Advancing Insights into the Structures and Functions of Interfacial Proteins and Peptides Using Spectroscopic Techniques

by

Wen Guo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2023

Doctoral Committee:

Professor Zhan Chen, Chair
Associate Professor Andrew P. Ault
Professor Kenichi Kuroda
Professor E. Neil G. Marsh

Wen Guo

weguo@umich.edu

ORCID iD: 0000-0003-2527-545X

Dedication

To my family

Acknowledgements

I would first like to express my gratitude to my advisor Professor Zhan Chen for his continued support throughout the past five years. Professor Chen not only provides extensive scientific suggestions on my research, but also guided me to think deeply and critically, to present my research work with logic and confidence, and to become an independent researcher in the field of chemistry. Professor Chen respected any ideas I came across and he was always willing to sit down to discuss any challenges and problems I encountered. Professor Chen also gave me tremendous freedom in time- and project-management and taught me how to keep work-life balance, and to embrace the ups and downs in research, and most importantly, in everyday life.

I would like to extend my thanks to all my lab fellows who helped me at different stages of my research. For the moment I joined Chen lab as an exchange undergraduate student in 2017, I was immediately captivated by the friendly, inclusive, and supportive environment within the group. The first two lab members I came across, Dr. Nathan Ulrich and Dr. John Andre, warmly welcomed me on my very first day. Their expertise on polymer science provided me with invaluable insights into advanced polymer materials. I would also like to thank Dr. Minyu Xiao, Dr. Xingquan Zou, Dr. Joshua Jasensky and Dr. Shan Xu as my mentors for SFG experiments and my research related to surface-immobilized peptides and proteins. It was through these experiences as an exchange student that my interests in interfacial biologics was nurtured, ultimately leading me to pursue my Ph.D in Chen lab. I would like to thank Dr. Tieyi Lu, who has extensive experience in studying interfacial structures and functions of proteins and peptides, gave me

tremendous help on solving my research problems. Additionally, I also want to thank Dr. Daniel Rossi and Dr. Bolin Li, true optics engineers, who patiently introduced me to the wonder of optics and engaged in meaningful scientific conversations on all various topics. To Dr. Ralph Crisis, Dr. Ruiying Guo, Dr. Jun Zhao, Shuqing Zhang, Yuchen Wu, and Ting Lin, Jinpeng Gao, Guangyao Wu, Zahra Gandhi, Xuhong Chen and Najae Escoffery, I am deeply appreciative of all your help and insightful discussions in the past five years.

Furthermore, I want to thank Associate Professor Tao Wei from Howard University for his incredible guidance in the field of molecular dynamic simulations. I am also thankful to Professor E. Neil G. Marsh, Professor Kenichi Kuroda and Associate Professor Andrew P. Ault for being my dissertation committee members. I extend my sincere appreciation to Rackham Graduate School for the predoctoral fellowship and to the Department of Chemistry for the fellowships awarded during one winter semester and one summer semester.

Finally, to my family: If you never abandon, I will be in life and death.

Wen Guo

07/26/2023

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables ix

List of Figures x

Abstract xvii

Chapter 1 Introduction 1

- 1.1 Attribution Information 1
- 1.2 Motivation and Background 1
- 1.3 SFG Spectroscopy 6
  - 1.3.1 Basic Principals 6
  - 1.3.2 SFG Experimental Setup 8
- 1.4 Supplementary Techniques 10
  - 1.4.1 ATR-FTIR Spectroscopy 10
  - 1.4.2 CD Spectroscopy 11
  - 1.4.3 Dynamic Light Scattering 12
- 1.5 Molecular Orientation Determination of Proteins and Peptides 12
  - 1.5.1 Bond Additivity Model for α-helical Segments 12
  - 1.5.2 Hamiltonian Approach 16
- 1.6 Atomistic Molecular Dynamic Simulation 20
- 1.7 Presented Research 20
- 1.8 References 23

Chapter 2 Molecular Interactions Between Surface-Immobilized Antimicrobial Peptides and Lipopolysaccharides 28

2.1 Attribution Information.... 28

2.2 Background and Motivation 28

2.2.1 Antimicrobial Peptides (AMPs).... 28

2.2.2 Lipopolysaccharides (LPS).... 29

2.2.3 SMAP29.... 30

2.3 Experimental Materials and Methods 31

2.4 Results and Discussion 34

2.4.1 Surface-immobilized SMAP29 Variants in Water 34

2.4.2 Quantitative Analysis of Immobilized SMAP29 Variants.... 40

2.4.3 Dynamics from the Time-Dependent SFG Spectra 44

2.5 Conclusion 45

2.6 References.... 45

Chapter 3 Molecular Structure of the Surface-Immobilized Super Uranyl Binding Protein....48

3.1 Attribution Information.... 48

3.2 Background and Motivation 48

3.2.1 Super Uranyl Binding Protein (SUP).... 48

3.2.2 Methodology of Structural Determination of Interfacial SUP.... 49

3.3 Structures of Immobilized SUP on SAM Surface with and without the Presence of Uranyl Ions.... 50

3.3.1 Experimental Materials and Methods 50

3.3.2 Parameters for Hamiltonian Approach 51

3.3.3 Score System for Spectral Comparison between Experimental Measurements and Hamiltonian Calculations.... 52

3.3.4 DMD Simulations 53

3.3.5 Results and Discussion 54
3.4 Conclusion 64
3.5 References 65
Chapter 4 Determine GB1 Conformation and Orientation at Solid/Liquid Interface 69
4.1 Attribution Information 69
4.2 Background and Motivation 69
4.2.1 B_1 Domain of Protein G (GB1) 70
4.2.2 Isotope Labeling 71
4.3 Experimental Materials and Methods 72
4.3.1 Preparation of WT GB1 and MT GB1 with and without Isotopic Labels 72
4.3.2 Preparation of PS Thin Film 72
4.3.3 SFG Experimental Setup 73
4.3.4 Atomistic MD Simulations 73
4.3.5 Hamiltonian Approach and Scoring System 75
4.4 SFG Results 76
4.5 Atomistic MD Simulation Results 79
4.6 Hamiltonian Matching Results 82
4.6.1 Spectral Matching for WT GB1 82
4.6.2 Spectral Matching for MT GB1 85
4.6.3 Crystal Structure for SFG Spectral Calculation 88
4.7 Comparison between Interfacial WT GB1 and Interfacial MT GB1 90
4.8 Conclusion 92
4.9 References 94
Chapter 5 Conformation and Orientation of Antimicrobial Peptides MSI-594 and MSI-594A in a Lipid Membrane 96
5.1 Attribution Information 96

5.2 Background and Motivation 96
  5.2.1 Antimicrobial Peptide (AMP)-Lipid Membrane Interaction Mechanisms 96
  5.2.2 MSI-594, MSI-594A and Model Membranes 97
  5.2.3 Combination of SFG and ATR-FTIR Measurements 98
5.3 Experimental Materials and Methods 99
  5.3.1 Materials 99
  5.3.2 Hamiltonian Approach and Scoring System 100
5.4 Results and Discussion 102
  5.4.1 SFG and ATR-FTIR Measurements 102
  5.4.2 Hamiltonian Matching Results 103
5.5 Conclusion 119
5.6 Reference 121
Chapter 6 Conclusions and Future Work 124
6.1 Summary of the Achievements in This Thesis Research 124
6.2 Future Directions 127
  6.2.1 Experimental Measurements by Vibrational Spectroscopic Techniques 127
  6.2.2 In-silico Modeling Methods 129
  6.2.3 Hamiltonian Approach 131
  6.2.4 Score System 132
  6.2.5 Development of Other Useful Methodologies 133
6.3 Limitations and Challenges 133
6.4 References 134

List of Tables

Table 2-1 χzzz/χyyz Ratios Deduced from Fitted SFG Spectra 39

Table 2-2 Possible Orientation Angle Ranges of NT−LPS and CT−LPS Pairs. 43

Table 3-1 Fitting results of the collected SFG spectra displayed in Figure 3-2 55

Table 5-1 Highest scores obtained from the final score maps shown in Figure 5-3 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined helical hairpin structure of MSI-594 (PDB ID: 2K98). 105

Table 5-2 Highest scores obtained from the final score maps shown in Figure 5-7 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined structure of MSI-594A (PDB ID: 2L36). 109

Table 5-3 Highest scores obtained from the final score maps shown in Figure 5-10 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined helical hairpin structure of MSI-594 (PDB ID: 2K98). 111

Table 5-4 Highest scores obtained from the final score maps shown in Figure 5-12 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined structure of MSI-594A (PDB ID: 2L36). 111

List of Figures

Figure 1-1 (a) SFG energy diagram. (b) SFG near-total reflection sample geometry. 8

Figure 1-2 Schematic of multi-reflection ATR-FTIR experimental setup. 11

Figure 1-3 Schematic of CD experimental setup to study interfacial proteins and peptides. 12

Figure 1-4 Schematic of defined orientation of the transition dipole (red), the principal coordinate axes of the Raman polarizability (purple) and molecular frame coordinate (orange) in an individual amide unit. 17

Figure 2-1 Archetypal LPS structures from: (A) E. coli O55:B5 (B) E. coli O111:B4 and (C) P. aeruginosa. 30

Figure 2-2 Schematic diagram of SFG sample geometry. 33

Figure 2-3 CD spectra collected from NT, CT, and Y17C in the free-solution state (A) and CD spectra collected from surface-immobilized NT (B), CT (C), and Y17C (D) in water and various LPS solutions. 35

Figure 2-4 SFG ppp (red) and ssp (black) amide I spectra (dots) collected from surface-immobilized (A) NT, (B) CT, and (C) Y17C in water. The fitting results are shown as lines..... 36

Figure 2-5 SFG ppp (red) and ssp (black) amide I spectra (dot) collected from the interfaces between surface-immobilized NT and (A) O55:B5, (B) O111:B4, and (C) P.A. solutions; between surface-immobilized CT and (D) O55:B5, (E) O111:B4, and (F) P.A. solutions; and between surface-immobilized Y17C and (G) O55:B5, (H) O111:B4, and (I) P.A. solutions. The spectral fitting results are shown as lines. The nonzero baselines of some spectra are due to the nonresonant signal contributions. 39

Figure 2-6 Illustration of a bent alpha-helix structure. 40

Figure 2-7 Relationship between the SFG $\chi zzz/\chi yyz$ ratio and tilt angles of a bend alpha-helical SMAP29 with the first alpha-helix segment of 12 residues and the second alpha-helix segment of 9 residues. The tilt angle is defined as the angle between the surface normal and the first fragment. Curves in different colors represent different bent angles between the first and the second segments. 41

Figure 2-8 Size distributions of LPS vesicles measured by the DLS experiment. 42

Figure 2-9 Schematics of NT (bent angle at 60°) in contact with (A) O55:B5, (B) O111:B4, and (C) P.A., CT (bent angle at 60°) in contact with (D) O55:B5, (E) O111:B4, and (F) P.A., and (G) Y17C in contact with LPS solutions. Y17C adopts nonhelical structures when in contact with LPS solutions; therefore, the orientations (tilt angles and bent angles) are not determined. 43

Figure 2-10 SFG time dependence ppp signal intensity measured at 1655 cm−1 of (A) NT/O111:B4, (B) NT/O55:B5, (C) NT/P.A., (D) CT/ O111:B4, (E) CT/O55:B5, (F) CT/P.A., (G) Y17C/O111:B4, (H) Y17C/O55:B5, and (I) Y17C/P.A. interfaces. 45

Figure 3-1 Schematic of the SFG sample geometry. 51

Figure 3-2 SFG ssp and ppp amide I spectra collected from the interfaces between surface-immobilized (a) SUP-C105 in contact with PB solution and (b) SUP-C105 in contact with a 100 μM uranyl ion PB solution. Dots represent experimental data, while lines are fitting curves. Fitted resonant SFG spectra from the immobilized (c) SUP-C105 in contact with PB solution and (d) SUP-C105 in contact with a 100 μM uranyl ion PB solution. Black: ssp spectra. Red: ppp spectra. The ssp spectra in (d) were divided into the SUP contribution (blue) and the SAM contribution (green, this fitted signal from SAM is very small). It is worth noting that no SFG signal can be detected from the SAM surface (before SUP immobilization) in the amide I signal frequency region. 55

Figure 3-3 Simulated results of surface-immobilized SUP-C105 on the maleimide-terminated SAM surface using the DMD approach. The surface immobilization site (cysteine) is labeled as a yellow ball. Protein−surface interaction energies E_{ps} for the four cases are as follows: −26.27 kcal/mol (case 1), −27.89 kcal/mol (case 2), −18.95 kcal/mol (case 3), and −16.02 kcal/mol (case 4). Note: a negative value represents an attractive interaction energy, and a positive value stands for a repulsion energy. The differences of chain−surface interaction energies are as follows: 14.23 kcal/mol for $\Delta E_{1,3} = E_{chain1} - E_{chain3}$ and 14.41 kcal/mol for $\Delta E_{2,3} = E_{chain2} - E_{chain3}$. 57

Figure 3-4 Final score map showing the matching scores between SFG experimental data and calculated spectra of various DMD structures (case 1 to case 4) using the Hamiltonian approach as a function of orientation angles of immobilized SUP-C105 before uranyl ion binding based on all five matching criteria. 60

Figure 3-5 Visualized structures with the best matching scores with experimental data using various DMD structures as inputs for the Hamiltonian approach calculation: (a) case 1 (40°,90°), (b) case 2 (55°,75°), (c) case 3 (50°, 280°), and (d) case 4 (70°, 140°) of surface-immobilized SUP-C105 before uranyl binding. 61

Figure 3-6 Final score maps showing the matching scores between SFG experimental data and calculated spectra of various DMD structures (case 1 to case 4) using the Hamiltonian approach as a function of orientation angles of immobilized SUP-C105 after uranyl ion binding based on all five matching criteria. 61

Figure 3-7 Final score map of immobilized SUP-C105 before and after binding with uranyl ions based on (a) 4FZP and (b) 4FZP. 63

Figure 3-8 Schematics of the SUP-C105 orientation: (a) surface-immobilized SUP-C105 before uranyl ion binding deduced based on the crystal structure and (b) DMD results, (c) surface-immobilized SUP-C105 after uranyl ion binding deduced based on the crystal structure and (d) DMD results. 63

Figure 4-1 Schematic of the SFG prism geometry used in this study to collect SFG spectra from the PS/protein solution interfaces.......... 73

Figure 4-2 Snapshots of initial configurations of (a) a WT GB1 and (b) a MT GB1 on a PS surface in an aqueous environment with counterions (Na^+). The PS molecules are shown in cyan; counterions are shown in orange and water molecules are shown in gray. The thickness of the simulated PS film is around 5.5 nm. According to the simulation results of PS films[22], the mobile free surface is 2 nm (less than 5.5 nm used in this study), showing that our MD simulation of PS film can be justified.......... 74

Figure 4-3 SFG ssp spectra collected from proteins (a) WT NL, (b) WT Leu, (c) MT NL, (d) MT Leu, (e) MT Val, (f) MT Phe, (g) MT Lys and (h) MT Ile absorbed at the PS/protein solution interfaces. Black dots are experimental data points and red lines are fitted spectra....... 78

Figure 4-4 SFG ppp spectra collected from proteins (a) WT NL, (b) WT Leu, (c) MT NL, (d) MT Leu, (e) MT Val, (f) MT Phe, (g) MT Lys and (h) MT Ile absorbed at the PS/protein solution interfaces. Black dots are experimental data points and red lines are fitted spectra....... 78

Figure 4-5 SFG ssp (a) and ppp (b) spectra of polystyrene/air interface (black) and polystyrene/D_2O interface (red).......... 79

Figure 4-6 Distance between the protein (a) WT GB1 or (b) MT GB1 and the PS surface as a function of time in simulation. WT GB1 and MT GB1 were released at ~ 2.0 nm from the PS surface. The adsorption processes happened quickly for both cases (~20 ns for WT GB1 and ~ 10 ns for MT GB1). After landing on the PS surface, the protein-surface distance has almost no change for each case along the entire simulation period. 80

Figure 4-7 Radius of gyration (Rg) fluctuations of (a) WT GB1 and (b) MT GB1 along the simulation time. The mean Rg value of WT GB1 is similar to the mean Rg value of MT GB1, with slightly larger variations. This indicates that the atom distribution along the principal axis of inertia of WT GB1 is similar to that of MT GB1 along the entire simulation period. This result infers that the mutation of Q32A and N35A has little to no effect on the folding of GB1. 80

Figure 4-8 Root mean square deviation (RMSD) between the protein structure at 0 ns and the protein structure from 0 ns to 1000 ns. (a) is the RMSD plot of WT GB1 and (b) is the RMSD plot of MT GB1. WT GB1 and MT GB1 possess similar mean value and variation of RMSDs, indicating that the mutation of Q32A and N35A causes little to no conformation changes due to the GB1 – PS interactions.......... 80

Figure 4-9 Secondary structure map of (a) WT GB1 and (b) MT GB1 as a function of the WT GB1 and MT GB1 have similar conformations in the entire simulation time. The major conformational difference is that for MT GB1, N8, L12 and K13 formed β sheet structure when stabilized, while for WT GB1 these three residues could sometimes be coil structure. 81

Figure 4-10 (a) θa is defined as the angle between the z axis and the sum of the amide I vectors of α-helix (from residue 21 to residue 36, pointing from near N-terminal to near C-terminal) of protein GB1. (b) and (d) are the changes of θa of the simulated WT GB1 and MT GB1, respectively, from 0 ns to 1000 ns with the increment of 1 ns. (c) and (e) are the θa distributions of last 400 ns of the simulations of WT GB1 and MT GB1, respectively.............. 82

Figure 4-11 (a) Final score map of the spectral matching between the reconstructed experimentally collected WT GB1 SFG spectra (after deconvoluting the non-resonant contribution in ssp and ppp spectra and PS contribution in the ppp spectra) and the calculated WT GB1 SFG spectra as a function of protein orientation based on the simulated WT GB1 structure at 781 ns. The orientations at (30°, 50°) and (150°, 230°) in the map possess the highest matching scores (0.66). The spectral comparisons (b) between the reconstructed WT NL SFG spectra and the calculated WT NL spectra (based on the simulated structure at 781 ns) and (c) between the reconstructed WT Leu SFG spectra and the calculated WT Leu spectra (based on the simulated structure at 781 ns) at an orientation of (30°, 50°) (or (150°, 230°)) which have the highest matching scores. The orientation visualizations of (d) WT GB1 (with the 781 ns simulation structure) at (30°, 50°) and (e) WT GB1 (with the 781 ns simulation structure) at (0°, 0°). The (0°, 0°) orientation is the protein orientation obtained from the MD simulation result without further rotating the protein. 85

Figure 4-12 (a) Final score map of spectral matching between the reconstructed experimentally collected MT GB1 SFG spectra (after deconvoluting the non-resonant contribution in ssp and ppp spectra and PS contribution in the ppp spectra) and the calculated MT GB1 SFG spectra as a function of orientation based on the simulated MT GB1 structure at 972 ns. The orientations at (30°, 100°) and (150°, 280°) shown in the map possess the highest matching score (0.19). The spectral comparisons between the reconstructed experimental spectra and calculated spectra using the simulated MT structure at 972 ns with an orientation of (30°, 100°) (or (150°, 280°)) for (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. The orientation visualizations of (h) MT GB1 with a simulated structure at 972 ns with the most likely orientation at (30°, 100°) and (i) MT GB1 with a simulated structure at 972 ns without rotation at (0°, 0°). 87

Figure 4-13 (a) Final score heat map of spectral matching between the reconstructed WT SFG spectra and the calculated WT SFG spectra by rotating the crystal structure 3gb1. The orientations at (55°, 325°) and (125°, 145°) possess highest matching scores (score = 0.56). The spectral comparisons between the reconstructed experimental spectra and the calculated spectra using 3gb1 at orientation of (55°, 325°) (or (125°, 145°)) of (b) WT NL and (c) WT Leu. By using the crystal structure, the best matched calculated ppp spectra could not match the reconstructed experimental ppp spectra well, leading to lower matching scores. 88

Figure 4-14 (a) Final score heat map between the reconstructed experimental MT SFG spectra and the calculated MT SFG spectra using crystal structure 3gb1. The orientations at (60°, 0°) and (120°, 180°) possess highest matching scores (score = 0.15). The spectral comparisons between reconstructed experimental spectra and calculated spectra by rotating 3gb1 at orientation of (60°, 0°) (or (120°, 180°)) of the case of (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. By using the crystal structure, the calculated spectra and the experimental reconstructed spectra could not match well. For example, almost all the ppp spectra except MT Leu could not match well in peak intensity, peak center, etc., leading to lower matching scores.. 89

Figure 4-15 Orientation comparisons of GB1 at the PS/protein solution interfaces: (a) and (b) the best matched orientations of WT GB1 based on the SFG data analysis using the GB1 crystal structure, (c) the best matched WT GB1 orientation after rotating all the simulated structures (based on the simulated structure at 781 ns) – replotted from Fig. 4-11 (d). (d) and (e) The best matched orientations of MT GB1 based on the SFG data analysis using the GB1 crystal structure, (f) the best matched MT GB1 orientation after rotating all the simulated structures (based on the simulated structure at 972 ns) – replotted from Fig. 4-12 (h)............... 90

Figure 4-16 Orientation visualizations of (a) the simulated WT GB1 structure at 781 ns of (30°, 50°) and (b) the simulated MT GB1 structure at 972 ns of (30°, 100°). Residues Q32 and N35 are shown in magenta sticks. Residues F30, Y33 and N37 are shown in red sticks.................... 92

Figure 4-17 Visualization of Contact Areas (in dark blue) of (a) WT GB1 and (b) MT GB1 on PS. The contact area of WT GB1 on PS is obviously larger than the contact area of MT GB1 on PS. 93

Figure 5-1 SFG (a, c, e, g) and ATR-FTIR (b, d, f, h) spectra of (a, b) MSI-594 associated with a POPC lipid bilayer; (c, d) MSI-594A associated with a POPC lipid bilayer; (e, f) MSI-594 associated with a 7:3 POPC/POPG lipid bilayer; and (g, h) MSI-594A associated with a 7:3 POPC/ POPG lipid bilayer. SFG ppp and ssp spectra are shown in open dots and filled black dots, respectively (top row). ATR-FTIR p and s spectra are shown in circles and crosses, respectively (bottom row). The solid black lines are fitted spectra. 103

Figure 5-2 Modified structures of (a) MSI-594 based on PDB ID: 2K98 and (b) MSI-594A based on PDB ID: 2L36. Residues 12−24 were colored in cyan for all of the structures. Residues 1−11 were colored in red for the original NMR structures, in blue for the structures with a bent angle change of −5° from the original NMR structures, in light purple for the structures with a bent angle change of −10° from the original NMR structures, in orange for the structures with a bent angle change of +5° from the original NMR structures, and in yellow for the structures with a bent angle change of +10° from the original NMR structures. 105

Figure 5-3 Final score maps that quantify the degree of similarity between the reconstructed spectra (SFG and ATR-FTIR) of MSI-594 associated with a POPC lipid bilayer and the corresponding calculated spectra by using the input structure as (a) NMR structure (PDB ID: 2K98), (b) modified NMR structure with −5° bent angle change, (c) modified NMR structure with +5° bent angle change, (d) modified NMR structure with −10° bent angle change, and (e) modified NMR structure with +10° bent angle change. 105

Figure 5-4 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and ATR-FTIR (bottom)) of MSI-594 associated with a POPC lipid bilayer and calculated spectra at the best-matched orientations by using input structures as (a, b) the NMR structure (PDB ID: 2K98); (c, d) the modified NMR structure with −5° bent angle change; (e, f) the modified NMR structure with +5° bent angle change; (g, h) the modified NMR structure with −10° bent angle change; and (i, j) the modified NMR structure with +10° bent angle change.. 105

Figure 5-5 Membrane orientations deduced based on the modified NMR structure of MSI-594 (PDB ID: 2K98) with +5° bent angle change associated with the POPC or 7:3 POPC/POPG lipid bilayer with the highest matching score (score = 293 of the case of POPC and score =311 of the case of 7:3 POPC/POPG). The orientations of N- and C-termini segments are shown in deep orange (residues 1− 11) and cyan (residues 12−24), respectively, with respect to the lipid bilayer surface (i.e., the xy plane). 107

Figure 5-6 The membrane orientations deduced based on the modified NMR structure of MSI-594 (PDB ID: 2K98) with +10° bent angle change associated with POPC or 7:3 POPC:POPG lipid bilayer with a high matching score (score >= 264 for the case of POPC and score >= 280 for the case of 7:3 POPC:POPG). The orientations of N- and C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane). 107

Figure 5-7 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594A associated with a POPC lipid bilayer and the corresponding calculated spectra by using the input structure as (a) the NMR structure (PDB ID: 2L36); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change. 108

Figure 5-8 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594A associated with a POPC lipid bilayer and the calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2L36); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g, h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change. 109

Figure 5-9 Best matched membrane orientations of the modified NMR structure of MSI-594A (PDB ID: 2L36) with +10° bent angle change associated with POPC lipid bilayer with highest matching score (score=425). The orientations of N- and C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane). 109

Figure 5-10 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594 associated with a 7:3 POPC:POPG lipid bilayer and the corresponding calculated spectra by using the input structure as (a) the NMR structure (PDB ID: 2K98); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change. 110

Figure 5-11 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594 associated with a 7:3 POPC:POPG lipid bilayer and the calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2K98); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g,

h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change. 111

Figure 5-12 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594A associated with a 7:3 POPC:POPG lipid bilayer and the corresponding calculated spectra by using the input structure as (a) the NMR structure (PDB ID: 2L36); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change. 111

Figure 5-13 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594A associated with a 7:3 POPC:POPG lipid bilayer and calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2L36); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g, h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change. 112

Figure 5-14 Best matched membrane orientations of the original NMR structure of MSI-594A (PDB ID: 2L36) in 7:3 POPC:POPG lipid bilayer with highest matching score (score=387). The orientations of N- and C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane). 112

Figure 5-15 Possible configurations of (a) MSI-594 associated with POPC or 7:3 POPC:POPG lipid bilayer; (b) MSI-594A associated with POPC lipid bilayer and (c) MSI-594A associated with 7:3 POPC:POPG lipid bilayer. Grey solid spheres represent the lipid headgroups, grey curved lines represent the lipid hydrophobic tails, and the positively charged residue Lys is colored in purple sticks. 115

Figure 5-16 (a) Front view (left) and top view (right) of a “complete lying down” (top) representation and a “complete standing up” (bottom) representation of MSI-594-a and (b) front view (left) and top view (right) of a “complete lying down” (top) representation and a “complete standing up” (bottom) representation of MSI-594-b. The membrane surface is parallel to the xy plane. 115

Figure 5-17 The peptides shown in Figure 5-6 are displayed along with the POPC or 7:3 POPC:POPG lipid bilayer. Grey solid spheres represent the lipid headgroups, grey curved lines represent the lipid hydrophobic tails, and the positively charged Lys residues are colored in purple sticks. 116

Figure 5-18 (a) MSI-594A-b1 associated with a POPC bilayer without insertion and (b) the MSI-594A-b2 associated with a 7:3 POPC:POPG bilayer without insertion. 118

Abstract

Biomolecules such as proteins/peptides at interfaces play important roles in many applications ranging from biomedical implants, antifouling coatings, biosensors/biochips, to antibody drugs. Understanding the interfacial structures of proteins/peptides is the crucial first step towards unraveling their interfacial functions. For the first time, this dissertation developed a systematic method to determine molecular structures of proteins/peptides at buried solid/liquid interfaces *in situ* using a nonlinear optical spectroscopic technique, sum frequency generation (SFG) vibrational spectroscopy, alongside linear spectroscopic methods, computer simulation, isotope labeling, and Hamiltonian spectral calculation.

Before this systematic methodology was developed, the data analysis of the collected vibrational spectra was limited to specific secondary structures contained by the protein/peptide sample. This was illustrated in this thesis research by a study of surface immobilized antimicrobial peptides (AMPs) for bacteria capture. Using a helical model to approximate the AMPs' interfacial conformation, the bond additivity model correlated their vibrational responses to their interfacial tilt angles while interacting with different model bacteria systems. It was found that stronger electrostatic interactions between the peptide residues and the bacteria membrane components led to smaller tilt angles of the AMPs. Therefore, the measured tilt angles of the AMPs can potentially be used to classify bacterial membrane components.

With the newly developed systematic approach, detailed conformation and orientation information of a variety of proteins/peptides (regardless of their secondary structures) can be determined. For this systematic method, vibrational spectroscopies provide experimental data for studying interfacial behaviors of proteins/peptides. The Hamiltonian approach considers the couplings between all the amino acids within a protein/peptide (regardless of secondary structures) and calculates its amide I signals as a function of interfacial orientation. The calculated spectra can then be compared to the results obtained from experimental measurements using the developed score systems. Thus, the possible interfacial orientations of a protein/peptide for a specific conformation can be deduced. The input structures for the Hamiltonian approach can be obtained from published crystal structures or can be further refined by molecular dynamic (MD) simulations to better estimate the interfacial conformation of a protein/peptide. This thesis also discussed approaches to improve this systematic method. For example, employing isotope labeling and/or combining various spectroscopic techniques can increase independently measured parameters for more accurately determining interfacial peptide/protein structure.

The developed methodology was applied to study surface immobilized proteins for uranyl ion binding. The results suggested that this protein lies down upon immobilizing onto a hydrophobic surface and thus proposed that a hydrophilic surface may be a better candidate to alter protein interfacial orientation for effective uranyl harvesting. This systematic method was also employed to study wild-type protein GB1 (WT) and its mutant (MT) upon adsorption onto a polystyrene (PS) surface. The results showed that MT oriented differently from the WT on the PS surface, which indicated that mutation on some key amino acids can manipulate the interfacial orientation of protein. This systematic method was also used to study AMPs interacting with model cell membranes and the deduced peptide structures were successfully correlated to their membrane disrupting mechanisms. The results showed that mutation-induced peptide structure change can further modulate the membrane translocation efficiency of the peptide.

Overall, the developed systematic method is powerful and can be widely applicable to study detailed structures and molecular interactions of peptides/proteins at interfaces, advancing our insights into their potential interfacial functions.

Chapter 1 Introduction

1.1 Attribution Information

Detailed calculations about the bond additivity model covered in Section 1.4.1 were adapted with permission from Guo, W.; Xu, S.; Reichart, T. M.; Xiao, M.; Lu, T.; Mello, C.; Chen, Z. Probing Molecular Interactions between Surface-Immobilized Antimicrobial Peptides and Lipopolysaccharides *In Situ*. *Langmuir* **2020**, *36* (41), 12383–12393. Copyright © 2020 American Chemical Society.

Detailed Hamiltonian calculations (Section 1.4.2) for SFG spectra were adapted with permission from Guo, W.; Zou, X.; Jiang, H.; Koebke, K. J.; Hoarau, M.; Crisci, R.; Lu, T.; Wei, T.; Marsh, E. N. G.; Chen, Z. Molecular Structure of the Surface-Immobilized Super Uranyl Binding Protein. *J. Phys. Chem. B* **2021**, *125* (28), 7706–7716. Copyright © 2021 American Chemical Society.

Detailed Hamiltonian calculations (Section 1.4.2) for ATR-FTIR spectra were adapted with permission from Yang, P.; Guo, W.; Ramamoorthy, A.; Chen, Z. Conformation and Orientation of Antimicrobial Peptides MSI-594 and MSI-594A in a Lipid Membrane. *Langmuir* **2023**, *39* (15), 5352–5363. Copyright © 2023 American Chemical Society.

1.2 Motivation and Background

Proteins and peptides have long been a central topic in both academia and industry due to their important biological functions, such as molecule recognition[1,2], signal transduction[3,4], antimicrobial acitivity,[5,6] and more. The versatile functions of proteins and peptides are determined by their structures, which can be changed by the surrounding environments. Therefore, to understand the unique function of a protein or a peptide in a defined environment, efforts are needed to elucidate the structure of the protein/peptide in such an environment. Most commonly used analytical techniques to study the structures of proteins and peptides are X-ray crystallography[7,8], nuclear magnetic resonance (NMR) spectroscopy[9–11] and cryogenic electron microscopy (cryo-EM)[12]. These techniques are powerful to elucidate the structures of proteins and peptides in solid states or solution states with atomistic-level resolution. However, the required sample preparation conditions of these techniques can be far from the “real situations” (or physiological conditions) of the proteins and peptides in interest. For example, the conformations and tertiary structures of the crystallized proteins or peptides can be very different from those of their solvated form. Such a structural difference can be more dramatic for small peptides which have higher structural flexibility. Additionally, due to the complication of the sample preparation, these techniques also often fail to monitor the dynamics and kinetics of structural changes of proteins and peptides *in situ*.

More specifically, in many real-world applications, it is more important to observe the structure changes of proteins and peptides associated with other materials. For example, proteins and peptides are widely employed in the field of bio-sensing[13–17] because of their ability of bio-recognition. For such applications, proteins and peptides are often immobilized onto a solid support. The immobilization process and the substrate interactions of the proteins/peptides could alter the structures of immobilized proteins and peptides from their original solution structures, leading to the destruction, reduction, or enhancement of the original functions of these proteins and peptides. Therefore, it is crucial to understand the structures of the surface-immobilized proteins and peptides to facilitate further design of biosensors. Another example comes from the physical adsorption of proteins and peptides onto a surface. Adsorption of proteins and peptides happened in a wide range of scenarios, such as implantation of biomaterials[18], marine fouling formation[19–21] and storage of proteins and peptides[22,23]. To further understand the behavior changes of the surface-adsorbed proteins and peptides, the first step is to monitor the structural changes of such proteins and peptides on surfaces and at interfaces *in situ*. The abovementioned techniques often fail to probe interfacial proteins and peptides such as surface-immobilized and physically adsorbed proteins and peptides. Other analytical techniques, such as attenuated total reflectance infrared spectroscopy (ATR-FTIR)[24], surface-enhanced Raman spectroscopy (SERS)[25] and surface plasmon resonance (SPR)[26,27] are often chosen to be alternative techniques to study interfacial proteins and peptides. However, surfaces and interfaces only involve monolayer or sub-monolayer materials; these techniques could not provide sufficient sensitivity or detailed structural information, or require complicated sample preparation procedures.

In this thesis, efforts have been made to address the abovementioned challenges regarding the examination of interfacial structures of proteins and peptides. Spectroscopic methods were applied to experimentally observe the interfacial proteins and peptides and computational methods were also developed to facilitate the experimental data interpretation. To achieve molecular sensitivity, noninvasive and real time investigation of interfacial proteins and peptides, a non-linear optical technique named sum frequency vibrational generation (SFG) vibrational spectroscopy was adopted as the major tool to study interfacial proteins and peptides in this thesis research. SFG is a second order nonlinear optical process, which probes the surface susceptibility $\chi^{(2)}$. Similar to other vibrational spectroscopic methods, such as infrared (IR) spectroscopy and Raman spectroscopy, SFG can also provide vibrational information of proteins and peptides (e.g.

amide I band ranging from 1600 cm^{-1} to 1690 cm^{-1}) with noninvasive nature, but with higher surface sensitivity (sub-monolayer level). Other advantages of SFG spectroscopy include the absence of water background (compared to IR spectroscopy) and simpler sample preparation method (compared to SERS).

In the past 30 years, SFG has been widely used to study interfacial proteins and peptides such as surface immobilized antimicrobial peptides (AMPs)[28–30], adsorption and aggregation of proteins and peptides on polymer surfaces[23,31–33], antifouling coatings[34–36], and more. More recently, isotope labeling methods have been introduced into SFG studies to increase the independent experimental measurements.[37] For example, isotopic substitution of ^{13}C in the backbone of proteins could shift the amide I band by ~ 40 cm^{-1}.[37,38] Such a frequency shift induced by isotope labeling distinguishes originally overlapped signals and therefore generates more independent spectra measurements. Other spectroscopic techniques, such as ATR-FTIR spectroscopy and circular dichroism (CD) spectroscopy, are often used to supplement SFG measurements for more experimental insights.[39–41]

A systematic method was developed to correlate SFG experimental results to conformations and orientations of interfacial proteins and peptides. To find the most possible orientation of a given protein or peptide at an interface, our previous strategy was to construct a theoretical map of the measured $\chi^{(2)}$ of the amide I signals from all the α-helical domains in the peptide/protein as a function of protein/peptide orientation (defined by a tilt angle and a twist angle) and then match the calculated map with the experimentally deduced $\chi^{(2)}$ (of the α-helical domains) to generate a heat map. Bond additivity model was used to calculate the $\chi^{(2)}$ map.[28,39,40,42] This strategy is simple in calculation and works well for simple α-helical peptides and proteins with extensive α -helical structures. Later, an improved strategy for more complicated protein structures has been developed, which considers the signal contributions from all the peptide units, not only the α-helices. Here, it is necessary to simulate the entire SFG spectra contributed from all the peptide units as a function of orientation of the target protein or peptide. A scoring system can be used to quantitatively compare the similarity between the calculated SFG spectra with the experimental SFG data and select the best matchings as the most likely orientations of the target.[23,43,44] Hamiltonian approach is one way to construct the SFG spectra, which addresses vibrational mode couplings between amide units better than the bond additivity model.[37,45,46]

Amino acids are the basic building blocks of proteins and peptides, and amide units are the functional units that link amino acids together in proteins and peptides. In vibrational spectroscopy, amide units coupled with each other and show different vibrational bands, such as amide I, amide II, amide III and amide A. This thesis focuses on the analysis of the amide I band, which is the most intense absorption band in proteins and peptides in infrared spectroscopy.[47] Amide I mode primarily contains the stretching vibrational modes of C=O (70-85%) and C-N (10-20%) bonds, presenting vibrational signatures between 1600 cm^{-1} and 1690 cm^{-1}.[48,49] An isolated amide unit possesses a vibrational frequency, noted as local frequency, of around 1640 cm^{-1}.[37] Mode coupling of the amide I local modes give rise to the observed vibrational peaks of proteins and peptides, noted as normal modes, which have different frequencies from local modes. Bond additivity model simplifies normal modes for α-helical structure to two SFG active modes, with one (A mode) parallel and the other (E_1) perpendicular to the principal axis of the helix.[50] One-exciton Hamiltonian approach is another way to better interpret the energy rearrangement from local modes to normal modes.[37,45,46] After obtaining normal modes, SFG spectra and ATR-FTIR spectra can be constructed theoretically under the defined orientation. Detailed theories of the bond additivity model and Hamiltonian approach will be introduced in later sections.

It is worth noting that, both bond additivity model and Hamiltonian approach require a given three-dimensional protein/peptide structure as input. The initial protein/peptide structure could be obtained from X-ray (crystal structure) and NMR (solution structure) experiments or molecular dynamic (MD) simulations (simulated interfacial structure). These structures/conformations could also be scored to search for the best candidate(s). If detailed three-dimensional structure is not available, initial structure used for SFG spectra calculation can be predicted from protein/peptide sequence, which will be discussed in more detail below.

This thesis aims to demonstrate that the combination of spectroscopic experimental measurements (SFG, ATR-FTIR and CD measurements) and computational data analysis (bond additivity calculations, Hamiltonian calculations and MD simulations) is powerful to study a wide range of interfacial proteins and peptides, improving the design and development of materials and devices in many applications involving interfacial peptides and proteins.

1.3 SFG Spectroscopy

1.3.1 Basic Principals

SFG is a second order nonlinear optical process where two photons (with different frequencies) are combined to generate one output photon which sums the frequencies of the two input photons. In our studies, one input beam is a visible beam with a fixed frequency at 532 nm, and the other input beam is a frequency tunable mid-IR beam which can be tuned from 1100 cm^{-1} to 4300 cm^{-1}. The energy diagram of a SFG process is shown in Figure 1-1 (a). The selection rule indicates that SFG can only happen where the inversion symmetry is broken, such as surfaces and interfaces. Most bulk media are centrosymmetric, thus cannot generate SFG signals. This allows

SFG to be an intrinsic technique to selectively study vibrational spectroscopic features of interfacial molecules.

The signal intensity of SFG (I_{SFG}) can be expressed as:[51]

$$I_{SFG} = I_{VIS} I_{IR} \left|\chi_{eff}^{(2)}\right|^2 \qquad (1 - 1)$$

where I_{VIS} and I_{IR} are the intensities of the input visible and IR beams, respectively. $\chi_{eff}^{(2)}$ is the effective second order nonlinear optical susceptibility of the probed surface/interface, which contains the non-resonant $\chi_{NR}^{(2)}$ contribution and the resonant contribution: [51]

$$\chi_{eff}^{(2)} = \chi_{NR}^{(2)} + \sum_q \frac{A_q}{\omega_{IR} - \omega_q + i\Gamma_q} \qquad (1 - 2)$$

where the resonant contribution is the sum of Lorentzian peaks with the peak amplitude of A_q, the peak center of ω_q, and the peak width of $i\Gamma_q$ for a peak q. The peak center ω_q should equal the ω_{IR} as a resonant state, where the probed molecular vibrational frequency is in resonant with the input IR frequency.

Different $\chi_{eff}^{(2)}$ components can be probed using different polarization combinations of the input and output beams in the SFG experiment. For example, $\chi_{eff,ssp}^{(2)}$ is measured when the SFG signal beam is s-polarized, the input visible beam is s-polarized, and the input IR beam is p-polarized. $\chi_{eff}^{(2)}$ observed in the Jones frame can be converted to $\chi_{IJK}^{(2)}$ (I, J, K = x, y, z) in Cartesian coordinate of a fixed lab frame, expressed as:[52]

$$\chi_{eff,ssp}^{(2)} = F_{ssp/yyz}\chi_{yyz}^{(2)} \qquad (1 - 3)$$

$$\chi_{eff,ppp}^{(2)} = F_{ppp/zzz}\chi_{zzz}^{(2)} + F_{ppp/xxz}\chi_{xxz}^{(2)} + F_{ppp/xzx}\chi_{xzx}^{(2)} + F_{ppp/zxx}\chi_{zxx}^{(2)} \qquad (1 - 4)$$

where $F_{Jones/Cartesian}$ represents the following:[52]

$$F_{ssp/yyz} = L_{yy}(\omega_{SUM})L_{yy}(\omega_{VIS})L_{zz}(\omega_{IR}) \sin\delta_2 \qquad (1 - 5)$$

$$F_{ppp/zzz} = L_{zz}(\omega_{SUM})L_{zz}(\omega_{VIS})L_{zz}(\omega_{IR}) \sin\delta \sin\delta_1 \sin\delta_2 \qquad (1 - 6)$$

$$F_{ppp/xxz} = -L_{xx}(\omega_{SUM})L_{xx}(\omega_{VIS})L_{zz}(\omega_{IR})\cos\delta\cos\delta_1\sin\delta_2 \quad (1-7)$$

$$F_{ppp/xzx} = -L_{xx}(\omega_{SUM})L_{zz}(\omega_{VIS})L_{xx}(\omega_{IR})\cos\delta\sin\delta_1\cos\delta_2 \quad (1\text{ - }8)$$

$$F_{ppp/zxx} = L_{zz}(\omega_{SUM})L_{xx}(\omega_{VIS})L_{xx}(\omega_{IR})\sin\delta\cos\delta_1\cos\delta_2 \quad (1\text{ - }9)$$

where L_{ii} (i = x, y, or z) are the Fresnel coefficients, and δ is the incident angle between the surface normal and the incoming light of visible beam or IR beam.

The deduced susceptibility $\chi_{IJK}^{(2)}$ of the surface/interface is the macro-average of the hyperpolarizability $\beta_{xyz}^{(2)}$ (x, y, z = a, b, c) in the molecular frame (a, b, c) of interfacial molecules, expressed as:[46]

$$\chi_{IJK}^{(2)} = N\, f(\theta,\psi,\varphi)\, \beta_{xyz}^{(2)} \quad (1\text{ - }10)$$

where N is the molecular surface density, and $f(\theta,\psi,\varphi)$ is the Euler rotation matrix to convert the optical response in the molecular frame to the lab frame.

The hyperpolarizability $\beta_{xyz}^{(2)}$ is the product of IR transition dipole μ_z and the Raman polarizability α_{xy}:[37,46]

$$\beta_{xyz}^{(2)} = \alpha_{xy} \otimes \mu_z \quad (1\text{ - }11)$$

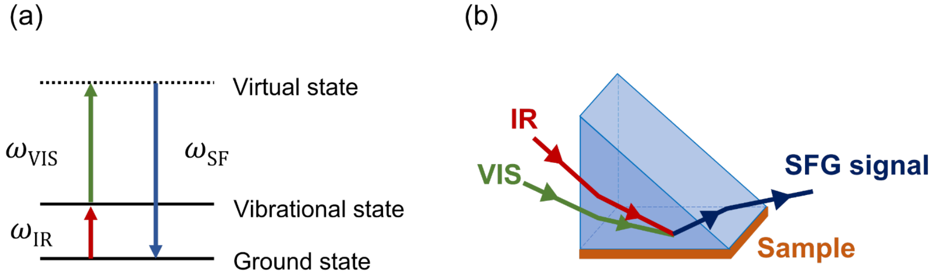

Figure 1-1 (a) SFG energy diagram. (b) SFG near-total reflection sample geometry.

