2016

CRYSTALLOGRAPHIC STUDY OF MUTANT INFLUENZA MATRIX (M1) PROTEIN AND AFFINITY STUDY OF SMALL MOLECULE INHIBITORS TOWARD M1 AND GROWTH FACTORS

Bashayer althufairi

CRYSTALLOGRAPHIC STUDY OF MUTANT INFLUENZA MATRIX (M1) PROTEIN AND AFFINITY STUDY OF SMALL MOLECULE INHIBITORS TOWARD M1 AND GROWTH FACTORS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

By

Bashayer Althufairi
B pharm, University of Kuwait, Jabria, Kuwait 2011

Director: UMESH R. DESAI
PROFESSOR OF MEDICINAL CHEMISTRY AND DIRECTOR OF INSTITUTE OF STRUCTURAL BIOLOGY AND DRUG DISCOVERY AND DEVELOPMENT

Virginia Commonwealth University
Richmond, Virginia
September 2016

Acknowledgements

I'd like to thank Dr. Umesh Desai for his great help and support during my study at Virginia Commonwealth University. I also like to acknowledge my lab members for their useful discussion and trouble shooting. Special thanks to Dr. Faik and Yingzi for their help in Crystallography and Surface Plasmon Resonance studies. I would love to thank my committee members Dr. Martin Safo and Dr. Adam Hawkridge. I'm very appreciative to Kuwait University for funding and support to get my degree and knowledge. I'd like to acknowledge Department of Medicinal Chemistry and Institute of Structural Biology and Drug Discovery and Development for the knowledge and scientific experience they provide for the students and scholars.

I also like to thank my parents and friends for their support and encouragement during the time I've spent in Richmond, VA.

Table of Contents

ACKNOWLEDGEMENTS II

LIST OF TABLES V

ABSTRACT X

CHAPTER 1. INFLUENZA A VIRUS (IAV) 1
- 1.1 OVERVIEW 1
- 1.2 INFLUENZA VIRUS 4
- 1.3 STRUCTURAL VIRAL PROTEINS OTHER THAN MATRIX M1 PROTEIN 9
  - *1.3.1 Hemagglutinin (HA) 9*
  - *1.3.2 Neuraminidase (NA) 9*
  - *1.3.3 M2 Protein 10*
- 1.4 MATRIX M1 PROTEIN 12
- 1.5 STRUCTURE AND FUNCTIONS 13
- 1.6 MUTATION STUDIES ON M1 PROTEIN 16
- 1.7 CRYSTALLOGRAPHIC STUDIES ON M1 PROTEIN 19

CHAPTER 2. FUNDAMENTALS OF PROTEIN X-RAY CRYSTALLOGRAPHY 23
- 2.1 OVERVIEW 23
- 2.2 PRINCIPLES OF X-RAY STRUCTURE DETERMINATION 23
- 2.3 CRYSTALLIZATION TECHNIQUES 28
- 2.4 CRYSTAL SYMMETRY 32
- 2.5 X-RAY DIFFRACTION OF A CRYSTAL 37
- 2.6 ELECTRON DENSITY AND PHASE PROBLEM 41
  - *2.6.1 Heavy Atom Method 42*
  - *2.6.2 Anomalous Scattering 42*
  - *2.6.3 Molecular Replacement 43*
- 2.7 STRUCTURE REFINEMENT AND VALIDATION 43
  - *2.7.1 Ramachandran Plot 44*

CHAPTER 3. BIOPHYSICAL TECHNIQUES: MICROSCALE THERMOPHORESIS (MST) AND SURFACE PLASMON RESONANCE (SPR) 46
- 3.1 OVERVIEW 46
- 3.2 PRINCIPLES OF THERMOPHORESIS 46
  - *3.2.1 Microscale Thermophoresis (MST) 47*
  - *3.2.2 Experiment Aspects of MST 49*
  - *3.2.3 Mathematics of MST 50*
- 3.3 SURFACE PLASMON RESONANCE (SPR) 52
  - *3.3.1 Principle of Surface Plasmon Resonance 52*
  - *3.3.2 Strategies of ligand chemical immobilization 55*
- 3.4 OTHER TECHNIQUES USED FOR AFFINITY STUDIES 56

CHAPTER 4. CRYSTAL STRUCTURE OF MATRIX PROTEIN (M1) MUTANT (B-M1) 59
- 4.1 INTRODUCTION 59
- 4.2 MATERIAL 60
- 4.3 CRYSTALLIZATION AND DATA COLLECTION 61

4.4 STRUCTURE DETERMINATION 62
4.5 RESULTS AND DISCUSSION 65
4.5.1 Crystallization and Structure of the Truncated N-terminal Domain of M1 65
4.5.2 Structure Comparison with Other M1 Structures....................73
4.5.3 Physiological relevance of M1 Conformation and Oligomerization 79
4.5.4 Crystal Packing and M1 Oligomerization 83
4.6 CONCLUSION....................87
4.7 FUTURE PLANS....................87
CHAPTER 5. INTERACTION OF PROTEINS WITH SMALL MOLECULES USING BIOPHYSICAL TECHNIQUES. 88
5.1 STUDY OF M1 BINDING TOWARDS INHIBITOR AND ITS DERIVATIVES. 88
5.1.1 Introduction....................88
5.1.2 Elucidating of M1 binding affinity using SPR experiments 89
5.1.3 Elucidating of M1 binding affinity using MST....................94
5.2. AFFINITY STUDY OF GROWTH FACTORS TOWARDS INHIBITOR G2.2. 100
5.2.1 Material....................105
5.2.2 Result 106
5.2.3 Discussion 111
5.2.4 Conclusion....................111
REFERENCE: 113

List of Tables

Table 2.1. The seven crystal systems with crystal symmetry elements and parameters of cell dimension.
Table 4.1. Data collection of x-ray reflections.
Table 4.2. Statistical data of B-M1 protein structure refinement.
Table 4.3. M1-M1 interface interaction with buried interface residues of B-M1, and 1AA7.
Table 4.4. M1-M1 interface interaction with buried interface residues of 1EA3.
Table 5.1. Off rate (dissociation rate constant) and on rate (association rate constant) of M1-M1 interaction, M1-PHE interaction, and M1-M1-PHE.
Table 5.2. Equilibrium dissociation constant (K_d) of M1-M1 interaction, M1-PHE interaction, and M1-M1-PHE interaction.
Table 5.3. Summary of K_d values of PHE and all four derivatives with their response amplitude.

List of Figures

Figure 1.1. Viruses have different shape and structural components.
Figure 1.2. Statistic of most occurring influenza types by Global Influenza. Surveillance and Response System (GISRS) at World Wide Health (WHO).
Figure 1.3. Schematic representation of influenza virus structure.
Figure 1.4. Influenza A virus life cycle.
Figure 1.5. Reassortment process of the genetic material.
Figure 1.6. Structure of N-terminal of matrix M1 protein.
Figure 1.7. In vitro characterization of different M1 mutants.
Figure 1.8. Dimer structure of wt-M1 in acidic pH (PDB: 1AA7).
Figure 1.9. Dimer structure of wt-M1 in neutral pH (PDB: 1EA3).
Figure 2.1. From top to bottom a diagram depicting macromolecular structure solution.
Figure 2.2 The European Synchrotron Research Facility (ESRF), Grenoble, France.
Figure 2.3. Process of crystal formation.
Figure 2.4. Crystallization techniques.
Figure 2.5. Unit cell with vectors and angles between them.
Figure 2.6. Bravais lattices composed of the seven crystal systems with translational symmetry.
Figure 2.7. Example of Miller indices.
Figure 2.8. Elucidating of intensities from diffraction of crystal geometry.
Figure 2.9. The paths of x-rays reflected by consecutive planes of same atoms.
Figure 2.10. Ewald Sphere with reciprocal lattice.
Figure 2.11. Ramachandran plot.
Figure 3.1. Schematic representation of MST signal and instrument.
Figure 3.2. Schematic representation of detailed MST signal.
Figure 3.3. Schematic representation of SPR experiment.
Figure 3.4. Typical SPR sensorgram.
Figure 4.1. Three monomers of B-M1 in asymmetric unit and detailed structure of one monomer.
Figure 4.2. COOT modeling of Gly 136 and final Ramachandran plot.
Figure 4.3. COOT modeling of fitting electron density after Gln 158 in monomer B.
Figure 4.4. COOT modeling of fitting electron density after Gln 158 in monomer C.
Figure 4.5. COOT modeling of fitting electron density of Glu 88 and Ser 105.
Figure 4.6. 1EA3 dimer interface.
Figure 4.7. 1AA7 dimer interface.
Figure 4.8. B-M1 dimer interface.
Figure 4.9. Overlapping 1AA7 and B-M1.
Figure 4.10. Mutation at interface of B-M1 dimer.
Figure 4.11. Tetramers of 4PUS 1AA7.
Figure 4.12. Tetramers of B-M1.
Figure 5.1. Chemical structures of PHE.
Figure 5.2. SPR binding curves of M1-PHE, and M1-M1 interactions.
Figure 5.3. MST binding curves of PHE, compound 1, compound 2, compound 3, and compound 4.

Figure 5.4. Irregular fluorescence signal of PHE, compound 2, and compound 3.
Figure 5.5. MST experimental parameters of compound 3.
Figure 5.6. Models of tumor propagation.
Figure 5.7. Chemical structure of G2.2.
Figure 5.8. Schematic representation displays the site of thiol coupling with maleimide functional group
Figure 5.9. Binding curves of G2.2 with EGF and TGFB2.
Figure 5.10. MST traces of EGF and TGFB upon G2.2 binding.
Figure 5.11. Overlapping of MST capillaries shape of (A) EGF and (B) TGFB2 upon G2.2 binding.

Abbreviation

1. Å: Angstrom.
2. aa: amino acids.
3. Ala (A): Alanine.
4. Arg (R): Arginine.
5. Asn (N): Aspargine.
6. Asp (D): Aspartate.
7. BMP4: Bone morphogenetic protein 4
8. BSA: Bovine serum albumin.
9. CBER: Center for biologics evaluation and research.
10. CRAC: Cholesterol recognition/interaction amino acid consensus.
11. CRC: Colorectal cancer.
12. CSC: Cancer stem cell.
13. d: inter-plane spacing.
14. 3D: Three-dimensional.
15. 2D: Two-dimentional.
16. DLS: Dynamic light scattering.
17. ds: Double-stranded.
18. EGF: Epidermal growth factor.
19. F cold: Initial fluorescence.
20. FDA: U.S. food and drug administration.
21. FGF: Fibroblast growth factor.
22. F hot: Steady state fluorescence.
23. ΔF_{norm}: Normalized change in fluorescence.
24. GAG: Glycosaminoglycan.
25. GISRS: Global Influenza Surveillance and Response System.
26. Gly (G): Glycine.
27. Gln (Q): Glutamine.
28. Glu (E): Glutamic acid.
29. H: Helix.
30. HA: Hemagglutinin.
31. His (H): Histidine.
32. HIV: human immunodeficiency virus.
33. HS: Heparan sulfate.
34. IAV: Influenza A virus.
35. ITC: Isothermal titration calorimetry.
36. k_a: Association rate constant.
37. K_d: Affinity constant or equilibrium dissociation constant.
38. k_d: Dissociation rate constant.
39. kDa: Kilo Dalton.
40. K_2HPO_4: Potassium phosphate dibasic.
41. KH_2PO_4: Potassium phosphate monobasic.
42. L: Loop.
43. Leu (L): Leucine.
44. Lys (K): Lysine.

45. M: Membrane gene.
46. M1: Matrix protein 1
47. M2: Matrix protein 2
48. MAD: Multiwavelength anomalous diffraction.
49. Met (M): Methionine.
50. MST: Microscale thermophoresis.
51. mRNA: messenger RNA.
52. NA: Neuraminidase.
53. NaCl: Sodium chloride.
54. NCS: Non-crystallographic symmetry.
55. NEP: Nuclear export protein.
56. NLS: Nuclear Localization Signal.
57. NP: Nucleoprotein.
58. NPC: Nuclear pore complex.
59. NS2: Nonstructural protein 2.
60. NSGM: Non-saccharide glycosaminoglycan mimetic.
61. PDB: Protein data bank.
62. PEG: Polyethylene glycol.
63. PHE: 2-(4-(3-(4-acetyl-3-hydroxy-2-propylphenoxy)propoxy)phenoxy)acetic acid
64. Phe (F): Phenylalanine.
65. Rmax: Maximum resonance unit.
66. RU: Resonance unit.
67. SC: Stem cell.
68. Ser (S): Serine.
69. SPR: Surface plasmon resonance.
70. ss: Single-stranded.
71. (-)ssRNA: Negative sense single strand RNA.
72. (+)ssRNA: Positive sense single strand RNA.
73. TCEP: tris (2-carboxyethyl) phosphine.
74. TEM: Transmission electron microscope.
75. TGF β: Transforming growth factor-β.
76. Thr (T): Threonine.
77. T-jump: Temperature-jump.
78. Tris: Tris(hydroxymethyl)aminomethane.
79. Tyr (Y): Tyrosine.
80. Val (V): Valine.
81. WHO: World health organization.
82. Wnt: Wingless-int
83. wt: Wild type.
84. λ: wavelength of the x-ray.
85. θ: angle between incident or diffracted ray with corresponding crystal plane.

Abstract

CRYSTALLOGRAPHIC STUDY OF MUTANT INFLUENZA MATRIX (M1) PROTEIN AND AFFINITY STUDY OF SMALL MOLECULE INHIBITORS TOWARD M1 AND GROWTH FACTORS.

By Bashayer Althufairi, MS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

Virginia Commonwealth University, 2011

Major Director: Umesh R. Desai
Professor, Department of Medicinal Chemistry

Influenza A virus (IAV) is a seasonal infectious agent that could cause a major worldwide catastrophe. Due to its genetic properties, IAV generates new viral particles that resist the body's immune defense and antiviral drug therapy. This occurs when a host cell has been co-infected by different IAV strains leading to the generation of hybrid viruses. This process is called reassortment. The majority of these new IAVs contain genetically altered hemagglutinin (HA) and neuraminidase (NA). Unfortunately, all current IAV drug therapies target the highly mutated proteins, HA and NA, which is not very useful.

Influenza matrix protein 1 (M1) is a structural protein that accounts for a number of critical viral events. It displays a highly conserved sequence compared to other proteins, HA and NA. It is the most abundant structural viral protein. M1 has a key role in viral replication and viral assembly. During all viral steps of cellular invasion, not a single step appears to occur without the contribution of M1 in one way or another. M1 protein forms a layer underneath the lipid bilayer membrane, which contributes to vital integrity and provides an intact viral entity. Upon cellular viral entry, the M1 layer dissociates to release an RNA genome that migrates to the nucleus to utilize the host's cellular machinery for synthesizing viral proteins. More interestingly, M1 protein exhibits different structural conformations that correlate with its physiological activity. These conformational changes come with a variety of M1-M1 interactions. Crystallographic structures have revealed a tremendous amount of information regarding the M1 mechanism in self-oligomerization and depolymerization.

Various crystal structures of M1 are available. Our collaborator at the FDA identified an M1 mutant with G88E substitution, which is unable to form an intact M1 layer as wt-M1. In order to understand the role of a single mutated residue, M1 protein (G88E-M1) has been crystallized and its crystal structure was resolved by the groups of Desai and Safo. This crystal forms three monomers in an asymmetric unit. G88E-M1 concentration was 15 mg/mL in a buffer of 55 mM $KH_2PO_4/K_2HPO_4/H_3PO_4$, 0.2 M NaCl, pH 3.4. The condition of the reservoir was 0.1 M Tris, pH 8.5, 8% PEG (8K). The estimated pH of the crystallization drop was 6.2. In combination with the literature, significant structural manifestations were observed in different pH conditions. Under acidic conditions, this M1 mutant forms a face-to-face dimer, which is stabilized by hydrophobic interactions as well as hydrogen bond interactions. Although the monomers have less hydrophobic interactions at the monomer-monomer interface due to mutation of Gly88 into a polar amino acid, Glu88, it forms a stable dimer. That is because Glu88 generates at the interface a number of hydrogen bond interactions with Tyr100, Lys104 and Arg134.

M1 is an attractive a therapeutic target protein. Recently, the Desai's group has identified through computer-based drug design a promising anti-IAV drug candidate, called PHE that interferes with M1 layer formation leading to defects in cellular production of new viral particles. However, PHE binding affinity to M1 was unknown. Experiment of PHE-M1 binding affinity was performed using surface plasmon resonance with NeutrAvidin gold chip on which biotinylated M1 was immobilized under neutral pH. An affinity constant (K_d) of ~ 1 μM was determined. Likewise, PHE-M1 affinity was studied using microscale thermophoresis (MST), which yielded an affinity constant (K_d) of ~ 1.5 μM.

Another project undertaken in this study is to evaluate the affinity of small-molecule inhibitors that bind to signaling proteins. Small-molecules that could interfere with signaling pathways are highly valued in cancer therapy. G2.2, which is a highly sulfated molecule, has previously shown anticancer activity. It seems to be safe, potent, and selective toward colorectal cancer. The mechanism of action of G2.2 mainly triggers multiple important signaling pathways of cancer stem cells including fibroblast growth factor, epidermal growth factor (EGF), bone morphogenetic protein 4, wingless-int, and transforming growth factor-β (TGFβ). EGF and TGFβ were labeled with reactive dye NT-647 on the free thiol group of cysteine residues. MST experiments were performed using phosphate buffer (pH 7.4) and serial dilution of G2.2 (1 mM as the highest concentration). MST binding studies have revealed K_d of 80 μM and 54 μM for EGF and TGFβ, respectively.

In conclusion, this project has elucidated the crystal structure of G88E-M1 protein with valuable structural manifestations. PHE has high affinity for M1, which was confirmed using two different biophysical techniques. Moreover, G2.2 seems to be a promising drug therapy that targets cancer stem cells through inhibition of growth factors and cytokines associated with their survival and activity. However, G2.2 has low affinity for EGF and TGFβ.

Literature Survey

Chapter 1. Influenza A virus (IAV)

1.1 Overview

Viral infections could be more disastrous and problematic than any war because they can deceive medical preparation from health organizations around the world and resist available antiviral agents. Thus, critical understanding of virus epidemiology, function of its structural components, pathogenesis, and its types is of great importance to prevent, control and treat any outbreak of the disease. Historically, influenza pandemic in 1918 is considered the worst devastating event, taking 40 million lives around the world.[1, 2] Following that, another two influenza A virus (IAV) pandemics took place in 1957 and 1968 arising from H2N2 subtype and H3N2 subtype, respectively. Four decades later, another pandemic of influenza virus of H1N1 subtype took place.[3]

Viruses simply comprise of genetic material coated with capsid that collectively are called nucleocapsid. Some viruses could have additional lipid bilayer membrane. They could cause curable and incurable infectious, such as Herpes and HIV respectively. Viral transmission among people commonly takes place through many ways: in food and water, e.g. hepatitis A virus, via inhaled droplet, e.g. influenza viruses, by direct transfer from infected human, e.g. HIV, and from bites of vector arthropods, e.g. yellow fever virus.[4]

Despite the fact that viruses share key component, nucleocapsid, they are significantly different in terms of shape, structural composition, and complexity. Viruses could be classified into enveloped viruses and non-enveloped viruses according to the absence and presence of lipid bilayer, which is derived from host cell membrane. In Figure 1.1, non-enveloped virus, e.g. parvovirus, has a simplest virus structural entity of nucleocapsid.[5] Additionally, non-enveloped virus can be more structural complex by having extra proteins embedded in capsid shell, e.g. bacteriophage Φ29. Lastly, enveloped viruses are more advanced with lipid bilayer, called envelope, which contains multilayer of lipid and proteins, e.g. influenza virus.

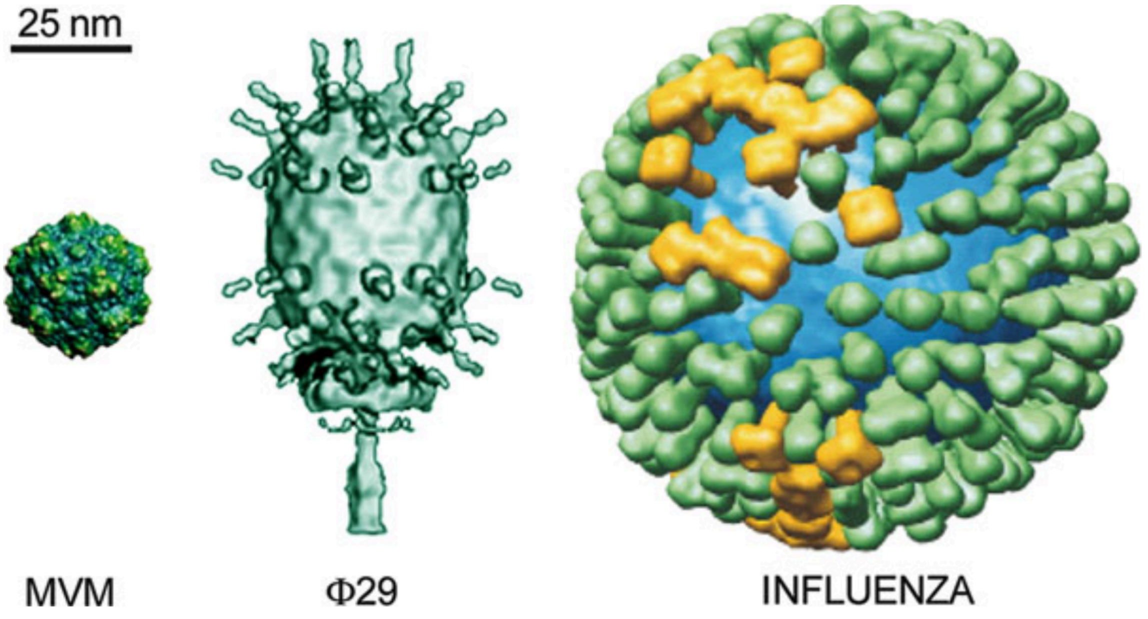


Figure 1.1. On left, parvovirus virus (MVM), having nucleocapsid. In the middle, bacteriophage Φ29, having nucleocapsid with more protein in the shell. In the right, influenza virus with extra lipid bilayer and structural proteins.[5] Taken from Goering et al. *Mim's medical microbiology*; 2013.

Remarkably, viruses are infectious organisms considered unique from other microorganisms in some characteristics, namely their reproduction. They are inert although they own DNA or RNA carrying genetic information. Instead, they exploit host's cellular synthetic machinery to process their genetic information and eventually replicate. The process of viral replication is quite complex and varied depending on the nature of the viral genetics material, either RNA or DNA but never both. More interesting, genome could be single stranded (ss), or double-stranded (ds), linear or circular RNA or DNA.

The principal step in viral gene expression is the production of messenger RNA (mRNA). Viruses containing RNA genome have diverse pathways in their gene replication. Positive sense single strand RNA ((+) ssRNA), for example coronaviruses, is first transcribed by viral RNA ploymerase to corresponding negative sense single strand RNA ((-)ssRNA), which undergoes transcription to produce complementary (+)ssRNA acting as mRNA or viral genome, while, (-)ssRNA, such as influenza virus is transcribed by viral RNA polymerase into corresponding (+)ssRNA, which can act as mRNA or is replicated to yield complementary (-)ssRNA as viral genome for new viral particles. On the other hand, retroviruses, e.g. human immunodeficiency virus (HIV), follow totally different route of replication, in which (+)ssRNA is transcribed into corresponding (-)ssDNA by viral reverse transcriptase. Then, (-)dsDNA is integrated into host's DNA and transcribed by host polymerase into mRNA. In dsRNA viruses, genome first is transcribed by viral polymerase into mRNA and one (+)ssRNA strand can act as a template for the synthesis of (-)ssRNA to form new dsRNA. Lastly, viruses containing dsDNA, their replication occurs by utilizing host's RNA polymerase to produce mRNA. Although it seems to be simple and conventional process, replication process of viral mRNA is highly organized, in which genome and its related proteins are first synthesized, and then, structural proteins are synthesized for final viral particle formation and release.[5, 6]

1.2 Influenza Virus

Influenza virus belongs to *orthmyxoviridae* genus, which is characterized by outer envelope or lipid bilayer containing eight segments (-)ssRNA genome. There are three types of seasonal influenza viruses, A, B, and C according to antigenic variation that is caused by Nucleoproteins (NP). The most common causative agents of respiratory infection or seasonal flu are influenza A and B viruses, while influenza type C causes local outbreaks with mild illness. They are transmitted mainly through droplet released by sneezing or coughing. The primary site of infection is tracheobronchial tree, although the nasopharynx is also involved. Typically, influenza virus results in destruction and desquamation of the superficial mucosa, giving rise to signs and symptoms including sudden onset of high fever, cough, headache, muscles and joint pain, vomiting, sore throat, and runny noise. Although signs and symptoms seem to be controllable and manageable and hardly no one has experienced flu symptoms in his/her life, influenza infection is inherently contagious and could sometimes be life-threatening infection, especially patients with heart or lung disease. Seasonal flu occurs often in winter, yet it can lead to influenza outbreaks around the world resulting in a major health concern. Influenza type A is also divided into subtypes according to surface proteins existing on the envelope. This classification also contributes to the nomenclature of influenza strains. The name of influenza virus is written as X/Y/Z/K/L/M, of which X is the antigenic type (e.g. A, B, C), Y is host of origin (e.g. swine, chicken, if human origin virus, there is no need to write it), Z is the geographical origin (e.g. Denver), K is strain number, L is the year of isolation and lastly M describes IAV strain, particularly hemagglutinin (HA) and neuraminidase (NA) (e.g. H1N1).[1-4]

Global Influenza Surveillance and Response System (GISRS) at World Health Organization (WHO) has collected more than 61285 specimens from national influenza centers or laboratories in 84 countries across the world during time period from 16 May 2016 to 29 May 2016. Among these samples, 4320 specimens were positive for influenza viruses, of which 1276 (29.5%) was IAV positive and the remained major cases, 3044 (70.5%), were influenza B positive.[4]

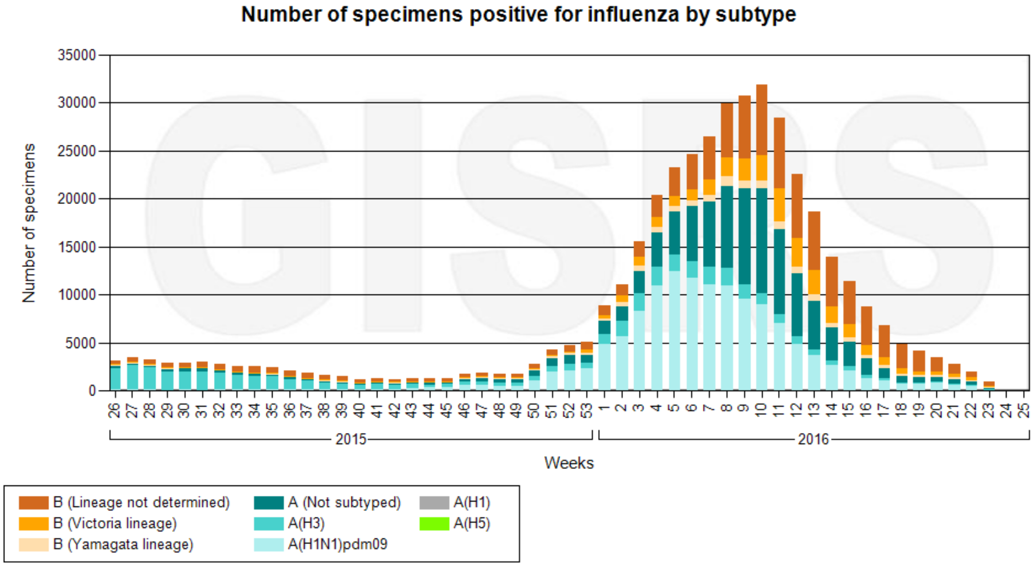


Figure 1.2. 29.5% (1276) of cases were for influenza A (71% for H1N1 and 29% for H3N2), 70.5% (3044) of cases were for influenza B (30.4% were for B-Yamagata lineage and 69.6% for B-Victoria lineage).[3] Taken from World Health Organization.

The structure of influenza virus A is quite simple, but could be complex at molecular level due to mutations. As shown in Figure 1.3, influenza virus A basically consists of structural proteins, including HA, NA, matrix protein 2 (M2) and matrix protein 1 (M1). They all exist on the envelope except for the latter lies just underneath the envelope. Genome is consists of ribonucleoprotein (RNP) complex, which comprise of viral RNA segments and NP. Moreover, polymerases are proteins associated with RNA segments, such as polymerase basic 1 (PB1), polymerase basic 2 (PB2), and polymerase acid (PA). Another protein is nuclear export protein (NEP), also called as nonstructural protein 2 (NS2). The eight segments RNA can encode up to 11 viral proteins.[1, 6]

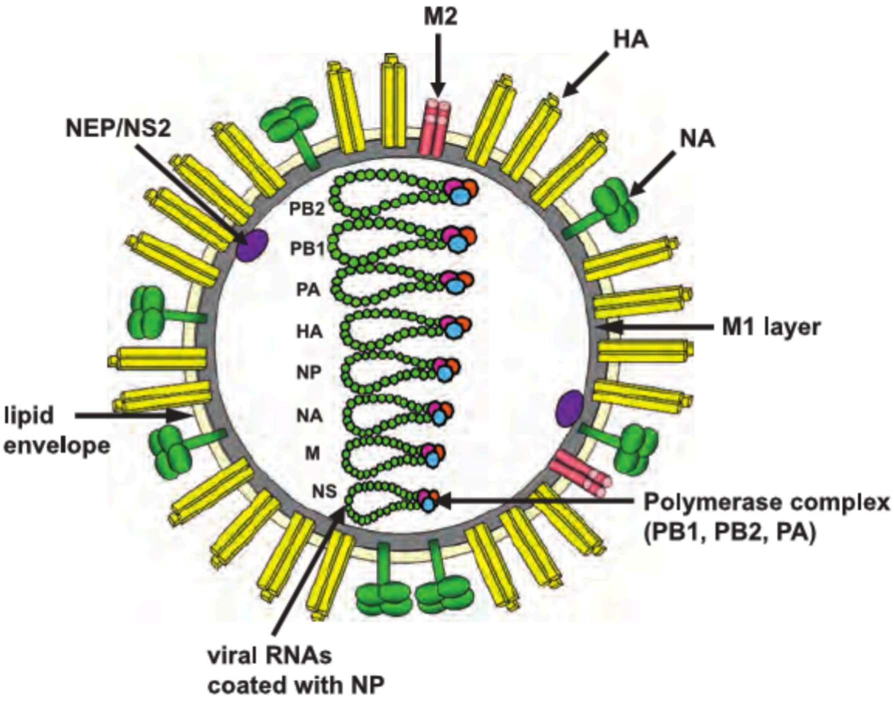


Figure 1.3. Structure of influenza virus.[1] Taken from Reid et al. *Nature Rev. Microbiol.* **2004**, *2*, 909-914.

