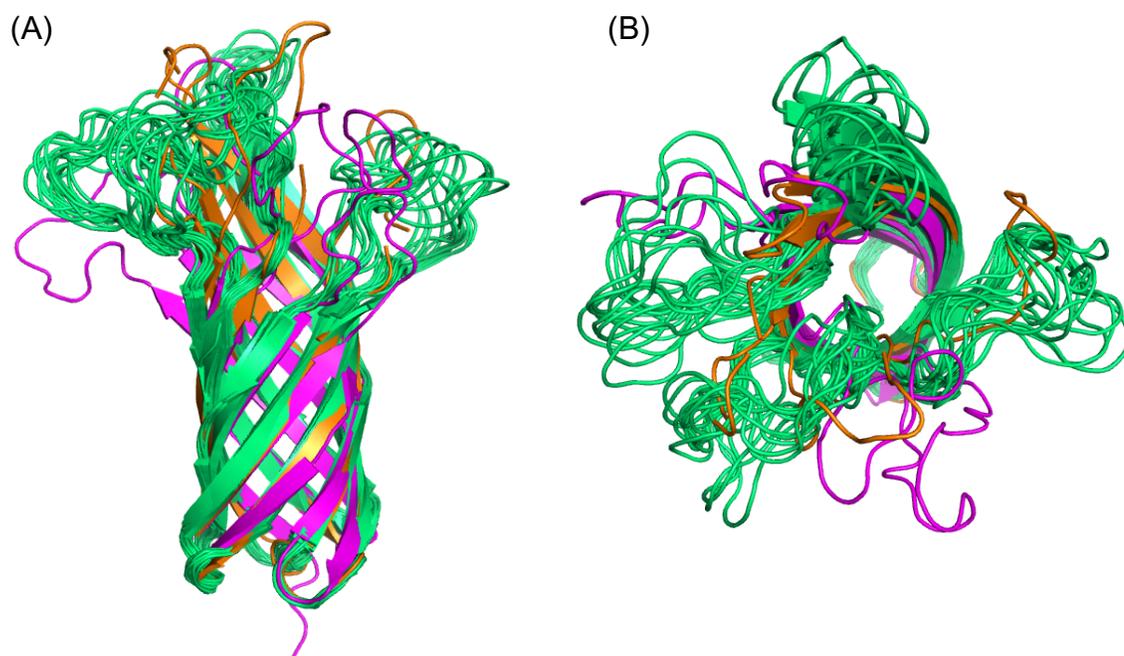


Figure 2.1: 15 OmpA conformations for docking: 13 conformations from 1.5 Å RMSD distance clustering of 2500 snapshots from 2.5 ns MD (green), 1 X-ray crystal structure (orange, 1BXW), 1 NMR lowest DREIDING force field energy structure (magenta, 1G90\_3). (A) side view, (B) top view.

These 15 OmpA conformations have a minimum RMSD diversity of 1.5Å in the loop regions. Next, we examined the crystal structure of FcγRIa (PDB number: 3RJD)<sup>36</sup>. The interaction between the D1 and D2 fragments in the crystal is well defined (Figure 2.2A). Only the Ig-like domains D1 and D2 of the FcγRIa crystal structure were taken for docking purposes based on the experimental evidence about the N-glycosylation sites. Then the Man-Manβ1–4GlcNAc β1–4GlcNAcβ1 portion of N-glycans to each of the five N-glycosylation sites on D1 and D2 (residues 59, 78, 152, 159, and 163, predicted by NetNGlyc 1.0 Server) was added. Further, we relaxed the backbone of D1 and D2 with 1ns MD in a solvent box of explicit water (including 154 mM NaCl or 0.9% w/w NaCl) while the side chains including the N-glycans were relaxed with 5ns of MD in explicit water. The snapshot closest to the average conformation was then chosen for docking (Figure 2.2B).

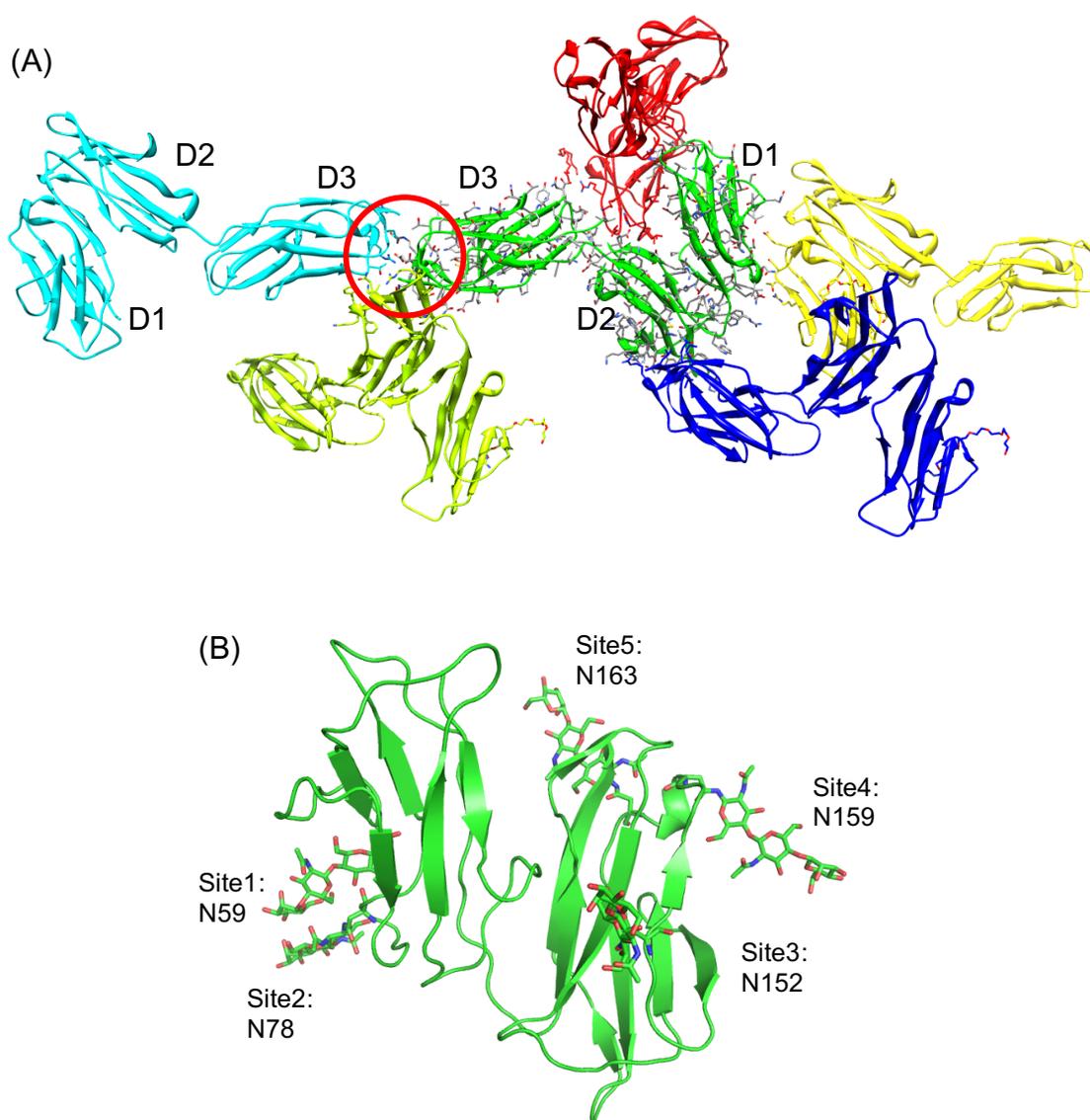


Figure 2.2. (A) In the crystal packing of FcγR1, two neighboring D3 domains form 4 salt-bridges, which may cause artifact in the conformation of D3 domain. (B) 5 N-glycosylation sites in D1 and D2 domains of FcγR1. Glycan conformations are taken from the centroid conformation of 5ns MD trajectory.

To describe the glycosylated Asn consistently with natural amino acids in the MD, the AMBER ff99SBildn (improved side-chain torsion potentials for the Amber ff99SB) forcefield for the natural residues in the protein and the general AMBER force field

(GAFF) for the glycosylated Asn were used<sup>37,38</sup>. The Antechamber auxiliary program was used to prepare the GAFF library of the glycosylated Asn<sup>39</sup>. The TIP3P water model and the AmberTools<sup>40</sup> including a terminal interface called tLEAP were used to add the ions and the water solvation box (<http://ambermd.org/>).

Next, all 15 OmpA conformations were docked to the FcγRIa D1 and D2 structure using only the protein-protein complex pose generation part of ZDOCK<sup>41,42</sup>. The ZDOCK shape complementarity algorithm generated 54,000 complex poses with exhaustive rotational and translational sampling for each configuration of OmpA. Combining this with the 15 separate docking procedures yielded 810,000 complex poses (Table 2.1A). We then eliminated all poses in which FcγRIa would have a likely clash with the virtual membrane where OmpA is buried. This virtual membrane screening left us 496,100 poses. These poses were analyzed to determine which glycans and loops are involved in binding and the number of inter-protein salt-bridges within each pose (Figure 2.3). After eliminating poses that did not agree with the experimental evidence, we obtained 326 poses (Table 2.1B).

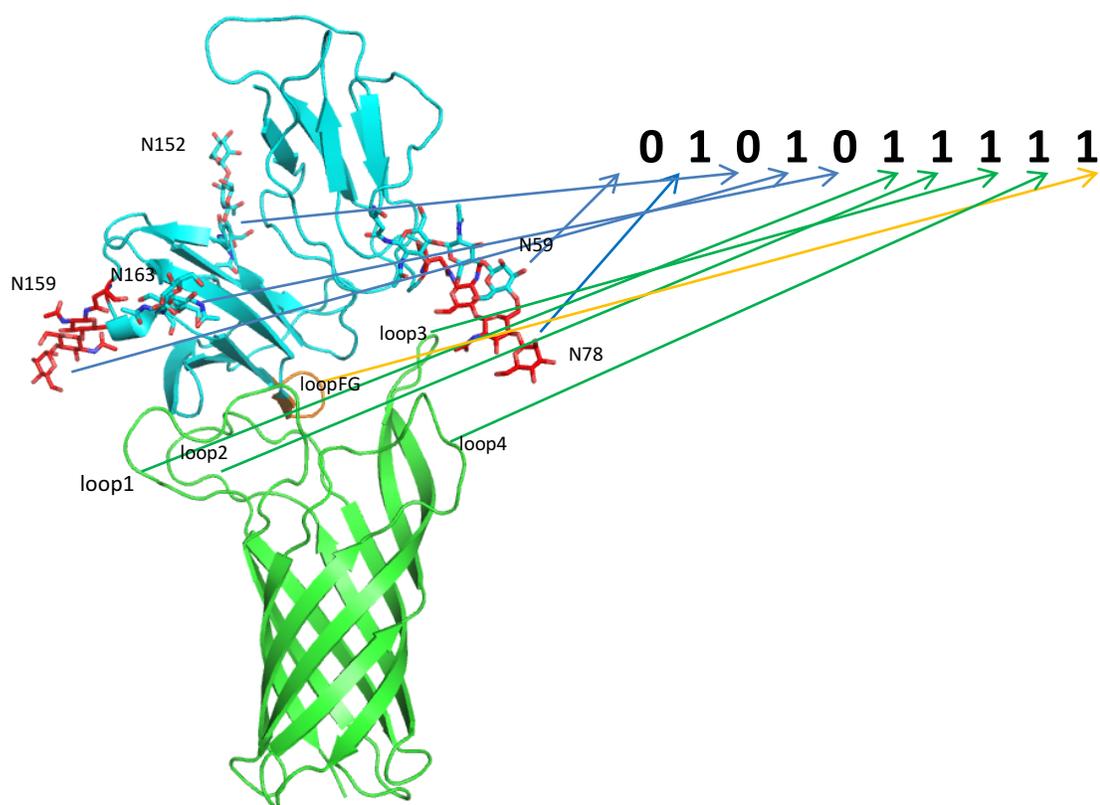


Figure 2.3: For each protein-protein complex pose that has passed virtual membrane screening, we generated a 10-digit code that represents whether the five glycans of FcγR1 (D1 and D2 only), the four OmpA loops and the FG loop of FcγR1 are involved in binding. If one of the above ten moieties (in the order it is mentioned) is involved in binding, the digit in the corresponding position is 1, otherwise it is 0. Also, the number of inter-protein salt-bridges and the residues that compose the salt-bridges are recorded for each pose. In this example, the analysis shows that glycan 2 and 4 (N78 and N156 highlighted in red), OmpA loop 1, 2, 3, and 4, and the FG loop of FcγR1 are involved in binding.

(A)

<b>Input code: -----</b>				
(Use '-' if binding is undetermined)				
Total <b>496100</b> /496100= 100%				
SB#	POSE#	ACCUM#	POSE%	ACCUM%
17	2	2	0.00%	0.00%
16	4	6	0.00%	0.00%
15	14	20	0.00%	0.00%
14	20	40	0.00%	0.01%
13	71	111	0.01%	0.02%
12	196	307	0.04%	0.06%
11	428	735	0.09%	0.15%
10	948	1683	0.19%	0.34%
9	1886	3569	0.38%	0.72%
8	4234	7803	0.85%	1.60%
7	8211	16014	1.70%	3.20%
6	15269	31283	3.10%	6.30%
5	28897	60180	5.80%	12%
4	53039	113219	11%	23%
3	82712	195931	17%	39%
2	103276	299207	21%	60%
1	108100	407307	22%	82%
0	88793	496100	18%	100%

(B)

<b>Input code: 1--111-11-</b>				
Info: Glycan1,4,5, loop1,3,4 in binding				
Glycan2,3, loop2, loopFG undetermined				
Total <b>326</b> /496100= 0.066%				
SB#	POSE#	ACCUM#	POSE%	ACCUM%
8	1	1	0.31%	0.31%
7	4	5	1.20%	1.50%
6	10	15	3.10%	4.60%
5	38	53	12%	16%
4	37	90	11%	28%
3	69	159	21%	49%
2	75	234	23%	72%

1	40	274	12%	84%
0	52	326	16%	100%

Table 2.1: The output files from using experimental results as an input to search for the matching poses. (A) If none of the ten pieces of experimental binding information is used (represented by ten '-'s as undetermined input), all 496100 poses in the pool would be selected. This is served as a sanity check of the code and algorithm. (B) The positive experimental binding information is used to select the matching poses. Input code '1--111-11-' means that FcγR1 N-glycan sites 1, 4, and 5 ('1's in first 5 digits), OmpA loops 1, 3, and 4 ('1's in 6<sup>th</sup> to 9<sup>th</sup> digits) are involved in binding, while binding information of FcγR1 N-glycan sites 2 and 3, OmpA loop 2, and FG loop of FcγR1 are assumed as unknown.

**Poses matching experimental input:** These 326 hits were then ranked by their number of inter-protein salt bridges, leading to one pose with a maximum of 8 salt-bridges (Figure 2.4). This pose of OmpA-FcγRIa (D1-D2 only) was selected as the structure likely to have the strongest protein-protein structure that is consistent with all experimental data. Then the D3 domain of FcγRIa as it is in the crystal structure was added and the consequent clashes between several loops were resolved by resampling the loop conformations using the DREIDING forcefield. This new structure is referred to as OmpA-FcγRIa for the remainder of the manuscript. We then subjected this OmpA-FcγRIa structure to 50 ns of NPT constant temperature-pressure ensemble at 300K and 1 atm MD simulation in a periodic box with explicit lipid membrane (POPC) and explicit water with 154 mM NaCl, where the position of OmpA in the membrane is calculated according to the implicit solvent model of the lipid bilayer in the Orientations of Proteins in Membranes (OPM) database<sup>43</sup>. In order to allow the predicted pose to relax any strains that might arise in the newly formed protein-protein interface from our rigid-body docking procedure, we carried out molecular dynamics (MD) for 50 ns on the full system including explicit water and ions. We found that all 8 salt bridges and the other polar

interactions between the subunits are maintained during the 50 ns of MD, indicating that the predicted protein-protein complex structure is stable (Table 2.2).

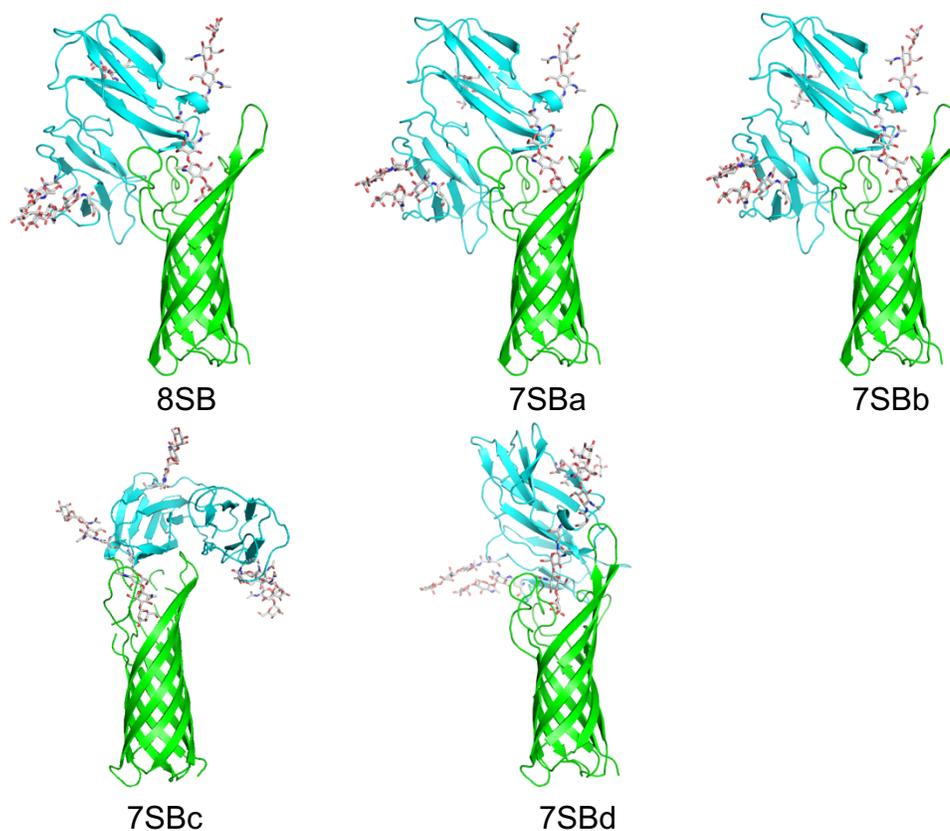


Figure 2.4: Poses 8SB, 7SBa, 7SBb, 7SBc, and 7SBd have a similar overall binding mode between OmpA and FcγR1.

