UCLA
UCLA Electronic Theses and Dissertations

Title
Discovery and Design of Multifunctional Membrane-Active and Immunomodulatory Peptides and Proteins

Permalink
https://escholarship.org/uc/item/4k31t3fb

Author
Lee, Ernest

Publication Date
2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Discovery and Design of Multifunctional Membrane-Active and Immunomodulatory Peptides and Proteins

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Bioengineering

by

Ernest Lee

2018

© Copyright by

Ernest Lee

2018

ABSTRACT OF THE DISSERTATION

Discovery and Design of Multifunctional Membrane-Active and Immunomodulatory Peptides and Proteins

by

Ernest Lee
Doctor of Philosophy in Bioengineering
University of California, Los Angeles, 2018
Professor Gerard Chee Lai Wong, Chair

The human innate immune system consists of several components, which include antimicrobial peptides, pattern-recognition receptors, cytokines, and immune cells. In this dissertation, we explore unifying themes underlying the antimicrobial, membrane remodeling, and immunomodulatory behaviors of antimicrobial peptides and related molecules, and their interactions with microbial and mammalian cells. We utilize machine learning on antimicrobial peptides to examine the physicochemical parameters characteristic of membrane curvature generation, and develop a search tool to discover hidden antimicrobial and membrane-remodeling activity in new and existing taxonomies of multifunctional peptides and proteins, including mitochondrial fission proteins, histones, and neuropeptides. Using structural characterization and calibrating immune cell stimulation experiments, we outline molecular rules for antimicrobial peptide-mediated immunomodulation via ligand clustering of nucleic acids. Antimicrobial peptides condense naked DNA, nucleosomal DNA, and dsRNA into nanocrystalline immunocomplexes, which drastically amplify inflammation via multivalent binding to Toll-like receptors in immune cells. This work has broad implications for the deterministic control of inflammation in the contexts of infection, chronic inflammation, and autoimmune disease.

The dissertation of Ernest Lee is approved.

Tom Chou

Jeffery Floyd Miller

Carlos Portera-Cailliau

Jacob J. Schmidt

Andrea M. Kasko

Gerard Chee Lai Wong, Committee Chair

University of California, Los Angeles

2018

To my family.

TABLE OF CONTENTS

1 Introduction to antimicrobial peptides, membrane biophysics, membrane curvature, and machine learning 1

1.1 Antimicrobial Peptides 1

1.2 Phospholipid Membranes and Membrane Curvature 4

1.3 Induction of Membrane Curvature and Negative Gaussian Curvature by Diverse Peptides and Proteins 5

1.4 Introduction to Machine Learning 7

1.5 Machine Learning on AMPs 9

1.6 Summary 12

2 Mapping membrane activity in undiscovered peptide sequence space using machine learning 13

2.1 Abstract 13

2.2 Introduction 14

2.3 Results and Discussion 16

2.3.1 SVM Model Development, Monte-Carlo Sampling, and Pareto Analysis 16

2.3.2 Distance to hyperplane of known AMPs does not correlate with antimicrobial efficacy 22

2.3.3 Predicted distance-to-hyperplane of synthesized test peptides correlates with strength of negative Gaussian curvature generation 26

2.3.4 Directed search of the sequence space of physicochemically restricted and unrestricted peptides discovers diverse families of membrane curvature-generating peptides 33

2.3.5 Pareto-optimal sequences and newly discovered curvature-generating sequences follow the saddle-splay selection rule. 34
2.4 Conclusions 37
2.5 Methods 39
2.5.1 Computational Methods 39
2.5.2 SAXS Experiments 50
2.5.3 Antimicrobial Assays 52
2.5.4 Statistical Analysis 52
2.6 Acknowledgements 52

3 Machine learning-enabled discovery of membrane-active domains in naturally occurring peptides and proteins 54
3.1 Introduction 54
3.2 Molecular Motor Dnm1 Synergistically Induces Membrane Curvature To Facilitate Mitochondrial Fission 54
3.2.1 Abstract 54
3.2.2 Introduction 55
3.2.3 Results and Discussion 58
3.2.4 Conclusion and Prospects 71
3.2.5 Methods 71
3.2.6 Acknowledgements 74
3.3 Histone H4 permeates membranes by generating negative Gaussian curvature 75
3.3.1 Introduction 75
3.3.2 Machine learning predicts that membrane activity of histone H4 is localized to the N-terminus 75