IAV attacks any vital cell you would think of on this earth. As Figure 1.4 depicts, the life cycle of IAV starts when HA protein binds to host cell receptor molecule on plasma membrane, sialic acid. Following this binding, entire viral particle enters the cell by endocytosis (step 1). Inside endosome, acidity of the medium causes HA molecule to change conformations, which lead to fusion between virion and endosomal membrane. Afterward, M2 ion channel acidifies the interior of viral particle, thereby, causing vRNPs to dissociate from M1 layer matrix leading to the release of the former into the cytoplasm. As a result, vRNPs are ready for transportation into genome expression machinery, nucleus, by nuclear pore complex (NPC) (step 2). Inside the nucleus, the most important molecule to produce is mRNA, which is then transported to the cytoplasm, particularly host ribosome, as being translated into viral proteins. After being synthesized in the cytoplasm, those proteins involving PB1, PB2, PA, nucleoprotein (NP), M1, M2, NA, HA, and (NEP), are transported into the nucleus (step 3). What remains for full viral physical entity is the genome replication as in step 4 with the association of NP, PB1, PB2, and PA to form final eight segments of vRNPs. Additionally, M1 and NEP aid the transportation of progeny vRNPs from nucleus into cytoplasm as in step 5. In step 6, RAB11 vesicles, which are microtubule cargo in eukaryotic cells, bind and act as a carrier of vRNPs through the cytoplasm to plasma membrane.[21] Viral proteins and vRNPs are converged near plasma membrane for budding and new virions release to the surrounding environment and neighboring cells for more replication.[7, 8]

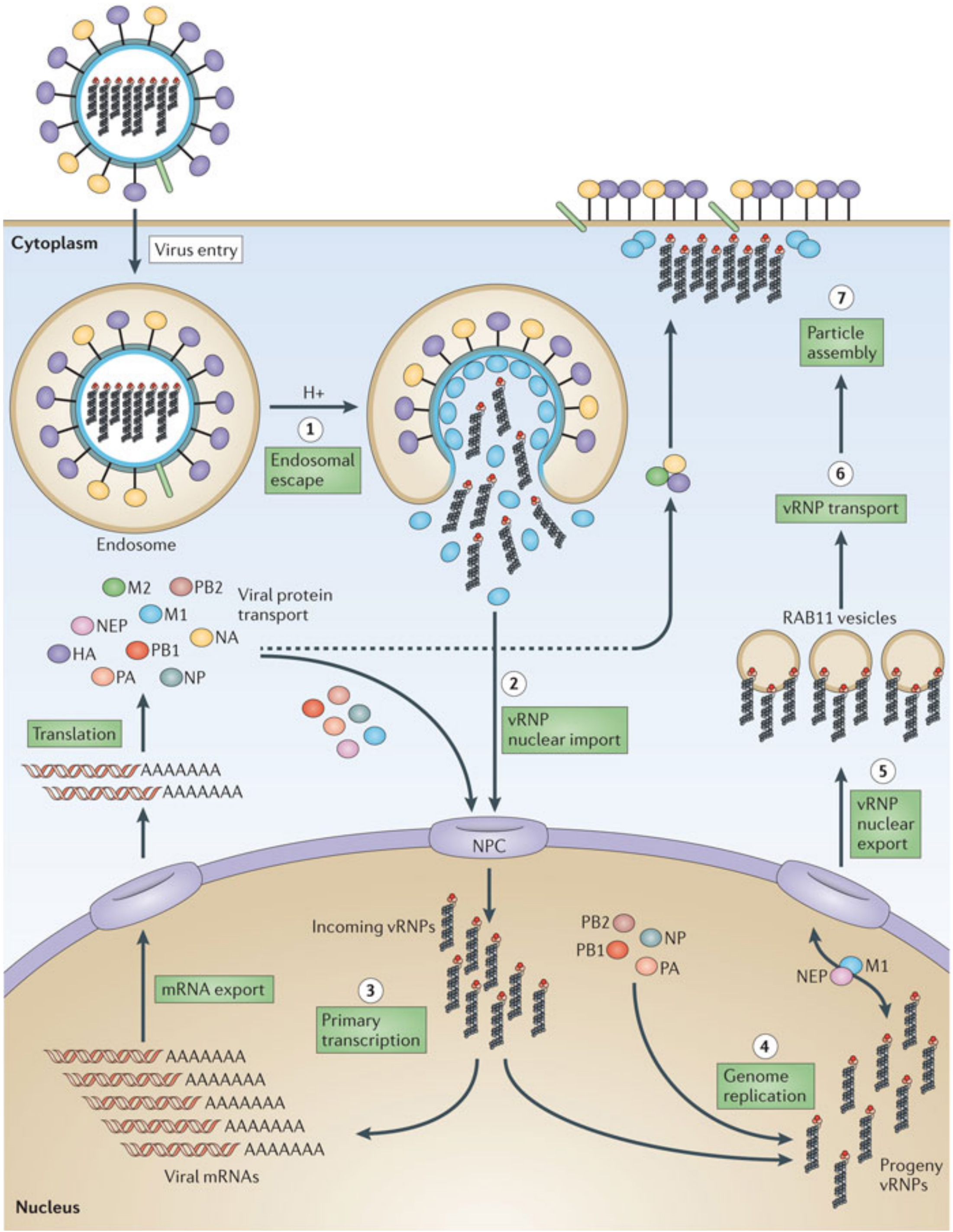


Figure 1.4. IAV life cycle: cell entry, release into cytoplasm and viral genome entry into nucleus, mRNA transcription and translation into viral proteins, genome replication, and vRNPs cytoplasmic transportation, and finally viral particle formation.[7] Taken from Eisfeld, et al. *Nature Rev. Microbiol.* **2015**, *13*, 28-41.

1.3 Structural Viral Proteins Other than Matrix M1 Protein

1.3.1 Hemagglutinin (HA)

Hemagglutinin (HA) is an integral membrane protein, which is a glycoprotein having trimeric rod-shaped molecule. The viral particle remains inactive until HA is cleaved into HA1 and HA2 subunits by trypsin-like proteolytic enzyme found at the respiratory tract. Each subunit has a distinctive function.[9] The function of HA1 is the attachment to specific sialyl sugar chain on host plasma membrane, whereas, HA2 is responsible for pH-dependent fusion between viral envelope and host endosomal membrane.[10] Once acidity is built up inside the endosome, conformational changes of both subunits take place leading to exposure of a fusogenic domain at the *N*-terminus of HA2 allowing fusion of the host endosomal membrane with the viral envelope to eventually release vRNPs into the cytoplasm. Moreover, HA is a primary viral antigen introducing new antigenic properties to the virus, thus, allowing viral reinfection of human population.[9, 10] There are 18 different HA subtypes up to date (H1 through H18).[11]

1.3.2 Neuraminidase (NA)

Neuraminidase (NA) is another major integral membrane protein. It is tetramer of identical subunits. NA can facilitate the final step of viral replication, the liberating of newly made virions from host cell membrane. It does that through removing sialic acid from cellular glycoproteins and glycolipids on host plasma membrane.[12] This is extremely important, otherwise, HA would bind to sialic acid on cellular membrane during budding, resulting in aggregation of new viral particles close to the interior side of host membrane preventing their release. Tamiflu is an NA inhibitor approved as a prophylactic drug, but viral resistant unfortunately has been developed.[13] Like HA, NA is a glycoprotein and could contribute to antigenic activity.[8, 14] There are 11 different NA subtypes up to date (N1 through N11).[11]

1.3.3 M2 Protein

M2 protein is an integral transmembrane protein embedded in the lipid bilayer membrane. It has been found that M2 possesses ion channel activity. Its major role is to conduct proton ions from acidified endosome into the interior of the viral particle, promoting pH-based conformations of some proteins, such as M1, and efficient release of vRNP from the rest components of the virus. M2 protein is tetrameric molecule, in which four helices are shaped in lipid bilayer membrane to form a pore like ion channel proteins. Furthermore, M2 protein protects some highly acid-sensitive HA, e.g. H5 and H7, from degrading by maintaining stable pH.[15-17]

IAV has phenomenal genome characteristics, by which it can generate unique surface proteins acting as antigens against existing antibodies in a process called antigenic shift.[18] This biological process can lead to an explosive global pandemic of approximately one-third of the population. It can infect many animals, such as pigs and poultry. Upon viral transmission, IAV also has the propensity to arrange and reassort its genome across different species, such as pigs and poultry, leading to novel combination of viral genome producing new influenza strain. Another way that IAV can rearrange its genome is through reassortment as shown in Figure 1.5, which takes place when co-infection of a cell with two different IAV strains leads to generation of hybrid viruses that encompass some segments from parent viruses. For example, a major H1N1 pandemic occurred in 2009 as a result of the reassortment of North American avian-like PB2 and PA segments, Eurasian avian-like swine NA and M segments, a human H3N2–like PB1 segment, and classic swine H1N1-like HA, NP, and NS segments, producing a new rapid transmitted virus that has never been seen by human population. Furthermore, IAV has the ability to rapidly mutate two or three amino acids of surface proteins, such as HA and NA. This type of mutation is known as antigenic drift, causing body immune system unable to recognize surface antigens by the corresponding circulating antibodies from previous influenza infection.[7-11]

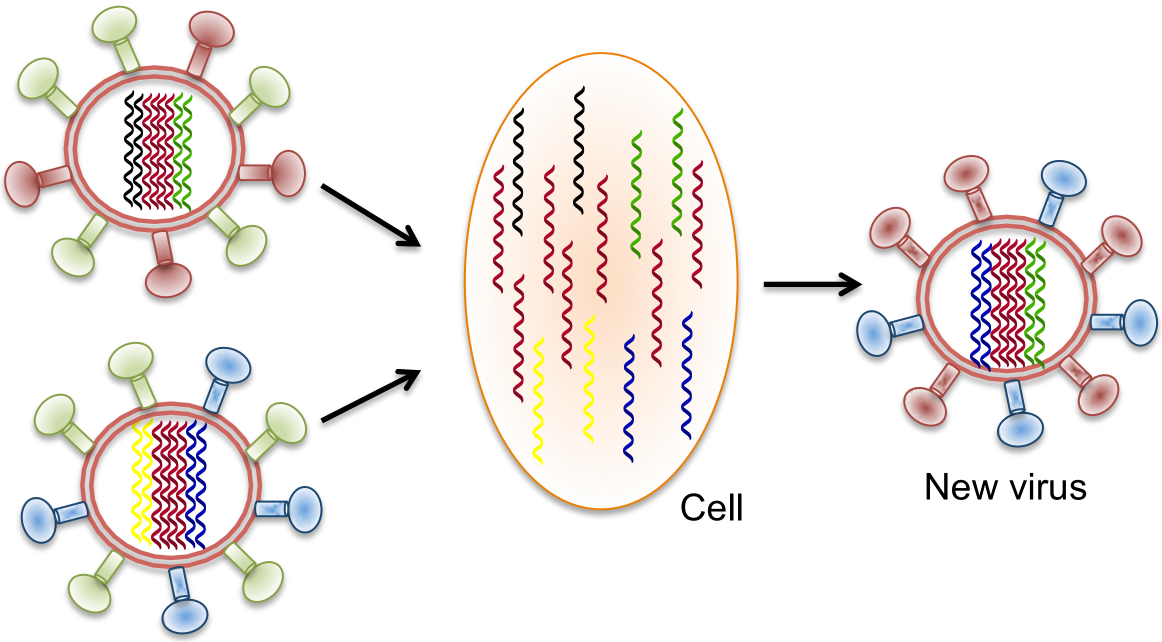


Figure 1.5. Reassortment of the genetic material from two different influenza particles within an infected cell to produce a new virus.[8] Taken from Greber et al. *Biophys. J.* **2014**, *106*, 2317–2321.

Owing to mutations and reassortment with different subtypes of viruses, it’s hard to maintain complete intervention of IAV, including drug therapy or preventing vaccine. The current medications target the most commonly mutated surface proteins, HA and NA, thereby drug resistant can be easily developed and hinder recovery from the infection. Amantadine and Rimantadine are M2 ion channel blockers and they are specific to IAV not B. Most of antiviral drugs are used as prophylaxis, which means that they ideally should be administered early in the course of the disease. Vaccines are solely effective when they well matched circulating viruses. Viruses keep instantly changing its structural components by the advantage of mutations; thus, necessitate the replacement of the vaccine every one to three years.[19]

1.4 Matrix M1 Protein

Type A influenza matrix protein M1 is encoded by RNA seventh segment. It’s the most abundant viral protein accounting for 40% of all viral proteins content. Virus generally consists of two components: transport and replication components. Electron microscopy of influenza virus has shown that M1 protein is a 6 nm long, thin rod of which one end touches the lipid membrane.[1-3]

It has been found that HA, NA, and M2 found on the lipid membrane are extremely important for viral transportation, namely viral entry and release; on the other hand, vRNP and its components represent viral replication part. Both components are linked by M1 protein.[20] M1 protein forms a layer underneath lipid membrane forming a bridge between viral inner core and membrane glycoproteins. Unlike HA and NA, M1 protein has the most conserved primary structure of influenza virus proteins when its membrane (M) gene was com pared to other genes among antigenic shifts in human pandemics.[21] Our body forms immunity through synthesizing particular antibodies directed against surface glycoproteins, HA and NA, of a particular viral strain. Due to the fact that HA and NA are highly mutated surface proteins, thereby, one may have multiple influenza infections within short period of time. Taken all together, these could be strong reasons to consider M1 protein as novel drug target, since M1 has been also shown to have a great contribution to different aspect of viral life cycle and essential role in virus survival.[14]

1.5 Structure and Functions

M1 protein consists of 252 amino acids having molecular weight of approximately 28 kDa, forming basically two helical domains, N-terminal domain (1-164) and C-terminal domain (165-252) linked by protease sensitive region. Only the N-terminal domain had intensively been studied and the crystal structure has revealed that N-terminal domain is divided further into two domains: N-domain and M-domain. Each domain consists of four helical bundles that are joined by short helix (H5). N-domain is mapped from H1 to H4, while M-domain encompasses H6-H9.[22]

M1 generates a multiple network of lipid and protein interactions: lipid-M1, HA-M1, NA-M1, RNPs-M1 and M1-M1.[16] Remarkably, each domain has distinguishable interactions with other viral components. N-terminal domain significantly contributes to most of M1 interactions. Each viral protein has nuclear localization sequence (NLS) motif, which enables viral proteins to interact with transporters and enter the nucleus after being synthesized in the host's ribosomes. For M1 protein, NLS, located at N-terminal domain, first was identified as Arg 101, Lys 102, Leu 103, Lys 104, Arg 105.[23] Later, this motif was found to be part of an important interactive positive patch of certain amino acids including Lys 95, Ala 96, Val 97, Lys 98, Leu 99, Tyr 100, Arg 101, Lys 102, Leu 103, Lys 104, and Arg 105.[22] Owing to positively charged region of N-terminal domain, M1 protein makes contact with negatively charged regions of NP and RNA through electrostatic interaction. These positively charge residues on M1 protein indeed play a principal role in M1-M1 interaction or M1 oligomer state.

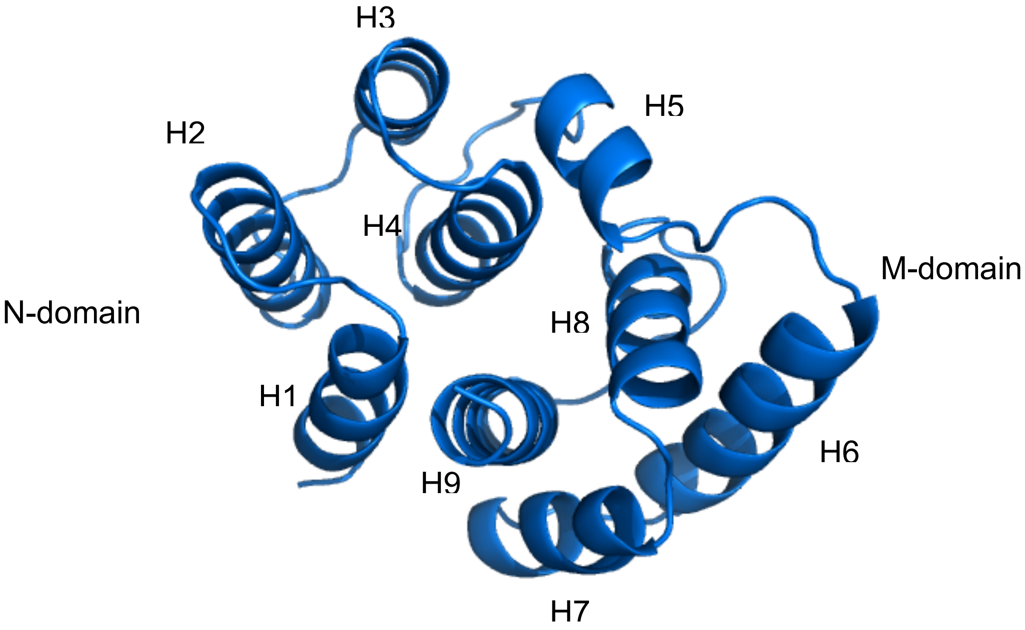


Figure 1.6. There are no β-strands in N-terminal domain of M1. Typically, N and M domains are main domains in N-terminal of M1. N-domain (2-67 aa) contains helices (H1-H4), separated by loops L1-L3. M-domain (91-158 aa) has H6-H9, separated by loops L6-L8. A short helix joins N and M domains and attaches through L4 and L5 to each domain respectively.[22] Taken from Sha, et al. *Nat. Struct. Mol. Biol.* **1997**, *4*, 239–244.

In addition to that, M1 protein also mediates RNPs exit from the nucleus and hinders their reentering into the nucleus.[24] During the step of releasing vRNPs upon viral cellular entry, dissociated M1 and vRNPs migrate independently into the nucleus, where mRNAs and new replicated vRNPs are synthesized. M1 and NEP are needed for vRNPs export from the nucleus. This occurs through the binding of N-terminal domain of M1 protein, particularly Nuclear localization sequence (NLS) region, with negative charge region in the C-terminal domain of NEP, while vRNPs attach along with this complex by M1 protein C-terminal domain.[20] Of NEP, N-terminal binds to chromosome region maintenance protein 1 (CRM1) which facilitate the export of the whole complex through nuclear pores.[15, 25]

Furthermore, M1 also has multiple roles in viral assembly, viral budding and release. M1 binds to cytoplasmic tails of transmembrane proteins, HA, NA, and M2, on the outer side of the virus and core vRNP on the inner side, which collectively brings these parts in one structurally intact influenza virus during preparation for viral budding.[26-28] Not only that but also M1 causes membrane bending in the very beginning of budding through binding to itself leading to asymmetry in the lipid bilayer. It has been found that M1 is important for bud closing and release. There is a correlation between low levels of M1 with reduction in virus release.[29] Moreover, mutations in M1 protein may lead to elongated virus particle as well as budding defects.[30] Another distinct charge-charge interaction was found to take place between Arg 77 and Arg 78 of M1 and phosphate head groups of phospholipid membrane.[31, 32] It should be pointed out that the primary interaction between M1 protein and the membrane is hydrophobic binding, which further strengthens M1-membrane contact.[31] Taken all together, it's obvious that M1 has a critical structural function in that it acts as a basement and anchor for surface glycoproteins and primarily provides integrity of virus to be functionally active.[33]

1.6 Mutation Studies on M1 Protein

There is no doubt that certain regions of amino acids are critical for a variety of M1 interactions and functionality. In order to outline various functions of M1, series of M1 mutants were produced using reverse genetic approaches. For example, one study found that deletion of RKLKR or substitution of Lys 102 or 104 with Asn (K102N or K104N) causes virus replication to fail.[34] In the same study, mutation of Arg at position 101 or 105, but not both, in RKLKR motif didn't significantly affect nuclear export of vRNP.[23, 34] In another study, double mutations of R101S and R105S, designated as SKLKS-mutant, showed a major impairment of viral replication.[35-37] Mutation of both R101S and R105S in NLS region of A/WSN/33 results in viral generation of an additional third mutation in neighboring region, G88R in SKLKS-mutant leading to more stable triple mutant, designated as NLS-88R. The latter has similar characteristics as wt-M1.[38] Figure 1.7 summarizes different experiments of comparing NLS-88R and wt-M1, for example both showed similar temperature sensitivity at 33 °C, 37 °C, 39 °C, whereas double mutant, SKLKS, showed instability at highest temperature. Thus, NLS-88R possesses comparable functionality and stability with those of wt-M1 (Figure 1.7-A). As expected, NLS-88R maintained viral replication close to the level of wt-M1 *in vitro* (Figure 1.7-B). From pathogenesis standpoint, NLS-88R was attenuated in mice with 100% survival, while wt-M1 infection resulted in a decrease in body weight prior to causing mortality of all mice group (Figure 1.7-C). Further proof of this finding, *in vivo* lung test revealed that NLS-88R virus was cleared by the seventh day of the infection, while wt-M1 virus persisted to grow and replicate in respiratory tract at the same time.[36] These results support the notion of considering NLS-88R mutant a live attenuated vaccine.[38]

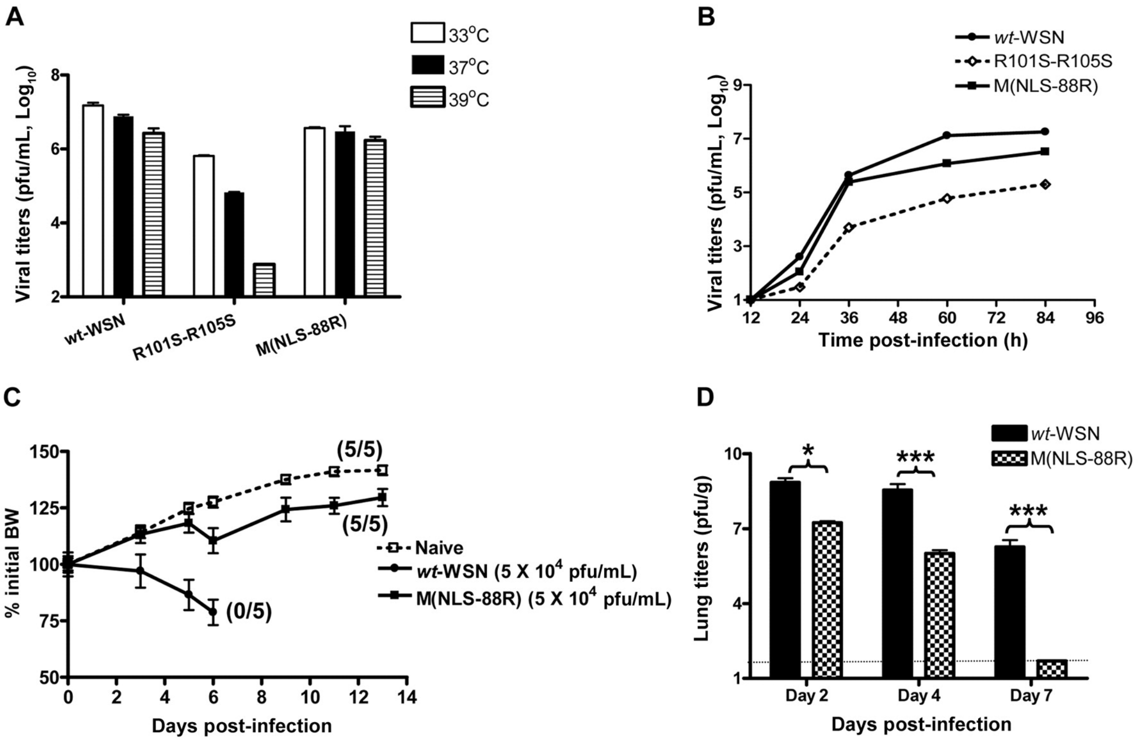


Figure 1.7. *In vitro* characterization of NLS-88R: (A) The temperature-sensitive (ts) phenotype. (B) Replication assay was determined by plaque assay at plaque-forming units (pfu)/mL. *In vivo* characterization of NLS-88R: (C) Four-week-old mice were infected intranasally with wt-WSN (wt-M1) or NLS-88R at 5×104 pfu/50 μL of phosphate-buffered saline/mouse. Body weight (BW) was monitored for two weeks post-infection. Survival is expressed in parentheses as the no. of survivors/total no. of mice weighed. (D) Entire lungs from groups of 3 mice were collected on days 2, 4, and 7 post-infection. Pulmonary viral replication was determined as (B) and expressed as the mean number of plaque-forming units per gram. *P<0.05 and ***P<0.001, by unpaired Student's t test performed on log-transformed data. The lower detection limit is denoted by the dashed horizontal line.[36] Taken from Xie et al. *J. Infect. Dis.* **2009**, *200*, 1874–1883.

Other triple mutants were produced keeping the same mutation pattern of having Ser substitutions at 101 and 105 positions instead of Arg in NLS region, but substituting different AA with different polarity at position 88 instead of Gly in order to approve that NLS-88R mutant exclusively is as active as wt-M1. Under transmission electron microscope (TEM), viruses of NLS-88V and NLS-88E mutants, substitution of Gly 88 with Val and Glu respectively, displayed irregular thickness of M1 single layer, while viruses of NLS-88R resembled wt-M1 viruses in forming uniform thickness of M1 bilayer. Despite that, all mutants except NLS-88R showing low viral replication, NLS-88E was the least efficient mutant to replicate in mouse lung. Since M1 self-association is facilitated by charge-charge interaction, computational studies suggested that NLS-88R mutant may form salt bridge during M1-M1 binding that contributes to its efficient viral activity and overcomes the loss of electrostatic interactions of NLS region resulted from double mutation of Arg at 101 and 105 positions.[38, 39]

1.7 Crystallographic Studies on M1 Protein

M1 protein has the ability to adapt number of various structural conformations depending on the pH of the surrounding environment. Beside the standard research conducted for all proteins relating to pH effect on their conformational structures, changing in pH is in fact a crucial factor that efficiently effects launching viral replication. Prior to the release of vRNP from M1 oligomer, an increase in the acidity of M1 layer surrounding initiates and facilitates M1 structural changes required for vRNP liberation. It’s noteworthy to direct research investigations of IAV pathogenesis into pH effect on conformational changes of viral proteins during viral replication. Definitely, M1 protein displays plenty of structural conformations that can be visualized by crystallographic structures.[22-27]

Many crystals of M1 protein have been deposited in protein data bank (PDB) with different crystallization conditions, particularly pH structural changes. However, none of them provides the full length structure of M1 protein rather they have revealed N-terminal domain (1-164) with molecular weight of 18 kDa, because C-terminal domain (165-252) underwent proteolysis during purification step.[22, 31] The first crystal structure of the wild type (wt) M1 protein was for N-terminal domain, which remains stable in solutions as well as crystal preparations, making it an independent domain. The crystal of N-terminal domain was obtained in acidic pH, similar to the physiological condition in which M1 protein dissociates from lipid membrane liberating RNPs for genome replication in the host's nucleus after host's endocytosis. Asymmetric unit contains a dimer in which each monomer interacts in face-to-face fashion, meaning that amino acids in one side of a monomer is facing their counterparts in the other monomer as in Figure 1.8.[22]

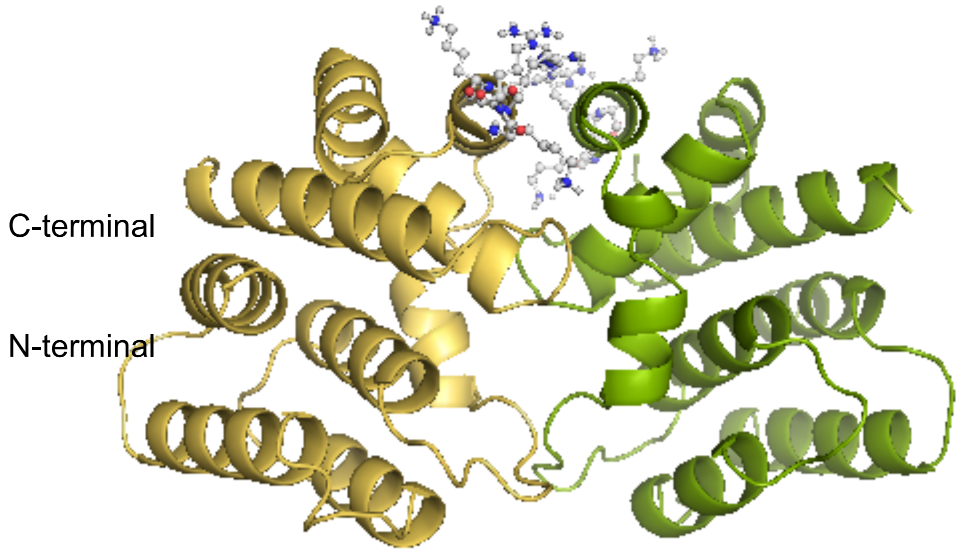


Figure 1.8. Dimer structure of wt-M1 in acidic pH (PDB: 1AA7). N-terminal and C-terminal are shown as well as positive aa of each monomers (Arg 101, Arg 105, Lys 102, and Lys 104) which are important for electrostatic interactions of M1. Two monomers are packed face-to-face pattern.[22] Taken from Sha, et al. *Nat. Struct. Mol. Biol.* **1997**, *4*, 239–244.

Another crystal was revealed in neural pH, a condition in which the M1 forms an intact layer underneath lipid bilayer surrounding genetic core of the virus. Similar to pH 4, asymmetric unit contains a two monomers, however, they are stacked into face-to-back fashion as in Figure 1.9.[39]

Safo *et al*. revealed another crystal structure of M1 in acidic condition with different dimer interface when compared to the acidic M1 crystal structure by Sha *et al*. Although the two crystals structures dimerize in a face-to-face fashion, the relative arrangements between monomers are different. The fact that M1 can adopt a monomeric structure or dimer structure but with different dimer interface indicates that M1 adopts several conformations during viral life cycle enabling other viral proteins movement and function as well as supporting overall viral geometry.[40]

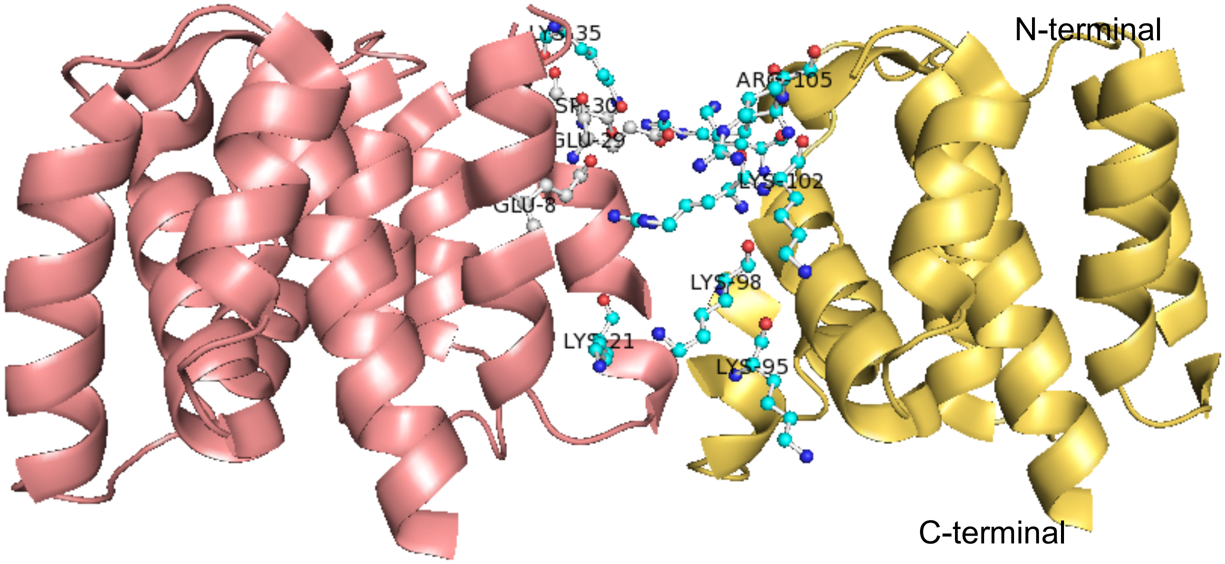


Figure 1.9. Dimer of wt-M1 in neutral pH (PDB: 1EA3). N-terminal and C-terminal are shown. Two monomers are packed face to back pattern and dimerize by electrostatic interaction. One side mainly carrying positive aa, such as Lys98, Arg105 (shown with cyan colored carbons), and the other side mainly carrying negative aa, such as Glu8 and Glu29 (shown with white colored carbon).[39] Taken from Arzt, et al. *Virology* **2001**, *279*, 439–446.

Chapter 2. Fundamentals of Protein X-ray Crystallography

2.1 Overview

Proteins are biological macromolecules of interest for most scientists in different fields. It is of great importance to study the structure of proteins since they are different in their structure geometry, which controls protein's activity and function and also varies according to different disease states for the same protein. Over a century, crystallography, a technique simply used to elucidate macromolecular structures, has been considered the most powerful technique in structural biology. Utilizing such a technique can reveal protein structure and in conditions mimicking physiological or pathological states that tailored by crystallographers. Typically, an experiment conducted to study protein structure involves different steps such as, crystallization, data collection, and solving phase problem and structural data.