Mutation index	OmpA	FcγR1	Type of interaction	Compare to experiments	Suggested mutations (To combine mutations on both proteins)		
					OmpA	FcγR1	Effect of mutation
1	ASP20 (1c)	LYS186 (D2)	Salt-bridge	OmpA loop 1c region is important for bacterial survival in microphage [1], thus is involved in binding FcγR1	D20A	K186A	Decrease binding
2	ASN27 (1a)	ARG220 (D3)	Hydrogen bond	1a is involved in binding	N27D		Increase binding
3	LYS113 (in loop3, but not in 3a)	GLU187 (D2)	Salt-bridge		K113A	E187A	Decrease binding
4	ARG156 (in loop4, but not in 4abc)	GLU82 (D1)	Salt-bridge	All-ALA mutant 4c (157-159 on loop4) exhibited Bacterial survival ability in microphage, because salt-bridge involving ARG156 compensates the lost salt-bridge by mutating ASP158 to alanine	R156A	E82A	Decrease binding
5	ASP158 (4c)	ARG84, ARG87 (D2)	Salt-bridge		D158A	R84A, R87A	Decrease binding

Table 2.2: Specific interactions in the protein-protein interface of the best pose.

## 2.3 Results and Discussion

### Molecular simulation of OmpA and glycosylated FcγRIa interaction supports that

**NG1, NG4 and NG5 sites are important for binding:** Based on the experimental data

that NG sites 1, 4, and 5 in FcγRIa are critical to interacting with OmpA; simulation studies were performed to unravel molecular bases for this interaction. As described in experimental procedures, the FcγRIa structure was assembled initially using D1-D2 domains and the D3 domain was added finally by resolving loop clashes (Figure 2.2).

The orientation of NG sites 1 to 5 obtained from the centroid conformation of 5ns MD trajectory is shown in Figure 2.2. The modeling predicted a large number (810,000) of docking poses for the protein-protein complex, which was done without inputting

information from experiment (Table 2.1A). These structures were analyzed on the basis of the experiments to obtain 326 poses consistent with the experimental mutation studies (Table 2.1B). We then analyzed these to find that 1 led to 8 salt bridges, while 4 led to 7 salt bridges, while all others had fewer (Figure 2.4). We found that 4 of these 5 structures were similar and chose the one with 8 salt bridges to examine in more detail. The resulting 3D structure supports interpretations of the experiments that N-glycan sites 1, 4, and 5 interact with OmpA. The transmembrane region of OmpA was then embedded in the POPC membrane with the rest of the complex and solvated in water. Then protein-protein complex was then subjected to 50ns of MD simulation. The 50ns MD trajectory on the OmpA-FcγRIa complex led to stable and reasonable interactions for the 5 important salt bridges, each of which remained stable during 50ns of MD (Figure 2.5, Table 2.2). In addition to many new water-mediated hydrogen bonds showing up in the MD, we found that NG1 and NG5 formed direct non-covalent interactions with OmpA (Figure 2.6). Although NG4 is close to OmpA, the 50ns MD simulation of OmpA-FcγRIa did not lead to any direct interaction between the glycan of NG4 and OmpA (Figure 2.7). Instead the predicted NG4 conformation was locked into a hydrogen bond within FcγRIa between the side chain carbonyl group of Asn-159 and the backbone amino group of Ser-161. This hydrogen bond remained stable during 50ns MD trajectory. This lack of interaction of NG4 with OmpA was inconsistent with the experimental data as lack of NG4 also prevented the bacterial invasion into macrophages. However, we speculated that the strong interaction with Asn 159 and Ser 161 might be accidental, caused by our initial configuration. To test this we rotated the side chain torsional angles of Asn-159 (the  $\chi_1$  angle from 66 degrees to 145 degrees and the  $\chi_2$  angle from 75 degrees to 30

degrees) to form direct interactions between NG4 and OmpA. This caused the hydrogen bond between Asn-159/Ser-161 to break, forming instead three new hydrogen bonds (one between carbonyl of Asn-159 side chain and backbone amino group of Ile-160 within FcγRIa, and another two between the second GlcNAc moiety in NG4 and Asn-109 on loop 3 of OmpA). This new conformation of NG4 does not clash with any of the protein-protein interfacial peptides. We carried out 50 ns of simulation with this new structure and found it to be stable (Figure 2.8).

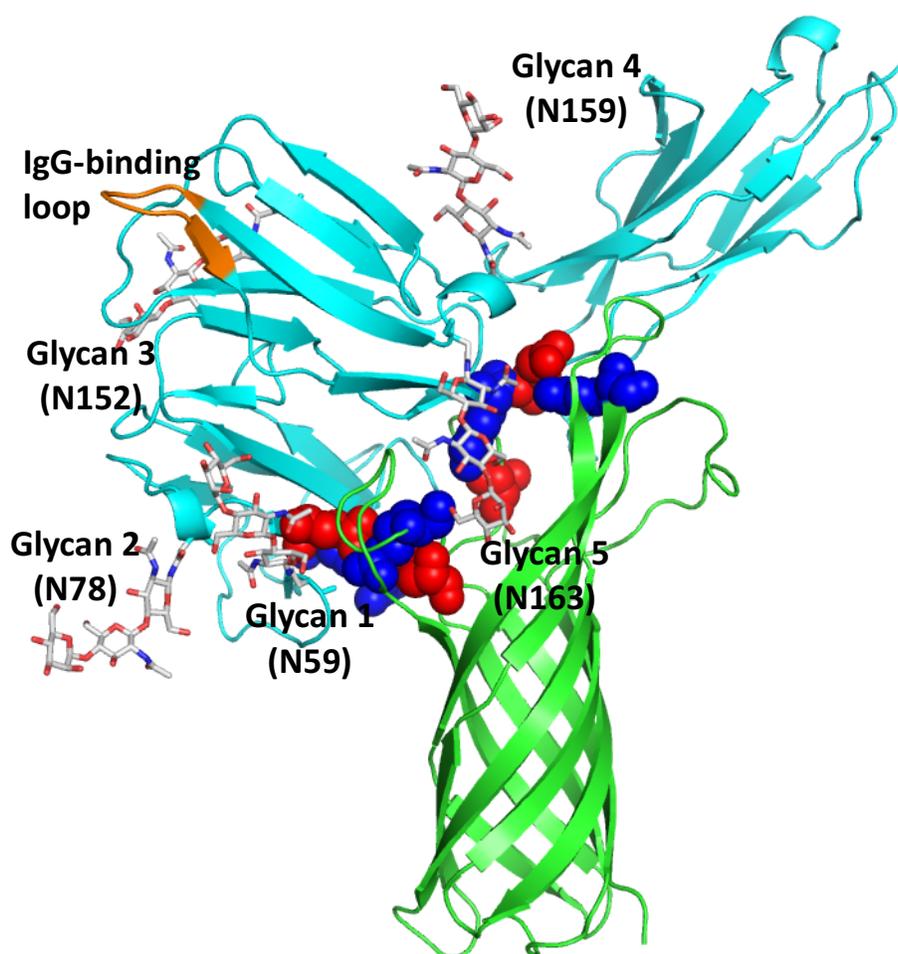


Figure 2.5: N-linked glycans 1, 4, and 5 (grey) of D1 and D2 of Fc $\gamma$ R1 are involved in binding OmpA. N-linked glycan 5 sticks into the  $\beta$ -barrel mouth of OmpA. OmpA forms salt-bridges in all corners with Fc $\gamma$ R1, stabilizing the protein-protein interface (positively charged residue in blue, negatively charged in red). The IgG-binding loop of Fc $\gamma$ R1 is colored orange and is not involved in binding OmpA.

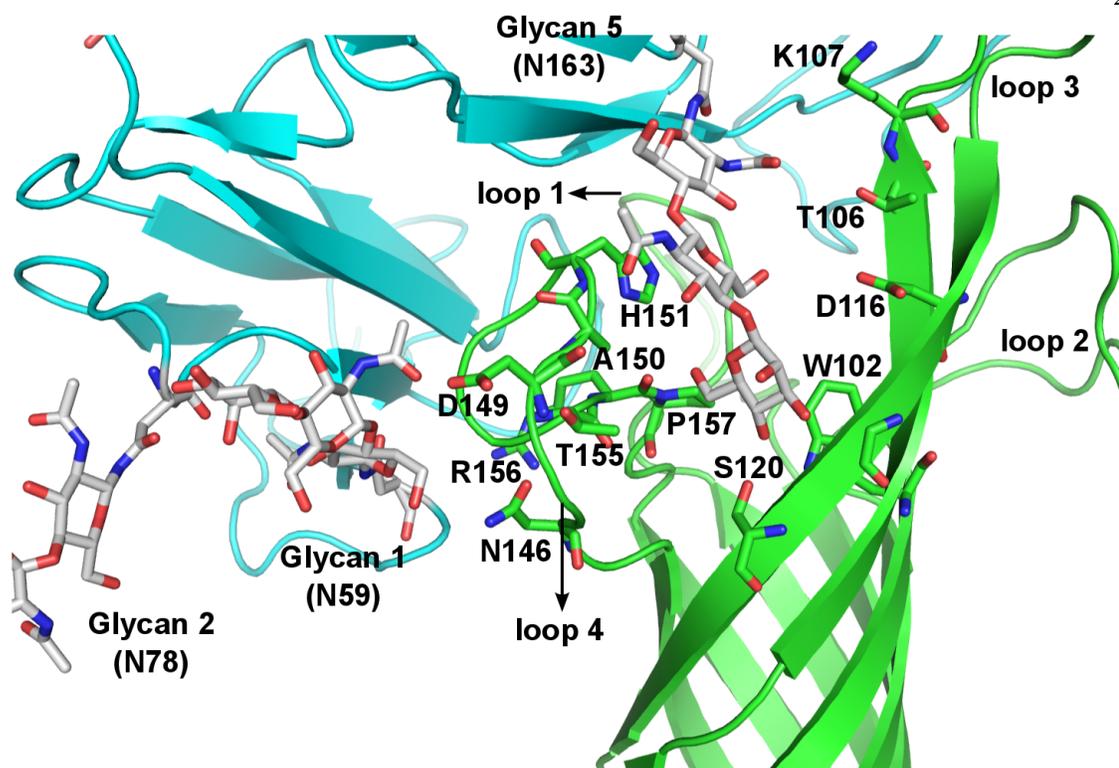


Figure 2.6: Glycans 1 forms a hydrogen bond with D149 of OmpA and is in proximity to N146 and R156 of OmpA to form water mediated hydrogen bonds. Glycan 5 is in the barrel binding pocket between loops 3 and 4 of OmpA and forms direct non-covalent interaction with W102, S120, A150, H151, T155, and P157 of OmpA. Glycan 5 could also form water-mediated interaction with D116, T106, and K107 of OmpA. Glycan 2 is too far away from OmpA and does not have any possibility to form direct or indirect interaction with OmpA.

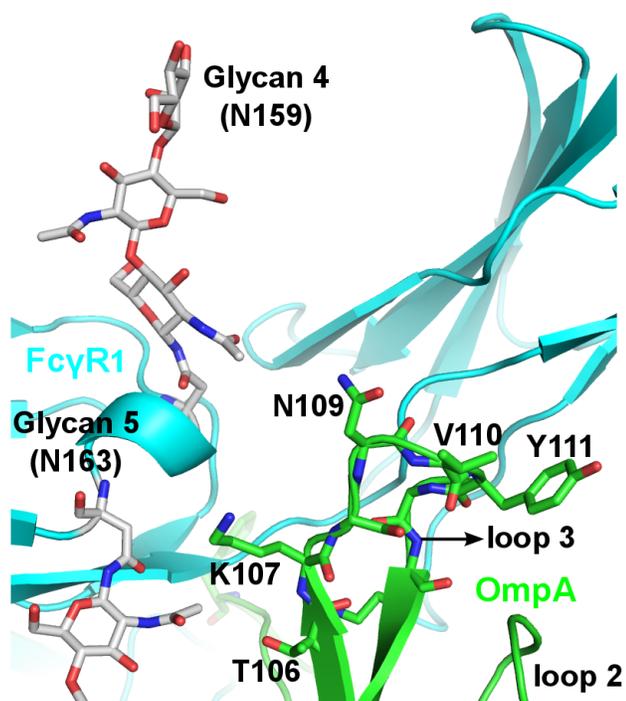


Figure 2.7: NG4 does not form any direct non-covalent interactions with OmpA, but is in proximity to OmpA to form indirect interaction like water-mediated hydrogen bonds.

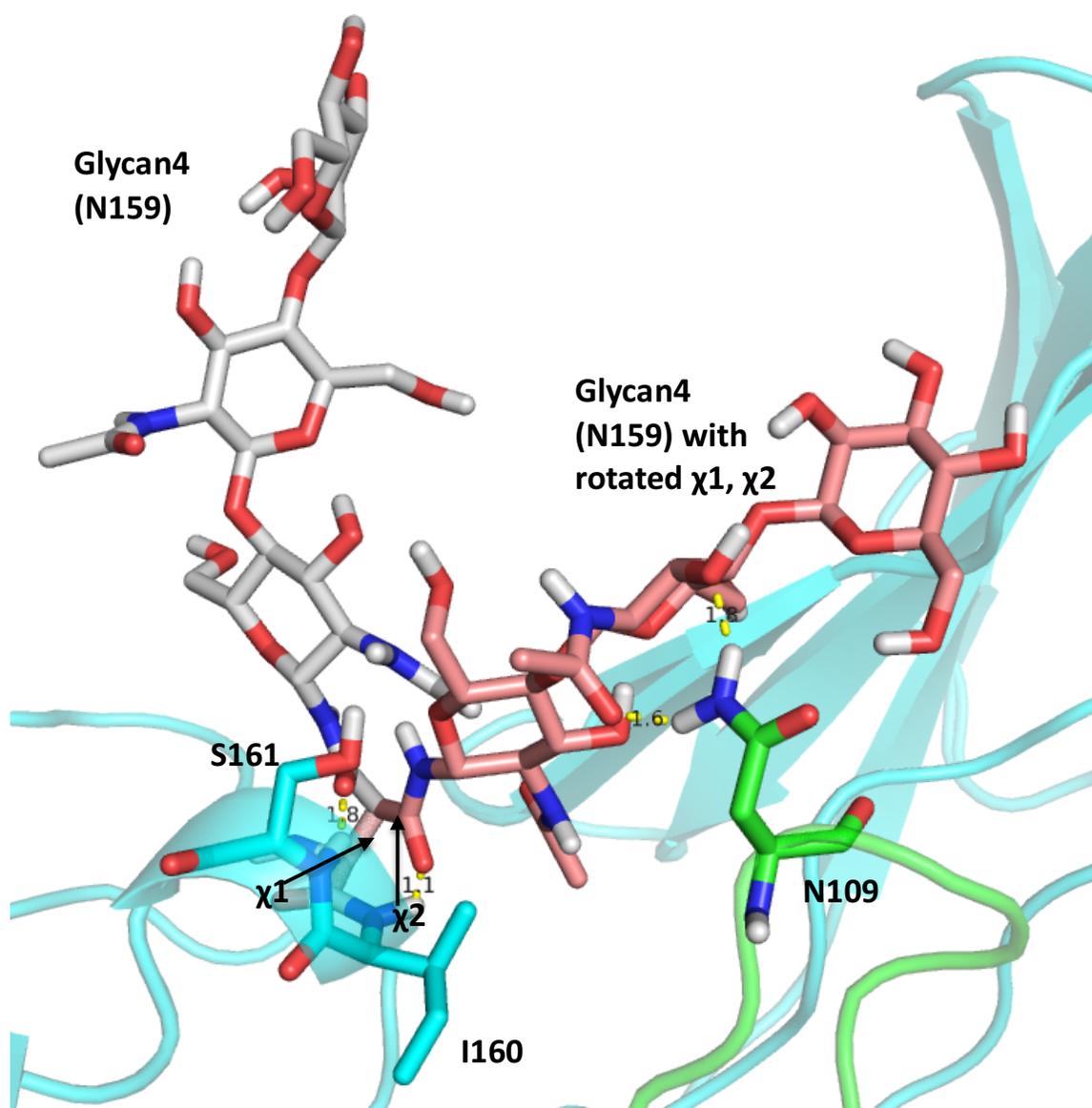


Figure 2.8: Rotating the torsional angles  $\chi_1$  and  $\chi_2$  can break the hydrogen bond between NG4 (silver conformation) and Ser-161 of Fc $\gamma$ R1 (colored cyan), replaced by a new hydrogen bond between NG4 (salmon-color conformation) and Ile-160 of Fc $\gamma$ R1, while also forming two more new hydrogen bonds that contribute to the direct interaction between NG4 and Asn-109 (highlighted yellow) on loop 3 of OmpA (colored green).

The NG5 binding site in the barrel between loop 3 and loop 4 of OmpA predicted here coincides with the barrel binding site of GlcNAc1-4GlcNAc (chitobiose) epitopes predicted in our previous work in which we docked small ligand GlcNAc1-4GlcNAc (chitobiose) to OmpA<sup>22</sup>. To test how sensitive this barrel binding site is to the sugar unit when attached to the protein, we added the common N-glycan core sugar sequence Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1 to D1 and D2 domains of Fc $\gamma$ RIa. In this predicted binding mode of the 3-sugar unit Man1-GlcNAc2 at the NG5 site, the two GlcNAc moieties maintain their interaction with OmpA as in GlcNAc $\beta$ 1-4GlcNAc binding mode (Figure 2.6). Subsequently, the D3 domain and Man $\alpha$ 1-6(Man $\alpha$ 1-3)-Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1 were also included in the minimization, which did not change the interaction of GlcNAc $\beta$ 1-4GlcNAc with the OmpA. This suggests that mannoses play a less important role than GlcNAc epitopes in N-glycan binding to OmpA, consistent with the previous experimental observations<sup>26</sup>. Taken together, the simulation studies support the experimental evidence that NG1, NG4 and NG5 interact with the loops 1, 3, and 4 of OmpA for binding, and provides a detailed 3D structure that can be used to better understand the nature of the protein-protein complex, which we expect will help us develop molecules that might block formation of this complex and hence prevent neonatal meningitis.

## 2.4 Conclusions

Recent developments in the application of dynamics and simulation have helped unravel temporal progress of bacterial-host interactions. Simulation studies have identified novel mechanisms of ‘molecular mimicry’ by pathogens that modify their virulence factors to

resemble host ligands on receptors<sup>44</sup>. Moreover, simulation of receptor N-glycosylation has strengthened our understanding of host cell manipulation by parasites such as *Trypanosoma brucei*<sup>45</sup>. However, little information is available on interaction of bacterial virulence factors with N-glycosylated host receptors. Indeed, molecular simulation studies of an N-glycosylation protein with a protein ligand had not been attempted previously. Thus, our study introduces a novel method of integrating molecular modeling and experimental evidence to predict the protein-protein interface of OmpA with glycosylated FcγRIa. This predicted structure presents an atomistic detail of the interactions that provide a mechanistic understanding of one aspect of the bacterial invasion process in neonatal meningitis. It can also be used to formulate further experimental and computational tests to obtain deeper understanding of the interactions of OmpA with the peptide regions of FcγRIa beyond the glycosylation portions. This would provide the basis for structure based design of small molecule inhibitors to prevent *E. coli* K1 interaction with FcγRIa, thereby preventing the bacterial multiplication in macrophages.