1.3.2 SFG Experimental Setup

A picosecond SFG system was applied for the research covered in this thesis. This SFG system (EKSPLA, Lithuania) has a Nd:YAG laser source to generate laser pulses at 1064 nm with a 50 Hz repetition rate. Part of this fundamental 1064 nm laser beam was further used to generate a fixed frequency visible beam at 532 nm (pulse width ~ 30 ps) via a second harmonic generation (SHG, also called frequency doubling) process. Part of the generated 532 nm visible beam was directly used in SFG experiments. Rest of the fundamental 1064 nm beam and the rest of the 532 nm visible beam were further used to generate a frequency-tunable IR beam via optical parametric generation/amplification (OPG/OPA) and difference frequency generation (DFG) process. Near total internal reflection geometry (as shown in Figure 1-1 (b)) of SFG experiments was used throughout the studies covered by this thesis. The incident angles of input visible beam and IR beam are 60° and 55° with respect to the surface normal, respectively. Two input beams overlapped temporarily and spatially at the bottom of the right-angle prism to generate the SFG signal beam of the interfacial molecules. The overlapped beam size was around 500 μm in diameter. The energies of the input visible beam and the input IR beam were ~ 40 μJ and ~ 60 μJ in the amide I region, respectively. The SFG signal beam was reflected and then collected by a monochromator and photomultiplier tube (PMT) as a function of the input IR frequency.

SFG amide I signals from proteins and peptides are in the wavenumber range of 1600 cm^{-1} – 1690 cm^{-1}. Different secondary structures result in different amide I peak centers. Usually, a peak at ~ 1650 cm^{-1} is contributed by α-helical structure and a peak at ~1630 cm^{-1} can be assigned to β-sheet structure.[48,49]

The dynamics of a vibrational peak can be obtained by SFG time dependent spectra. The intensity trace of SFG signal at a certain frequency can be monitored over time. A normal SFG spectrum (SFG intensity is recorded as a function of IR frequency) should be taken after the system is equilibrated. The equilibrium usually means a stable signal plateau observed in the SFG time dependent spectra.

1.4 Supplementary Techniques

1.4.1 ATR-FTIR Spectroscopy

ATR-FTIR spectroscopy is a linear optical spectroscopic method commonly used to study a wide range of materials, including biological samples and polymers. Similar to SFG, ATR-FTIR provides vibrational spectroscopic information about functional groups in the sample. In an ATR-FTIR experiment, a sample is placed against a material with a high refractive index, such as a germanium crystal or a zinc selenide crystal, to achieve total internal reflection at the sample/crystal interface.[53] The multi-bounce reflection ATR-FTIR setup (shown in Figure 1-2) was used throughout the research carried out in this thesis, to achieve greater signal to noise ratio. ATR-FTIR can probe the sample surface down to the magnitude of μm[53] and functions as a less surface-sensitive technique compared to SFG. Different polarizations (p- or s-polarization) can also be used in ATR-FTIR measurements to deduce orientation information of the detected proteins and peptides. Bond additivity model and Hamiltonian approach mentioned above can also be used to correlate the ATR-FTIR measurements to the orientation of interfacial proteins and peptides. Different from SFG measurements, ATR-FTIR measurements suffer from the influence of water O-H bending mode in the amide I region. To avoid such signal confusion, D_2O is often used to make protein/peptide solution for ATR-FTIR study. Nevertheless, when the interface is thin (e.g., only a monolayer or sub–monolayer of protein/peptide) and the background contribution from bulk solution is negligible, ATR-FTIR can serve as a supplementary technique to SFG spectroscopy to more accurately determine the orientations of interfacial proteins and peptides.

The actual orientation of the studied proteins and peptides on surfaces or at interfaces should be in agreement with both SFG and ATR-FTIR measurements.

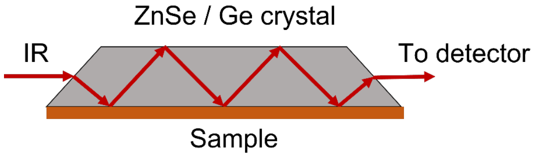

Figure 1-2 Schematic of multi-reflection ATR-FTIR experimental setup.

1.4.2 CD Spectroscopy

CD spectroscopy measures the optical chiral properties of a molecule (e.g. amino acid, the building block for peptides/proteins) in solution. The distinct three-dimensional arrangements of different secondary structures within a protein or a peptide contribute uniquely to a CD spectrum. For example, an α-helix displays a positive band at around 193 nm and two negative bands at around 208 nm and 222 nm respectively in a CD spectrum, while an anti-parallel β-sheet structure has a positive band at around 195 nm and a negative band at around 218 nm.[54] Measurements and analysis of CD spectra is much faster than X-ray and NMR experiments, therefore, CD often becomes an alternative way to examine the secondary structures of proteins and peptides in solution state. However, it is challenging to use CD to probe interfacial proteins and peptides because of the low signal to noise ratio generated from only a thin layer of interfacial molecules and the potential light scattering effects at far UV range.[55] To address these challenges, UV-grade quartz slides were used to chemically immobilize or physically absorb proteins/peptides to reduce light scattering and multiple slides were stacked for the CD testing to increase signal to noise ratio (as shown in the Figure 1-3).[55,56] In this thesis, CD spectroscopy was adopted as a supplemental technique to examine the secondary structures of surface-immobilized proteins and peptides, and the structure difference between the free state proteins/peptides in solution and the interfacial proteins/peptide.

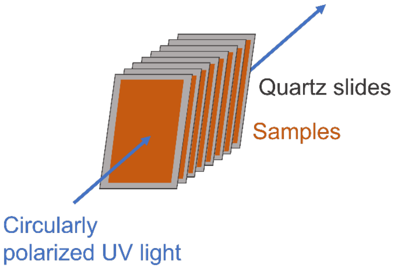

Figure 1-3 Schematic of CD experimental setup to study interfacial proteins and peptides.

1.4.3 Dynamic Light Scattering

Dynamic light scattering (DLS) is widely used to characterize the size and size distribution of microscopic particles, such as lipid vesicles, suspended in a liquid fluid. DLS measures the fluctuations in light intensity caused by the random Brownian motion of particles in solution. By analyzing the correlation between these fluctuations, the size distribution of the particles can be determined. DLS can be used to study a wide range of particle sizes, ranging from sub-nanometer to micrometer.[57] In this thesis, DLS was used to measure the size distribution of membrane-like vesicles to facilitate the understanding of the interaction between such vesicles and the studied peptides and proteins.

1.5 Molecular Orientation Determination of Proteins and Peptides

1.5.1 Bond Additivity Model for α-helical Segments

Proteins and peptides are biomolecules consisting of amino acids linked by amide bonds. Amide units can have vibrational modes in different ranges, such as amide I band (1600 cm^{-1} –

1690 cm^{-1}), amide II band (1470 cm^{-1} – 1570 cm^{-1}), amide III band (1250 cm^{-1} – 1350 cm^{-1}) and amide A band (3300 cm^{-1} – 3500 cm^{-1}).[48,49] Amide I vibrational modes can be SFG active and each such a mode usually consists of ~80% of C=O stretching and ~20% of C-N stretching.[48,49] For an α-helical structure, the amide I vibration of each amide unit can be projected into three modes: A mode (parallel to the principal axis of the α-helix), E_1 and E_2 modes (perpendicular to the principal axis of the α-helix).[50] Among all the three modes, only A mode and E_1 mode are SFG active, because E_2 mode is IR inactive - A vibrational mode is SFG active when this mode is both IR and Raman active. Thus, the bond additivity model of α-helix needs to consider the overall SFG response of both A mode and E_1 mode for all the amide units in the entire helical chain.[50] The amide I peak centers of the A and E1 modes of a helix are similar (at ~1650 cm^{-1}), so the SFG signal can be calculated as the sum of the contributions of the two modes. For a protein with multiple helical chains, the overall SFG response is the sum of the contributions from the A mode and E_1 mode for all the helical components.[40]

In the molecular frame (abc) of an α-helix structure, the c axis can be defined as the principal axis of the helix. As a result, the twist angle of the helix can be averaged out due to the structural symmetry along the c axis. The tilt angle (θ) can be defined as the angle between the c axis and the z axis in the laboratory frame (xyz). Since a surface or interfaces is isotropic, thus it is reasonable to believe that the interfacial proteins/peptides have the azimuthal symmetry - The azimuthal angle can also be averaged out. As a result, for a simple structure like α-helix, the SFG response can be expressed as a function of tilt angle, surface coverage and hyperpolarizability.

For an α-helix, the relationship between $\chi^{(2)}$ and A mode is as follows:[50]

$$\chi^{(2)}_{xxz,A} = \chi^{(2)}_{yyz,A} = \frac{1}{2}N[(1+r) < cos\theta > -(1-r) < cos^3\theta >]\beta_{ccc} \qquad (1 - 12)$$

$$\chi^{(2)}_{zzz,A} = N[r < cos\theta > +(1-r) < cos^3\theta >]\beta_{ccc} \qquad (1 - 13)$$

where $\beta_{a,b,c}$ is a tensor component of the hyperpolarizability of the amide I of the surface molecule in the molecular frame (abc), and $r = \beta_{aac}/\beta_{ccc}$.

The relationship between $\chi^{(2)}$ and E_1 mode is as follows:[50]

$$\chi^{(2)}_{xxz,\mathrm{E1}} = \chi^{(2)}_{yyz,\mathrm{E1}} = -N[< cos\theta > - < cos^3\theta >]\beta_{aca} \tag{1 - 14}$$

$$\chi^{(2)}_{zzz,\mathrm{E1}} = 2N[< cos\theta > - < cos^3\theta >]\beta_{aca} \tag{1 - 15}$$

The values of related Raman tensor and IR transition dipole moment components for both A and E1 modes are needed to calculate the values of β_{aac}, β_{aca} *and* β_{ccc}. For a linear helix model of n residues, the amide I Raman tensor can be calculated as:[50,58,59]

For A mode:

$$\left[\frac{\partial\alpha}{\partial Q_A}\right](0^\circ) = \sum_{n=0}^{n}\left[R\begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix}^T \cdot \begin{pmatrix} 0.624 & 0 & 0.4 \\ 0 & 0.05 & 0 \\ 0.4 & 0 & 0.577 \end{pmatrix} \cdot \begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix} R^T\right] \tag{1 - 16}$$

For E_1 mode:

$$\left[\frac{\partial\alpha}{\partial Q_{\mathrm{E1}}}\right](100^\circ) = \sum_{n=0}^{n}\left[R\begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix}^T \cdot \begin{pmatrix} 0.624 & 0 & 0.4 \\ 0 & 0.05 & 0 \\ 0.4 & 0 & 0.577 \end{pmatrix} \cdot \begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix} R^T \cdot e^{i\cdot 100\cdot n\cdot\pi/180}\right] \tag{1 - 17}$$

$$\left[\frac{\partial\alpha}{\partial Q_{\mathrm{E1}}}\right](-100^\circ) = \sum_{n=0}^{n}\left[R^T\begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix}^T \cdot \begin{pmatrix} 0.624 & 0 & 0.4 \\ 0 & 0.05 & 0 \\ 0.4 & 0 & 0.577 \end{pmatrix} \cdot \begin{pmatrix} \cos(\pi\cdot 100\cdot\frac{n}{180}) & \sin(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ -\sin(\pi\cdot 100\cdot\frac{n}{180}) & \cos(\pi\cdot 100\cdot\frac{n}{180}) & 0 \\ 0 & 0 & 1 \end{pmatrix} R^T \cdot e^{-i\cdot 100\cdot n\cdot\pi/180}\right] \tag{1 - 18}$$

Note that the E_1 mode consists of two degenerate vibrations, one corresponds to the phase difference of 100° and the other corresponds to the phase difference of -100° between the adjacent peptide units.

The IR transition dipole moment of a linear helical peptide can be calculated as:[50,58,59]

For A mode:

$$\left[\frac{\partial\mu}{\partial Q_A}\right](0^{\circ}) = \sum_{n=0}^{n}\left[R\begin{pmatrix}\cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ -\sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ 0 & 0 & 1\end{pmatrix}^{T}\cdot\begin{pmatrix}\sin\left(\frac{42\pi}{180}\right)\\ 0\\ \cos\left(\frac{42\pi}{180}\right)\end{pmatrix}\right] \qquad (1 - 19)$$

For E1 mode:

$$\left[\frac{\partial\mu}{\partial Q_{\mathrm{E1}}}\right](100^{\circ}) = \sum_{n=0}^{n}\left[R\begin{pmatrix}\cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ -\sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ 0 & 0 & 1\end{pmatrix}^{T}\cdot\begin{pmatrix}\sin\left(\frac{42\pi}{180}\right)\\ 0\\ \cos\left(\frac{42\pi}{180}\right)\end{pmatrix}\cdot e^{i\cdot 100\cdot n\cdot\pi/180}\right] \qquad (1 - 20)$$

$$\left[\frac{\partial\mu}{\partial Q_{\mathrm{E1}}}\right](-100^{\circ}) = \sum_{n=0}^{n}\left[R\begin{pmatrix}\cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ -\sin\left(\pi\cdot 100\cdot\frac{n}{180}\right) & \cos\left(\pi\cdot 100\cdot\frac{n}{180}\right) & 0\\ 0 & 0 & 1\end{pmatrix}^{T}\cdot\begin{pmatrix}\sin\left(\frac{42\pi}{180}\right)\\ 0\\ \cos\left(\frac{42\pi}{180}\right)\end{pmatrix}\cdot e^{-i\cdot 100\cdot n\cdot\pi/180}\right] \qquad (1 - 21)$$

The R matrix is equal to 1 when all n residues form a linear helix. When there is a bend or kink in the middle of a helix (noted as a non-linear helical model), the first segment of the helix can be calculated as a linear helix, and a R matrix (equation 1 - 22) is needed to consider the phase change between the first segment of helix and the second segment of the helix when calculating the response for the second segment.[59]

$$R = \begin{pmatrix}\cos\left(\pi\cdot\frac{kk}{180}\right) & 0 & -\sin\left(\pi\cdot\frac{kk}{180}\right)\\ 0 & 1 & 0\\ \sin\left(\pi\cdot\frac{kk}{180}\right) & 0 & \cos\left(\pi\cdot\frac{kk}{180}\right)\end{pmatrix} \qquad (1 - 22)$$

where for a bent helical model $kk = \pi - \theta_{bent}$ (θ_{bent} is the bent angle between two helical segments) to consider the continuum between two helical segments, and for a kink helical model $kk = 100\cdot n_1$ ($n_1$is the residue number of the first helical segment) without considering the continuum between two helical segments.

After the Raman tensor and the IR transition dipole moment are calculated, the hyperpolarizability can be constructed according to equation (1 - 11). For example, β_{aca} of A mode is:

$$\beta_{aca} \propto \frac{\partial\alpha_{ac}^{*}}{\partial Q_A}(0°)\frac{\partial\mu_{a}^{*}}{\partial Q_A}(0°) \qquad (1 - 23)$$

And β_{aca} of E1 mode is:

$$\beta_{aca} \propto \frac{\partial\alpha_{ac}^{*}}{\partial Q_{\mathrm{E1}}}(100°)\frac{\partial\mu_{a}^{*}}{\partial Q_{\mathrm{E1}}}(100°) + \frac{\partial\alpha_{ac}^{*}}{\partial Q_{\mathrm{E1}}}(-100°)\frac{\partial\mu_{a}^{*}}{\partial Q_{\mathrm{E1}}}(-100°) \qquad (1 - 24)$$

Then the obtained hyperpolarizability components can be plugged into the equation (1 - 10) and the relationship between $\chi^{(2)}$ ratio is then a function of the tilt angle θ of the model helix:

$$\frac{\chi_{zzz}^{(2)}}{\chi_{yyz}^{(2)}} = \frac{|\chi_{zzz,A}^{(2)} + \chi_{zzz,\mathrm{E1}}^{(2)}|}{|\chi_{yyz,A}^{(2)} + \chi_{yyz,\mathrm{E1}}^{(2)}|} \qquad (1 - 25)$$

1.5.2 Hamiltonian Approach

The bond additivity model of α-helix is too simplified when it is applied to study proteins, because it only considers the contributions from the α-helical domains and the couplings contributed from helical structures. It ignores the signal contributions and couplings of other secondary structures (*e.g.* sheets and turns) as well as the couplings between different secondary structures. To better consider the overall signal contributions and couplings of all the amino acids in a protein, a Hamiltonian approach was proposed. For the Hamiltonian method, each amide unit in the protein was used as a basis to construct the Hamiltonian matrix. The peak centers of the uncoupled amide I vibration mode are used as the diagonal terms of the Hamiltonian matrix. Coupling between each pair of amide units can be calculated according to the distance and orientation of each local amide I vibration mode (with the known protein structure) and used as the off-diagonal terms of the Hamiltonian matrix. The Hamiltonian matrix can then be diagonalized, with the diagonal terms (eigen values) as the new peak centers and each new basis (eigen function) for each mode is a combination of the original local amide modes. The SFG signal strength of each mode (eigen vector) is then calculated for each new peak center. Using a reasonably assumed peak width for each mode, the SFG spectrum of the protein can be constructed for a specific orientation. The assigned local frequencies and peak widths for the spectral calculation can be obtained from experimental SFG fitted results. Then, the protein can be rotated, and the zzz and yyz SFG spectra are calculated as a function of the protein orientation (defined by the tilt angle θ and the twist angle ψ). These calculated orientation dependent SFG spectra are then compared to the experimental SFG data to deduce the interfacial orientation.

The molecular frame (xyz) of a single amide unit (shown in Figure 1-4) is defined to locate the CO bond as the z axis with the O-C-N unit on the yz plane. A local vibrational mode has contributions from vibrations along CO bond and CN bond, and the effective center vector of the local vibrational mode is defined as: [37,43]

$$\vec{r}_{amideI} = \vec{r}_C + 0.665\vec{r}_{CO} + 0.256\vec{r}_{CN}$$

where $\vec{r}_C$ is the unit vector of the carbon atom, $\vec{r}_{CO}$ is the unit vector of the CO bond and $\vec{r}_{CN}$ is the unit vector of the CN bond. This effective center vector is used to quantify the distance of the coupled modes described below.

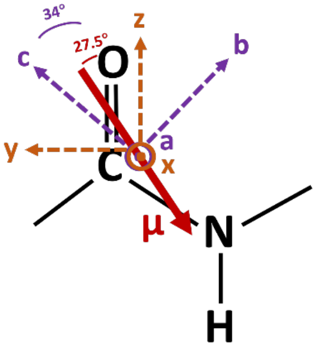

Figure 1-4 Schematic of defined orientation of the transition dipole (red), the principal coordinate axes of the Raman polarizability (purple) and molecular frame coordinate (orange) in an individual amide unit.

The purple axes (abc) shown in Figure 1-4 are the principal axes of the Raman polarizability. The c axis is 34° from the CO bond and the transition dipole (μ) is oriented 27.5° from the CO bond. Therefore, the μ and the transition polarizability (α) for an individual amide unit in the molecular frame are defined as following:[37,43]

$$\mu = \begin{pmatrix} 0 \\ -sin\frac{6.5\pi}{180} \\ -cos\frac{6.5\pi}{180} \end{pmatrix} \begin{pmatrix} 0 \\ sin\frac{27.5\pi}{180} \\ -cos\frac{27.5\pi}{180} \end{pmatrix}^T \tag{1 - 26}$$

$$\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\frac{34\pi}{180} & -sin\frac{34\pi}{180} \\ 0 & sin\frac{34\pi}{180} & cos\frac{34\pi}{180} \end{pmatrix} \begin{pmatrix} 0.25 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\frac{34\pi}{180} & -sin\frac{34\pi}{180} \\ 0 & sin\frac{34\pi}{180} & cos\frac{34\pi}{180} \end{pmatrix}^T \tag{1 - 27}$$

Hamiltonian program obtains the atom positions of a protein/peptide from the input MD simulated or crystal structure and utilizes these atom positions to calculate the molecular coordinate, amide center, μ and α contributions (mentioned above) of each local amide unit. The couplings between the local amide units are calculated based on the transition dipole coupling model:[37,43]

$$\kappa_{ij} = \frac{1}{4\pi\varepsilon_0}\left[\frac{\vec{\mu}_i \cdot \vec{\mu}_j}{\vec{r}_{ij}^3} - 3\frac{(\vec{r}_{ij} \cdot \vec{\mu}_i)(\vec{r}_{ij} \cdot \vec{\mu}_j)}{\vec{r}_{ij}^5}\right] \tag{1 - 28}$$

where κ_{ij} quantifies the coupling between the local mode i and local mode j, $\vec{\mu}_i$ ($\vec{\mu}_j$) is the transition dipole of the local mode i (j), and $\vec{r}_{ij}$ is the vector between the amide centers of two modes.

Then the Hamiltonian matrix can be constructed by setting the local mode frequencies as the experimental measured uncoupled peak centers (e.g. 1640 cm^{-1}) for the diagonal terms and setting the calculated coupling terms as the off-diagonal elements. The matrix can be diagonalized. The normal mode IR transition dipoles and normal mode Raman polarizabilities can be calculated by using the diagonalized Hamiltonian matrix:

$$\mu_N = \sum_m C_{Nm}\,\mu_m, \;\; \alpha_N = \sum_m D_{Nm}\,\alpha_m \tag{1- 29}$$

where the μ_m and α_m are the local mode IR transition dipole and local mode Raman polarizability, respectively. The μ_N and α_N are the normal mode transition dipole and normal mode transition polarizability, respectively. C_{Nm} (D_{Nm}) is the eigenvector matrix element for IR transition dipole (Raman polarizability). Then the Kronecker product of the normal mode IR transition dipole and the normal mode Raman polarizability generates the normal mode SFG hyperpolarizability as shown in the equation (1 - 11).

The normal mode hyperpolarizability tensor was then rotated into the lab frame by different orientation combinations (θ, ψ) to obtain the lab frame response susceptibility χ tensor. To construct an SFG spectrum, we expanded χ of each normal mode into a Lorentzian peak centered at each normal mode center frequency based on equation (1 - 2) and sum over all the contributions from the Lorentzian peaks. The damping coefficients are obtained from the experimental fitted peak widths.

IR spectra can also be constructed from the Hamiltonian approach based on equation (1 – 29). Azimuthal angle (φ) average was applied to the transition dipole μ in the lab frame (x, y, z):

$$\mu_{x,y,z}^{2}(\theta,\psi) = \int_{0}^{2\pi} \mu_{x,y,z}^{2}(\theta,\psi,\varphi)d\varphi \tag{1 - 30}$$

Then the transition dipole μ in the lab frame (x, y, z) was projected onto the Jones frame (s, p):[60]

$$\mu_s = \mu_y * E_y \tag{1 - 31}$$

$$\mu_p = \cos(\alpha) * \mu_x * E_x + \sin(\alpha) * \mu_z * E_z \tag{1 - 32}$$

where $\alpha = 45°$ is the incident angle, and the values of the correction terms $E_{x,y,z}$ were deduced from ref. 60.

Finally, the transition dipole μ of each normal mode was expanded into a Gaussian shape as shown in equation (1 - 33) and the final ATR-FTIR spectrum was constructed by the sum of all the Gaussian peaks.

$$I_{FTIR} = y + \sum_q A_q e^{-(\frac{x-x_q}{2w_q})^2} \tag{1 - 33}$$

where A_q, x_q, and w_q are the ATR-FTIR signal amplitude, the vibrational frequency (peak center), and the peak width of the vibrational mode q, respectively. y is an offset term to consider the spectral translation in the vertical direction, which can be set to 0 during the IR spectra calculation for convenience.

1.6 Atomistic Molecular Dynamic Simulation

Atomistic molecular dynamics (MD) simulation is a computational technique used to study the behavior of chemical and biological systems at an atomistic level of detail. In atomistic MD simulations, the interactions between all the atoms in a system are modeled, allowing for the prediction of the motion and behavior of complex systems with high accuracy. In atomistic MD simulations, the equations of motion for each atom in the system follow Newton's law and are integrated overtime, taking into account the interactions between the atoms, such as bonds, angles, and van der Waals forces.[61] Atomistic MD simulations can provide valuable insights into the detailed behavior of complex molecular systems, e.g., the underlying mechanisms of biological processes. In this thesis, atomistic MD simulations were applied to examine the behavior of proteins or peptides at interfaces over time, including the structural changes of proteins and peptides, and the energetics and kinetics of the interfacial systems. The structures of proteins and peptides obtained from the atomistic MD simulations can be used as input structures for the Hamiltonian program to generate SFG and ATR-FTIR spectra as a function of orientation.

1.7 Presented Research

In Chapter 2, the bond additivity model for a bent helix was employed to analyze the orientation of surface-immobilized AMPs SMAP29 variants in contact with the bacteria cell membrane component, lipopolysaccharides (LPS). It was found that the orientations of the surface-bound AMPs can be used as indicators to capture and distinguish different bacteria strains. This study provides valuable insights into AMP-LPS interactions at the molecular level, and explains how immobilization sites affect the structural characteristics of SAMP29 and lead to different

AMP-LPS binding behavior. Such knowledge is crucial for further AMP design in bacteria recognition.

In Chapter 3, the Hamiltonian approach along with discrete MD (DMD) simulation was used to favor the data interpretation of SFG measurements and both the conformation and orientation of the surface-immobilized uranyl harvesting protein were determined. This study found that upon binding to a hydrophobic surface via C-terminal, the uranyl binding protein (SUP) tends to lie down on the surface. No substantial conformation and orientation changes of the immobilized SUP were observed with and without the presence of uranyl ions. This study proposed that tuning the hydrophobicity of the surface for SUP immobilization or adjusting the hydrophobicity of the SUP itself could be future developments of SUP engineering to favor a standing-up pose of SUP at the interface to optimize its accessibility to uranyl ions. Additionally, the developed systematic methodology for simultaneous determination of conformation and orientation of interfacial proteins and peptides can be further employed in the future to screen more protein candidates and surfaces with different hydrophobicity to achieve better uranyl binding performance. The developed method can also be applied to study a wide range of important biological systems.

To improve the accuracy of interfacial protein structural determination, new methods in both experimental measurements and computational modeling were introduced in Chapter 4. Isotope labeling of ^{13}C for various amino acids in the protein sample was used to increase independently measured SFG parameters. Atomistic MD simulation was used to replace DMD for better sampling and precision. A systematic methodology with these new methods along with Hamiltonian approach was developed to deduce most likely conformation and orientation of B_1 domain of protein G (GB1) and its mutant on the polystyrene surface. It was found that only by mutating two key amino acids in GB1, the orientation of the protein charged substantially on the surface. This research demonstrated that the manipulation of protein orientation can be achieved by the mutation of key amino acids that in responsible for protein-surface interaction. Such knowledge is important to control the adsorbed protein orientation on surface to achieve optimized function. The systematic methodology developed is generally applicable to determine conformation and orientation of peptides and proteins on surfaces with great accuracy.

In Chapter 5, the developed method of combining SFG measurements, atomistic MD simulation and Hamiltonian spectra calculation was further developed by including the ATR-FTIR data to supplement the SFG measurement. The FTIR spectra calculation was implemented in the Hamiltonian program. This further developed systematic method was applied to reveal the structural changes (conformational and orientational changes) of AMP MSI-594 and its analogue MSI-594A associated with different lipids. The conformation and orientation differences between MSI-594 and MSI-594A were successfully correlated to their different effectiveness against Gram negative bacteria, which provides useful insights into future design of potent AMPs. The combination of SFG and ATR-FTIR measurements, along with the data interpretation methodology, can be powerful to understand the membrane-disrupting properties of a variety of biomolecules including amyloid peptides/proteins and fusogenic viral peptides/proteins.

In summary, this thesis first applied a previously developed method to study surface immobilized peptides interacting with components from bacteria cell membrane. Then, this thesis research focused on further developing a systematic methodology which combines SFG measurements, ATR-FTIR measurements, isotope labeling, Hamiltonian spectral calculation, and experimental-calculated spectra matching (detailed procedures will be discussed in Chapters 3, 4 and 5) to successfully determine protein/peptide conformations and orientations at buried interfaces. Such a method has been applied to study surface-immobilized super uranyl binding proteins, adsorbed B_1 domain of protein G on polystyrene surface, and AMP MSI-594 and its analogue MSI-594A associated with different lipid membranes. The above applications of this method on a variety of peptide/protein systems indicate that it is powerful to provide important understanding on structure-function relationships of interfacial peptides and proteins. The developed method is widely applicable to study many different peptides and proteins at interfaces, facilitating research and applications involving interfacial biomolecules such as those in biomaterials, antifouling coatings, membranes for biological separations, antimicrobial drugs, biosensors etc.

1.8 References

(1) Nguyen, H. H.; Lee, S. H.; Lee, U. J.; Fermin, C. D.; Kim, M. Immobilized Enzymes in Biosensor Applications. *Materials (Basel)* **2019**, *12* (1), 121. https://doi.org/10.3390/ma12010121.

(2) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97* (5), 1359–1472. https://doi.org/10.1021/cr960370z.

(3) Lee, M. J.; Yaffe, M. B. Protein Regulation in Signal Transduction. *Cold Spring Harb Perspect Biol* **2016**, *8* (6), a005918. https://doi.org/10.1101/cshperspect.a005918.

(4) Syrovatkina, V.; Alegre, K. O.; Dey, R.; Huang, X.-Y. Regulation, Signaling, and Physiological Functions of G-Proteins. *Journal of Molecular Biology* **2016**, *428* (19), 3850–3868. https://doi.org/10.1016/j.jmb.2016.08.002.

(5) Alves, D.; Olívia Pereira, M. Mini-Review: Antimicrobial Peptides and Enzymes as Promising Candidates to Functionalize Biomaterial Surfaces. *Biofouling* **2014**, *30* (4), 483–499. https://doi.org/10.1080/08927014.2014.889120.

(6) Costa, F.; Gomes, P.; Martins, M. C. L. Antimicrobial Peptides (AMP) Biomaterial Coatings for Tissue Repair. In *Peptides and Proteins as Biomaterials for Tissue Regeneration and Repair*; Elsevier, 2018; pp 329–345. https://doi.org/10.1016/B978-0-08-100803-4.00013-9.

(7) Drenth, J.; Mesters, J. *Principles of Protein X-Ray Crystallography*, 3rd ed.; Springer: New York, 2007.

(8) Roessle, M. Basics of X-Ray Scattering. **2012**, 49.

(9) Öster, C.; Kosol, S.; Hartlmüller, C.; Lamley, J. M.; Iuga, D.; Oss, A.; Org, M.-L.; Vanatalu, K.; Samoson, A.; Madl, T.; Lewandowski, J. R. Characterization of Protein–Protein Interfaces in Large Complexes by Solid-State NMR Solvent Paramagnetic

Relaxation Enhancements. *J. Am. Chem. Soc.* **2017**, *139* (35), 12165–12174. https://doi.org/10.1021/jacs.7b03875.

(10) Verardi, R.; Traaseth, N. J.; Masterson, L. R.; Vostrikov, V. V.; Veglia, G. Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins. In *Isotope labeling in Biomolecular NMR*; Atreya, H. S., Ed.; Advances in Experimental Medicine and Biology; Springer Netherlands: Dordrecht, 2012; Vol. 992, pp 35–62. https://doi.org/10.1007/978-94-007-4954-2_3.

(11) Yu, B.; Pletka, C. C.; Iwahara, J. NMR Observation of Intermolecular Hydrogen Bonds between Protein Tyrosine Side-Chain OH and DNA Phosphate Groups. *J. Phys. Chem. B* **2020**, *124* (6), 1065–1070. https://doi.org/10.1021/acs.jpcb.9b10987.

(12) Zhang, Y.; Sun, B.; Feng, D.; Hu, H.; Chu, M.; Qu, Q.; Tarrasch, J. T.; Li, S.; Sun Kobilka, T.; Kobilka, B. K.; Skiniotis, G. Cryo-EM Structure of the Activated GLP-1 Receptor in Complex with a G Protein. *Nature* **2017**, *546* (7657), 248–253. https://doi.org/10.1038/nature22394.

(13) Faccio, G. From Protein Features to Sensing Surfaces. *Sensors* **2018**, *18* (4), 1204. https://doi.org/10.3390/s18041204.

(14) Choi, M. M. F. Progress in Enzyme-Based Biosensors Using Optical Transducers. *Microchimica Acta* **2004**, *148* (3–4), 107–132. https://doi.org/10.1007/s00604-004-0273-8.

(15) Emran, M. Y.; Mekawy, M.; Akhtar, N.; Shenashen, M. A.; EL-Sewify, I. M.; Faheem, A.; El-Safty, S. A. Broccoli-Shaped Biosensor Hierarchy for Electrochemical Screening of Noradrenaline in Living Cells. *Biosensors and Bioelectronics* **2018**, *100*, 122–131. https://doi.org/10.1016/j.bios.2017.08.050.

(16) Kulagina, N. V.; Lassman, M. E.; Ligler, F. S.; Taitt, C. R. Antimicrobial Peptides for Detection of Bacteria in Biosensor Assays. *Analytical Chemistry* **2005**, *77* (19), 6504–6508. https://doi.org/10.1021/ac050639r.

(17) Liu, Y.; Yu, J. Oriented Immobilization of Proteins on Solid Supports for Use in Biosensors and Biochips: A Review. *Microchimica Acta* **2016**, *183* (1), 1–19. https://doi.org/10.1007/s00604-015-1623-4.

(18) Fu, J.; Zhu, W.; Liu, X.; Liang, C.; Zheng, Y.; Li, Z.; Liang, Y.; Zheng, D.; Zhu, S.; Cui, Z.; Wu, S. Self-Activating Anti-Infection Implant. *Nat Commun* **2021**, *12* (1), 6907. https://doi.org/10.1038/s41467-021-27217-4.

(19) Barry, M. E.; Davidson, E. C.; Zhang, C.; Patterson, A. L.; Yu, B.; Leonardi, A. K.; Duzen, N.; Malaviya, K.; Clarke, J. L.; Finlay, J. A.; Clare, A. S.; Chen, Z.; Ober, C. K.; Segalman, R. A. The Role of Hydrogen Bonding in Peptoid-Based Marine Antifouling Coatings. *Macromolecules* **2019**, *52* (3), 1287–1295. https://doi.org/10.1021/acs.macromol.8b02390.

(20) Callow, J. A.; Callow, M. E. Trends in the Development of Environmentally Friendly Fouling-Resistant Marine Coatings. *Nature Communications* **2011**, *2* (1), 244. https://doi.org/10.1038/ncomms1251.

(21) Cao, S.; Wang, J.; Chen, H.; Chen, D. Progress of Marine Biofouling and Antifouling Technologies. *Chin. Sci. Bull.* **2011**, *56* (7), 598–612. https://doi.org/10.1007/s11434-010-4158-4.

(22) Kannan, A.; Shieh, I. C.; Negulescu, P. G.; Chandran Suja, V.; Fuller, G. G. Adsorption and Aggregation of Monoclonal Antibodies at Silicone Oil–Water Interfaces. *Mol. Pharmaceutics* **2021**, *18* (4), 1656–1665. https://doi.org/10.1021/acs.molpharmaceut.0c01113.

(23) Lu, T.; Guo, W.; Datar, P. M.; Xin, Y.; Marsh, E. N. G.; Chen, Z. Probing Protein Aggregation at Buried Interfaces: Distinguishing between Adsorbed Protein Monomers, Dimers, and a Monomer–Dimer Mixture *in Situ*. *Chem. Sci.* **2022**, *13* (4), 975–984. https://doi.org/10.1039/D1SC04300E.

(24) Fabian, H.; Mantele, W. Infrared Spectroscopy of Proteins. 27.

(25) Ma, H.; Tang, X.; Liu, Y.; Han, X. X.; He, C.; Lu, H.; Zhao, B. Surface-Enhanced Raman Scattering for Direct Protein Function Investigation: Controlled Immobilization and Orientation. *Anal. Chem.* **2019**, *91* (14), 8767–8771. https://doi.org/10.1021/acs.analchem.9b01956.

(26) Reichart, T. M.; Uzarski, J. R.; Mello, C. M. Differential Presentation of a Single Antimicrobial Peptide Is Sufficient to Identify LPS from Distinct Bacterial Samples. *The Analyst* **2019**, *144* (24), 7242–7249. https://doi.org/10.1039/C9AN01781J.

(27) Uzarski, J. R.; Mello, C. M. Detection and Classification of Related Lipopolysaccharides via a Small Array of Immobilized Antimicrobial Peptides. *Analytical Chemistry* **2012**, *84* (17), 7359–7366. https://doi.org/10.1021/ac300987h.

(28) Xiao, M.; Jasensky, J.; Foster, L.; Kuroda, K.; Chen, Z. Monitoring Antimicrobial Mechanisms of Surface-Immobilized Peptides in Situ. *Langmuir* **2018**, *34* (5), 2057–2062. https://doi.org/10.1021/acs.langmuir.7b03668.

(29) Xiao, M.; Jasensky, J.; Gerszberg, J.; Chen, J.; Tian, J.; Lin, T.; Lu, T.; Lahann, J.; Chen, Z. Chemically Immobilized Antimicrobial Peptide on Polymer and Self-Assembled Monolayer Substrates. *Langmuir* **2018**, *34* (43), 12889–12896. https://doi.org/10.1021/acs.langmuir.8b02377.

(30) Han, X.; Zheng, J.; Lin, F.; Kuroda, K.; Chen, Z. Interactions between Surface-Immobilized Antimicrobial Peptides and Model Bacterial Cell Membranes. *Langmuir* **2018**, *34* (1), 512–520. https://doi.org/10.1021/acs.langmuir.7b03411.

(31) Weidner, T.; Castner, D. G. SFG Analysis of Surface Bound Proteins: A Route towards Structure Determination. *Phys. Chem. Chem. Phys.* **2013**, *15* (30), 12516. https://doi.org/10.1039/c3cp50880c.

(32) Hosseinpour, S.; Roeters, S. J.; Bonn, M.; Peukert, W.; Woutersen, S.; Weidner, T. Structure and Dynamics of Interfacial Peptides and Proteins from Vibrational Sum-Frequency Generation Spectroscopy. *Chem. Rev.* **2020**, *120* (7), 3420–3465. https://doi.org/10.1021/acs.chemrev.9b00410.

(33) Guo, W.; Lu, T.; Gandhi, Z.; Chen, Z. Probing Orientations and Conformations of Peptides and Proteins at Buried Interfaces. *J. Phys. Chem. Lett.* **2021**, *12* (41), 10144–10155. https://doi.org/10.1021/acs.jpclett.1c02956.

(34) Leng, C.; Hung, H.-C.; Sun, S.; Wang, D.; Li, Y.; Jiang, S.; Chen, Z. Probing the Surface Hydration of Nonfouling Zwitterionic and PEG Materials in Contact with Proteins. *ACS Applied Materials & Interfaces* **2015**, *7* (30), 16881–16888. https://doi.org/10.1021/acsami.5b05627.

(35) Zhang, C.; Parada, G. A.; Zhao, X.; Chen, Z. Probing Surface Hydration and Molecular Structure of Zwitterionic and Polyacrylamide Hydrogels. *Langmuir* **2019**, *35* (41), 13292–13300. https://doi.org/10.1021/acs.langmuir.9b02544.

(36) Huang, H.; Zhang, C.; Crisci, R.; Lu, T.; Hung, H.-C.; Sajib, M. S. J.; Sarker, P.; Ma, J.; Wei, T.; Jiang, S.; Chen, Z. Strong Surface Hydration and Salt Resistant Mechanism of a New Nonfouling Zwitterionic Polymer Based on Protein Stabilizer TMAO. *J. Am. Chem. Soc.* **2021**, *143* (40), 16786–16795. https://doi.org/10.1021/jacs.1c08280.

(37) Ding, B.; Laaser, J. E.; Liu, Y.; Wang, P.; Zanni, M. T.; Chen, Z. Site-Specific Orientation of an α-Helical Peptide Ovispirin-1 from Isotope-Labeled SFG Spectroscopy. *The Journal of Physical Chemistry B* **2013**, *117* (47), 14625–14634. https://doi.org/10.1021/jp408064b.

(38) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. 2D IR Line Shapes Probe Ovispirin Peptide Conformation and Depth in Lipid Bilayers. *Journal of the American Chemical Society* **2010**, *132* (8), 2832–2838. https://doi.org/10.1021/ja9101776.

(39) Yang, P.; Glukhova, A.; Tesmer, J. J. G.; Chen, Z. Membrane Orientation and Binding Determinants of G Protein-Coupled Receptor Kinase 5 as Assessed by Combined Vibrational Spectroscopic Studies. *PLOS ONE* **2013**, *8* (11), e82072. https://doi.org/10.1371/journal.pone.0082072.

(40) Boughton, A. P.; Yang, P.; Tesmer, V. M.; Ding, B.; Tesmer, J. J. G.; Chen, Z. Heterotrimeric G Protein <em>β</Em>$_1$<em>γ</Em>$_2$ Subunits Change Orientation upon Complex Formation with G Protein-Coupled Receptor Kinase 2 (GRK2) on a Model Membrane. *Proc Natl Acad Sci USA* **2011**, *108* (37), 15019.

(41) Guo, W.; Xu, S.; Reichart, T. M.; Xiao, M.; Lu, T.; Mello, C.; Chen, Z. Probing Molecular Interactions between Surface-Immobilized Antimicrobial Peptides and Lipopolysaccharides *In Situ*. *Langmuir* **2020**, *36* (41), 12383–12393. https://doi.org/10.1021/acs.langmuir.0c02492.

(42) Xiao, M.; Wei, S.; Li, Y.; Jasensky, J.; Chen, J.; Brooks, C. L.; Chen, Z. Molecular Interactions between Single Layered MoS $_2$ and Biological Molecules. *Chemical Science* **2018**, *9* (7), 1769–1773. https://doi.org/10.1039/C7SC04884J.

(43) Guo, W.; Zou, X.; Jiang, H.; Koebke, K. J.; Hoarau, M.; Crisci, R.; Lu, T.; Wei, T.; Marsh, E. N. G.; Chen, Z. Molecular Structure of the Surface-Immobilized Super Uranyl Binding Protein. *J. Phys. Chem. B* **2021**, *125* (28), 7706–7716. https://doi.org/10.1021/acs.jpcb.1c03849.

(44) Guo, W.; Lu, T.; Crisci, R.; Nagao, S.; Wei, T.; Chen, Z. Determination of Protein Conformation and Orientation at Buried Solid/Liquid Interfaces. *Chem. Sci.* **2023**, 10.1039.D2SC06958J. https://doi.org/10.1039/D2SC06958J.

(45) Hamm, P.; Zanni, M. Concepts and Methods of 2D Infrared Spectroscopy. 298.

(46) Roeters, S. J.; van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J. Phys. Chem. A* **2013**, *117* (29), 6311–6322. https://doi.org/10.1021/jp401159r.

(47) Ji, Y.; Yang, X.; Ji, Z.; Zhu, L.; Ma, N.; Chen, D.; Jia, X.; Tang, J.; Cao, Y. DFT-Calculated IR Spectrum Amide I, II, and III Band Contributions of *N* -Methylacetamide Fine Components. *ACS Omega* **2020**, *5* (15), 8572–8578. https://doi.org/10.1021/acsomega.9b04421.

(48) Arrondo, J. L. R.; Muga, A.; Castresana, J.; Goñi, F. M. Quantitative Studies of the Structure of Proteins in Solution by Fourier-Transform Infrared Spectroscopy. *Progress in Biophysics and Molecular Biology* **1993**, *59* (1), 23–56. https://doi.org/10.1016/0079-6107(93)90006-6.

(49) Arrondo, J. L. R.; Goñi, F. M. Structure and Dynamics of Membrane Proteins as Studied by Infrared Spectroscopy. *Progress in Biophysics and Molecular Biology* **1999**, *72* (4), 367–405. https://doi.org/10.1016/S0079-6107(99)00007-3.

(50) Zou, X.; Wei, S.; Jasensky, J.; Xiao, M.; Wang, Q.; Brooks, C. L.; Chen, Z. Molecular Interactions between Graphene and Biological Molecules. *J. Am. Chem. Soc.* **2017**, *139* (5), 1928–1936. https://doi.org/10.1021/jacs.6b11226.

(51) Lambert, A. G.; Davies, P. B.; Neivandt, D. J. Implementing the Theory of Sum Frequency Generation Vibrational Spectroscopy: A Tutorial Review. *Applied Spectroscopy Reviews* **2005**, *40* (2), 103–145. https://doi.org/10.1081/ASR-200038326.

(52) Lu, X.; Zhang, C.; Ulrich, N.; Xiao, M.; Ma, Y.-H.; Chen, Z. Studying Polymer Surfaces and Interfaces with Sum Frequency Generation Vibrational Spectroscopy. *Anal. Chem.* **2017**, *89* (1), 466–489. https://doi.org/10.1021/acs.analchem.6b04320.