2.2 Principles of X-ray Structure Determination

The process of revealing the three-dimensional (3D) structure of a protein has historically been an issue because the structure of protein carries crucial information including, but not limited to its function, folding, shape, and stability. Over a century, many of discoveries have beneficially impacted the development of crystallography making biologically important macromolecules visible. One of great accomplishments is the discovery of x-rays, by Wilhelm Röntgen who won Nobel Prize in physics 1901.[41, 42] In 1914, Max von Laue won Noble prize in physics for discovery of diffraction of x-rays from a crystal.[43] One problem that scientists faced at that time was “how to process x-rays diffraction pattern to yield chemical structure”. Max Laue had also discovered that x-rays are electromagnetic in nature, which opened the door later to William Henry Bragg and William Lawrence Bragg to develop Braggs law (see section 2.3.). After that, number of experiments on salts crystal was conducted successfully to determine the chemical structures. At that time, none of biological macromolecules structure was determined until 1957 when John Kendrew has resolved for the first time the atomic structure of myoglobin, a globular protein storing oxygen in muscles.[44, 45]

For studying molecules, X-ray structure is of great importance. It can provide explanation of chemical or physical properties, for example, hardness of the diamond and softness of graphite.[46] By resolving a crystal structure, one could see the arrangements of atoms in molecules according to the crystallization conditions. Additionally, knowledge obtained from x-ray crystal structure of protein about its structure and careful study of its folding, cavities and amino acids (AA) sequences and orientation, exclusively answers questions regarding protein function, protein behavioral changes in both disease and normal states.[47-49] Furthermore, x-ray crystallography has amenably been used in accordance with fragment-based drug design approach for the purpose of finding potential inhibitor for a target protein.[50, 51] Potential binding pocket can be visualized and utilized in a co-crystallized complex of protein and ligand to design a lead compound under an approach called structure-based drug design.[52, 53]

Five important milestones steps for the determination of x-ray crystal structure of a macromolecule include: [54]

1. Suitable expression system produced enough large quantity of a protein.
2. Efficient purification protocol to ensure homogenous protein sample.
3. Crystallization, screening and optimization of conditions to get high-quality diffracted crystal
4. Collection and processing of considerably diffraction data for building a good density map.
5. Finally, model building, refinement, and quality assessment.

Figure 2.1 summarizes the overall process of x-ray crystallography. High-throughput crystallography has been recently emerged to accelerate determination of atomic structure of a protein because in old days one could spend approximately three years only to calculate electron density.[55] Nevertheless, the growth of diffraction-quality crystal is still a major bottleneck despite the advent in studying proteins properties. There isn’t yet a possible way to predict the appropriate condition for a protein of interest to produce crystal.[56] As medicinal chemists, we are interested more in the last three events of crystallography technique, where we could apply chemical knowledge.[57, 58]

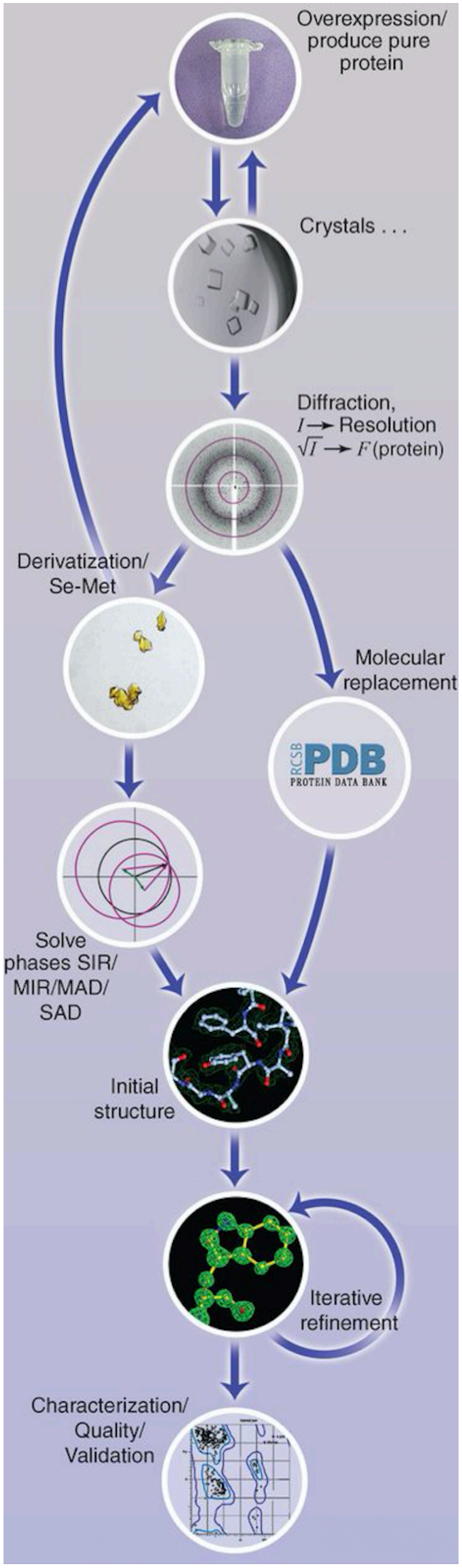


Figure 2.1. From top to bottom a diagram depicting macromolecular structure solution.[58] Taken from German et al. *Science* **2014,** *343*, 1102-1108.

Nowadays, sources of x-ray radiations are more powerful with a variety of intensities and uses comparing to decades ago. Many types of X-ray generators have been developed including in-house x-ray generator (rotating anode) and synchrotron radiation. Laboratory x-ray device is basically equipped with hard x-ray source, x-ray optics, a goniostat to position and rotate, and x-ray detector. In addition, other important components are added for specific practical purpose, such as cryo-cooler (cold nitrogen gas) to reduce radiation damage on the crystal, and a microscope focused onto center of the instrument to align and center the crystal on the x-ray beam path. For protein crystallography, copper is usually used as an anode material that emits x-ray radiation of a wavelength of 1.54 Å.[59] Other metals, such as chromium, iron, and molybdenum, with various characteristic wavelengths can also be used. For example, molybdenum (0.71 Å) is used for small molecule x-ray crystallography because the crystal has small unit cell and the diffraction spots are closer together, which makes molybdenum convenient to use with area detectors of small molecules. In parallel, advancements in detectors have also taken place. Two-dimensional (2D) detectors, such as charge-coupled device or pixel array detector, replace photographic film detector due to their good detection efficiency and spatial resolution.[60, 61]

Synchrotron (Figure 2.1) radiations are very high intensity x-ray beams. In this case, electrons are injected into a storage ring, which is a circular arrangement of bending magnets causing change in electrons direction. This leads to generation of x-ray radiations. Unlike x-ray tube, synchrotron reveals useful and high resolution data from very small crystals due its high energy beam and brilliance, which show extensive details of the crystal in short time compared to traditional in-house x-ray generator. It's also a sophisticated facility to detect very short-lived crystal structures.[59-63]

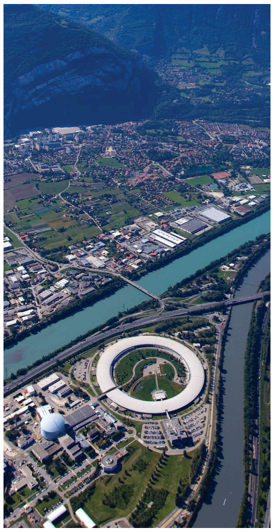


Figure 2.2. The European Synchrotron Research Facility (ESRF), Grenoble, France.[58] Taken from German et al. *Science* **2014,** *343*, 1102-1108.

2.3 Crystallization Techniques

Before setting up crystallization experiments, it's important to ensure certain considerations. First, it's essential to ensure the crystallizability of the target protein; some proteins are hardly crystallized, e.g. membrane proteins.[64] A protein of interest has to be expressed in a biological system that guarantees the correct folding, AA sequence, and functionality of the protein as if it is expressed in its original biological system. Another significant factor is protein purification. Highly purified protein sample is required due to the fact that impurities could interfere with the growth of the crystal, which is a highly ordered 3D array of molecules.[54, 65]

Obtaining a crystal of protein is a matter of trial and error, because each protein has its own structural and biological characteristics making it's hard to establish defined universal protocol for protein crystallization. Multiple parameters, however, could govern the growth and formation of the crystal.[54] Physical properties, such as temperature, chemical properties, like precipitant type, and biochemical properties, such as sample purity, can considerably affect crystal formation. More interestingly, a change of pH values by 0.5, or sometimes even by 0.1, can lead to dramatic change in protein solubility, which in turn affects nucleation formation and crystal growth. There are numerous conditions of crystallization screening commercially available. They are available with multiple combinations of typical components of crystal screening, such as precipitant, buffer, salt, or different pH units.[65]

Three main steps in the protein crystallization, which are nucleation, crystal growth, and cessation growth, take place to produce final crystal as in Figure 2.3. In nucleation, protein molecules aggregate to form thermodynamically stable particle, which is called "critical nucleus" stabilized by specific interactions. Then, protein molecules diffuse and orderly assemble on the critical nucleus into a bigger particle. This step represents the stage of crystal growth, which is a characteristic of well-ordered 3D array lattice of the protein crystal. Finally, the crystal growth ends when protein concentration efficiently drops in the solution.[66]

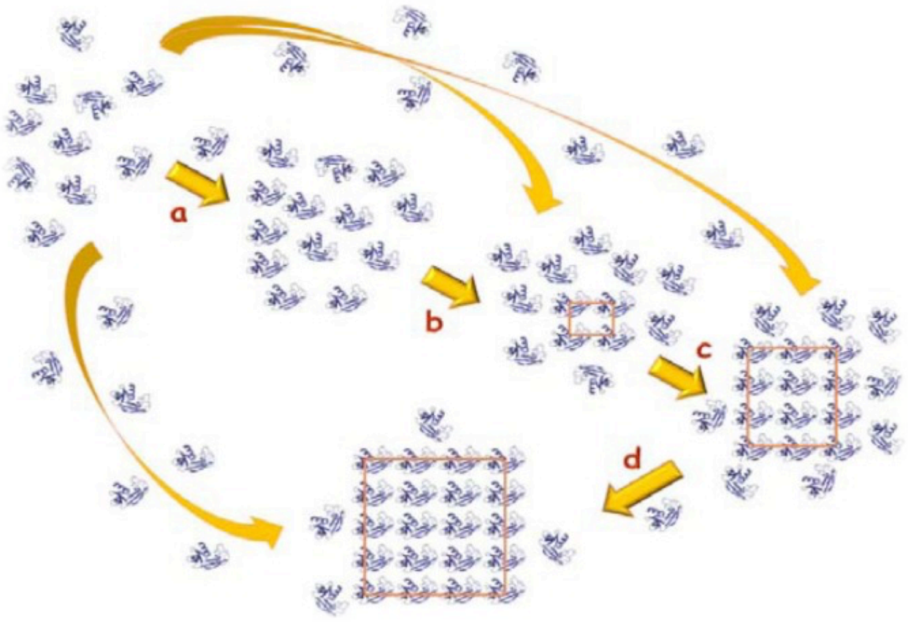


Figure 2.3. Process of crystal formation, (a) protein molecules disorderly aggregate, (b) critical nucleus formation, (c) development of the true crystal, (d) free protein molecules adsorb on the lattice to increase crystal volume.[66] Taken from Krauss et al. *Int. J. Mol. Sci.* **2013**, *14*, 11643-11691.

Experimentally, crystallization is performed by number of techniques: vapor diffusion, microbatch in oil, and dialysis. Vapor diffusion is a common technique in which a drop of protein is mixed with a drop of reservoir solution (buffer, precipitant, salts, and other additives).[67] The drops mixture is placed in different positions in the reservoir solution wells either by sitting drop, or hanging drop or sandwich drop as in Figure 2.4 (a-c). Microbatch in oil is another method (Figure 2.4 (d)), in which a protein is mixed with the reservoir components and allowed to be placed under oil, which controls the evaporation rate of the reservoir solution from protein/reservoir mixture. Lastly, dialysis is mainly based on diffusion of contents of the reservoir through semipermeable membrane as in Figure 2.4 (e, f).[65]

Since crystallization step depends on trials and errors, it’s essential to find a mean to automate this step to expedite the process of obtaining high quality crystal.[65] Number of robots has been developed, by which an experiment of crystallization with 96-well plate could be done in less than a minute.[68] Moreover, robot could help in resolve one of unavoidable problems with human-performed experiment of crystallization, in which crystallographer could draw variable volumes affecting overall mixing ratio. Usually, protein and reservoir solution of a condition is mixed as a ratio of 1:1. [69, 70] In order to get a faster growth of the crystal, it’s recommended to mix volumes in a nano-scale, however; this could be accurately achieved by robotic crystallization.[71]

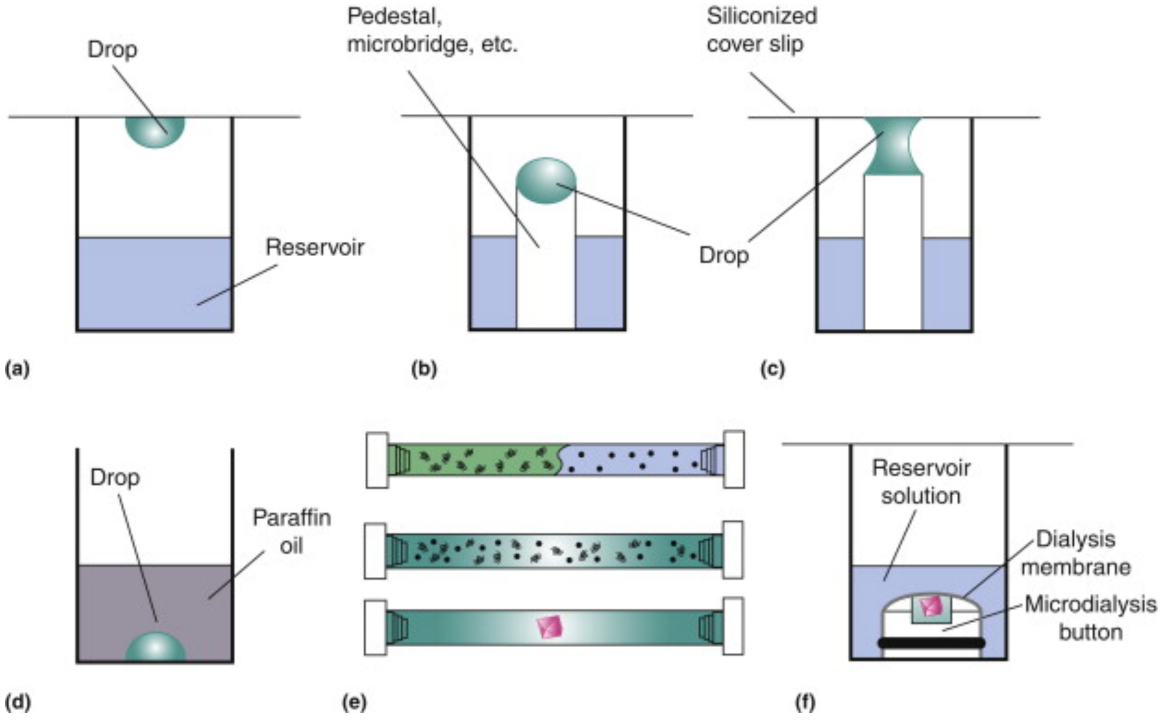


Figure 2.4. Crystallization techniques, (a-c) Vapor diffusion (Hanging drop, Setting drop, Sandwich drop, respectively). (d) Microbatch in oil, (e,f) Dialysis.[67] Taken from Benvenuti et al. *Nat. Protoc.* **2007**, *2*, 1633-1651.

2.4 Crystal Symmetry

As mentioned before, crystal is geometrically composed of organized 3D array of protein molecules. Each protein molecule is placed at precise position relative to each other. The basic building block of a crystal is called unit cell, which is characterized by three vectors (a, b, and c) and three angles between them (α, β, and γ) as in Figure 2.5.[72] Moreover, unit cell can be further simplified into two or more asymmetric units that are related by symmetry element.[73]

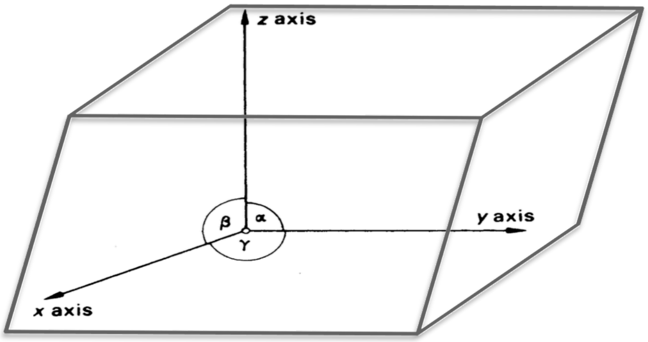


Figure 2.5. Unit cell with vectors and angles between them.[73] Taken from Hammond, C. *The basics of crystallography and diffraction*, 2009.

There are seven geometrical crystal systems (Table 2.1). Applying translational symmetry to the seven crystal systems gives rise to 14 Bravais lattices, which are described by four different ways of lattice points (Figure 2.6).[74]

1. Primitive (P): lattice points are placed solely at the corners of the unit cell.
2. Faced-centered (F): lattice points are placed at the corners and one lattice point is placed in the middle of every face of the unit cell.
3. Base-centered (C): lattice points are placed at the corners and one lattice point is placed at each of the two relevant faces in the unit cell
4. Body-centered (I): lattice points are placed at the corners and one lattice point is placed in the middle of the unit cell.

Translational symmetry is defined by a translation, while point symmetry elements consist of four symmetry operations: [75]

1. Inverse center: when applied, the resulting object is a mirror image of the original, rotated upside-down.
2. Reflection: when applied, the resulting object is a mirror image of the original; both lie in opposite sides with equal distance from the mirror plane.
3. Rotation: rotation takes place counterclockwise by an angle of 360/n (n could be 1, 2, 3, 4, 6).
4. Rotation-inversion: it consists of two-symmetry elements rotation (360/n), followed by an inversion.

Table 2.1. The seven crystal systems with crystal symmetry elements and parameters of cell dimension.[73, 76]

Crystal system	Essential symmetry of crystal	Cell parameters
Triclinic	None	None
Monoclinic	One diad axis (2-fold rotation) or mirror plane (inverse diad axis)	α = γ = 90°
Orthorhombic	Three orthogonal diad axes or inverse diad axes	α = β = γ = 90°
Tetragonal	One tetra axis (4-fold rotation) or inverse tetrad axes	a = b; α = β = γ = 90°
Trigonal	One triad (3-fold rotation) axis or inverse triad axis	a = b; α = β = 90°; γ = 120°
Hexagonal	One hexad (5-fold rotation) axis or inverse hexad axis	a = b; α = β = 90°; γ = 120°
Cubic	Four triad axes or inverse triad axes	a = b = c; α = β = γ = 90°

When point symmetry and translations are combined, they generate two symmetry operations: [75]

1. Glide plane: it’s combination of two symmetry operations, reflection in a mirror plane followed by translation.
2. Screw axis: it’s combination of two symmetry operations, rotations followed by translation.

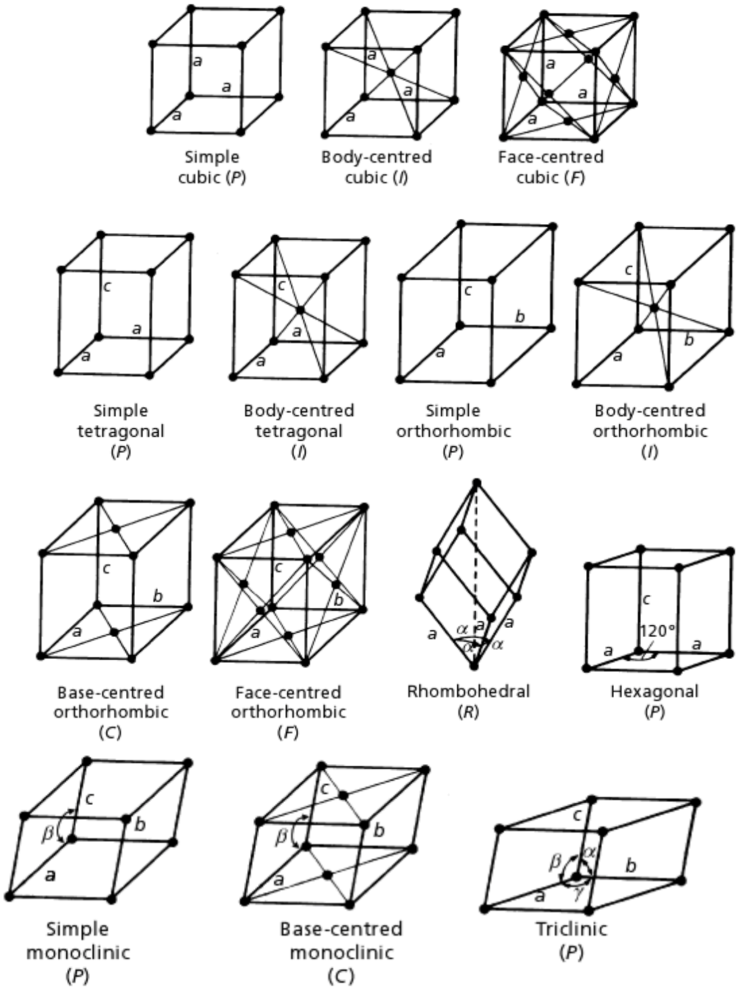
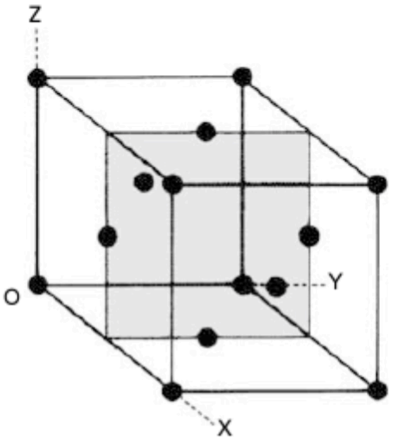


Figure 2.6. Bravais lattices composed of the seven crystal systems with translational symmetry.[73] Taken from Hammond, C. *The basics of crystallography and diffraction*, 2009.

There are 230 crystallographic space groups as a result of combining the Bravais lattices and the point, translational and space symmetry. Proteins are chiral macromolecules most composed of chiral amino acids as building blocks, therefore; mirror plane or an inversion center can't be applied into protein crystal. As a result, proteins generally crystallize into 65 chiral space groups.[73-76]

To be able to solve the crystal lattice and subsequently unit cell, imaginary planes have to be devised. These planes are designated by Miller indices, *h*, *k*, and *l*. These are integers of the inverse ratio of the intercepts on a, b, and c axes of the unit cell (Figure 2.7). [73-76]



$$\begin{array}{ccc} h & k & l \\ \frac{1}{1/2} & \frac{1}{\infty} & \frac{1}{\infty} \end{array} = (2,0,0)$$

Figure 2.7. Example of Miller indices.[73] Taken from Hammond, C. *The basics of crystallography and diffraction*, 2009.

2.5 X-ray Diffraction of a Crystal

As mentioned, crystal contains many protein molecules aligned with high degree of order. Upon x-ray scattering from a crystal, many diffraction pattern images are generated, in which each point characterizes one reflection. X-ray diffraction patterns represent 2D array of reflections, which is a cross section of 3D array lattice of reflections.[7] It's important to collect as much diffraction patterns as possible through different crystal orientations with regard to x-ray beam eventually for the purpose of measuring directions and intensities of reflections as in Figure 2.8.[59, 77]

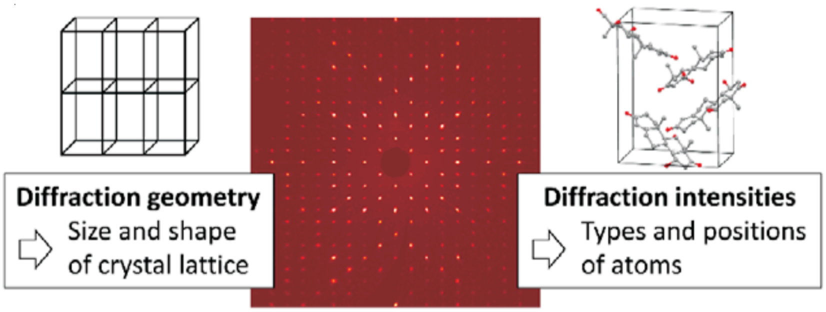


Figure 2.8. Elucidating of intensities from diffraction of crystal geometry.[77] Taken rom Andrew, B. *Resonance* **2014**, *19*, 1087-1092

Two major parameters that crystallographers seek from x-ray diffraction pattern, are the position of *h, k, l* and intensity of each reflection.[75] These are essential data in order to proceed into the electron density of the macromolecular structure. Not all x-ray reflections are recorded, they have to be instead resulted from waves that are constructive interfered or obey Braggs law (equation 2.1).[74]

$$n\lambda = 2d_{hkl}\sin\theta$$ Equation **2.1**

Where n is an integer, λ is the wavelength of the x-rays, d is the spacing distance between the planes of the unit cell, and θ is the angle at which x-rays are reflected. Figure 2.9 displays the fact that atoms with symmetry-equivalent position in each imaginary plane within the lattice are separated by *d* distance, which is vertical distance between the two planes. The path difference between x-rays reflected at the same atoms existing between two planes has to equal an integer multiple of the wavelength.[59-62, 74]

It is impossible to correctly arrange spots which one is presenting which plane in the diffraction pattern. Instead, reciprocal lattice is an abstract of unit cell lattice providing a mean to view crystal structure. Reciprocal lattice is a grid of points in which each point represents a family of planes and is labeled with Miller indices.[76] For diffraction to occur, number of reciprocal lattice points have to comes into contact with Ewald Sphere as in Figure 2.10 Experimentally, this could happen upon rotation of the crystal during x-ray irradiation, as a result, reciprocal lattice will rotate with it, and eventually reciprocal lattice points will touch Ewald Sphere.[73-76]

Ewald Sphere is a sphere with a radius $1/\lambda$, where λ is the wavelength of the x-rays. In Figure 2.10, when a crystal scatters the incident x-ray (shown in blue), a vector joining the reciprocal lattice point on Ewald Sphere to the scattered x-ray is called reciprocal lattice spacing, which is exactly equal to $1/d$. That's why it's called reciprocal lattice.[59]

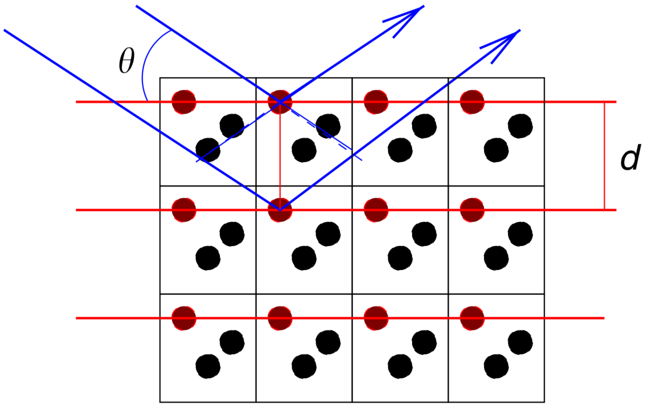


Figure 2.9. The paths of x-rays reflected by consecutive planes of same atoms.[60]

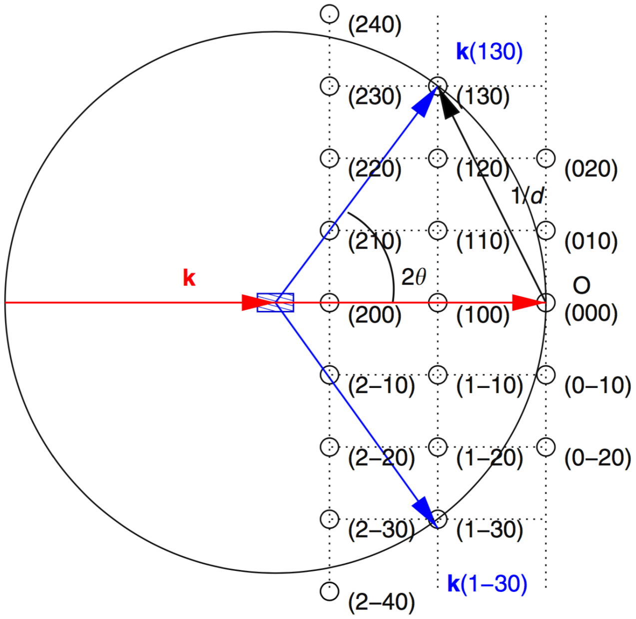


Figure 2.10. Ewald Sphere with reciprocal lattice. Incident x-ray beam (red line) hits a crystal in the origin of the sphere and gets diffracted on a reciprocal lattice point that lies on the sphere. Reciprocal lattice spacing (1/*d*) extends from edge of sphere in the direction of incident beam to the reciprocal lattice point.[60] Taken from González et al. *Comprehensive Biophysics,* 2012.

2.6 Electron Density and Phase Problem

When a crystal diffracts x-rays, a diffraction pattern is subsequently generated, which encompasses information about the amplitude and the direction of the incident x-ray, however, the phase is lost. Structure factor, F_{hkl}, is the sum of each atomic structural factor, f_{hkl}, corresponding to certain reflection *h k l* and described by Fourier series shown in equation 2.2. Structural factor equation is consistent with Bragg's law, in which atoms located on a set of equivalent parallel lattice planes diffract in phase with each other.[75]

$$F_{hkl} = \sum_{j=1}^{n} fj \ \exp 2\pi i \,(hxj + kyj + lzj) \qquad \text{Equation } \mathbf{2.2}$$

Where *fj* is scattering factor of "*j*" atom, *h, k, and l* are unit cell axes and (*xj*, *yj*, *zj*) is the atomic coordinate of atom "*j*". Additionally, Fourier transformation of structural factors, amplitudes, and phases express electron density as in equation 2.3. Intensities of diffraction spots are easily obtained, however, phases have failed to be determined during x-ray data collection.[59, 74]

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl}. \ \exp\left[-2\pi i(\mathrm{hx} + \mathrm{ky} + \mathrm{lz} - \alpha_{\mathrm{hkl}})\right] \qquad \text{Equation } \mathbf{2.3}$$

Where V is the volume of the unit cell, ρ is the electron density of an atom located at (x,y,z), *h, k, and l* are indices of reflection *hkl*, F_{hkl} is the structural factors defining reflection, and α is phase angle. Phase is determined based on phase angle, at which incident x-ray reaches the crystal. In order to resolve phases, a number of methods has been developed, like heavy atom method, anomalous scattering, and molecular replacement.[76]

2.6.1 Heavy Atom Method

Based on the knowledge that each atom in the unit cell contributes to each reflection in the diffraction pattern. Heavy atom method or isomorphous replacement is a method used often for structure solution of unknown protein. Technically, protein crystal is immersed in a solution of heavy atom, such as mercuric (II) chloride, selenomethionine, and platinum potassium chloride. Then, heavy atom position is determined and located by Patterson method, in which a vector between two atoms can be calculated and located; accordingly, other atoms' position is also defined. In this technique, heavy atom derivative crystals should have the same unit cell and symmetry as that of native protein crystal so that difference in electron density can be calculated and phases of the new protein might be determined.[59, 74]

2.6.2 Anomalous Scattering

Like isomorphous replacement, heavy atoms are also utilized to create difference in electron density resulting in anomalous signal. Conveniently, anomalous scattering requires one crystal, which can serve as both heavy atom derivative crystal and native protein crystal. This crystal is often exposed to x-rays of multiple wavelengths; this is why is often called multiwavelength anomalous diffraction (MAD). Expectedly, the likelihood of being radiologically damaged is quite high.[78] In fact, heavy atoms have the characteristic of behaving distinctively from light atoms, e.g. carbon, oxygen, and nitrogen, in that x-ray absorption drops severely at wavelengths just below their characteristic emission wavelength. This abrupt change in absorption resulting from change in λ called absorption edge, thus heavy atoms display anomalous scattering.[74]

2.6.3 Molecular Replacement

Isomorphous replacement and anomalous scattering are experimental methods for phasing, whereas, molecular replacement is a method of solving the atomic structure utilizing computer software, like *Phaser*.[79] It's the most common technique in resolving a protein structure given that a known protein crystal structure has to be already revealed and sharing nearly 40% of identical AA with a protein of interest. In other words, crystallographers often use the phases of a known protein structure to calculate initial phases of the new protein using six variables, three rotational and three translational. These variables transform coordinates of known protein by first rotating and then translating into new orientation.[59] This could be carried out by either Patterson method or maximum likelihood method. The latter is most commonly used, which predicts the best model; the one is most in keeping with findings.[76]

2.7 Structure Refinement and Validation

Electron density map is created once phases are determined using Fourier transformations and electron density equation (equation 1.3) to decipher atomic positions in protein structure.[76] The quality of the electron density and subsequent structural details rely significantly on resolution. Crystallographers are familiar with molecular graphics programs, such as crystallographic object-oriented toolkit (COOT) to fit protein's chain into the electron density map.[80, 81] The electron density maps that I've used in this thesis are 2Fo-Fc map and Fo-Fc map, which is the difference between the observed structure factor and the calculated structure factor of the model. One of the most important markers for refinement and prediction of a good quality 3D model for an atomic structure is R-factor, which is mathematically determined as in equation 2.4.

$$R = \frac{\sum \big| |Fobs| - |Fcalc| \big|}{\sum |Fobs|}$$ Equation **2.4**

R factor is a fraction computed by the sum of all difference between the observed and calculated structure factors divided by all observed structure factors.[75] During refinement, the main goal is to minimize the difference between the observed and calculated structure factors, which indicates that the model is in a great agreement with the experimental observation. Another independent factor that is calculated spontaneously with R factor is R_{free}, which is derived from the calculation of 5% of the reflections that are excluded from refinement. This is unbiased marker for the accuracy of the refined structure.[82, 83]

2.7.1 Ramachandran Plot

It is a method of validating protein structure, which examines protein structure based on angular parameters *phi* and *psi*, which are torsion angles between free rotatory αC-N and αC-C of the main chain, respectively. Amino acids of protein tend to cluster in certain regions in the plot, such as favored, allowed, and outlier regions.[84] In order to consider a valid structure, most of AA should fit the allowed or favored regions.[59] Figure 2.11 demonstrates Ramachandran plot layout.