*Chapter 3*THE INTERACTION OF NATURAL SWEETENER STEVIA GLYCOSIDES WITH  
HUMAN BITTER TASTE RECEPTORS TAS2R4 AND TAS2R14**Abstract**

Steviol glycosides, the sweet principle of *Stevia Rebaudiana* (Bertoni) Bertoni, have recently been approved as a food additive in the EU. The herbal non-nutritive high-potency sweeteners perfectly meet the rising consumer demand for natural food ingredients in Europe. Experiments have revealed that two human bitter taste receptors, hTAS2R4 and hTAS2R14, mediate the bitter off-taste of steviol glycosides. We computationally predicted the structures of the transmembrane domain of these two receptors and how various agonists, including steviol glycosides, bind to the binding sites. From the structural insights and statistical analysis of the sugar number distribution on two sides of the steviol glycosides, we proposed an activation mechanism of how steviol glycosides activate these two bitter receptors. These results might contribute to the prediction of preferentially least bitter tasting *Stevia* extracts by an optimization of breeding and postharvest downstream processing.

### 3.1 Introduction

The mammalian taste sensation provides valuable information about the nature and quality of food. Taste transduction involves the interaction of tastant molecules with taste receptor-expressing cells that reside in the taste buds located on the papillae of the tongue. Taste buds relay information to the brain about the nutrient content of food. At present, there are five basic tastes, sweet, salt, sour, umami, and bitter, and among these, sweet, umami, and bitter taste sensations are sensed by G-protein-coupled receptors (GPCRs)<sup>46</sup>.

Human bitter taste receptors (T2Rs) that reside in the taste buds of the tongue are responsible for bitter taste perception, but they are also found in other parts of the body, playing important roles for nutrient sensing in the gut and in the regulation of metabolic processes, which makes them potential therapeutic targets for disease<sup>47,48</sup>. T2Rs are non-Class A members of the GPCR superfamily, with very limited structural information. Despite extensive studies, the precise mechanisms of GPCR activation are still poorly understood.

Steviol glycosides, a class of non-nutritional and natural sweetener<sup>49</sup>, have bitter aftertaste due to its agonicity towards human bitter receptors 2R4 and 2R14<sup>50</sup>. Based on molecular docking results of several agonists including rubusoside to our predicted structures of 2R4 and 2R14 transmembrane (TM) helical bundle, we proposed an activation mechanism of these two receptors that could explain the bitterness of steviol glycosides. Our activation model leads to a selection rule that could qualitatively predict the potential non-bitterness of a steviol glycoside molecule.

### 3.2 Simulation Details

We used our GEnSeMBLE method<sup>51-55</sup> to predict an ensemble of 25 low energy structures for each of 2R4 and 2R14 receptors. The structure of 2R4 is based on the homology model of the turkey Beta1 adrenergic receptor X-ray crystal structure<sup>56</sup> with 40.7% sequence similarity. The structure of 2R14 is based on the homology model of the human dopamine D3 X-ray crystal structure<sup>57</sup> with 39.6% sequence similarity. For each of the two bitter receptors, the helix shapes were optimized within their respective template, sampled all  $(12)^7 = 35$  million rotations of each of the seven helices independently, followed by sampling of 13 trillion combinations of helix rotations and tilts to select by neutral inter-helical energy the best conformation of the seven helix bundles.

We then docked various known agonists to the best conformation of the seven helix bundle for each receptor. To obtain a diverse set of ligand conformations for docking, we generated a total of 10 ligand conformations for each ligand. The molecular structures of the ligands were constructed with Maestro software or constructed from existing coordinates from the crystallized structures deposited in the Cambridge Structural Database. A conformational search was performed with MacroModel software. Systematic and extended torsional sampling options were used where the selected rotatable bonds are rotated  $360^\circ$  in  $30^\circ$  increments. Ligand conformations that fall within an energy window of 10 kcal/mol and an RMSD diversity of 0.5 Å were saved for subsequent steps. The conformational search was conducted with the OPLS 2005 force field<sup>58</sup> and in a dielectric of 80.37 to match water. Subsequently, we performed two rounds of clustering for each ligand; in the first round we clustered ligands with a 2.0 Å diversity followed by another

round of clustering with a 1.0 Å diversity. The Mulliken populations of each atom were calculated with Jaguar software<sup>59</sup> using Density Functional Theory (DFT) with the B3LYP functional and the 6-31G\*\* basis set. This led to the selection of 10 ligand conformations. Each ligand was minimized using the Surface Generalize Born (SGB) solvation model<sup>60</sup> for 100 steps or to a convergence threshold of 0.2 kcal/mol/Å RMS force with the MPSim program<sup>61</sup>. For each of the ten conformations of a ligand, we used the DarwinDock and GenDock<sup>51,62-65</sup> methods to predict the optimum binding site for each of the lowest energy protein structures predicted by GEnSeMBLE for the 2R4 and 2R14 receptors.

The DarwinDock method aims at generating a complete set of poses for the binding pocket while using RMSD clustering of the poses to dramatically reduce the computational cost. To provide flexibility and space for the ligand to identify favorable binding sites, we replaced the seven bulky hydrophobic residues (FILMYVW) with alanines. The mutated residues are called alanized residues, and the mutated protein is called alanized protein. For each of the best 100 ligand poses in the alanized protein, we then dealanized the mutated residues back to their original hydrophobic identity and optimized their positions along with those of other residues in the binding site using SCREAM<sup>66</sup>. This leads to a unique set of optimized residue side chains for each of the 100 ligand poses.

In the pose generation step of DarwinDock, a ligand pose is acceptable if it clashed or bumped the receptor residues at six positions or less. First, we used Dock6<sup>67</sup> to generate

5,000 ligand poses (without evaluating an energy) and clustered them into families, where every family member is within a 2.00 Å RMSD of each other. Then, we added 5,000 more ligand poses from Dock6 and reclustered. This procedure of adding 5,000 poses and reclustering is repeated until the number of new families generated is less than 2% of the total number of families in the preceding iteration. Typically, 45,000 poses were generated leading to 6000–9000 2.0 Å families. At this point DarwinDock scores the energies of one representative from each family, the family head, and selects the 10% of family heads with the lowest Dreiding energies. All members of these respective families are then scored energetically. From this list of approximately 5000 poses, we select the lowest 50 by each of three criteria: lowest hydrophobic energy, lowest polar energy, and lowest total energy, giving at most 150 poses.

GenDock was used to refine the 150 docked ligand–receptor poses generated by DarwinDock. In the SCREAM step, “alanized” residues were replaced with the original hydrophobic residues but using the optimum side chain rotamers to avoid any clashes with the ligand and other protein side chains. Then the entire complex was minimized for 10 steps for each case to remove any bad contacts. Next, the receptor was neutralized, so that the acidic residues (aspartic acid and glutamic acid) each gained a proton, and the basic residues (lysine and arginine) each lost a proton. The resulting receptor–ligand complexes were minimized for 60 steps using the Dreiding III FF. Then for each of the ten ligand conformations, we selected the lowest binding energy structure, including strain and ligand solvation, which we expected to best represent the binding affinity. The binding energy with strain and ligand solvation is defined as the energy difference

between the complex and the sum of the receptor and ligand energies with ligand strain and ligand solvation included. The selected complex was minimized with the Dreiding III FF with the LJ vdW term (Dreiding III-LJ FF) in vacuum for 50 steps or to an RMS force threshold of 0.5 kcal/mol/Å using the MPSim program.

### **3.3 Results and Discussion**

Our prediction of the 2R4 and 2R14 transmembrane (TM) bundle structures suggested that the inactive protein conformations are characterized by interactions between the intracellular end of TM helices 3 and 6 (IC-TM 3-6 coupling), and IC-TM 2-6 coupling, as highlighted in Figures 3.1 and 3.2. The IC-TM 3-6 couplings through hydrogen bonding in both 2R4 and 2R14 are similar to the intracellular 3-6 “ionic lock” observed as an characteristic salt-bridge in Class A GPCRs<sup>68,69</sup>. The IC-TM 2-6 couplings are similar to the predicted and experimentally validated intracellular 2-6 salt-bridge in CB1 cannabinoid receptor, which also stabilizes the inactive state of the receptor<sup>70</sup>.

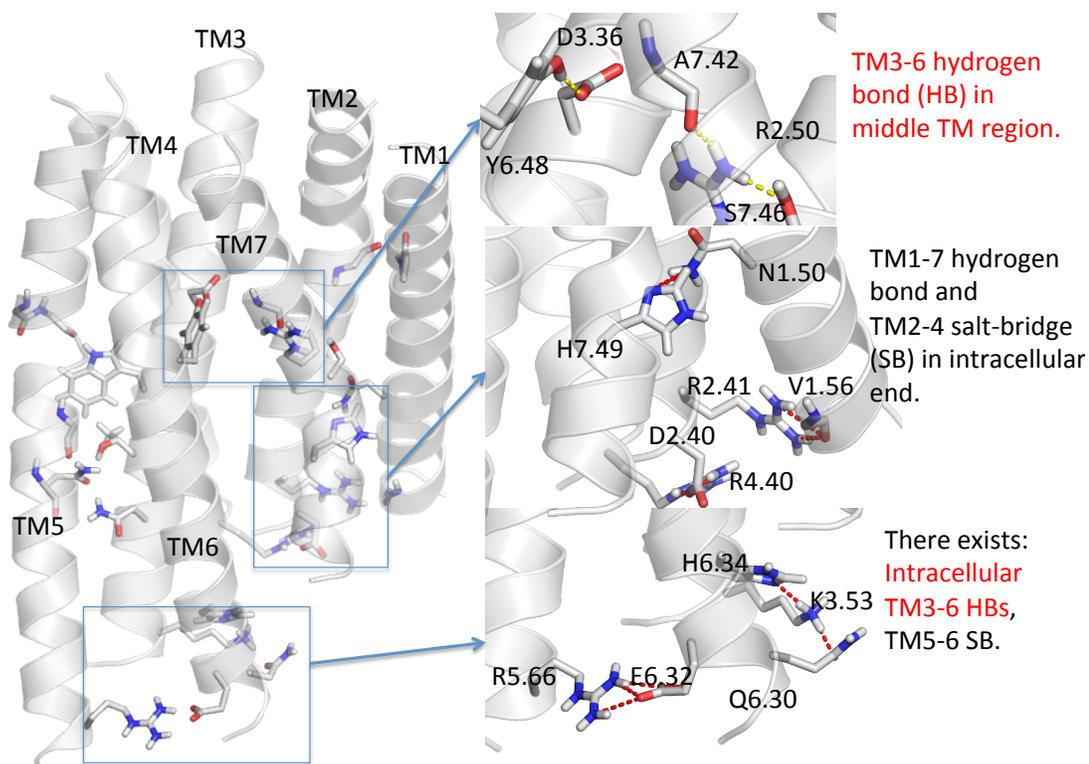


Figure 3.1: Predicted TM bundle structure of TAS2R4. TM3-6 coupling interactions in both binding site and the intracellular end are highlighted in red.

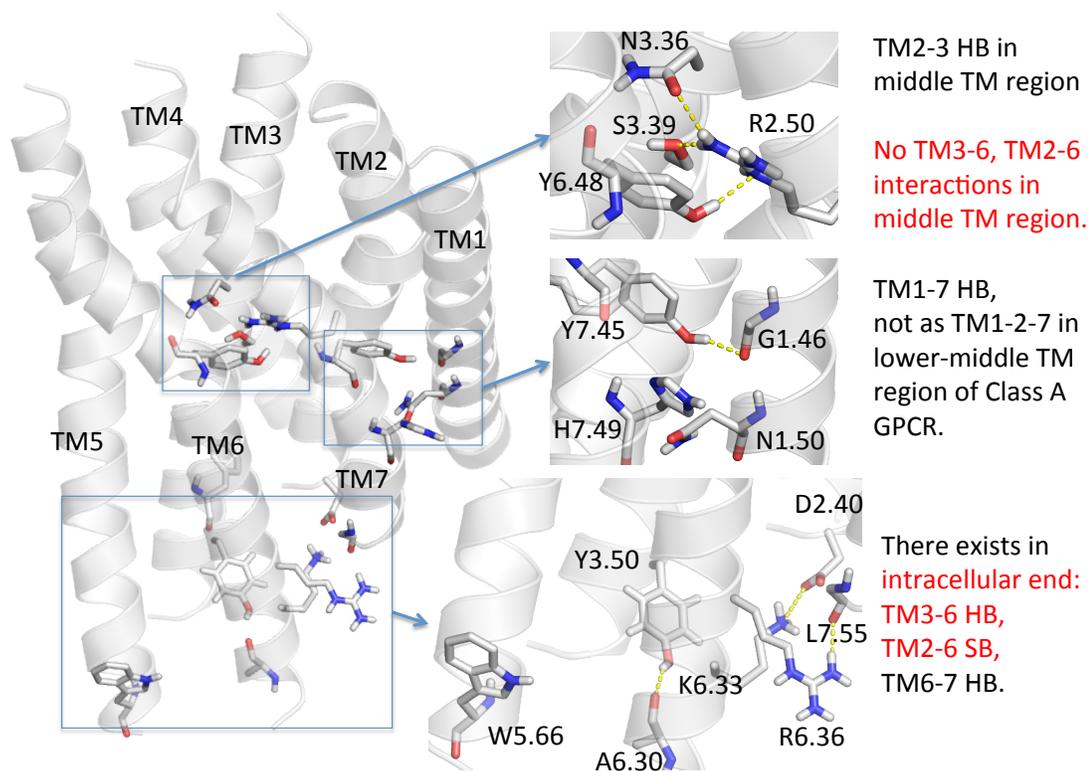


Figure 3.2: Predicted TM bundle structure of TAS2R14. TM3-6 and TM2-6 coupling interactions in the intracellular end are highlighted in red.

By docking various known agonists to the TM binding sites of the two receptors, we could validate the accuracy of the predicted binding sites. The results showed that agonists binding changes the conformations of several key residues in the binding site, yielding strong TM 3-6 and 2-6 couplings in the mid-TM binding site as shown in Figures 3.3 and 3.4, which could potentially break the intracellular TM 3-6 and 2-6 couplings and thus lead to activation. This mechanical view of coupled helical motions serves as our working hypothesis for activation mechanism of these two bitter receptors.

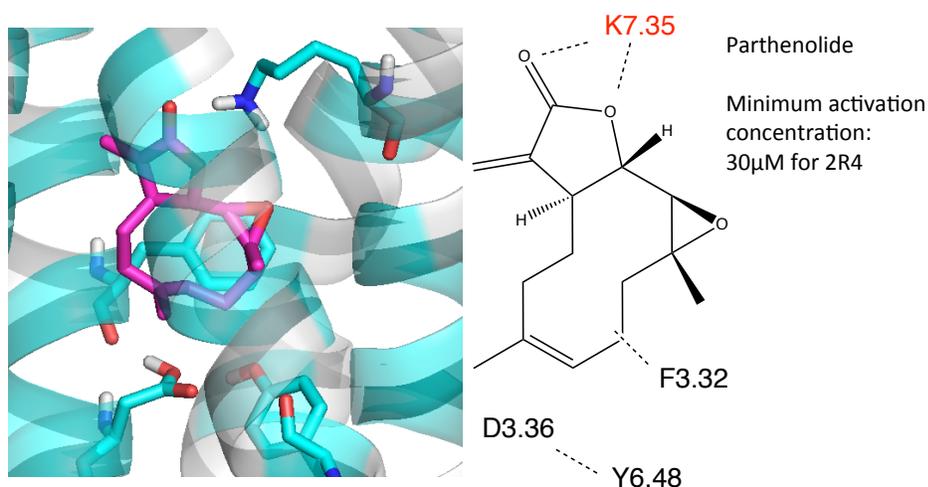


Figure 3.3: Predicted structure of 2R4 bound with its agonist parthenolide. Binding of an agonist strengthens TM3-6 coupling in the binding site of 2R4.

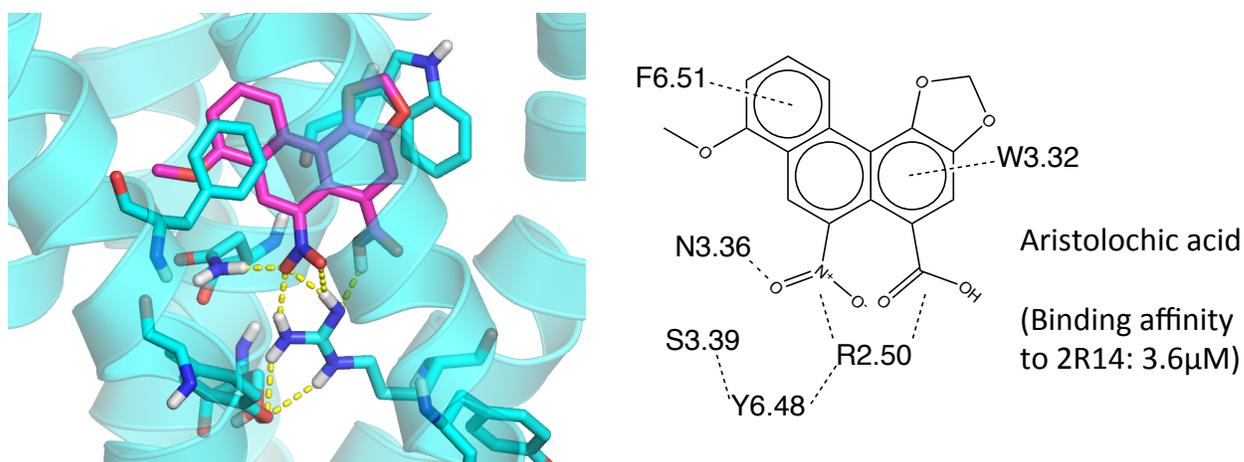


Figure 3.4: Predicted structure of 2R14 bound with its agonist aristolochic acid. Binding of an agonist breaks TM2-3 coupling and forms TM3-6 and TM2-6

Our prediction of the binding interactions of rubusoside with 2R4 and 2R14 provided a key insight on how the whole class of steviol glycosides binds to 2R4 and 2R14. As shown in Figure 3.5 (left), a steviol glycoside molecule comprises three parts: a stevia backbone, an  $-OH$  linked sugar moiety  $R_B$ , and a  $-COOH$  linked sugar moiety  $R_A$ . The

representation of the molecule is further simplified in Figure 3.5 (right). Rubusoside binds to 2R4 and 2R14 with the A side (one glucose) inserted into the TM binding site facing towards the intracellular end, while the B side (one glucose) is facing the extracellular end, as shown in Figures 3.6 and 3.7. The A side directly interacts with the key residues at the bottom of the binding site and mediates TM 3-6 and 2-6 couplings within these residues.