(53) Barbeş, L.; Rădulescu, C.; Stihi, C. ATR-FTIR SPECTROMETRY CHARACTERISATION OF POLYMERIC MATERIALS.

(54) Yang, J. T.; Wu, C.-S. C.; Martinez, H. M. [11] Calculation of Protein Conformation from Circular Dichroism. In *Methods in Enzymology*; Elsevier, 1986; Vol. 130, pp 208–269. https://doi.org/10.1016/0076-6879(86)30013-2.

(55) Hoarau, M.; Badieyan, S.; Marsh, E. N. G. Immobilized Enzymes: Understanding Enzyme – Surface Interactions at the Molecular Level. *Org. Biomol. Chem.* **2017**, *15* (45), 9539–9551. https://doi.org/10.1039/C7OB01880K.

(56) Wang, Q.; Wei, S.; Wu, J.; Zou, X.; Sieggreen, O.; Liu, Y.; Xi, C.; Brooks, C. L.; Chen, Z. Interfacial Behaviors of Antimicrobial Peptide Cecropin P1 Immobilized on Different Self-Assembled Monolayers. *The Journal of Physical Chemistry C* **2015**, *119* (39), 22542–22551. https://doi.org/10.1021/acs.jpcc.5b06882.

(57) Jia, Z.; Li, J.; Gao, L.; Yang, D.; Kanaev, A. Dynamic Light Scattering: A Powerful Tool for In Situ Nanoparticle Sizing. *Colloids and Interfaces* **2023**, *7* (1), 15. https://doi.org/10.3390/colloids7010015.

(58) Nguyen, K. T.; Le Clair, S. V.; Ye, S.; Chen, Z. Orientation Determination of Protein Helical Secondary Structures Using Linear and Nonlinear Vibrational Spectroscopy. *J. Phys. Chem. B* **2009**, *113* (36), 12169–12180. https://doi.org/10.1021/jp904153z.

(59) Ding, B.; Soblosky, L.; Nguyen, K.; Geng, J.; Yu, X.; Ramamoorthy, A.; Chen, Z. Physiologically-Relevant Modes of Membrane Interactions by the Human Antimicrobial Peptide, LL-37, Revealed by SFG Experiments. *Sci Rep* **2013**, *3* (1), 1854. https://doi.org/10.1038/srep01854.

(60) Bechinger, B.; Ruysschaert, J.-M.; Goormaghtigh, E. Membrane Helix Orientation from Linear Dichroism of Infrared Attenuated Total Reflection Spectra. *Biophysical Journal* **1999**, *76*.

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

Chapter 2 Molecular Interactions Between Surface-Immobilized Antimicrobial Peptides and Lipopolysaccharides

2.1 Attribution Information

Materials covered in this chapter were adapted with permission from Guo, W.; Xu, S.; Reichart, T. M.; Xiao, M.; Lu, T.; Mello, C.; Chen, Z. Probing Molecular Interactions between Surface-Immobilized Antimicrobial Peptides and Lipopolysaccharides *In Situ*. *Langmuir* **2020**, *36* (41), 12383–12393. Copyright © 2020 American Chemical Society. This research was in collaboration with Combat Capabilities Development Command Soldier Center, Natick.

T.M.R. and C. M. provided peptide and LPS samples. W.G. and X.S. conducted SFG experiments and spectra fitting. W.G. conducted CD and DLS experiments and carried out quantitative data analysis based on SFG results.

2.2 Background and Motivation

2.2.1 Antimicrobial Peptides (AMPs)

Bacterial infection is one of the leading medical problems worldwide, which can cause severe immune response and even lead to mortality in patients. Conventional antibiotic treatments for microbial infection successfully saved millions of patients when they were first discovered in the 20th century. Antibiotics inhibit the DNA/RNA replications and protein synthesis to kill or slow the growth of bacteria.[1] This mechanism is vulnerable to adaptive resistance of bacteria, leading to diminished effectiveness of antibiotics over time.[2] Making the situation worse, the invention of new antibiotics is time consuming while the revolution of antibiotic-resistance bacteria is rapid especially with the abusive usage of antibiotics. Therefore, next generation of anti-infection drugs needs to be ungently developed which can overcome bacterial drug resistance and kill bacteria effectively, specifically, and rapidly.

Antimicrobial peptides (AMPs) are the promising candidates for combating bacterial infection and resistance because of their biocompatibility and excellent bacterial-targeting and -killing properties.[3] Different from traditional antibiotics, AMPs kill bacteria by disrupting the bacterial membrane via electrostatic and hydrophobic interactions[4–7], which is hard for bacteria to develop drug resistance. Due to these excellent properties, except for functioning as bacteria-killing agents, AMPs have also been widely used for bio-recognition to capture and distinguish bacteria. In practical applications, a biosensor requires two main parts, a bio-recognition moiety and a transduction platform. The bio-recognition elements, such as AMPs, are usually immobilized onto the platform for better sensitivity, portable ability, and reusability. A good bio-recognition element needs to be stable and be able to conduct original functions after the immobilization process. Extensive research proved that AMPs could retain their bacterial-killing functions after immobilizing onto a solid support, even though they failed to penetrate the bacterial cell membranes.[8–10]

2.2.2 Lipopolysaccharides (LPS)

Lipopolysaccharides (LPS), also known as endotoxin, are found at the outer membrane of Gram-negative bacteria. LPS can be released during the bacterial infection and can cause severe immune response, leading to inflammation, sepsis, and even death. Therefore, detection and classification of LPS at the early stage of bacterial infection (before the formation of biofilms) are crucial in clinical trials.

The structure of LPS mainly consists of three parts as shown in Figure 2-1.[11] The head O antigen is hydrophilic glycan-based polymer varied among bacterial strains. The middle is hydrophilic oligosaccharide chain, linked by KDO (3-deoxy-D-manno-oct-2-ulosonic acid) and non-sugar components such as phosphate and amino acids. The tail is hydrophobic phosphorylated glucosamine disaccharide, also called lipid A, containing multiple fatty acids. Overall, LPS carries negative charges, which in opposite of AMPs (usually contains several arginine (R) and Lysine (K) to be positively charged), making them attracted to each other via electrostatic interactions. Therefore, AMPs are naturally capable of recognizing LPS and developed as anti-bacterial agents.

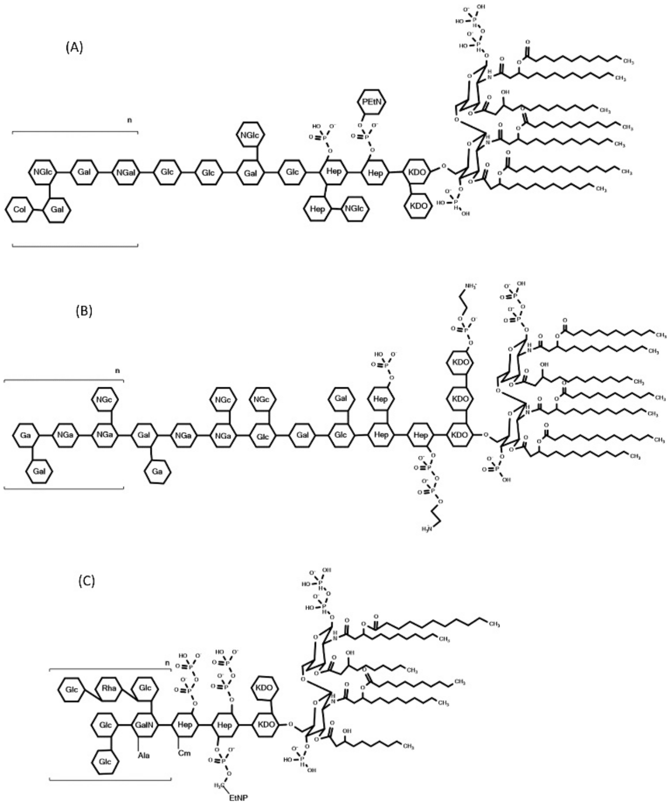

Figure 2-1 Archetypal LPS structures from: (A) E. coli O55:B5 (B) E. coli O111:B4 and (C) P. aeruginosa.

Col: colitose; Hep: L-glycerol-Dmanno-heptose; Gal: galactose; Glc: glucose; KDO: 2-keto-3-deoxyoctonic acid; NGa: N-acetyl-galactosamine; NGcl: N-acetyl-glucosamine; GlcN: glucosamine; PEtN: phosphorylethanolamine; Ala: alanine; Rha: rhamnose; GalN: galactosamine; Cm: carbamoyl; and EtnP: ethanolamine phosphate. (ref. 12 - 15)

2.2.3 SMAP29

SMAP29, an AMP extracted from sheep, exhibited excellent bacterial-killing and LPS-recognizing ability.[16,17] Previously, Uzarski and Mello used an immobilized AMP surface plasmon resonance (SPR) platform associated with linear discriminant data analysis method to successfully detect and classify LPS from various bacterial strains with detection limits ranging from 0.1 to 1.0 μM and 80% success rate.[18] SPR studies found that the combination of surface immobilized SMAP29 variants (immobilized onto surface via different residues) had better LPS recognition ability than the combination of non-SMAP29 AMP variants.[18,19] More molecular-level insights are needed to explain the strong binding affinity between SMAP29 with LPS molecules. CD studies found that SMAP29 was disordered in aqueous solutions and transited into helix-hinge-helix motif in LPS environments.[20] Two helical segments near N- and C- termini were identified as two independent LPS-binding sites, which explains the strong binding affinity between SMAP29 and LPS. However, this structural study using CD spectroscopy only involved SMAP29 in bulk solution. Real applications required the detailed structure analysis of surface immobilized SMAP29. Therefore, in this research, surface sensitive technique SFG was used to reveal interfacial structures of the immobilized SMAP29 variants with and without the binding of LPS in the molecular level. Furthermore, a semi-quantitative data interpretation method (bond additivity model for non-linear helix) was applied to differentiate the orientation preferences among the immobilized SMAP29 variants when contacting different LPS samples.

2.3 Experimental Materials and Methods

Right angle CaF_2 prisms with 100 nm silica coating were used as the solid platform to support SMAP29 immobilization. A layer of maleimide-terminated self-assembly monolayer (SAM) was grown on the silica-coated prism to react with the cysteine-modified SMAP29 to achieve peptide immobilization. This SAM was well-studied and showed excellent flexibility to prevent peptide/protein denaturalization.[8,21–23] The preparation of SAM was documented in previous literature.[8,22,23]

Three SMAP29 variants were designed: (1) SMAP29 with cysteine added at the N-terminus, noted as NT (sequence: **C**RGLRRLGRKIAHGVKKYGPTVLRIIRIAG), which exposed LPS binding site on the C-terminus after immobilization, (2) SMAP29 with cysteine added at the C-terminus, noted as CT (sequence: RGLRRLGRKIAHGVKKYGPTVLRIIRIAG**C**), which exposed LPS binding site on the N-terminus after immobilization, and (3) SMAP29 with cysteine modification at the middle site Y17, noted as Y17C (sequence: RGLRRLGRKIAHGVKK**C**GPTVLRIIRIAG), which exposed both N- and C- termini LPS binding sites after immobilization. All three samples were purchased from Peptide 2.0 and used as received.

LPS from three bacterial strains: O111:B4, O55:B5 and P.A. (structures shown in Figure 2-1) were purchased from Sigma and used as received. O111:B4 and O55:B5 were derived from *E. coli*, and *P.A.* were derived from *pseudomonas aeruginosa*. Each LPS sample was dissolved in ultrapure water (resistance > 18.2 mΩ) resulting in a concentration of 1.0 mg/mL. The final pH of O111:B4, O55:B5 and P.A. solutions were 7.86, 7.98 and 7.75 at 21 °C, respectively. All LPS solutions were stored at -40 °C before use.

SFG was applied to probe the interfacial structures of each SMAP29 variant in pure water and in the LPS solutions. Initially, the SFG ssp and ppp spectra were collected for each SMAP29 variant in contact with water. Then the SFG signal intensity change of 1655 cm^{-1} was monitored when replacing water with LPS solutions for each SMAP29 variant. Such SFG intensity change records the dynamics of the structural change of the amide I helical component of a peptide for each peptide-LPS pair (9 pairs in total). Detailed SFG setup was described in Chapter 1 and will not be repeated here. The SFG sample geometry used in this study is shown in Figure 2-2.

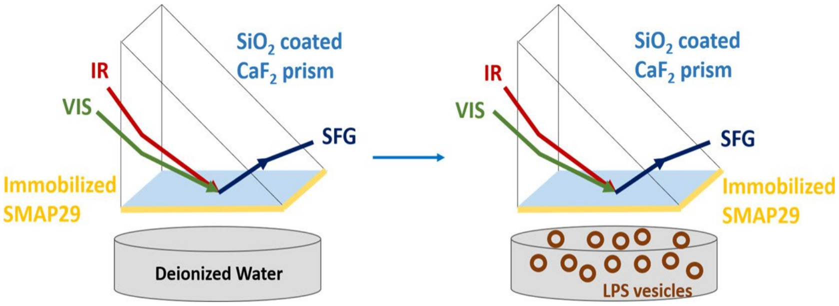

Figure 2-2 Schematic diagram of SFG sample geometry.

A J-1500 CD spectropolarimeter (Jasco Inc., Japan) was used to examine the secondary structures of surface-immobilized SMAP29 variants in both solution state and immobilization state. For solution state, the peptides were tested in deionized water with a concentration of 30 µM. For immobilized state, the peptides were immobilized on UV-grade quartz slides (the same immobilization process as that on the silica-coated CaF_2 prisms), and seven slides were stacked upon each other during the CD testing to generate detectable signal. 30 uL of solutions (in water or in LPS solution) were added into the gap between each slide before the CD testing. The CD spectra were collected under a nitrogen atmosphere, ranging from 240 to 190 nm with 1 nm resolution and 20 nm/min scan rate. Each CD spectra was scanned with 5 accumulations.

Dynamic light scattering (DLS) experiments (DynaPro NanoStar, WYATT Tech., US.) were conducted to determine the size of each LPS sample. Before DLS experiments, LPS samples were filtered through 0.45 µm tube filters to remove oversized vesicles. Each spectrum was obtained by 5 s recording time and 10 accumulations. A global protein model was used for spectral fitting.

2.4 Results and Discussion

2.4.1 Surface-immobilized SMAP29 Variants in Water

In solution state, CD data (Figure 2-3) showed that all three SMAP29 variants (NT, CT and Y17C) are randomly structured, same as the wild-type SMAP29.[20] This indicates that the mutations of SMAP29 variants in this research have minor effects on the structure of SMAP29. After immobilization, CD data of immobilized NT and CT showed obvious helical features, indicating the structures of NT and CT changed from random coil to helix after the surface immobilization. This result indicates that the substrate surface could limit the motion of NT and CT and the interactions with the SAM surface could favor the formation of helical structures of NT and CT. The CD data of immobilized Y17C still presented a random coil feature, indicating that the immobilization via a middle site may shorten the distance between the two termini and the SAM surface therefore hindering the formation of helical structures. The structure difference between NT (or CT) and Y17C indicates that the immobilization site can strongly affect the structure of the peptide, as discovered in previous reported studies.[21]

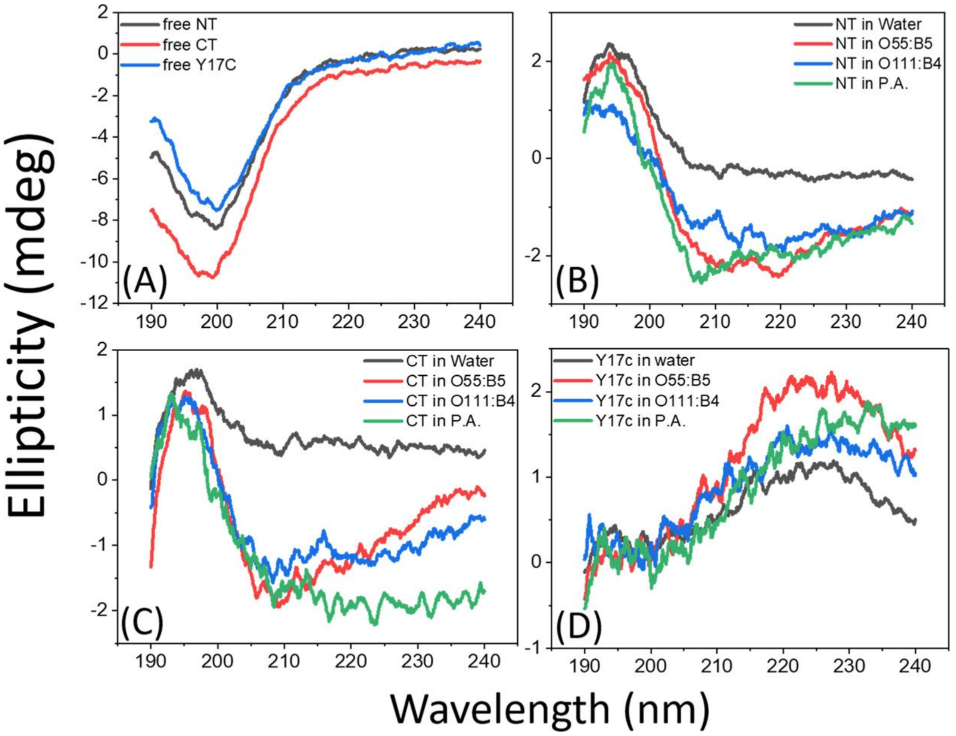

Figure 2-3 CD spectra collected from NT, CT, and Y17C in the free-solution state (A) and CD spectra collected from surface-immobilized NT (B), CT (C), and Y17C (D) in water and various LPS solutions.

SFG was also used to probe the interfacial structures of NT, CT and Y17C in contact with water. The experimental sample setup is shown in Figure 2-2. SFG spectra collected from SMAP29 variants in water had distinguishable features, as shown in Figure 2-4. The SFG spectra of NT and CT are dominated by a single peak ~1650 cm^{-1}, respectively, indicating that NT and CT possess α-helical structures at the interface. This result is in agreement with CD results presented above. The deduced χ_{zzz}/χ_{yyz} ratios are 1.35 and 1.01 for NT and CT, respectively. Since χ_{zzz}/χ_{yyz} ratios are related to the interfacial orientations, this result indicates that NT and CT adopted different orientations at the interface. Our previous studies also observed that same peptide could possess different interfacial orientations when being immobilized onto a maleimide-terminated SAM via different termini.[22,24,25] Such orientation preference is a result of tuning the electrostatic and hydrophobic interactions between the surface and the residues near the surface area. More detailed quantitative orientation analysis using the bond additivity model will be discussed below.

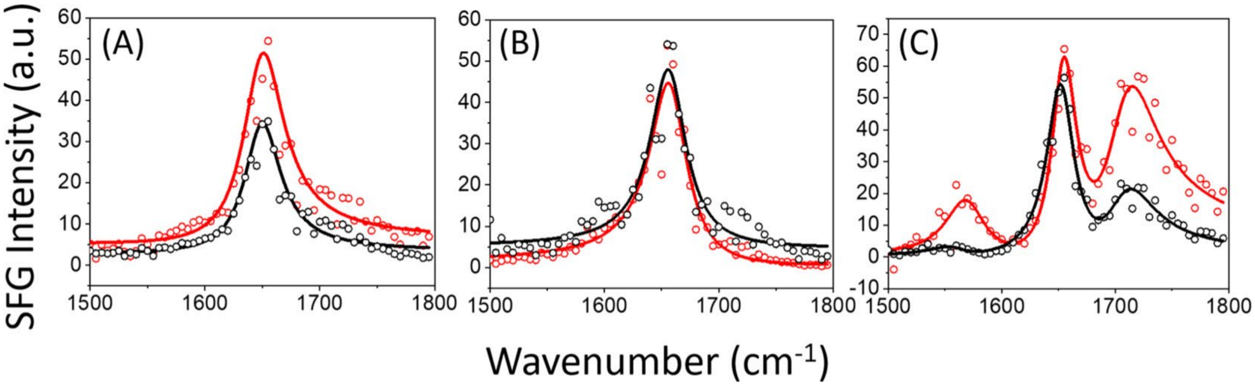

Figure 2-4 SFG ppp (red) and ssp (black) amide I spectra (dots) collected from surface-immobilized (A) NT, (B) CT, and (C) Y17C in water. The fitting results are shown as lines.

Here in this study, SMAP29 has different charge and hydrophobicity distribution of N- and C- termini, leading to different interactions with the SAM surface. The C-terminus of SMAP29 (RGLRRLGRKIAHGVKKYG) has 7 basic amino acids (R and K are positively charged at pH ~ 7) and a hydrophobicity of 0.045 while the N-terminus of SMAP29 (HGVKKYGPTVLRIIRIAG) has 4 basic amino acids and a hydrophobicity of 0.440. Therefore, the C-terminus carries more charge and favors the interactions with the solvent than with the SAM surface, while the N-terminus has more hydrophobic residues that favor the interactions with the SAM surface than with the solvent. As a result, NT should tilt towards surface more because its near-surface portion (N-terminus) is more likely to interact with the surface. On the contrary, CT should stand up more because its near-surface portion (C-terminus) is highly charged and more likely to interact with the solvent than with the surface.

The CD result (Figure 2-3) clearly showed that Y17C is still unordered after immobilization. In the SFG spectra of Y17C in water, the dominant peak is also at ~1650 cm^{-1}.

However, different from NT and CT, this peak is not assigned to a helix feature but a random coil feature when correlating to the CD results. The 1650 cm^{-1} peak in Y17C SFG spectra could come from the coil structures strongly interacted with the surface to form a pattern, which is SFG active, or from the coil structures which trapped a layer of water between the immobilized peptide and the SAM surface to create detectable water bending SFG signal. Other than the peak dominant at ~ 1650 cm^{-1} in SFG spectra, SFG spectra of Y17C in water exhibited two additional peaks located at ~1550 cm^{-1} and ~1710 cm^{-1}. The peak ~1550 cm^{-1} could come from the side chains of the charged residue, such as R and K, strongly interacted with the SAM surface, therefore adopted a preferred orientation, and generated SFG signal. The peak ~1710 cm^{-1} could come from the carbonyl group of the SAM surface interacting with the immobilized peptides.

In summary, the combination of CD and SFG data revealed the interfacial structural differences among three SMAP29 variants (NT, CT and Y17C), indicating that one could well-tune the interfacial structure of the same peptide by altering its immobilization site.

2.3.2. Surface-immobilized SMAP29 Variants in LPS Solutions

Interfacial structures of three SMAP29 variants (NT, CT and Y17C) in contact with three strains of LPS (O111:B4, O55:B5 and P.A.) were also studied by CD spectroscopy and SFG spectroscopy.

In CD spectra (Figure 2-3), the alpha helical features of NT and CT in the presence of LPS samples were more obvious than that in pure water environment, while the random coil features of Y17C remained the same as in the pure water environment.

Compared with SFG spectra of SMAP29 variants in water, SFG spectra of SAMP29 variants contacting with LPS samples (Figure 2-5) had the same peak assignments but various measured χ_{zzz}/χ_{yyz} ratios. Table 2-1 summarized the χ_{zzz}/χ_{yyz} ratios obtained from the fitting parameters of the SFG spectra in Figure 2-5. NT in contact with LPS samples possessed higher χ_{zzz}/χ_{yyz} ratios than that in pure water, and the same trend was also observed for CT. The χ_{zzz}/χ_{yyz} ratios for SFG spectra collected from the NT (or CT) in contact with various LPS samples followed a decreased trend from O111:B4 (2.08 for NT and 1.73 for CT) to O55:B5 (1.81 for NT and 1.61 for CT) to P.A. (1.62 for NT and 1.45 for CT). On the contrary, the χ_{zzz}/χ_{yyz} ratios of Y17C in LPS samples are relatively lower than the ratios of Y17C in water. All the χ_{zzz}/χ_{yyz} ratios of Y17C are lower than the ratios of NT and CT. Combined with CD data, Y17C continued to adopt a random structure after in contact with LPS samples. These observations clearly demonstrate that immobilized Y17C has different interactions with LPS molecules and the SAM surface from the immobilized NT and CT. Y17C could enable more complicated interactions at the interfaces with LPS (e.g., both termini can interact with LPS at the same time) because it is immobilized onto the surface via a middle site.

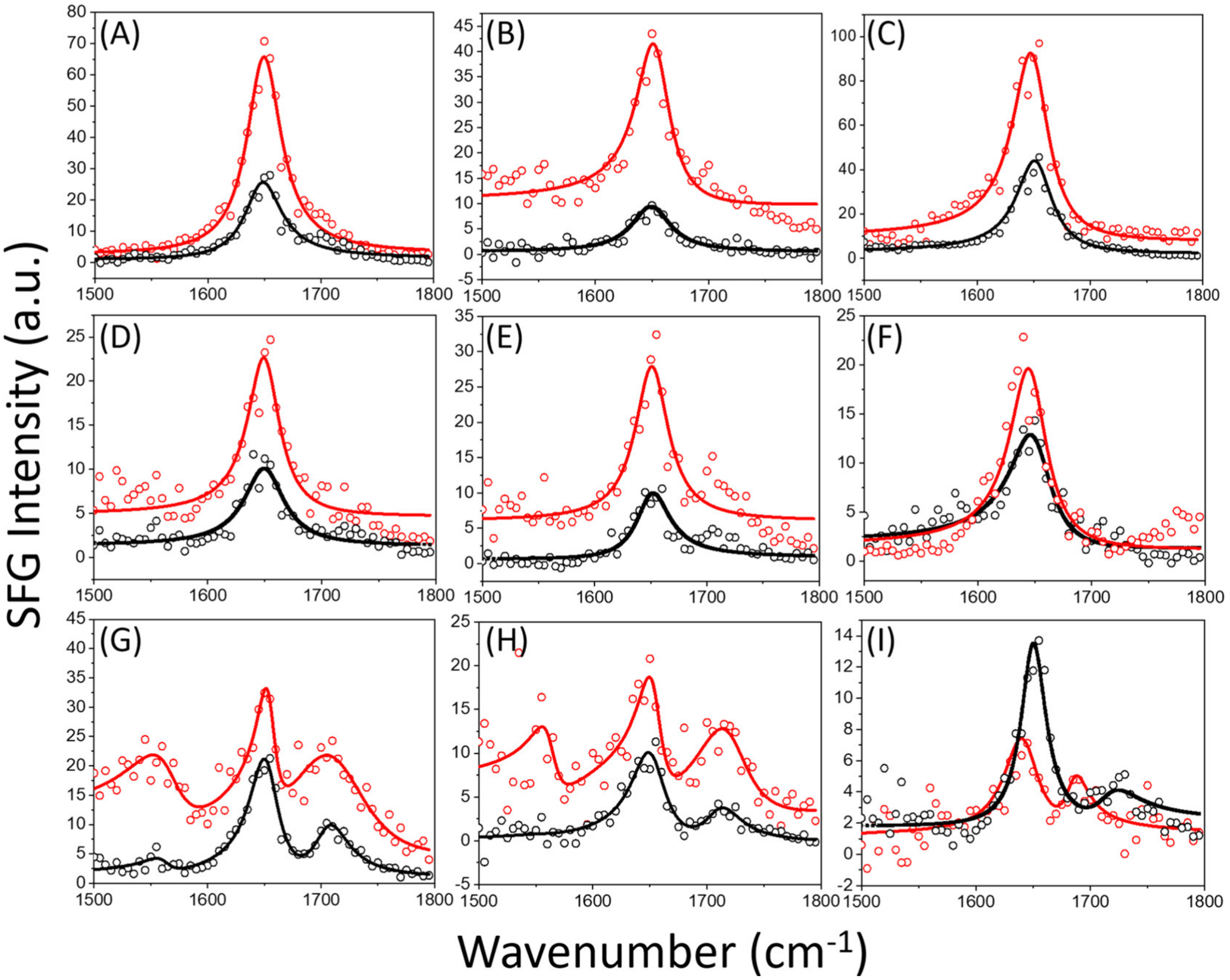

Figure 2-5 SFG ppp (red) and ssp (black) amide I spectra (dot) collected from the interfaces between surface-immobilized NT and (A) O55:B5, (B) O111:B4, and (C) P.A. solutions; between surface-immobilized CT and (D) O55:B5, (E) O111:B4, and (F) P.A. solutions; and between surface-immobilized Y17C and (G) O55:B5, (H) O111:B4, and (I) P.A. solutions. The spectral fitting results are shown as lines. The nonzero baselines of some spectra are due to the nonresonant signal contributions.

Table 2-1 χ_{zzz}/χ_{yyz} Ratios Deduced from Fitted SFG Spectra

	DEDUCED χ_{zzz}/χ_{yyz} AT 1650 CM^{-1}			
	water	O55:B5	O111:B4	P.A.
NT	1.35 (±0.21)	1.81(±0.22)	2.08(±0.47)	1.62(±0.22)
CT	1.01(±0.04)	1.61(±0.07)	1.73(±0.08)	1.45(±0.30)
Y17C	1.07(±0.37)	0.72(±0.36)	0.93(±0.71)	0.72(±0.47)

2.4.2 Quantitative Analysis of Immobilized SMAP29 Variants

Previously, it was found that SMAP29 possessed two helical ends with a middle bend in the membrane-like environment, such as LPS vesicles.[20] The two helical segments are located at residues 8-17 (10 amino acids) and residues 20-28 (9 amino acids). Therefore, we applied the bond additivity model introduced in Chapter 1 for a 12-9 bent helix (first segment has 10 helical residues and two turn residues, and second segment has 9 helical residues with the bent site at Pro19, illustrated in Figure 2-6) for a 'semi-quantitative' orientation analysis for SFG spectra collected from SMAP29 variants/LPS sample interfaces. The bent helical model assumes that the tilt angle of the tested peptide (the main axis of the first segment versus the surface normal) has a δ-distribution and the bent angle (the angle between the principal axis of the first helical segment and the principal axis of the second helical segment) is fixed. The bent angle was varied from 0 – 90° when considering the χ_{zzz}/χ_{yyz} ratio as a function of tilt angle (Figure 2-7).

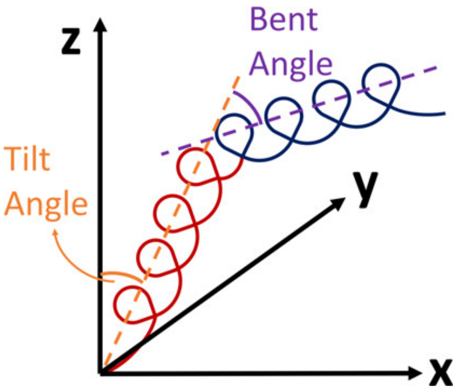

Figure 2-6 Illustration of a bent alpha-helix structure.

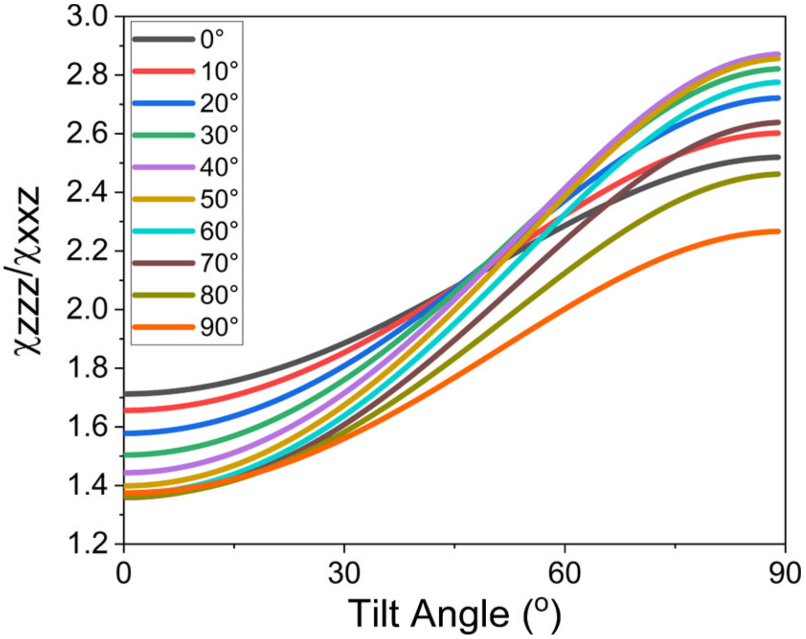

Figure 2-7 Relationship between the SFG χ_{zzz}/χ_{yyz} ratio and tilt angles of a bend alpha-helical SMAP29 with the first alpha-helix segment of 12 residues and the second alpha-helix segment of 9 residues. The tilt angle is defined as the angle between the surface normal and the first fragment. Curves in different colors represent different bent angles between the first and the second segments.

The deduced χ_{zzz}/χ_{yyz} ratios of NT and CT in water are beyond the range of all the curves shown in Figure 2-7, meaning that NT and CT in water may adopt multiple orientations[8] or they are not assembled as the assumed bent helix peptides. As presented above, according to the discussions on the CD and SFG data, we believe that immobilized Y17C adopts a random structure in water or in LPS samples. Here, the deduced χ_{zzz}/χ_{yyz} ratios measured for both Y17C in water and in LPS samples are beyond the possible ranges of a bent alpha helix, showing that the structure of the surface immobilized Y17C is not a bent helix.

The bent helical model can be used to deduce the possible orientation ranges of NT and CT in contact with LPS samples. The deduced bent angle range of NT in LPS samples is from 20° to 90° and the bent angle range of CT in LPS samples is from 40° to 90°. LPS was assembled into vesicles in the solution with a concentration of 1.0 mg/mL (as evidenced by the DLS data Figure

2-8). For LPS vesicles, the hydrophilic/charged portion is exposed while the hydrophobic core is buried inside. Therefore, the electrostatic interaction should be dominant when immobilized NT and CT encountered LPS vesicles. The first segment of the bent SMAP29 model carries 7 positive charges while the second segment carries 4 positive charges. Therefore, the CT should have stronger interaction with LPS samples because its N-terminus (which carries more charges) is closer to LPS vesicles. On the other hand, a weaker electrostatic interaction between NT and LPS samples resulted in a wider bent range for NT to behave more flexibly.

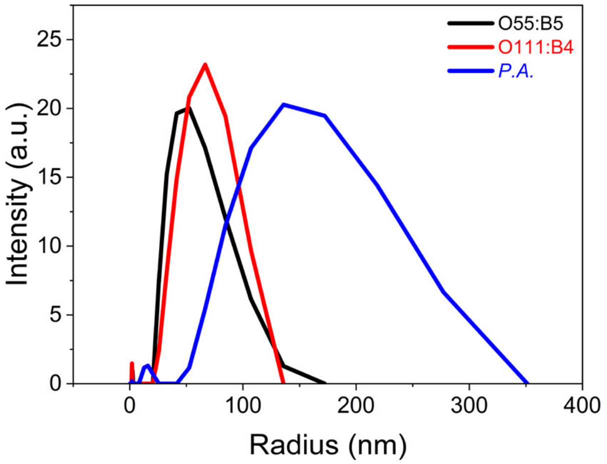

Figure 2-8 Size distributions of LPS vesicles measured by the DLS experiment.

For the purpose of comparison, the tilt angles of NT and CT with a fixed bent angle of 60° are listed in Table 2-2. With a fixed bent angle, the trend of the tilt angle can be concluded as following: (1) the tilt angles of the same peptide (NT or CT) in different LPS solutions followed the order of O111:B4 > O55:B5 > P.A. Among three LPS strains, P.A. exhibits the highest charge distribution per molecule and has the strongest electrostatic interactions with NT and CT to favor them to exhibit a tilting up presentation on the surface. (2) NT displayed a lager tilt angle than CT when interacting with the same LPS sample. This result is in correlation with the above discussion that NT should tilt towards to the surface more than CT, and CT is favored by LPS to stand up on the SAM surface because CT can have stronger Coulombic interaction between its N-termini and the LPS samples. Figure 2-9 summarizes the schematics of the peptide structures discussed above assuming the bent angle of 60° for NT and CT.

Table 2-2 Possible Orientation Angle Ranges of NT−LPS and CT−LPS Pairs.

	NT-O55:B5	**NT-O111:B4**	**NT-P.A.**	**CT-O55:B5**	**CT-O111:B4**	**CT-P.A.**
ORIENTATION RANGE/DEG (BENT ANGLE RANGE/DEG)	30-48 (20-90)	46-66 (20-90)	13-35 (20-90)	23-34 (40-90)	31-43 (40-90)	4-19 (40-90)
ORIENTATION/DEG (60° BENT ANGLE)	39	50	29	28	35	16

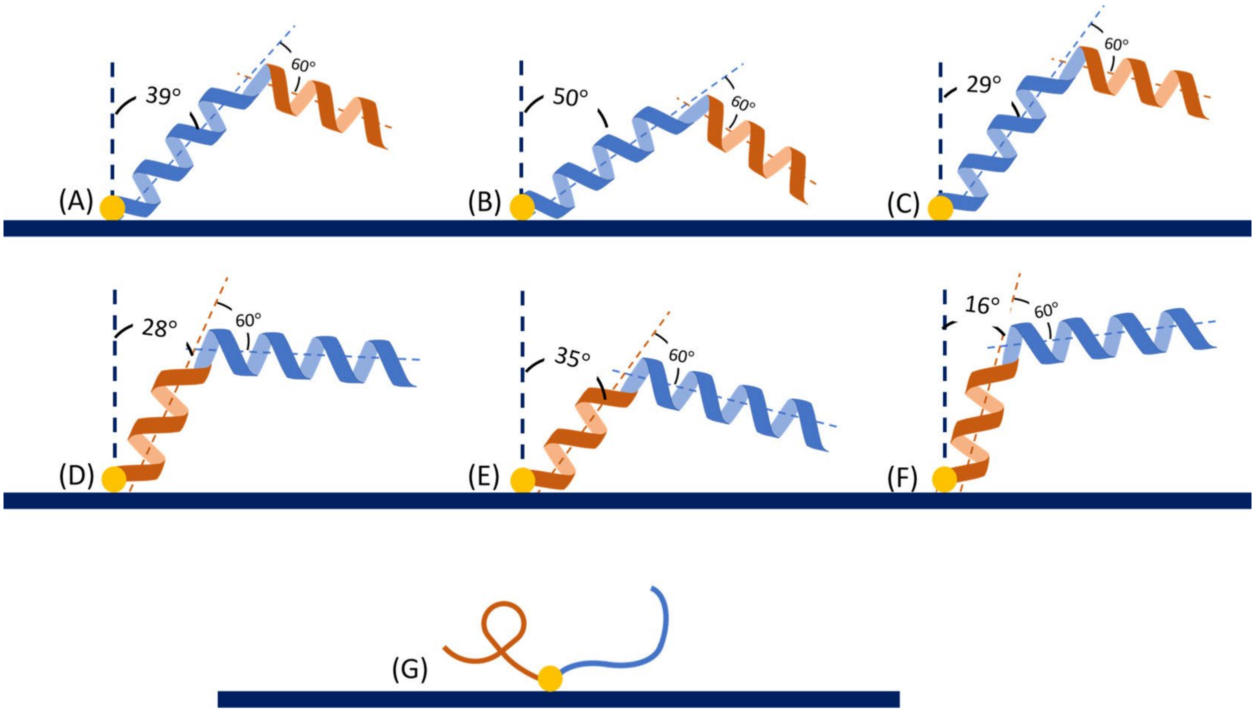

Figure 2-9 Schematics of NT (bent angle at 60°) in contact with (A) O55:B5, (B) O111:B4, and (C) P.A., CT (bent angle at 60°) in contact with (D) O55:B5, (E) O111:B4, and (F) P.A., and (G) Y17C in contact with LPS solutions. Y17C adopts nonhelical structures when in contact with LPS solutions; therefore, the orientations (tilt angles and bent angles) are not determined.

2.4.3 Dynamics from the Time-Dependent SFG Spectra

SFG ppp spectral intensity change at 1655 cm^{-1} was recorded as a function of time when switching the peptide contact solution from water to LPS samples. SFG time dependent spectra for each peptide-LPS pair are shown in Figure 2-10. SFG time dependent spectra of NT, CT and Y17C all showed a downward trend when in contact with O111:B4 or O55:B5. However, they all showed an upward trend when interacting with P.A. According to the DLS data shown in Figure 2-8, the P.A. has a larger vesicle size than NT and CT, which may be responsible for such an upward trend. In addition, the larger size of P.A. may induce more complicated interactions with the immobilized SMAP29 variants. Therefore, a longer time was needed for peptide signal to reach equilibrium when interacting with P.A. than other two LPS strains.

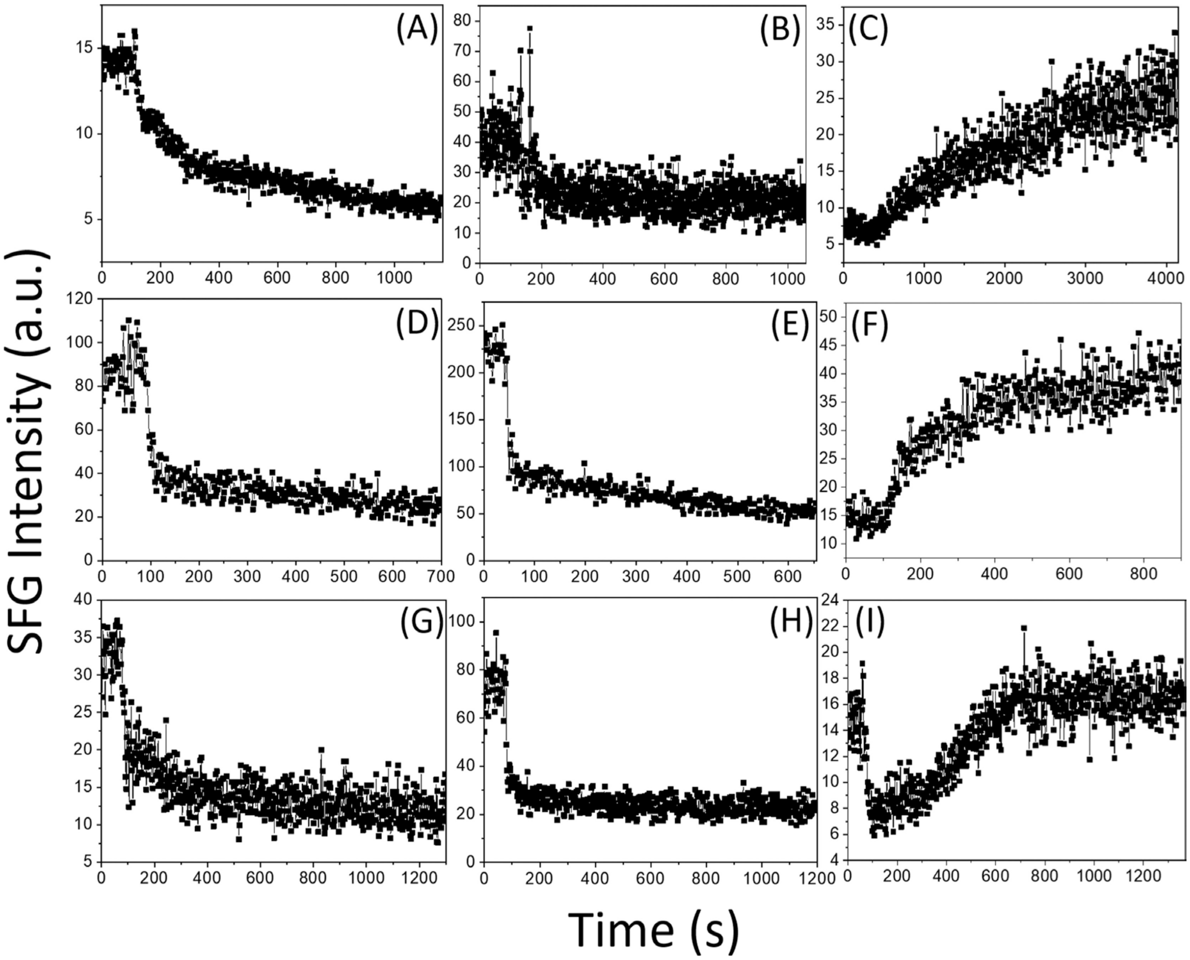

Figure 2-10 SFG time dependence ppp signal intensity measured at 1655 cm−1 of (A) NT/O111:B4, (B) NT/O55:B5, (C) NT/P.A., (D) CT/ O111:B4, (E) CT/O55:B5, (F) CT/P.A., (G) Y17C/O111:B4, (H) Y17C/O55:B5, and (I) Y17C/P.A. interfaces.