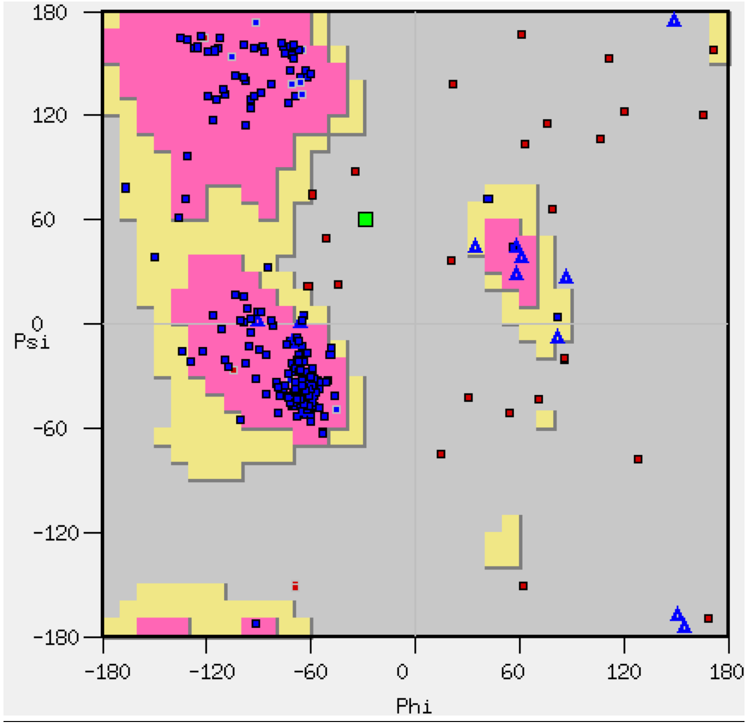


Figure 2.11. Ramachandran plot utilized for validating atomic structure of a protein. Pink colored regions are favored regions, yellow colored regions are allowed regions, and grey colored regions are disallowed or outlier regions. Blue triangles are glycine residue, and red squares represent aa that need to refine their backbone polypeptide to fit allowed and favored regions as aa represented by blue squares. Taken from initial COOT modeling of influenza virus Matrix (M1) protein.

Chapter 3. Biophysical techniques: Microscale Thermophoresis (MST) and Surface Plasmon Resonance (SPR).

3.1 Overview

Binding of biological molecules with vital endogenous ligands or drugs is of great importance for understanding physiological events at molecular level in different medical fields related to human health, including structural biology, pathology, pharmacology, and diagnostics. Binding affinity has also crucial rule in drug design process, readily filtering libraries of lead compounds. Predication of, for example, small molecules and protein target affinity would give an insight of which molecule could be further qualified for biological investigations. Also, affinity studies could answer biological phenomena, cellular networks and communications or explain disease state, and thus, guide researcher towards the appropriate disease treatment.

3.2 Principles of Thermophoresis

Thermophoresis is defined as directed movement of molecules under temperature gradient effect.[85] German Ludwig first described this phenomenon in 1858, yet its theoretical concept is still under debate. Thermophoresis can be expressed physically by the following equation 33.1.[86]

$$J = c\,DT\,grad\,T$$

Equation **3.1.**

Where J is the molecular flow, c molecule concentration, D_T is the thermal diffusion coefficient, and T is the temperature. Molecular flow proportionally correlates to temperature gradient with proportionality constant D_T. Once heat is generated in one spot of solution, molecules concentrate either at cold or hot region until steady state is reached where steady state concentration ratio is calculated as in equation 3.2.[87]

$$\frac{c\ hot}{c\ cold} = \exp(-ST\ \Delta T) \qquad \text{Equation } \mathbf{3.2.}$$

Where c_{hot} and c_{cold} are the molecule concentrations in the hot and cold areas. Soret coefficient (S_T) describes the degree of the molecule separation. Mathematically, ST is described as a fraction of thermal diffusion coefficient (D_T) relative to the normal diffusion coefficient (D). Depending on the sign of D_T, S_T could be positive, where molecules focus at cold region, or negative, where molecules focus at hot region.[88]

3.2.1 Microscale Thermophoresis (MST)

Based on thermophoresis theory, microscale thermophoresis (MST) has been developed and designed as a tool for biomolecular affinity quantification. In principle, MST is a technique used for quantitative analysis of macromolecular interactions in free solution with high sensitivity for changes in molecular characteristics during binding event, like size, charge, hydration shell and conformation. In essence, MST is operated through two fundamental actions: fluorescence and heat application via infrared (IR) laser, which is selected as heat source because it coincides with the restriction of the temperature gradient of thermophoresis.[89] Water molecules in the sample are primarily responsible for providing temperature gradient when they absorb IR radiations. According to MST application, IR laser of 1480 nm is linked to fluorescence pathway. Three readings of fluorescence in each well are recorded in accordance to the change in sample temperature, before, during, and after IR laser is turned on. Figure 3.1 depicts MST device, its components, and MST signal.[90]

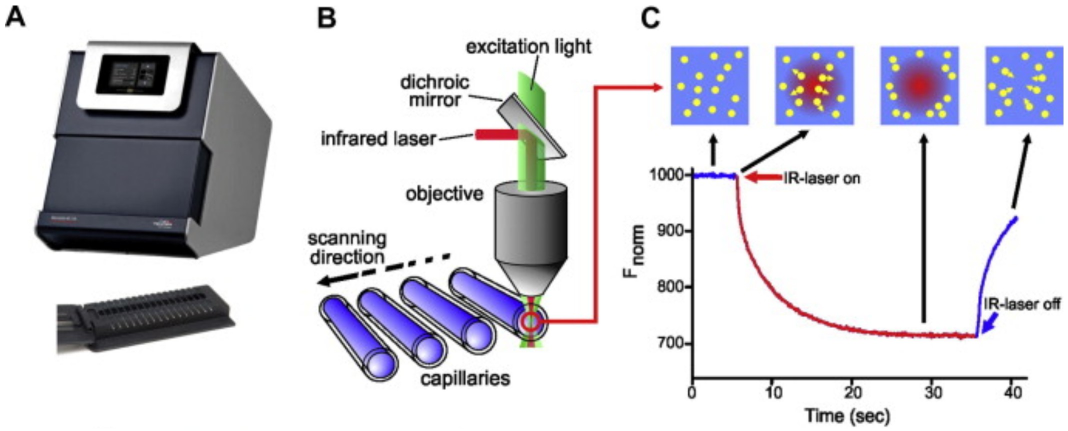


Figure 3.1. A) Image of MST device from NanoTemper technology GmbH. B) MST setup showing a focused IR laser is combined with fluorescence light excitation and emission and both IR laser and excitation and emission light are directed to a specific spot or volume of a capillary. C) Typical MST signal. Before turning on IR laser, initial fluorescence is recorded, where homogenous sample exists. Once IR laser is turned on, temperature jump (T-Jump) is noticed as a result of rapid change in temperature and fluorescence intensity of molecules. Then, fluorescently labeled molecules start to move outside the heated spot and thermophoresis is recorded till it reaches steady state shown as plateau. Afterward, when IR laser is turned off, an inverse T-Jump is observed where molecules moves back by effect of mass diffusion.[89] Taken from Jerabek-Willemsen et al. *J. Mol. Struct.* **2014**, *1077*, 101–113.

3.2.2 Experiment Aspects of MST

MST is an optical and solution-based technique used to define biomolecules characteristics. In MST, non-fluorescent partner is often titrated against constant concentration of fluorescent partner. Titration is performed by serial dilution in either 96 or 384 wells plate that facilitates the insertion of glass capillaries into wells to allow samples loading through capillary action for MST measurement. [90, 91]

With respect to fluorescence, MST can be utilized into two approaches: label free approach or intrinsic fluorescence and labeling-based approach or extrinsic fluorescence. The former is very simple approach in which the fluorescence of one reactant is utilized to monitor binding effect. Most of the time intrinsic fluorescence of the protein itself is used, arising from aromatic AA (Tyrosine (Tyr), Phenylalanine (Phe), Tryptophan (Trp). The other approach is commonly used when both reactants could interfere fluorescently with each other at the intended emission wavelength making it hard to perform the experiment without using external fluorescent molecule emitting light at a wavelength that none of reactants could have recognized for its emission spectrum. In label free, both excitation and emission wavelengths are fixed at 280 nm and 360 nm, respectively, which are related especially to protein excitation and emission, thus, limits using this approach for molecules other than proteins.[85]

From labeling approach standpoint, visible light is used to excite fluorescent molecule via three types of LED-filter: blue (excitation 460-480 nm, emission 515-530 nm), green (excitation 515-525 nm, emission 560-585 nm), and red (excitation 605-645 nm, emission 680-685 nm). There are certain AA in protein molecules considered potential sites for fluorescent dye labeling including Lysine (Lys) or Arginine (Arg) and Cysteine (Cys).[87] In these cases, an additional purification step to eliminate free dye molecules prior to binding assay is required. One important thing one has to keep in mind that location of labeling site in protein molecule is not important for thermophoretic analysis, however, it is extremely important to make sure labeling doesn't interfere with binding site. In addition to that, other fluorescence labeling approach can be used; including fluorescent fusion protein, such as green fluorescent protein. This protein is coupled to specific tag of AA sequence on protein of interest providing site specific labeling as well as purification-free approach.[85-87]

3.2.3 Mathematics of MST

MST can quantify biomolecular interaction and provides relevant interpretation within the corresponding biological system. Affinity constant (K_d), which is the equilibrium dissociation constant of the binding event is feasibly derived from MST traces. Upon microscopic temperature gradients, thermophoretic movement of fluorescent molecule (unbound) comparing to that of fluorescent molecule complex with non-fluorescent molecules (bound state) lead to different MST traces which could be utilized to estimate K_d.[85] In fact, change in thermophoresis is correlated to normalized change in fluorescence ΔF_{norm}, which is described as fraction of F_{hot}, a fluorescence value at steady state, relative to F_{cold}, an initial fluorescence value before IR laser starts working. Equation 3.3 is used to derive K_d of the binding event. Figure 3.2 demonstrates the concept of thermophoresis and fluorescence correlation.[85, 89]

$$[AL]=1/2*(([A_0]+[L_0]+ K_d)-(([A_0]+[L_0]+ K_d)^2-4*[A_0]*[L_0])^{1/2}) \qquad \text{Equation } \mathbf{3.3.}$$

Where [AL] is the concentration of the fluorescent and ligand molecules complex, $[A_0]$ is fluorescent molecule initial concentration, which is kept constant in all samples, $[L_0]$ is the initial concentration of ligand or non-fluorescent molecule, though its concentration is adjusted during serial dilution, and K_d is the equilibrium dissociation constant.

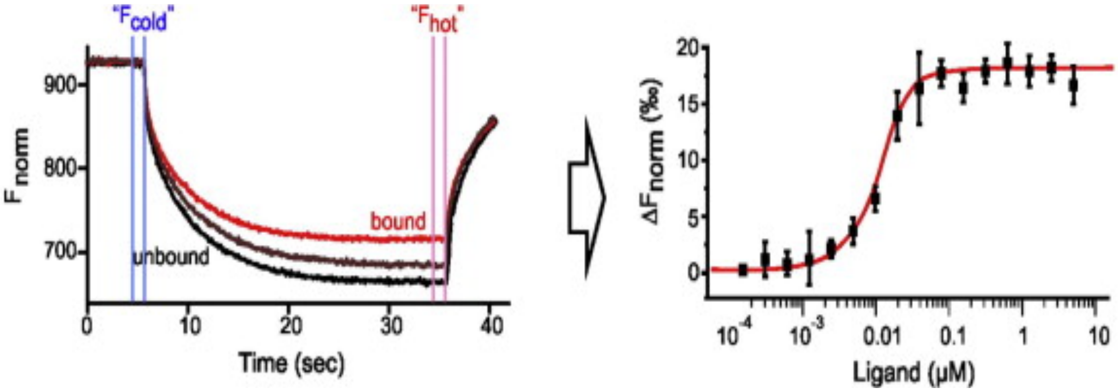


Figure 3.2. On the left side: Different MST traces are plotted for fluorescent molecule (shown in black; unbound), and bound state of fluorescent molecule with ligand (shown in red). On the right side: ΔF_{norm} (F_{hot} / F_{cold}) of each sample is plotted against concentrations of ligand (non-fluorescent molecule) from which K_d is calculated.[89] Taken from Jerabek-Willemsen et al. *J. Mol. Struct.* **2014**, *1077*, 101–113.

3.3 Surface Plasmon Resonance (SPR)

Surface plasmon resonance (SPR) is a biosensor-based optical technique, which allows various types of biomacromolecules, such as proteins and nucleic acids, to interact with other molecules in label free fashion. Kinetics binding parameters, like association rate (*k*a) and dissociation rate (*k*d) constants can be obtained along with K_d because interaction is monitored in real-time featuring SPR with a unique characteristic. Simply, SPR experiment is performed by injecting a series of analyte concentrations over a protein of interest that is chemically immobilized on a metal chip, most probably gold due to its high ability to support surface plasmons.[92]

3.3.1 Principle of Surface Plasmon Resonance

Resonance, generated physically by photons, occurs at the interface between dielectric medium (usually glass) and metal (gold or silver) when only the propagation constant of surface plasmons matches that of incident light waves at particular angle, known as resonance angle.[92] To ensure the coupling of propagation constants, prism coupler is used. Then, intensity of reflected light decreases as a result of absorption detected at that particular angle, which is specifically related to the refractive index of the medium. Upon binding events, refractive index changes for a given change in chemical or biological species at the surface, which produces change in resonance angle (Figure 3.3.).[93]

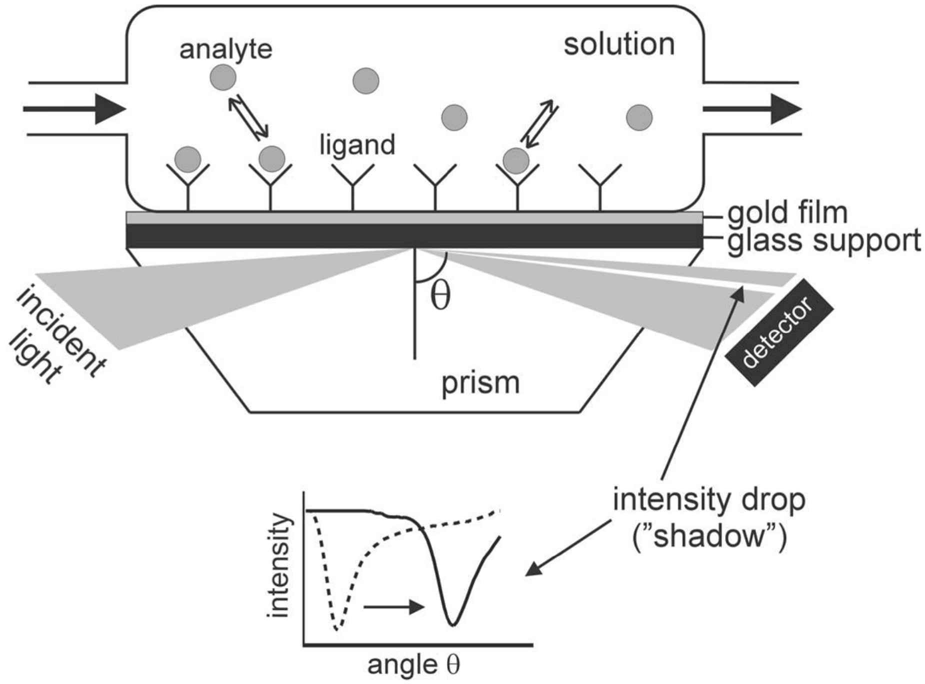


Figure 3.3. Schematic representation of SPR experiment. The instrument measures the angle of resonance at a minimum observed intensity, which changes according to change in refractive index of the medium, such as a ligand-protein binding.[93] Taken from Hahnefeld et al*. Methods Mol. Med.* **2004**, *94*, 299–320.

SPR sensorgram is a SPR signal measured in resonance units [RU] over a period of time. It represents the three primary phases during one interaction of an analyte over immobilized ligand. First, the ligand has to be immobilized on the metal chip by various chemical immobilization strategies (please see 3.3.2 section).[93, 94] Ligand immobilization would give a certain resonance unit that is considered the baseline from which analyte response is measured. In first phase, which is called association phase, analyte sample is injected over a period of time allowing binding which could be seen as an increase in SPR signal until it reaches saturation (Rmax). Depending on analyte-ligand affinity, this

SPR increase could be fast/slow or low/high in signal magnitude. Second phase is dissociation phase in which the system stops injecting analyte sample and allows buffer running only. Sometimes baseline of ligand can't be achieved so to eliminate remaining analyte molecules from the surface before injecting the next analyte sample, a regeneration solution is used, which represents regeneration phase.[95, 96] Figure 3.4 displays SPR signal with the three phases.

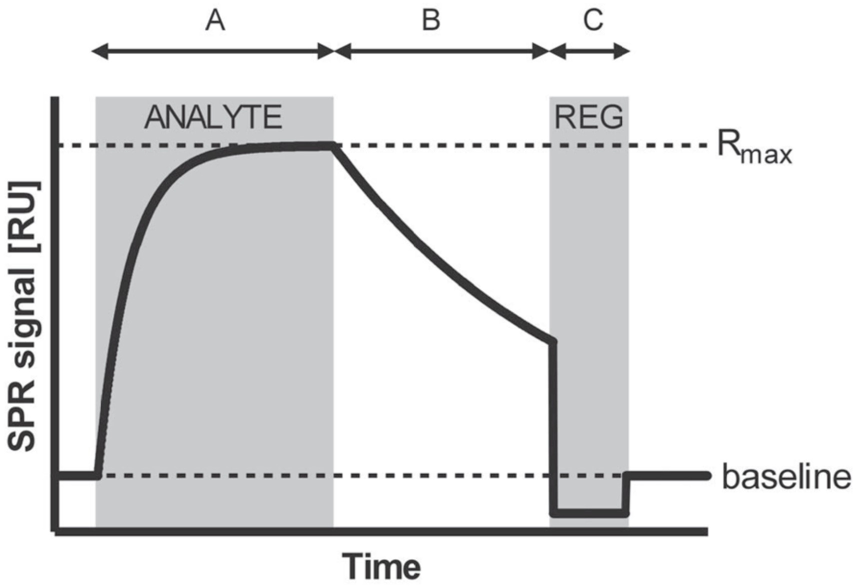


Figure 3.4. Typical SPR sensorgram with y-axis denoting SPR signal in [RU] and x-axis denoting time. (A), (B), and (C) represent association phase, dissociation phase, and regeneration phase, respectively.[93] Taken from *Hahnefeld et al. Methods Mol. Med.* **2004**, *94*, 299–320.

3.3.2 Strategies of ligand chemical immobilization

The bottleneck step in SPR experiment is the immobilization of ligand on the metal chip with taking into consideration the caveats that ligand inactivity or improper site of immobilization could happen. Great knowledge of ligand structure for determining the appropriate chemical mean of immobilization is required to successfully complete this step. There are two types of coupling chemistry: covalent and non-covalent immobilization.[97]

3.3.2.1 Covalent immobilization

It's also known as direct immobilization in which a covalent bond is formed between ligand and metal surface. This strategy involves coupling of the following chemical functional groups: amine, thiol, maleimide, and aldehyde coupling. The most common chemical way is amine coupling in which free primary amine, such as Lys, bonds covalently to carboxyl dextran chip. Sample of ligand with high purity is required. Thiol and maleimide are both thiol coupling, however, the former involves exchange reaction between thiol and active disulfide groups, while the latter involves formation of thioether. Both require the ligand to have a free thiol group (usually cysteine residue). Aldehyde coupling is useful for site directed immobilization in which a ligand possesses either inherent or induced aldehyde group. Among variety of chemical immobilizing ways, one has to ensure that the binding site of a ligand with the chip is clear of interfering with ligand active site by neither introducing conformational change nor modifying it chemically.[92-96]

3.3.2.2 Non-covalent immobilization

Non-covalent or indirect immobilization strategy involves high affinity capturing of ligand to a molecule binding covalently with a metal chip by previously mentioned covalent immobilizing reactions. For example, biotin-streptavidin interaction is widely used as non-covalent strategy in which biotinylated molecules are immobilized on streptavidin metal chip. Non-covalent strategy overweighs covalent one by a number of advantages. It's not required for ligand to be chemically modified which help to avoid improper immobilization. Unlike covalent immobilization, capturing approach is selective technique, which doesn't require very purified ligand sample. Ligand can be removed with analyte after regeneration phase providing fresh binding sites of new ligand sample. This process could consume more ligand sample at each analyte injection. However, this could be considered beneficial if the ligand molecule gets structurally changed after each analyte binding event.[93, 98]

Normally, SPR instrument comprises of three main components: sensor device, sample delivery system, sensor surface (metal chip). Sensor device, where physical and optical principles can be utilized, consists of LED-near infrared light, a glass prism, and position-sensitive diode array detector.[92, 99]

3.4 Other Techniques Used for Affinity Studies

There are plenty of biophysical techniques available for studying protein interactions but it's hard to find one that is widely applicable for most of protein interactions types. One important factor gives rise to the right and successful binding assay experiment is the awareness of researchers about strengths and weaknesses of biophysical methods and selecting the right technique, consistent with the purpose of experiment. Isothermal titration calorimetry (ITC) relies on heat change, which is either released or absorbed during biomolecular binding event, by titrating one reactant into adiabatic cell containing fixed amount of the other reactant. Good part about this approach is conducting the experiment without requiring labeling. However, it is a low sensitive technique because it depends on alteration of heat during experiment, which means high amounts of protein samples are required to produce efficient heat signal. Another label-free biophysical technique is dynamic light scattering (DLS), which is limited to certain binding effect that causes change in the average particle size. It detects time dependent variations in light scattering by binding partners. Hydrodynamic radius of each reactant has to be similar or at maximum differs in factor of two because calculation depends on substantial difference in the hydrodynamic radius of unbound reactant related to the bound state radius providing low sensitive and limited results.[89]

Surface plasmon resonance (SPR) is somehow unique in that interaction between molecules takes place on a special designed metal chip. SPR is an optical method relying on change in refractive index of the medium closed to metal chip surface, which is used to observe interaction event. It has the advantages of measuring binding events in real time, and measuring on- and off-rates with affinities range of nM to mM. Nevertheless, producing the appropriate chip is quite challenging and time consuming. Also immobilization of a protein on chip could affect binding pocket or alter protein conformation and thus inaccurate binding event assessment could happen.[96-99]

Conversely, MST provides almost all possible conditions and can be adjusted to coincide with experimental parameters. Protein sample could be prepared in solutions or in cell lysate resembling *in vivo* conditions. That is, MST could analyze protein sample regardless of its purity. One can also customize protein environment or solution with various potential factors such as co-factors, salts, and other important additives that might contribute to or reveal protein function and behavior in context of its cellular, extracellular or nuclear surroundings. Additionally, some proteins, such as membrane proteins, are only stable in complex solutions or buffers in order to assess its binding constant. Variety of molecules types involving small molecules, peptide, and proteins could be tested for binding effect with the protein of interest. With respect to maintenance and cleaning, MST instrument doesn’t require any sort of cleaning and sterilization for device parts, since application of buffer addition and titration is completed outside the device.[88-90]

Chapter 4. Crystal structure of Matrix protein (M1) Mutant (B-M1)

4.1 Introduction

Influenza virus is a negative single stranded RNA, which is composed of eight RNA segment, protected by two main layers: nucleocapsid protein and matrix M1 protein, and membrane envelope with HA and NA glycoproteins. The influenza virus can adopt various morphologic conformations owing to the protein protective shells. During virus' evolution, supplemental protein layers are developed as a barrier machinery to protect genetic material. These layers of protein also participate in viral entry and invasion. These proteins have evolved in a symmetry-related interaction to enable its functions.[33, 100, 101]

Influenza matrix protein 1 (M1) is a viral protein shell that recently has gained much attention due to its vital role in virus replication. It was assumed that M1 only functions as a structural protein, which supports virus particle shape and protects genome, however, importance of M1 protein exceeds that assumption to a broader scope involving nuclear transportation of viral genome. Additionally, many metabolic functions remain to be discovered that are proposed to involve M1 in one way or another.[9, 103] Biochemical assays indicate that M1 is quite important for virion integrity and formation. Generally, M1 provides entire physical entity of viral particle and anchored layer for membrane proteins. It also facilitates the recruitment of viral proteins into the newly formed virions.

When it’s the time for virus to release its genomic material, the interior of the virus is acidified, and in response M1 protein oligomeric layer dissociates.[101] On the other hand, M1 oligomeric layer is formed and packed again at neutral pH in host cellular cytoplasm when virus particles are assembled for budding. Different structural conformations of M1 during virus lifecycle result from direct influence of pH changes.[26, 102-105]

We were interested in studying much more structural details of M1 protein in conditions of various pH. Crystallographic study, therefore, were planned by our lab in collaboration with lab of FDA to examine the effect of mutation of G88E, R101S, and R105S on geometry of M1 structure and its oligomerization, and elucidate any structural changes that may contribute to its biological function. These mutations deliberately took place in specific regions containing charged residues and mediate M1 interactions. One of the mutation that the virus itself generates is G88R and was effective as wt-M1. Furthermore, NLS region consists of basic residues, such as Arg 101 and Arg 105, which are found critical for M1 layer formation. Surprisingly, triple mutation of R101S and R105S with G88R maintain virion integrity and activity. Therefore, G88E mutant was generated along with mutation in NLS region (R101S and R105S) to investigate the reason of G88R mutant activity and confirm the importance of positive charged residues of NLS and its neighboring regions.

4.2 Material

The N-terminal domain (1–165 AA) of mutant M1 was produced through pHW2000 plasmid expressing full-length M1 complementary DNA gene of influenza virus

A/WSN/33 (H1N1). Three point mutations were inserted into the plasmid including R101S, R105S, and G88E, located at the NLS motif and its neighboring region, respectively.[34-35, 103] The mutant is denoted as G88E-M1, was confirmed by the Center for Biologics Evaluation and Research (CBER) at U.S. Food and Drug Administration (FDA).[38]

4.3 Crystallization and Data Collection

The crystals of $N^{1–165}$-domain of G88E-M1 were obtained by automated sitting-drop vapor diffusion technique using Crystal GRYPHON crystallization robot using 96 wells, which can be sealed in less than two minutes. In each well, one drop of protein (0.2 μL) was mixed with one drop of the reservoir (0.2 μL) at 1:1 ratio, leaving the mixture to equilibrate with 58 μL of the same reservoir solution at 20 °C. Various commercially available crystallization conditions were also screened including Screen 1 & 2 from Hampton research, Wizard 1,2,3, and 4 from Rigaku. The sample of G88E-M1 had a protein concentration of 15 mg/ml in a buffer of 55 mM $KH_2PO_4/K_2HPO_4/H_3PO_4$, 0.2 M NaCl, 2mM tris(2-carboxyethyl) phosphine (TCEP), pH 3.4. The condition of the reservoir was 0.1 M Tris, pH 8.5, 8% PEG (8K). The estimated pH of crystallization drop was 6.2.

For X-ray data collection, the M1 crystal was cryoprotected with mixture of its mother liquid and 25% glycerol prior to data collection in liquid nitrogen stream. The X-ray data set was obtained at 100 K on an R-axis IV++ image plate detector using CuKα X-ray (λ=1.5417) from a Rigaku Micro-MaxTM-007 X-ray source equipped with Varimax confocal optics operating at 40 kV and 20 mA (Rigaku, The Woodlands, TX). Crystal diffracted to 2.5 Å resolution, and the data set was processed and scaled using Rigaku D*TREK software.

4.4 Structure Determination

The structure was solved through molecular replacement technique using *Phaser*-MR (simple interface) within the *Phenix* software package (version1.9).[79] Using the monomeric structure of N-terminal domain of M1 (A/WSN/33 strain) with a PDB entry of 1EA3, molecular replacement resulted in a solution of three monomers per asymmetric unit; LLG= 1173, TFZ= 13.7, R-value= 48.2. Initial refinement was performed by *Phenix.refine* within the *Phenix* software package. The refinement revealed disorder in different regions in all three monomers. After multiple cycle of refinement, the current refined structure resulted in R_{work} / R_{free} of 22.35 / 31.83 with three water molecules. Table 4.1 and Table 4.2 summarize data collection and refinement statistics. This crystal structure of N-terminal domain of M1 is abbreviated as B-M1 throughout this chapter.

Table 4.1. Data collection of x-ray reflections.

Data collection statistics	
Space group	$P\ 2_1\ 2_1\ 2$
Cell dimensions (Å)	a = 85.61 b = 133.26 c = 39.31
Resolution (Å)	29.44 - 2.50 (2.59 - 2.50)
Number of measured reflections	79188
Unique reflections	14742
Redundancy	5.37 (5.51)
1/σI	16.69 (5.01)
Completeness (%)	90.7 (91.9)
R_{merge} (%)	0.077 (0.449)

Table 4.2. Statistical data of G88E M1 protein structure refinement.

Structure refinement	
Resolution limit (Å)	26.23 - 2.5 (2.589 - 2.5)
Number of reflections	14742 (1473)
R_{work} (%)	22.35 (29.6)
R_{free} (%)	31.83 (41.68)
R.m.s.d. standard geometry	
Bond lengths (Å)	0.010
Bond angles	1.42
Dihedral angles (%)	
Most favored regions	97
Allowed regions	2.33
Average B-factors ($Å^2$)	
All atoms	59.50
Protein alone	59.50
Water	49.00

$\mathrm{Rmerge} = \Sigma \mathrm{hkl} \frac{\Sigma \mathrm{i}}{\mathrm{Ihkli}} - < \mathrm{Ihkl} > / \Sigma \mathrm{hkl}\ \Sigma \mathrm{i} < \mathrm{Ihkli} >.$

R_{free} was calculated with 5% excluded reflection from the refinement.[82]

4.5 Results and Discussion

4.5.1 Crystallization and Structure of the Truncated N-terminal Domain of M1

Truncated N terminal domain (1-165) of M1 was crystallized in a pH close to neutral (~ 6.4) but still acidic. The space group was found to be *P* 2_1 2_1 2 with unit cell parameters, a= 85.61, b= 133.26, c= 39.31 Å and α=β=γ 90°, and a trimer (monomers A, B, C) in the asymmetric unit as shown in Figure 4.1. Molecular replacement was processed based on the fact of having three monomers in the asymmetric unit using a search model 1EA3, a monomeric structure of N-terminal domain of wt-M1 in neutral pH. The final refinement produced R_{work} / R_{free} 22.35 / 31.83.