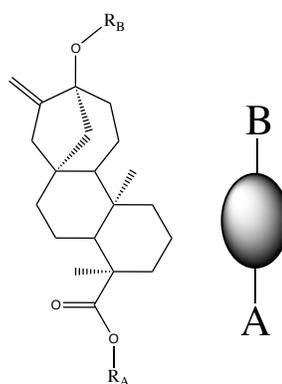


Figure 3.5: The general framework of steviol glycosides (left) and its simplified representation (right).

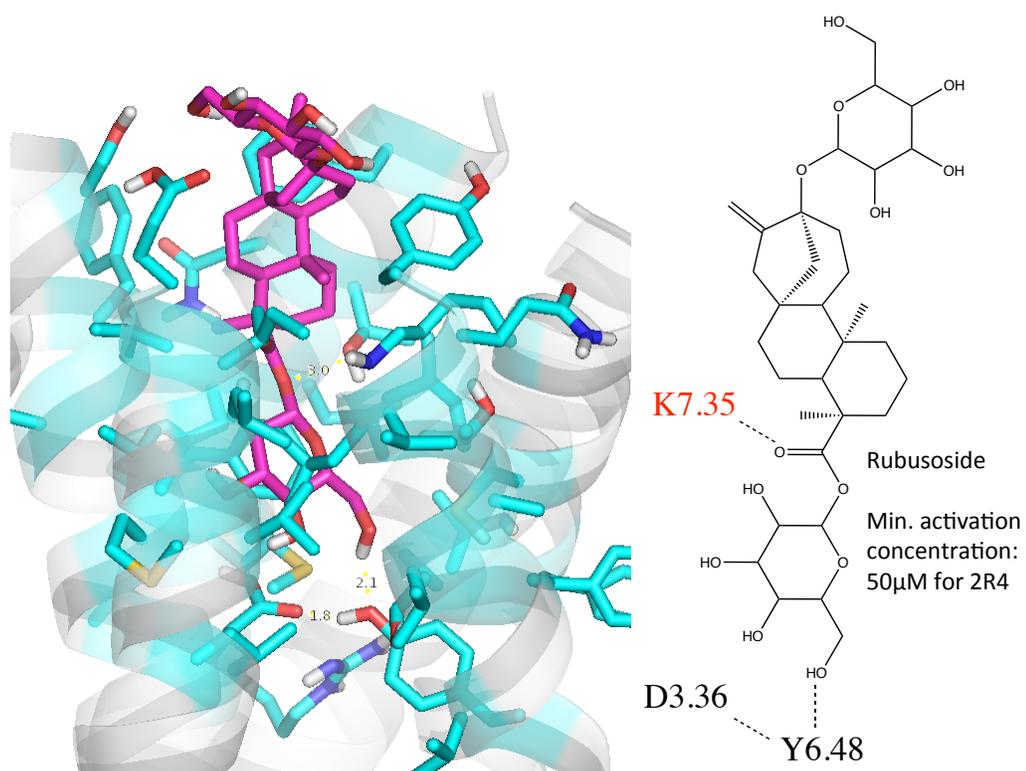


Figure 3.6: Predicted structure of 2R4 bound with rubusoside. Binding of rubusoside strengthens TM3-6 coupling in the binding site of 2R4.

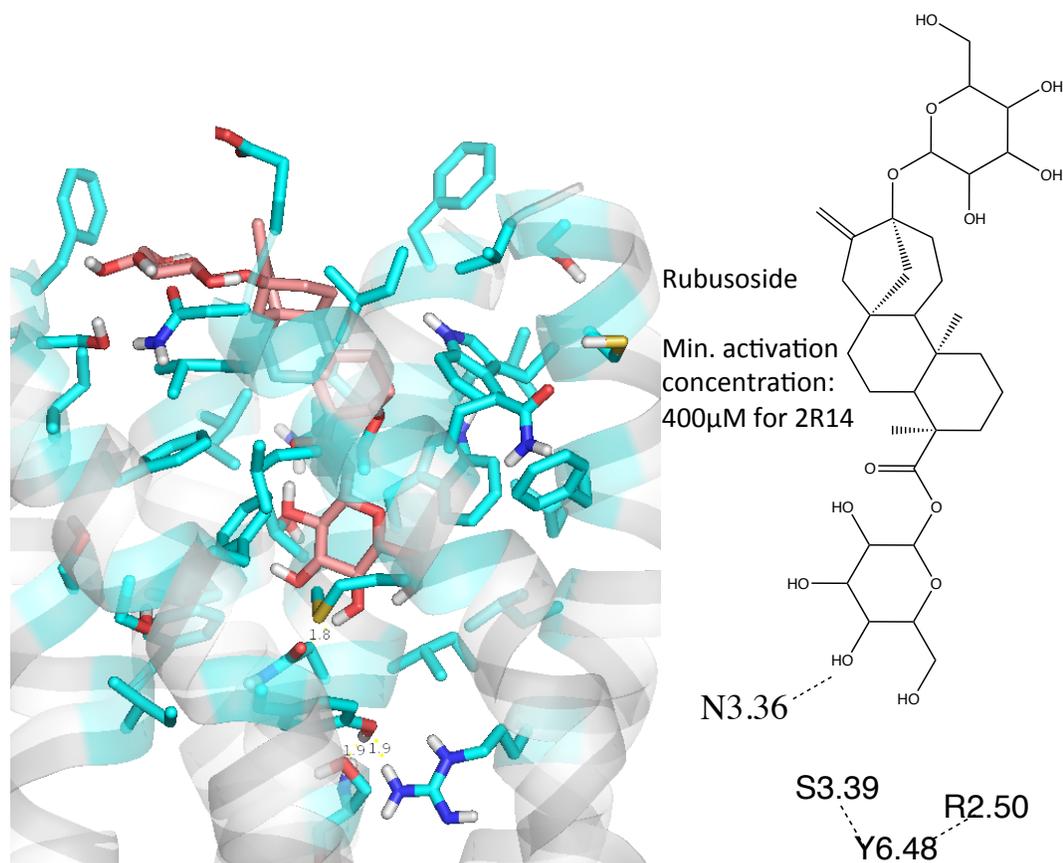


Figure 3.7: Predicted structure of 2R14 bound with rubusoside. Binding of rubusoside breaks TM2-3 coupling and forms TM3-6 and TM2-6 coupling in the binding site of 2R14.

We observed that the 11 steviol glycoside molecules reported in reference 71 all have group A (0-2 monosaccharides) smaller than or equal to their group B (1-3 monosaccharides). Due to the spatial constraint of the funnel-like shape of the TM binding site, we proposed that this A-down-B-up binding model is general for the whole class of steviol glycosides.

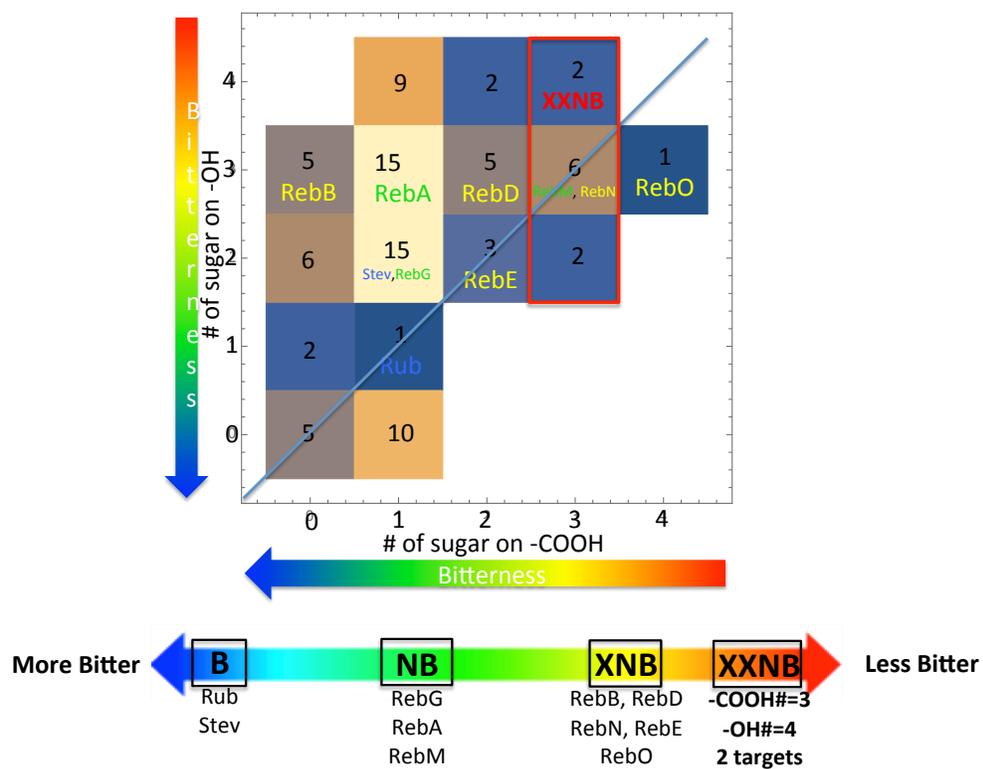


Figure 3.8: Histogram of (COOH sugar #, OH sugar #) distribution of 94 molecules in Cargill priority list 1. Molecules in the red box except RebM and RebN are predicted to be low bitter targets to be tested experimentally.

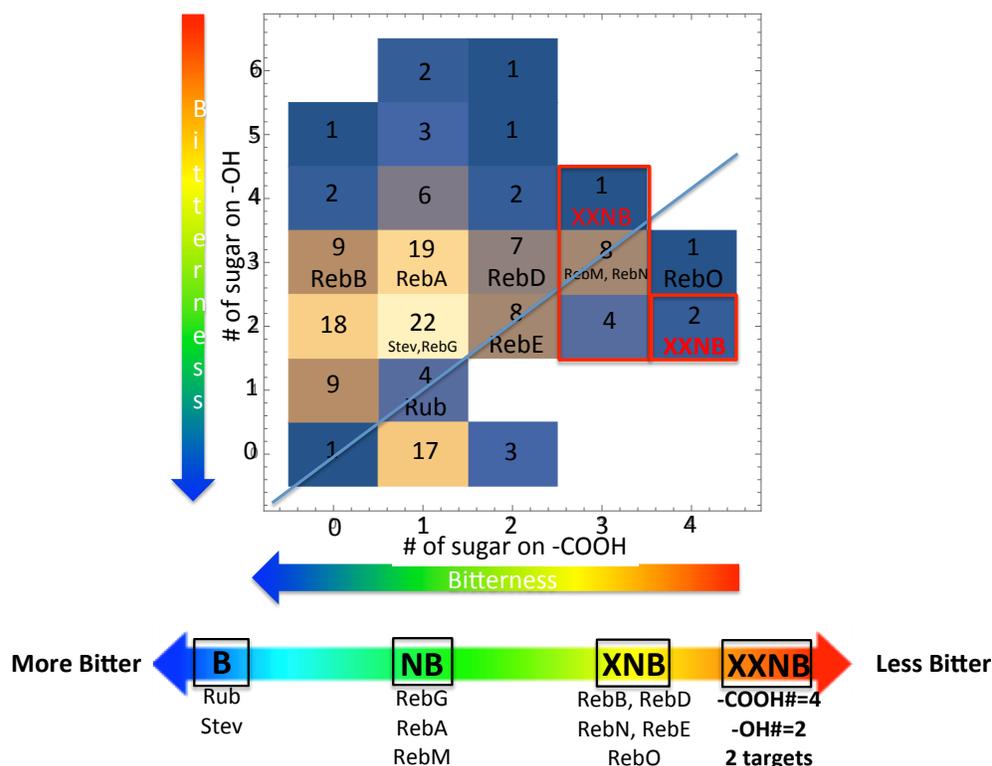


Figure 3.9: Histogram of (COOH sugar #, OH sugar #) distribution of 151 molecules in Cargill priority list 2. Molecules in the red box (not including RebM and RebN) are predicted to be low bitter targets to be tested experimentally.

Given two lists of stevia glycosides from the Cargill company, we did further analysis on a total number of 94 entries in priority list 1 and 151 entries in priority list 2. We first did a statistical analysis on how the stevia glycosides in each list distribute in a two-dimensional histogram where the x-axis is the number of sugar groups (monosaccharides) on the -COOH side and the y-axis is the number of sugar groups on the -OH side, as shown in Figure 3.8 for priority list 1 and Figure 3.9 for Priority list 2. The number in a box of coordinate (x, y) is the number of entries in a list with x sugars in group A and y sugars in group B (Figure 3.5). As mentioned, stevia glycosides tend to have more sugar groups on the -COOH side than on the -OH side with a few exceptions. By placing the

10 stevia glycosides with experimental bitterness measurement from Cargill onto the histogram and color-coding them by their relative bitterness, we found that the more sugar groups on either the  $-\text{COOH}$  side or the  $-\text{OH}$  side, the less bitter a stevia glycoside is. This statistical trend suggested that the 23 molecules in the red box in the two histograms would be on the less bitter end of the bitterness spectrum.

This statistical observation could be explained by the structural insight we obtained from modeling. The glycan on the  $-\text{COOH}$  side with more sugar groups is difficult to fit into the binding pocket and form strong interactions with the two residues that couple TM3-6 in the binding site. The glycan on the  $-\text{OH}$  side with more sugar groups will have stronger interactions with the flexible extracellular loop region, which prevents the whole ligand from reaching deeper into the bottom of the binding site where the two residues couple TM3-6. These two interpretations lead to two hypothetical rules to select the best candidates with the least bitterness. Rule #1 says that a large group B on the  $-\text{OH}$  side would lead to less bitterness, as explained by Figure 3.10. According this rule, 2 molecules in the Priority 1 list and 1 molecule in the Priority 2 list with coordinate  $(3_{\text{COOH}}, 4_{\text{OH}})$  will be among the best targets for low bitterness. Rule #2 says that a large group A on the  $-\text{COOH}$  side would lead to less bitterness, as illustrated in Figure 3.11. Thus the 2 molecules in the Priority 2 list with coordinate  $(4_{\text{COOH}}, 2_{\text{OH}})$  will be among the best targets.

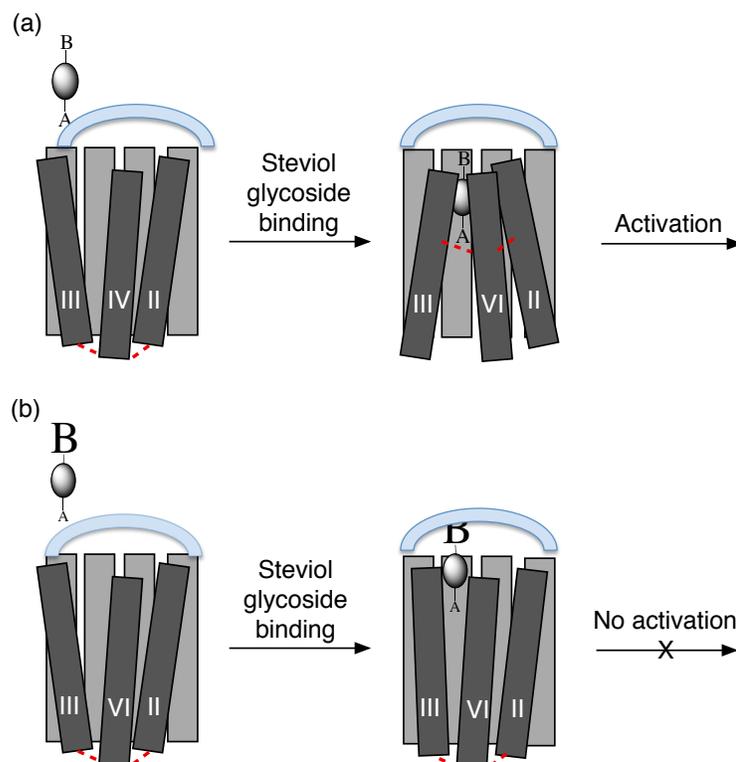


Figure 3.10. Rule #1 for selecting low bitter targets. If  $-OH$  sugar group (group B) is large, it will be held by the loop region, making the  $-COOH$  sugar group difficult to form stable interaction with the two residues that couple TM3-6 in the binding site.

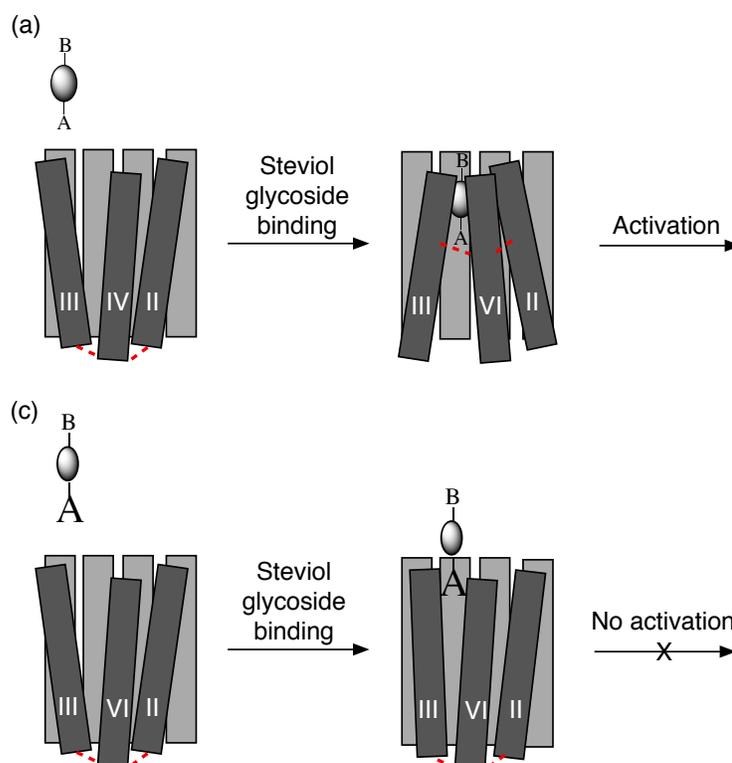


Figure 3.11. Rule #2 for selecting low bitter targets. It is difficult for large  $\text{-COOH}$  sugar group (group A) to fit into the binding site and form stable interaction with the two residues that couple TM3-6 in the binding site.

### 3.4 Conclusions

We predicted the structures of 2R4 and 2R14 and how various ligands including stevia glycosides with diversity in structure and bitterness bind to these two receptors. We also distilled the key insight on activation mechanism from these structures and came up with qualitative criteria which could potentially predict the relative bitterness of experimentally untested stevia glycosides. Given two lists of stevia glycosides with a total number of 235 entries from the Cargill company, we predicted that 23 out of 235 molecules are of low bitterness and 5 out of 23 are of extra low bitterness.