2.5 Conclusion

In this research, we demonstrated that the same peptide SMAP29 immobilized with different sites, N-terminus (NT), C-terminus (CT), and a middle site (Y17C) on a surface, has different structures in pure water environments and such differences further lead to different interactions with LPS samples from three bacterial strains, E. coli O111:B4, E. coli O55:B5, and P. aeruginosa. The different interfacial structures (including conformations and orientations) of three surface-immobilized SMAP29 variants during the peptide−LPS interactions were investigated using SFG and CD spectroscopy. A bond additivity model for a 12-9 bent helix was used for detailed orientation analysis for SFG measurements. The deduced orientation preferences for three immobilized SAMP29 variants when in contact with LPS samples, along with the dynamics/kinetics of these peptide-LPS interactions (monitored by the time-dependent SFG measurements), can serve as indicators to classify different peptide-LPS pairs. This research proposed that one could modulate a single AMP by differential immobilization to yield distinct interfacial structural characteristics of the AMP for bio-recognition.

2.6 References

(1) Kohanski, M. A.; Dwyer, D. J.; Collins, J. J. How Antibiotics Kill Bacteria: From Targets to Networks. *Nat Rev Microbiol* **2010**, *8* (6), 423–435. https://doi.org/10.1038/nrmicro2333.
(2) Ventola, C. L. The Antibiotic Resistance Crisis. 7.
(3) Hoyos-Nogués, M.; Gil, F. J.; Mas-Moruno, C. Antimicrobial Peptides: Powerful Biorecognition Elements to Detect Bacteria in Biosensing Technologies. *Molecules* **2018**, *23* (7), 1683. https://doi.org/10.3390/molecules23071683.

(4) Lee, T.-H.; N. Hall, K.; Aguilar, M.-I. Antimicrobial Peptide Structure and Mechanism of Action: A Focus on the Role of Membrane Structure. *Current Topics in Medicinal Chemistry* **2015**, *16* (1), 25–39. https://doi.org/10.2174/1568026615666150703121700.

(5) Domingues, T. M.; Perez, K. R.; Riske, K. A. Revealing the Mode of Action of Halictine Antimicrobial Peptides: A Comprehensive Study with Model Membranes. *Langmuir* **2020**, *36* (19), 5145–5155. https://doi.org/10.1021/acs.langmuir.0c00282.

(6) Zhang, S.; Fu, L.; Wan, M.; Song, J.; Gao, L.; Fang, W. Peripheral Antimicrobial Peptide Gomesin Induces Membrane Protrusion, Folding, and Laceration. *Langmuir* **2019**, *35* (40), 13233–13242. https://doi.org/10.1021/acs.langmuir.9b02175.

(7) Zhang, Y.; Chen, T.; Pan, Z.; Sun, X.; Yin, X.; He, M.; Xiao, S.; Liang, H. Theoretical Insights into the Interactions between Star-Shaped Antimicrobial Polypeptides and Bacterial Membranes. *Langmuir* **2018**, *34* (44), 13438–13448. https://doi.org/10.1021/acs.langmuir.8b02677.

(8) Xiao, M.; Jasensky, J.; Foster, L.; Kuroda, K.; Chen, Z. Monitoring Antimicrobial Mechanisms of Surface-Immobilized Peptides in Situ. *Langmuir* **2018**, *34* (5), 2057–2062. https://doi.org/10.1021/acs.langmuir.7b03668.

(9) Xiao, M.; Jasensky, J.; Gerszberg, J.; Chen, J.; Tian, J.; Lin, T.; Lu, T.; Lahann, J.; Chen, Z. Chemically Immobilized Antimicrobial Peptide on Polymer and Self-Assembled Monolayer Substrates. *Langmuir* **2018**, *34* (43), 12889–12896. https://doi.org/10.1021/acs.langmuir.8b02377.

(10) Han, X.; Zheng, J.; Lin, F.; Kuroda, K.; Chen, Z. Interactions between Surface-Immobilized Antimicrobial Peptides and Model Bacterial Cell Membranes. *Langmuir* **2018**, *34* (1), 512–520. https://doi.org/10.1021/acs.langmuir.7b03411.

(11) Raetz, C. R. H. BIOCHEMISTRY OF ENDOTOXINS. 44.

(12) Rodriguez-Loureiro, I.; Latza, V. M.; Fragneto, G.; Schneck, E. Conformation of Single and Interacting Lipopolysaccharide Surfaces Bearing O-Side Chains. *Biophysical Journal* **2018**, *114* (7), 1624–1635. https://doi.org/10.1016/j.bpj.2018.02.014.

(13) Magalhães, P. O.; Lopes, A. M.; Rangel-Yagui, C.; Penna, T. C. V.; Jr, A. P. Methods of Endotoxin Removal from Biological Preparations: A Review. **2007**, 17.

(14) Barkleit, A.; Moll, H.; Bernhard, G. Interaction of Uranium(vi) with Lipopolysaccharide. *Dalton Trans.* **2008**, No. 21, 2879. https://doi.org/10.1039/b715669c.

(15) Bystrova, O. V.; Lindner, B.; Moll, H.; Kocharova, N. A.; Knirel, Y. A.; Zahringer, U.; Pier, G. B. Full Structure of the Lipopolysaccharide of Pseudomonas Aeruginosa Immunotype 5. *Biochemistry (Moscow)* **2004**, *69* (2), 170–175. https://doi.org/10.1023/B:BIRY.0000018947.60328.8d.

(16) Dawson, R. M.; Liu, C.-Q. Cathelicidin Peptide SMAP-29: Comprehensive Review of Its Properties and Potential as a Novel Class of Antibiotics. *Drug Development Research* **2009**, *70* (7), 481–498. https://doi.org/10.1002/ddr.20329.

(17) Skerlavaj, B.; Benincasa, M.; Risso, A.; Zanetti, M.; Gennaro, R. SMAP-29: A Potent Antibacterial and Antifungal Peptide from Sheep Leukocytes. *FEBS Letters* **1999**, *463* (1–2), 58–62. https://doi.org/10.1016/S0014-5793(99)01600-2.

(18) Uzarski, J. R.; Mello, C. M. Detection and Classification of Related Lipopolysaccharides via a Small Array of Immobilized Antimicrobial Peptides. *Analytical Chemistry* **2012**, *84* (17), 7359–7366. https://doi.org/10.1021/ac300987h.

(19) Reichart, T. M.; Uzarski, J. R.; Mello, C. M. Differential Presentation of a Single Antimicrobial Peptide Is Sufficient to Identify LPS from Distinct Bacterial Samples. *Analyst* **2019**, *144* (24), 7242–7249. https://doi.org/10.1039/C9AN01781J.

(20) Tack, B. F.; Sawai, M. V.; Kearney, W. R.; Robertson, A. D.; Sherman, M. A.; Wang, W.; Hong, T.; Boo, L. M.; Wu, H.; Waring, A. J.; Lehrer, R. I. SMAP-29 Has Two LPS-Binding Sites and a Central Hinge. *European Journal of Biochemistry* **2002**, *269* (4), 1181–1189. https://doi.org/10.1046/j.0014-2956.2002.02751.x.

(21) Hoarau, M.; Badieyan, S.; Marsh, E. N. G. Immobilized Enzymes: Understanding Enzyme – Surface Interactions at the Molecular Level. *Org. Biomol. Chem.* **2017**, *15* (45), 9539–9551. https://doi.org/10.1039/C7OB01880K.

(22) Han, X.; Liu, Y.; Wu, F.-G.; Jansensky, J.; Kim, T.; Wang, Z.; Brooks, C. L.; Wu, J.; Xi, C.; Mello, C. M.; Chen, Z. Different Interfacial Behaviors of Peptides Chemically Immobilized on Surfaces with Different Linker Lengths and via Different Termini. *The Journal of Physical Chemistry B* **2014**, *118* (11), 2904–2912. https://doi.org/10.1021/jp4122003.

(23) Wang, Q.; Wei, S.; Wu, J.; Zou, X.; Sieggreen, O.; Liu, Y.; Xi, C.; Brooks, C. L.; Chen, Z. Interfacial Behaviors of Antimicrobial Peptide Cecropin P1 Immobilized on Different Self-Assembled Monolayers. *The Journal of Physical Chemistry C* **2015**, *119* (39), 22542–22551. https://doi.org/10.1021/acs.jpcc.5b06882.

(24) Li, Y.; Wei, S.; Wu, J.; Jasensky, J.; Xi, C.; Li, H.; Xu, Y.; Wang, Q.; Marsh, E. N. G.; Brooks, C. L. I.; Chen, Z. Effects of Peptide Immobilization Sites on the Structure and Activity of Surface-Tethered Antimicrobial Peptides. *J. Phys. Chem. C* **2015**, *119* (13), 7146–7155. https://doi.org/10.1021/jp5125487.

(25) Wang, Z.; Han, X.; He, N.; Chen, Z.; Brooks, C. L. I. Molecular Structures of C- and N-Terminus Cysteine Modified Cecropin P1 Chemically Immobilized onto Maleimide-Terminated Self-Assembled Monolayers Investigated by Molecular Dynamics Simulation. *J. Phys. Chem. B* **2014**, *118* (21), 5670–5680. https://doi.org/10.1021/jp5023482.

Chapter 3 Molecular Structure of the Surface-Immobilized Super Uranyl Binding Protein

3.1 Attribution Information

Materials covered in this chapter were adapted with permission from Guo, W.; Zou, X.; Jiang, H.; Koebke, K. J.; Hoarau, M.; Crisci, R.; Lu, T.; Wei, T.; Marsh, E. N. G.; Chen, Z. Molecular Structure of the Surface-Immobilized Super Uranyl Binding Protein. *J. Phys. Chem. B* **2021**, *125* (28), 7706–7716. Copyright © 2021 American Chemical Society. This research was in collaboration with Prof. Neil March's group at the University of Michigan and Prof. Tao Wei's group at Howard University.

M.H. and K.J.K contributed to the preparation of the protein samples used in this research. X.Z. and W.G. collected the SFG spectra with the help of H.J. and R.C. W.G. contributed to the spectra fitting, Hamiltonian calculations and spectra matching. W.G. and T.L. contributed to the development of the Hamiltonian program. T.W. conducted DMD simulations.

3.2 Background and Motivation

3.2.1 Super Uranyl Binding Protein (SUP)

Uranium is one of the most important sources for nuclear energy generation, however, it is considered a non-renewable nuclear source and it is lethal toxic to the environment if the release of uranium is not well-controlled. Therefore, it is necessary to develop methods to gather uranium sources from a diluted reservoir (such as harvesting uranyl ions from the ocean) for energy generation, and to sensitively detect uranium and uranyl ions for toxicity control and diagnosis. Many analytical techniques, including spectroscopy (e.g. fluorescence[1] and vibrational[2,3]), mass spectrometry[4] and electrochemical methods[5,6], have been developed for the detection of uranyl ions. More recently, biological molecules such as proteins and peptides have also been designed to detect uranyl ions because of their structure flexibility for ion-binding. In 2014, a protein named super uranyl binding protein (SUP) was designed to bind uranyl ions with femtomolar binding affinity and excellent selectivity (>10,000-fold affinity to uranyl ion over other metal ions.[7] This remarkable efficiency of SUP for uranyl ion selection and capture makes it promising to harvest uranyl ions from the sea water. The structures of the native SUP and the uranyl-bound SUP were resolved by X-ray crystallography, with PDB entry of 4FZO and 4FZP, respectively. These crystal structures of SUP with and without the presence of uranyl ions are superimposed, meaning that binding with uranyl ions has minor effect on SUP structure.[7,8] Like other biosensors, in real applications (such as the uranyl-affinity test for SUP), the native proteins are needed to be immobilized onto a solid support for better localization and transportation.[7] However, the interfacial structure of SUP has not been determined yet. Therefore, it is significant to study the structural changes of the immobilized SUP before and after contacting uranyl ions, which can help to better understand the binding affinity of the surface-immobilized SUP with uranyl ions. Such information is crucial for the design of SUP immobilization in real applications.

3.2.2 Methodology of Structural Determination of Interfacial SUP

In this study, a systematic method that combines the surface-sensitive SFG spectroscopy and computational methods (SFG spectra simulation and MD simulation) was developed to resolve both the conformation and orientation of the surface-immobilized SUP. This method can also be widely applied to study other interfacial proteins and peptides.

Compared to the bond additivity method that used to determine interfacial orientation of surface-immobilized SMAP29 variants when interacting with LPS samples (Chapter 2), a more advanced method, named Hamiltonian approach, was applied to study the conformation and orientation of surface-immobilized SUP. Different from the bond additivity model, which only considers the signal contributions of α-helical structures within the studied protein or peptide, the Hamiltonian approach considers the signal contributions of all the amino acids and mode coupling between each amino acid within the studied protein or peptide, and can simulate the SFG spectra as a function of orientation of the entire protein/peptide molecule. By comparing the experimental spectra with the calculated spectra under different orientations, one could resolve the orientation of the target protein or peptide with higher accuracy. Detailed mechanism of the Hamiltonian approach was discussed in Chapter 1, Section 1.4.2. Here, discrete molecular dynamics (DMD) simulations were employed to simulate the conformations of the surface-immobilized SUP and to provide the structure inputs for the SFG spectra calculations using the Hamiltonian approach.

3.3 Structures of Immobilized SUP on SAM Surface with and without the Presence of Uranyl Ions

3.3.1 Experimental Materials and Methods

In this study, SUP was mutated to SUP C31S C87S I105C (SUP-C105) for the purpose of immobilization via thiol-maleimide interaction.[9–11] The entire SUP-C105 sequence is as follows: SLDSRERIEKDLEDLEKELMEMSIKLSDDEEAVVERALNYRDDSVYYLEKGDHITSFGSI TYAEGLTDSLRMLHRICEG. The protocol for the expression and purification of SUP-C105 was reported previously.[8] These protein samples were stored at -80 °C until use. The immobilization substrate is maleimide-terminated SAM, the same as that used in the research presented in Chapter 2 above.

Near-total reflection sample geometry was used in SFG spectra collection (Figure 3-1). A layer of maleimide-terminated SAM was firstly grown on the bottom of a right-angle silica-coated CaF_2 prism. Then SUP-C105 was immobilized onto the SAM via maleimide-thiol interaction. The prism with immobilized SUP-C105 was placed into contact with a phosphate buffer (PB) solution (5 mM, pH = 7.4) and SFG ssp and ppp spectra were collected from the SUP-C105/PB solution interface in the frequency region of 1500 – 1800 cm^{-1}. Then the PB solution was replaced by a uranyl ion solution (100 μM uranyl ion in PB solution) and SFG signal was monitored. After the SFG signal became stable, the ssp and ppp SFG spectra of the SUP-C105/uranyl ion solution interface were collected in the same frequency region. The measured $\chi_{ppp}^{(2)}/\chi_{ssp}^{(2)}$ can be obtained from the fitted experimentally collected ssp and ppp SFG spectra, and the $\chi_{zzz}^{(2)}/\chi_{yyz}^{(2)}$ can then be deduced by the formula: $\chi_{ppp}^{(2)}/\chi_{ssp}^{(2)} = 0.9 \times \chi_{zzz}^{(2)}/\chi_{yyz}^{(2)}$ [12].

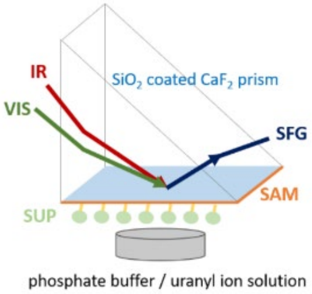

Figure 3-1 Schematic of the SFG sample geometry.

3.3.2 Parameters for Hamiltonian Approach

Hamiltonian approach introduced in Chapter 1, Section 1.4.2, can be used to simulate SFG spectra as a function of orientation (tilt angle θ, twist angle ψ) for an input protein/peptide structure. The local mode frequencies which were used as the diagonal terms in the Hamiltonian matrix were set to the same as the peak centers obtained from experimental spectra fitting, as listed in Table 3-1. The peak width of each amide I mode (normal mode) was also assumed as the corresponding peak width obtained from experimental spectral fitting, as listed in Table 3-1.

3.3.3 Score System for Spectral Comparison between Experimental Measurements and Hamiltonian Calculations

To evaluate the similarity between the calculated and experimentally collected SFG spectra, a quantitative score system (noted as parameter comparison score system) was developed. The parameter comparison score system is to firstly obtain the fitting parameters of the calculated SFG spectra and then compare these obtained parameters to the ones deduced from the experimentally collected spectra. For a single peak SFG ssp or ppp spectrum, these fitting parameters (also considered as matching standards) include their peak center, the peak width and the $\chi^{(2)}$ ratio between the SFG ssp and ppp spectra. To save computational time, for experimentally collected ppp or ssp spectrum which only contains one peak, the calculated SFG spectra (zzz or yyz) that have more than one peak and whose r^2 values of the fitting are lower than 0.8 were firstly filtered out. Spectra that survived from this filtering process are possible candidates for the next step, which then can be used to match against the experimental spectra. Score 1 and score 2 can be defined to evaluate the deviation between experimentally deduced peak centers in ppp and ssp spectra and fitted peak centers of the calculated zzz and yyz spectra, respectively. The deviation between the calculated and observed peak centers within ± 5 cm^{-1} is considered as a good matching with score > 0.9. Score 3 and score 4 assessed the propinquity between experimental deduced peak widths of the ppp and ssp spectra and fitted peak widths of the calculated zzz and yyz spectra, respectively. The tolerance of the deviation was set to ± 20% of the experimental value, and a score higher than 0.9 means that the deviation is less than ∼ 5 cm^{-1}. Score 5 evaluated the $\chi^{(2)}_{ppp}/\chi^{(2)}_{ssp}$ ratios between experimentally deduced and theoretically deduced values, with a tolerance of ± 20% of the experimentally deduced value. The final matching score is a product of all the five scores defined above. A good overall matching is defined as a final matching score higher than $0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.59$.

3.3.4 DMD Simulations

DMD simulations were performed using the software package (sDMD) developed by Zheng and co-workers.[13,14] In our simulations, the intra- and intermolecular interactions of protein(s) in the aqueous environment were represented precisely in an implicit water environment using well-calibrated all-atom DMD forcefield parameters published in the literature.[15] Interactions between protein residues and the substrate surface were computed efficiently using the coarse-grained Go-like model[16–18] In DMD, discontinuous step functions of the interparticle distance were adopted and particles were moved at constant velocities between steps.

The systems were simulated in a box of $10.0 \times 10.0 \times 10.0\ \mathrm{nm}^3$, and surfaces were all on the X−Z plane. Since DMD is event-driven and the solvent is represented only implicitly, it is not straightforward to correlate the simulation time and the temperature with the real time and temperature.[13] To address this issue, instead of applying the real units, we used time step t and reduced temperature $T^* = T/T_s$ in the simulations.[14] We assigned $T_s = 503.2$ K because we set $Nk_BT_s = E$, where k_B stands for Boltzmann's constant, N for Avogadro's number, and E for one unit of energy, taken to be 1 kcal/mol.[14] We chose $T^* = 0.5$ (approximately 250 K) because at such a temperature the SUP protein does not denature in the bulk water. A protein was initially placed above the surface with a gapping distance of ~2.0 nm with negligible protein−surface interactions. As a protein reaches the surface's vicinity within a cutoff distance of 0.3 nm, a harmonic potential between the SAM surface and the residue's immobilization site was formed to mimic the grafting of a protein onto the SAM surface via site Cys 105. The initial velocity of each atom was selected from the Maxwell−Boltzmann distribution at $T = 0.4$. After simulation with 5000 time steps, the system was stabilized. Then, the system was heated by a series of short runs that gradually increased T^* from 0.4 to 0.5, and during each run, we simulated for 20,000 time steps. DMD simulations were then carried out at 0.5 (approximately 250 K) for 1 × 107 time steps. Multiple independent simulations with different initial orientations and velocity profiles were carried out for each system to assess the effect of initial conditions. The initial orientations were randomly assigned.

3.3.5 Results and Discussion

3.3.5.1 Experimentally Collected SFG Spectra of Surface-Immobilized SUP-C105 in Buffer Solution and in Uranyl Ion Solution

Figures 3-2 (a) and (b) show the SFG ssp and ppp spectra collected from the surface immobilized SUP-C105 in contact with PB solution and with uranyl ion solution. The reconstructed ssp and ppp spectra without non-resonant backgrounds were also shown in Figures 3-2 (c) and (d). Only one peak at around 1650 cm^{-1} was detected in ssp or ppp spectrum of SUP-C105 before uranyl ion binding, which is contributed by the protein amide I signal. The ppp spectrum of SUP-C105 in contact with uranyl ions also contains only one protein peak. Differently, the ssp spectrum of SUP-C105 in contact with uranyl ions contains an additional peak centered at ~1700 cm^{-1}, in addition to the protein amide I peak centered at ~1650 cm^{-1}. Such a high frequency peak is rising from the stretching mode of the ordered C=O groups from the SAM surface, as a result of the interaction between the immobilized SUP-C105 and the SAM surface. Only the strong protein amide I peak was considered for each case when compared to the calculated protein spectra by Hamiltonian approach, as discussed below.

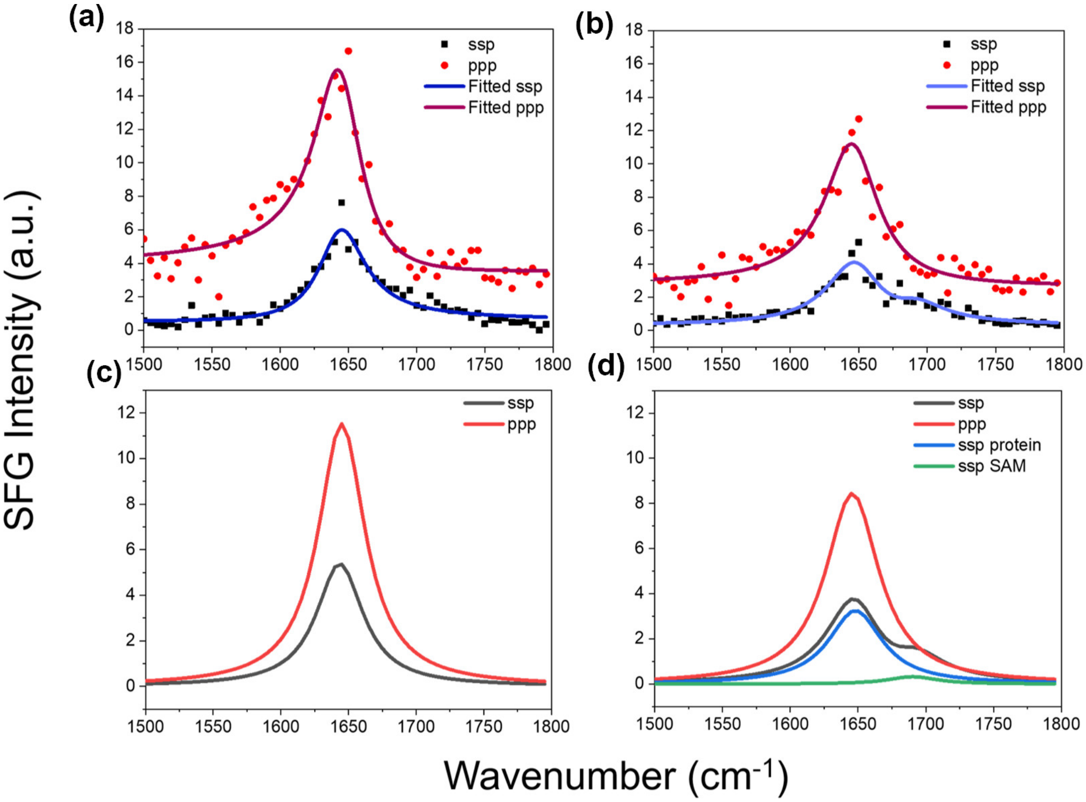

Figure 3-2 SFG ssp and ppp amide I spectra collected from the interfaces between surface-immobilized (a) SUP-C105 in contact with PB solution and (b) SUP-C105 in contact with a 100 μM uranyl ion PB solution. Dots represent experimental data, while lines are fitting curves. Fitted resonant SFG spectra from the immobilized (c) SUP-C105 in contact with PB solution and (d) SUP-C105 in contact with a 100 μM uranyl ion PB solution. Black: ssp spectra. Red: ppp spectra. The ssp spectra in (d) were divided into the SUP contribution (blue) and the SAM contribution (green, this fitted signal from SAM is very small). It is worth noting that no SFG signal can be detected from the SAM surface (before SUP immobilization) in the amide I signal frequency region.

Table 3-1 Fitting results of the collected SFG spectra displayed in Figure 3-2

	Amplitude	**Peak center (cm^{-1})**	**Peak width (cm^{-1})**

Before uranyl ions binding, ppp	71.3	1645.0	21.0
Before uranyl ions binding, ssp	48.8	1643.5	21.0
After uranyl ions binding, ppp	69.4	1645.8	23.9
After uranyl ions binding, ssp	43.3	1648.0	24.0

3.3.5.2 DMD Simulation Results of SUP-C105 without the Presence of Uranyl Ions

The native SUP structure with (PDB entry: 4FZO) and without (PDB entry: 4FZP) the presence of uranyl ions was resolved by X-ray crystallography, and it was found that binding with uranyl ions has minor effects on the SUP structure. To examine the SUP structural transition from a solid bulk state (deduced by X-ray crystallography) to solid/liquid interface, and the effect of mutation from native SUP to SUP-C105, DMD simulation was conducted. Since the SUP structure only changes slightly when binding with uranyl ions, here SUP-C105 without the presence of uranyl ions was examined by DMD simulations to represent both the uranyl ion binding and unbinding cases. Structures output from the DMD simulations were used to approximate the interfacial structures of both uranyl-bound and -unbound SUP-C105 and these structures were input into the Hamiltonian program for SFG spectra calculations. The Hamiltonian matching process will then be used to determine the most possible orientations of each input structure. Crystal structures (4FZO for uranyl-unbound SUP and 4FZP for uranyl-bound SUP) were also used as inputs for Hamiltonian calculations. The final matching result illustrated below showed that higher matching scores can be obtained by using DMD simulated structures instead of the crystal structures, indicating that DMD simulations can provide better candidates to represent the interfacial structures of SUP-C105.

Four runs of DMD simulations with different initial velocity profiles and protein orientations were performed and the final representations of all the cases are shown in Figure 3-3. Three out of the four cases represent a "lying down" representation of SUP-C105 with chain 3 (here, we defined chain 1 as 6Ser− 26Ser, chain 2 as 30Ser−53Lys, and chain 3 as 55Asp−76Leu based on the published SUP crystal structures 4FZO and 4FZP) touching the SAM surface, due to the strong surface-protein interactions. The calculated protein−surface interaction energies (E_{ps}) between different helical chains indicate that the interaction between chain 3 and the SAM surface has the lowest energy. The comparison of E_{ps} among different cases shows that in case 2, the protein-surface interaction is the strongest among all the four cases. In contrast, case 4 of surface-immobilized SUP-C105 shows a "standing" posture. Detailed analysis below ruled out the possibility of a standing-up pose of SUP-C105 on the SAM surface.

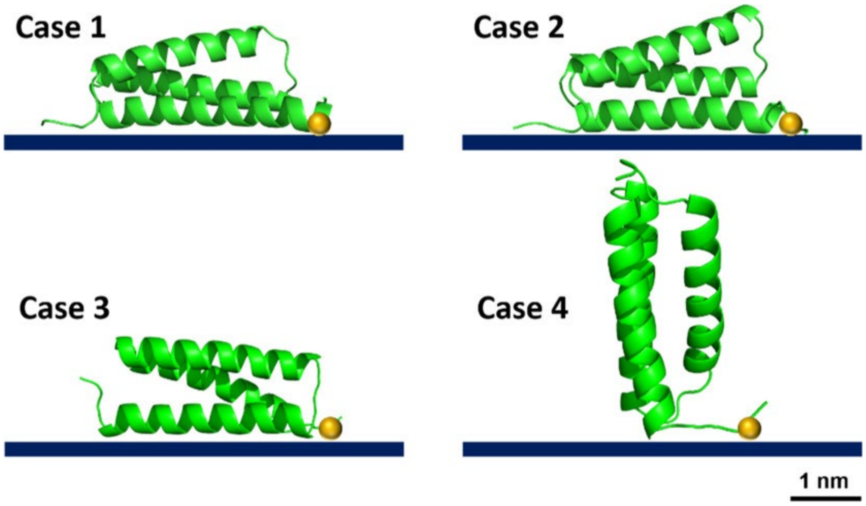

Figure 3-3 Simulated results of surface-immobilized SUP-C105 on the maleimide-terminated SAM surface using the DMD approach. The surface immobilization site (cysteine) is labeled as a yellow ball. Protein−surface interaction energies E_{ps} for the four cases are as follows: −26.27 kcal/mol (case 1), −27.89 kcal/mol (case 2), −18.95 kcal/mol (case 3), and −16.02 kcal/mol (case 4). Note: a negative value represents an attractive interaction energy, and a positive value stands for a repulsion energy. The differences of chain−surface interaction energies are as follows: 14.23 kcal/mol for $\Delta E_{1,3} = E_{chain1} - E_{chain3}$ and 14.41 kcal/mol for $\Delta E_{2,3} = E_{chain2} - E_{chain3}$.

3.3.5.3 Possible Interfacial Orientations of SUP-C105 with and without the Presence of Uranyl Ions, Determined via DMD Structures

The protein structure obtained from each case of the DMD simulation was input to Hamiltonian program to calculate the SFG spectra as a function of protein orientation. As a result, four overall score maps (including all the scores 1-5 defined above) were obtained by comparing the calculated spectra with the reconstructed experimental spectra of SUP-C105 without uranyl ion binding, shown in Figure 3-4. The best matched orientations for each DMD case are case 1 (40°, 90°) and (140°, 270°), case 2 (55°, 75°) and (125°, 255°), case 3 (50°, 280°) and (130°, 100°), and case 4 (70°, 140°) and (130°, 100°). It is worth noting that here the azimuthal angle average was applied when calculating the SFG spectra. Thus, the obtained final score map represents a symmetric feature: the score of (θ, ψ) is equal to the one of $(180° - \theta, 180° + \psi)$, which means that the absolute orientation (up or down) cannot be differentiated in the heat map. Therefore, the final score map always generates a pair of the most possible orientations. However, for a surface-immobilized protein, the immobilization site should always be adjacent to the surface. Such a criterion can be used to further differentiate the right orientation from the selected pair from the final score map. Based on this criterion, case 1 (140°, 270°), case 2 (125°, 255°), case 3 (130°, 100°) and case 4 (130°, 100°) have the immobilization site far away from the surface and should be excluded. All other four matched results (shown in Figure 3-5) indicated that the immobilized SUP-C105 lies down on the SAM surface with the absence of uranyl ions. When compared carefully, the matched orientations of case 1 – 3 show similar results: the chain 3 of SUP-C105 is the closest chain to the surface, the chain 2 is the furthest chain from the surface, and the chain 1 is titled with the N-terminus close to the surface. The best matched orientation of case 4 shows a lying down posture, which is different from its initial standing posture from simulation. This indicates that the Hamiltonian matching result fails to be aligned with the DMD simulation result. Therefore, case 4 should be ruled out from the possible structure of SUP-C105 without the presence of uranyl ions. Among cases 1, 2, and 3, we chose case 2 (55°, 75°) as the best matching result because it has the highest matching score.

The hydrophobic residues of SUP are evenly distributed among all three helical chains while the charged residues are mainly distributed on chain 1 and chain 2. This allows all three chains to have hydrophobic interactions with the surface and favors the chain 1 and chain 2 to interact with the solvent molecules. In addition, the immobilization site is located near the C-terminus of SUP and restricts the rotation of chain 2 from being close to the surface. The protein-surface interaction energies and chain-surface interaction energies were calculated and the result shows that case 2 has the lowest protein-surface energy and chain 3 has the lowest chain-surface energy among all the three chains, which means that case 2 is the best case to describe the conformation and the orientation of immobilized SUP-C105 on SAM surface (without binding with uranyl ions) and the chain 3 should be the closest chain in contact with the SAM surface. Therefore, we concluded that in consistent with both DMD simulation and Hamiltonian matching results, chain 3 of case 2 should have the highest affinity with the SAM surface.

The SFG spectra calculated from the four DMD structures were also compared to the reconstructed experimental SFG spectra of SUP-C105 after binding with uranyl ions. The best matched orientations were found to be case 1 (35°, 80°), case 2 (55°, 85°), case 3 (60°, 285°) and case 4 (25°, 250°) (only the matched orientations with immobilization site near surface will be discussed here and below). The matching score maps are shown in Figure 3-6. These matched orientations are very similar to the ones of SUP-C105 before uranyl ion binding, which further confirms that uranyl binding has minor effects on the SUP structure. Among cases 1, 2, and 3, we chose case 2 (55°, 85°) (shown in Figure 3-8 (d)) as the best matched orientation because it has the highest matching score.

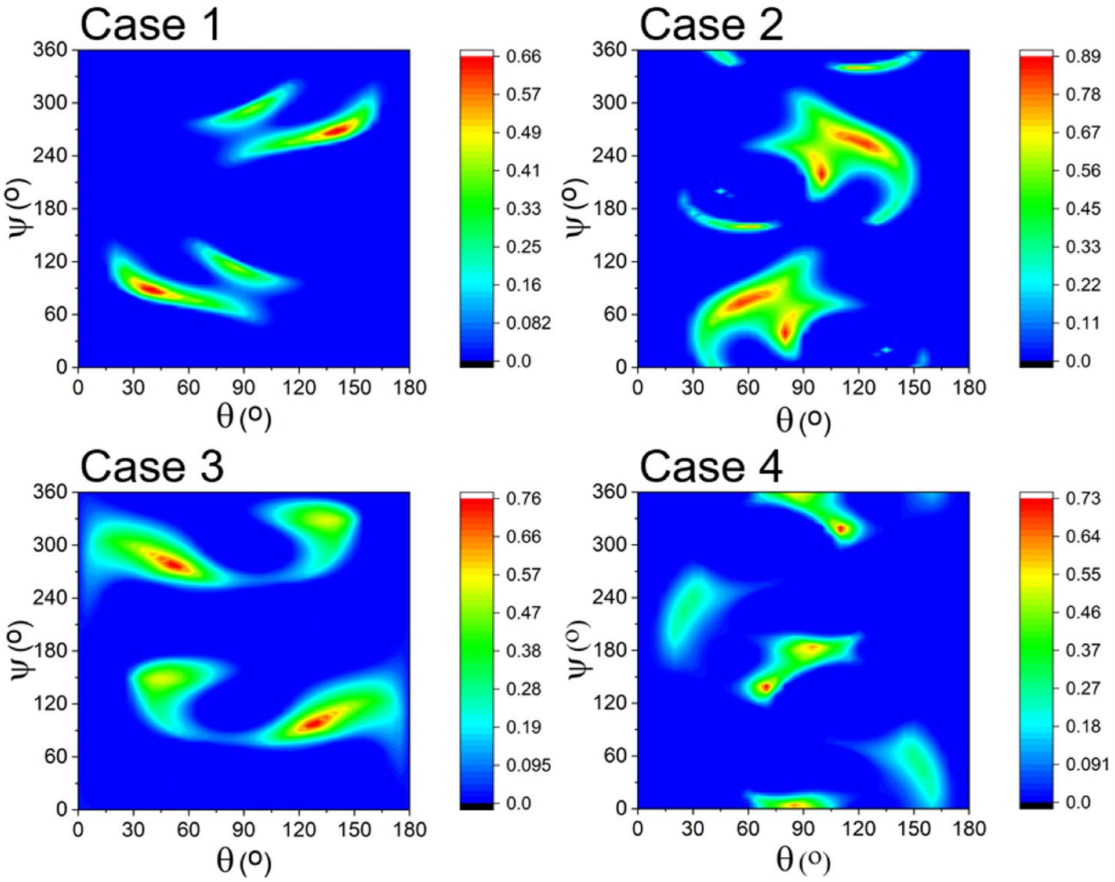

Figure 3-4 Final score map showing the matching scores between SFG experimental data and calculated spectra of various DMD structures (case 1 to case 4) using the Hamiltonian approach as a function of orientation angles of immobilized SUP-C105 before uranyl ion binding based on all five matching criteria.

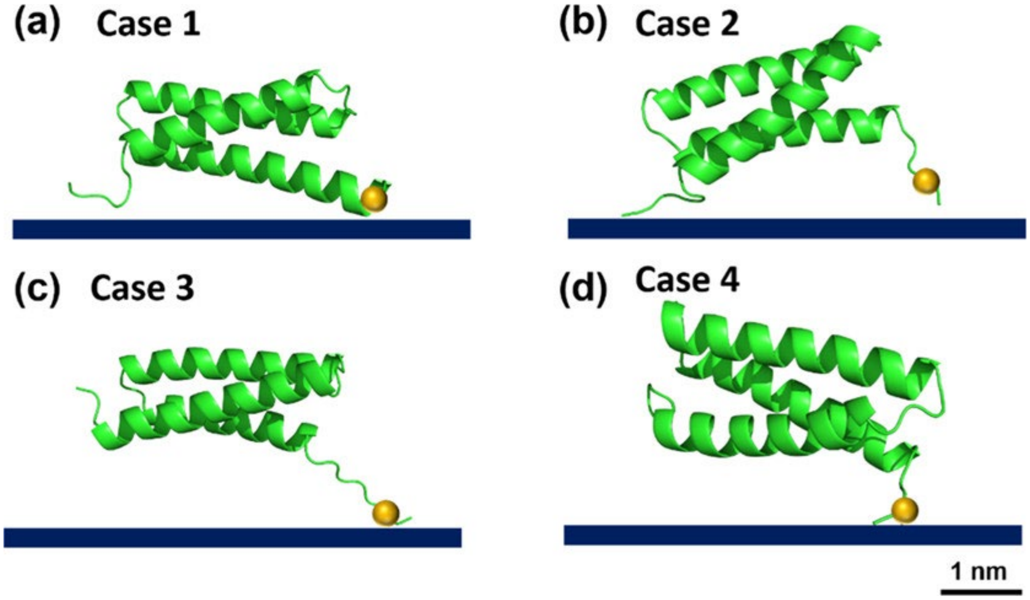

Figure 3-5 Visualized structures with the best matching scores with experimental data using various DMD structures as inputs for the Hamiltonian approach calculation: (a) case 1 (40°,90°), (b) case 2 (55°,75°), (c) case 3 (50°, 280°), and (d) case 4 (70°, 140°) of surface-immobilized SUP-C105 before uranyl binding.

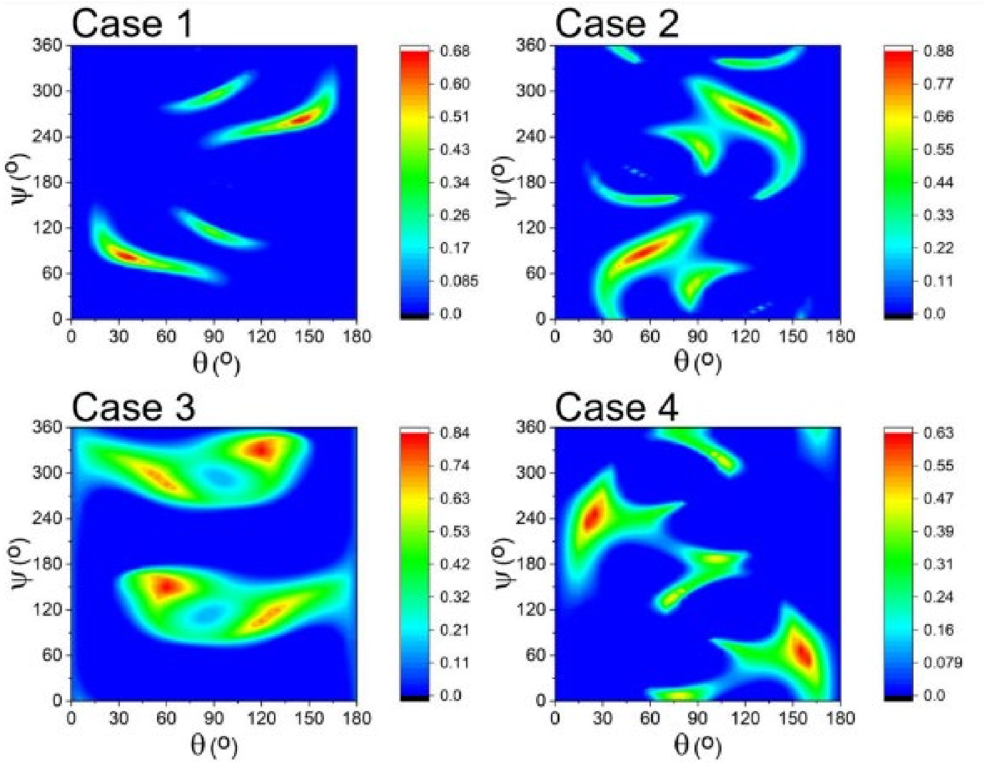

Figure 3-6 Final score maps showing the matching scores between SFG experimental data and calculated spectra of various DMD structures (case 1 to case 4) using the Hamiltonian approach as a function of orientation angles of immobilized SUP-C105 after uranyl ion binding based on all five matching criteria.

.

3.3.5.4 Possible Interfacial Orientations of SUP-C105 with and without the Presence of Uranyl Ions, Determined via Crystal Structures

The crystal structures deduced by X-ray crystallography and NMR spectroscopy were often used to approximate interfacial structures of proteins. This approximation is reasonable when the studied proteins are rigid, as shown in our previous publication.[16,17] Here, the crystal structure of native SUP without (4FZO) and with (4FZP) the presence of uranyl ions were also used as inputs for Hamiltonian program, and the calculated spectra as a function of protein orientation (θ, ψ) were compared to the corresponding reconstructed experimental spectra. The matching results can then be compared with the results obtained by using the DMD simulated structures. The two final score maps (final score map is the multiplication of five score maps defined in Section 3.2.2.) from the spectral comparisons are shown in Figure 3-7. Here, the most possible orientation of surface-immobilized SUP-C105 before uranyl ion binding by using 4FZO as the input structure is (75°, 105°), visualized in Figure 3-8 (a). The most possible orientation (90°, 110°) of surface-immobilized SUP-C105 after uranyl ion binding by using 4FZP as the input structure is also shown in Figure 3-8 (c).

The matching results show that the SUP-C105 lies down regardless of the presence of uranyl ions. However, in these matching results, chain 1 is the closest chain to the surface, which contrasts with the DMD results. As discussed above, chain 3 should be the chain that has the strongest interaction with the surface, and the crystal structure matching results fail to agree with such a conclusion. In addition, the best matching scores obtained by using the crystal structures are lower than those obtained by using the DMD structures for SUP-C105 both with and without uranyl binding, respectively, further indicating that DMD simulation can consider the interfacial effect to well-tune the crystal structure and therefore provide better candidates to represent the structures of interfacial proteins.

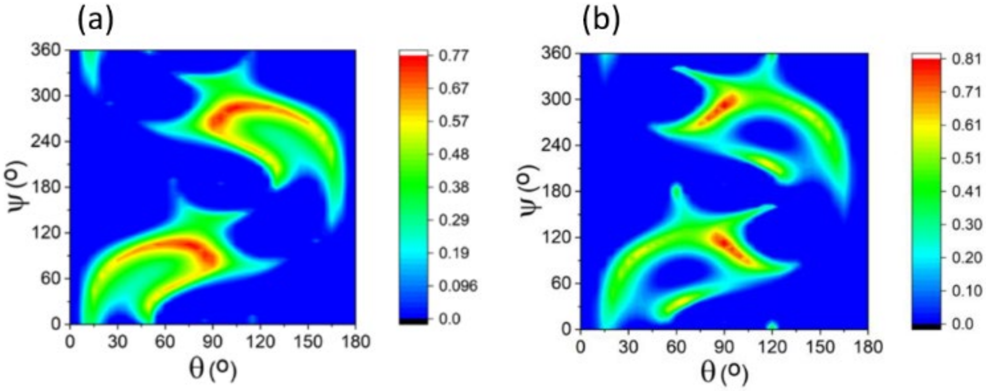

Figure 3-7 Final score map of immobilized SUP-C105 before and after binding with uranyl ions based on (a) 4FZP and (b) 4FZP.

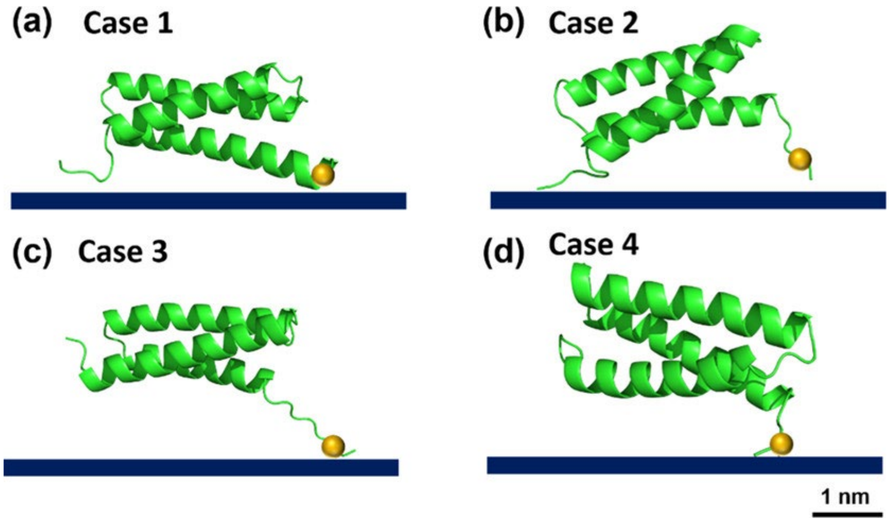

Figure 3-8 Schematics of the SUP-C105 orientation: (a) surface-immobilized SUP-C105 before uranyl ion binding deduced based on the crystal structure and (b) DMD results, (c) surface-immobilized SUP-C105 after uranyl ion binding deduced based on the crystal structure and (d) DMD results.