Space group *P* 2_1 2_1 2 is an orthorhombic crystal with two-fold screw axis along axes (a) and (b) and two-fold rotation along axis (c). Each monomer, A, B, and C, is related to each other by two-fold non-crystallographic symmetry (NCS) as in Figure 4.1 (A). Monomers A and B in B-M1 are packed into face-to-face fashion resembling that of wt-M1 in acidic pH. Monomers A and B form a physiological dimer arranged in the so-called “face-to-face” fashion as previously observed for M1 dimers obtained at low-pH of crystallization. Monomer C also forms a dimer with its symmetry-related counterpart, also in a face-to-face fashion. It should be noted that M1 structures form physiological monomers when crystallized at pHs over 7. The different oligomerization states of M1 at different pH have been proposed to be related to its multiple function.[40]

The N-terminal domain of M1 used for the structural study has a molecular weight of ~18 kDa and consists of nine helices with no β-strands. The nine helices are further divided into two domains N and M domains and each comprises of four helices. N-domain contains H1 (3-12), H2 (19-32), H3 (39-48), and H4 (54-67), while M domain consists of H6 (90-105), H7 (106-116), H8 (121-132), and H (140-157). Both domains are linked by small helix H5 (78-84). Moreover, each helix is separated from the other by a loop. Each of the three monomers has two distinguished faces; one face occupied mostly by positive amino acids, while the opposite face is covered mostly by negative AA. Figure 4.1 (B) demonstrates the two opposite sides, in which positively changed AA, NLS motif or 101RKLKR105 and Lys95, Lys98, and Arg134 are located on one side and on the other side negatively charged AA are lined up, such as Glu8, Glu23, Glu29, Asp30, Asp38, and Glu44. Primarily, these charged AA define electrostatic interactions of M1 with several viral proteins and in M1-M1 oligomerization at high pH, where the monomers associate in a face-to-back fashion.[22, 39]

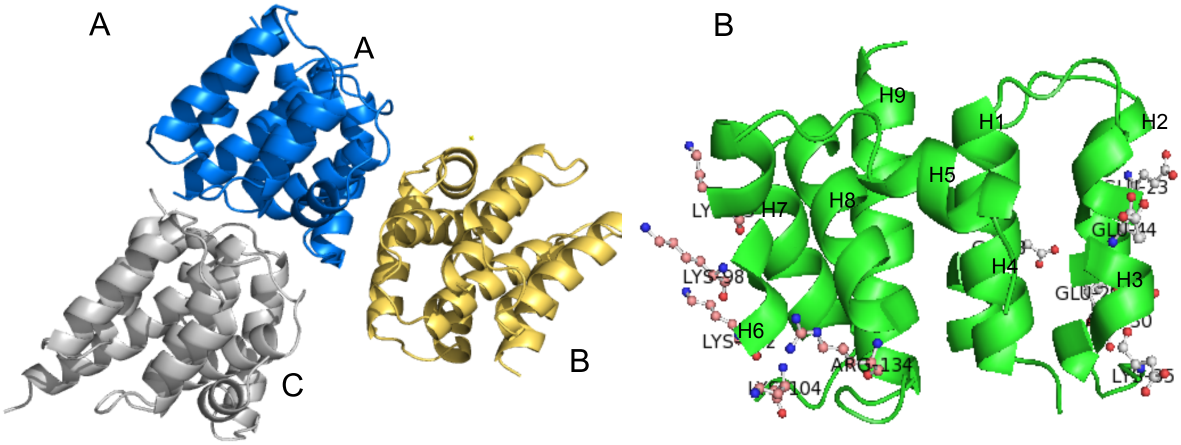


Figure 4.1. (A) Three monomers of B-M1 in asymmetric unit (A monomer in blue, B monomer in gold, and C monomer in silver). (B) A monomer geometry with two bundles of four helices: N domain (H1-H4) and M domain (H6-H9). The side of H2 and H3 contains mainly acidic aa where is the side of H6 and H7 contains mainly

During structure validation by COOT software version 0.8, each monomer showed high degree of disorder in certain AA side chains. Generally, N-domain in all monomers has shown higher disorder degree comparing to M-domain. The calculated average-B factors for N-domain vs M-domain are in the following order for monomer A, B, and C: 61.12 Å^2 / 57.57 Å^2, 59.27 Å^2 / 53.22 Å^2, and 67.42 Å^2 / 57.88 Å^2. However, monomer C has the highest average-B factor among B-M1 monomers with average-B factor of 62.49 Å^2 comparing to A and B monomers, whose average-B factors are 59.83 Å^2 and 56.16 Å^2, respectively. Despite these figures, all monomers have similar average-B factor range, which is expected with the obtained resolution of 2.5 Å. Furthermore, monomers displayed similar geometry in their structures when they are superimposed to each other with RMSD ranged from 0.35 Å to 0.5 Å, except for some difference in loops folding.

The AA sequence, 69-75, is highly disordered and lacked continuous electron density in all monomers of B-M1, and was therefore deleted from the structure. It's important to note that similar observation has been reported in the same region in previously resolved M1 crystal structures.[22, 39] Ser2, Leu3, and Leu4 of monomers A and C but not B had highly constrained peptide bond. Application of multiple cycles of deleting AA, refinement, and adding AA was implemented until acceptable electron density was achieved to fit the peptide bond of the three AA.

To improve refinement, the occupancy of Gly136 was adjusted to zero in monomer C since it is disordered as shown in Figure 4.2 (A). In monomer A and C, Asn36, Thr37, Asn87, and Asn91 were removed and rebuilt manually in order to improve overall model validity, due to steric hindrance of *phi* and *psi* angles conformations, which eventually results in acceptable Ramachandran plot Figure 4.2 (B).

The intrinsically disordered regions of M1 could perform special biological function, contributing to vital interaction. Similar observations have been observed for many proteins that also contain intrinsically unstructured regions. These regions highlight proteins with multifunctionality and plasticity to accommodate different conformations required for interactions with different proteins.[106-108]

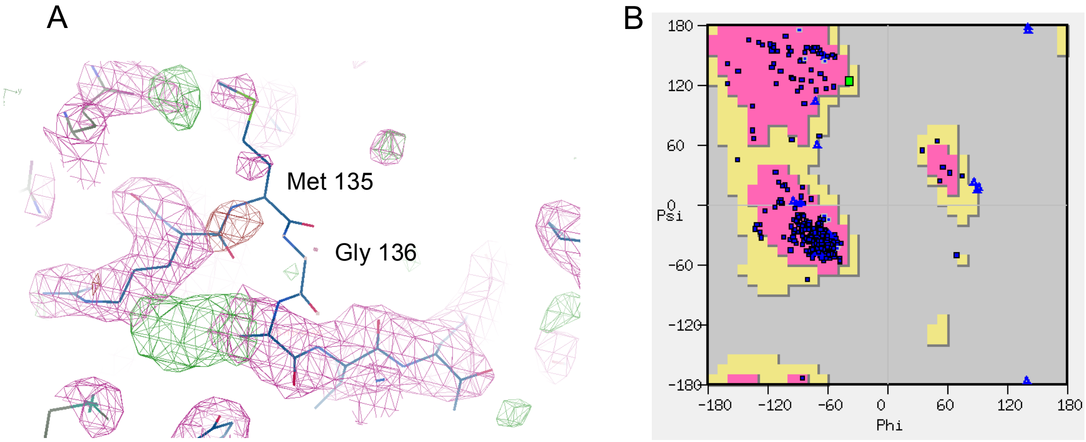


Figure 4.2. (A) AA Gly 136 and Met 135 are out of density and hard to fit them in density. (B) Final modeling with COOT resulted in zero outliers, 96.44% and 3.56% of AA are in preferred and allowed regions of Ramachandran plot, respectively.

Although this structure of M1 is His-tagged at its N terminus, we didn't observe any relevant density in all B-M1 monomers. However, positive helical electron densities at the end of monomers B and C were found after Gln158 and filled with the corresponding AA. In monomer B, AA sequence of His159, Arg160, and Ser161 was fitted except for Arg160 side chain which was mutated to Ala, while monomer A showed no density to any AA after Gln158 as depicted in Figure 4.3. Moreover, AA sequence: His159, Arg160, Ser161, His162, Arg163, Glu164, Met165 was built in monomer C, however, side chain of His159 and Arg160 lacked for electron density, so they were mutated to Ala. These AA were conjoined into atomic structure of monomers B and C at an RMSD contour of 0.9 σ as shown in Figure 4.4.

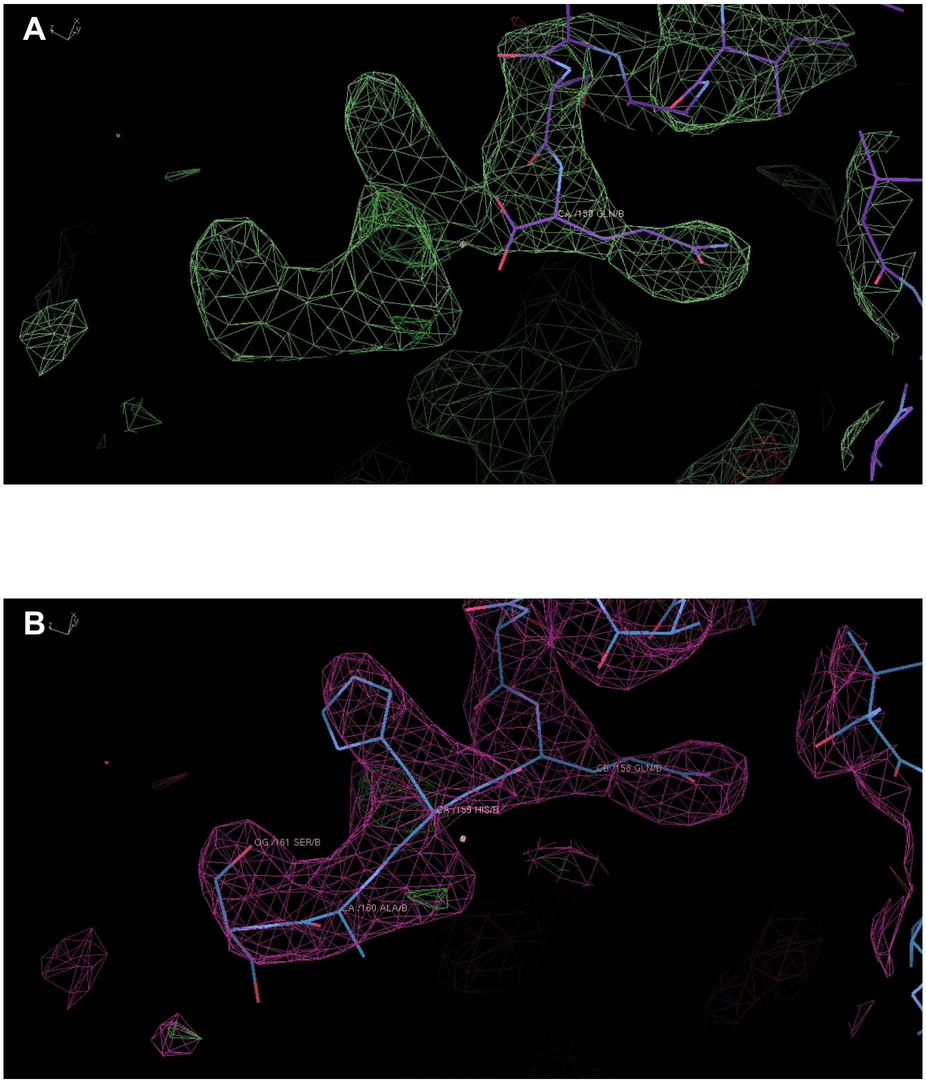


Figure 4.3. (A) The electron density at the end of Gln158 of monomer B. (B) Fitted aa sequence of His159, Ala160, and Ser161 into electron density.

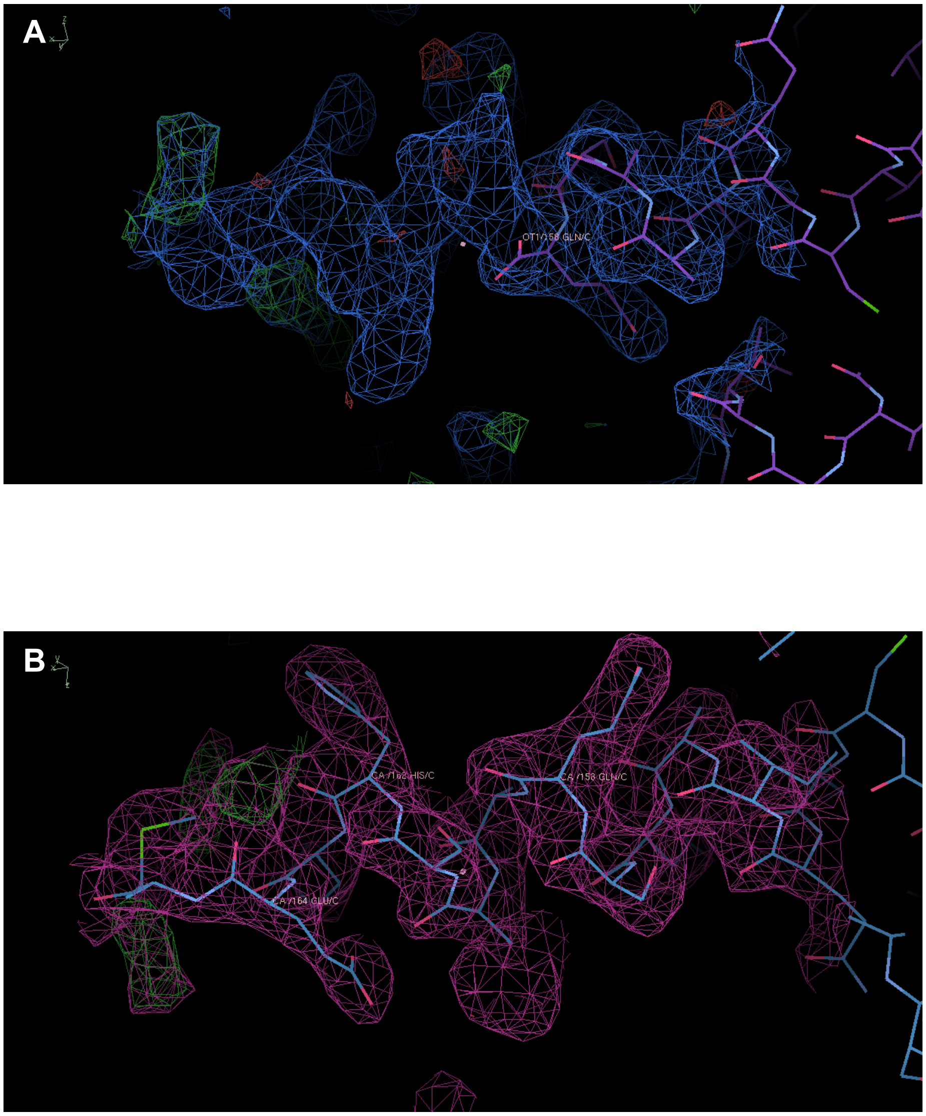


Figure 4.4. (A) The electron density at the end of Gln158 of monomer C. (B) Fitted aa sequence of Ala159, Ala160, and Ser161, His162, Arg163, Glu164, and Met165 into electron density.

Another important modification step during refinement was mutating G88, R101, and R105 to E88, S101, S105 since the search model is wt-M1 with these residues being Gly88, Arg101, and Arg105. This mutation was also observed through electron density change, for example, density of both Arg101 and Arg105 was missing and side chains of serine residues were nicely fitted into their corresponding electron densities. The same thing took place for Gly88 replacement with Glu as depicted in Figure 4.5.

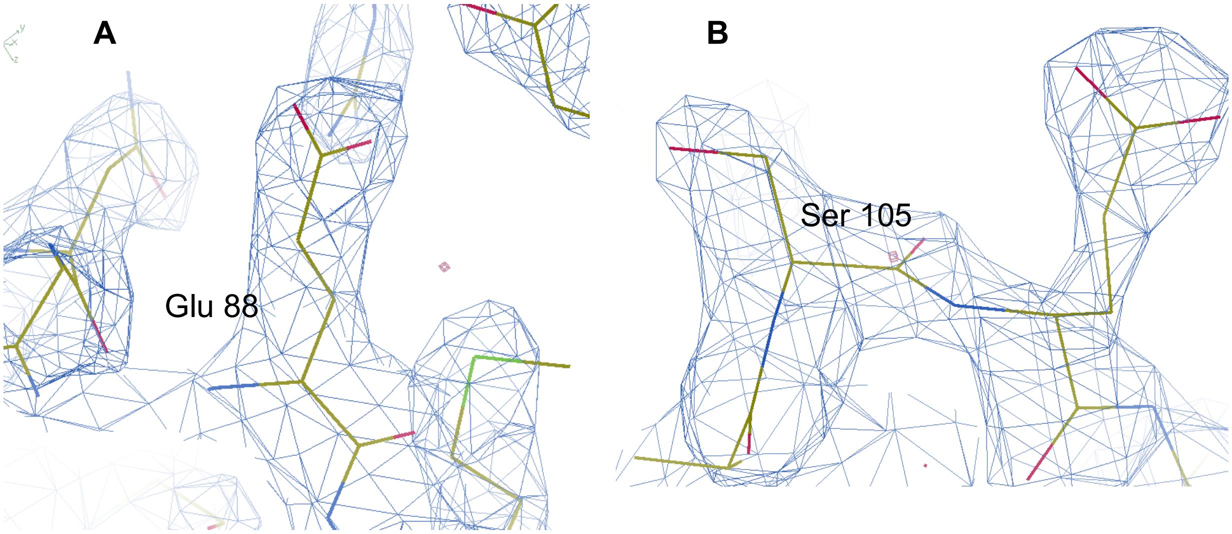


Figure 4.5. COOT modeling. (A) Gly88 was replaced with Glu88 and its side chain was well fitted into its relative electron density. (B) Smaller electron density projected from main electron density path corresponding to Ser105 side chain instead of Arg105.

4.5.2 Structure Comparison with Other M1 Structures

There are several M1 crystal structures deposited in PDB of which only three are known for their detailed conditions, including 1AA7 (pH~4.3), 4PUS (pH~4.7), and 1EA3 (pH~7.5), which form dimer in their asymmetric unit except 1EA3. When comparing the structure of 1AA7 and 1EA3, 1AA7 dimerizes into face-to-face fashion, where, 1EA3 dimerizes into face-to-back fashion.[40, 110] The interaction involved in monomer-monomer association determines the pattern in which M1 molecules dimerize. For example, 1EA3 dimerization is mainly derived by electrostatic interaction between patch of positive charge AA from monomer A including Lys95, Lys98, Arg101, Lys102, Lys104, Arg105, and Arg134 as well as Asp94 with negative charged AA from monomer B including Glu8, GLu23, Glu29, and Asp30 plus Lys35 as shown in Figure 4.6.[39] Conversely, at acidic medium, the M1-M1 dimer structure, such as observed with 1AA7 and 4PUS occur due to mainly hydrophobic interaction in such a way that Pro90, Met93, Val97, Leu130, and Met135 on monomer A interact with the same AA on monomer B as depicted in Figure 4.7 (A). Additionally, hydrogen bond interactions take place between both monomers, summarized in (Table 4.3).[22] Equally, B-M1 dimerizes like 1AA7, however, they differ in many aspects in terms of interface alignment, monomer-monomer interaction, and buried surface area (Table 4.3). We noticed largely unaligned loop L8 portion of B-M1 with 1AA7. This is primarily due to the replacement of Gly88 in 1AA7 with Glu88 in B-M1 resulting in the disturbance of the hydrophobic AA cluster in the M-domain that is involved in the dimerization of the M1 monomers. A closer look at the interface of 1AA7 shows loop L8 with the hydrophobic AA, Met135, Gly136, Ala137, and Val138 protruding into the hydrophobic interface and interacting with another hydrophobic AA as depicted in Figure 4.7 (B).[22] In case of B-M1, this loop is shifted inside a monomer due to projection of polar Glu88 side chain into the interface from the other monomer as demonstrated in Figure 4.8. For example, Met135 folds inside a monomer to interact with hydrophobic AA, such as Leu66 and Phe62 on H4. To further investigate the interface differences, solvent accessible buried surface area (computed by Chimera UCSF software) and structure alignment (computed by PyMol software) were estimated.[109] The buried surface area of 1AA7, B-M1, and 1EA3 are 2272 Å^2, 1403 Å^2, and 1063 Å^2. Interestingly, the buried area of B-M1 is far less than that of 1AA7 as a result of reduction of the interfacial hydrophobic interaction though they have the same interface of dimerization. Least-squares superposition of monomer A of B-M1 with 1AA7 showed similar monomeric structures with RMSD value of 0.6 Å, whereas the superposition of all Cα residues resulted in RMSD value of 1.35 Å. To align monomer B of B-M1 onto its counterpart of 1AA7 a screw rotation angle/translation of 10.1°/0.22 Å is needed. Alignment of 1AA7 and B-M1 is visualized in Figure 4.9. Despite the perturbation in hydrophobic interaction zone caused by mutation of G88E, glutamate forms a net of hydrogen bond and electrostatic interaction with Tyr100, Lys104, and Arg134 as shown in Figure 4.10.

Furthermore, structural difference other than Loop L8 between 1AA7 and B-M1 was placed on folding of Loop regions. Of B-M1, Loops, L1, L2, L3, L6, and L8, were not well aligned between the two structures, among which L8 is the most unlayered with its counterpart on 1AA7. On the other hand, L5 and L7 are well aligned.

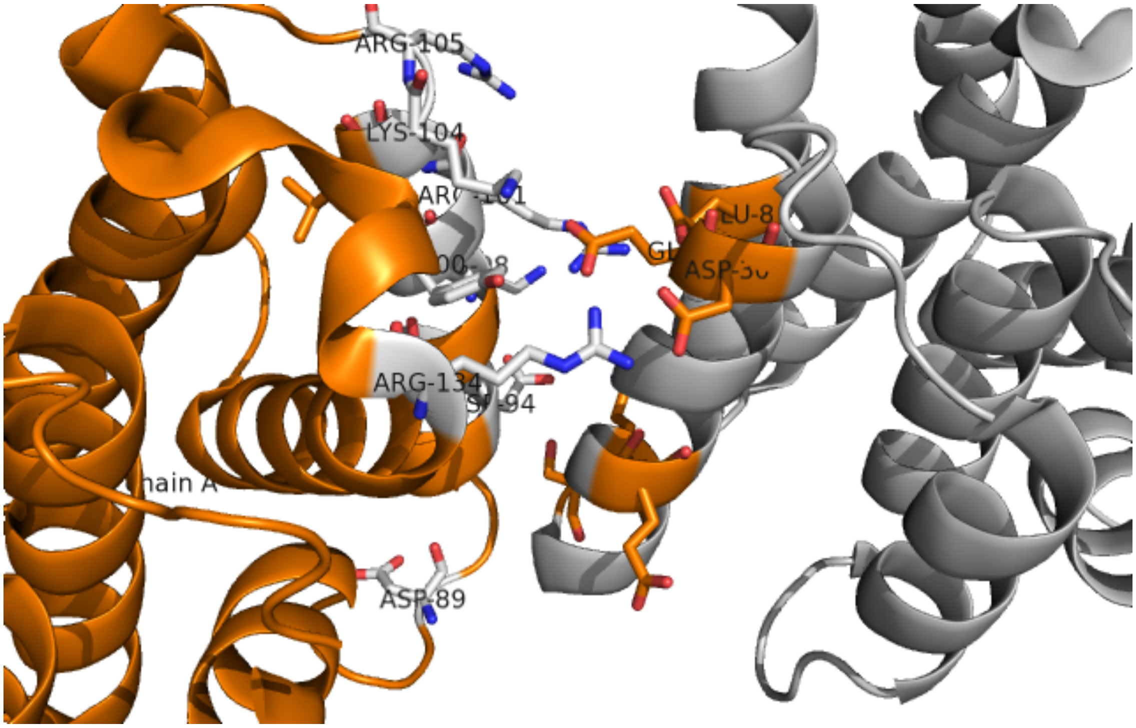


Figure 4.6. 1EA3 dimer interface, which is mainly maintained through Hydrogen bond or electrostatic interactions which are summarized in Table 4.4 (monomer A in orange (Carbon: Silver, Nitrogen: Blue, Oxygen: Red) and monomer B in silver (Carbon: Orange, Nitrogen: Blue, Oxygen: Red)

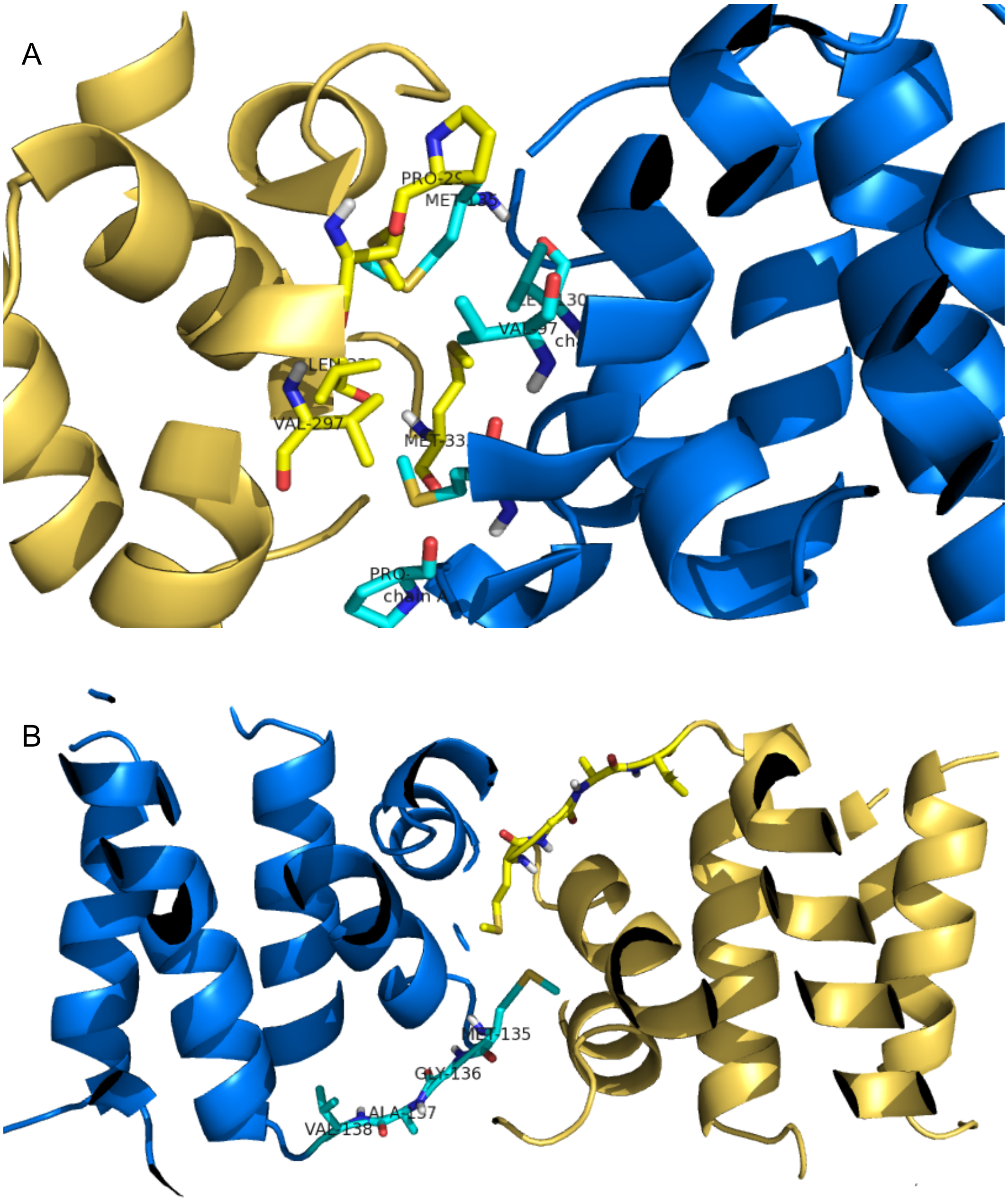


Figure 4.7. 1AA7 dimer interface that (A) is stabilized through hydrophobic interactions of Pro 90, Met 93, Val 97, Leu 130, and Met 135 from both monomers. (B) Loop L8 (Met 135, Gly 136, Ala 137, and Val 138) of each monomer pointed toward interface (A monomer in blue (Carbon: cyan, Oxygen: red, Hydrogen: white, Sulfur: gold, and Nitrogen: blue), and B monomer in gold (Carbon: yellow, Oxygen: red, Hydrogen: white, Sulfur: gold, and Nitrogen: blue)).

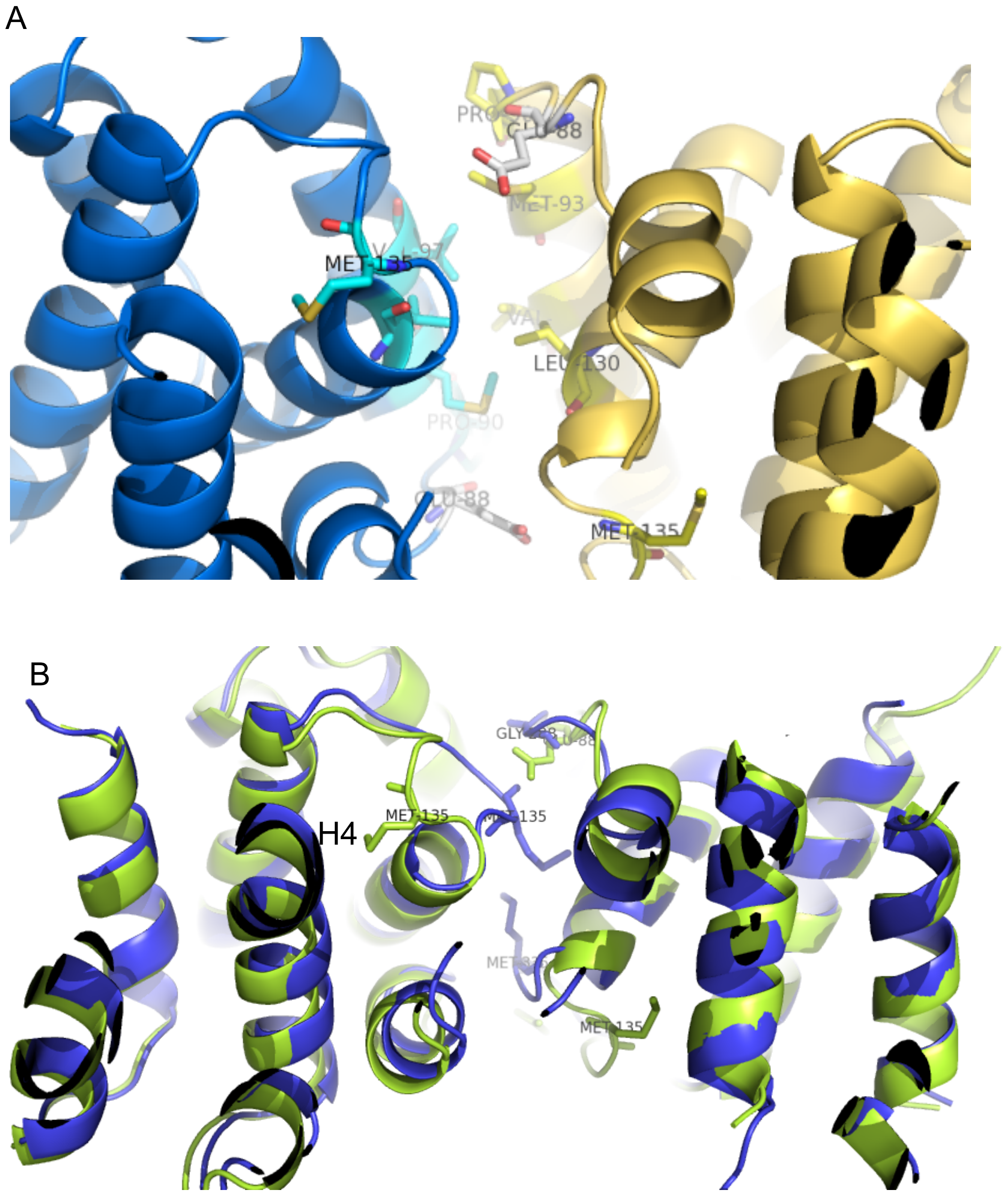


Figure 4.8. B-M1 monomers interface is (A) stabilized through hydrophobic interactions of Pro 90, Met 93, Val 97, and Leu 130 from both monomers. (A monomer in blue (Carbon: cyan, Oxygen: red, Hydrogen: white, Sulfur: gold, and Nitrogen: blue), and B monomer in gold (Carbon: yellow, Oxygen: red, Hydrogen: white, Sulfur: gold, and Nitrogen: blue)). (B) Difference in L8 between 1AA7 (blue) and B-M1 (green). L8 in 1AA7 protrudes into monomers interface, and in B-M1 it folds back inside the monomer to interact with H4 hydrophobic residues.