*Chapter 4*

## REACTION MECHANISMS OF METHANE OXIDATION BY THE BIOMIMETIC TRICOPPER CLUSTER COMPLEXES

**Abstract**

We report QM studies of the reaction mechanism for biomimetic tricopper complexes to convert methane to methanol after activation of O<sub>2</sub>. Based on analyses of particulate methane monooxygenase (pMMO), Chan and coworkers developed biomimetic tricopper complexes that carry out efficient oxidation of methane to methanol after activation by dioxygen. We used density functional theory (DFT) and wavefunction based methods (CASSCF) to determine the reaction mechanisms of these tricopper complexes for the entire thermodynamic cycle, which consists of (a) C-H bond activation, (b) sacrificial reduction of H<sub>2</sub>O<sub>2</sub>, and (c) regeneration of the active tricopper cluster by O<sub>2</sub> or H<sub>2</sub>O<sub>2</sub>. Our results reveal that the facile methane oxidation is mediated by transformations between various structural and spin states of the tricopper cluster. Analyses of the electronic structures and energetics of the spin-state surfaces along the reaction pathway provide detailed reaction mechanisms for all three steps.

## 4.1 Introduction

The activation and functionalization of an inert C-H bond of alkane molecules is a great challenge for both scientific and industrial communities, because alkanes are the main constituents of natural gas and an efficient way of such transformation of alkanes to liquid fuels like alcohols could have a direct economic impact<sup>71-73</sup>. For the strong C-H bond (~100 kcal/mol) in methane, the current industrial chemical processes require extreme conditions which are costly and inefficient, due to the fact that the inertness of the strong C-H bond requires significant energy for both the homo- and hetero-lytic C-H bond cleavage. Oxygenation of the C-H bonds in alkanes, especially methane oxidation to produce methanol, still remains as a challenging problem<sup>74</sup>.

However, methanotrophic bacteria accomplish this feat readily under ambient conditions using metalloenzymes called methane monooxygenases (MMOs)<sup>75</sup>. The MMOs catalyze oxidation of methane to methanol utilizing O<sub>2</sub> and two reducing equivalents from NAD(P)H or quinols to split the O-O bond. One oxygen atom of the O<sub>2</sub> is incorporated into the hydrocarbon substrate to yield methanol and the other oxygen atom is reduced to water with the uptake of two protons. Two types of MMO enzymes are found in methanotrophic bacteria: the soluble form (sMMO) present in the form of a cytoplasmic complex<sup>76,77</sup>, and particulate MMO (pMMO) which is in membrane-associated form<sup>78</sup>.

While it is now well established that in sMMO the activation of methane takes place in a diiron active site with possible mechanisms figured out from spectroscopic measurements and computational studies<sup>77</sup>, little is known about the methane oxidation process catalyzed

by pMMO partially due to a long-standing lack of detailed information regarding its structure. The first x-ray crystal structure of pMMO was isolated and purified from *Methylococcus capsulatus* (Bath), and reported by the laboratory of Amy Rosenzweig in 2005<sup>79</sup>. According to this work, the protein crystallizes as a trimer of  $\alpha\beta\gamma$  monomers with three copper ions and one zinc ion per monomer. A ribbon diagram of the structure is shown in Figure 4.1, with different colors used to depict the three subunits. PmoA and PmoC are mostly transmembrane, each with approximately six transmembrane segments. The N- and C-terminal subdomains of PmoB are exposed to the cytosol and are anchored at the water-membrane interface by two  $\alpha$ -helical 10 transmembrane segments inserted into the membrane. The three copper ions and the zinc ion reported in the crystal structure are also shown in Figure 4.1, with the three copper ions at site A and site B in blue and the zinc ion at site C in brown. Highlighted in the transmembrane domain of the crystal structure at site D is a cavity consisting of a hydrophilic cluster of potential metal-ligating residues, including His38, Met42, Met45, Asp47, Trp48, Asp49, and Glu100 from PmoA and Glu154 from PmoC. This “cluster of hydrophilic residues” has been discounted as a metal-binding site. However, the electrostatic energy of sustaining this cavity would be extremely high without metal counter-ions to balance the negative charges of the hydrophilic residues. The metal ions must have been stripped away from site D during the purification of the protein for crystallographic analysis<sup>80</sup>.

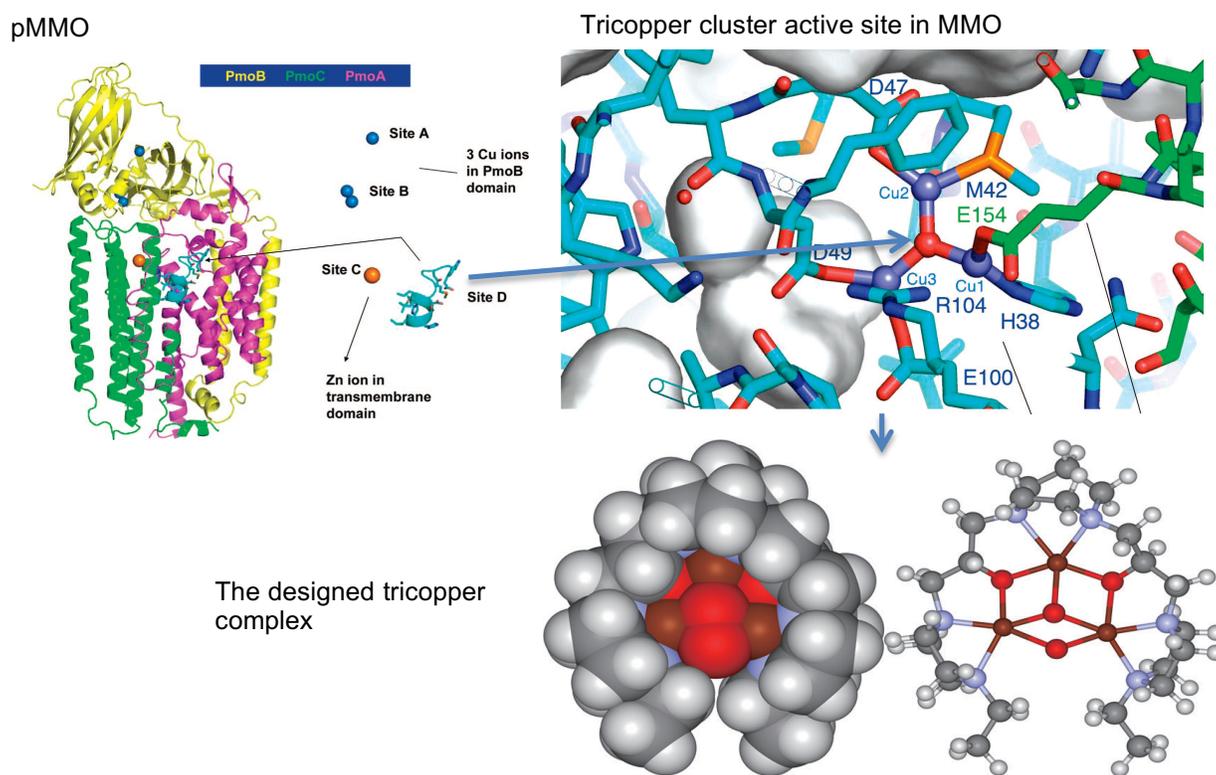


Figure 4.1: Left: the X-ray crystal structure of pMMO isolated and purified from *Methylococcus capsulatus* (Bath). Top right: tricopper cluster modeled as Cu(II)Cu(II)Cu(II) with capping “oxo” at site D of the crystal structure. Top bottom: the designed ligand environment of a given tricopper complex.

Assuming that the overall fold of the protein structure has not been dramatically compromised by the loss of the copper ions, attempts have been made to rebuild the coppers back into the protein scaffold<sup>81</sup>. This effort has led to a tricopper cluster at site D (Figure 4.1). The ligands to the copper atoms in the model are as follows: PmoC Glu154 and PmoA His38 for Cu1; PmoA Met42 and Asp47 for Cu2; and PmoA Asp49 and Glu100 for Cu3. In building this model, the Cu(II)Cu(II)Cu(II) cluster is assumed to be capped with a  $\mu$ -oxo ( $O^{2-}$ ) to ensure electrical neutrality at the site. The coordinated

ligands and the geometry of the cluster, including the Cu–Cu and Cu–O distances, are all reasonable demonstrating the feasibility of pMMO to accommodate a tricopper cluster. The trinuclear CuIICuIICuII structure modeled here would correspond to that of the fully oxidized cluster after turnover by dioxygen in the absence of hydrocarbon substrate, the so-called “dead-end” species<sup>82</sup>.

Inspired by the proposal that the catalytic site might be a tricopper cluster, a series of tricopper complexes have been developed and have been shown to be capable of supporting facile catalytic oxidation of hydrocarbons<sup>83–85</sup>. The oxidation of CH<sub>4</sub> mediated by the tricopper complex is summarized in Figure 4.2. A single turnover is obtained when this Cu(I)Cu(I)Cu(I) complex **3** is activated by excess dioxygen in the presence of excess CH<sub>4</sub> (Figure 4.2). The reaction is complete within ten minutes, clearly indicating that the oxidation is very rapid. In accordance with the single turnover, the kinetics of the overall process is pseudo first-order with respect to the concentration of the fully reduced tricopper complex. The process can be rendered catalytic by adding the appropriate amounts of H<sub>2</sub>O<sub>2</sub> to regenerate the “spent” catalyst **3** after O-atom transfer from the activated tricopper complex **1a** to CH<sub>4</sub> and form an intermediate state **2**.

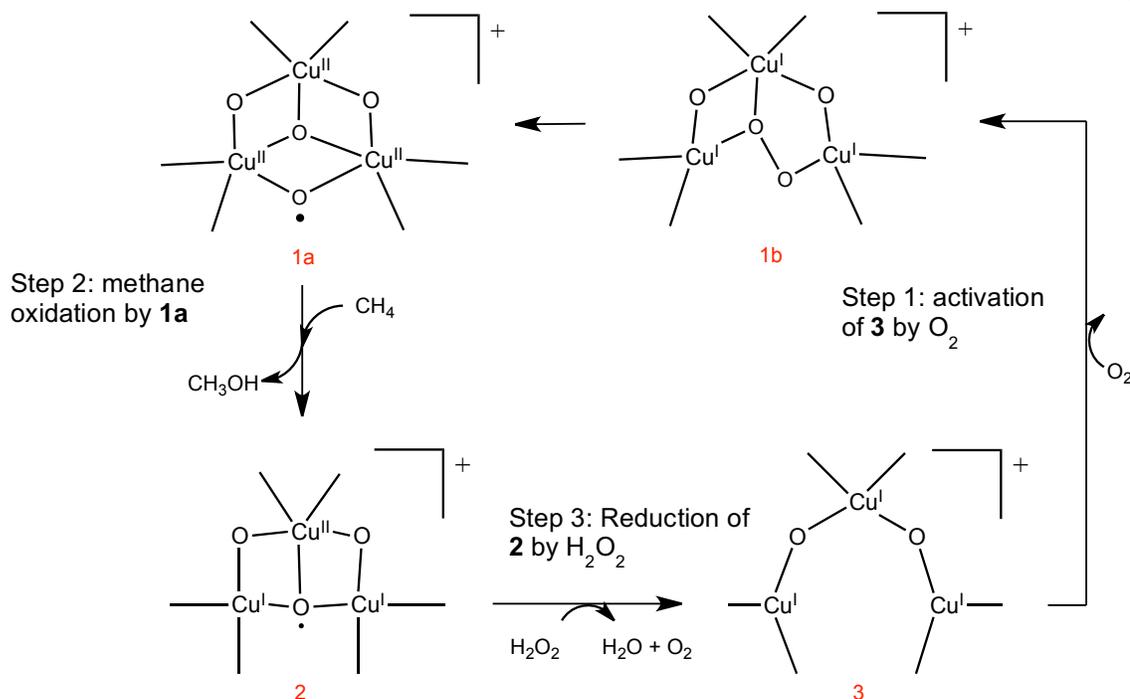


Figure 4.2: Productive cycling in the oxidation of methane by  $O_2$ , mediated by the  $Cu(I)Cu(I)Cu(I)$  complex **3** in the presence of  $H_2O_2$  as the sacrificial reductant.

In this work, we investigated the reaction mechanisms the oxygen activation, the methane hydroxylation and the sacrificing reduction of  $H_2O_2$  catalyzed by the tricopper complex by quantum chemistry simulations. Density functional theory with B3LYP functionals was used to calculate the structures and energies of all the intermediates and transition states of the relevant mechanism. The reactivity of different spin states is discussed and the wave-function based multireference method, namely Complete Active Space Self Consistent Field (CASSCF) is used to investigate the nature of chemical bond of the active state of the catalyst.

## 4.2 Computation Details

The DFT calculation of the tricopper complex is performed using the Jaguar ab initio electronic structure program<sup>59</sup>, which is a commercial software currently produced and maintained by Schrödinger Inc. Keywords in the Jaguar input file is included in the parenthesis follow the description of the detailed method for readers who are interested to reproduce the results. B3LYP functional<sup>16,17</sup> is used combining D3 dispersion correction<sup>86</sup> (dftname=B3LYP-D3). For geometry optimization, a medium and polarized basis set LACVP\*\*<sup>87</sup> is used (basis=LACVP\*\*). For single point energies of the optimized geometries, a large basis set cc-pvtz(-f)<sup>88-91</sup> is used (basis=cc-pvtz(-f)). Both geometry optimization and single point energy calculations are done with Poisson-Boltzmann solvation model (isolv=2, nogas=2) for acetonitrile, which uses a surrounding dielectric continuum medium of dielectric constant of 37.5 (epsout=37.5) and probe radius of 2.18 (radprb=2.18). The complexes were optimized in the gas phase, and the optimized structures were subsequently subjected to a single-point energy calculation using the PBF solvation model. Frequencies and thermodynamic contribution to the free energy were calculated for the optimized geometries in the gas phase at 373.15 K. The reported free energies were calculated by adding the thermodynamic correction, including zero-point energy, and the dispersion correction to the single-point solution phase energy. The presented transition states (TSs) were verified to have only one imaginary frequency and were verified to be connected to the presented succeeding and preceding energy minima, respectively. All stationary minima were verified to have no imaginary frequencies. The Complete Active Space Self Consistent Field (CASSCF) calculation is performed using the ORCA software<sup>92</sup> with the same basis set for DFT

single point energy.

The accuracy of DFT calculations on the multinuclear copper oxygen clusters have been evaluated by several previous computational studies<sup>93–95</sup>. Recently, Liakos and Neese used highly ab initio correlated methods, including relativistic and solvation effects, to show that the bis( $\mu$ -oxo)Cu(III)<sub>2</sub> is in fact more stable than ( $\mu$ - $\eta^2$ : $\eta^2$ -peroxo)Cu(II)<sub>2</sub>, in agreement with the experimental results<sup>96</sup>. Furthermore, Liakos and Neese also investigated the performance of different exchange–correlation functionals to describe the relative energies involved in the peroxo–bis-oxo isomerization and found that the B3LYP-D functional presented the best agreement with the LPNO-CCSD results. The mean absolute error computed for the deviation of the B3LYP results in relation to LPNO-CCSD was 4.4 kcal/mol. The summary of these results shows that the use of B3LYP functional including dispersion effects provides relative energies in reasonable agreement with correlated ab initio methods. Application of highly correlated ab initio methods for studies involving bioinorganic systems still remains challenging, mainly with respect to the computational demand required. Therefore, DFT methods remain as the main choice for investigations of biological systems of moderate size. The results reported by Liakos and Neese support the B3LYP functional including dispersion effects as a good computational method for studies involving trinuclear copper complexes like the one investigated here.

## **4.3 Results and Discussions**

### **4.3.1 The oxygen activation**

The structure model of the most reduced Cu(I)Cu(I)Cu(I) singlet state **3** in the reaction

cycle was drawn from a derivative of the interested ligand and its geometry was minimized as the starting point of investigating the species involved in the reaction cycle (Figure 4.3). The central Cu(I) has a coordination number of 4 with a tetrahedral configuration. The bottom two Cu(I)s have a coordination number of 3 and could bind the solvent molecule acetonitrile to saturate their coordination. Upon adding a triplet dioxygen molecule 3 Å away from the the bottom two coppers in such non-bonding distance from the complex **3**, further geometry optimization of ~140 minimization steps yielded a stable intermediate state **1b**, where the O-O bond decreased from 1.21 Å to 1.33 Å and the distance between the bottom two coppers decreased from 5.41 Å to 5.14 Å. The change of these bond lengths and the analysis of the molecular orbitals showed that one electron transferred from the Cu(I)Cu(I)Cu(I) cluster to the dioxygen, partially breaking the O-O bond and leaving only one of the two antibonding pi orbitals of the dioxygen that is perpendicular two the plane where the three coppers reside singly occupied. Spin density analysis showed that the other unpaired spin mainly resides on the central copper, making it a 5-coordinated Cu(II), while the bottom two coppers are 4-coordianted Cu(I)s. The electronic binding affinity of O<sub>2</sub> in the nonbonding distance to the state **3** in the initial van der Walls complex is -7.7 kcal/mol. The electronic energy decrease from the van der Waals complex to the state **1b** is -24.9 kcal/mol, which could pay for the 15 kcal/mol entropy penalty of stabilizing molecular oxygen in gas phase (Figure 4.3).

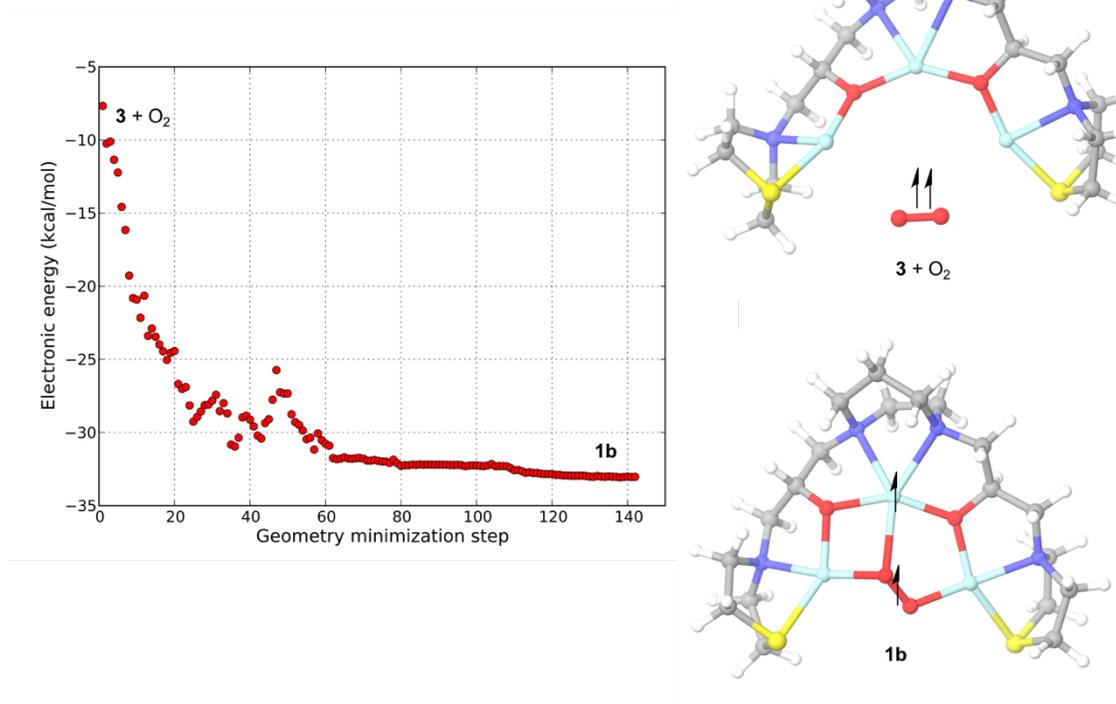


Figure 4.3: Dioxygen activation of the Cu(I)Cu(I)Cu(I) complex **3** to form the Cu(I)Cu(I)Cu(II)(superoxo) species **1b**.