Previously, circular dichroism (CD) spectroscopy was used to investigate changes in the SUP secondary structure upon uranyl ion binding in solution. It was found that after adding 1 equiv uranyl ion to the SUP solution, the CD spectrum of the SUP molecules noticeably changed, with a loss of the 208 nm characteristic α helix band, implying that SUP changed its conformation or secondary structure dramatically after binding with uranyl ions in the solution state.[8] On the contrary, published crystal structures show that the SUP secondary structure is similar before and after uranyl ion binding.[7] Combined with the SFG and DMD simulation data discussed above, it can be concluded that the structural changes of the surface-immobilized SUP-C105 before and after uranyl ion binding are very small, with no conformational change and only a slight change in orientation. This conclusion is different from the case in free solution and similar to the conclusions reached from crystallography. This is reasonable because the surface-immobilized proteins have much less freedom compared to the proteins in free solution due to strong surface−protein interactions, similar to the constrained state of crystallized protein molecules. Here, our results show that the surface-immobilized SUP-C105 was lying down to a similar extent with or without the presence of uranyl ions (before or after uranyl ion binding) due to the strong restriction and hydrophobic interaction created by the SAM surface.

3.4 Conclusion

In this study, a more systematic method, combining experimental measurements of SFG vibrational spectroscopy and computational interpretation (MD simulations and Hamiltonian approach for SFG spectra calculations) was developed and applied to understand both the conformation and orientation of the surface-immobilized SUP-C105 on the SAM surface. Because of the strong hydrophobic interactions between SUP and the SAM surface, immobilized SUP-

C105 adopts a lying-down orientation. The orientation of the surface-immobilized SUP varies slightly upon uranyl ion binding, but the effects are not substantial. This is in contrast to the effect of uranyl ion binding on free SUP in solution (where large secondary structure changes were observed) and is similar to that of the crystal structures of SUP before and after uranyl ion binding (where there is almost no change). This research suggests that by using a less hydrophobic surface for SUP immobilization or by mutating some hydrophobic amino acids of SUP with hydrophilic ones (such amino acids should be far away from the binding domain so that the mutation would not interfere with the uranyl ion binding), the SUP−surface hydrophobic interaction should be reduced, which may favor a standing-up pose of the SUP to better expose uranyl ion binding sites.

The developed Hamiltonian method has higher accuracy than the previous developed bond additivity model because it considers more vibrational couplings into the calculation. It considers the contributions of all the amino acids to the SFG amide I spectrum, which enables the comparison between the experimentally collected SFG measurements and calculated spectra more accurately. In addition, MD simulation was used to model the interfacial structures of SUP-C105 on the SAM surface. The output structures from the MD simulation can be used as inputs to the Hamiltonian program to calculate the theoretical SFG response of these structures, then these calculations can be compared with the experimental observations to determine both the conformation and orientation of proteins at interfaces. This systematic method is general and can be used to study a wide range of interfacial proteins and peptides.

3.5 References

(1) Nivens, D. A.; Zhang, Y.; Angel, S. M. Detection of Uranyl Ion via Fluorescence Quenching and Photochemical Oxidation of Calcein. *Journal of Photochemistry and*

Photobiology A: Chemistry **2002**, *152* (1–3), 167–173. https://doi.org/10.1016/S1010-6030(02)00229-0.

(2) Lu, G.; Forbes, T. Z.; Haes, A. J. Evaluating Best Practices in Raman Spectral Analysis for Uranium Speciation and Relative Abundance in Aqueous Solutions. *Anal. Chem.* **2016**, *88* (1), 773–780. https://doi.org/10.1021/acs.analchem.5b03038.

(3) Lu, G.; Johns, A. J.; Neupane, B.; Phan, H. T.; Cwiertny, D. M.; Forbes, T. Z.; Haes, A. J. Matrix-Independent Surface-Enhanced Raman Scattering Detection of Uranyl Using Electrospun Amidoximated Polyacrylonitrile Mats and Gold Nanostars. *Anal. Chem.* **2018**, *90* (11), 6766–6772. https://doi.org/10.1021/acs.analchem.8b00655.

(4) Savina, M. R.; Isselhardt, B. H.; Kucher, A.; Trappitsch, R.; King, B. V.; Ruddle, D.; Gopal, R.; Hutcheon, I. High Useful Yield and Isotopic Analysis of Uranium by Resonance Ionization Mass Spectrometry. *Anal. Chem.* **2017**, *89* (11), 6224–6231. https://doi.org/10.1021/acs.analchem.7b01204.

(5) Becker, A.; Tobias, H.; Mandler, D. Electrochemical Determination of Uranyl Ions Using a Self-Assembled Monolayer. *Anal. Chem.* **2009**, *81* (20), 8627–8631. https://doi.org/10.1021/ac901092t.

(6) Peled, Y.; Krent, E.; Tal, N.; Tobias, H.; Mandler, D. Electrochemical Determination of Low Levels of Uranyl by a Vibrating Gold Microelectrode. *Anal. Chem.* **2015**, *87* (1), 768–776. https://doi.org/10.1021/ac503719r.

(7) Zhou, L.; Bosscher, M.; Zhang, C.; Özçubukçu, S.; Zhang, L.; Zhang, W.; Li, C. J.; Liu, J.; Jensen, M. P.; Lai, L.; He, C. A Protein Engineered to Bind Uranyl Selectively and with Femtomolar Affinity. *Nature Chemistry* **2014**, *6* (3), 236–241. https://doi.org/10.1038/nchem.1856.

(8) Hoarau, M.; Koebke, K. J.; Chen, Z.; Marsh, E. N. G. Probing Metal Ion Discrimination in a Protein Designed to Bind Uranyl Cation With Femtomolar Affinity. *Front. Mol. Biosci.* **2019**, *6*, 73. https://doi.org/10.3389/fmolb.2019.00073.

(9) Hoarau, M.; Badieyan, S.; Marsh, E. N. G. Immobilized Enzymes: Understanding Enzyme – Surface Interactions at the Molecular Level. *Org. Biomol. Chem.* **2017**, *15* (45), 9539–9551. https://doi.org/10.1039/C7OB01880K.

(10) Han, X.; Liu, Y.; Wu, F.-G.; Jansensky, J.; Kim, T.; Wang, Z.; Brooks, C. L.; Wu, J.; Xi, C.; Mello, C. M.; Chen, Z. Different Interfacial Behaviors of Peptides Chemically Immobilized on Surfaces with Different Linker Lengths and via Different Termini. *The Journal of Physical Chemistry B* **2014**, *118* (11), 2904–2912. https://doi.org/10.1021/jp4122003.

(11) Xiao, M.; Jasensky, J.; Gerszberg, J.; Chen, J.; Tian, J.; Lin, T.; Lu, T.; Lahann, J.; Chen, Z. Chemically Immobilized Antimicrobial Peptide on Polymer and Self-Assembled Monolayer Substrates. *Langmuir* **2018**, *34* (43), 12889–12896. https://doi.org/10.1021/acs.langmuir.8b02377.

(12) Zou, X.; Wei, S.; Jasensky, J.; Xiao, M.; Wang, Q.; Brooks, C. L.; Chen, Z. Molecular Interactions between Graphene and Biological Molecules. *J. Am. Chem. Soc.* **2017**, *139* (5), 1928–1936. https://doi.org/10.1021/jacs.6b11226.

(1) Nivens, D. A.; Zhang, Y.; Angel, S. M. Detection of Uranyl Ion via Fluorescence Quenching and Photochemical Oxidation of Calcein. *Journal of Photochemistry and Photobiology A: Chemistry* **2002**, *152* (1–3), 167–173. https://doi.org/10.1016/S1010-6030(02)00229-0.

(2) Lu, G.; Forbes, T. Z.; Haes, A. J. Evaluating Best Practices in Raman Spectral Analysis for Uranium Speciation and Relative Abundance in Aqueous Solutions. *Anal. Chem.* **2016**, *88* (1), 773–780. https://doi.org/10.1021/acs.analchem.5b03038.

(3) Lu, G.; Johns, A. J.; Neupane, B.; Phan, H. T.; Cwiertny, D. M.; Forbes, T. Z.; Haes, A. J. Matrix-Independent Surface-Enhanced Raman Scattering Detection of Uranyl Using Electrospun Amidoximated Polyacrylonitrile Mats and Gold Nanostars. *Anal. Chem.* **2018**, *90* (11), 6766–6772. https://doi.org/10.1021/acs.analchem.8b00655.

(4) Savina, M. R.; Isselhardt, B. H.; Kucher, A.; Trappitsch, R.; King, B. V.; Ruddle, D.; Gopal, R.; Hutcheon, I. High Useful Yield and Isotopic Analysis of Uranium by Resonance Ionization Mass Spectrometry. *Anal. Chem.* **2017**, *89* (11), 6224–6231. https://doi.org/10.1021/acs.analchem.7b01204.

(5) Becker, A.; Tobias, H.; Mandler, D. Electrochemical Determination of Uranyl Ions Using a Self-Assembled Monolayer. *Anal. Chem.* **2009**, *81* (20), 8627–8631. https://doi.org/10.1021/ac901092t.

(6) Peled, Y.; Krent, E.; Tal, N.; Tobias, H.; Mandler, D. Electrochemical Determination of Low Levels of Uranyl by a Vibrating Gold Microelectrode. *Anal. Chem.* **2015**, *87* (1), 768–776. https://doi.org/10.1021/ac503719r.

(7) Zhou, L.; Bosscher, M.; Zhang, C.; Özçubukçu, S.; Zhang, L.; Zhang, W.; Li, C. J.; Liu, J.; Jensen, M. P.; Lai, L.; He, C. A Protein Engineered to Bind Uranyl Selectively and with Femtomolar Affinity. *Nature Chemistry* **2014**, *6* (3), 236–241. https://doi.org/10.1038/nchem.1856.

(8) Hoarau, M.; Koebke, K. J.; Chen, Z.; Marsh, E. N. G. Probing Metal Ion Discrimination in a Protein Designed to Bind Uranyl Cation With Femtomolar Affinity. *Front. Mol. Biosci.* **2019**, *6*, 73. https://doi.org/10.3389/fmolb.2019.00073.

(9) Hoarau, M.; Badieyan, S.; Marsh, E. N. G. Immobilized Enzymes: Understanding Enzyme – Surface Interactions at the Molecular Level. *Org. Biomol. Chem.* **2017**, *15* (45), 9539–9551. https://doi.org/10.1039/C7OB01880K.

(10) Han, X.; Liu, Y.; Wu, F.-G.; Jansensky, J.; Kim, T.; Wang, Z.; Brooks, C. L.; Wu, J.; Xi, C.; Mello, C. M.; Chen, Z. Different Interfacial Behaviors of Peptides Chemically Immobilized on Surfaces with Different Linker Lengths and via Different Termini. *The Journal of Physical Chemistry B* **2014**, *118* (11), 2904–2912. https://doi.org/10.1021/jp4122003.

(11) Xiao, M.; Jasensky, J.; Gerszberg, J.; Chen, J.; Tian, J.; Lin, T.; Lu, T.; Lahann, J.; Chen, Z. Chemically Immobilized Antimicrobial Peptide on Polymer and Self-Assembled Monolayer Substrates. *Langmuir* **2018**, *34* (43), 12889–12896. https://doi.org/10.1021/acs.langmuir.8b02377.

(12) Zou, X.; Wei, S.; Jasensky, J.; Xiao, M.; Wang, Q.; Brooks, C. L.; Chen, Z. Molecular Interactions between Graphene and Biological Molecules. *J. Am. Chem. Soc.* **2017**, *139* (5), 1928–1936. https://doi.org/10.1021/jacs.6b11226.

(13) Zheng, S.; Javidpour, L.; Sahimi, M.; Shing, K. S.; Nakano, A. SDMD: An Open Source Program for Discontinuous Molecular Dynamics Simulation of Protein Folding and Aggregation. *Computer Physics Communications* **2020**, *247*, 106873. https://doi.org/10.1016/j.cpc.2019.106873.

(14) Zheng, S.; Sajib, M. S. J.; Wei, Y.; Wei, T. Discontinuous Molecular Dynamics Simulations of Biomolecule Interfacial Behavior: Study of Ovispirin-1 Adsorption on a

Graphene Surface. *J. Chem. Theory Comput.* **2021**, *17* (3), 1874–1882. https://doi.org/10.1021/acs.jctc.0c01172.

(15) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* **2008**, *16* (7), 1010–1018. https://doi.org/10.1016/j.str.2008.03.013.

(16) Wei, S.; Zou, X.; Tian, J.; Huang, H.; Guo, W.; Chen, Z. Control of Protein Conformation and Orientation on Graphene. *Journal of the American Chemical Society* **2019**, *141* (51), 20335–20343. https://doi.org/10.1021/jacs.9b10705.

(17) Xiao, M.; Wei, S.; Chen, J.; Tian, J.; Brooks III, C. L.; Marsh, E. N. G.; Chen, Z. Molecular Mechanisms of Interactions between Monolayered Transition Metal Dichalcogenides and Biological Molecules. *J. Am. Chem. Soc.* **2019**, *141* (25), 9980–9988. https://doi.org/10.1021/jacs.9b03641.

(18) Wei, S.; Knotts, T. A. A Coarse Grain Model for Protein-Surface Interactions. *The Journal of Chemical Physics* **2013**, *139* (9), 095102. https://doi.org/10.1063/1.4819131.

Chapter 4 Determine GB1 Conformation and Orientation at Solid/Liquid Interface

4.1 Attribution Information

Materials covered in this chapter were reproduced from Guo, W.; Lu, T.; Crisci, R.; Nagao, S.; Wei, T.; Chen, Z. Determination of Protein Conformation and Orientation at Buried Solid/Liquid Interfaces. *Chem. Sci.* **2023**, *14 (11)*, 2999–3009. Copyright © 2023 Royal Society of Chemistry. This research was in collaboration with Prof. Tao Wei from Howard University and Dr. Satochi Nago from University of Hyogo.

S.N. contributed to the protein sample preparation and characterization. W.G. conducted SFG and CD experiments, spectra fitting, atomistic MD simulations and analysis, Hamiltonian calculations, and spectra matching. W.G. and T.L. contributed to the development of the Hamiltonian program.

4.2 Background and Motivation

In Chapter 3, a systematic methodology, which combines the SFG measurement, Hamiltonian calculation and MD simulation, was developed and used to study the conformations and orientations of proteins and peptides at solid/liquid interfaces. Here in this chapter, this method will be further optimized to increase the accuracy of both experimental observation and data analysis. Isotope labeling method was applied to selectively label the studied proteins or peptides to increase the independent experimental measurements. Atomistic MD simulation was used to replace the DMD simulation to monitor the detailed dynamics of the interfacial proteins and peptides and to increase the accuracy of the entire modeling process. The parameter comparison score system used to quantify the deviation between the calculated SFG spectra and the experimental collected SFG spectra was further developed to a point-by-point comparison score system as described below. In this chapter, the systematic method with the abovementioned optimizations was used to examine the conformation and orientation of the adsorption of the B_1 domain of protein G (GB1) and its mutant Q32A, N35A GB1 onto the polystyrene (PS) thin film surface. The result shows that the protein orientation can be manipulated on a specific surface by rationally mutating some key amino acids.

4.2.1 B_1 Domain of Protein G (GB1)

GB1 is a small protein of 56 residues. Although small, GB1 forms two anti-parallel β-sheet structures near each terminus and an α-helix structure in the middle. Such a rich secondary structure with small size makes GB1 an excellent model for many protein studies.[1–4] Previously, we studied the behavior of GB1 on a graphene surface and showed that residues Q32 and N35, located at the helical portion of GB1, play an important role in the strong protein – surface interaction with graphene, leading to the denaturation of the protein.[5] The mutation of these two residues into alanine could greatly reduce the protein – graphene interaction and retain the native GB1 structure on graphene, preventing the denaturation of GB1.[5] In the above study, we deduced the mutant GB1 orientation based on the GB1 crystal structure (PDB entry: 3gb1), and used the SFG signals from the α-helical domain in the orientation analysis. Here in this study, we used the I-TASSER homology simulation[6–8] to model the theoretical structure of the Q32A, N35A mutated GB1 to further differentiate the conformation between the wild-type GB1 (WT GB1) and mutant GB1 (MT GB1). The atomistic MD simulation[9–11] was adopted to model the protein adsorption processes onto a PS thin film using the initial protein structures as the crystal structure for WT GB1 and the simulated structure for MT GB1, respectively. The refined interfacial structures obtained from the atomistic MD simulation will then input into the Hamiltonian program for SFG spectra calculations and comparisons.

4.2.2 Isotope Labeling

It has been demonstrated that isotope labeling can generate many more independent spectral measurements by distinguishing originally overlapping signals.[12–14] For example, isotopic substitution of ^{13}C in the backbone of proteins could shift the amide I band by ~ 40 cm^{-1} for a vibrational spectroscopic technique.[15,16] Selective isotope labeling can also help vibrational spectroscopy to study proteins and peptides in residue-by-residue resolution.[16–18] Therefore, isotope labeling can facilitate the determination of complex structures of biological molecules.

In the current study, we isotope labeled WT GB1 and MT GB1 and applied SFG spectroscopy to record interfacial behavior of both non-isotope-labeled and isotope-labeled samples. It is assumed that isotope labeling has minor effects on protein structure, therefore, both non-isotope-labeled WT GB1 (or MT GB1) and isotope-labeled WT GB1 (or MT GB1) should have the same conformation and orientation when interacting with PS surface. That is to say, even though isotope labeling samples can provide distinct SFG spectral features, the most possible orientation(s) deduced from the Hamiltonian approach for isotope labeled protein samples should be the same as the orientations deduced from the non-isotope-labeled protein samples. Therefore, same final score maps (as a function of orientation) should be expected for the same protein regardless of isotope labeling or not. To determine more reliable matched orientation(s) of WT GB1 (or MT GB1), the overlap region from the final score maps of all the non-isotope-labeled and isotope-labeled WT GB1 (or MT GB1) was found. The highest matching score from the overlap region among all the input structures determines the most likely conformation (the input structure used) and orientation of WT GB1 or MT GB1 at the interface.

4.3 Experimental Materials and Methods

4.3.1 Preparation of WT GB1 and MT GB1 with and without Isotopic Labels

^{13}C-labeled amino acids at the carbonyl position were used to isotope-labeled WT GB1 and MT GB1. Eight protein samples (two WT GB1 samples and six MT GB1 samples) including non-labeled WT GB1 (WT NL), Leu-labeled WT GB1 (WT Leu), non-labeled MT GB1 (MT NL), Leu-labeled MT GB1 (MT Leu), Val-labeled MT GB1 (MT Val), Phe-labeled MT GB1 (MT Phe), Lys-labeled MT GB1 (MT Lys), and Ile-labeled GB1 (MT Ile) were prepared. Detailed protein expression, purification and verification can be found in our previous publication and will not be repeated here.[19] All proteins were stored at -40°C until use. Each protein sample was dissolved in D_2O to a concentration of 0.1 mg/mL for SFG experiments.

4.3.2 Preparation of PS Thin Film

PS thin film was prepared by spin-coating method. Right-angle CaF_2 prisms (Altos Photonics, Bozeman, MT) were cleaned by toluene rinse, N_2 dry, and 10 min plasma cleaning (Plasma Etch Inc. model: PE-50). Solid PS with Mw=60,000 (Scientific Polymer, Inc.) was used to prepare a 3% PS (w/w) solution in anhydrous toluene for the spin-coating (P-6000 spin coater, Speedline Technologies, Franklin, MA). The PS solution was spin-coated onto clean CaF_2 prisms at 2000 rpm for 1 min, forming a PS thin film (~ 150 nm). The PS-coated prisms were annealed at 120°C for 3 hrs and then cooled down to room temperature before use.

4.3.3 SFG Experimental Setup

SFG signals in the protein amide I frequency region (1500 cm^{-1} to 1800 cm^{-1}) were collected from the PS/protein solution interfaces using the experimental near-total reflection geometry presented in Figure 4-1. SFG ssp and ppp spectra were collected for each protein sample after the adsorption process reached equilibrium.

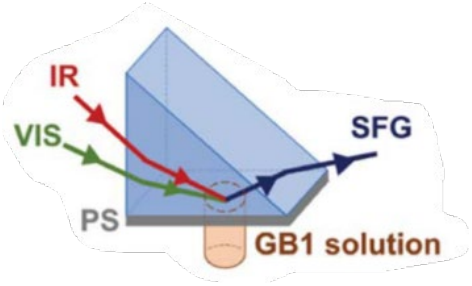

Figure 4-1 Schematic of the SFG prism geometry used in this study to collect SFG spectra from the PS/protein solution interfaces.

4.3.4 Atomistic MD Simulations

Gromacs[20] (version 2019.6) simulation package was used to conduct atomistic MD simulations. The CHARMM 36 force field[21] combined with TIP3P water model was applied to describe the interactions in the system. The Velocity Verlet algorithm was used for the dynamic equations with a time step of 1 fs. The long-range electrostatic interactions were calculated by the particle mesh Ewald (PME) summation with a cutoff of 1.2 nm. A spherical cutoff of 1.2 nm was imposed on Lennard-Jones interactions. Parrinello-Rahman method was used to keep the NPT ensemble at 1 bar. The temperature was maintained at 298.15 K by Berendsen thermostat. A periodic boundary condition was applied to the system, along the X and Y directions only.

For the WT GB1, we used a crystal structure (PDB code: 3gb1) from the Protein Data Bank (http://www.ncbi.nlm.nih.gov/). The MT GB1 protein structure was predicted by I-TASSER[6–8] based on the crystal structure of WT GB1. With amino acids lysine (Lys) protonated, glutamate (Glu) and aspartate (Asp) deprotonated, and uncapped N- and C-terminals, both WT and MT GB1 proteins have a net charge of -4e at pH of 7. Both proteins were relaxed in water at 298.15 K in NVT ensemble before assembled with the PS substrate surface.

An amorphous PS surface was constructed with 78 polystyrene chains. Each chain consists of 20 monomers. The PS surface was initially relaxed in vacuum with a series of simulations of 30 ns for a total of 270 ns, which were carried out using a stepwise heating/annealing protocol starting from 298.15 K, reaching 700 K and then recovering to 298.15 K. Then the surface was exposed to water solution for further relaxation for 150 ns. The equilibrated PS surface is around 6.847×6.797 nm^2 in the X-Y plane with a thickness of 5.506 nm. After that, the hydrated PS surface was assembled with a protein, leaving an initial gap of at least 2.0 nm between the PS surface and the protein. The total dimension of the solvation box along the Z direction is 13.787 nm. Sodium counter ions were added to neutralize the system. The assembled initial configurations of WT GB1/PS and MT GB1/PS systems are shown in Figure 4-2.

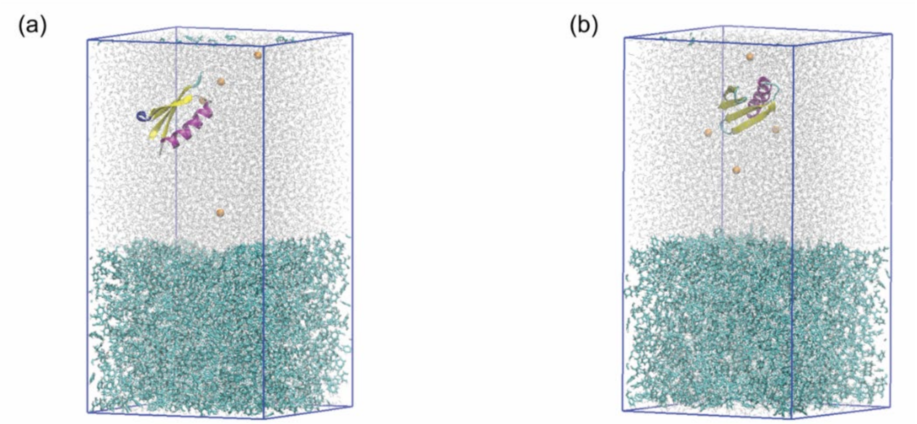

Figure 4-2 Snapshots of initial configurations of (a) a WT GB1 and (b) a MT GB1 on a PS surface in an aqueous environment with counterions (Na^+). The PS molecules are shown in cyan; counterions are shown in orange and water molecules are shown in gray. The thickness of the simulated PS film is around 5.5 nm. According to the

simulation results of PS films[22], the mobile free surface is 2 nm (less than 5.5 nm used in this study), showing that our MD simulation of PS film can be justified.

4.3.5 Hamiltonian Approach and Scoring System

4.3.5.1 Parameters for Hamiltonian Approach

A Hamiltonian matrix was constructed to obtain the SFG signal frequency and the strength of each amide I normal mode of a protein by using the couplings among the local modes of the amide units. Here, the diagonal terms of the Hamiltonian matrix, representing the local mode frequencies, were set to 1640 cm^{-1} for non-isotope-labeled amide units and set to 1610 cm^{-1} for isotope-labeled amide units. The off-diagonal terms of the Hamiltonian matrix were calculated as demonstrated in Chapter 1, equation (1 - 28). The peak width of each amide mode (normal mode) was assumed as 10 cm^{-1} in this study when constructing SFG spectra.

4.3.5.2 Point-by-point Comparison Score System

In Chapter 3, a parameter comparison score system was used to automatically fit the calculated SFG spectra then compare the obtained fitting parameters with the experimental fitting parameters. This score system works well when only a single peak was observed in SFG spectrum. However, when comparing SFG spectra with two or more peaks, the auto-fitting process in the parameter comparison score system is less accurate and less efficient. A new score system, point-by-point score system, was developed to quantify the intensity difference between two spectra with multiple peaks at each frequency effectively and accurately.

In the point-by-point score system, the reconstructed experimental ssp SFG spectra and the calculated ssp SFG spectra were normalized to [0, 1] and meanwhile the reconstructed experimental ppp SFG spectra and the calculated SFG ppp spectra were scaled to maintain the original ppp/ssp intensity ratios (the ratio was determined by the maximum of I_{ppp} divided by the maximum of I_{ssp}). The comparison (difference or diff.) between the reconstructed experimental ssp (or ppp) SFG spectra and the calculated ssp (or ppp) SFG spectra was done by using a point-to-point least square method, as shown in equation (4 - 1):

$$diff. = \sum_{i=1}^{i=301} \sqrt[2]{(I_{exp,i} - I_{cal,i})^2} \tag{4 - 1}$$

where i means the i-th data point in the range from 1500 cm^{-1} to 1800 cm^{-1} (increment of 1 cm^{-1}, 301 data points in one spectrum). $I_{exp,i}$ and $I_{cal,i}$ are the experimental intensity and calculated SFG intensity at the i-th frequency, respectively.

Then the score of the spectral comparison of a protein with a specific orientation is defined with equation (4 - 2):

$$Score_{protein\ x} = \sqrt{\frac{1}{ssp\ diff.} \times \frac{1}{ppp\ diff.}} \tag{4 - 2}$$

The final score of WT GB1 or MT GB1 with a specific orientation can be obtained by the following equation (4 - 3):

$$Final\ Score = \sqrt[n]{\prod_{1}^{n} Score_{protein\ x}} \tag{4 - 3}$$

Here for WT GB1, n = 2 (for two protein samples WT NL, WT Leu). For MT GB1, n = 6 (for 6 protein samples MT NL, MT Leu, MT Val, MT Phe, MT Lys and MT Ile). The final score can be plotted as a function of protein orientation (θ, ψ) for each protein conformation to generate a final score map.

4.4 SFG Results

Figures 4-3 and 4-4 display the experimentally collected and then fitted SFG ssp and ppp spectra for all the eight protein samples (WT NL, WT Leu, MT NL, MT Leu, MT Val, MT Phe, MT Lys and MT Ile) at the PS/protein solution interfaces. In both SFG ssp and ppp spectra of each of the eight protein samples, a positive peak at around 1625 cm^{-1} and a negative peak at around 1655 cm^{-1} were observed, and these two peaks can be assigned to the β-sheet structure and α-helix structure respectively.[2,23] A small shoulder peak at around 1690 cm^{-1} was detected in the ssp spectra of MT GB1 samples and could be assigned to β-turns.[23,24] Another weak peak observed at around 1740 cm^{-1} in both SFG ssp and ppp spectra of all eight samples could be assigned to protein side chains and will not be included in the amide I spectra comparison discussed below.

A peak centered at ~1600 cm^{-1} was dominant in all the SFG ppp spectra, which could be assigned to the C=C stretching mode of the phenyl groups on the PS side chains, or the amide I mode (~80% contributed by the $^{13}C=O$ stretching) of the isotope labeled amide units of the protein. This peak was also detected in the ppp spectra collected from the PS/D_2O (without proteins) interface (Figure 4-5 (b)) and the PS/non-isotope-labeled protein (WT NL or MT NL) solution interfaces (Figures 4-4 (a) and 4-4 (c)). The intensity of this peak was twice more intense than that of a normal protein amide I peak (e.g., a peak assigned to a β-sheet structure or an α-helix structure). Generally, the intensity of an isotope-labeled amide I peak of a protein should be very weak because of the low population of such amide units (which can be proved from the spectral calculation shown below). Therefore, here we believe that the intense peak observed at ~1600 cm^{-1} in the ppp spectra mainly comes from the phenyl groups of the PS side chains, instead of the isotope-labeled units of proteins. We thus decided that it is reasonable to exclude this peak when reconstructing the experimentally measured ppp spectra for later spectral comparisons with the calculated ppp spectra. For SFG, isotope labeling not only shifts the peak center of the amide I signal from the isotope labeled units, it also changes the signal couplings between the isotope labeled units and other non-labeled units, leading to the changes of spectral features and intensities of the amide I signals from the non-labeled units of the protein. Thus, examining the spectral changes of the main protein amide I peaks (from non-labeled β-sheet/turn structure and α-helix structure) should also be able to quantify the isotope labeling effects.

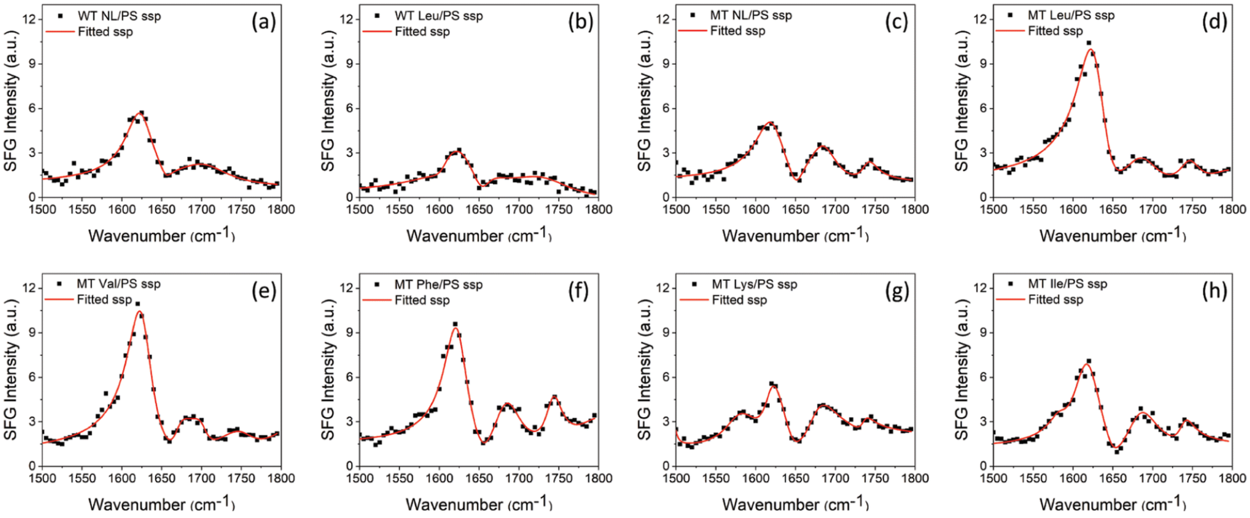

Figure 4-3 SFG ssp spectra collected from proteins (a) WT NL, (b) WT Leu, (c) MT NL, (d) MT Leu, (e) MT Val, (f) MT Phe, (g) MT Lys and (h) MT Ile absorbed at the PS/protein solution interfaces. Black dots are experimental data points and red lines are fitted spectra.

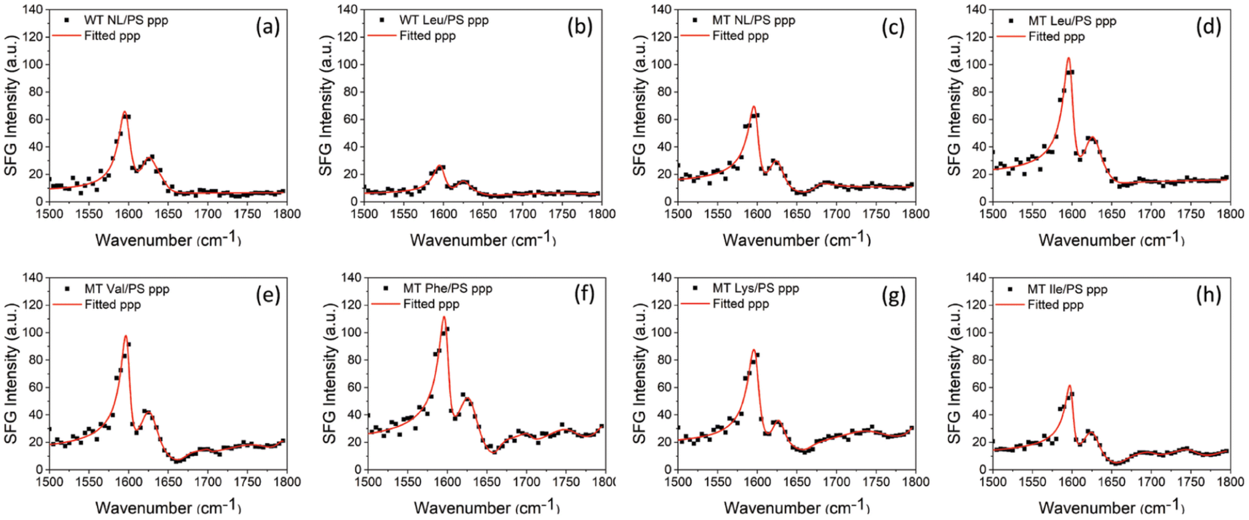

Figure 4-4 SFG ppp spectra collected from proteins (a) WT NL, (b) WT Leu, (c) MT NL, (d) MT Leu, (e) MT Val, (f) MT Phe, (g) MT Lys and (h) MT Ile absorbed at the PS/protein solution interfaces. Black dots are experimental data points and red lines are fitted spectra.

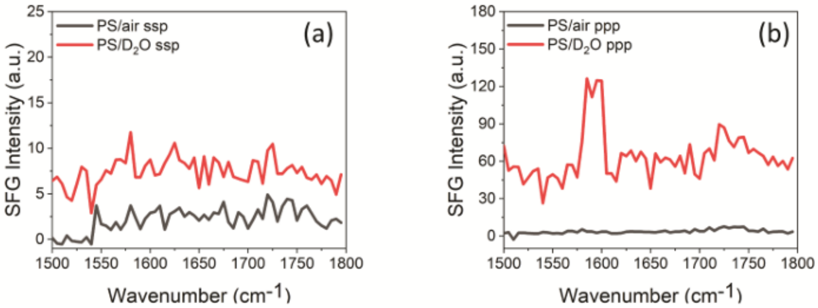

Figure 4-5 SFG ssp (a) and ppp (b) spectra of polystyrene/air interface (black) and polystyrene/D_2O interface (red).

4.5 Atomistic MD Simulation Results

Here atomistic MD simulations were conducted to study the WT GB1 and MT GB1 samples without introducing isotope-labeling. We assume that isotope-labeling has negligible effects on the protein structure in solution and protein – PS interactions. Therefore, the non-labeled and isotope labeled WT NL (or MT NL) samples have the same structure (same conformation and orientation) at the PS/protein solution interface. To ensure that we capture the equilibrated protein – polymer interaction, the MD run was performed for 1000 ns. Figure 4-6 shows the distance between the protein (WT GB1 or MT GB1) and the PS surface as a function of time. According to this distance plot, the adsorption process happened quickly for both cases (~20 ns for WT GB1 and ~ 10 ns for MT GB1). For MT GB1 and WT GB1, the mean values and fluctuations of the radius of gyration (R_g) and the root mean square deviation (RMSD) are similar over the course of simulation (shown in the Figure 4-7 and Figure 4-8), indicating that the mutation of Q32A and N35A would not cause substantial conformation changes of MT GB1 due to the GB1 – PS interactions. The structures of WT GB1 and MT GB1 were similar and stable after GB1's landing on the PS surface. The secondary structure revolution maps (Figure 4-9) show that WT GB1 and MT GB1 have similar conformations along the entire simulation time.

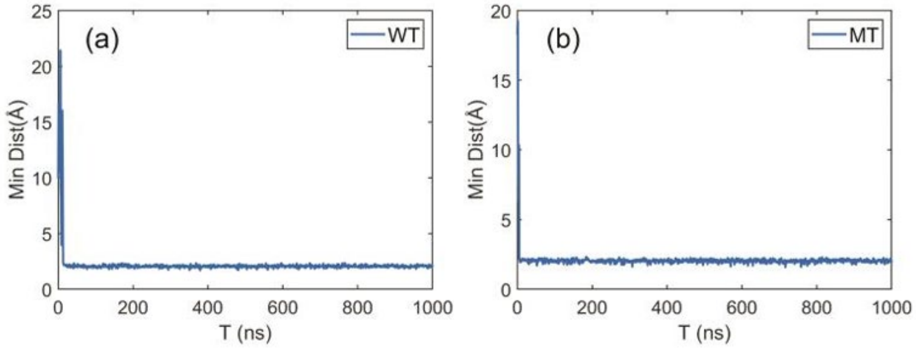

Figure 4-6 Distance between the protein (a) WT GB1 or (b) MT GB1 and the PS surface as a function of time in simulation. WT GB1 and MT GB1 were released at ~ 2.0 nm from the PS surface. The adsorption processes happened quickly for both cases (~20 ns for WT GB1 and ~ 10 ns for MT GB1). After landing on the PS surface, the protein-surface distance has almost no change for each case along the entire simulation period.

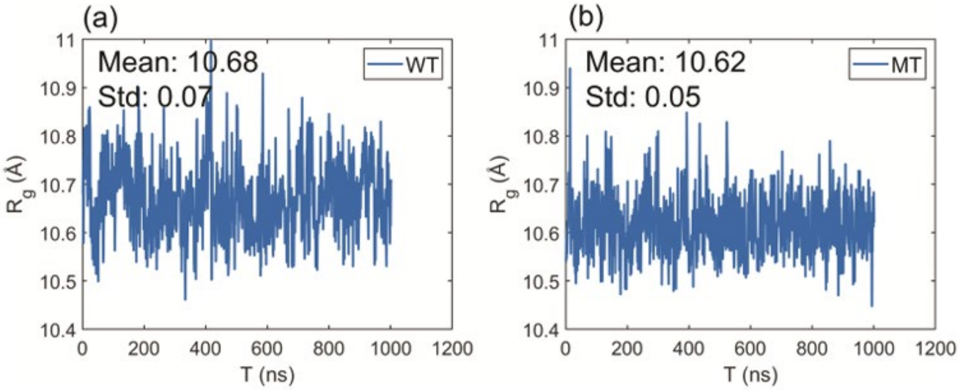

Figure 4-7 Radius of gyration (Rg) fluctuations of (a) WT GB1 and (b) MT GB1 along the simulation time. The mean Rg value of WT GB1 is similar to the mean Rg value of MT GB1, with slightly larger variations. This indicates that the atom distribution along the principal axis of inertia of WT GB1 is similar to that of MT GB1 along the entire simulation period. This result infers that the mutation of Q32A and N35A has little to no effect on the folding of GB1.

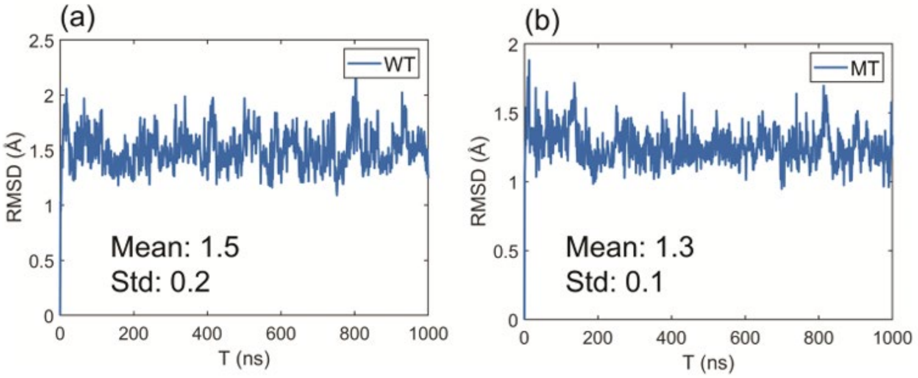

Figure 4-8 Root mean square deviation (RMSD) between the protein structure at 0 ns and the protein structure from 0 ns to 1000 ns. (a) is the RMSD plot of WT GB1 and (b) is the RMSD plot of MT GB1. WT GB1 and MT GB1 possess similar mean value and variation of RMSDs, indicating that the mutation of Q32A and N35A causes little to no conformation changes due to the GB1 – PS interactions.

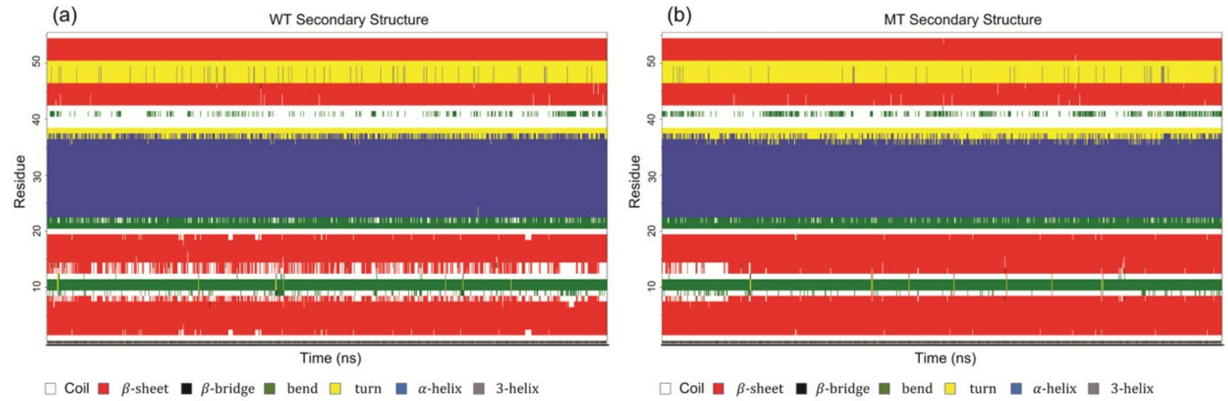

Figure 4-9 Secondary structure map of (a) WT GB1 and (b) MT GB1 as a function of the WT GB1 and MT GB1 have similar conformations in the entire simulation time. The major conformational difference is that for MT GB1, N8, L12 and K13 formed β sheet structure when stabilized, while for WT GB1 these three residues could sometimes be coil structure.

To quantify the orientation fluctuation of the α-helical portion of WT GB1 and MT GB1, we defined a tilt angle θ_a to evaluate the angle between the z axis (the PS/protein solution interface normal) and the sum of amide I vectors from residue 21 to residue 36 (the α-helical portion, pointing from near the N- terminal to near the C-terminal) of protein GB1. The time-dependent θ_a plots of WT GB1 and MT GB1 (shown in the Figure 4-10) indicated that θ_a reached equilibrium after 600 ns for both proteins. However, the θ_a distributions of the last 400 ns of the simulations of WT GB1 and MT GB1 exhibited different mean values with similar standard deviations, indicating that the α-helical structures in WT GB1 and MT GB1 adopted different preferred orientations on the PS surface although they had similar conformations. It is interesting to observe that only mutating two amino acids of GB1 could change the protein orientation on PS substantially. This suggests that the mutation of Q32A and N35A alters the interactions between the GB1 protein and the PS surface, and the planar residues Q32 and N35 play an important role in such interactions. Similar phenomena were previously observed from WT GB1 and MT GB1 on graphene[5], which will be discussed in more detail later.