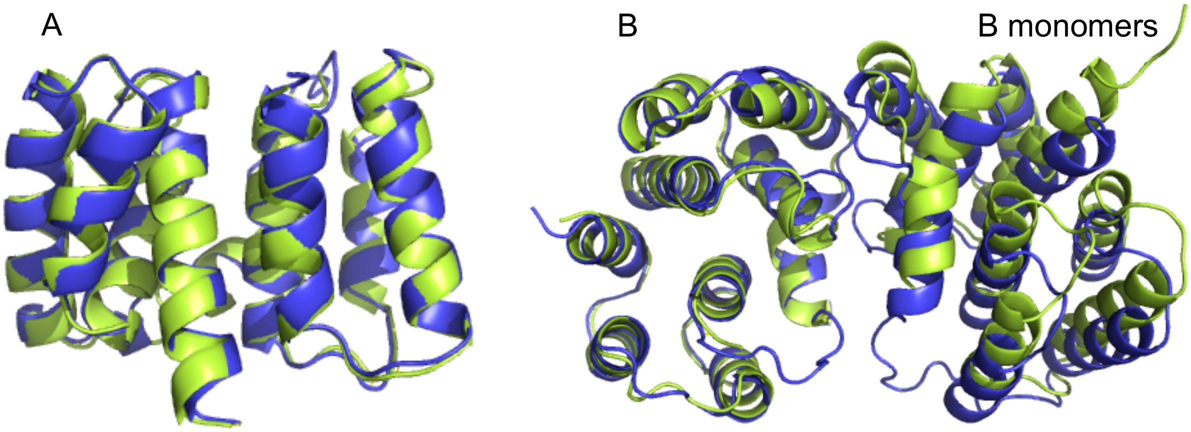


Figure 4.9. (A) Monomers A of both 1AA7 and B-M1 are aligned (1AA7 in blue, B-M1 in green). (B) Overlapping of monomers “A” of the two dimers and monomers “B” showed more deviation than monomer A.

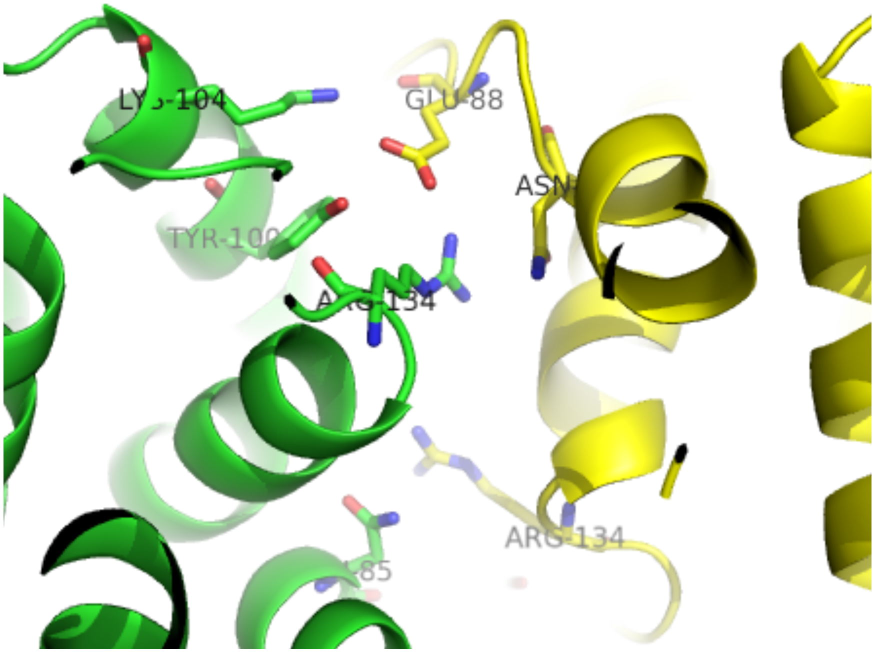


Figure 4.10. Interface of B-M1. Glu88 interacts with Lys104, Tyr100, and Arg134. Another Hydrogen bond is between Asn85 and Arg134. (monomer A in green, monomer B in yellow)

4.5.3 Physiological relevance of M1 Conformation and Oligomerization

As described above, dimeric M1 structures show significant differences in their interface arrangements that might represent several intermediate conformations of M1 during cell invasion in accordance to subtle changes in pH, which can also determine the oligomerization state, either face-to-face or face-to-back.[40] At neutral pH, all acidic and basic residues of the two monomers get ionized and favor electrostatic interaction leading to face-to-back dimerization fashion (Table 4.4).[110] Conversely, at low pH, acidic AA become neutralized leaving basic AA ionized on same side of a monomer, such as Lys35 and Lys21, generating repulsive force with positively charged AA, such as Arg134 and Lys98, on another monomer leading to different dimerization form.[109] In acidic pH condition, the main driving force for dimerization of face-to-face fashion is hydrophobic interaction involving helices H6 and H8 and Loop L8.[22]

Oligomer is the super structure of M1 layer when there is an intact viral particle. Crystal structures were used to predict the model of M1 layer. It was predicted that M1 layer forms either coil or strands.[110] Another gel filtration study has revealed that M1 oligomerization is a pH dependent process in which M1 stays in oligomeric state at neutral pH and dimer in acidic pH. Upon pH neutralization, M1 maintains dimerization instead of oligomerization, which could be a result of pH irreversible change of M1 structure.[117] The smallest oligomerization state that M1 can form is a dimer.[109] Despite mutation, B-M1 keeps the same packing fashion, face-to-face, in acidic condition. Although, some hydrophobic interactions disappear at the dimer interface, Glu88 generates hydrogen bond interactions with Tyr100, Lys104 and Arg134 that account for stabilized monomer-monomer association.

Table 4.3. M1-M1 interface interaction with buried interface residues of B-M1, and 1AA7.

Monomer A	**Monomer B**	**Distance (Å)**	**Buried Surface Area (Å^2)**
B-M1			1403
Glu 81 [NE2]	Asn 133 [O]	2.4	
Arg 76 [NH2]	Asn 133 [OD1]	3	
Arg 134 [NH2]	Asn 85 [OD1]	3.4	
Arg 134 [NH1]	Glu 88 [OE2]	2.7	
Tyr 100 [OH]	Glu 88 [OE1]	2.9	
Lys 98 [NZ]	Asp 94 [OD1]	3.4	
Asp 94 [OD2]	Lys 98 [NZ]	3.5	
Lys 104 [NZ]	Glu 88 [OE1]	3	
Glu 88 [OE1]	Tyr 100 [OH]	3.2	
Asn 133 [O]	Gln 81 [NE2]	2.5	
1AA7			2272
Gln 75 [OE1]	Arg 78 [N]	2.9	
Arg 78 [N]	Gln 75 [OE1]	3.0	
Arg 76 [O]	Gln 75 [NE2]	3.2	
Gln 81 [OE1]	Arg 134 [NE]	3.1	
Gly 88 [O]	Tyr 100 [OH]	2.9	
Tyr 100 [OH]	Gly 88 [O]	2.7	
Arg 101 [NH2]	Asn 91 [OD1]	3.0	
Asn 91 [OD1]	Arg 101 [NH1]	3.4	
Asn 133 [O]	Asn 85 [ND2]	3.1	
Asn 85 [ND2]	Arg 134 [O]	3.1	
Arg 134 [O]	Asn 85 [ND2]	3.2	
Gln 75 [NE2]	Arg 77 [NH1]	3.4	
Asn 133 [ND2]	Asn 133 [O]	3.6	

Table 4.4. M1-M1 interface interaction with buried interface residues of 1EA3.

Monomer A	Monomer B	Distance (Å)	Buried Surface Area (Å^2)
Asp 94 [OD2]	Lys 21 [NZ]	3.2	1063
Tyr 100 [OH]	Glu 29 [OE1]	3.5	
Tyr 100 [OH]	Glu 29 [OE2]	3.7	
Arg 134 [NH1]	Asp 30 [OD1]	2.7	
Arg 134 [NH2]	Asp 30 [OD2]	2.5	
Arg 101 [NH2]	Glu 8 [OE1]	3.4	
Arg 101 [NH2]	Glu 8 [OE2]	3.4	
Lys 104 [NZ]	Glu 29 [OE2]	3.1	
Asp 94 [OD2]	Ser 17 [OG]	3.1	

Self-polymerization and depolymerization of M1 are two major features that mediate M1 functions and interactions.[100] Self polymerization of M1 is important to build a full matrix layer that supports viral shape and skeleton, however, during virus lifecycle it is necessary for M1 to breakdown, thereby, viral particle and its components can be fused into host cell.[111] Accordingly, conformational changes of M1 in response take place of which each structural conformation uniquely interact with other proteins or act as an intermediate.[40] Specific AA sequences are responsible for a number of M1 interactions either hydrophobic or hydrophilic interaction, thus, M1 is characterized as amphitropic molecule.[112] At neutral pH, M1 exists in continuous oligomer underneath lipid bilayer where positively charged areas are lined up towards the core of the virus to interact and hold vRNP in virus core through quantum electrostatic interactions. A segment of AA from 80-111 in helices H6 and H7 is critical for vRNP binding owing to positively charged NLS motif and other positive charge neighboring AA that bind to negative charged molecules of RNP. There is controversy concerning whether negative charge comes from the phosphate group of RNA or acidic AA on NP.[20, 22] Although, NLS motif contributes to dimer and oligomer formation at neutral pH, its interaction is not essential for M1 to dimerize or associate with membrane. A site-directed mutagenesis study has suggested that polybasic residues of NLS motif are not mandatory for membrane association, but it could be trivial.[113]

Sha et al had suggested that N domain involves mainly in M1-lipid interaction. Helices H1 and H4 has hydrophobic interactions with lipid bilayer via Leu3, Leu4, Val7, Tyr10, Val11, and Ile14 from H1 and Pro54, Leu55, Ile59, Phe62, Val63, and Leu66 from H4.[22] Like 1AA7, H1 and H4 of B-M1 are not exposed to bulk solvent; hence they are in non-binding conformation. Sequence of Arg76, Arg77, and Arg78, which are located right after H4, has also found to facilitate electrostatic interaction with lipid membrane, in particular phosphate heads.[32] This charge-charge interaction, formed either by NLS residues or 76-78 AA or both, has been prevented *in vitro* by salt dependent study. It's clearly that multitude of M1 sites contribute to membrane binding forming final strong association of the two layers. One recent study has shown that M1 contains cholesterol recognition/interaction amino acid consensus (CRAC) motifs, involving Val31, Met93, Val97, Lys98, Tyr100, Lys104, and Leu130 that provides multi-point interactions with viral membrane.[102] With respect to B-M1, mutation of G88E could also affect membrane binding. Since it is acidic and polar, it may deteriorate M1-membrane binding through two events: first, polar side chain causes bending and hiding of hydrophobic residues inside M1 protein inhibiting their exposure to the membrane. Second factor is the acidic characteristic of Glu that repulses interaction with phosphate groups on the membrane. This was observed when M1 layer was disturbed in virion with B-M1 under TEM.[38]

Normally, M1 dissociate between pH 6.0 and 5.5.[103, 114] This dissociation takes place when viral particle gets acidified at fusion step causing M1 layer disappearance and release of M1 dimer or monomer, and subsequently, the release of genetic material due to conformational changes in M1 structure. M1 has a role in newly synthesized vRNP transportation from nucleus.[100] In this case, M1 should exist in dimer or monomeric state to expose NLS motif for charge interaction with glutamate residues on NEP/NS2.[32] In the same time, C terminal of M1 (166-252 AA) binds to vRNP complex and eventually the triple complex travel across nuclear pores to the cytoplasm for viral particle generation steps.[24, 115] A site directed mutagenesis of four basic residues in NLS, where Arg and Lys replaced with Ala (101AALAA105), resulted in severe reduction in M1-NEP/NS2 binding affinity.[113, 116] From B-M1 standpoint, mutation of R101S and R105S reduces basicity nature of M1, thus could weaken M1-NEP/NS2 interaction.

4.5.4 Crystal Packing and M1 Oligomerization

Crystal packing could be used to interpret M1 oligomerization and build structural model that is required to maintain virus integrity and M1 interactions. At acidic pH, two M1 monomers dimerize into face-to-face fashion by hydrophobic interactions. Acidic pseudo tetramer is formed by two stacked 1AA7 dimers as one dimer on the top of the other and related through non crystallographic two-fold axis, while 4PUS had failed to show fully stacked pseudo tetramer due to 2_1 symmetry related molecule.[40] On the other hand, two monomers of M1 at neutral pH dimerize into face-to-back fashion through electrostatic interaction. 1EA3 dimer at neutral pH forms pseudo tetramer by translational stacking interaction that clusters patch of basic AA on one side and acidic AA on the other side as in Figures 4.11.[39] Only neutral pH tetramer generates the intact M1 layer that is truthfully functional M1 layer. At neutral pH, each side of monomer is available for other M1 molecule binding and eventually all monomers form final M1 polymer. In comparison, acidic pH tetramer fails to produce intact M1 layer since its monomers are assembled into face-to-face fashion in which interactive M1 sides are not accessible for binding with other M1 monomer. Expectedly, B1-M1 structure generates pseudo tetramer at acidic pH condition in the same manner as 1AA7. However, a crystallographic symmetry related copy of monomer C was generated to build up the tetramer, as depicted in Figure 4.12. One important point is that this tetramer lacks any physiological relevance since the tetramer at neutral pH is the one that has been shown to form M1 oligomeric layer.[101] If we consider B-M1 tetramer as two dimers, they are related by NCS two-fold axis.

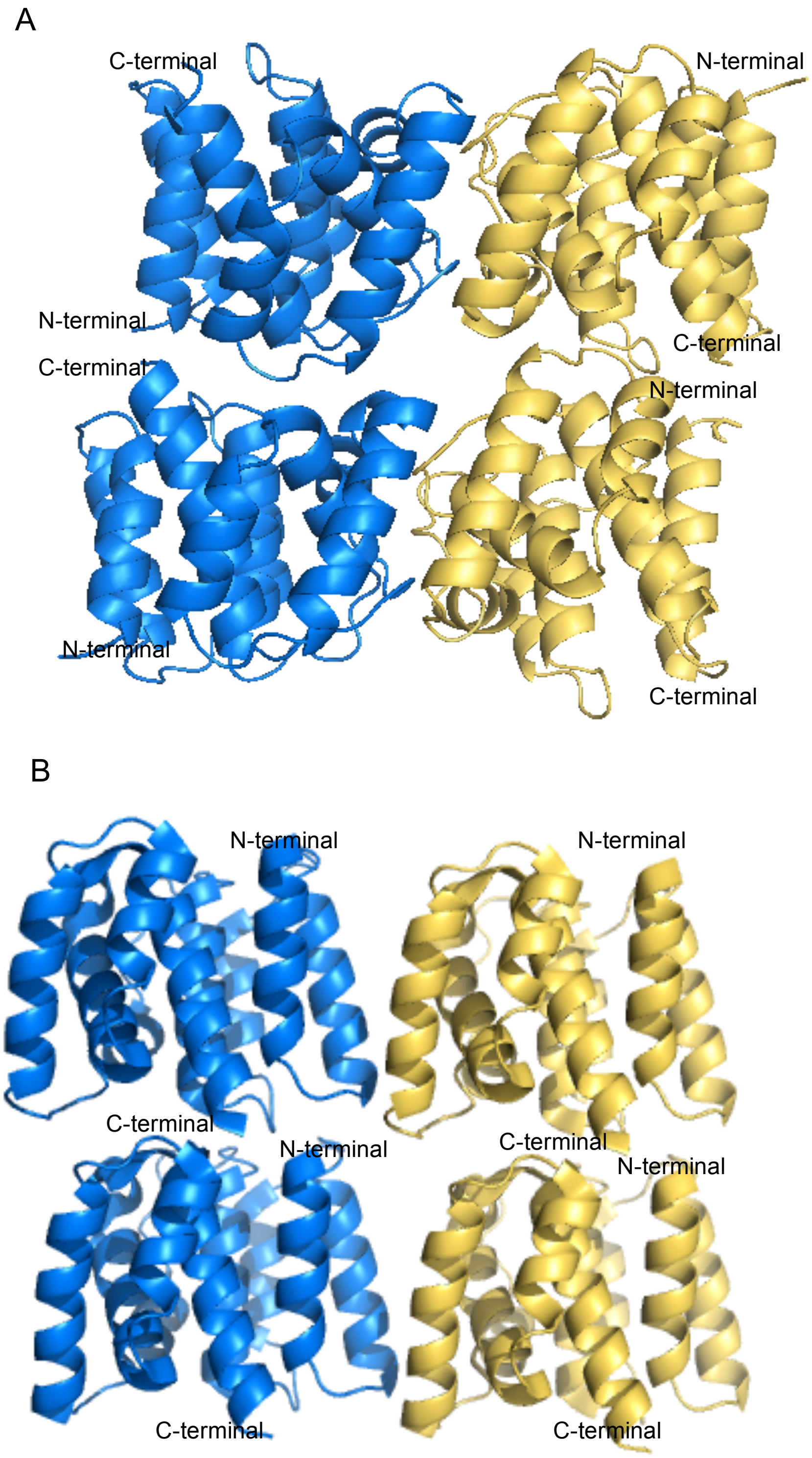


Figure 4.11. Tetramers of (A) 1AA7 and (B) 1EA3. (monomer A in blue and monomer B in gold)

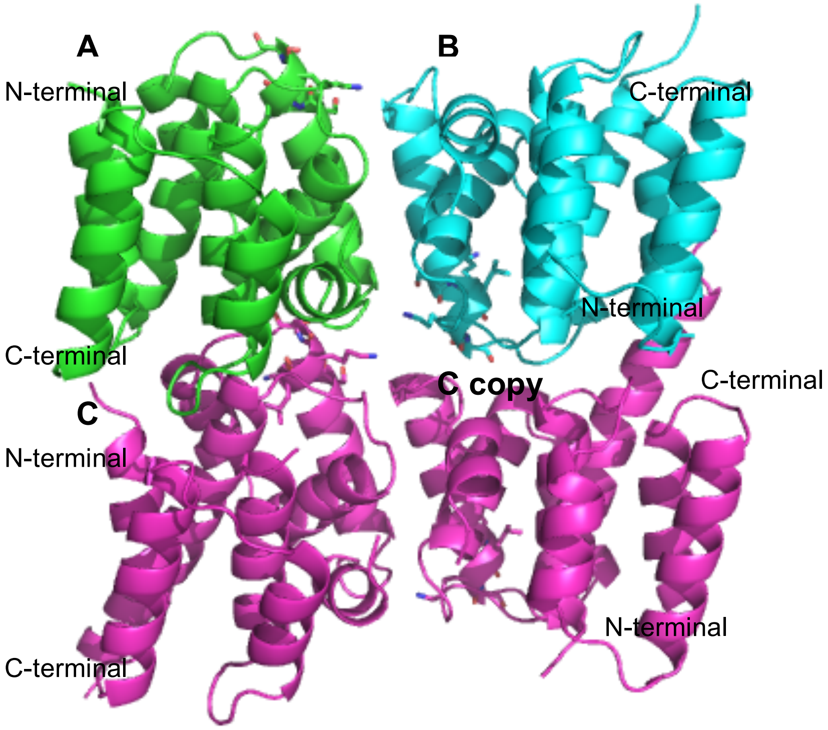


Figure 4.12. B-M1 tetramer is arranged as two dimers (A and B as one dimer) and (C and C copy as the other dimer) in such a way one dimer is on the top of the other.

4.6 Conclusion

We have crystallized M1 at pH 6.5, and as expected the structure crystallized as a dimer. Nonetheless, the dimer arrangement shows significant differences with 1AA7 which also crystallized as a dimer at low pH. Surprisingly, it delivers the structural keys relevant to dimerization of M1 in acidic condition despite the three point mutations of Gly88, Arg101, and Arg105. However, it also revealed the importance of having Gly at position 88 because it stabilizes hydrophobic characteristic of M1 for membrane binding. Glu88 creates hydrogen bond interactions with AA in close proximity The take home message is pH factor plays a vital role in structural features of M1 that are correlated to its physiological behavior.

4.7 Future plans

Viral morphology detected by cryo-electron microscopy and cryo-electron tomography have shown that pleomorphic sizes and shapes of the virus are manifestations of conformational changes of structural proteins, such as HA and M1 under acidic or neutral conditions. These changes in structural conformations of M1 supposedly could rely on C-terminal structural changes because at different pH conditions N-terminal crystal structures are relatively similar. Nevertheless, N-terminal of M1 largely contributes to M1 oligomerization and interactions. It is important, therefore, to investigate structure conformational changes of C-terminal domain.[114, 118] Our lab has computationally developed library of small molecules to destabilize M1 layer. Co-crystallization of wt-M1 with an inhibitor is highly appreciated to detect binding mode of drug candidate with M1.

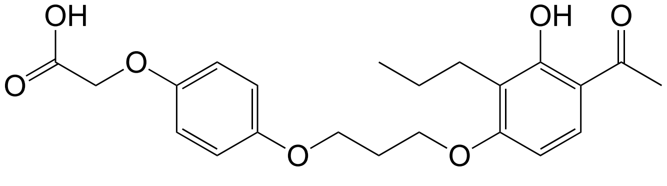
Chapter 5. Interaction of Proteins with Small Molecules using Biophysical Techniques.

5.1 Study of M1 binding towards inhibitor and its derivatives.

5.1.1 Introduction

The M1 protein is considered a promising target protein because it has an advantageous profile of low antigenic shift or mutation compared to HA and NA. However, M1 is a structural protein, which is located between viral core and envelope. Indeed, this layer is important for viral replication and integrity.[18-21] Yet, designing a drug to block M1 activity is the hardest task, because M1 lacks for well-defined binding cavity. Our lab computationally found small molecule inhibitor, called 2-(4-(3-(4-acetyl-3-hydroxy-2-propylphenoxy)propoxy)phenoxy)acetic acid (PHE), through virtual screening of a drug-like molecules library (Figure 5.1). Interestingly, virtual screening of libraries is usually performed on a single molecule of target protein. However, this is not the case for M1, therefore, screening was based on the nature forces of M1-M1 interaction aiming to destabilize M1 layer by molecule having wedge like action.[119]

We studied the affinity of M1 with PHE, M1-M1, and M1-M1-PHE interaction using biophysical techniques, SPR and MST. The results revealed a fairly high affinity of M1 with PHE as well as very high affinity of M1 for self-association. We studied same interaction using MST. Additionally, we predicted affinity constants of M1 with PHE derivatives (Figure 5.1), which are selected from ZINC docking library based on structural similarity with PHE. Then, we predicted structure-activity relationship through criterion of affinity constant.



PHE

Figure 5.1. Chemical structures of PHE.

5.1.2 Elucidating of M1 binding affinity using SPR experiments

5.1.2.1 Materials and Result

The interaction of M1-M1 in the presence and absence of PHE was studied using a Reichert SR7500DC optical biosensor. NeutrAvidin sensor chips (Reichert Technologies, Depew, NY, USA) were used for capturing biotinylated M1 and Scrubber (Version 2.0c, 2008, BioLogic Software) was used for processing the data. Two biotinylated M1 protein chips were prepared utilizing recombinant M1 protein expressing the N^{1-165}-domain of WSN/33 and a biotin-avidin immobilization strategy. One chip was used to study M1–PHE interaction. Biotinylated M1 protein (50 μM) in 10 mM HEPES buffer, pH 7.4 containing 150 mM NaCl, 3.4 mM EDTA, and 0.05% Tween 80 was injected onto NeutrAvidin sensor chip for ~40 min with a constant flow of 5 μL/min at 10 °C. An immobilization level of ~700 μRIU was achieved. For M1–PHE interaction study, PHE at varying concentrations (1.67, 3.33, 12.5, 25 and 50 μM) was injected over the biotinylated M1 protein chip at flow rate of 20 μL/min and room temperature. Association and dissociation of PHE was monitored for 1–3 min. The increase in SPR signal was proportional to the PHE concentration. The association and dissociation rates were used to calculate the affinity constant.

To study M1-M1 self-association, biotinylated M1 (10 μM) was injected over the NeutrAvidin sensor chip for ~5 min with a constant flow of 30 μL/min at 10 °C. An immobilization response of 150 μRIU was achieved. For the M1-M1 interaction study, M1 at 71, 143, 300 and 570 nM concentrations were injected as described above for M1–PHE interaction at flow rate of 30 μL/min and room temperature. Association and dissociation kinetics was monitored for over 5 min and the affinity was calculated by averaging the affinities at each concentration. Binding curves showing association and dissociation phases are displayed in Figure 5.2. Lastly, constant concentration of M1 protein was pre-mixed in samples with different PHE concentrations (75 nM, 150 nM, 300 nM, and 600 nM) for the M1-M1-PHE interaction study.[119] Dissociation rate constant of each concentration of M1-PHE, M1-M1, and M-M1-PHE interactions are depicted in Table 5.1.

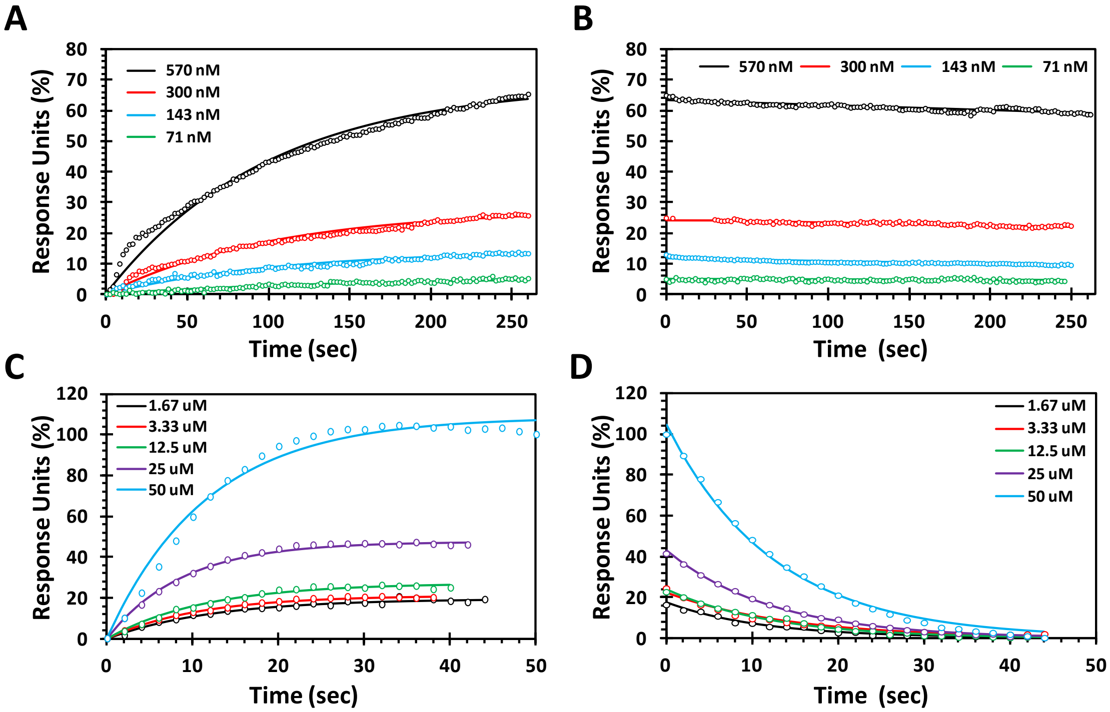


Figure 5.2. Binding curves of M1-M1 interaction (A) is the association phase, (B) is the dissociation phase. Association and dissociation of M1-PHE interaction represented by (C) and (D), respectively. (Binding curves are plotted as response unit against time in (sec)).[119]

Table 5.1. Off rate (dissociation rate constant (s^{-1})) and On rate (association rate constant ($\mu M^{-1}.s^{-1}$)) of M1-M1 interaction (yellow table), M1-PHE interaction (blue table), and M1-M1-PHE interaction (green table).

M1-M1 interaction

Conc (µM)	kOn	kOff
0.3	0.0096	0.00042
0.57	0.0100	0.00027
1.2	0.0163	0.00043
2.4	0.0686	0.00086

M1-M1-PHE interaction

Conc (µM)	kOn	kOff
0.075	0.0454	0.0292
0.15	0.0453	0.0235
0.3	0.0408	0.0318
0.6	0.0553	0.0260

M1-PHE interaction

Conc (µM)	kOn	kOff
1.67	0.0812	0.0902
3.33	0.0941	0.0702
6.67	0.103	0.0795
1.25	0.0879	0.0806
2.50	0.114	0.0809
5.00	0.0854	0.0802

Table 5.2. Equilibrium dissociation constant (K_d) in μM of M1-M1 interaction (yellow table), M1-PHE interaction (blue table), and M1-M1-PHE interaction (green table).

PHE (μM)	K_d (μM)
1.67	1.11±0.21
3.33	0.75±0.05
6.67	0.77±0.04
12.5	0.92±0.10
25	0.71±0.04
50	0.94±0.08

M1 (μM)	K_d (μM)
0.3	0.043±0.003
0.57	0.027±0.004
1.2	0.026±0.003
2.4	0.013±0.002

PHE (μM)	K_d (μM)
0.075	0.643±0.038
0.15	0.519±0.023
0.3	0.779±0.045
0.6	0.470±0.0187

5.1.3 Elucidating of M1 binding affinity using MST

5.1.3.1 Material

His-tagged N-terminal of M1 protein (0.5 mg/ml) was obtained from Origene. Buffer used in experiment was 20 nM Na phosphate, 100 mM NaCl, 0.1% PEG (8K), 0.1 mM EDTA, and pH 7.4. PHE and their derivatives were obtained from Sigma-Aldrich. Protein labeling kit His-tag labeling kit RED-tris-NTA from NanoTemper Technologies GmbH was used to label M1.

5.1.3.2 Result

PHE and their derivatives contain aromatic moiety that fluorescently interfere with M1 fluorescence. Therefore, M1 was labeled with His-Tag Labeling Kit RED-tris-NTA since M1 is a His-tagged protein. The labeling was carried out as ratio of 2:1 of protein and dye. Then, the mixture was left in dark for 30 min. The protein wasn't further purified and the sample immediately was used. Serial dilution of 16 samples of PHE (2 µM as initial highest concentration) were mixed with 50 nM of M1. Change in fluorescence revealed an affinity constant (K_d) of 1.5 µM and that is comparable to the one predicted by SPR (K_d ~1µM) as in Figure 5.3 (A). This experiment was done four times and only one showed good fluorescence signal. Irregular fluorescence signal is shown in Figure 5.4 (A).