Further decrease of the distance between the bottom two coppers from **1b** will fully break the O-O bond and form a stable state **1a** through a barrier of 12.9 kcal/mol (electronic energy) on the triplet surface (Figure 4.4). At the geometry of state **1b**, the quintet energy is 84.3 kcal/mol (electronic energy) higher than the triplet energy. At the geometry of state **1a**, the quintet energy is 1.2 kcal/mol (electronic energy) lower than the triplet energy. As the distance between the bottom coppers decreases from **1b** to **1a**, the quintet-triplet gap decreases and undergoes a spin-crossing close to **1a**. The quintet state of **1a** has three 5-coordinated Cu(II), with the the 4th unpaired spin on the bottom oxygen

stabilized by the exchange interactions with there same spin electrons in the Cu(II)s.

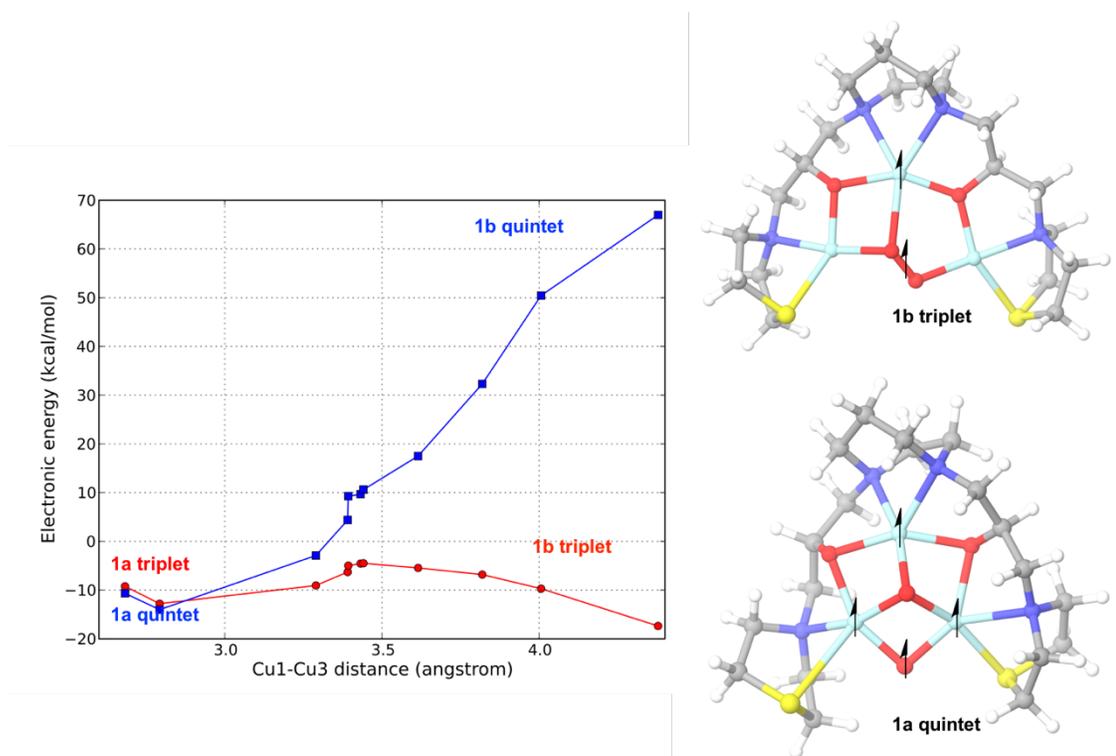


Figure 4.4: Transformation from Cu(I)Cu(I)Cu(II)(superoxo) species **1b** to Cu(II)Cu(II)Cu(II)(oxygen radical) species **1a**.

To confirm that this small energy gap between the triplet and quintet states at **1a** geometry is not an artifact of the broken-symmetry property of the DFT calculation used for the triplet state, we carried out Complete Active Space Self Consistent Field (CASSCF) calculation. For the 4 electrons in 4 orbitals (4e/4o) calculation which uses the localized highest energy alpha-spin orbitals from the quintet state DFT calculation, two separate CASSCF calculations were performed, converging to the lowest triplet and quintet states respectively. The energy of the lowest triplet state from the calculation converging to that state is 0.048 kcal/mol higher than the energy of the lowest quintet

state from the calculation converging to that state. A larger active space calculation with 18 electrons in 11 orbitals (18e/11o) was performed converging to the lowest quintet state. The energy of the lowest triplet state is 0.049 kcal/mol higher than the lowest quintet state from the same calculation. These two CASSCF calculations confirmed the near-degeneracy of the low lying quintet and triplet states, revealing that the quintet ground state is stabilized by the exchange energy of the 4 unpaired electron with the same spin (Figure 4.5).

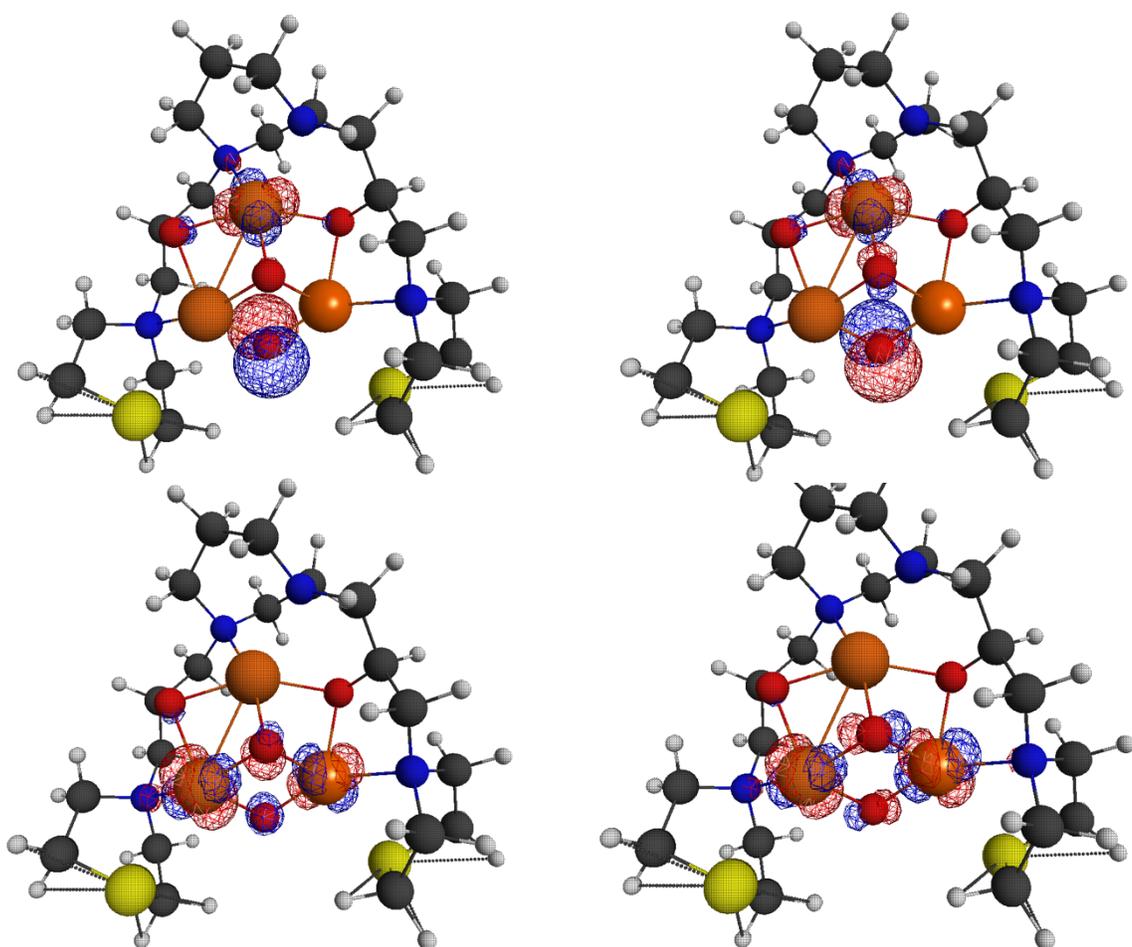


Figure 4.5: The four natural orbitals from 4 electrons in 4 orbitals CASSCF calculation converging to the quintet ground state.

### 4.3.2 The methane activation

Both the triplet and quintet states of **1** have a radical character on the bottom oxygen.

This O-radical is stabilized by the exchange interaction with the unpaired electron with same spin on the two bottom Cu(II)s. Binding an equivalence of CH<sub>4</sub> gives a free energy decrease of 13.3 kcal/mol from separate **1** and gas phase CH<sub>4</sub> (MC0) to a van der Waals complex between **1** and CH<sub>4</sub> (MC1) on the triplet surface and 13.8 kcal/mol on the quintet surface (Figure 7). The next step is the hydrogen atom abstraction by the bottom O-radical through a linear O-H-C transition state. The free energy barrier for this step is 11.5 kcal/mol for the triplet surface and 11.3 kcal/mol for the quintet surface. After the hydrogen atom transfer, the methyl radical forms a van der Waals complex with the O-radical quenched compound (MC2), where the newly formed O-H bond points towards the methyl radical. In the next step, the O-H bond formed in the previous hydrogen atom transfer step reverses its direction, going through a barrier of 10.8 kcal/mol on the triplet surface and that of 10.7 kcal/mol on the quintet surface, forming a van der Waals complex with the O-H bond pointing away from the methyl radical (MC3). On the triplet surface, the methyl radical could rebound to the -OH group in the compound, go through a barrier of 6.2 kcal/mol and end up with a neutral methanol product coordinated on one of the bottom coppers (MC4t). On the quintet surface, however, the singlet coupling between the methyl radical and the -OH group is energetically unfavorable, as MC4q is ~80 kcal/mol higher than MC4t. The triplet and quintet surfaces undergoes another spin crossing between TS2 and MC2. The quintet surface goes up to a thermally inaccessible energy level for the temperature of the reaction condition. This shows that the methyl radical rebound mechanism is spin-allowed on the quintet surface, but spin-forbidden on

the triplet surface, though the earlier hydrogen abstraction and the -OH bond inversion steps could happen on both surfaces.

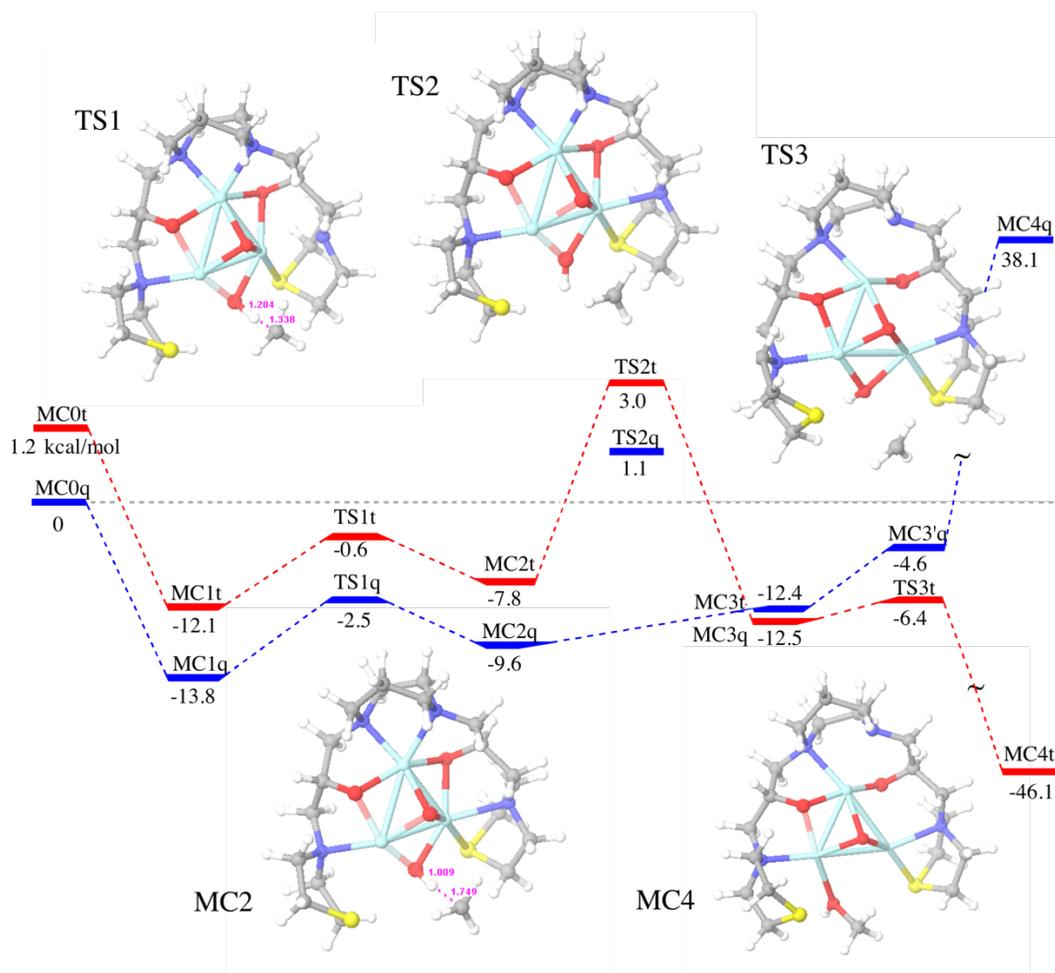


Figure 4.6: Free energy profile of the methane activation at the B3LYP-D3 level of theory. The values in are given in kcal/mol and were obtained relative to the quintet state of the 1a with separate a methane molecule.

### 4.3.3 Hydrogen peroxide activation

After the methanol product dissociates from the complex in the previous step, the tricopper compound still has two remaining oxidizing power. An equivalence of  $\text{H}_2\text{O}_2$  could bound to the compound and serve as a sacrificing reductant, regenerating the most

reduced Cu(I)Cu(I)Cu(I) state **3** and an equivalence of H<sub>2</sub>O and O<sub>2</sub>. The energies in Figure 9 are enthalpies comparing to the separate compound **3**, H<sub>2</sub>O and O<sub>2</sub>.

In PC1, H<sub>2</sub>O<sub>2</sub> bound to the tricopper compound through a hydrogen bond after methanol dissociation. The complex then undergoes the first hydrogen atom transfer through a barrier of 4.0 kcal/mol (TS4). The intermediate from the first hydrogen transfer PC2 undergoes an isomerization where the oxygen in the –OH group breaks one of its bonding to the right bottom copper and the hydrogen in the –OH group forms a hydrogen bond with the –OOH group bound to the right bottom copper. Despite our effort, the transition state for this isomerization TS5 is not found. After that, the second hydrogen could transfer from the –OOH group to the –OH group through TS6 with a barrier of 2.4 kcal/mol. The products from the second hydrogen transfer, H<sub>2</sub>O and O<sub>2</sub>, could dissociate from the tricopper compound, gaining 30 kcal/mol entropy to pay for the lost of -37.9 kcal/mol enthalpy in PC5.

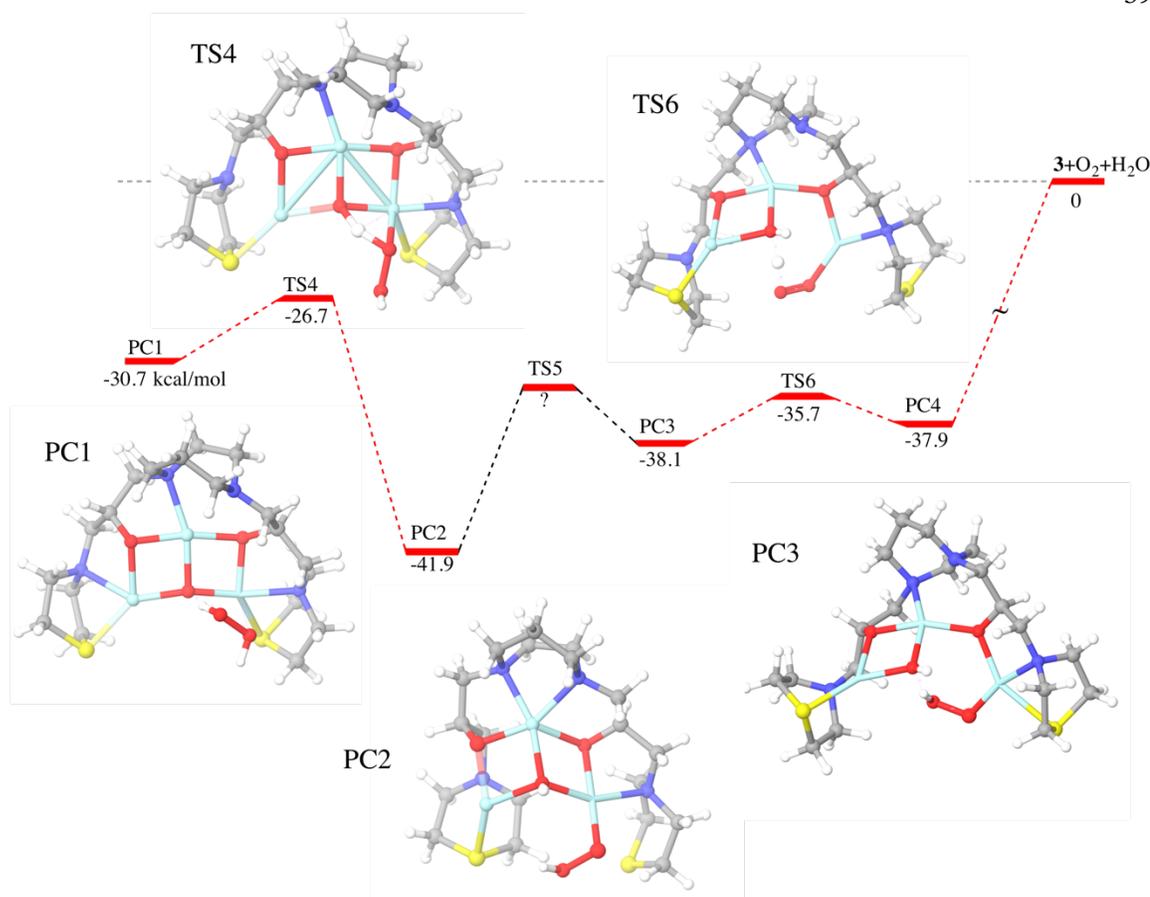


Figure 4.7: Enthalpy profile of the hydrogen peroxide activation at the B3LYP-D3 level of theory. The values in are given in kcal/mol and were obtained relative to the state **3** with separate H<sub>2</sub>O and O<sub>2</sub>.