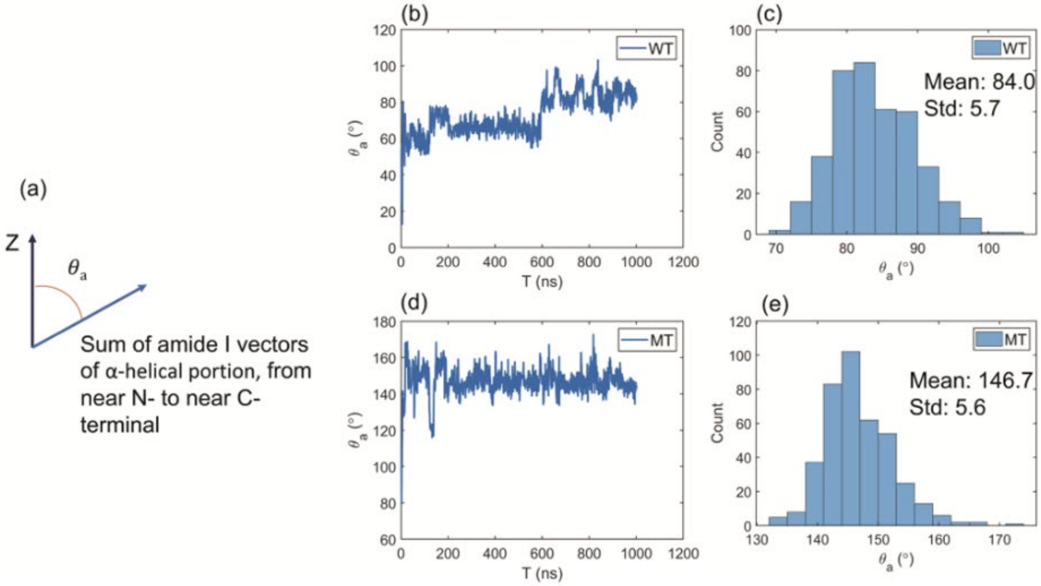

Figure 4-10 (a) θa is defined as the angle between the z axis and the sum of the amide I vectors of α-helix (from residue 21 to residue 36, pointing from near N-terminal to near C-terminal) of protein GB1. (b) and (d) are the changes of θa of the simulated WT GB1 and MT GB1, respectively, from 0 ns to 1000 ns with the increment of 1 ns. (c) and (e) are the θa distributions of last 400 ns of the simulations of WT GB1 and MT GB1, respectively.

4.6 Hamiltonian Matching Results

4.6.1 Spectral Matching for WT GB1

To determine the interfacial protein conformation, for the Hamiltonian spectral analysis, protein configurations in the last 400 ns of the simulation were used. Due to the fluctuations and roughness of the soft surfaces of the PS film and protein GB1, the orientation of a landed protein can have small changes at a microsecond temporal scale. The atomistic MD simulation (≤ 1 μs) might not catch the slow tumbling (i.e. side rotation of a non-principle axis) of an adsorbed GB1 molecule on the PS surface.[11,25–28] In contrast, SFG experiments measure the averaged conformations of multiple proteins at the macro-scale. To overcome the defect of insufficient sampling in MD simulations, for the Hamiltonian SFG spectra calculation, we first rotated each MD simulated configuration of WT GB1 in the last 400 ns (1 ns per configuration) to (θ, ψ) from its initial orientation (0°, 0°) to compensate the orientation fluctuations ($\theta \in [0°, 180°]$, $\psi \in [0°, 360°]$, 5° as a step). It is worth noting that, although an orientation grid search was conducted, the final deduced orientation which best matches the experimentally measured spectra should be close to its initial state to match the atomistic MD simulation result. All 400 protein configurations and their rotated structures were used to calculate the ssp and ppp SFG spectra for WT NL and WT Leu (WT NL and WT Leu used the same protein structure but different Hamiltonian matrix). The calculated spectra were scored by comparing them to the experimentally reconstructed spectra for both spectral features and relative ssp/ppp spectra intensity. A heat map showing such matching scores as a function of protein orientation can be obtained for each of the 400 simulated configurations. The overall heat map for each simulated configuration can be obtained by combining the heat maps of the non-labeled and all isotope labeled samples of each configuration. The simulated configuration with the overall heat map with the maximum matching score among the 400 generated overall heat maps was chosen as the most likely conformation. The orientation with the highest matching score from the chosen heat map was selected as the most likely orientation. For WT GB1, considering the results for both WT NL and WT Leu, it was found that among all the possible candidates, the simulated WT GB1 structure at 781 ns with the orientation of (30°, 50°) or (150°, 230°) has the highest score (0.66). The heat map showing the matching scores of WT GB1 based on the simulated structure at 781 ns as a function of protein orientation is shown in Figure 4-11. Clearly the heat map shows that the orientations (30°, 50°) and (150°, 230°) have the highest scores. In this study we performed homodyne SFG measurements to measure SFG signal intensity, thus the orientations of (θ, ψ) and (180° - θ, 180° + ψ) could not be differentiated. Here (30°, 50°) and (150°, 230°) have the opposite absolute orientations.

For the WT GB1 with the orientation of the highest matching score, we plotted the calculated SFG ssp and ppp spectra from WT NL (Figure 4-11 (b)) and WT Leu (Figure 4-11 (c)) proteins. For comparison, re-constructed experimentally detected SFG spectra are also displayed in Figure 4-11 (b) and 4-11 (c). From the spectral comparison results, one could find that the spectral features (peak center, peak intensity ratio of ppp/ssp, peak width, etc.) could be well-matched between the calculations and the experimental measurements, except the isotope-labeled peak at ~1600 cm^{-1} in the WT Leu's case for the ppp spectrum. The calculated ppp spectrum of WT Leu shows the isotope labeled signal at ~1600 cm^{-1}, which is missing in the reconstructed ppp spectrum. As we mentioned above, the mismatch could be due to the removal of the ~1600 cm^{-1} peak during the experimental spectral reconstruction since such a peak could come from the PS surface. Although the isotope-labeled peak was removed during the experimental spectra reconstruction due to its overlapping with the PS signal, the match based on the main amide I peaks (from β-sheet/turn structure and α-helix structure) is still satisfactory.

Figures 4-11 (d) and 4-11 (e) display the orientation visualizations of the deduced most likely (or best matched) orientation (30°, 50°) and the orientation obtained from the MD simulation without rotation (0°, 0° - the input structure for the Hamiltonian calculation) respectively. The best matched orientation (30°, 50°) showed that the α-helical structure lies down on the PS surface, while the β-sheet portion faces towards the solvent. This orientation is similar to the initial (0°, 0°) orientation obtained from the MD simulation at 781 ns without rotation, indicating that the orientation deduced from the SFG experimental data corroborates the MD simulation data. Furthermore, for the deduced best matched orientation, the θ_a value was deduced to be 90.2°. This value falls within the θ_a distribution of simulated WT GB1 for the last 400 ns of the MD simulations (Figure 4-10), further demonstrating that the deduced orientation from the experimental data well matches the simulation results. Meanwhile, the other best matched opposite absolute orientation (150°, 230°) with the α-helical structure far from the PS surface, was very different from its initial simulated orientation, which should be excluded because it failed to match the atomistic MD simulation results.

The close contact of the alpha-helical structure of WT GB1 with the PS surface is reasonable because of the strong interactions between the PS phenyl groups and the “planar” amino acids on the helix. More detailed discussion on the GB1 – PS interactions will be presented below.

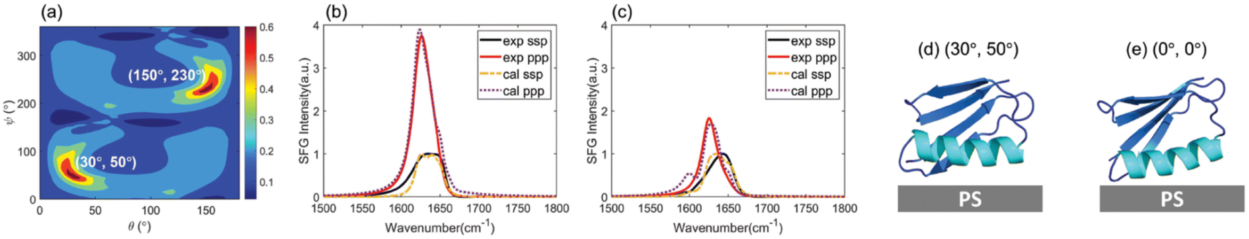

Figure 4-11 (a) Final score map of the spectral matching between the reconstructed experimentally collected WT GB1 SFG spectra (after deconvoluting the non-resonant contribution in ssp and ppp spectra and PS contribution in the ppp spectra) and the calculated WT GB1 SFG spectra as a function of protein orientation based on the simulated WT GB1 structure at 781 ns. The orientations at (30°, 50°) and (150°, 230°) in the map possess the highest matching scores (0.66). The spectral comparisons (b) between the reconstructed WT NL SFG spectra and the calculated WT NL spectra (based on the simulated structure at 781 ns) and (c) between the reconstructed WT Leu SFG spectra and the calculated WT Leu spectra (based on the simulated structure at 781 ns) at an orientation of (30°, 50°) (or (150°, 230°)) which have the highest matching scores. The orientation visualizations of (d) WT GB1 (with the 781 ns simulation structure) at (30°, 50°) and (e) WT GB1 (with the 781 ns simulation structure) at (0°, 0°). The (0°, 0°) orientation is the protein orientation obtained from the MD simulation result without further rotating the protein.

4.6.2 Spectral Matching for MT GB1

We then used a similar approach to that discussed above for the WT GB1 study to deduce the MT GB1 structure on PS. The last 400 MT GB1 configurations obtained from the MD simulations and their rotated structures were used to calculate the ssp and ppp SFG spectra for MT NL and isotope-labeled MT GB1 samples. An overall heat map showing the matching scores for each simulated configuration can be obtained by considering the matching qualities of the spectra generated from the non-labeled and all the isotope labeled samples including MT NL, MT Leu, MT Val, MT Phe, MT Lys and MT Ile. The heat map with the maximum matching score among all the heat maps of the 400 configurations (with rotations) was selected, as shown in Figure 4-12.

The best matched result for MT GB1 was found to be the simulated structure at 972 ns with an orientation of (30°, 100°) or (150°, 280°). The comparisons between the calculated spectra from different samples at this orientation with this conformation and the reconstructed experimental data are shown in Figure 4-12 (b) – (g). The best matched structure (30°, 100°) and the simulated structure (0°, 0°) are displayed in Figure 4-12 (h) and 4-12 (i). The initial orientation of the simulated MT GB1 at 972 ns without rotation (Figure 4-12 (i)) showed that the α-helix was tilted up (θ_a = 144.2°) on the PS surface and the anti-parallel β-sheet near the N-terminal was close to the PS surface, which is similar to its initial orientation with θ_a = 148.2°, while the anti-parallel β-sheets lie down towards the surface slightly more. The other best matched MT GB1 simulated structure at 972 ns of (150°, 280°) has an opposite absolute orientation to its initial orientation obtained from simulation with a θ_a of 31.8° and the anti-parallel β-sheet near the N-terminal is far away from the surface. Because of such discrepancies between the simulated orientation and experimentally deduced orientation, we believe that the orientation of (150°, 280°) is unlikely the possible orientation and should be excluded.

From our above analysis, clearly WT GB1 and MT GB1 adopt different structures on the PS surface. The difference could be confirmed by directly looking at their SFG spectra. The SFG spectral differences between WT NL and WT Leu (Figure 4-11) are not the same as the spectral differences between MT NL and MT Leu (Figure 4-12). For example, the SFG ppp/ssp intensity ratio (~ 1625 cm^{-1}) of WT NL (~ 4) is twice as the ratio of WT Leu (~ 2), while this ratio is ~ 4 for both MT NL and MT Leu. Such spectral differences inferred the structure variations (i.e.

different conformations and/or orientations) of WT GB1 and MT GB1 at the PS/protein solution interface.

For structure variations, we first compare their deduced best matched conformations. The conformation difference between WT GB1 781 ns and MT GB1 972 ns is not obvious with RMSD = 2.0 Å (shown in the Figure 4-8), but it is larger than the conformational fluctuations of both WT GB1 and MT GB1 along the entire simulation time (RMSD mean of WT is 1.5 Å and RMSD mean of MT is 1.3 Å). We believe that the conformation difference between WT GB1 and MT GB1 is real. For orientation differences, the α-helical component of WT GB1 is more or less lying down on the PS surface, which is substantially different from the standing-up pose of the α-helical structure of MT GB1.

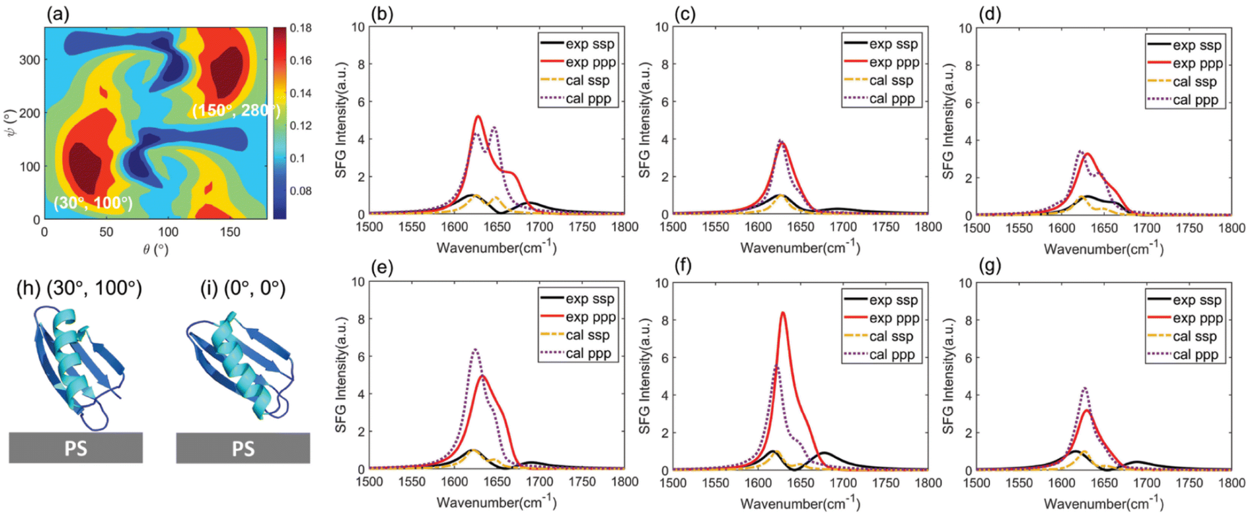

Figure 4-12 (a) Final score map of spectral matching between the reconstructed experimentally collected MT GB1 SFG spectra (after deconvoluting the non-resonant contribution in ssp and ppp spectra and PS contribution in the ppp spectra) and the calculated MT GB1 SFG spectra as a function of orientation based on the simulated MT GB1 structure at 972 ns. The orientations at (30°, 100°) and (150°, 280°) shown in the map possess the highest matching score (0.19). The spectral comparisons between the reconstructed experimental spectra and calculated spectra using the simulated MT structure at 972 ns with an orientation of (30°, 100°) (or (150°, 280°)) for (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. The orientation visualizations of (h) MT GB1 with a simulated structure at

972 ns with the most likely orientation at (30°, 100°) and (i) MT GB1 with a simulated structure at 972 ns without rotation at (0°, 0°).

4.6.3 Crystal Structure for SFG Spectral Calculation

To confirm that MD simulated structure may provide a better structural input for SFG spectra calculation, crystal structure of WT GB1 (PDB ID: 3gb1) was also used to calculate the SFG ssp and ppp spectra of WT GB1 and MT GB1 using the Hamiltonian method for reference and comparison. The final score heat maps obtained based on the crystal structures of WT GB1 and MT GB1, along with the spectral comparisons of the top-ranking orientations are shown in Figure 4-13 and Figure 4-15, respectively. The best matched results (Figure 4-15 (a) and (b)) of WT GB1 determined using the crystal structure input showed lying-down orientations, while for MT GB1, the best matched protein orientations (Figure 4-15 (d) and (e)) are standing-up poses. Although using the crystal structure for spectral calculations could still differentiate the orientations between WT GB1 and MT GB1 on the PS surface, the score of best matched orientations of WT GB1 (0.56) or MT GB1 (0.15) is lower than that obtained by using the rotated simulated structures of WT GB1 (0.66) or MT GB1 (0.19), respectively. This indicates that the analysis accuracy by using the crystal structure is not as high as that obtained by using the simulated GB1 structures. Thus, we believe that the atomistic MD simulation could capture the protein structure deviation from the crystal structure when considering the protein – surface interaction, improving the accuracy of the SFG data analysis.

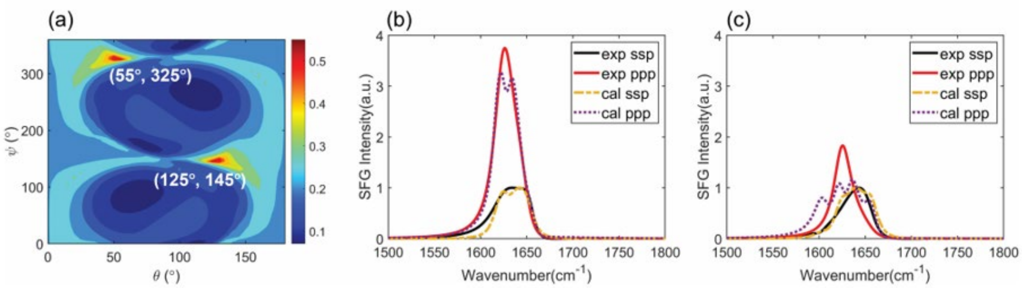

Figure 4-13 (a) Final score heat map of spectral matching between the reconstructed WT SFG spectra and the calculated WT SFG spectra by rotating the crystal structure 3gb1. The orientations at (55°, 325°) and (125°, 145°)

possess highest matching scores (score = 0.56). The spectral comparisons between the reconstructed experimental spectra and the calculated spectra using 3gb1 at orientation of (55°, 325°) (or (125°, 145°)) of (b) WT NL and (c) WT Leu. By using the crystal structure, the best matched calculated ppp spectra could not match the reconstructed experimental ppp spectra well, leading to lower matching scores.

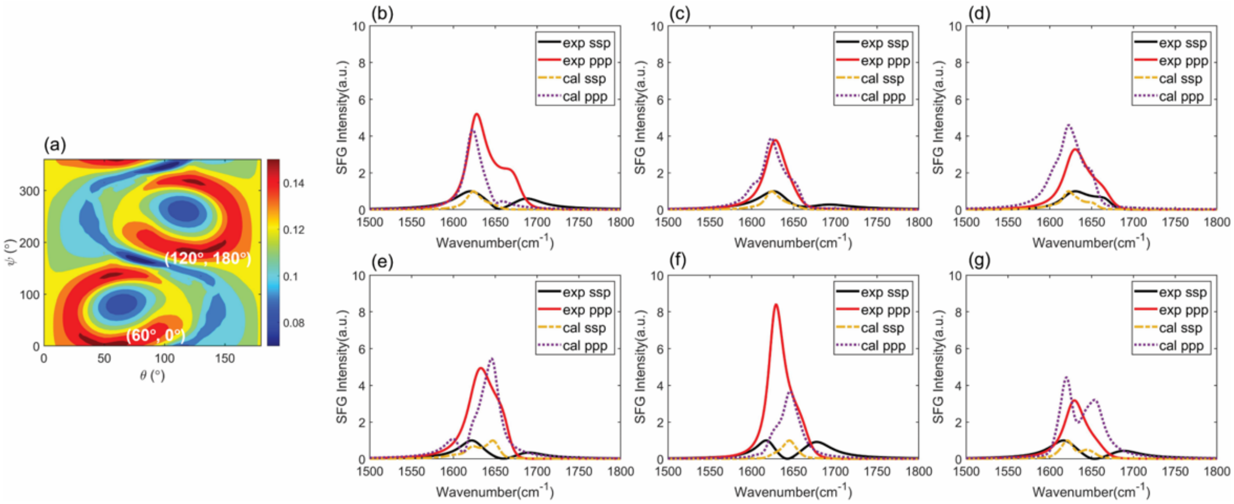

Figure 4-14 (a) Final score heat map between the reconstructed experimental MT SFG spectra and the calculated MT SFG spectra using crystal structure 3gb1. The orientations at (60°, 0°) and (120°, 180°) possess highest matching scores (score = 0.15). The spectral comparisons between reconstructed experimental spectra and calculated spectra by rotating 3gb1 at orientation of (60°, 0°) (or (120°, 180°)) of the case of (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. By using the crystal structure, the calculated spectra and the experimental reconstructed spectra could not match well. For example, almost all the ppp spectra except MT Leu could not match well in peak intensity, peak center, etc., leading to lower matching scores.

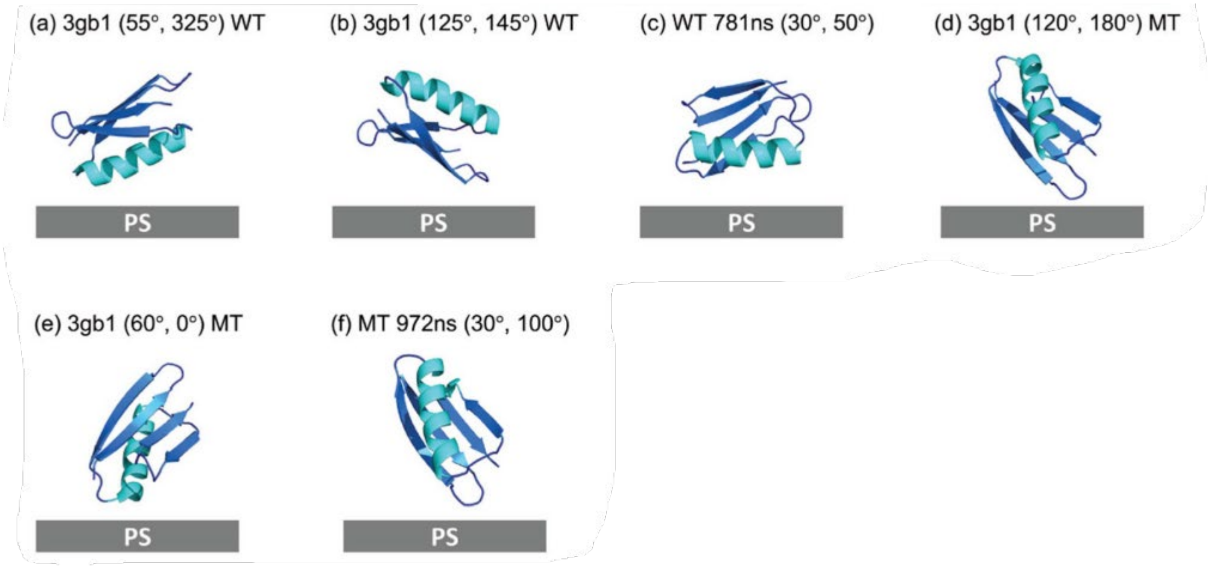

Figure 4-15 Orientation comparisons of GB1 at the PS/protein solution interfaces: (a) and (b) the best matched orientations of WT GB1 based on the SFG data analysis using the GB1 crystal structure, (c) the best matched WT GB1 orientation after rotating all the simulated structures (based on the simulated structure at 781 ns) – replotted from Fig. 4-11 (d). (d) and (e) The best matched orientations of MT GB1 based on the SFG data analysis using the GB1 crystal structure, (f) the best matched MT GB1 orientation after rotating all the simulated structures (based on the simulated structure at 972 ns) – replotted from Fig. 4-12 (h).

4.7 Comparison between Interfacial WT GB1 and Interfacial MT GB1

In our previous studies of GB1 on graphene, we found that the planar amino acid residues play dominant roles in protein-graphene interactions through π-π interactions.[5] Five out of the ten non-charged planar residues of WT GB1 (F30, Q32, Y33, N35, and N37) are in the α-helical chain (shown in Figure 4-16). They have strong interactions with the graphene surface to cause a complete denaturation of WT GB1 on graphene, shown by both coarse grained MD simulation and the absence of detected SFG signal.[5] Here PS has many phenyl groups, which can have strong π-π interactions with GB1 planar amino acids, but not as strong as those between graphene and WT GB1. Therefore, SFG signals of WT GB1 on PS could still be detected, showing that the WT GB1 does not denature completely. On graphene, it is possible that WT GB1 needs to unfold for its side chains to well contact with the rigid planar graphene surface. However, in the case of PS, the PS surface roughness and molecular flexibility may be able to accommodate the protein-surface interaction better.

Among the five planar residues in the helical structure, our SFG studies and MD simulations showed that Q32 and N35 are facing towards the PS surface for better interaction with the PS surface. After mutating the residues Q32 and N35 to alanine, the π-π interactions between the α-helical chain of MT GB1 and PS were greatly reduced. Therefore the α-helix does not lie down on the PS surface anymore. Our results agreed with this analysis and showed that the MT GB1 helix more or less stands up on the PS surface (Figure 4-16 (b)).

To consider the experimental errors and the possible structural fluctuations of proteins at interfaces, we plot all possible WT GB1 and MT GB1 structures which have matching scores with 90% or higher compared to the highest matching scores. Out of the $400 \times 180/5 \times 360/5 = 1{,}036{,}800$ orientations, seven WT GB1 and one hundred MT GB1 structures were found. All the WT GB1 structures are similar, with the α-helix lying down, while all the MT GB1 structures are similar, with the α-helix more or less standing up. MT GB1 represents more possible structures on the PS surface than WT GB1 (one hundred vs. seven), indicating that MT GB1 has greater structure flexibility and has less interaction with the PS surface.

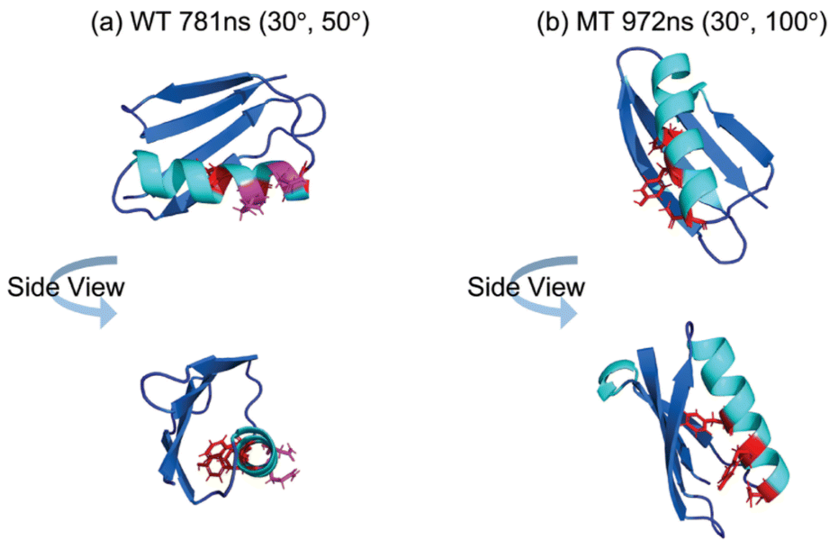

Figure 4-16 Orientation visualizations of (a) the simulated WT GB1 structure at 781 ns of (30°, 50°) and (b) the simulated MT GB1 structure at 972 ns of (30°, 100°). Residues Q32 and N35 are shown in magenta sticks. Residues F30, Y33 and N37 are shown in red sticks.

4.8 Conclusion

In this Chapter, the systematic methodology developed in Chapter 3, which combines SFG measurements, MD simulations and Hamiltonian calculations, was further improved in accuracy. Isotope labeling was incorporated into the protein sample preparations to enrich independent SFG measurements. Atomistic MD simulation was employed to replace DMD simulations to continuously model the dynamic process during protein/surface interaction and provide better sampling of interfacial protein structures. A point-by-point score system was created to compare the calculated spectra and the experimental reconstructed spectra with multiple peaks, which has higher efficiency and accuracy than the parameter comparison score system. This further developed methodology successfully deduced the conformations and orientations of WT GB1 and MT GB1 at the PS/protein solution interface in situ.

It was found that the WT GB1 adopts an orientation with its α-helix lying down on the PS surface due to the strong π-π interactions between the PS phenyl groups and the planar residues in the helical structure of WT GB1. It was also found that the MT GB1 has a different interfacial orientation, with the helix more or less standing up at the interface. These two orientations have different contact areas of the proteins with PS (shown in Figure 4-17). With Q32 and N35 amino acids replaced by alanine, the π-π interactions between the PS phenyl groups and the helix in MT GB1 were greatly reduced, enabling MT GB1 to stand up. This research shows that by mutating a very small number of some key amino acids in a protein, it is feasible to greatly vary the protein interfacial interactions to mediate interfacial protein structure.

Atomistic MD simulation is a powerful tool to capture protein - surface interactions to determine protein interfacial structure. However, it is necessary to validate the simulation results with experimental data, which is challenging. This study provides a systematic approach using SFG to determine protein interfacial structure with simulated results as inputs for SFG data analysis. The deduced protein structures with SFG measurements are more reliable than those using the protein crystal structure as the input structure.

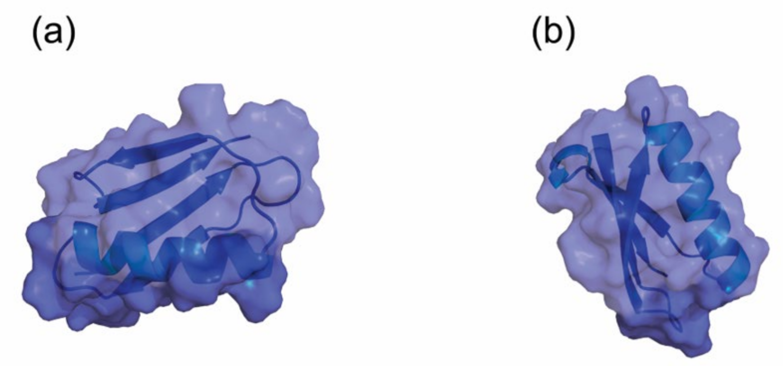

Figure 4-17 Visualization of Contact Areas (in dark blue) of (a) WT GB1 and (b) MT GB1 on PS. The contact area of WT GB1 on PS is obviously larger than the contact area of MT GB1 on PS.

4.9 References

(1) Kirsten Frank, M.; Dyda, F.; Dobrodumov, A.; Gronenborn, A. M. Core Mutations Switch Monomeric Protein GB1 into an Intertwined Tetramer. *Nat Struct Biol* **2002**. https://doi.org/10.1038/nsb854.

(2) Harrison, E. T.; Weidner, T.; Castner, D. G.; Interlandi, G. Predicting the Orientation of Protein G B1 on Hydrophobic Surfaces Using Monte Carlo Simulations. *Biointerphases* **2017**, *12* (2), 02D401. https://doi.org/10.1116/1.4971381.

(3) Berkovich, R.; Mondal, J.; Paster, I.; Berne, B. J. Simulated Force Quench Dynamics Shows GB1 Protein Is Not a Two State Folder. *J. Phys. Chem. B* **2017**, *121* (20), 5162–5173. https://doi.org/10.1021/acs.jpcb.7b00610.

(4) Wei, S.; Ahlstrom, L. S.; Brooks, C. L. Exploring Protein-Nanoparticle Interactions with Coarse-Grained Protein Folding Models. *Small* **2017**, *13* (18), 1603748. https://doi.org/10.1002/smll.201603748.

(5) Wei, S.; Zou, X.; Tian, J.; Huang, H.; Guo, W.; Chen, Z. Control of Protein Conformation and Orientation on Graphene. *Journal of the American Chemical Society* **2019**, *141* (51), 20335–20343. https://doi.org/10.1021/jacs.9b10705.

(6) Yang, J.; Zhang, Y. I-TASSER Server: New Development for Protein Structure and Function Predictions. *Nucleic Acids Res* **2015**, *43* (W1), W174–W181. https://doi.org/10.1093/nar/gkv342.

(7) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, *12* (1), 7–8. https://doi.org/10.1038/nmeth.3213.

(8) Zheng, W.; Zhang, C.; Li, Y.; Pearce, R.; Bell, E. W.; Zhang, Y. Folding Non-Homologous Proteins by Coupling Deep-Learning Contact Maps with I-TASSER Assembly Simulations. *Cell Reports Methods* **2021**, *1* (3), 100014. https://doi.org/10.1016/j.crmeth.2021.100014.

(9) Wei, T.; Ma, H.; Nakano, A. Decaheme Cytochrome MtrF Adsorption and Electron Transfer on Gold Surface. *The Journal of Physical Chemistry Letters* **2016**, *7* (5), 929–936. https://doi.org/10.1021/acs.jpclett.5b02746.

(10) Zhang, T.; Wei, T.; Han, Y.; Ma, H.; Samieegohar, M.; Chen, P.-W.; Lian, I.; Lo, Y.-H. Protein–Ligand Interaction Detection with a Novel Method of Transient Induced Molecular Electronic Spectroscopy (TIMES): Experimental and Theoretical Studies. *ACS Cent. Sci.* **2016**, *2* (11), 834–842. https://doi.org/10.1021/acscentsci.6b00217.

(11) Sarker, P.; Sajib, M. S. J.; Tao, X.; Wei, T. Multiscale Simulation of Protein Corona Formation on Silver Nanoparticles: Study of Ovispirin-1 Peptide Adsorption. *J. Phys. Chem. B* **2022**, *126* (3), 601–608. https://doi.org/10.1021/acs.jpcb.1c08267.

(12) Zhang, T. O.; Grechko, M.; Moran, S. D.; Zanni, M. T. Isotope-Labeled Amyloids via Synthesis, Expression, and Chemical Ligation for Use in FTIR, 2D IR, and NMR Studies. In *Protein Amyloid Aggregation*; Eliezer, D., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2016; Vol. 1345, pp 21–41. https://doi.org/10.1007/978-1-4939-2978-8_2.

(13) Decatur, S. M. Elucidation of Residue-Level Structure and Dynamics of Polypeptides via Isotope-Edited Infrared Spectroscopy. *Accounts of Chemical Research* **2006**, *39* (3), 169–175. https://doi.org/10.1021/ar050135f.

(14) Arkin, I. T. Isotope-Edited IR Spectroscopy for the Study of Membrane Proteins. *Curr Opin Chem Biol* **2006**, *10* (5), 394–401. https://doi.org/10.1016/j.cbpa.2006.08.013.

(15) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. 2D IR Line Shapes Probe Ovispirin Peptide Conformation and Depth in Lipid Bilayers. *Journal of the American Chemical Society* **2010**, *132* (8), 2832–2838. https://doi.org/10.1021/ja9101776.

(16) Ding, B.; Laaser, J. E.; Liu, Y.; Wang, P.; Zanni, M. T.; Chen, Z. Site-Specific Orientation of an α-Helical Peptide Ovispirin-1 from Isotope-Labeled SFG Spectroscopy. *The Journal of Physical Chemistry B* **2013**, *117* (47), 14625–14634. https://doi.org/10.1021/jp408064b.

(17) Ding, B. Studying Orientation and Conformation of α-Helical Peptides at Interfaces.

(18) Ding, B.; Panahi, A.; Ho, J.-J.; Laaser, J. E.; Brooks, C. L. I.; Zanni, M. T.; Chen, Z. Probing Site-Specific Structural Information of Peptides at Model Membrane Interface In Situ. *J. Am. Chem. Soc.* **2015**, *137* (32), 10190–10198. https://doi.org/10.1021/jacs.5b04024.

(19) Guo, W.; Lu, T.; Crisci, R.; Nagao, S.; Wei, T.; Chen, Z. Determination of Protein Conformation and Orientation at Buried Solid/Liquid Interfaces. *Chem. Sci.* **2023**, 10.1039.D2SC06958J. https://doi.org/10.1039/D2SC06958J.

(20) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(21) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(22) Li, B.; Zhang, S.; Andre, J. S.; Chen, Z. Relaxation Behavior of Polymer Thin Films: Effects of Free Surface, Buried Interface, and Geometrical Confinement. *Progress in Polymer Science* **2021**, *120*, 101431. https://doi.org/10.1016/j.progpolymsci.2021.101431.

(23) Ding, B.; Jasensky, J.; Li, Y.; Chen, Z. Engineering and Characterization of Peptides and Proteins at Surfaces and Interfaces: A Case Study in Surface-Sensitive Vibrational Spectroscopy. *Acc. Chem. Res.* **2016**, *49* (6), 1149–1157. https://doi.org/10.1021/acs.accounts.6b00091.

(24) Tan, J.; Zhang, J.; Luo, Y.; Ye, S. Misfolding of a Human Islet Amyloid Polypeptide at the Lipid Membrane Populates through β-Sheet Conformers without Involving α-Helical Intermediates. *J. Am. Chem. Soc.* **2019**, *141* (5), 1941–1948. https://doi.org/10.1021/jacs.8b08537.

(25) Wei, T.; Carignano, M. A.; Szleifer, I. Lysozyme Adsorption on Polyethylene Surfaces: Why Are Long Simulations Needed? *Langmuir* **2011**, *27* (19), 12074–12081. https://doi.org/10.1021/la202622s.

(26) Wei, T.; Carignano, M. A.; Szleifer, I. Molecular Dynamics Simulation of Lysozyme Adsorption/Desorption on Hydrophobic Surfaces. *The Journal of Physical Chemistry B* **2012**, *116* (34), 10189–10194. https://doi.org/10.1021/jp304057e.

(27) Nakano, C. M.; Ma, H.; Wei, T. Study of Lysozyme Mobility and Binding Free Energy during Adsorption on a Graphene Surface. *Applied Physics Letters* **2015**, *106* (15), 153701. https://doi.org/10.1063/1.4918292.

(28) Jahan Sajib, M. S.; Wei, Y.; Mishra, A.; Zhang, L.; Nomura, K.-I.; Kalia, R. K.; Vashishta, P.; Nakano, A.; Murad, S.; Wei, T. Atomistic Simulations of Biofouling and Molecular Transfer of a Cross-Linked Aromatic Polyamide Membrane for Desalination. *Langmuir* **2020**, *36* (26), 7658–7668. https://doi.org/10.1021/acs.langmuir.0c01308.

Chapter 5 Conformation and Orientation of Antimicrobial Peptides MSI-594 and MSI-594A in a Lipid Membrane

5.1 Attribution Information

Materials covered in this chapter were adapted with permission from Yang, P.; Guo, W.; Ramamoorthy, A.; Chen, Z. Conformation and Orientation of Antimicrobial Peptides MSI-594 and MSI-594A in a Lipid Membrane. *Langmuir* **2023**, *39* (15), 5352–5363. Copyright © 2023 American Chemical Society. This research was in collaboration with Prof. Ayyalusamy Ramamoorthy's group from the University of Michigan.

The SFG and ATR-FTIR experimental measurements included in this research were conducted by P.Y. The data interpretation of the collected experimental data was conducted by W.G.

5.2 Background and Motivation

5.2.1 Antimicrobial Peptide (AMP)-Lipid Membrane Interaction Mechanisms

AMPs are promising candidates for next generation anti-bacteria drugs as mentioned in Chapter 2. To favor the design of AMPs, understanding their antibiotic mechanism is necessary. Extensive research showed that AMPs can selectively disrupt bacterial membranes resulting in cell lysis via three main modes: barrel-stave mode, toroidal pore mode, and carpet mode.[1] Barrel-stave mode and toroidal mode propose that AMPs can induce transmembrane pores in lipid bilayers via different interaction pathways. In the barrel-stave mode, peptides can insert into the lipid bilayers and form a barrel-like buddle perpendicular to the lipid membrane with a central lumen.[2] The hydrophobic side of the AMPs is facing the hydrophobic aryl chain of the lipid while the hydrophilic side of the AMPs is exposed to the water-filled channel.[3] Different from the barrel-stave mode, the toroidal mode proposes that AMPs are always associated with the lipid head groups and can lead to positive curvature of the lipid membrane to accommodate such a strong interaction between the AMPs and the lipid head groups.[2] Such an interaction can also lead to multiple orientation distributions of the AMPs.[3] When AMPs' concentration is sufficiently high, AMPs can disrupt the lipid membrane via the carpet mode. In the carpet mode, AMPs parallelly cover the entire lipid membrane and disrupt the membrane like detergents/surfactants.[2] It has been suggested that toroidal pore formation may be an intermediate state prior to carpet mode since the carpet mode also requires the lipid head group to associate with the AMPs during the micellization process.[3] The membrane-bound AMPs orient differently in each membrane disruption mode. Therefore, investigations on membrane orientations of AMPs are essential to understand the membrane-disrupted antimicrobial mechanisms of AMPs and to favor further AMP designs.

5.2.2 MSI-594, MSI-594A and Model Membranes

MSI peptides are a series of AMPs derived from magainin and melittin, which exhibit a broad-spectrum of activities against both Gram-positive and Gram-negative bacteria.[4,5] MSI-594 is a synthesized lysine-rich peptide with 24 amino acids and exhibits high antimicrobial potency. NMR experiments indicated that MSI-594 forms a helical hairpin structure interacting with LPS.[6,7] Its derivative MSI-594A with a mutation of Phe-5 to alanine, exhibits a decreased π–π inter-helical interaction and lower activities against Gram-negative bacteria compared to MSI-594. Comparing the membrane interactions of these two peptides can significantly enhance the understanding of mutation-induced structural changes and the relationships between structure and activity.

To probe AMP - membrane interactions, model membranes of phosphate lipids are being widely used. The neutral POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipids were often used to construct a model system for mammalian cell membrane.[6,8] The anionic 7:3 POPC/POPG (anionic 1-palmitoyl-2-oleoyl-sn-glycero3-phospho(1'-rac-glycerol)) and cationic 7:3 POPC/POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine) model membranes are used to mimic Gram-negative and Gram-positive bacterial lipid membranes, respectively.[5,6,8] Typical experimental models for membranes with these lipid compositions are vesicles/liposomes, Langmuir monolayers, and solid supported bilayers (SLBs).[9] Lipid vesicles are widely used to observe membrane breakdown and lysis.[5,6] Langmuir monolayer model is easy to control and modulate, and is suitable to study the lipid packing changes induced by AMPs.[10] The solid supported bilayers are lipid bilayers constructed on a solid platform, which can be more readily characterized by surface-sensitive techniques, such as SFG and ATR-FTIR, than free-floating vesicle/liposome models.[11–14] In this study, we aimed to determine the orientations of MSI-594 and its analogue MSI-594A associated with two different lipid bilayers: neutral POPC and anionic 7:3 POPC/POPG.

5.2.3 Combination of SFG and ATR-FTIR Measurements

As mentioned previously, ATR-FTIR could serve as a supplemental tool to SFG to provide more detailed structural information of interfacial peptides and proteins, when the background contributions from the surface and the bulk solution are negligible. Since the thickness of a SLB is ~ 10 nm (much smaller than the penetration depth of IR evanescent wave)[8] and the IR absorption of water in the amide I frequency region can be greatly reduced if D_2O is used as a substitute solvent, ATR-FTIR can be a suitable tool to study AMPs adsorbed onto a SLB surface. The combination of SFG and ATR-FTIR can be powerful to probe peptide−lipid bilayer interactions at a very low peptide concentration in situ in real time, as illustrated in previous publications.[11,15–17]

Here in this chapter, the Hamiltonian program was further developed to include IR spectra calculations and the scoring system was also upgraded to include IR spectra comparison between experimental and calculated spectra. The final score map (as a function of peptide orientation) considers spectra comparisons from both SFG and FTIR measurements since the actual orientation of the studied proteins and peptides on surfaces or at interfaces should be in agreement with both SFG and ATR-FTIR data.

5.3 Experimental Materials and Methods

5.3.1 Materials

MSI-594 (GIGKFLKKAKKGIGAVLKVLTTGL) and MSI-594A (GIGKALKKAKKGIGAVLKVLTTGL) were purchased from Genscript (Piscataway, NJ). The POPC and POPG lipids were purchased from Avanti Polar Lipids (Alabaster, AL). They are both in the fluid phase at room temperature (~24° C). Right angle CaF_2 prisms were purchased from Altos Photonics (Bozeman, MT) and used for SFG experiments. The Langmuir−Blodgett and Langmuir−Schaefer (LB/LS) method was used to deposit the proximal and distal leaflets of a single lipid bilayer onto a CaF_2 prism. A detailed procedure to prepare the SBLs was well-documented in previous publications.[13,18] A KSV2000 LB system (Biolin Scientific, Stockholm, Stockholms Lan, Sweden) was used for lipid bilayer preparation. For SFG experiments, phosphatebuffered saline (PBS, 10.0 mM, pH =7.4) was used in our experiment to make the subphase peptide solution. The lipid bilayers were then placed in contact with the buffer solution inside a 2 mL reservoir throughout the entire experiment. For peptide−lipid bilayer interaction experiments, an appropriate amount of peptide stock solution was injected into the reservoir to achieve 4.0 μM peptide solution concentration and allowed peptide molecules to diffuse to and interact with the lipid bilayer over 1 h. The SFG ssp and ppp spectra were then collected in frequency range of 1550 – 1750 cm^{-1} after the peptide−lipid bilayer interaction reached equilibrium and the SFG signal became stable. For ATR-FTIR experiments, the lipid bilayer was prepared on clean ZnSe substrates (Specac, U.K.) for ATR-FTIR measurement by using the vesicle fusion method.[12,13] Peptide samples were injected into the subphase to match the concentration we used in the SFG study. To avoid the water background in the amide I region (1600−1690 cm−1) in ATR-FTIR spectra, D_2O-based buffer was used in all the ATR-FTIR experiments. ATR-FTIR spectra were then collected in the p and s polarizations in frequency range 1575 – 1725 cm^{-1} on a Nicolet 6700 spectrometer with an ATR-FTIR accessory.