Compound 1, 2, 3, and 4, derivatives of PHE with chemical variation of linker length, and aromaticity character, were tested for their affinity toward M1. Among them, compounds 2 and 3 failed to show consistent change in fluorescence unless bovine serum albumin (BSA) is added. Irregular MST signal of compound 2 and 3 without BSA are shown in Figure 5.4 (B and C). Compound 1, 2, 3, and 4 had shown K_d values of 37.87 μM, 24.5 μM, 6 μM, and 60 μM, respectively. Change in fluorescence signals of the four compounds, characterizing their binding curves, is shown in Figure 5.3. Table 5.3 summarizes affinity constants of the ligands. Similar to PHE, experiments of compound 1, 2, 3, and 4 were carried out three times but their fluorescence signals were irregular. One experiment of good signal of each compound is presented in this thesis. Experiment settings were set at MST power of 20% and excitation power of 20%. The standard evaluation strategy of the experiments was set up to “thermophoresis with T-Jump”, which compares initial fluorescence (F cold) to the steady state fluorescence (F hot).

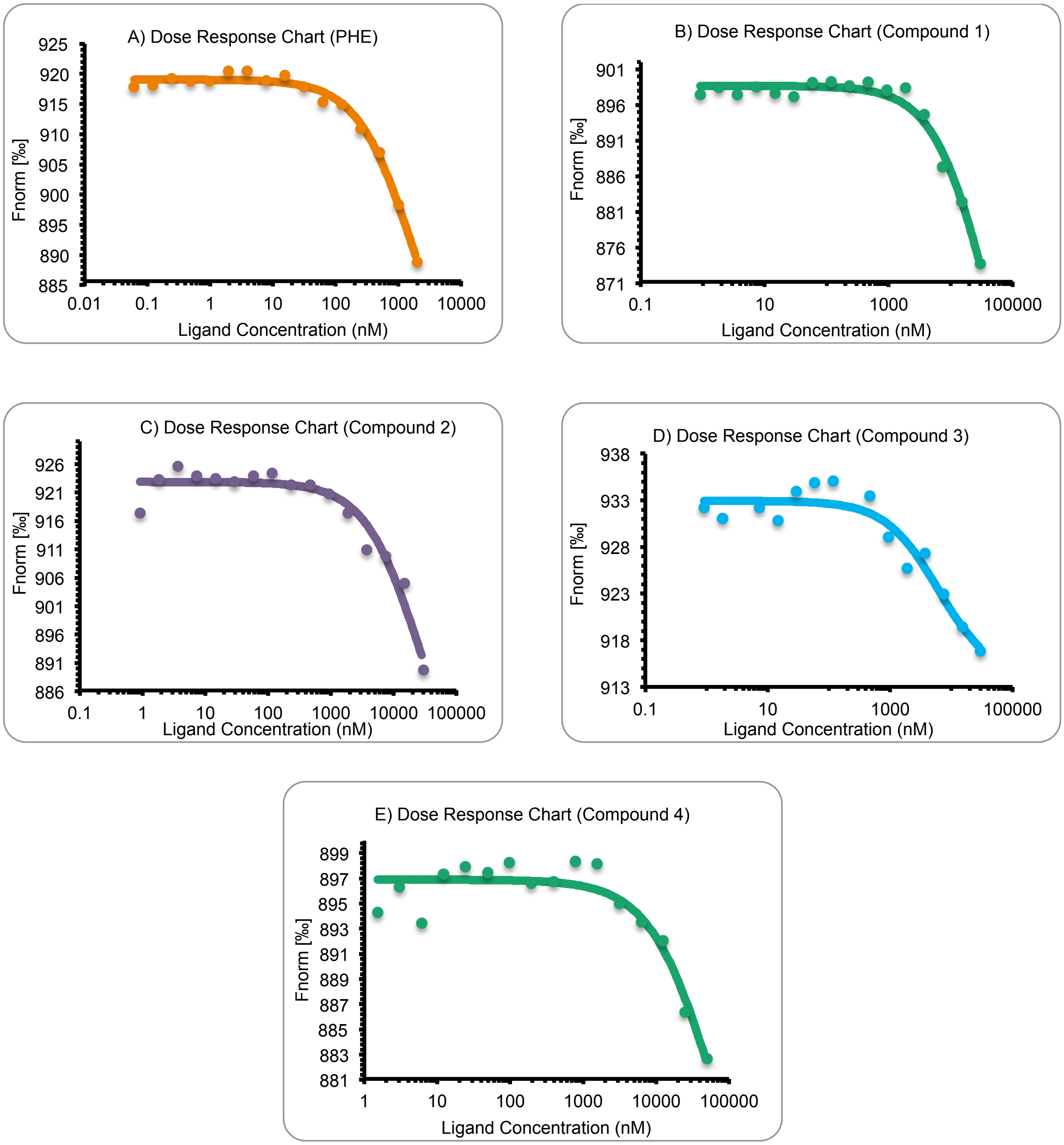


Figure 5.3. MST binding curves of wt-M1 with (A), (B), (C), (D), and (E) represent PHE, compound 1, compound 2, compound 3, and compound 4 respectively. (Binding curves are plotted as normalized fluorescence (hot fluorescence/initial fluorescence) against ligand concentration (nM)).

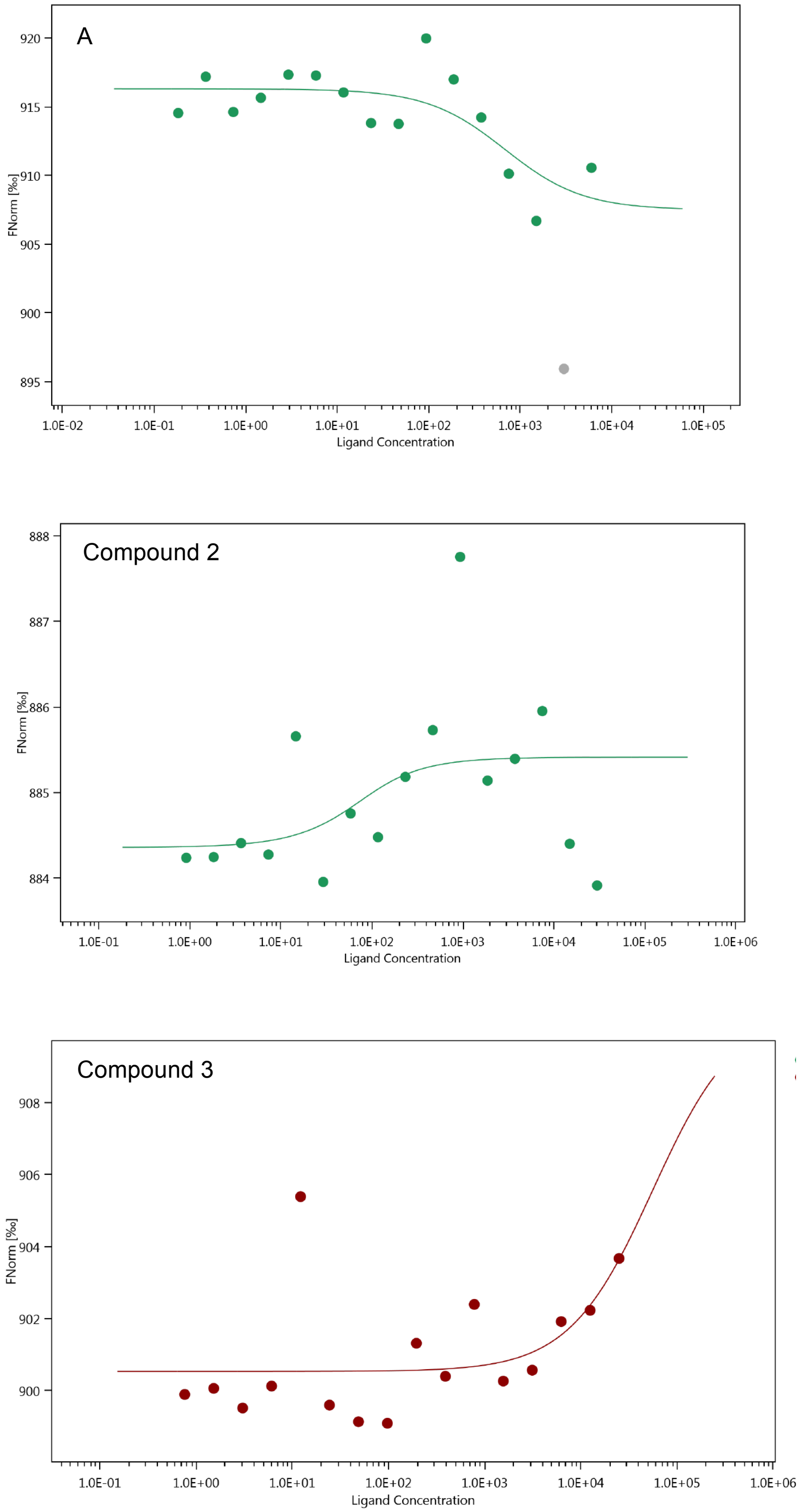


Figure 5.4. Irregular fluorescence signal of (A) PHE, (B) compound 2, and (C) compound 3. Both Compound 2 and 3 experiments were carried out without BSA.

Table 5.3. Summary of K_d values of PHE and all four derivatives with their response amplitude.

Compound	Affinity constant (K_d)	Response amplitude
PHE	1.5 μM ± 0.1	53
Compound 1	37.87 μM ± 1.23	57
Compound 2	24.5 μM ± 2.5	57
Compound 3	6 μM ± 1.8	19
Compound 4	60 μM ±1.6	32

With respect to compound (3) analysis, two data points (capillaries no. 8, and 14) were deleted because they were largely deviated. Upon eye-inspection of these capillaries, there were small bubbles inside, which probably contribute to irregular fluoresence reading as in Figure 5.5.

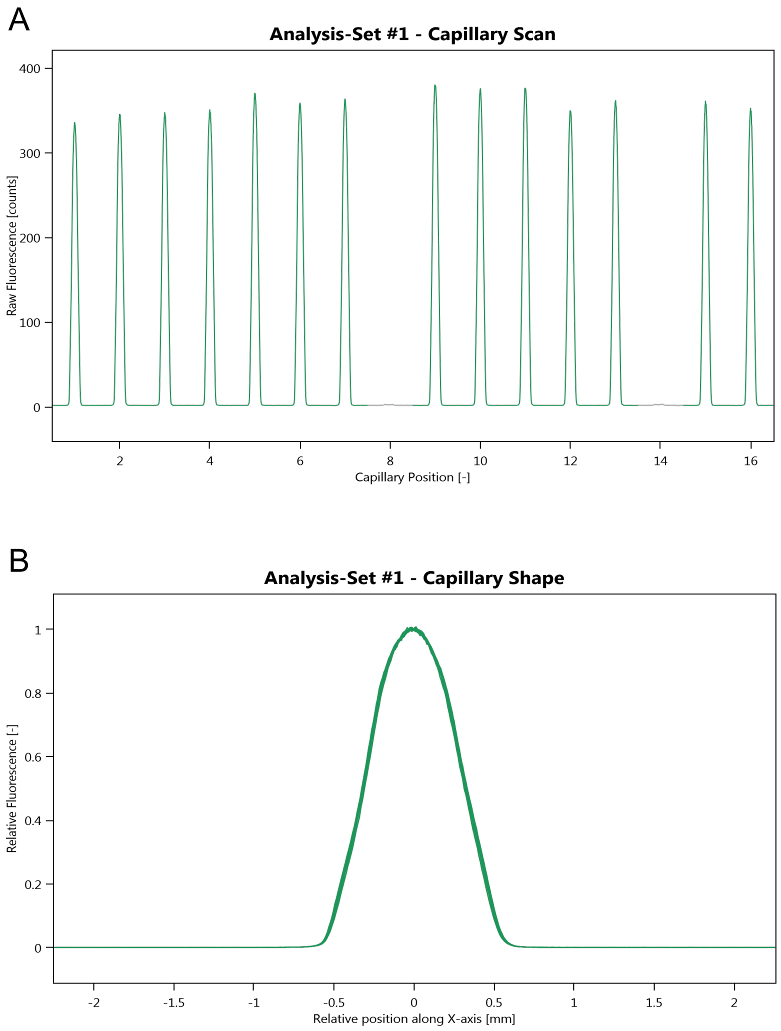


Figure 5.5. MST experimental parameters of compound 3 (A) All capillaries showed similar fluorescence magnitude except for capillaries no. 8 and 14 which were deleted. (B) There was slight adsorption on capillaries that was noticed with all experiments with or without BSA.

5.1.3.3 Discussion

Affinity of PHE for M1 is fairly high (K_d ~ 1 μM) which is confirmed by two biophysical techniques, SPR and MST. Moreover, M1-M1 interaction is so strong to have a K_d of 50 pM. Relating off-rate and on-rate of the three type of experiments in table 5.1 reveals that dissociation of M1-M1 interaction took place at faster rate in presence of PHE. PHE has acidic and hydrophobic characteristics explicated by carboxyl and aromatic groups, respectively. These chemical features are complementary to their counterparts of M1. For example, acidic character may interact with basic residues of NLS motif or its neighboring region that play a crucial rule in M1 oligomerization at neutral pH. Additionally, hydrophobic moiety of PHE could interfere with hydrophobic interactions that maintain M1-M1 association. Unexpectedly, the affinity of compound 1 that has longer linker by only a methylene group and shares almost all PHE chemical features, dramatically dropped. Similarly, compound 2, which has double bond instead on an ether group binds to M1 with K_d of 24.4 µM, while compound 3, possessing the highest affinity to M1 after PHE, has a piperidine group instead of 2-phenoxycarboxyl group. Although, compound 3 lacks for carboxyl group, it still binds to M1 similar to PHE. Finally, compound 4 poorly binds to M1 among all molecules with K_d of 60 µM. We can conclude from all compounds that linker length is extremely important for positioning two chemical groups at the end of PHE to their counterpart AA residues of M1. Furthermore, it's likely that acidic group doesn't add large difference in M1 binding comparing to hydrophobic moiety.[119]

5.2. Affinity study of Growth factors towards inhibitor G2.2.

Our body maintains its tissue growth and homeostasis through stem cells (SCs), which are found in many tissue such as, haematopoietic lineages, digestive tract, brain tissue and skin.[120] These cells are characterized by their long-term self-renewal and ability to produce different cell lineages that serve the specific function of a specific tissue. In digestive tract, for example, intestinal cells proliferate and grow from SCs that reside in a particular microenvironment called “niche”, which provides necessary factors needed for stem cells to work.[121] Also this environment limits the expansion of SC growth to localized regions in the intestine and controls its activity. Under normal physiological condition, SCs are developed into progenitor cells, which further differentiate into different cells lineages that have shorter life span, forming cellular hierarchy as shown in Figure 5.6 (a). However, it has been found that malignant tumors occur in tissue containing long life span and self-renewing SCs.[121, 122]

Colorectal cancer (CRC) is considered a third leading death in the world with 8% five-years relative survival rate.[123] Many health issues could contribute to CRC including obesity and visceral adiposity, but potential complexity of CRC is mainly due to disease metastasis and tumor reoccurrence.[124] These manifestations were poorly understood till the hypothesis of cancer stem cells (CSCs) has emerged.[125] Primarily, there are two models for tumor propagation, clonal evalution model and CSCs model, described in Figure 5.6 (b, d). Normal SCs or progenitor cells could be developed into CSCs, which are responsible for initiating and maintaining tumor growth. Moreover, CSCs are distinctive from other tumor cells by their renewal capacity as in Figure 5.6 (c, d). In other words, the role of CSCs in the growth of cancer tissue is analogous to that of normal SCs in maintaining growth of body tissue whenever there is a need.[126]

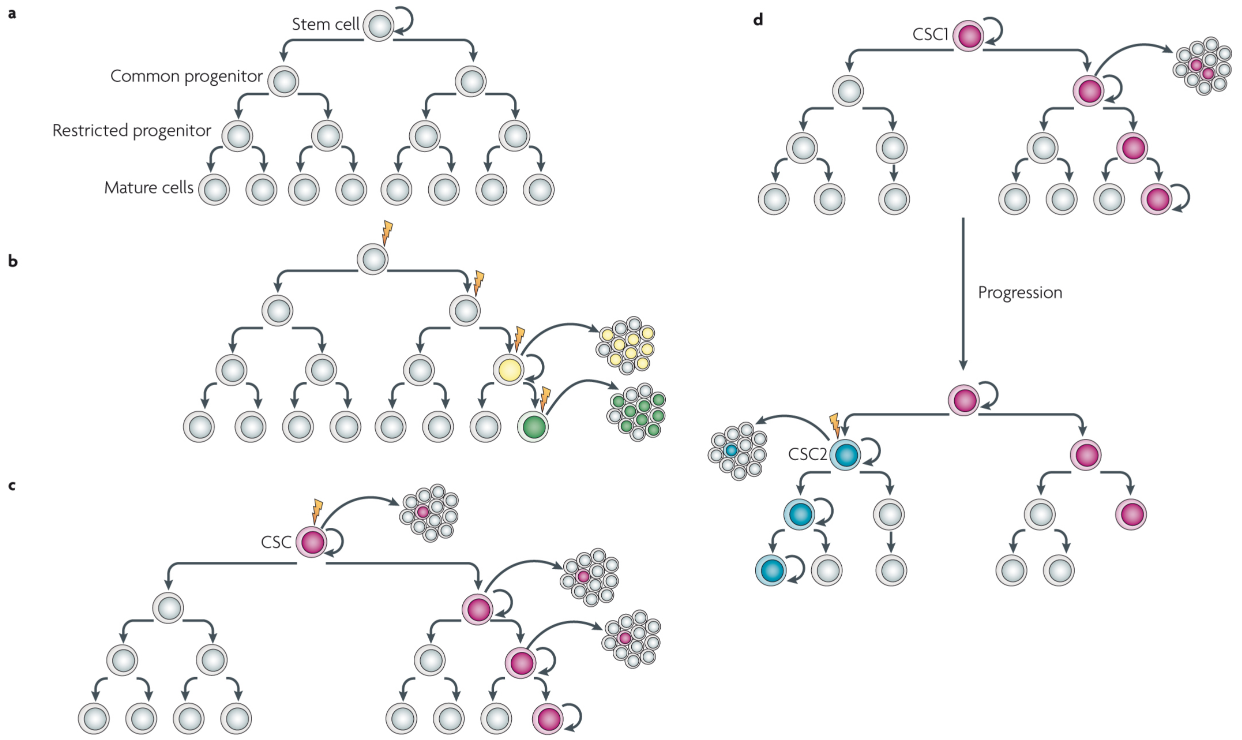


Figure 5.6. Models of tumor propagation. (a) is the normal hierarchy of cell growth in which stem cells produce progenitor cells that eventually differentiate into mature cells, (b) is the clonal evolution model where all cells equally tend to acquire cancerous activity. (c) CSCs model is hierarchal model, in which only CSCs develop tumor. (d) Both models may contribute to tumorigenesis. First, a specific CSC (CSC1) will cause tumor growth. As tumor progressed, another distinct CSC (CSC2) may appear due to clonal evolution of CSC1, resulted from the acquisition of an additional mutation or epigenetic modification. This more aggressive CSC2 becomes dominant and initiates tumor formation.[126] Taken from Visvader, et al. *Nat. Rev. Cancer* **2008**, *8*, 755–768.

Failure of cancer therapy and disease relapses are most likely due to CSCs, which are resistance to chemotherapy by the virtue of their mutation.[122] Consequently, CSCs could be an attractive paradigm for cancer therapy through targeting biological molecules that facilitate CSCs cellular activity.[126] Treatment should be highly selective to cancerous cells without affecting other healthy cells. It also should be potent to a degree that ensures no potential cancerous cells left behind, which might lead to disease recurrence and metastasis.[123]

Both SCs and CSCs are regulated through signaling molecules through different signaling pathways that mediate cell proliferation and differentiation. Some of signaling pathways are more active than the other, which are identified by elevation of their corresponding molecules, such as fibroblast growth factor (FGF), epidermal growth factor (EGF), bone morphogenetic protein 4 (BMP4), wingless-int (Wnt), and transforming growth factor-β (TGF β). These growth factors are secreted from epithelial cells. These cells are associated always with SCs and CSCs in their niches. Therefore, SCs and CSCs can get vital soluble signaling molecules in extremely closed proximity to empower their communication and replication in a very short time.[120-123, 127]

EGF and FGF are important for cells purification and renewal. One study had shown that EGF promotes tumor sphere formation in dose dependent manner,[128] while FGF promotes angiogenesis and maintenance of cancerous cells.[129] On the other hand, TGFβ enhances the production of IL11 from stromal cells, which is a megakaryopoietic cytokine that stimulates platelet production.[130] Platelets are a rich source of TGFβ, which has been found to be hiding circulating tumor from immune response.[131] Thus, TGFβ enhances tumor metastasis. In one study, it has been shown that inhibition of TGFβ pathways leads to cell differentiation, which is a treatment strategy to enhance tumor regression.[132] Other important signaling molecule is Wnt, which is an essential molecule in normal intestinal SCs. However, in CRC it gets mutated leading to active Wnt isoform that causes benign polyps and CSCs renewal and proliferation.[133, 134] Unlike previously mentioned growth factors, activation of BMP pathway leads to suppression of tumor growth. It has been found that when BMP pathway is disturbed, it results in development of adenomas.[135]

These growth factors communicate through glycosaminoglycans (GAGs) and heparan sulfate (HS).[136] Our lab had devised a unique class of compounds, called sulfated monosaccharide GAG mimetic (NSGM). Of this class, one compound, called G2.2, has shown anti-cancer activity by inhibiting a number of growth factors (Figure 5.7).[137] Interestingly, it is a selective inhibitor of colon CSCs. Unlike available chemotherapy, G2.2 had exclusively displayed a great safety profile in that mice could tolerate a dose of G2.2 up to 400 mg/kg (unpublished data). We believe that G2.2 acts by cutting off cellular communicating pipeline of CSCs; hence, preventing their activity and limiting tumor propagation. We are interested in elucidating affinity of G2.2 to a library of various proteins that might be involved in CRC.

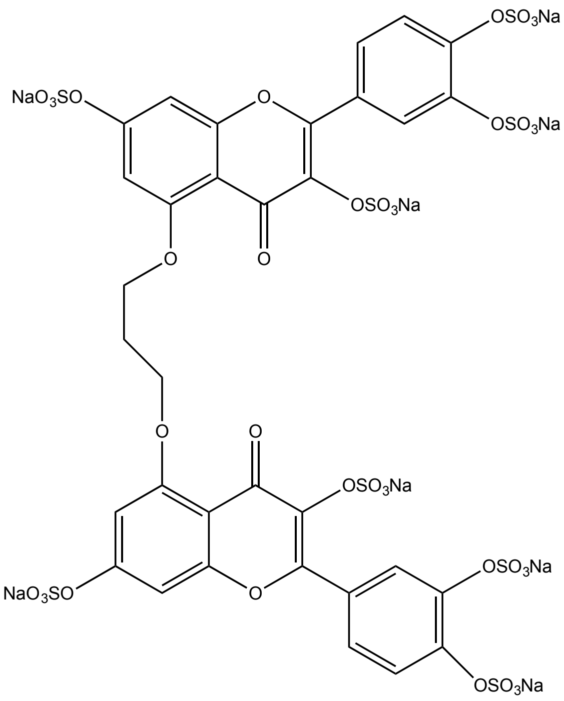


Figure 5.7. Chemical structure of G2.2.

5.2.1 Material

TGFβ2 protein sample of 10 µg (EMD Millipore) and EGF of 100 µg (EMD Millipore) were constituted with the same buffer used in experiment (20 nM Na phosphate, 100 mM NaCl, 0.1% PEG (8K), 0.1 mM EDTA, pH 7.4). G2.2 was used with a concentration of 1 mM being the highest concentration used in serial dilution of 12 capillaries. Protein labeling kit RED-MALEIMIDE (cysteine reactive) from NanoTemper Technologies GmbH was used to label EGF and TGFβ2.

5.2.2 Result

Since G2.2 emits fluorescence simultaneously with protein fluorescence emission range, extrinsic fluorescence was needed. Labeling of EGF and TGFβ2 was done with maleimide labeling through free thiol group, such as reduced cysteine (Cys) residues as in Figure 5.8.

Figure 5.8. Schematic representation displays the site of thiol coupling with maleimide functional group that attaches to red dye absorbing near red region in visible light wavelength.[138]

Importantly, EGF and TGFβ2 structures were analyzed for availability of free Cys residues, which have no interference with G2.2 binding site. A 20 μM of EGF and 4 μM TGFβ2 of 100 uL were separately mixed with cysteine red dye of concentration 43 μM of 100 uL and left to react for 30 min in dark. Then, labeled EGF and TGFβ2 were purified from other reaction components by gravity gel filtration. Small fractions of 200 μL were collected and UV/VIS absorbance of protein as well as dye was determined at 280 nm and 640 nm respectively. Protein concentration of EGF was 0.8 μM and degree of labeling was 0.1, while the concentration of TGFβ2 was 50 nM and degree of labeling was 0.2. Affinity constants of both EGF/TGFβ2 with G2.2 binding were determined using automated MST instrument from NanoTemper Technologies GmbH. Figure 5.9 displays binding curves of EGF and TGFβ2.

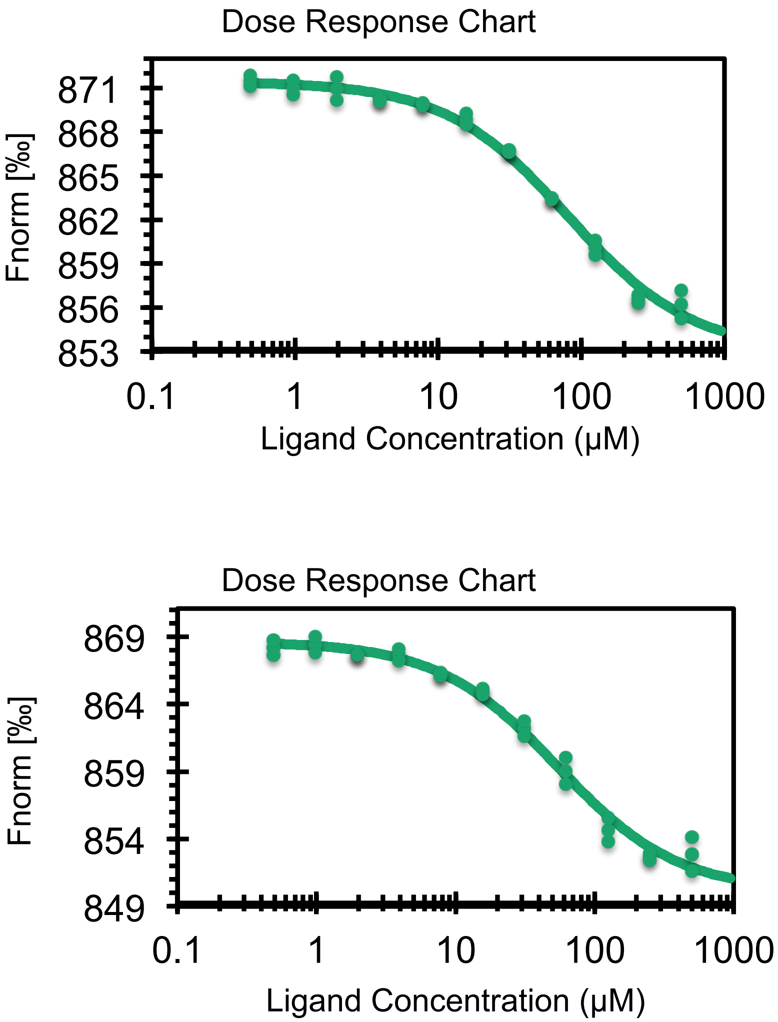


Figure 5.9. Binding curves of (A) EGF and (B) TGFβ2 with G2.2. (Fluorescence change is plotted against G2.2 (µM)).

The experiments were performed in duplicate. Experiment settings were adjusted to MST power of 20% and excitation power of 20%. Thermophoresis with T-Jump was chosen as the standard evaluation strategy of the experiments, which compares initial fluorescence to the steady state fluorescence as in Figure 5.10. There are some important experimental factors that have to be checked to ensure accuracy and avoid experiment caveats, such as capillaries adsorption.

In Figure 5.11, capillary shapes of all samples are overlapped nicely indicating no adsorption of protein on capillaries. Change in fluorescence magnitude was 21% and equilibrium dissociation constant were calculated as 80 μM and 54 μM for EGF and TGFβ2, respectively.

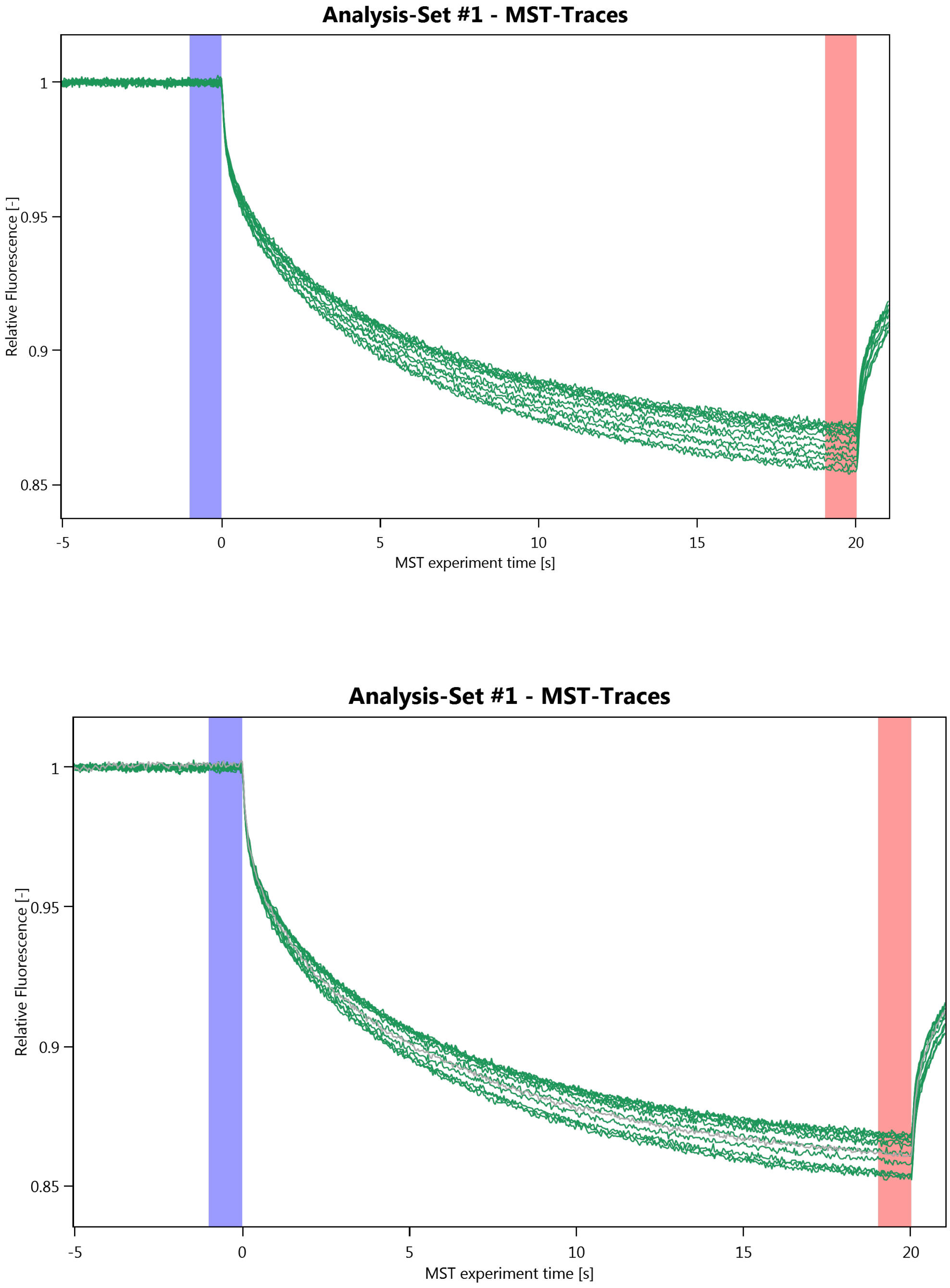


Figure 5.10. MST traces of (A) EGF and (B) TGFβ2 upon G2.2 binding. Blue bar represents initial fluorescence and red bar represents hot fluorescence.