#### 4.3.4 Discussion

The reaction mechanisms of the tricopper complex studies in this work highlighted that the active state for C-H activation has a multiradical character where the non-overlapping unpaired electrons in the singly occupied  $d_{x^2-y^2}$  orbitals of each Cu(II) in stabilized the O-radical with the same spin by exchange interactions. The exchange-stabilized O-radical is able to abstract the H-atom in the C-H bond through hydrogen atom transfer and form the final methanol product by methyl rebound mechanism if the singlet bond coupling is

spin-allowed. This mechanism is different from the the “singlet oxene-insertion” mechanism<sup>97</sup> proposed by the Chan group who designed the tricopper complex. In Chan’s model, the active state is a Cu(II)Cu(II)Cu(III) complex which could activate the C–H bond through a direct insertion of singlet oxene into the C–H bond in a concerted manner.

#### 4.4 Conclusions

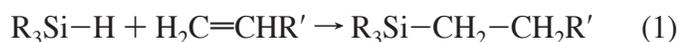
In this study, DFT calculations at the level of B3LYP-D3 functional were carried out to the reaction mechanisms of the methane hydroxylation by a biomimetic tricopper complex. The reaction mechanisms involve three key steps: (i) the oxygen activation, (ii) the methane activation, and (iii) hydrogen peroxide activation. During the first two steps, two potential energy surfaces of the triplet and quintet states undergo spin-crossing twice. The first spin-crossing in the oxygen activation is due to the multiradical character and the weak coupling between the back Cu(II) and the strongly coupled quartet in the bottom dicopper cluster. The second spin-crossing in the methane activation is due to the high spin quintet state does not favor the methyl rebound mechanism. Both the O-radical hydrogen abstraction followed by methyl rebound mechanism and the previously proposed “oxene C-H bond concerted insertion” mechanism are kinetically possible based on computational insights. The computational results in this work are not enough to prove the proposed mechanism, but only to provide a possibility for an alternative mechanism.

*Chapter 5*MECHANISM FOR THE KARSTEDT CATALYZED HYDROSILYLATION  
REACTION**Abstract**

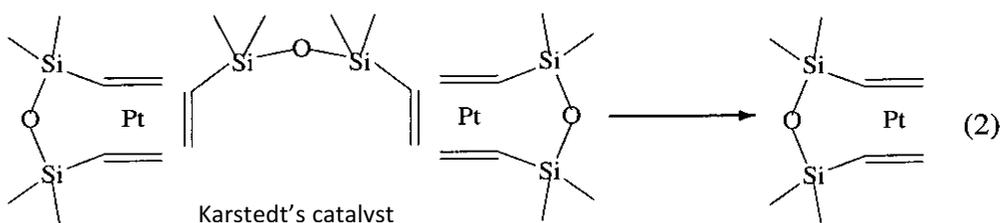
The activity of Pt-based Karstedt's catalyst competent for olefin hydrosilylation has been explored by computing free energy surfaces for reaction mechanisms composed of elementary transformations with literature precedent. Methoxydimethyl(vinyl)silane and trimethyl silane were used as model substrates, instead of using most simplified substrates of silicon hydrides and siloxo vinyls as in the previous computational studies. An unexpected rate-determining step of agostic bond dissociation is found in between the elementary reaction steps proposed previously. The regioselectivity of the products are studied. An alternative reaction cycle which is kinetically unflavored is proposed. Since little data is available on the decomposition or oxidation products of unstable olefin hydrosilylation catalysts, we considered reactions of O<sub>2</sub> with catalytic intermediates.

## 5.1 Introduction

Hydrosilylation reactions (eq 1) are among the most important reactions in silicone and siloxane chemistry<sup>98</sup>. These type of reactions are extensively used to form Si-C bonds, with an important application to cross link hydrosiloxanes with vinyl terminated dimethylsiloxanes for the production of protective backing surfaces in pressure sensitive adhesive labels. The catalysts which could catalyze these type of reactions usually involve late transition metals, most notably Pt, Pd, Ni, Rh, and Co



Low valent Pt compounds are typically used for the cross linking. Of these, one of the best characterized is Karstedt's catalyst 2:



Karstedt's catalyst 2 is composed of  $\text{Pt}_2[(\text{H}_2\text{C=CH})(\text{CH}_3)_2\text{SiOSi}(\text{CH}_3)_2-(\text{CH=CH}_2)]_3$ .

Upon solvation, the bridging ligand is released, yielding the more active form

$\text{Pt}[(\text{H}_2\text{C=CH})(\text{CH}_3)_2\text{SiOSi}(\text{CH}_3)_2-(\text{CH=CH}_2)]$ .

Chalk and Harrod<sup>99,100</sup> proposed a mechanism for platinum catalyzed hydrosilylation reactions based on simple elementary steps commonly observed in organometallic chemistry, such as oxidative addition and reductive elimination (Figure 5.1, route a with green arrows). A variation of the Chalk-Harrod mechanism was also proposed to explain the formation of vinylsilanes.<sup>101</sup> This modified version assumes that an olefin attacks on

the Pt-Si bond which has never been observed experimentally. Although both the original and the modified Chalk-Harrod mechanisms correctly describe the hydrosilylation reaction, they could not account for a number of phenomena observed mainly at the beginning (induction period) and the end (change in color) of the reaction.

To explain the induction period, color change, and several other effects, Lewis proposed a different mechanism which suggested that the actual catalytic species are colloidal platinum particles instead of organometallic complexes.<sup>102,103</sup> The Lewis mechanism regarded the reaction as a heterogeneous process that happens at the surfaces, or edges, of the Pt colloidal particles. Later, he realized that hydrosilylation does not depend on the existence of colloidal particles, though most experiments result in the formation of them. When the original Lewis mechanism is translated from the platinum surface to an organometallic compound, it becomes a homogeneous process and is similar to the older Chalk-Harrod mechanism.<sup>104</sup> The main difference of the Lewis mechanism from the Chalk-Harrod mechanism is the order in which different ligands are coordinated and activated (Figure 1, route b with blue arrows).

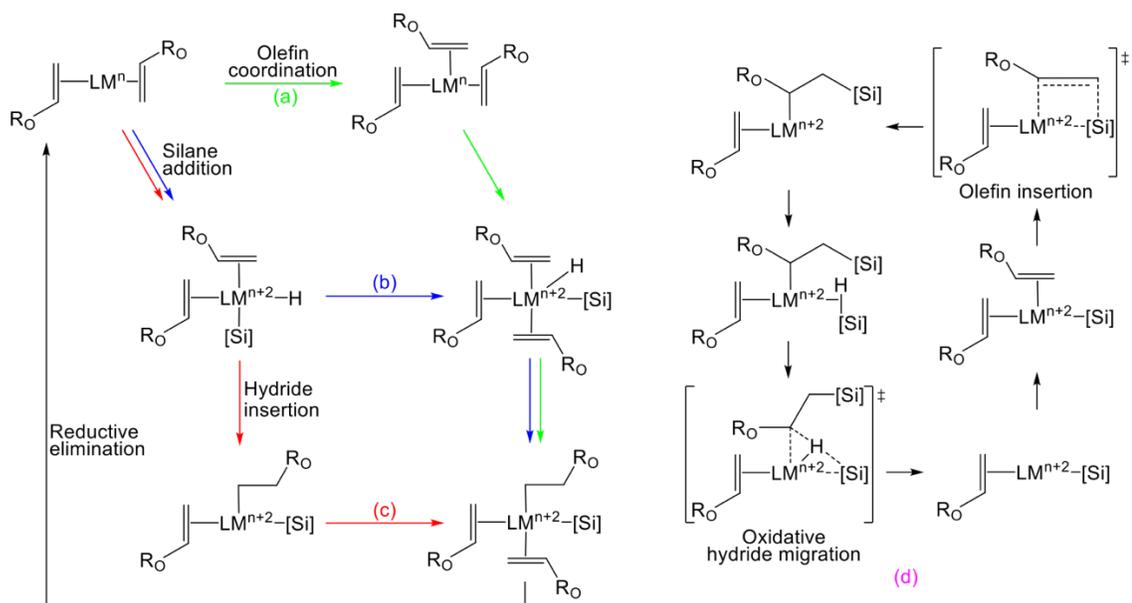


Figure 5.1: Redox cycles for olefin hydrosilylation (a-c) and (d) the non-redox cycle based on the established mechanisms.

Previous computational studies have investigated each of the elementary steps involved in the Chalk-Harrod, Lewis, and an alternative mechanism (Figure 1, route c with red arrows), with the most simplified reactants  $\text{SiH}(\text{OH})_2$  and  $\text{SiH}_2\text{OH}$ , where the long chain alkyl groups of the substrates are all substituted by hydrogen.<sup>105</sup> On the basis of these results, it was concluded that all three mechanisms represent minor variations of the same catalytic cycle and that the activation energy is small.

In this work, we used two more realistic substrates sets (Figure 5.2) to computationally study the reaction mechanisms involved in the catalytic cycle and tried to explain qualitatively the variation of regioselectivity due to the change in the substrates. We also studied a possible mechanism which involves silyl group migration as illustrated in

Figure 5.1, route d. At last, we looked into the oxygen binding affinity of pre-catalyst and the resting state in the reaction cycle to study the atmospheric stability of the catalyst.

## 5.2 Computation Details

The DFT calculations are performed using the Jaguar ab initio electronic structure program<sup>59</sup>, which is a commercial software currently produced and maintained by Schrödinger Inc. Keywords in the Jaguar input file is included in the parenthesis follow the description of the detailed method for readers who are interested to reproduce the results. B3LYP<sup>16,17</sup> functional is used combining D3 dispersion correction<sup>86</sup> (dftname=B3LYP-D3). For geometry optimization, a medium and polarized basis set LACVP\*\*<sup>87</sup> is used (basis=LACVP\*\*). For single point energies of the optimized geometries, a large basis set cc-pvtz(-f)<sup>88-91</sup> is used (basis=cc-pvtz(-f)). Both geometry optimization and single point energy calculations are done with Poisson-Boltzmann solvation model (isolv=2, nogas=2) for solvent dichloromethane to mimic the polymer matrix environment. The complexes were optimized in the gas phase, and the optimized structures were subsequently subjected to a single-point energy calculation using the PBF solvation model. Frequencies and thermodynamic contribution to the free energy were calculated for the optimized geometries in the gas phase at 373.15 K. The reported free energies were calculated by adding the thermodynamic correction, including zero-point energy, and the dispersion correction to the single-point solution phase energy. The presented transition states (TSs) were verified to have only one imaginary frequency and were verified to be connected to the presented succeeding and preceding energy minima, respectively. All stationary minima were verified to have no imaginary frequencies.

### 5.3 Results and Discussion

The reaction mechanisms of the Karstedt's platinum catalyst are modeled using the two sets of simplified substrate side-chains, as illustrated in Figure 5.2. For the olefin substitution  $R_0$  group, set 2 uses trimethoxysilyl group and set 1 uses methoxy(diethyl)silyl group. For the silane substrate, set 1 is trimethylsilane, while one of the three methyl groups is substituted to a dimethylsiloxy group. Overall, the substrates in set 2 are bulkier than the substrates in set 1, representing the local environment of the reaction centers in a polymer matrix. Experiments have shown that the bulkier substrate set 2 yields higher anti-Markovnikov vs. Markovnikov product ratio than that of substrate set 1, with the exact numbers uninformed for this project.

We compared several competing reaction routes in Figure 1 and found that the sequence of silane addition, hydride migration and reductive elimination is kinetically the most preferred route. The rate-limiting step of this route is the transition state B in Fig. 2 where the agostic bond breaks after hydride migration. This elementary step is not explicitly included in the previously proposed mechanisms. The difference in this free energy barrier,  $\Delta\Delta G$  between Markovnikov and anti-Markovnikov transition states, determines the selectivity of the two regioisomers. For substrate set 1, a  $\Delta\Delta G(M-TM)$  of 0.7 kcal/mol gives an AM/M ratio of 77/23 at room temperature. For the bulkier substrate set 2, a decreased  $\Delta\Delta G(M-TM)$  of 0.2 kcal/mol gives a reduced AM/M ratio of 55/45 at room temperature, which agrees with the experimental trend.

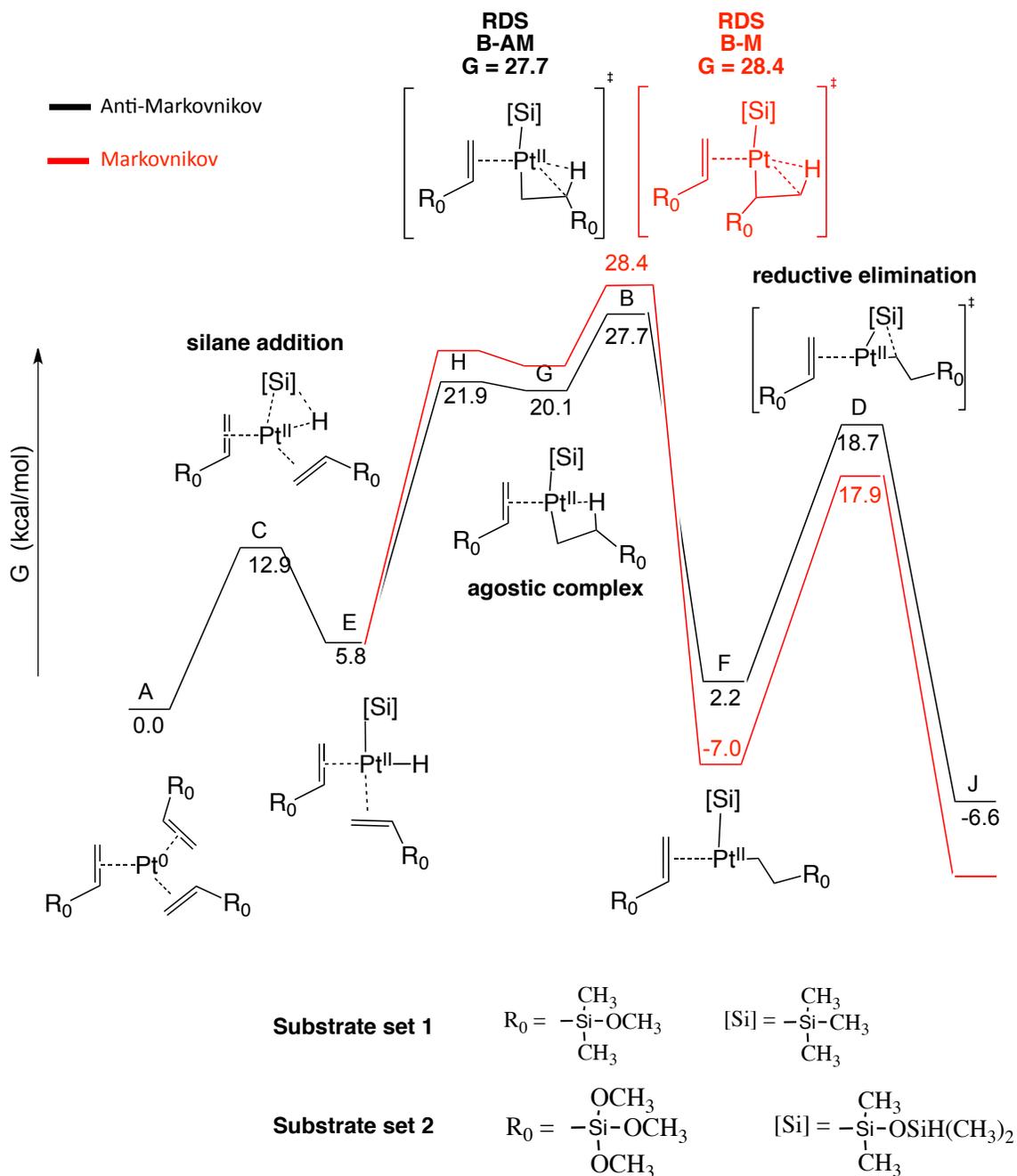


Figure 5.2: The free energy profile of the Pt(II) reductive elimination route of the Karstedt catalyst, using the simplified substrate set 1. The transition state B is the rate determining step (RDS) and the difference in free energy between the anti-Markovnikov route (black) and the Markovnikov route (red) in state B gives rise to the regioselectivity of this reaction.

The reaction mechanism is further complicated by several Pt(II) routes and potential Pt(IV) routes competing unfavorably with the reductive elimination route, as mapped out in Fig. 3. These alternative pathways were only explored for the substrate set 1. Several possible reaction mechanisms could be drawn from state F, which is the intermediate after the rate-determining agonist bond breaking step. Besides the reductive elimination route from F to J to form product, F could have four alternative fates: (i) F undergoes silyl migration to the  $\alpha$ -carbon of the olefin with transition state L, (ii) F undergoes silyl group migration to the  $\beta$ -carbon of the olefin with transition state N, (iii) F binds another silane substrate and undergoes hydride migration with transition state M, and (iv) A silane exchange with the olefin bound in F and gets activated in P, then hydride transfer in transition state Q to form the product. Comparisons between these five routes showed that the reductive elimination has the lowest barriers of 24.9 kcal/mol for forming Markovnikov product and 20.9 kcal/mol for forming anti-Markovnikov product. The next lowest barrier route is from F to Q which is 2.0 kcal/mol higher for Markovnikov product and 3.1 kcal/mol higher for anti-Markovnikov product.

If the reaction goes through the kinetically unfavorable route of transition state Q, the product could dissociate and another olefin could bind the Pt catalyst and form the state R which is thermodynamically more favorable. The state R could be a sink state once it is formed because of the high barrier to get out through the transition state S of silyl migration to further cycle through silane activation and hydride migration via transition state U.