5.3.2 Hamiltonian Approach and Scoring System

The experiments reported in Chapters 2 and 3 included SFG measurements, not the ATR-FTIR experiments. Therefore, the developed Hamiltonian approach focused on the SFG spectra calculations, and the score system was only used for spectra comparison between collected SFG spectra and the calculated SFG spectra in those two chapters. Here in this study, ATR-FTIR measurements were also included to increase independent experimental measurements. Therefore, the Hamiltonian approach was further developed to incorporate FTIR spectra calculations for both s and p polarizations. The point-by-point score system was also upgraded with weighted least square comparison to match the calculated SFG and FTIR spectra with experimental data separately. The final score map was then generated by combining SFG and FTIR score maps.

When calculating the SFG and ATR-FTIR spectra, parameters such as peak center and peak width for a single uncoupled amide unit are needed. These parameters were chosen from the peak center range [1640 cm^{-1}, 1660 cm^{-1}] for both SFG and ATR-FTIR spectra calculation, and peak width range [10.0 cm^{-1}, 20.0 cm^{-1}] for SFG spectra calculation and [5.0 cm^{-1}, 15.0 cm^{-1}] for ATR-FTIR spectra calculation to generate the best outcome for each case.

The calculated SFG ssp spectra were normalized to [0, 1] and meanwhile the calculated SFG ppp spectra were scaled to maintain the original ppp/ssp intensity ratios (the ratio was determined by the maximum of I_{ppp} divided by the maximum of I_{ssp}). The calculated ATR-FTIR s spectra were normalized to [0, 1] and meanwhile the calculated ATR-FTIR p spectra were scaled to maintain the original p/s intensity ratios (the ratio was determined by the maximum of I_p divided by the maximum of I_s). The experimental spectra were also normalized in the same fashion.

Spectra comparison between the reconstructed experimental spectra and the Hamiltonian calculated spectra for each type of spectra (SFG ssp spectra, SFG ppp spectra, ATR-FTIR s spectra and ATR-FTIR p spectra) was done by weighted least square comparison of each wavenumber (1 cm^{-1} as increment) under a given orientation (equation (5 – 1)). The experimental spectra were used as weights in such comparison. The reciprocal of the least square comparison was used as a score to quantify the similarity between the two compared spectra (equation (5 – 2)). As a result, for one input structure for the Hamiltonian program, a total of four orientation-dependent score maps can be obtained (one for each type of spectra). The final score map can be obtained by multiplying all the four score maps (and 4th root was applied after the multiplication, see equation

(5 – 3)). In the final overall score map, the best matched orientations are in the region with the highest score (marked in red).

$$LS_k = \sqrt{\frac{\Sigma_i\left[\left(x_{exp,i} - x_{cal,i}\right) \times \frac{x_{exp,i}}{\Sigma_i x_{exp,i}}\right]^2}{(n-1)_k}} \qquad (5-1)$$

$$Score_k = \frac{1}{LS_k} \qquad (5-2)$$

$$Final\ Score = \sqrt[4]{\prod Score_k} \qquad (5-3)$$

where LS_k is the least square comparison between the calculated spectrum ($x_{cal,i}$) (under a certain orientation) and the experimental spectrum ($x_{exp,i}$) of a specific type of the spectrum (ssp SFG, ppp SFG, p FTIR or s FTIR), and n is the total number of points of the spectrum (n=201 points for an SFG spectrum and n=151 points for a FTIR spectrum).

5.4 Results and Discussion

5.4.1 SFG and ATR-FTIR Measurements

SFG ssp and ppp spectra, and ATR-FTIR s and p spectra were collected at amide I frequency region of the four interfaces: MSI-594/POPC, MSI-594A/POPC, MSI-594/7:3 POPC:POPG, and MSI-594A/7:3 POPC:POPG, as shown in Figure 5-1. The measured SFG ppp and ssp signal strength ratio (or $\chi^{(2)}_{ppp}/\chi^{(2)}_{ssp}$) is 1.39 ± 0.14 for MSI-594 associated with a POPC lipid bilayer. Unlike SFG, which mainly detects signals where inversion symmetry is broken and the signals usually are dominated by the α-helical structure contribution, both α-helical and β-turn structures have contributed to the detected ATR-FTIR signals. The ATR-FTIR spectra contained the vibrational peak at 1650 cm^{-1} from the α-helical structure, as well as the 1674 cm^{-1} peak from the β-turn structure. The ATR-FTIR spectral fitting results indicated that the fitted R^{ATR} is 1.47 ± 0.15 for the α-helical structure. We obtained the measured SFG ppp and ssp signal strength ratio (or $\chi_{ppp}^{(2)}/\chi_{ssp}^{(2)}$) of 1.28 ± 0.13 for MSI-594A associated with a POPC lipid bilayer. The ATR-FTIR spectra from MSI-594A contained vibrational peaks at 1635 cm^{-1} from the 3_{10}-helical structure; 1650 cm^{-1} from the α-helical structure, and 1674 cm^{-1} from the β-turn structure. We carefully fit the spectra and obtained the fitted R^{ATR} = 1.37 ± 0.14 for the α-helical structure. In the cases of MSI-594 and MSI-594A associated with 7:3 POPC:POPG lipid membrane, we fitted these spectra and obtained the $\chi_{ppp}^{(2)}/\chi_{ssp}^{(2)}$ = 1.36 ± 0.14 and R^{ATR} = 1.45 ± 0.15 for the MSI-594 α-helical structure, and $\chi_{ppp}^{(2)}/\chi_{ssp}^{(2)}$ = 1.22 ± 0.12 and R^{ATR} = 1.35 ± 0.14 for the MSI-594A α-helical structure.

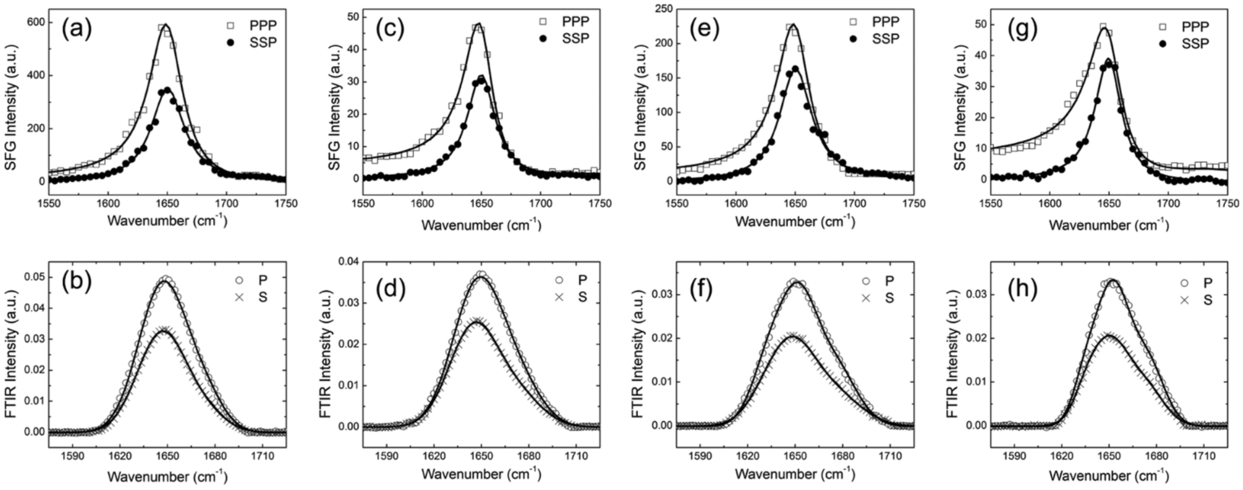

Figure 5-1 SFG (a, c, e, g) and ATR-FTIR (b, d, f, h) spectra of (a, b) MSI-594 associated with a POPC lipid bilayer; (c, d) MSI-594A associated with a POPC lipid bilayer; (e, f) MSI-594 associated with a 7:3 POPC/POPG lipid bilayer; and (g, h) MSI-594A associated with a 7:3 POPC/ POPG lipid bilayer. SFG ppp and ssp spectra are shown in open dots and filled black dots, respectively (top row). ATR-FTIR p and s spectra are shown in circles and crosses, respectively (bottom row). The solid black lines are fitted spectra.

5.4.2 Hamiltonian Matching Results

The Hamiltonian method was then used to determine the conformation and orientation of MSI-594 on the POPC lipid bilayer surface. The NMR structure of MSI-594 (PDB ID: 2K98, shown in Figure 5-2) was used to approximate the interfacial structure of MSI-594 and used as the input structure for the Hamiltonian program. The best-matched orientations with the highest matching score between the calculated and reconstructed experimental spectra of this MSI-594 structure on the POPC lipid bilayer surface are (130, 320°) and (50, 140°) with a score of 240, as shown in Figure 5-3. The spectral comparisons between the reconstructed SFG and ATR-FTIR spectra from the experimental data and the calculated SFG and ATR-FTIR spectra based on the NMR structure (PDB ID: 2K98) at the orientation of (130, 320°) (or (50, 140°)) are shown in Figures 5-4 (a) and (b). The matching qualities of the ATR-FTIR spectra are good, while the matching qualities of the SFG ssp and ppp spectra are not satisfactory. The calculated SFG ssp spectrum has an additional side peak with a frequency higher than 1650 cm^{-1}, which was not observed in the experimentally collected SFG ssp spectrum. Also, the intensity ratio between the calculated SFG ssp and ppp spectra does not match the experimental measurement (the difference is ~10%). This means that perhaps the input NMR structure (PDB ID: 2K98) for the Hamiltonian program may not be a good candidate to approximate the interfacial structure of MSI-594 associated with a POPC lipid bilayer. As a result, the spectra calculated with the highest matching quality based on such a structure still cannot reasonably match the reconstructed experimental spectra.

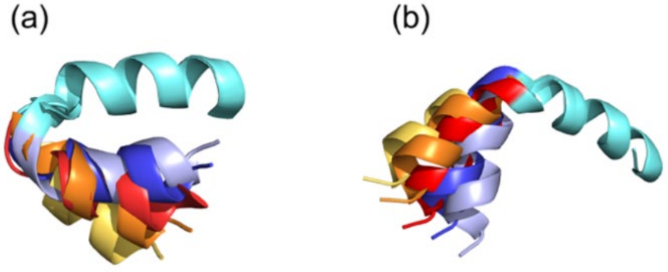

Figure 5-2 Modified structures of (a) MSI-594 based on PDB ID: 2K98 and (b) MSI-594A based on PDB ID: 2L36. Residues 12−24 were colored in cyan for all of the structures. Residues 1−11 were colored in red for the original NMR structures, in blue for the structures with a bent angle change of −5° from the original NMR structures, in light purple for the structures with a bent angle change of −10° from the original NMR structures, in orange for the structures with a bent angle change of +5° from the original NMR structures, and in yellow for the structures with a bent angle change of +10° from the original NMR structures.

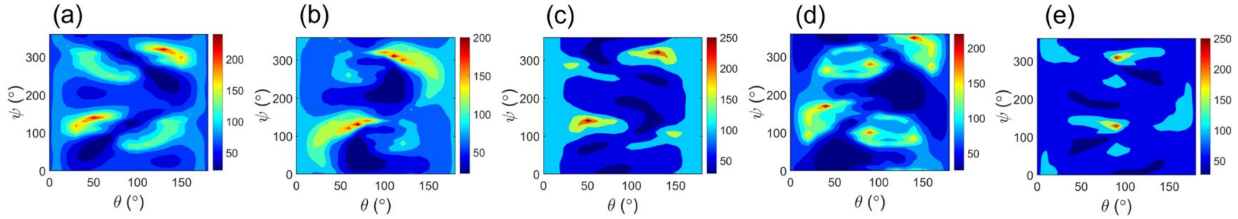

Figure 5-3 Final score maps that quantify the degree of similarity between the reconstructed spectra (SFG and ATR-FTIR) of MSI-594 associated with a POPC lipid bilayer and the corresponding calculated spectra by using the input structure as (a) NMR structure (PDB ID: 2K98), (b) modified NMR structure with −5° bent angle change, (c) modified NMR structure with +5° bent angle change, (d) modified NMR structure with −10° bent angle change, and (e) modified NMR structure with +10° bent angle change.

Table 5-1 Highest scores obtained from the final score maps shown in Figure 5-3 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined helical hairpin structure of MSI-594 (PDB ID: 2K98).

	NMR	**NMR -5°**	**NMR +5°**	**NMR -10°**	**NMR +10°**
Highest Score	240	211	293	236	269

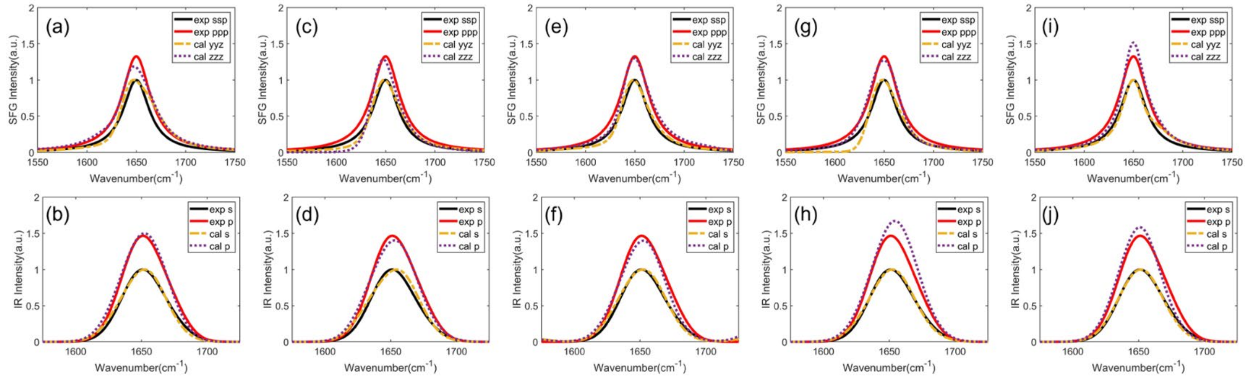

Figure 5-4 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and ATR-FTIR (bottom)) of MSI-594 associated with a POPC lipid bilayer and calculated spectra at the best-matched orientations by using input structures as (a, b) the NMR structure (PDB ID: 2K98); (c, d) the modified NMR structure with −5° bent angle change; (e, f) the modified NMR structure with +5° bent angle change; (g, h) the modified NMR structure with −10° bent angle change; and (i, j) the modified NMR structure with +10° bent angle change.

Since the NMR structure 2K98 was measured in a lipopolysaccharide (LPS) environment,[6] which is not identical to the planar POPC lipid surface, the bent angle between the two helical chains of MSI-594 may slightly vary. This needs to be considered to compensate for such an environmental difference. Therefore, in this study, the NMR structure (PDB ID: 2K98) was modified with the bent angle changes of −5, +5, −10, and +10°, as shown in Figure 5-2 (a). Each modified MSI-594 structure was then used as the input structure for the Hamiltonian program to calculate the orientation-dependent SFG and ATR-FTIR spectra. The calculated spectra were compared to the reconstructed experimental spectra to generate a final score map for each modified structure, as shown in Figures 5-3 (b)-(e). The highest matching scores obtained from the deduced final score maps are summarized in Table 5-1.

Among all five input structures (one original NMR structure and four modified structures), the results obtained from the Hamiltonian method indicated that the NMR structure with a bent angle change of +5° gives the highest matching score. The spectral matching results shown in Figure 5-4 also validate that the NMR structure with +5° bent angle modification can generate SFG and ATR-FTIR spectra that are more similar to the experimental reconstructed SFG and ATR-FTIR spectra, with better-matched spectra intensity ratios between different polarizations. The matching result shows that with a bent angle change of +5°, the best matching score increased to 293 at the orientations of (130, 320°) and (50, 140°). These matched orientations (130, 320°) and (50, 140°) are the same as the matched ones determined by the original structure of 2K98, but the matching score increased from 240 to 293, meaning that the structure of +5° modification of the bent angle is a better candidate to approximate the interfacial structure of MSI-594 than the original NMR structure. Figure 5-5 presents the matched orientations that possess the highest matching score. Overall, the helical chains are slightly tilted versus the lipid surface plane, with residues 1−11 (Figure 5-5 (a)) or residues 12−24 (Figure 5-5 (b)) touching the lipid surface, showing a near-parallel representation versus the lipid surface (more discussion below). Another pair of orientations (90, 130°) and (90, 310°) with a +10° deviation from the bent angle of the

NMR structure 2K98 also exhibited a high matching score. We define a matching score at 90% of the highest matching score and higher as a high matching score. For this case, the highest score is 293. Therefore, a score at or higher than 264 is a high matching score. These two orientations (shown in Figure 5-6) with a score of 269 can also be possible matching candidates. However, we excluded them after considering their interactions with the lipid bilayer surface. A more detailed discussion is presented below. This research demonstrates that the combined SFG and ATR-FTIR studies can be used to deduce MSI-594 conformation and orientation at the interface. Here, two opposite absolute orientations were determined to be the most likely orientations with the same highest matching score. Regular SFG and ATR-FTIR results cannot distinguish these two orientations. We will further discuss which orientation among the two is the most likely orientation below.

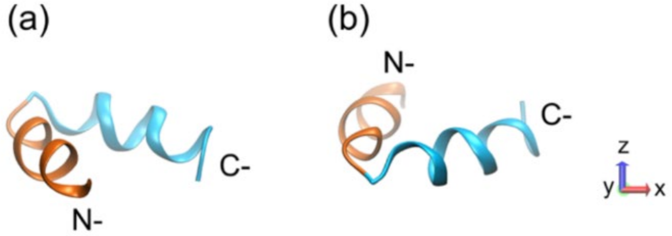

Figure 5-5 Membrane orientations deduced based on the modified NMR structure of MSI-594 (PDB ID: 2K98) with +5° bent angle change associated with the POPC or 7:3 POPC/POPG lipid bilayer with the highest matching score (score = 293 of the case of POPC and score =311 of the case of 7:3 POPC/POPG). The orientations of N- and C-termini segments are shown in deep orange (residues 1− 11) and cyan (residues 12−24), respectively, with respect to the lipid bilayer surface (i.e., the xy plane).

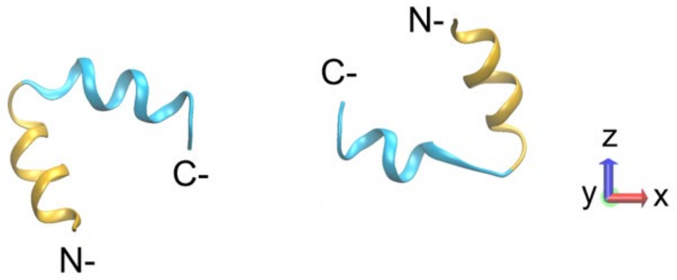

Figure 5-6 The membrane orientations deduced based on the modified NMR structure of MSI-594 (PDB ID: 2K98) with +10° bent angle change associated with POPC or 7:3 POPC:POPG lipid bilayer with a high matching score (score >= 264 for the case of POPC and score >= 280 for the case of 7:3 POPC:POPG). The orientations of N- and

C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane).

Similar to MSI-594, we reconstructed experimental SFG and ATR-FTIR spectra of MSI-594A using the spectral fitting parameters and compared them to the calculated spectra using the Hamiltonian method. This way we can determine the orientation of MSI-594A associated with a POPC lipid bilayer. Bent angle modifications (−5, +5, −10, and +10°) were also applied to the NMR structure of MSI-594A (PDB ID: 2L36), shown in Figure 5-2 (b). We calculated SFG and ATR-FTIR spectra of MSI-594A based on the NMR structure and these five modified structures as a function of MSI-594A orientation. These calculated spectra were compared to the above reconstructed experimental spectra to obtain a final matching score map for each input structure, as shown in Figure 5-7. The most likely peptide orientation angle regions are shown in red color. The results of the spectra comparison between the calculated spectra and reconstructed experimental spectra of MSI-594A associated with a POPC lipid bilayer at best matched orientations (or with the highest matching score) of each input structure are presented in Figure 5-8. The modified NMR structure (PDB ID: 2L36) with a bent angle change of +10° at the orientations of (160, 0°) and (20, 180°) gives the highest matching score. This is the only pair that possesses a high matching score for MSI-594A associated with the POPC bilayer. These two orientations are shown in Figure 5-9. Further discussion on which absolute orientation is the most likely MSI-794A orientation will be presented below.

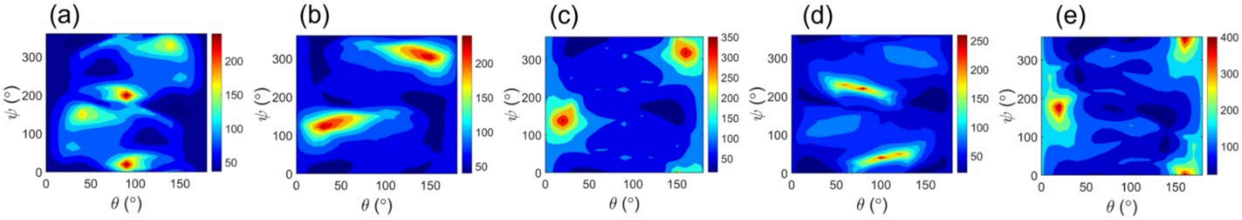

Figure 5-7 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594A associated with a POPC lipid bilayer and the corresponding calculated spectra by using

the input structure as (a) the NMR structure (PDB ID: 2L36); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change.

Table 5-2 Highest scores obtained from the final score maps shown in Figure 5-7 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined structure of MSI-594A (PDB ID: 2L36).

	NMR	NMR -5°	NMR +5°	NMR -10°	NMR +10°
Highest Score	256	280	360	261	425

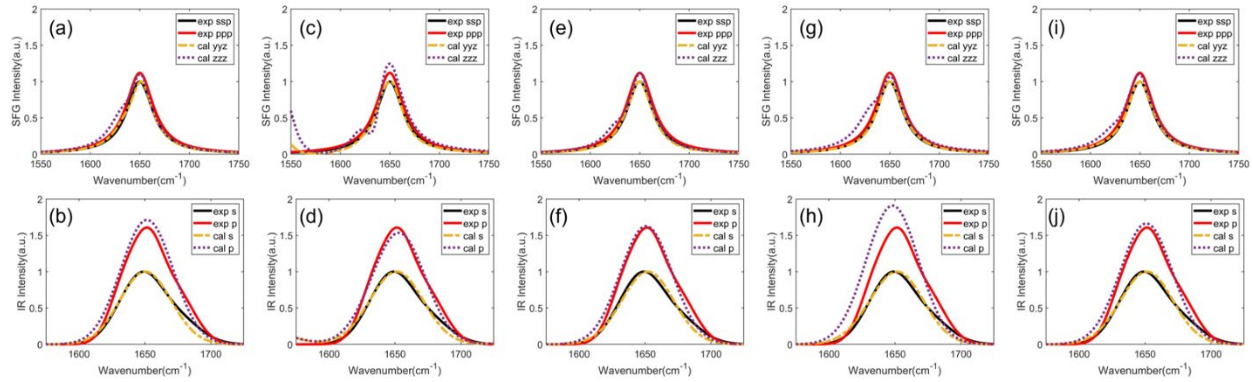

Figure 5-8 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594A associated with a POPC lipid bilayer and the calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2L36); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g, h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change.

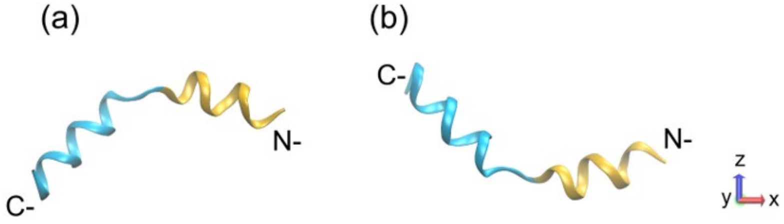

Figure 5-9 Best matched membrane orientations of the modified NMR structure of MSI-594A (PDB ID: 2L36) with +10° bent angle change associated with POPC lipid bilayer with highest matching score (score=425). The orientations of N- and C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane).

The orientation angles of MSI-594 and MSI-594A associated with a 7:3 POPC: POPG lipid bilayer were deduced using the same approach described above based on the original and modified

NMR structures. The final score maps and spectra comparisons are shown in Figure 5-10 to Figure 5-13. It was found that the modified NMR structure (PDB ID: 2K98) with a +5° change of the bent angle has the highest matching score for MSI-594, which is the same as the conformation that associated with pure POPC lipid bilayer. The deduced most possible membrane orientations of MSI-594 based on this structure are also the same as this peptide associated with pure POPC lipid bilayers (shown in Figure 5-6). Other possible orientations possessing high matching scores are shown in Figure 5-7, which are also the same as those of MSI-594 when interacting with a POPC lipid bilayer. The similar result obtained from the same peptide interacting with both lipids indicates that MSI-594 behaves similarly (has similar conformation and orientation) on the POPC lipid bilayer and the 7:3 POPC: POPG mixed lipid bilayer.

In the case of MSI-594A interacting with 7:3 POPC POPG lipid bilayer, it was found that the NMR structure (PDB ID: 2L36) with +10° bent angle modification generated the highest matching score between the calculated and reconstructed experimental spectra from the peptide on the mixed bilayer. This conformation is also the best matching for MSI-594A interacting with the zwitterionic POPC lipid bilayer. Figure 5-14 presents the most possible membrane orientations (150°, 20°) and (30°, 200°). The best matching orientations of MSI-594A on a 7:3 POPC: POPG lipid bilayer are similar to MSI-594A on a POPC lipid bilayer as presented above (shown in Figure 5-9).

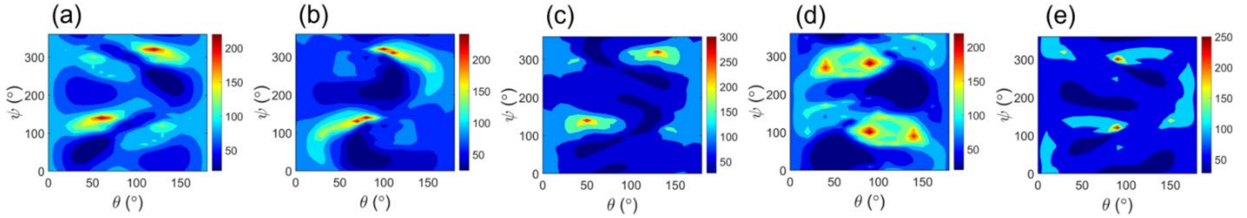

Figure 5-10 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594 associated with a 7:3 POPC:POPG lipid bilayer and the corresponding calculated spectra by using the input structure as (a) the NMR structure (PDB ID: 2K98); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change.

Table 5-3 Highest scores obtained from the final score maps shown in Figure 5-10 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined helical hairpin structure of MSI-594 (PDB ID: 2K98).

	NMR Struc.	NMR -5°	NMR +5°	NMR -10°	NMR +10°
Highest Score	231	255	311	235	283

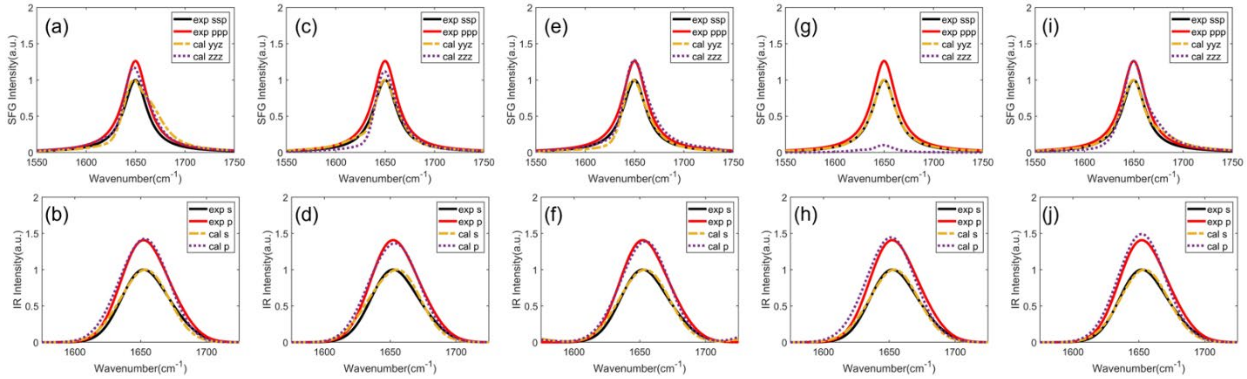

Figure 5-11 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594 associated with a 7:3 POPC:POPG lipid bilayer and the calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2K98); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g, h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change.

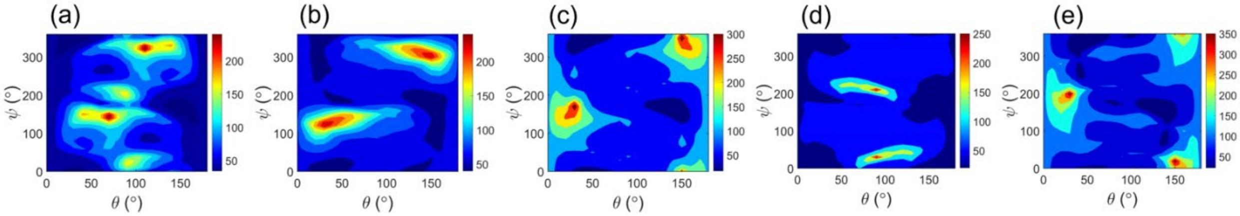

Figure 5-12 Final score maps that quantify the degree of similarity between the reconstructed experimental spectra (SFG and FTIR) of MSI-594A associated with a 7:3 POPC:POPG lipid bilayer and the corresponding calculated spectra by using the input structure as (a) the NMR structure (PDB ID: 2L36); (b) the modified NMR structure with bent angle -5° change; (c) the modified NMR structure with bent angle +5° change; (d) the modified NMR structure with bent angle -10° change; and (e) the modified NMR structure with bent angle +10° change.

Table 5-4 Highest scores obtained from the final score maps shown in Figure 5-12 for the bent angle deviation by 0°, -10°, -5°, +5° or +10° from the NMR determined structure of MSI-594A (PDB ID: 2L36).

	NMR	NMR -5°	NMR +5°	NMR -10°	NMR +10°
Highest Score	253	240	347	273	387

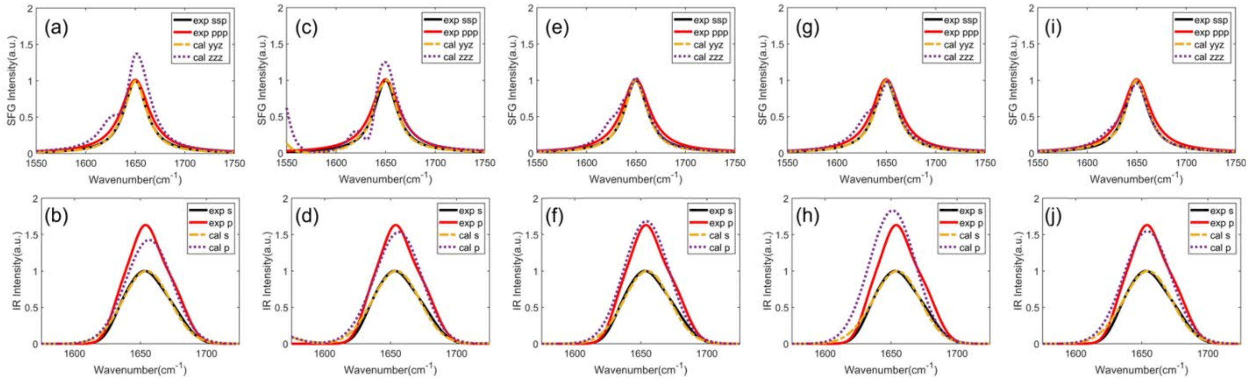

Figure 5-13 Spectra comparisons between the reconstructed experimental spectra (SFG (top) and FTIR (bottom)) of MSI-594A associated with a 7:3 POPC:POPG lipid bilayer and calculated spectra at best matched orientations (with the highest matching score) by using input structures as (a, b) the NMR structure (PDB ID: 2L36); (c, d) the modified NMR structure with bent angle -5° change; (e, f) the modified NMR structure with bent angle +5° change; (g, h) the modified NMR structure with bent angle -10° change; and (i, j) the modified NMR structure with bent angle +10° change.

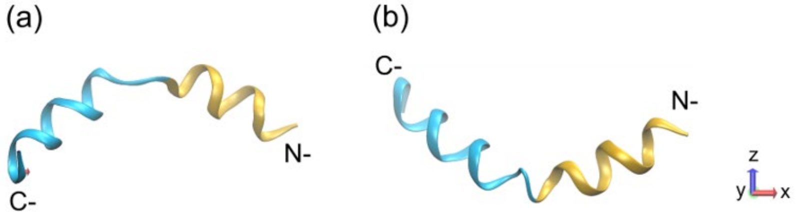

Figure 5-14 Best matched membrane orientations of the original NMR structure of MSI-594A (PDB ID: 2L36) in 7:3 POPC:POPG lipid bilayer with highest matching score (score=387). The orientations of N- and C- termini helical segments are shown in yellow (residue 1-11) and cyan (residue 12-24), respectively, with respect to the lipid bilayer surface (i.e. the xy plane).

5.3.3. Possible Peptide−Lipid Interaction Mechanisms

As presented above, from the SFG and ATR-FTIR results, we could deduce the most likely conformation (or the bending angle between the two helices) and orientations of MSI-594 and MSI-594A associated with the lipid bilayer at the interface. However, we could not differentiate the two most likely orientations with opposite absolute orientations (with the same highest matching score) for each case using our SFG and ATR-FTIR measurements. Phase-sensitive SFG measurements could determine the absolute orientation,[19,20] but most of the SFG phase-sensitive measurements were performed on exposed surfaces instead of buried interfaces, which is beyond the scope of this study. In the following, we will further discuss the possible structures, especially the possible absolute orientations of MSI-594 and MSI-594A associated with the lipid bilayer.

As mentioned above, the spectral matching results show that MSI-594 has the same deduced conformation and orientation on the POPC lipid bilayer and the 7:3 POPC/POPG lipid bilayer. The best-matched conformation is the modified NMR structure 2K98 with a bent angle change of +5°. The NMR structure 2K98 was determined by MSI-594 associated with LPS micelles.[6,7] The LPS micelles carry higher charges than the POPC and 7:3 POPC/POPG lipid bilayers, which may create stronger interactions with MSI-594, leading to a smaller bent angle and a more compact structure of the peptide. The two orientations (with opposite absolute orientations) of MSI594 with the highest matching score associated with the lipid bilayer (POPC or 7:3 POPC/POPG) are replotted in Figure 5-15 (a). The positively charged residue Lys is colored in purple. All Lys residues in MSI-594 are pointing to the same side of the peptide plane.[6] For the orientation shown on the left side of Figure 5-15 (a), noted as MSI-594-a, the positively charged plane faces the aqueous solution. For the orientation shown on the right side of Figure 5-15 (a), noted as MSI-594-b, the positively charged plane faces the lipid bilayer.

We believe it is unlikely that MSI-594 adopts both orientations MSI-594-a and MSI-594-b when associated with a lipid bilayer. For a7:3 POPC/POPG lipid bilayer, since the POPG lipid component is negatively charged, the entire mixed lipid bilayer is thus negatively charged. When MSI-794 is associated with a negatively charged lipid bilayer, it is more likely for the positively charged plane to face the lipid bilayer for a more favorable interaction. Therefore, for MSI-594 on the 7:3 POPC/POPG bilayer, MSI-594-b should be the most likely orientation. As presented above, SFG and ATR-FTIR spectra collected from MSI-594 on the POPC surface are similar to those on the 7:3 POPC/POPG bilayer; thus, the deduced MSI-594 on different lipid bilayers has the same conformation and the same (or opposite) possible orientations. It is unlikely that MSI-594 coincidently has the exact opposite orientations on POPC and on mixed lipid bilayers. We believe that MSI-594 also adopts the MSI-594-b orientation as shown in Figure 5-15 (a) when associated with a pure POPC bilayer. It is reasonable that the charged plane interacts with the POPC bilayer, since the POPC headgroups are hydrophilic, which could favorably interact with the charged hydrophilic side of MSI-594. In addition, it is also reasonable that MSI-594 adopts the same structure on POPC and 7:3 POPC/POPG bilayers, since the mixed bilayer studied here has POPC as the major component. Likely the MSI-594 molecules mainly interact with POPC in the mixed 7:3 POPC/POPG bilayer. It was shown that at very low peptide concentrations, the peptide can interact with the zwitterionic lipid via hydrophobic interactions without the need for an anionic lipid. So, the anionic lipid that is needed to attract the cationic peptide is not essential all the time.[5] Nevertheless, for MSI-594 on POPC and mixed lipid bilayers, the charged helical structure interacts with the lipids with a “face-on” orientation. Such a “face-on” interaction between MSI-594 and the lipid bilayer indicates that MSI-594 takes a carpet mechanism to perturb lipid bilayers. This observation matches the results reported in the literature.[1,5–8]

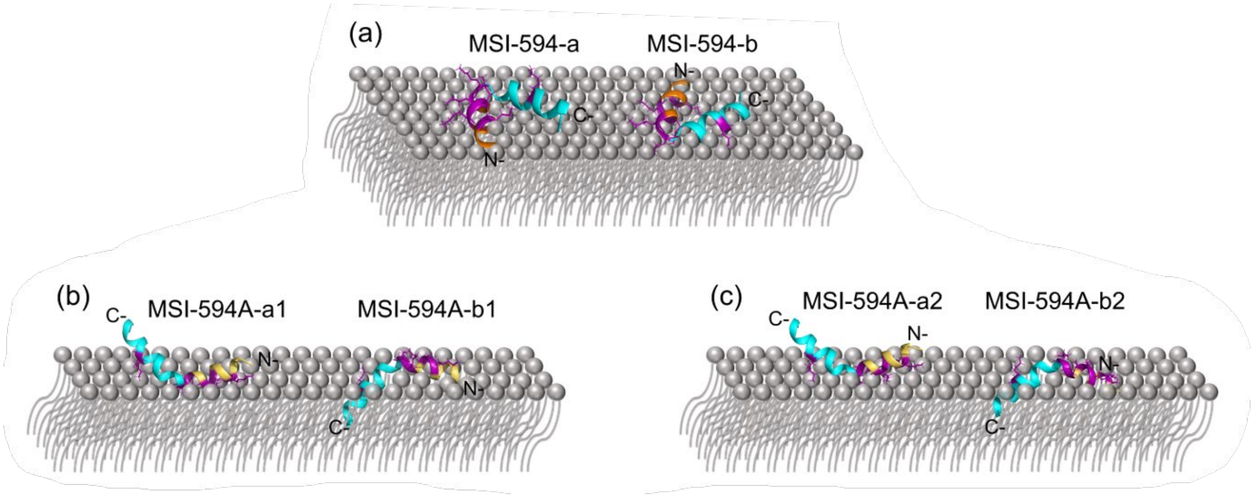

Figure 5-15 Possible configurations of (a) MSI-594 associated with POPC or 7:3 POPC:POPG lipid bilayer; (b) MSI-594A associated with POPC lipid bilayer and (c) MSI-594A associated with 7:3 POPC:POPG lipid bilayer. Grey solid spheres represent the lipid headgroups, grey curved lines represent the lipid hydrophobic tails, and the positively charged residue Lys is colored in purple sticks.

It is worth mentioning that the deduced "face-on" orientation of MSI-594 in this study is not a "complete lying down" orientation. To better visualize the deduced orientation of MSI-594 on lipid bilayers, "standing-up" and "complete lying down" orientations of MSI are displayed along with the deduced MSI-594 orientations in Figure 5-16 for comparison.

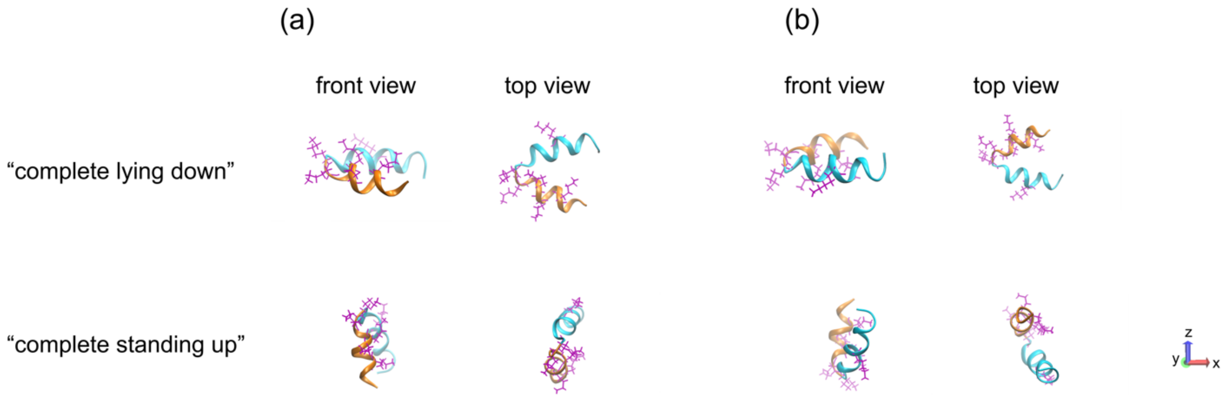

Figure 5-16 (a) Front view (left) and top view (right) of a "complete lying down" (top) representation and a "complete standing up" (bottom) representation of MSI-594-a and (b) front view (left) and top view (right) of a "complete lying down" (top) representation and a "complete standing up" (bottom) representation of MSI-594-b. The membrane surface is parallel to the xy plane.

As presented above, it was found that another pair of orientations (90, 130°) and (90, 310°) with +10° bent angle deviation from the NMR structure 2K98 also have a high matching score, which are therefore possible structures of MSI594 when associated with a POPC or 7:3 POPC/POPG lipid bilayer as well. These two orientations with and without the presence of lipid bilayers are illustrated in Figure 17. When interacting with a lipid bilayer, the peptide structure with the orientation (90, 130°) shows the charged N-terminal to insert into the hydrophobic lipid bilayer core, and the structure with the orientation (90, 310°) shows the hydrophobic C-terminal helix to interact with the hydrophilic lipid headgroup. Neither of these two orientations can be stabilized at the lipid bilayer surface. Therefore, these two membrane orientation possibilities are excluded and not discussed further in this study.

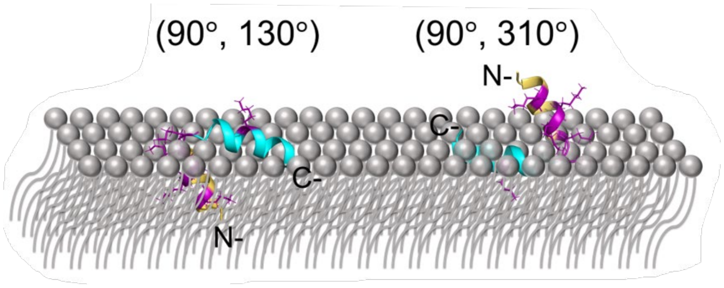

Figure 5-17 The peptides shown in Figure 5-6 are displayed along with the POPC or 7:3 POPC:POPG lipid bilayer. Grey solid spheres represent the lipid headgroups, grey curved lines represent the lipid hydrophobic tails, and the positively charged Lys residues are colored in purple sticks.

The best-matched conformation of MSI-594A on both POPC and 7:3 POPC/POPG lipid surfaces is the modified NMR structure 2L36 with a bent angle change of +10°. The mutation of F5A decreases the intrahelical chain interaction of MSI-594 and makes the bent region more flexible. Therefore, it is possible that MSI-594A opens more when interacting with a POPC lipid bilayer. The best-matched orientations (with the highest matching score) of MSI-594A associated with a POPC lipid bilayer are shown in Figure 5-15 (b). The two possible orientations with the opposite absolute orientation may represent two modes of action. The left orientation shown in Figure 5-15 (b) (noted as MSI-594A-a1) has the N-terminal helical chain (residues 1−11) lying down on the POPC distal leaflet with the charged Lys facing toward the lipid headgroup and the C-terminal helical chain (residues 12−24) pointing toward the aqueous solution. The right orientation in Figure 5-15 (b), noted as MSI-594A-b1, shows that the N-terminal helical chain (residues 1−11) interacts with the lipid bilayer horizontally and the C-terminal helical chain (residues 12−24) inserts into the POPC distal leaflet.