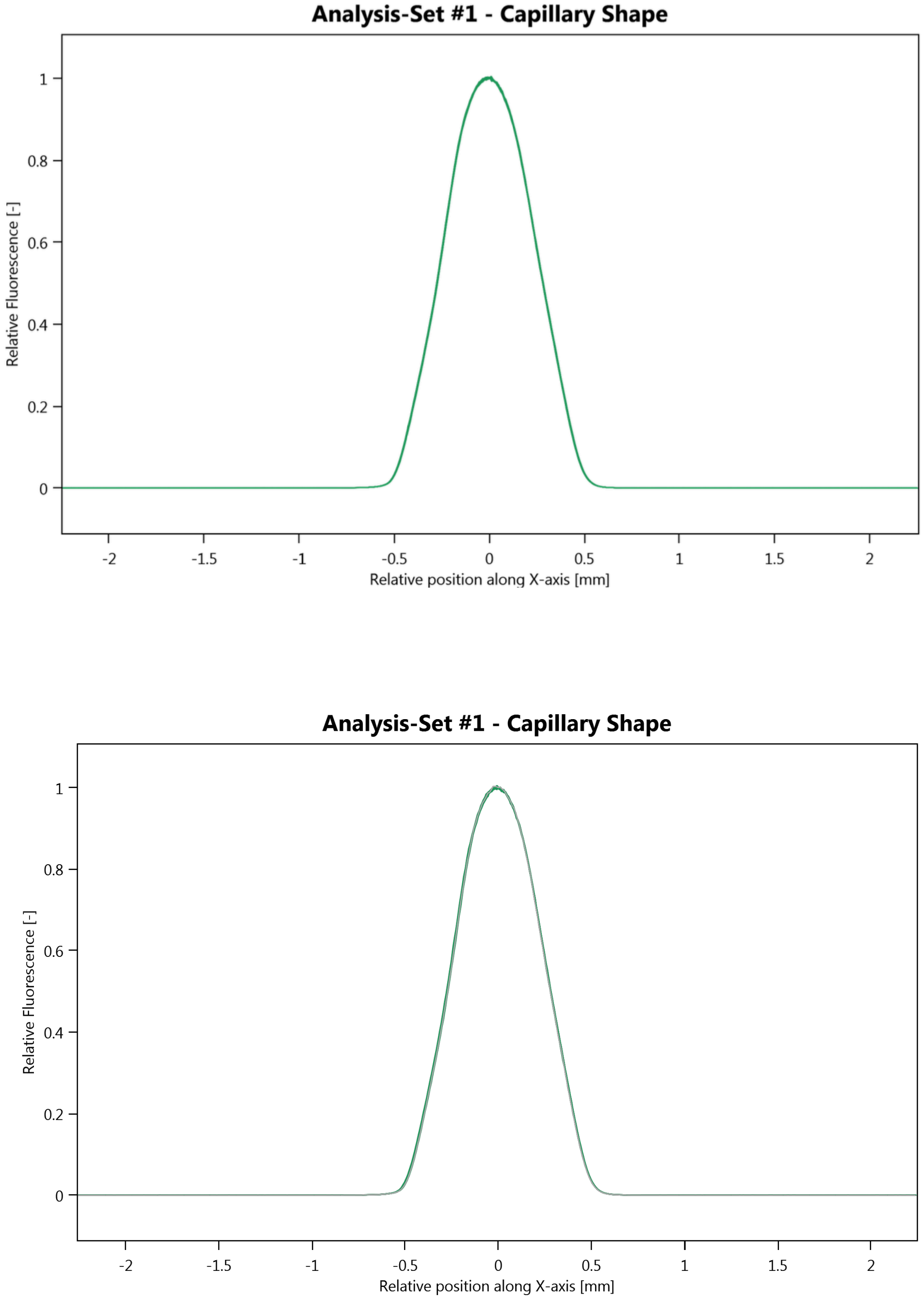


Figure 5.11. Overlapping of MST capillaries shape of (A) EGF and (B) TGFβ2 upon G2.2 binding, which show no adsorption on capillaries since all capillaries are nicely overlapped.

5.2.3 Discussion

G2.2 showed low affinity towards EGF and TGFβ2 compared to other growth factors such as FGF and IGF in which G2.2 binds with low K_d values (~1.8-50 μM) (unpublished data). Our computational studies initially indicated that G2.2 binds with low affinity toward EGF and TGFβ2, which was confirmed using MST. In the same study, G2.2 could have high affinity toward BMP4 and Wnt signaling molecules suggesting that G.2.2 targets key communicating proteins that are selective to CRC pathogenesis. We believe G2.2 occupies the same binding pocket of HS or GAGs that mediate signal transduction in aforementioned signaling pathways. G2.2 could interfere with HS or GAGs binding through electrostatic interaction of its sulfate groups with basic residues of signaling molecules along with hydrophobic interactions of its aromatic moieties with hydrophobic residues. Therefore, G2.2 targets multiple critical signaling pathways rather than single cellular route promoting its therapeutic value and synergistic effect. This is extremely important as therapeutic strategy in diseases prone to have cross-link between many cellular pathways, such as cancer.

5.2.4 Conclusion

PHE and G2.2 are promising small molecules that could be optimized into drug candidate for treatment of influenza infection and CRC, respectively. PHE had displayed sufficient inhibitory effect on viral replication in TEM studies and high affinity to M1 under neutral pH conditions. On the other hand, G2.2 seems to be a promising drug therapy that target CRC through inhibiting growth factors and cytokines associated with

CSCs activity. It has a high selectivity and safety profile, which enables it to compete with current CRC drug treatment. However, it has low affinity to EGF and TGFβ2. As future direction, it is quite important to investigate other protein binding affinity with G2.2, such as BMP4 and Wnt.

Reference:

1. Reid, A. H.; Taubenberger, J. K.; Fanning T. G. Evidence of an absence: The genetic origins of the 1918 pandemic influenza virus. *Nat. Rev. Microbiol*. **2004**, *2*, 909–914.
2. Shope, R. E. Influenza: History, epidemiology, and speculation: The R. E. Dyer lecture. *Public Health Rep*. **1958**, *73*, 165–178.
3. World Health Organization FluNet Summary Home Page. http://www.who.int/influenza/gisrs_laboratory/updates/summaryreport/en/ (accessed on Aug 03, 2016).
4. Barbara, J. Viruses. *Vox Sang*. **2004**, 87, S95–S97.
5. Goering, R. V.; Dockrell, H. M.; Zuckerman, M.; Chiodini, P. L.; Roitt, I. M. Mim's Medical Microbiology; Elsevier: China, 2013.
6. Mateu, M. G. Structure and Physics of Viruses an Integrated Textbook; Springer: Netherlands, 2013.
7. Eisfeld, A. J.; Neumann, G.; Kawaoka, Y. At the centre: Influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol*. **2015**, *13*, 28–41.
8. Centers for Disease Control and Prevention Home Page. http://www.cdc.gov.htm (accessed on Aug 03, 2016).
9. Greber, U. F. How cells tune viral mechanics—insights from biophysical measurements of influenza virus. *Biophys. J*. **2014**, *106*, 2317–2321.
10. Steinhauer, D. A. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* **1999**, *258*, 1–20.
11. Sriwilaijaroen, N.; Suzuli, Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci*. **2012**, *88*, 226–249.
12. Kim, J. H.; Resende, R.; Wennekes, T.; Chen, H. M.; Bance, N.; Buchini, S.; Watts, A. G.; Pilling, P.; Streltsov, V. A.; Petric, M.; Liggins, R.; Barrett, S.; McKimm-Breschkin, J. L.; Niikura, M.; Withers, S. G. Mechanism-based covalent neuraminidase inhibitors with broad-spectrum influenza antiviral activity. *Science* **2013**, *340*, 71–75.
13. Gamblin, S. J.; Skehel, J. J. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J. Biol. Chem*. **2010**, *285*, 28403–28409.
14. Rossman, J. S.; Jing, X.; Leser, G. P.; Lamb, R. A. Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* **2010**, *142*, 902–913.
15. Boulo, S.; Akarsu, H.; Ruigrok, R. W. H.; Baudin, F. Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. *Virus Res*. **2007**, *124*, 12–21.
16. Wang, J.; Qiu, J. X.; Soto, C.; DeGrado, W. F. Structural and dynamic mechanisms for the function and inhibition of the M2 proton channel from influenza A virus. *Curr. Opin. Struct. Biol*. **2011**, *21*, 68–80.
17. Gannagé, M.; Dormann, D.; Albrecht, R.; Dengjel, J.; Torossi, T.; Rämer, P. C.; Lee, M.; Strowig, T.; Arrey, F.; Conenello, G.; Pypaert, M.; Andersen, J.; García-Sastre, A.; Münz, C. Matrix protein 2 of influenza A virus blocks autophagosome fusion with lysosomes. *Cell Host Microbe* **2009**, *6*, 367–380.
18. Centers for Disease Control and Prevention: Influenza (FLU) Home Page. http://www.cdc.gov/flu/about/viruses/types.htm (accessed on Aug 06, 2016)

19. Huang, X.; Ye, Z. Evaluating the vaccine potential of an influenza A viral hemagglutinin and matrix double insertion DNA plasmid. *Vaccine* **2007**, *25*, 7649–7655.
20. Baudin, F.; Petit, I.; Weissenhorn, W.; Ruigrok, R. W. H. In vitro dissection of the membrane and RNP binding activities of influenza virus M1 protein. *Virology* **2001**, *281*, 102–108.
21. Ito, T.; Gorman, O. T.; Kawaoka, Y.; Bean, W. J.; Webster, R. G. Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins. *J. Virol*. **1991**, *65*, 5491–5498.
22. Sha, B.; Luo, M. Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nat. Struct. Mol. Biol*. **1997**, *4*, 239–244.
23. Ye, Z.; Robinson, D.; Wagner, R. R. Nucleus-targeting domain of the matrix protein (M1) of influenza virus. *J. Virol*. **1995**, *69*, 1964–1970.
24. Martin, K.; Heleniust, A. Nuclear transport of influenza virus ribonucleoproteins: The viral matrix protein (M1) promotes export and inhibits import. *Cell* **1991**, *67*, 117–130.
25. Akarsu, H. Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). *EMBO J*. **2003**, *22*, 4646–4655.
26. Chen, B. J.; Leser, G. P.; Jackson, D.; Lamb, R. A. The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *J. Virol*. **2008**, *82*, 10059–10070.
27. Pielak, R. M.; Chou, J. J. Flu channel drug resistance: a tale of two sites. Protein *Cell* **2010**, *1*, 246–258.
28. Ali, A.; Avalos, R. T.; Ponimaskin, E.; Nayak, D. P. Influenza Virus Assembly: Effect of Influenza Virus Glycoproteins on the Membrane Association of M1 Protein. *J. Virol*. **2000**, *74*, 8709–8719.
29. Bourmakina, S. V.; García-Sastre, A. The morphology and composition of influenza A virus particles are not affected by low levels of M1 and M2 proteins in infected cells. *J. Virol*. **2005**, *79*, 7926–7932.
30. Burleigh, L. M.; Calder, L. J.; Skehel, J. J.; Steinhauer, D. A. Influenza A Viruses with Mutations in the M1 Helix Six Domain Display a Wide Variety of Morphological Phenotypes. *J. Virol*. **2005**, *79*, 1262–1270.
31. Sha, B.; Luo, M. Crystallization and preliminary X-ray crystallographic studies of type A influenza virus matrix protein M1. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53*, 458–460.
32. Das, S. C.; Watanabe, S.; Hatta, M.; Noda, T.; Neumann, G.; Ozawa, M.; Kawaoka, Y. The highly conserved arginine residues at positions 76 through 78 of influenza A virus matrix protein M1 play an important role in viral replication by affecting the intracellular localization of M1. *J. Virol*. **2012**, *86*, 1522–1530.
33. Nayak, D. P.; Balogun, R. A.; Yamada, H.; Zhou, Z. H.; Barman, S. Influenza virus morphogenesis and budding. *Virus Res*. **2009**, *143*, 147–161.
34. Liu, T.; Ye, Z. Restriction of Viral Replication by Mutation of the Influenza Virus Matrix Protein. *J. Virol*. **2002**, *76*, 13055–13061.
35. Liu, T.; Ye, Z. Attenuating Mutations of the Matrix Gene of Influenza A/WSN/33 Virus. *J. Virol*. **2005**, *79*, 1918–1923.

36. Xie, H.; Liu, T. M.; Lu, X.; Wu, Z.; Belser, J. A.; Katz, J. M.; Tumpey, T. M.; Ye, Z. A live attenuated H1N1 M1 mutant provides broad cross-protection against influenza A viruses, including highly pathogenic A/Vietnam/1203/2004, in mice. *J. Infect. Dis*. **2009**, *200*, 1874–1883.
37. Liu, T.; Ye, Z. Introduction of a temperature-sensitive phenotype into influenza A/WSN/33 virus by altering the basic amino acid domain of influenza virus matrix protein. *J. Virol*. **2004**, *78*, 9585–9591.
38. Xie, H.; Lin, Z.; Mosier, P. D.; Desai, U. R.; Gao, Y. The compensatory G88R change is essential in restoring the normal functions of influenza A/WSN/33 virus matrix protein 1 with a disrupted nuclear localization signal. *J. Virol*. **2013**, *87*, 345–353.
39. Arzt, S.; Baudin, F.; Barge, A.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W. H. Combined results from solution studies on intact influenza virus M1 protein and from a new crystal form of its N-terminal domain show that M1 is an elongated monomer. *Virology* **2001**, *279*, 439–446.
40. Safo, M. K.; Musayev, F. N.; Mosier, P. D.; Zhou, Q.; Xie, H.; Desai, U. R. Crystal structures of influenza A virus matrix protein M1: Variations on a theme. *PLoS ONE* **2014**, *9*, e109510.
41. Nobel Prices and Laureates: Wilhelm Conrad Röntgen – Biographical Home Page. http://www.nobelprize.org/nobel_prizes/physics/laureates/1901/rontgen-bio.html (accessed Aug 01, 2016).
42. Hauptman, H. A. History of x-ray crystallography. *Struct. Chem*. **1990**, *6*, 617–620.
43. Nobel Prices and Laureates: Max von Laue – Biographical Home Page. http://www.nobelprize.org/nobel_prizes/physics/laureates/1914/laue-bio.html (accessed on Aug 01, 2016).
44. Nobel Prices and Laureates: John C. Kendrew – Facts Home Page. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/kendrew-facts.html Accessed on 08/01/2016.
45. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R.G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. Resolution. *Nature* **1960**, *185*, 422-427.
46. Meurig, T. J. The birth of x-ray crystallography. *Nature* **2012**, *491*, 186–187.
47. Sharff, A.; Jhoti, H. High-throughput crystallography to enhance drug discovery. *Curr. Opin. Chem. Biol*. **2003**, *7*, 340–345.
48. Blundell, T. L.; Patel, S. High-throughput x-ray crystallography for drug discovery. *Curr. Opin. Pharmacol*. **2004**, *4*, 490–496.
49. Goodwill, K. E.; Tennant, M. G.; Stevens, R. C. High-throughput x-ray crystallography for structure-based drug design. *Drug Discov. Today* **2001**, *6*, 113–118.
50. Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov*. **2002**, 1, 45–54.
51. Caliandro, R.; Belviso, D. B.; Aresta, B. M.; Candia, M.; Altomare, C. D. Protein crystallography and fragment-based drug design. *Future Med. Chem*. **2013**, *5*, 1121–1140.

52. Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discov. Today* **2002**, *7*, 522–527.
53. Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol*. **2002**, *9*, 268–272.
54. Stewart, L.; Clark, R.; Behnke, C. High-throughput crystallization and structure determination in drug discovery. *Drug Discov. Today* **2002**, 7, 187–196.
55. O'Reilly, M.; Vinkovic´, M.; Sharff, A.; Jhoti, H. High throughput protein crystallography: Developments in crystallisation, data collection and data processing. *Drug Discov. Today* **2006**, 3, 451–456.
56. Joachimiak, A. High-throughput crystallography for structural genomics. *Curr. Opin. Chem. Biol*. **2009**, *7*, 573–584.
57. Manjasetty, B. A.; Turnbull, A. P.; Panjikar, S.; Büssow, K.; Chance, M. R. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* **2008**, *8*, 612–625.
58. Garman, E. F. Developments in x-ray crystallographic structure determination of biological macromolecules. *Science* **2014**, *343*, 1102–1108.
59. Rupp, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology; Garland science: New York, 2010.
60. González, A. X-ray Crystallography: Data collection strategies and resources. In Comprehensive Biophysics; 1st ed.; Egelman, E. H., Eds.; Elsevier B.V.: New York, 2012; Vol. 1, pp 64–91.
61. Howard, J. A.; Probert, M. R. Cutting-edge techniques used for the structural investigation of single crystals. *Science* **2014**, *343*, 1098–1102.
62. Sayers, Z.; Avşar, B.; Cholak, E.; Karmous, I. Application of the advanced x-ray methods in life sciences. *Biochim. Biophys. Acta* [Online] **2016**, http://www.sciencedirect.com/science/article/pii/S0304416516301489 (accessed July 28, 2016).
63. Wasserman, S. R.; Koss, J. W.; Sojitra, S. T.; Morisco, L. L.; Burley, S. K. Rapid-access, high-throughput synchrotron crystallography for drug discovery. *Trends Pharmacol. Sci*. **2012**, *33*, 261–267.
64. Moraes, I.; Evans, G.; Sanchez-Weatherby, J.; Newstead, S.; Stewart, P. D. Membrane protein structure determination—The next generation. *Biochim. Biophys. Acta* **2013**, *1838*, 78–87.
65. Abola, E. E.; Cherezov, V.; Stevens, R. C. X-ray Crystallography: Crystallization. In Comprehensive Biophysics; 1st ed.; Egelman, E. H., Eds.; Elsevier B.V.: New York, 2012; Vol. 1, pp 34–63.
66. Krauss, I. R.; Merlino, A.; Vergara, A.; Sica, F. An overview of biological macromolecule crystallization. *Int. J. Mol. Sci.* **2013**, *14*, 11643–11691.
67. Benvenuti, M.; Mangani, S. Crystallization of soluble proteins in vapor diffusion for x- ray crystallography. *Nat. Protoc.* **2007**, *2*, 1633–1651.
68. Deller, M. C.; Rupp, B. Approaches to automated protein crystal harvesting. *Acta Crystallogr. F Struct. Biol. Commun*. **2014**, *70*, 133–155.

69. Viola, R.; Carman, P.; Walsh, J.; Frankel, D.; Rupp, B. Automated robotic harvesting of protein crystals— addressing a critical bottleneck or instrumentation overkill? *J. Struct. Funct. Genomics* **2007**, *8*, 145–152.
70. Hui, R.; Edwards, A. High-throughput protein crystallization. *J. Struct. Biol*. **2003**, *142*, 154–161.
71. Chayen, N. E. High-throughput protein crystallization. *Adv. Protein Chem. Struct. Biol*. **2009**, *77*, 1–22.
72. Smyth, M. S.; Martin, J. H. X ray crystallography. *J. Clin. Pathol: Mol. Pathol*. **2000**, *53*, 8–14.
73. Hammond, C. The Basics of Crystallography and Diffraction; Oxford University Press Inc.: New York, 2009.
74. Rhodes, G. Crystallography Made Crystal Clear: A Guide Users of Macromolecular Models; Academic Press: Orlando, 2000.
75. OOi, L. Principles of X-ray Crystallography; Oxford University Press Inc.: New York, 2010.
76. Ladd, M.; Palmer, R. Structure Determination by X-ray Crystallography; Springer: New York, 2013.
77. Andrew, B. Why do we trust x-ray crystallography? *Resonance* **2014**, *19*, 1087–1092.
78. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991**, *254*, 51–58.
79. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J.; IUCr Phaser crystallographic software. *J. Appl. Crystallogr*. **2007**, *40*, 658–674.
80. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
81. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr*. **2010**, *66*, 486–501.
82. Brünger, A. T. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472–475.
83. Stevens, R. C. High-throughput protein crystallization. *Curr. Opin. Struct. Biol.* **2000**, *10*, 558–563.
84. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol*. **1963**, *7*, 95–99.
85. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Comms*. **2010**, *1*, 100.
86. Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013**, *59*, 301–315.
87. Khavrutskii, L.; Yeh, J.; Timofeeva, O.; Tarasov, S. G.; Pritt, S.; Stefanisko, K.; Tarasova, N. Protein purification-free method of binding affinity determination by microscale thermophoresis. *J. Vis. Exp*. **2013**, *78*, e50541.

88. Platten, J. K. The Soret Effect: A Review of Recent Experimental Results. *J. Appl. Mech*. **2006**, *73*, 5–15.
89. Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D. MicroScale thermophoresis: Interaction analysis and beyond. *J. Mol. Struct*. **2014**, *1077*, 101–113.
90. Piazza, R. Thermophoresis: Moving particles with thermal gradients. *Soft Matter* **2008**, *4*, 1740–1744.
91. Duhr, S.; Braun, D. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19678–19682.
92. Wang, X.; Zhan, S.; Huang, Z.; Hong, X. Review: Advances and applications of surface plasmon resonance biosensing instrumentation. *Instrum. Sci. Technol.* **2013**, *41*, 574–607.
93. Hahnefeld, C.; Drewianka, S.; Herberg, F. W. Determination of kinetic data using surface plasmon resonance biosensors. Methods Mol. Med. 2004, 94, 299–320.
94. McDonnell, J. M. Surface plasmon resonance: Towards an understanding of the mechanisms of biological molecular recognition. *Curr. Opin. Chem. Biol*. **2001**, *5*, 572–577.
95. Lahiri, J.; Isaacs, L.; Tien, J.; Whitesides, G. M. A strategy for the generation of surfaces presenting ligands for studies of binding based on an active ester as a common reactive intermediate: A surface plasmon resonance study*. Anal. Chem*. **1999**, *71*, 777–790.
96. Rich, R. Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotechnol*. **2000**, *11*, 54–61.
97. Ciminelli, C.; Campanella, C. M.; Dell'Olio, F.; Campanella, C. E.; Armenise, M. N. Label-free optical resonant sensors for biochemical applications. *Prog. Quant. Electron*. **2013**, *37*, 51–107.
98. Fan, X.; White, I. M.; Shopova, S. I.; Zhu, H.; Suter, J. D.; Sun, Y. Sensitive optical biosensors for unlabeled targets: A review. *Anal. Chim. Acta* **2008**, *620*, 8–26.
99. Wilchek, M.; Bayer, E. A.; Livnah, O. Essentials of biorecognition: The (strept)avidin–biotin system as a model for protein–protein and protein–ligand interaction. *Immunol. Lett*. **2006**, *103*, 27–32.
100. Harrison, S. C. Viral membrane fusion. Nat. Struct. Mol. Biol. 2008, 15, 690– 698.
101. Rossman, J. S.; Lamb, R. A. Influenza virus assembly and budding. *Virology* **2011**, *411*, 229–236.
102. Tsfasman, T.; Kost, V.; Markushin, S.; Lotte, V.; Koptiaeva, I.; Bogacheva, E.; Baratova, L.; Radyukhin, V. Amphipathic alpha-helices and putative cholesterol binding domains of the influenza virus matrix M1 protein are crucial for virion structure organisation. *Virus Res*. **2015**, *210*, 114–118.
103. Li, S.; Sieben, C.; Ludwig, K.; Höfer, C. T.; Chiantia, S.; Herrmann, A.; Eghiaian, F.; Schaap, I. A. T. pH-controlled two-step uncoating of influenza virus. *Biophys. J*. **2014**, *106*, 1447–1456.
104. Hoffmann, E.; Krauss, S.; Perez, D.; Webby, R.; Webster, R. G. Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine* **2002**, *20*, 3165–3170.

105. Shtykova, E. V.; Baratova, L. A.; Fedorova, N. V.; Radyukhin, V. A.; Ksenofontov, A. L.; Volkov, V. V.; Shishkov, A. V.; Dolgov, A. A.; Shilova, L. A.; Batishchev, O. V.; Jeffries, C. M.; Svergun, D. I. Structural analysis of influenza A virus matrix protein M1 and its self-assemblies at low pH. *PLoS ONE* **2013**, *8*, e82431.
106. Calder, L. J.; Wasilewski, S.; Berriman, J. A.; Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10685–10690.
107. Ksenofontov, A. L.; Dobrov, E. N.; Fedorova, N. V.; Radyukhin, V. A.; Badun, G. A.; Arutyunyan, A. M.; Bogacheva, E. N.; Baratova, L. A. Intrinsically unstructured regions in the C domain of the influenza virus M1 protein. *Mol. Biol*. **2011**, *45*, 634–640.
108. Linding R., Jensen L.J., Diella F., Bork P., Gibson T.J., Russell R.B. Protein disorder prediction: Implications for structural proteomics. *Structure* **2003**, *11*, 1316– 1317.
109. Osztanyi Z., Csizmok V.P., Tompa P., Simon I. IUPred: Web server for the prediction intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434.
110. Harris, A.; Forouhar, F.; Qiu, S.; Sha, B.; Luo, M. The crystal structure of the influenza matrix protein M1 at neutral pH: M1–M1 protein interfaces can rotate in the oligomeric structures of M1. *Virology* **2001**, *289*, 34–44.
111. Lee, B.; Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol*. **1971**, *55*, 379IN3–400IN4.
112. Shishkov, A.; Bogacheva, E.; Fedorova, N.; Ksenofontov, A.; Badun, G.; Radyukhin, V.; Lukashina, E.; Serebryakova, M.; Dolgov, A.; Chulichkov, A.; Dobrov, E.; Baratova, L. Spatial structure peculiarities of influenza A virus matrix M1 protein in an acidic solution that simulates the internal lysosomal medium. *FEBS J*. **2011**, *278*, 4905–4916.
113. Thaa, B.; Herrmann, A.; Veit, M. The polybasic region is not essential for membrane binding of the matrix protein M1 of influenza virus. *Virology* **2009**, *383*, 150–155.
114. Fontana, J.; Cardone, G.; Heymann, J. B.; Winkler, D. C.; Steven, A. C. Structural changes in influenza virus at low pH characterized by cryo-electron tomography. *J. Virol*. **2012**, *86*, 2919–2929.
115. Brunotte, L.; Flies, J.; Bolte, H.; Reuther, P.; Vreede, F.; Schwemmle, M. The Nuclear Export Protein of H5N1 Influenza A Viruses Recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *J. Biol. Chem*. **2014**, *289*, 20067–20077.
116. Arzt, S.; Petit, I.; Burmeister, W. P.; Ruigrok, R. W. H.; Baudin, F. Structure of a knockout mutant of influenza virus M1 protein that has altered activities in membrane binding, oligomerisation and binding to NEP (NS2). *Virus Res*. **2004**, *99*, 115–119.
117. Zhang, K.; Wang, Z.; Liu, X.; Yin, C.; Basit, Z.; Bin Xia; Liu, W. Dissection of influenza A virus M1 Protein: pH-dependent oligomerization of N-terminal domain and dimerization of C-terminal domain. *PLoS ONE* **2012**, *7*, e37786.

118. Fontana, J.; Steven, A. C. At low pH, influenza virus matrix protein M1 undergoes a conformational change prior to dissociating from the membrane. *J. Virol*. **2013**, *87*, 5621–5628.
119. Mosier, P. D.; Chiang, M.-J.; Lin, Z.; Gao, Y.; Althufairi, B.; Zhou, Q.; Musayev, F.; Safo, M. K.; Xie, H.; Desai, U. R. Broad spectrum anti-influenza agents by inhibiting self-association of matrix protein 1. *Sci. Rep*. **2016**, *6*, 32340. http://www.nature.com/articles/srep32340 (accessed August , 06, 2016).
120. Lobo, N. A.; Shimono, Y.; Qian, D.; Clarke, M. F. The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol*. **2007**, *23*, 675–699.
121. Beck, B.; Blanpain, C. Unravelling cancer stem cell potential. *Nat. Rev. Cancer* **2013**, *13*, 727–738.
122. Crosnier, C.; Stamataki, D.; Lewis, J. Organizing cell renewal in the intestine: Stem cells, signals and combinatorial control. *Nat. Rev. Genet*. **2006**, *7*, 349–359.
123. Garza-Treviño, E. N.; Said-Fernández, S. L.; Martínez-Rodríguez, H. G. Understanding the colon cancer stem cells and perspectives on treatment. *Cancer Cell Int*. [Online] **2015**, *15*, 2. http://cancerci.biomedcentral.com/articles/10.1186/s12935-015-0163-7 (accessed August, 01, 2016).
124. Ailles, L. E.; Weissman, I. L. Cancer stem cells in solid tumors. *Curr. Opin. Biotechnol*. **2007**, *18*, 460–466.
125. Lapidot, T.; Sirard, C.; Vormoor, J.; Murdoch, B.; Hoang, T.; Caceres-Cortes, J.; Minden, M.; Paterson, B.; Caligiuri, M. A.; Dick, J. E. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **1994**, *367*, 645–648.
126. Visvader, J. E.; Lindeman, G. J. Cancer stem cells in solid tumours: Accumulating evidence and unresolved questions. *Nat. Rev. Cancer* **2008**, *8*, 755–768.
127. Izumi, D.; Ishimoto, T.; Sakamoto, Y.; Miyamoto, Y.; Baba, H. Molecular insights into colorectal cancer stem cell regulation by environmental factors. *J. Cancer Metastasis Treat*. **2015**, *1*, 156–162.
128. Feng, Y.; Dai, X.; Li, X.; Wang, H.; Liu, J.; Zhang, J.; Du, Y.; Xia, L. EGF signalling pathway regulates colon cancer stem cell proliferation and apoptosis. *Cell Prolif*. **2012**, *45*, 413–419.
129. Jibiki, N.; Saito, N.; Kameoka, S.; Kobayashi, M. Clinical significance of fibroblast growth factor (FGF) expression in colorectal cancer. *Int. Surg*. **2014**, *99*, 493–499.
130. Mishra, L.; Shetty, K.; Tang, Y.; Stuart, A.; Byers, S. W. The role of TGF-beta and Wnt signaling in gastrointestinal stem cells and cancer. *Oncogene* **2005**, *24*, 5775–5789.
131. Xu, Y.; Pasche, B. TGF-signaling alterations and susceptibility to colorectal cancer. *Hum. Genet*. **2007**, *16*, R14–R20.
132. Sakaki-Yumoto, M.; Katsuno, Y.; Derynck, R. TGF-β family signaling in stem cells. *Biochim. Biophys. Acta* **2013**, *1830*, 2280–2296.
133. Reguart, N.; He, B.; Taron, M.; You, L.; Jablons, D. M.; Rosell, R. The role of Wnt signaling in cancer and stem cells. *Future Oncol.* **2005**, *1*, 787–797.

134. Vermeulen, L.; De Sousa E Melo, F.; van der Heijden, M.; Cameron, K.; de Jong, J. H.; Borovski, T.; Tuynman, J. B.; Todaro, M.; Merz, C.; Rodermond, H.; Sprick, M. R.; Kemper, K.; Richel, D. J.; Stassi, G.; Medema, J. P. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat. Cell Biol*. **2010**, *12*, 468–476.
135. Puglisi, M. A.; Tesori, V.; Lattanzi, W.; Gasbarrini, G. B.; Gasbarrini, A. Colon cancer stem cells: Controversies and perspectives. *World J. Gastroenterol*. **2013**, *19*, 2997–3006.
136. Kraushaar, D. C.; Dalton, S.; Wang, L. Heparan sulfate: A key regulator of embryonic stem cell fate. *Biol. Chem*. **2013**, *394*, 741–751.
137. Patel, N. J.; Karuturi, R.; Al-Horani, R. A.; Baranwal, S.; Patel, J.; Desai, U. R.; Patel, B. B. Synthetic, non-saccharide, glycosaminoglycan mimetics selectively target colon cancer stem cells. *ACS Chem. Biol*. **2014**, *9*, 1826–1833.
138. Kim, Y,; Ho, S. O.; Gassman, N. R.; Korlann, Y.; Landorf, E. V.; Collart, F. R.; Weiss, S. Efficient site-specific labeling of proteins via cysteines. *Bioconjug. Chem*. **2008**, *19*, 786–791.