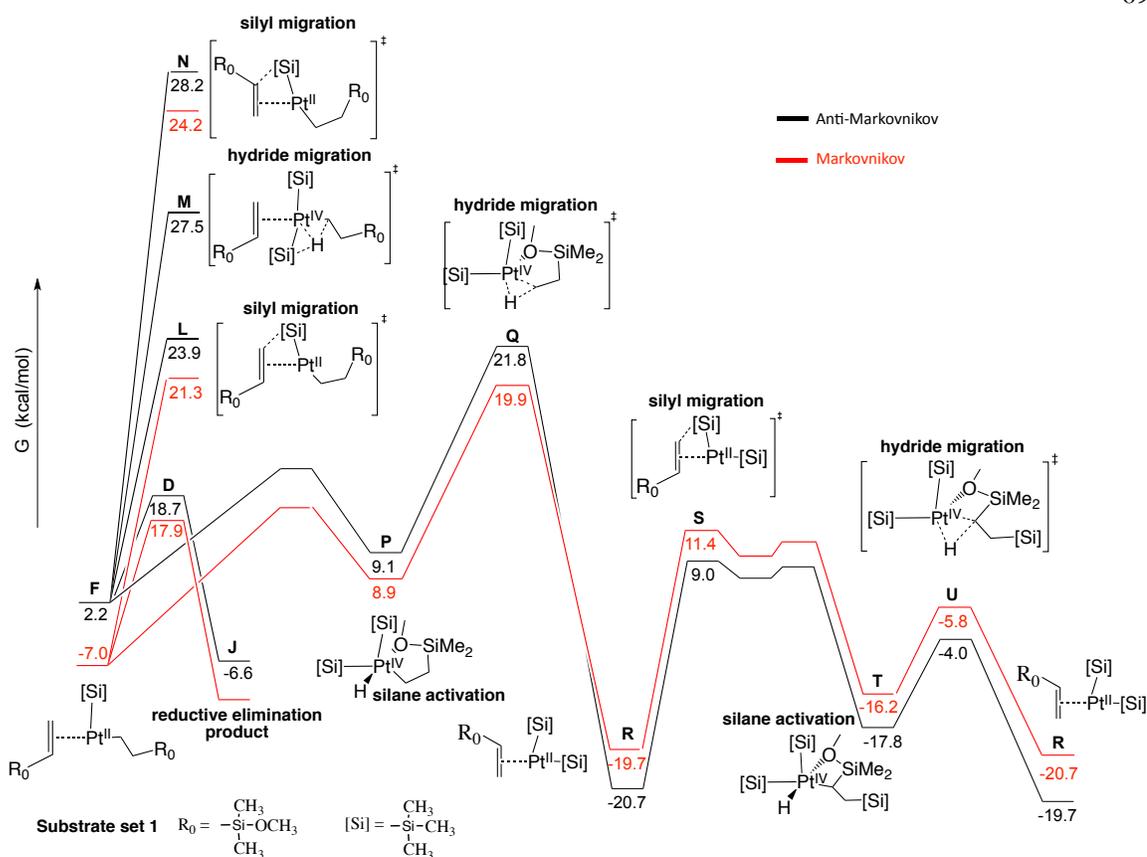


Figure 5.3: The free energy profile of several competing Pt(II) and Pt(IV) reaction routes of the Karstedt catalyst, using the simplified substrate set 1. The transition state Q involved in activating another silane from the state F has a much higher barrier than the transition state D involved in reductive elimination to enter a different turnover cycle  $R \rightarrow S \rightarrow U \rightarrow R$  via Pt(IV). But the state R could serve as a thermodynamic sink as it is the lowest state in free energy among all the calculated states and has a relatively higher barrier in the transition state S to make any products.

The oxygen sensitivity of the pre-catalyst state A and several resting states are also studied, as the examples shown in Fig. 4. The lowest  $\text{O}_2$  electronic binding energy is -9 kcal/mol, which could hardly beat  $\sim 15$  kcal/mol entropy penalty to form a physically-absorbed complex thus eliminate the possibility of chemical reactions between  $\text{O}_2$  and the catalyst.

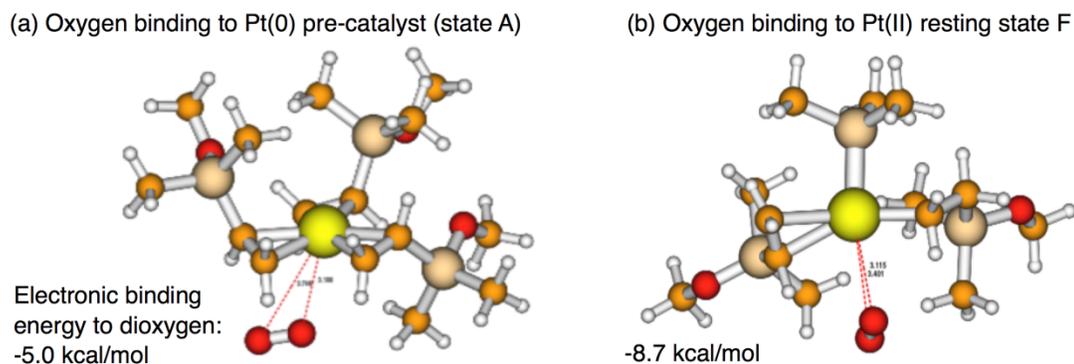


Figure 4. (a) The minimized structure of  $O_2$  bound Pt(0) pre-catalyst state A. The average distance between the dioxygen to the Pt atom is about  $3.4\text{\AA}$  and the electronic binding energy is  $-5.0\text{ kcal/mol}$ . (b) The minimized structure of  $O_2$  bound Pt(II) resting state F. The average distance between the dioxygen to the Pt atom is about  $3.2\text{\AA}$  and the electronic binding energy is  $-8.7\text{ kcal/mol}$ . In both cases the electronic binding energy for  $O_2$  could not beat the  $15\text{ kcal/mol}$  entropy penalty for binding a free  $O_2$  molecule.

## 5.4 Conclusions

In this study, the major points regarding the mechanism and selectivity of Karstedt's catalyst have been addressed with more realistic substrate sets. It was found that silane addition, hydride migration, and reductive elimination is kinetically the most preferred route. Between hydride migration and reductive elimination, the agnostic bond dissociation step unexpectedly turned out to be the rate-determining step in the reaction cycle. Free energy barrier differences between Markovnikov and anti-Markovnikov transition states are calculated for the two substrate sets respectively to explain the trend of the change of regioselectivity between the two substrate sets. Alternative competing reaction mechanisms were also explored outside the cycle of covered by previously studied Chalk-Harrod, Lewis, and other mechanisms. Instead of the reductive elimination step to form product, a

kinetically unflavored route through silane-olefin coordination exchange, silane activation and hydride migration could lead to a thermodynamically stable disilyl intermediate. This intermediate could be a thermodynamic sink state under room temperature but also cycle through Pt(IV) intermediate to form product under higher temperature. The pre-catalyst and the reaction intermediates from the catalytic cycle have very weak binding affinity to O<sub>2</sub>, supporting the experimentally observed O<sub>2</sub> stability. However, the role of O<sub>2</sub> in the alternative reaction cycle undergoing silyl migration with the stable disilyl intermediate should be explored to see if O<sub>2</sub> pressure plays any beneficial role.

## BIBLIOGRAPHY

- (1) Burkert, U.; Allinger, N. L. *Molecular mechanics*; ACS monograph; American Chemical Society: Washington, D.C, 1982.
- (2) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79* (22), 2361–2381.
- (3) Lifson, S. *J. Chem. Phys.* **1968**, *49* (11), 5116.
- (4) Allinger, N. L. In *Advances in Physical Organic Chemistry*; Elsevier, 1976; Vol. 13, pp 1–82.
- (5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (6) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187–217.
- (7) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666.
- (8) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory*; Dover Publications: Mineola, N.Y, 1996.
- (9) Saebo, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44* (1), 213–236.
- (10) Bartlett, R. J. *J. Phys. Chem.* **1989**, *93* (5), 1697–1708.
- (11) Siegbahn, P. E. M. *J. Chem. Phys.* **1981**, *74* (4), 2384.
- (12) Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*, 2nd ed., 5. reprint.; Wiley-VCH: Weinheim, 2008.
- (13) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136* (3B), B864–B871.
- (14) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138.
- (15) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.
- (16) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648.
- (17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37* (2), 785–789.
- (18) Rosenberger, C. M.; Finlay, B. B. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (5), 385–396.
- (19) Taylor, P. R.; Martinez-Pomares, L.; Stacey, M.; Lin, H.-H.; Brown, G. D.; Gordon, S. *Annu. Rev. Immunol.* **2005**, *23* (1), 901–944.
- (20) Hornef, M. W.; Wick, M. J.; Rhen, M.; Normark, S. *Nat. Immunol.* **2002**, *3* (11), 1033–1040.
- (21) Foster, T. J. *Nat. Rev. Microbiol.* **2005**, *3* (12), 948–958.
- (22) Pascal, T. A.; Abrol, R.; Mittal, R.; Wang, Y.; Prasadarao, N. V.; Goddard, W. A. *J. Biol. Chem.* **2010**, *285* (48), 37753–37761.
- (23) Shanmuganathan, M. V.; Krishnan, S.; Fu, X.; Prasadarao, N. V. *Microbes Infect.* **2014**, *16* (2), 134–141.

- (24) Mittal, R.; Krishnan, S.; Gonzalez-Gomez, I.; Prasadarao, N. V. *J. Biol. Chem.* **2011**, *286* (3), 2183–2193.
- (25) Datta, D.; Vaidehi, N.; Floriano, W. B.; Kim, K. S.; Prasadarao, N. V.; Goddard, W. A. *Proteins* **2003**, *50* (2), 213–221.
- (26) Krishnan, S.; Prasadarao, N. V. *Microbes Infect. Inst. Pasteur* **2014**, *16* (7), 540–552.
- (27) Pautsch, A.; Schulz, G. E. *Nat. Struct. Biol.* **1998**, *5* (11), 1013–1017.
- (28) Pautsch, A.; Schulz, G. E. *J. Mol. Biol.* **2000**, *298* (2), 273–282.
- (29) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843.
- (30) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926.
- (32) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33–38.
- (33) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.
- (34) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8* (4), 334–338.
- (35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94* (26), 8897–8909.
- (36) Lu, J.; Ellsworth, J. L.; Hamacher, N.; Oak, S. W.; Sun, P. D. *J. Biol. Chem.* **2011**, *286* (47), 40608–40613.
- (37) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78* (8), 1950–1958.
- (38) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- (39) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.
- (40) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210.
- (41) Chen, R.; Li, L.; Weng, Z. *Proteins* **2003**, *52* (1), 80–87.
- (42) Pierce, B. G.; Hourai, Y.; Weng, Z. *PLoS ONE* **2011**, *6* (9), e24657.
- (43) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. *Bioinforma. Oxf. Engl.* **2006**, *22* (5), 623–625.
- (44) Drayman, N.; Glick, Y.; Ben-nun-shaul, O.; Zer, H.; Zlotnick, A.; Gerber, D.; Schueler-Furman, O.; Oppenheim, A. *Cell Host Microbe* **2013**, *14* (1), 63–73.

- (45) Mehlert, A.; Wormald, M. R.; Ferguson, M. A. J. *PLoS Pathog.* **2012**, *8* (4), e1002618.
- (46) Zhang, Y.; Hoon, M. A.; Chandrashekar, J.; Mueller, K. L.; Cook, B.; Wu, D.; Zuker, C. S.; Ryba, N. J. *Cell* **2003**, *112* (3), 293–301.
- (47) Janssen, S.; Depoortere, I. *Trends Endocrinol. Metab.* **2013**, *24* (2), 92–100.
- (48) Kokrashvili, Z.; Mosinger, B.; Margolskee, R. F. *Am. J. Clin. Nutr.* **2009**, *90* (3), 822S – 825S.
- (49) Prakash, I.; Markosyan, A.; Bunders, C. *Foods* **2014**, *3* (1), 162–175.
- (50) Hellfritsch, C.; Brockhoff, A.; Stähler, F.; Meyerhof, W.; Hofmann, T. *J. Agric. Food Chem.* **2012**, *60* (27), 6782–6793.
- (51) Goddard, W. A.; Kim, S.-K.; Li, Y.; Trzaskowski, B.; Griffith, A. R.; Abrol, R. *J. Struct. Biol.* **2010**, *170* (1), 10–20.
- (52) Abrol, R.; Griffith, A. R.; Bray, J. K.; Goddard, W. A. *Methods Mol. Biol. Clifton NJ* **2012**, *914*, 237–254.
- (53) Abrol, R.; Bray, J. K.; Goddard, W. A. *Proteins* **2012**, *80* (2), 505–518.
- (54) Bray, J. K.; Goddard, W. A. *J. Mol. Graph. Model.* **2008**, *27* (1), 66–81.
- (55) Bray, J. K.; Abrol, R.; Goddard, W. A.; Trzaskowski, B.; Scott, C. E. *Proc. Natl. Acad. Sci.* **2014**, *111* (1), E72–E78.
- (56) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G. W.; Tate, C. G.; Schertler, G. F. X. *Nature* **2008**, *454* (7203), 486–491.
- (57) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Won Han, G.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. *Science* **2010**, *330* (6007), 1091–1095.
- (58) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.
- (59) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. *Int. J. Quantum Chem.* **2013**, *113* (18), 2110–2142.
- (60) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102* (52), 10983–10990.
- (61) Lim, K.-T.; Brunett, S.; Iotov, M.; McClurg, R. B.; Vaidehi, N.; Dasgupta, S.; Taylor, S.; Goddard, W. A. *J. Comput. Chem.* **1997**, *18* (4), 501–521.
- (62) Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A. *J. Med. Chem.* **2004**, *47* (1), 56–71.
- (63) Kim, S.-K.; Li, Y.; Abrol, R.; Heo, J.; Goddard, W. A. *J. Chem. Inf. Model.* **2011**, *51* (2), 420–433.
- (64) Kim, S.-K.; Riley, L.; Abrol, R.; Jacobson, K. A.; Goddard, W. A. *Proteins* **2011**, *79* (6), 1878–1897.
- (65) Nair, N.; Kudo, W.; Smith, M. A.; Abrol, R.; Goddard, W. A.; Reddy, V. P. *Bioorg. Med. Chem. Lett.* **2011**, *21* (13), 3957–3961.
- (66) Tak Kam, V. W.; Goddard, W. A. *J. Chem. Theory Comput.* **2008**, *4* (12), 2160–2169.

- (67) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161* (2), 269–288.
- (68) Kobilka, B. K. *Biochim. Biophys. Acta BBA - Biomembr.* **2007**, *1768* (4), 794–807.
- (69) Deupi, X.; Kobilka, B. In *Advances in Protein Chemistry*; Elsevier, 2007; Vol. 74, pp 137–166.
- (70) Ahn, K. H.; Scott, C. E.; Abrol, R.; Goddard, W. A.; Kendall, D. A. *Proteins Struct. Funct. Bioinforma.* **2013**, *81* (8), 1304–1317.
- (71) Bercaw, J. E.; Labinger, J. A. *Proc. Natl. Acad. Sci.* **2007**, *104* (17), 6899–6900.
- (72) Labinger, J. A.; Bercaw, J. E. *Nature* **2002**, *417* (6888), 507–514.
- (73) Hall, C.; Perutz, R. N. *Chem. Rev.* **1996**, *96* (8), 3125–3146.
- (74) Arndtsen, B. A.; Bergman, R. G.; Mobley, T. A.; Peterson, T. H. *Acc. Chem. Res.* **1995**, *28* (3), 154–162.
- (75) Hanson, R. S.; Hanson, T. E. *Microbiol. Rev.* **1996**, *60* (2), 439–471.
- (76) Sazinsky, M. H.; Lippard, S. J. *Acc. Chem. Res.* **2006**, *39* (8), 558–566.
- (77) Baik, M.-H.; Newcomb, M.; Friesner, R. A.; Lippard, S. J. *Chem. Rev.* **2003**, *103* (6), 2385–2420.
- (78) Hakemian, A. S.; Rosenzweig, A. C. *Annu. Rev. Biochem.* **2007**, *76* (1), 223–241.
- (79) Lieberman, R. L.; Rosenzweig, A. C. *Nature* **2005**, *434* (7030), 177–182.
- (80) Chan, S. I.; Nguyen, H.-H. T.; Chen, K. H.-C.; Yu, S. S.-F. *Methods Enzymol.* **2011**, *495*, 177–193.
- (81) Chan, S. I.; Wang, V. C.-C.; Lai, J. C.-H.; Yu, S. S.-F.; Chen, P. P.-Y.; Chen, K. H.-C.; Chen, C.-L.; Chan, M. K. *Angew. Chem. Int. Ed.* **2007**, *46* (12), 1992–1994.
- (82) Chan, S. I.; Yu, S. S.-F. *Acc. Chem. Res.* **2008**, *41* (8), 969–979.
- (83) Chan, S. I.; Lu, Y.-J.; Nagababu, P.; Maji, S.; Hung, M.-C.; Lee, M. M.; Hsu, I.-J.; Minh, P. D.; Lai, J. C.-H.; Ng, K. Y.; Ramalingam, S.; Yu, S. S.-F.; Chan, M. K. *Angew. Chem. Int. Ed.* **2013**, *52* (13), 3731–3735.
- (84) Chan, S. I.; Chien, C. Y.-C.; Yu, C. S.-C.; Nagababu, P.; Maji, S.; Chen, P. P.-Y. *J. Catal.* **2012**, *293*, 186–194.
- (85) Nagababu, P.; Maji, S.; Kumar, M. P.; Chen, P. P.-Y.; Yu, S. S.-F.; Chan, S. I. *Adv. Synth. Catal.* **2012**, *354* (17), 3275–3282.
- (86) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (87) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82* (1), 299.
- (88) Dunning, T. H. *J. Chem. Phys.* **1989**, *90* (2), 1007.
- (89) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96* (9), 6796.
- (90) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98* (2), 1358.
- (91) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100* (4), 2975.
- (92) Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (1), 73–78.
- (93) Gherman, B. F.; Cramer, C. J. *Coord. Chem. Rev.* **2009**, *253* (5-6), 723–753.

- (94) Malmqvist, P. Å.; Pierloot, K.; Shahi, A. R. M.; Cramer, C. J.; Gagliardi, L. *J. Chem. Phys.* **2008**, *128* (20), 204109.
- (95) Yanai, T.; Kurashige, Y.; Neuscamman, E.; Chan, G. K.-L. *J. Chem. Phys.* **2010**, *132* (2), 024105.
- (96) Liakos, D. G.; Neese, F. *J. Chem. Theory Comput.* **2011**, *7* (5), 1511–1523.
- (97) Chen, P. P.-Y.; Chan, S. I. *J. Inorg. Biochem.* **2006**, *100* (4), 801–809.
- (98) *Comprehensive handbook on hydrosilylation*; Marciniak, B., Ed.; Pergamon Press: Oxford ; New York, 1992.
- (99) Harrod, J. F.; Chalk, A. J. *J. Am. Chem. Soc.* **1964**, *86* (9), 1776–1779.
- (100) Chalk, A. J.; Harrod, J. F. *J. Am. Chem. Soc.* **1965**, *87* (1), 16–21.
- (101) Ojima, I. In *The Chemistry of Functional Groups*; Patai, S., Rappoport, Z., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 1989; pp 1479–1526.
- (102) Lewis, L. N. *J. Am. Chem. Soc.* **1990**, *112* (16), 5998–6004.
- (103) Lewis, L. N.; Lewis, N. *J. Am. Chem. Soc.* **1986**, *108* (23), 7228–7231.
- (104) Stein, J.; Lewis, L. N.; Gao, Y.; Scott, R. A. *J. Am. Chem. Soc.* **1999**, *121* (15), 3693–3703.
- (105) Faglioni, F.; Blanco, M.; Goddard, W. A.; Saunders, D. *J. Phys. Chem. B* **2002**, *106* (7), 1714–1721.