The best-matched orientations (with the highest matching score) of MSI-594A associated with a 7:3 POPC/POPG lipid bilayer are shown in Figure 5-15 (c). The membrane orientation shown in Figure 5-15 (c) (left; noted as MSI-594A-a2) is highly similar to the representation of MSI-594A-a1 shown in Figure 5-15 (b) (left), and the other orientation shown in Figure 5-15 (c) (right; noted as MSI-594A-b2) is highly similar to MSI-594A-b1 shown in Figure 5-15 (b) (right). As mentioned above, at very low peptide concentrations, an antimicrobial peptide can have a similar interaction with the bilayer surface of neutral or charged lipids.[5] Therefore, it is possible that similar to MSI-594, MSI-594A can adopt similar conformation and orientation on both POPC and 7:3 POPC/POPG lipid bilayer surfaces and can potentially disrupt both lipid membranes via a similar mechanism as discussed below.

Even though the positively charged helical N-terminus segment interacting with lipid headgroups stabilizes its conformation and orientation at the membrane interface for MSI-594A-a1 (or MSI-594A-a2), the exposure of the hydrophobic C-terminal helical segment to the aqueous environment does not favor a stable helical structure. Therefore, MSI-594Aa1 (or MSI-594A-a2) orientation is less likely to be a reasonable orientation of MSI-594A in the POPC (or 7:3 POPC/POPG) lipid bilayer. For MSI-594A-b1 (or MSI-594Ab2), on the other hand, while keeping the charged N-terminal interacting with the lipid headgroup, the hydrophobic C-terminal is inserted into the lipid hydrophobic core to interact with the lipid acyl chains. This “membrane-inserted” representation can stabilize both the N- and C-terminal helical segments and thus should be a more likely orientation to represent MSI-594A associated with POPC (or 7:3 POPC/ POPG) lipid bilayers. Such a “membrane-inserted” representation indicates that MSI-594A can disrupt the POPC or 7:3 POPC/POPG lipid bilayer via a carpet mechanism, which is in agreement with previously published results.[6]

Interestingly, even though both MSI-594 and MSI-594A can take a carpet mechanism to interact with both lipid bilayers (POPC and 7:3 POPC/POPG), MSI-594 adopts a “face-on” membrane representation, while MSI-594A adopts a "membrane-inserted" orientation to have interaction with the lipid bilayer surface. In addition, the best-matched conformation of MSI-594A (+10° bent angle deviation to the NMR structure 2L36) opens the bent angle more and allows the two helical ends to interact with lipids more flexibly (e.g., each terminal can interact with different regions of a lipid). In contrast, the best-matched conformation of MSI-594 (+5° bent angle deviation to the NMR structure 2K98) shows stronger intramolecular interaction between the two helical segments and restricts the peptide to interact with the bilayer surface via a "face-on" orientation. Such a difference in the conformation and orientation between MSI-594 and MSI594A is a result of the mutation F5A. The lack of the Phe-5 residue dramatically decreases the intramolecular interaction in the mutant peptide MSI-594A and allows it to interact with the lipid bilayer surface differently compared to MSI-594. It is worth mentioning that SFG can only deduce the peptide's conformation and orientation but cannot deduce the distance between the peptide and the lipid bilayer while the peptide is associated with a lipid bilayer. Figure 5-18 also shows the same orientations as MSI-594A-b1 and MSI-594Ab2 without inserting into the lipid bilayer. We believe that the interaction of peptides under such situations is not as favorable; thus, we exclude such possibilities.

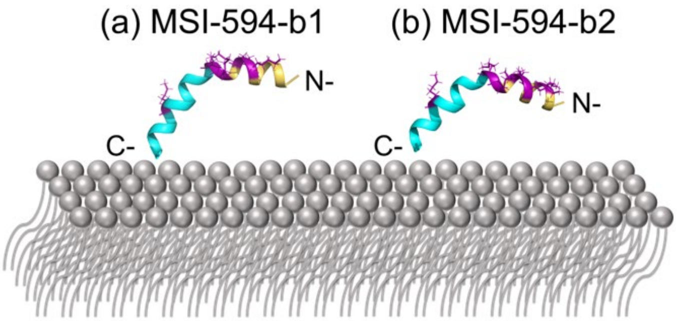

Figure 5-18 (a) MSI-594A-b1 associated with a POPC bilayer without insertion and (b) the MSI-594A-b2 associated with a 7:3 POPC:POPG bilayer without insertion.

Previously, dye-leakage experiments were conducted to confirm the membrane disruption activity of MSI-594 against mixed POPC/POPG lipids. Detailed ^{15}N solid-state NMR studies suggested that MSI-594 transitioned from a random coil structure to a helical structure in lipid environments and the peptide oriented nearly parallel to the lipid bilayer surface.[9] Therefore, MSI-594 most likely functions via a carpet mechanism of membrane disruption, which is in agreement with the findings reported in this study. The replacement of Phe-5 to Ala-5 can lower the antimicrobial ability of MSI-594A against Gram-negative bacteria compared to that exhibited by MSI-594 due to the limited permeabilization of a negatively charged lipid membrane.[6] Such a limited perturbation of MSI594A against a negatively charged lipid membrane can be caused by the fact that MSI-594A partly inserts into the lipid hydrophobic core for its structural and orientational stability, and therefore, MSI-594A has a weaker membrane translocation efficiency than MSI-594. These mechanistic insights are in agreement with previously published results.[6]

5.5 Conclusion

Investigation on the membrane orientation of AMPs is important to understand more details regarding the AMP's action on the cell membrane. SFG is a unique surface-specific technique and has been developed into a powerful tool to study peptide and protein conformation and orientation on the surfaces and at interfaces. In this study, we applied SFG and ATR-FTIR techniques, along with the computational Hamiltonian approach to deduce the preferable membrane conformations and orientations of MSI-594 and MSI-594A on lipid bilayers. Our results demonstrated that the MSI-594 molecule adopts a "face-on" orientation on the bilayer surface. Thus the mode of action of MSI-594 should be carpet mechanism. With only one mutated amino acid, MSI-594A adopts varied conformations and orientations when interacting with the lipid bilayers, even with the same mode of action. MSI-594A molecule adopts a "membrane-inserted" orientation on the bilayer surface and can disrupt the membrane via carpet model. Such result confirms the important role of Phe-5 residue for the membrane structure of MSI-594 peptide. Therefore, our approach can be powerful in understanding the role of a single amino acid residue in a peptide or protein.

In this study, the bent angle between the two helical components in MSI-594 or MSI-594A observed in NMR structures was varied to match the ATR and SFG experimental results that were obtained from peptides associated with lipids of different compositions. It is worth noting that the variation of the bend angle may not be the only possible way to better characterize the different interactions of a peptide with different bilayers. Other possibilities include[21–23]: (i) The end of the peptide may be fraying and/or disordered; (ii) The peptide may not adopt a uniform structural distribution; (iii) The peptides may aggregate to form oligomers at the membrane surface. To further improve the accuracy of the data interpretation, molecular dynamic simulation along with isotope labeling can be utilized to better determine the interfacial peptide/protein structure.[24] In addition, gathering more vibrational information from other regions, e.g. amide III region, can also help to capture the structural changes of peptides at the membrane surface.[25–27] This research showed that MSI-594 and MSI-594A exhibited larger bent angles when associated with a POPC/POPG lipid bilayer compared to those when they were in LPS environments, indicating that varying lipid compositions can cause interfacial structural changes of MSI-594 and its analogue. Thus, this research can only conclude that when associated with a POPC/POPG lipid bilayer, MSI-594 and MSI-594A can interact with the lipid via a carpet model under current experimental conditions. We cannot rule out the possibility that MSI-594 and MSI-594A can adopt pore-forming mechanisms when interacting with other lipids. In the future, more lipid compositions need to be tested to obtain a complete picture of membrane-disrupting mechanisms of MSI-594 and MSI-594A.

Nevertheless, our studies demonstrate that a better understanding of the membrane disrupting mechanisms by AMPs can be obtained by suitably combining the measurements using SFG and ATR-FTIR with the dynamic structures solved by NMR spectroscopy. Since the reported approach enables the use of a combination of different membrane components, it can be used to investigate the roles of lipid domains, clusters, or rafts to further address questions related to the lipid selectivity, membrane curvature, bilayer fluidity/rigidity, and hydrophobic thickness of the membrane. We expect the results reported in this study to be useful not only for the design of potent antimicrobial peptides, but also to better understand the membrane disrupting properties of a variety of biomolecules including amyloid peptides/proteins and fusogenic viral peptides/proteins.

5.6 Reference

(1) Ramamoorthy, A.; Thennarasu, S.; Lee, D.-K.; Tan, A.; Maloy, L. Solid-State NMR Investigation of the Membrane-Disrupting Mechanism of Antimicrobial Peptides MSI-78 and MSI-594 Derived from Magainin 2 and Melittin. *Biophysical Journal* **2006**, *91* (1), 206–216. https://doi.org/10.1529/biophysj.105.073890.

(2) Yang, L.; Harroun, T. A.; Weiss, T. M.; Ding, L.; Huang, H. W. Barrel-Stave Model or Toroidal Model? A Case Study on Melittin Pores. *Biophysical Journal* **2001**, *81* (3), 1475–1485. https://doi.org/10.1016/S0006-3495(01)75802-X.

(3) Sato, H.; Feix, J. B. Peptide–Membrane Interactions and Mechanisms of Membrane Destruction by Amphipathic α-Helical Antimicrobial Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2006**, *1758* (9), 1245–1256. https://doi.org/10.1016/j.bbamem.2006.02.021.

(4) Maloy, W. L.; Kari, U. P. Structure-Activity Studies on Magainins and Other Host Defense Peptides. *Biopolymers* **1995**, *37* (2), 105–122. https://doi.org/10.1002/bip.360370206.

(5) Lee, D.-K.; Bhunia, A.; Kotler, S. A.; Ramamoorthy, A. Detergent-Type Membrane Fragmentation by MSI-78, MSI-367, MSI-594, and MSI-843 Antimicrobial Peptides and Inhibition by Cholesterol: A Solid-State Nuclear Magnetic Resonance Study. *Biochemistry* **2015**, *54* (10), 1897–1907. https://doi.org/10.1021/bi501418m.

(6) Domadia, P. N.; Bhunia, A.; Ramamoorthy, A.; Bhattacharjya, S. Structure, Interactions, and Antibacterial Activities of MSI-594 Derived Mutant Peptide MSI-594F5A in Lipopolysaccharide Micelles: Role of the Helical Hairpin Conformation in Outer-Membrane Permeabilization. *J. Am. Chem. Soc.* **2010**, *132* (51), 18417–18428. https://doi.org/10.1021/ja1083255.

(7) Bhunia, A.; Ramamoorthy, A.; Bhattacharjya, S. Helical Hairpin Structure of a Potent Antimicrobial Peptide MSI-594 in Lipopolysaccharide Micelles by NMR Spectroscopy. *Chemistry – A European Journal* **2009**, *15* (9), 2036–2040. https://doi.org/10.1002/chem.200802635.

(8) Mukherjee, S.; Kar, R. K.; Nanga, R. P. R.; Mroue, K. H.; Ramamoorthy, A.; Bhunia, A. Accelerated Molecular Dynamics Simulation Analysis of MSI-594 in a Lipid Bilayer. *Phys. Chem. Chem. Phys.* **2017**, *19* (29), 19289–19299. https://doi.org/10.1039/C7CP01941F.

(9) Hollmann, A.; Martinez, M.; Maturana, P.; Semorile, L. C.; Maffia, P. C. Antimicrobial Peptides: Interaction With Model and Biological Membranes and Synergism With Chemical Antibiotics. *Front. Chem.* **2018**, *6*, 204. https://doi.org/10.3389/fchem.2018.00204.

(10) Knobloch, J.; Suhendro, D. K.; Zieleniecki, J. L.; Shapter, J. G.; Köper, I. Membrane–Drug Interactions Studied Using Model Membrane Systems. *Saudi Journal of Biological Sciences* **2015**, *22* (6), 714–718. https://doi.org/10.1016/j.sjbs.2015.03.007.

(11) Yang, P.; Ramamoorthy, A.; Chen, Z. Membrane Orientation of MSI-78 Measured by Sum Frequency Generation Vibrational Spectroscopy. *Langmuir* **2011**, *27* (12), 7760–7767. https://doi.org/10.1021/la201048t.

(12) Yang, P.; Glukhova, A.; Tesmer, J. J. G.; Chen, Z. Membrane Orientation and Binding Determinants of G Protein-Coupled Receptor Kinase 5 as Assessed by Combined Vibrational Spectroscopic Studies. *PLOS ONE* **2013**, *8* (11), e82072. https://doi.org/10.1371/journal.pone.0082072.

(13) Yang, P.; Boughton, A.; Homan, K. T.; Tesmer, J. J. G.; Chen, Z. Membrane Orientation of Gαiβ1γ2 and Gβ1γ2 Determined via Combined Vibrational Spectroscopic Studies. *J. Am. Chem. Soc.* **2013**, *135* (13), 5044–5051. https://doi.org/10.1021/ja3116026.

(14) Ding, B.; Soblosky, L.; Nguyen, K.; Geng, J.; Yu, X.; Ramamoorthy, A.; Chen, Z. Physiologically-Relevant Modes of Membrane Interactions by the Human Antimicrobial Peptide, LL-37, Revealed by SFG Experiments. *Sci Rep* **2013**, *3* (1), 1854. https://doi.org/10.1038/srep01854.

(15) Chen, X.; Wang, J.; Boughton, A. P.; Kristalyn, C. B.; Chen, Z. Multiple Orientation of Melittin inside a Single Lipid Bilayer Determined by Combined Vibrational Spectroscopic Studies. *J. Am. Chem. Soc.* **2007**, *129* (5), 1420–1427. https://doi.org/10.1021/ja067446l.

(16) Yang, P.; Wu, F.-G.; Chen, Z. Dependence of Alamethicin Membrane Orientation on the Solution Concentration. *J Phys Chem C Nanomater Interfaces* **2013**, *117* (7), 3358–3365. https://doi.org/10.1021/jp3099522.

(17) Nguyen, K. T.; King, J. T.; Chen, Z. Orientation Determination of Interfacial β-Sheet Structures in Situ. *J. Phys. Chem. B* **2010**, *114* (25), 8291–8300. https://doi.org/10.1021/jp102343h.

(18) Ye, S.; Li, H.; Wei, F.; Jasensky, J.; Boughton, A. P.; Yang, P.; Chen, Z. Observing a Model Ion Channel Gating Action in Model Cell Membranes in Real Time in Situ: Membrane Potential Change Induced Alamethicin Orientation Change. *J. Am. Chem. Soc.* **2012**, *134* (14), 6237–6243. https://doi.org/10.1021/ja2110784.

(19) Verreault, D.; Hua, W.; Allen, H. C. From Conventional to Phase-Sensitive Vibrational Sum Frequency Generation Spectroscopy: Probing Water Organization at Aqueous Interfaces. *The Journal of Physical Chemistry Letters* **2012**, *3* (20), 3012–3028. https://doi.org/10.1021/jz301179g.

(20) Ji, N.; Ostroverkhov, V.; Chen, C.-Y.; Shen, Y.-R. Phase-Sensitive Sum-Frequency Vibrational Spectroscopy and Its Application to Studies of Interfacial Alkyl Chains. *Journal of the American Chemical Society* **2007**, *129* (33), 10056–10057. https://doi.org/10.1021/ja071989t.

(21) Afrose, F.; McKay, M. J.; Mortazavi, A.; Suresh Kumar, V.; Greathouse, D. V.; Koeppe, R. E. Transmembrane Helix Integrity versus Fraying To Expose Hydrogen Bonds at a Membrane–Water Interface. *Biochemistry* **2019**, *58* (6), 633–645. https://doi.org/10.1021/acs.biochem.8b01119.

(22) Khandelia, H.; Ipsen, J. H.; Mouritsen, O. G. The Impact of Peptides on Lipid Membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2008**, *1778* (7–8), 1528–1536. https://doi.org/10.1016/j.bbamem.2008.02.009.

(23) Tan, J.; Zhang, J.; Luo, Y.; Ye, S. Misfolding of a Human Islet Amyloid Polypeptide at the Lipid Membrane Populates through β-Sheet Conformers without Involving α-Helical Intermediates. *J. Am. Chem. Soc.* **2019**, *141* (5), 1941–1948. https://doi.org/10.1021/jacs.8b08537.

(24) Guo, W.; Lu, T.; Crisci, R.; Nagao, S.; Wei, T.; Chen, Z. Determination of Protein Conformation and Orientation at Buried Solid/Liquid Interfaces. *Chem. Sci.* **2023**, 10.1039.D2SC06958J. https://doi.org/10.1039/D2SC06958J.

(25) Huang, J.; Tian, K.; Ye, S.; Luo, Y. Amide III SFG Signals as a Sensitive Probe of Protein Folding at Cell Membrane Surface. *J. Phys. Chem. C* **2016**, *120* (28), 15322–15328. https://doi.org/10.1021/acs.jpcc.6b05677.

(26) Ye, S.; Li, H.; Yang, W.; Luo, Y. Accurate Determination of Interfacial Protein Secondary Structure by Combining Interfacial-Sensitive Amide I and Amide III Spectral Signals. *J. Am. Chem. Soc.* **2014**, *136* (4), 1206–1209. https://doi.org/10.1021/ja411081t.

(27) Hu, X.; Tan, J.; Ye, S. Reversible Activation of PH-Responsive Cell-Penetrating Peptides in Model Cell Membrane Relies on the Nature of Lipid. *The Journal of Physical Chemistry C* **2017**, *121* (28), 15181–15187. https://doi.org/10.1021/acs.jpcc.7b03092.

Chapter 6 Conclusions and Future Work

In this thesis, a systematic methodology was developed to study the conformation and orientation of interfacial proteins and peptides at buried solid/liquid interfaces in situ. The methodology contains four main modules: (1) experimental measurements by vibrational spectroscopic techniques; (2) in-silico modeling of the structures and dynamics of interfacial proteins and peptides; (3) Hamiltonian approach of spectral calculations as a function of interfacial protein/peptide orientation; and (4) score system that assesses the similarity between the experimental measured spectra and the calculated spectra using the Hamiltonian approach. In addition, this thesis research also demonstrated that isotope labeling on proteins can greatly improve the accuracy of the results obtained by using the above systematic structural determination methodology. In this chapter, the current achievements for each module will be summarized and further directions of development will be proposed.

6.1 Summary of the Achievements in This Thesis Research

Interfacial proteins and peptides play important roles in many applications and research areas such as biomedical materials, antifouling coatings, biosensors, antimicrobial peptides, membrane proteins etc. The properties and functions of interfacial proteins/peptides are determined by their molecular structures. Therefore it is significant to elucidate such structures in situ in real time, which is challenging due to the lack of appropriate in situ interfacial sensitive analytical techniques. The systematic methodology which can be used to elucidate interfacial protein/peptide conformation and orientation added an important biophysical and analytical tool to investigate structures of biological molecules. This method is general and can be applied to study many different kinds of peptides and proteins at a variety of different interfaces. In this thesis, we applied this method to study several important interfacial protein/peptide systems including surface immobilized peptides and enzymes, membrane associated peptides, protein adsorbed on polymers, demonstrating the power of this method in studying molecular structures of interfacial proteins. and peptides.

In Chapter 2, we used an old data analysis method, bond additivity model, to study surface immobilized AMPs and their interactions with Gram-negative bacteria membrane components LPS. This research quantified the orientations of each immobilized AMP variant upon binding to various LPS samples with a hypothesized non-linear helical model. These deduced orientations reflected the binding affinity of different AMP-LPS pairs and can potentially be used as indicators to distinguish LPS samples. This research shows that engineering the immobilization sites of the AMPs can further modulate the interactions between the immobilized AMPs and their intended interaction partners, and therefore can tune their performance of bio-recognition.

In Chapter 3, a new systematic methodology was developed to better resolve both the conformation and orientation of interfacial proteins and peptides. MD simulation was incorporated into the new methodology to provide candidates to approximate interfacial protein/peptide structures with better accuracy. These simulated structures were input into the Hamiltonian program to calculate SFG spectra as a function of orientation. Compared to the previous bond additivity model, the Hamiltonian program offers significant improvements. Firstly, it enables the calculation of the entire SFG spectra of different polarization combinations. Thus, more fitting parameters (peak centers, peak widths, and $\chi^{(2)}$ ratios) can be obtained to compare with experimental measurements, whereas the bond additivity model only considered the $\chi^{(2)}$ ratios between SFG spectra of different polarization combinations. Secondly, the Hamiltonian program considers the signal contributions from all the peptide units in a peptide/protein along with the coupling between each amide unit within the protein/peptide, whereas the bond additivity model only examined contributions from α-helical structure.

This new systematic methodology was applied to deduce the conformations and orientations of C-terminus immobilized super uranyl binding protein (SUP-C105), and the results showed that SUP was lying prone on the surface, primary driven by the strong hydrophobic interactions with the surface. Moreover, the binding with uranyl ions has minimal impact on both the conformation and orientation of SUP-C105. The results likely suggest that fine-tuning the hydrophobicity of the substrate surface/protein interface could yield further improvements on the orientation control of immobilized SUP for more effective uranyl ion binding.

In Chapter 4, the systematic methodology underwent significant enhancements in terms of accuracy. Isotope labeling was introduced into protein sample preparation, leading to the split of vibrational coupling and thus enabling more independent SFG measurements. Additionally, the discrete MD simulation used in Chapter 3 was replaced by atomistic MD simulation to achieve higher precision. Furthermore, a point-by-point score system was developed and improved the efficacy for the comparison of multi-peak spectra. The optimized methodology was then utilized to study the conformations and orientations of B_1 domain of protein G (GB1) and its mutant. The findings highlighted a remarkable orientation transition in GB1, where the mutation of only two hydrophobic amino acids in the helical segments facilitated a shift from a lying down pose to a standing up pose of GB1 on the polystyrene surface. This research further developed the systematic approach to determine protein conformations and orientations at interfaces in situ with SFG experiments supplemented by in-depth data analysis and other methods.

In Chapter 5, the systematic method was further improved by incorporating ATR-FTIR spectroscopy to enrich experimental measurements. Then this method was used to investigate the interaction of AMP MSI-594 and its analogue MSI-594A with model cell membrane. The deduced membrane orientations of these two peptides were successfully correlated to their effectiveness of killing Gram-negative bacteria.

In conclusion, a systematic methodology to elucidate molecular structures of interfacial proteins and peptides was established in this thesis research, which exhibits remarkable power in determining the conformation and orientation of surface immobilized AMPs, surface immobilized enzymes, physically adsorbed proteins on polymers, and AMPs interacting with model cell membranes. This methodology is generally applicable which can be used to study many different kinds of interfacial proteins and peptides, establishing their interfacial structure-function relationships, providing valuable insights into their interfacial functions, and aiding in the improvement of their functions by optimizing their interfacial structures.

6.2 Future Directions

6.2.1 Experimental Measurements by Vibrational Spectroscopic Techniques

In this thesis, a picosecond SFG spectrometer was used as a main tool to collect vibrational spectra from the interfacial proteins and peptides with a sub-monolayer surface sensitivity. Equipped with a narrowband laser source, an SFG spectrum can be obtained by scanning the IR frequency. Using this scanning mode, taking one SFG amide I spectrum (1500 – 1800 cm^{-1}) requires ~ 20 minutes with adequate signal-to-noise (S/N) ratio. For time-dependent SFG studies, only SFG signals at one or two frequencies can be monitored simultaneously as a function of time, and it usually takes ~ 2 seconds to record a data point. Therefore, narrowband picosecond SFG spectroscopy is not a suitable technique to study fast structural changes (< 1 second) of interfacial proteins and peptides. To achieve better time resolution for dynamic studies, femtosecond broadband SFG spectroscopy can be applied. Shortening the pulse width of the laser source to femtoseconds broadens the IR bandwidth to hundreds of wavenumbers. Therefore, every IR pulse covers the entire amide I spectrum frequency range. Therefore, a complete scan obtained by a picosecond SFG system (which takes ~20 mins) can be easily obtained by a broadband SFG system in less than a second, making it possible to monitor the dynamics and kinetics of interfacial proteins and peptides. The structure determination for a dynamic process, e.g. protein adsorption onto a polymer surface, can be easily performed by using the developed methodology to analyze each SFG spectrum collected along the dynamic process.[1,2]

In a homodyne SFG (regardless of narrowband or broadband) experiments, only the information of squared $\chi^{(2)}$ can be obtained, therefore, the two protein or peptide orientations pointing to the absolute opposite directions cannot be differentiated by using SFG results alone. In this thesis research, with the help of detailed analysis and discussion on the physical restrictions of the interfacial peptides and proteins, we could identify the more appropriate orientation from the two orientations with opposite absolute orientations. Experiments, to differentiate these absolute opposite orientations, phase-sensitive SFG can be employed. In a phase-sensitive SFG system, the phase of the $\chi^{(2)}$ of the sample of interest can be deduced from its interference with a reference sample, as illustrated in previous publications.[3–5] Therefore, phase-sensitive SFG spectroscopy makes it promising to directly protein the absolute orientation of the interfacial proteins and peptides.

In Chapter 5, SFG spectroscopy was coupled with ATR-FTIR spectroscopy to study the membrane-bound AMPs. In the future, more vibrational spectroscopic methods, such as Raman spectroscopy, could be incorporated into the data interpretation loop for more independent experimental measurements.

In Chapter 4, we demonstrated the power of using isotope-labeled proteins with one type of amino acid labeled in each molecule to increase the independently measured parameters to determine the interfacial structures of proteins with higher precision. In the future, isotope-labeled proteins with multiple types of amino acids labeled in each molecule can be designed to further increase independently measured parameters. Additionally, isotope labeling can be employed in differentiating protein signals contributed from different components in protein conjugates or protein complexes, expanding investigations from a single protein to protein mixtures. As the isotope labeling technique advances, it will become a routine approach in sample preparation for our systematic methodology.

6.2.2 In-silico Modeling Methods

MD simulations can be roughly categorized into continuous MD simulation and discontinuous MD simulation. In continuous MD, a continuous potential was used to describe the system for each time step of the simulation. Take atomistic MD simulation as an example, Newton's equation was used to compute the net force applied on each atom by summing the pairwise forces between all the atoms in the system. Subsequentially, an accelerate vector can be computed to determine the new position and velocity of the next time step based on the current position and velocity. This process will iterate until the desired or feasible time scale is reached. To prevent data overflow, each time step in such a simulation is usually kept small in the magnitude of femtoseconds. Therefore, although the pairwise force and energy potential can be calculated precisely for each time step, the atomistic MD simulation can be time-intensive and expensive. Discontinuous MD simulations, on the other hand, use step-like functions to approximate the potential for each time step, meaning the atoms or particles will move in the same velocity until their inter-distance reaches a set limit. Although such an approximation may sacrifice the accuracy of atom/particle movements, the computational expense can be greatly reduced and longer time scale simulations are feasible. In conclusion, continuous MD and discontinuous MD each possess their own advantages and disadvantages. In Chapter 3, discontinuous MD simulation was employed to simulate the structure of surface-immobilized SUP for rapid parallel computations and in Chapter 4, continuous MD simulation was used to more accurately monitor and dynamics of GB1 adsorption onto PS surface. In the future, the combination of continuous MD and discontinuous MD simulation was a promising direction to maintain high accuracy in a longer time scale. For example, the time window of interest can be monitored by continuous MD simulations while other less important time windows can be expedited by discontinuous MD simulations, as illustrated in some publications.[6,7]

In this research, the molecular structures of interfacial peptides and proteins obtained from continuous MD simulations (chapter 4) and discontinuous MD simulations (chapter 3) were used as input structures for SFG spectral calculations. For such simulations, we used crystal structures (chapter 3 and 4) and predicted structures (chapter 4) as initial peptide/protein structures. The structural prediction (from sequence) methodologies, such as AlphaFold[8], have been developed very fast in recent years with the fast development of both hardware and software computer technologies. Such structural prediction method will be more and more widely used to study many new and complicated proteins and peptides in solutions, which will greatly help our data analysis on interfacial peptides and proteins.

6.2.3 Hamiltonian Approach

A Hamiltonian matrix consists of two main parts: diagonal terms (local modes of amide units) and off-diagonal terms (couplings between each amide unit). In our current Hamiltonian approach, the diagonal terms were simply set as a fixed value (~ 1640 cm^{-1}) for all amide units within a protein/peptide. However, the local frequency can be amino acid specific and can be shifted due to the solvent effects (e.g. hydrogen bonding).[9–11] For example, hydrogen bonding to the C=O bond within the protein/peptide groups increases the C=O bond length and thus leads to the red shift of the local amide frequency.[9–11] Moreover, the bond length change of C=O bond (and C-N bond) can also be caused by amino acid side chains and the connectivity between two amino acids, resulting in various local frequencies. In the future, amino acid specific local frequency can be approximated by density function theory (DFT) calculations[12], and the solvent effects can be considered by some empirical formulas published previously[9–11,13,14], or be estimated by MD simulations.

The off-diagonal terms in our current Hamiltonian approach are mainly considered by through-dipole couplings (TDC, equation (1-28)). However, previous research found that TDC alone may not be adequate to describe the couplings between two neighboring amide units. Some transferable maps, as a function of Ramachandran angles between neighboring amide units, can be constructed to improve the accuracy of TDC.[10,11,15,16] Additionally, in the consideration of electrostatic potential and electric fields, through-charge coupling (TCC) needs to be included if the coupling is dominant by electrostatic interactions.[11] Cho et al also proposed that the inter-strand coupling of β-sheets is not ignorable in some cases[14], and therefore suggests that the secondary structures of the proteins and peptides may need to be considered as factors that alter the amide unit couplings. Thus, additional considerations of coupling corrections are needed in our future studies.

In this thesis, the local mode transition dipole and polarizability used to construct the normal mode response were the average values for a typical protein.[17] It is necessary to better approximate the structure-specific transition dipole and polarizability for a unique protein/peptide. Obtaining these parameters from experiments can be time consuming. Computationally, Jiang et al. employed machine learning methods to train and predict transition dipole and polarizability for various structural descriptors, which can be introduced into our methodology in the future.[18,19]

Currently, only one structure was allowed as input to Hamiltonian calculations at a time, and it was assumed that the interfacial orientation of a protein/peptide followed a δ-distribution. However, in real situations, there should be a distribution of interfacial conformations and orientations for a protein/peptide. Therefore, multi-inputs[20] Hamiltonian program for spectral calculations under a Gaussian distribution can be further developed to better approximate the real situation.

6.2.4 Score System

In this thesis, a parameter comparison score system was developed for single peak spectra comparison, and a point-by-point score system (weighted and unweighted) was developed for multiple peak spectra comparison. In the future, it is necessary to set up uniform standards for spectra comparison and methods to access the quality of the matching.

6.2.5 Development of Other Useful Methodologies

Further developments including structural prediction methods using SFG spectra can be adopted. Possible extensions of current methodology can be (1) SFG-aided structure prediction/optimization, similar to NMR-aided prediction[21] and cryo-EM aided density based structure prediction[22]; and (2) SFG-based structure classification for interfacial proteins and peptides.

6.3 Limitations and Challenges

Although SFG is a powerful vibrational spectroscopic tool to investigate interfacial proteins/peptides, it is challenging to apply SFG to study the structure distribution of interfacial proteins/peptides. The current data analysis is based on a δ-distribution assumption of interfacial peptide/protein orientation. However, real situations can be much more complex. For example, when interacting with lipids, AMP may form peptide bundles/oligomers instead of adopting a single orientation/conformation. Development of statistical models to consider the peptide orientation/conformation distribution and the combined use of other techniques (e.g. imaging techniques[23,24]) can potentially overcome this limitation.

Additionally, the observed SFG spectra can be greatly affected by non-resonant (NR) backgrounds. NR background can distort resonant SFG spectra. Thus, removing such background interference from the observed SFG spectra is necessary in order to reliably analyze resonant SFG signals. Extracting NR contributions may need physical/mathematical modeling of signal contributions (such as equation (1 - 2)) followed by spectra fitting to determine NR term(s), which can be complicated and sometimes oversimplified[25]. More systematic effort is needed to extract/remove NR contributions in SFG spectra in the future.

6.4 References

(1) Lu, T.; Fu, L.; Qiu, Y.; Zhang, J.; Chen, Z. Probing Molecular Interactions of Antibody Drugs, Silicone Oil, and Surfactant at Buried Interfaces In Situ. *Anal. Chem.* **2022**, *94* (42), 14761–14768. https://doi.org/10.1021/acs.analchem.2c03425.

(2) Lu, T.; Chen, Z. Monitoring the Molecular Structure of Fibrinogen during the Adsorption Process at the Buried Silicone Oil Interface In Situ in Real Time. *J. Phys. Chem. Lett.* **2023**, *14* (13), 3139–3145. https://doi.org/10.1021/acs.jpclett.3c00331.

(3) Ji, N.; Ostroverkhov, V.; Chen, C.-Y.; Shen, Y.-R. Phase-Sensitive Sum-Frequency Vibrational Spectroscopy and Its Application to Studies of Interfacial Alkyl Chains. *Journal of the American Chemical Society* **2007**, *129* (33), 10056–10057. https://doi.org/10.1021/ja071989t.

(4) Sun, S.; Bisson, P. J.; Bonn, M.; Shultz, M. J.; Backus, E. H. G. Phase-Sensitive Sum-Frequency Generation Measurements Using a Femtosecond Nonlinear Interferometer. *J. Phys. Chem. C* **2019**, *123* (12), 7266–7270. https://doi.org/10.1021/acs.jpcc.9b00861.

(5) Chen, X.; Hua, W.; Huang, Z.; Allen, H. C. Interfacial Water Structure Associated with Phospholipid Membranes Studied by Phase-Sensitive Vibrational Sum Frequency Generation Spectroscopy. *Journal of the American Chemical Society* **2010**, *132* (32), 11336–11342. https://doi.org/10.1021/ja1048237.

(6) Mukherjee, G.; Nandekar, P. P.; Wade, R. C. An Electron Transfer Competent Structural Ensemble of Membrane-Bound Cytochrome P450 1A1 and Cytochrome P450 Oxidoreductase. *Commun Biol* **2021**, *4* (1), 55. https://doi.org/10.1038/s42003-020-01568-y.

(7) Huang, H.; Zhang, C.; Crisci, R.; Lu, T.; Hung, H.-C.; Sajib, M. S. J.; Sarker, P.; Ma, J.; Wei, T.; Jiang, S.; Chen, Z. Strong Surface Hydration and Salt Resistant Mechanism of a New Nonfouling Zwitterionic Polymer Based on Protein Stabilizer TMAO. *J. Am. Chem. Soc.* **2021**, *143* (40), 16786–16795. https://doi.org/10.1021/jacs.1c08280.

(8) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(9) Roeters, S. J.; van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J. Phys. Chem. A* **2013**, *117* (29), 6311–6322. https://doi.org/10.1021/jp401159r.

(10) la Cour Jansen, T.; Knoester, J. A Transferable Electrostatic Map for Solvation Effects on Amide I Vibrations and Its Application to Linear and Two-Dimensional Spectroscopy. *The Journal of Chemical Physics* **2006**, *124* (4), 044502. https://doi.org/10.1063/1.2148409.

(11) la Cour Jansen, T.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. Modeling the Amide I Bands of Small Peptides. *The Journal of Chemical Physics* **2006**, *125* (4), 044312. https://doi.org/10.1063/1.2218516.

(12) Gorbunov, R. D.; Kosov, D. S.; Stock, G. *Ab Initio* -Based Exciton Model of Amide I Vibrations in Peptides: Definition, Conformational Dependence, and Transferability. *The Journal of Chemical Physics* **2005**, *122* (22), 224904. https://doi.org/10.1063/1.1898215.

(13) Ham, S.; Kim, J.-H.; Lee, H.; Cho, M. Correlation between Electronic and Molecular Structure Distortions and Vibrational Properties. II. Amide I Modes of NMA–ND2O Complexes. 9.

(14) Lee, C.; Cho, M. Local Amide I Mode Frequencies and Coupling Constants in Multiple-Stranded Antiparallel β-Sheet Polypeptides. *J. Phys. Chem. B* **2004**, *108* (52), 20397–20407. https://doi.org/10.1021/jp0471204.

(15) Torii, H.; Tasumi, M. Application of the Three-dimensional Doorway-state Theory to Analyses of the Amide-I Infrared Bands of Globular Proteins. *The Journal of Chemical Physics* **1992**, *97* (1), 92–98. https://doi.org/10.1063/1.463528.

(16) Torii, H.; Tasumi, M. Three-dimensional Doorway-state Theory for Analyses of Absorption Bands of Many-oscillator Systems. *The Journal of Chemical Physics* **1992**, *97* (1), 86–91. https://doi.org/10.1063/1.463526.

(17) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. The Two-Dimensional IR Nonlinear Spectroscopy of a Cyclic Penta-Peptide in Relation to Its Three-Dimensional Structure. *Proceedings of the National Academy of Sciences* **1999**, *96* (5), 2036–2041. https://doi.org/10.1073/pnas.96.5.2036.

(18) Feng, C.; Xi, J.; Zhang, Y.; Jiang, B.; Zhou, Y. Accurate and Interpretable Dipole Interaction Model-Based Machine Learning for Molecular Polarizability. *J. Chem. Theory Comput.* **2023**, *19* (4), 1207–1217. https://doi.org/10.1021/acs.jctc.2c01094.

(19) Zhang, Y.; Jiang, J.; Jiang, B. Learning Dipole Moments and Polarizabilities. In *Quantum Chemistry in the Age of Machine Learning*; Elsevier, 2023; pp 453–465. https://doi.org/10.1016/B978-0-323-90049-2.00019-6.

(20) Lu, T.; Guo, W.; Chen, Z. Probing Protein Aggregation at Buried Interfaces: Distinguishing Adsorbed Protein Monomer, Dimer, and Monomer-Dimer Mixture in Situ.

(21) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **2016**, *428* (4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

(22) Chirigati, F. Predicting Protein Structure from Cryo-EM Data. *Nat Comput Sci* **2021**, *1* (2), 96–96. https://doi.org/10.1038/s43588-021-00035-w.

(1) Lu, T.; Fu, L.; Qiu, Y.; Zhang, J.; Chen, Z. Probing Molecular Interactions of Antibody Drugs, Silicone Oil, and Surfactant at Buried Interfaces In Situ. *Anal. Chem.* **2022**, *94* (42), 14761–14768. https://doi.org/10.1021/acs.analchem.2c03425.

(2) Lu, T.; Chen, Z. Monitoring the Molecular Structure of Fibrinogen during the Adsorption Process at the Buried Silicone Oil Interface In Situ in Real Time. *J. Phys. Chem. Lett.* **2023**, *14* (13), 3139–3145. https://doi.org/10.1021/acs.jpclett.3c00331.

(3) Ji, N.; Ostroverkhov, V.; Chen, C.-Y.; Shen, Y.-R. Phase-Sensitive Sum-Frequency Vibrational Spectroscopy and Its Application to Studies of Interfacial Alkyl Chains.

Journal of the American Chemical Society **2007**, *129* (33), 10056–10057. https://doi.org/10.1021/ja071989t.

(4) Sun, S.; Bisson, P. J.; Bonn, M.; Shultz, M. J.; Backus, E. H. G. Phase-Sensitive Sum-Frequency Generation Measurements Using a Femtosecond Nonlinear Interferometer. *J. Phys. Chem. C* **2019**, *123* (12), 7266–7270. https://doi.org/10.1021/acs.jpcc.9b00861.

(5) Chen, X.; Hua, W.; Huang, Z.; Allen, H. C. Interfacial Water Structure Associated with Phospholipid Membranes Studied by Phase-Sensitive Vibrational Sum Frequency Generation Spectroscopy. *Journal of the American Chemical Society* **2010**, *132* (32), 11336–11342. https://doi.org/10.1021/ja1048237.

(6) Mukherjee, G.; Nandekar, P. P.; Wade, R. C. An Electron Transfer Competent Structural Ensemble of Membrane-Bound Cytochrome P450 1A1 and Cytochrome P450 Oxidoreductase. *Commun Biol* **2021**, *4* (1), 55. https://doi.org/10.1038/s42003-020-01568-y.

(7) Huang, H.; Zhang, C.; Crisci, R.; Lu, T.; Hung, H.-C.; Sajib, M. S. J.; Sarker, P.; Ma, J.; Wei, T.; Jiang, S.; Chen, Z. Strong Surface Hydration and Salt Resistant Mechanism of a New Nonfouling Zwitterionic Polymer Based on Protein Stabilizer TMAO. *J. Am. Chem. Soc.* **2021**, *143* (40), 16786–16795. https://doi.org/10.1021/jacs.1c08280.

(8) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(9) Roeters, S. J.; van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J. Phys. Chem. A* **2013**, *117* (29), 6311–6322. https://doi.org/10.1021/jp401159r.

(10) la Cour Jansen, T.; Knoester, J. A Transferable Electrostatic Map for Solvation Effects on Amide I Vibrations and Its Application to Linear and Two-Dimensional Spectroscopy. *The Journal of Chemical Physics* **2006**, *124* (4), 044502. https://doi.org/10.1063/1.2148409.

(11) la Cour Jansen, T.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. Modeling the Amide I Bands of Small Peptides. *The Journal of Chemical Physics* **2006**, *125* (4), 044312. https://doi.org/10.1063/1.2218516.

(12) Gorbunov, R. D.; Kosov, D. S.; Stock, G. *Ab Initio* -Based Exciton Model of Amide I Vibrations in Peptides: Definition, Conformational Dependence, and Transferability. *The Journal of Chemical Physics* **2005**, *122* (22), 224904. https://doi.org/10.1063/1.1898215.

(13) Ham, S.; Kim, J.-H.; Lee, H.; Cho, M. Correlation between Electronic and Molecular Structure Distortions and Vibrational Properties. II. Amide I Modes of NMA–ND2O Complexes. 9.

(14) Lee, C.; Cho, M. Local Amide I Mode Frequencies and Coupling Constants in Multiple-Stranded Antiparallel β-Sheet Polypeptides. *J. Phys. Chem. B* **2004**, *108* (52), 20397–20407. https://doi.org/10.1021/jp0471204.

(15) Torii, H.; Tasumi, M. Application of the Three-dimensional Doorway-state Theory to Analyses of the Amide-I Infrared Bands of Globular Proteins. *The Journal of Chemical Physics* **1992**, *97* (1), 92–98. https://doi.org/10.1063/1.463528.

(16) Torii, H.; Tasumi, M. Three-dimensional Doorway-state Theory for Analyses of Absorption Bands of Many-oscillator Systems. *The Journal of Chemical Physics* **1992**, *97* (1), 86–91. https://doi.org/10.1063/1.463526.

(17) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. The Two-Dimensional IR Nonlinear Spectroscopy of a Cyclic Penta-Peptide in Relation to Its Three-Dimensional Structure. *Proceedings of the National Academy of Sciences* **1999**, *96* (5), 2036–2041. https://doi.org/10.1073/pnas.96.5.2036.

(18) Feng, C.; Xi, J.; Zhang, Y.; Jiang, B.; Zhou, Y. Accurate and Interpretable Dipole Interaction Model-Based Machine Learning for Molecular Polarizability. *J. Chem. Theory Comput.* **2023**, *19* (4), 1207–1217. https://doi.org/10.1021/acs.jctc.2c01094.

(19) Zhang, Y.; Jiang, J.; Jiang, B. Learning Dipole Moments and Polarizabilities. In *Quantum Chemistry in the Age of Machine Learning*; Elsevier, 2023; pp 453–465. https://doi.org/10.1016/B978-0-323-90049-2.00019-6.

(20) Lu, T.; Guo, W.; Chen, Z. Probing Protein Aggregation at Buried Interfaces: Distinguishing Adsorbed Protein Monomer, Dimer, and Monomer-Dimer Mixture in Situ.

(21) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **2016**, *428* (4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

(22) Chirigati, F. Predicting Protein Structure from Cryo-EM Data. *Nat Comput Sci* **2021**, *1* (2), 96–96. https://doi.org/10.1038/s43588-021-00035-w.

(23) Pleshakova, T.; Bukharina, N.; Archakov, A.; Ivanov, Y. Atomic Force Microscopy for Protein Detection and Their Physicochemical Characterization. *IJMS* **2018**, *19* (4), 1142. https://doi.org/10.3390/ijms19041142.

(24) Longchamp, J.-N.; Rauschenbach, S.; Abb, S.; Escher, C.; Latychevskaia, T.; Kern, K.; Fink, H.-W. Imaging Proteins at the Single-Molecule Level. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (7), 1474–1479. https://doi.org/10.1073/pnas.1614519114.

(25) Curtis, A. D.; Burt, S. R.; Calchera, A. R.; Patterson, J. E. Limitations in the Analysis of Vibrational Sum-Frequency Spectra Arising from the Nonresonant Contribution. *J. Phys. Chem. C* **2011**, *115* (23), 11550–11559. https://doi.org/10.1021/jp200915z.