3.3.3 Histone H4 N-terminus generates negative Gaussian curvature in cell membranes . . . 77
3.3.4 Methods . . . 79
3.4 Summary . . . 81

4 Introduction to innate immunity, Toll-like receptors, electrostatics of polyelectrolytes, and statistical mechanics of receptor-ligand interactions . . . 82
4.1 Innate Immunity, Cytokines, and Type I Interferons . . . 82
4.2 Toll-like Receptors . . . 83
4.3 Immunomodulatory Properties of AMPs . . . 85
4.4 AMP Interactions with TLRs and Relevance to Autoimmune Disease . . . 87
4.4.1 dsDNA binding and TLR9 activation . . . 88
4.4.2 dsRNA binding and TLR3 activation . . . 92
4.4.3 ssRNA binding and TLR7/8 activation . . . 92
4.5 Electrostatics of Biological Polyelectrolytes . . . 94
4.6 Statistical Mechanics of Multivalent, Superselective Receptor-Ligand Interactions . . . 96
4.7 Nanocrystalline dsDNA complexes influence the pathophysiology of autoimmune disease . . . 100
4.7.1 Bacterial amyloid curli acts as a carrier for DNA to elicit an autoimmune response via TLR2 and TLR9 . . . 100
4.7.2 CXCL4-DNA nanocrystal formation enables multivalent recognition of self-DNA in the pathogenesis of systemic sclerosis . . . 104
4.8 Summary . . . 112

5 Crystallinity of dsRNA-AMP immune complexes modulates TLR3-mediated inflammation . . . 113

5.1 Abstract . . . 113
5.2 Introduction . . . 114
5.3 Results and Discussion . . . 116
5.3.1 IL-6 production in keratinocytes through TLR3 activation depends on the inter-dsRNA spacing within self-assembled polycation-dsRNA electrostatic complexes . . . 116
5.3.2 Potential relationship between TLR3 activation and the inter-RNA spacing within dsRNA complexes . . . 121
5.3.3 Cytokine-inducing polycation-dsRNA complexes activate NHEK via TLR3 . . . 124
5.3.4 Coarse-grained model of TLR3 binding to RNA bundles exhibits dependence on inter-dsRNA spacing and repeat number in agreement with experimental observations . . . 125
5.4 Conclusions and Prospects . . . 133
5.5 Methods . . . 134
5.5.1 Preparation of self-assembled polycation-dsRNA complexes . . . 134
5.5.2 SAXS Experiments and Data Analysis . . . 136
5.5.3 Reverse transcription-quantitative PCR (RTqPCR) analyses . . . 138
5.5.4 Keratinocyte Stimulation Experiments . . . 138
5.5.5 NHEK Immunofluorostaining Experiments . . . 138
5.5.6 Statistical Analyses . . . 139
5.5.7 Analytical Theory Derivation . . . 139
5.5.8 Monte Carlo Simulation Model . . . 143
5.6 Acknowledgements . . . 143

6 Neutrophil extracellular traps and necrotic cell death: structural basis of

chromatin-mediated inflammation **145**

6.1 Abstract 145

6.2 Introduction 146

6.3 Results and Discussion 148

6.3.1 NCPs form columnar nanocrystalline aggregates in the presence of calcium and antimicrobial peptides and amplify TLR9-mediated inflammation 148

6.3.2 Coarse-grained simulations demonstrate that columnar aggregates of NCPs exhibit altered DNA wrapping states well-matched to the steric size of TLR9 152

6.3.3 LL37 organizes NCPs into hexagonal columnar nanostructures and amplifies TLR9-mediated inflammation 154

6.3.4 TLR9 activation by remodeled chromatin depends on inter-DNA spacing and size of the NCP columns 155

6.4 Conclusions and Prospects 159

6.5 Methods 160

6.5.1 Preparation of polycation-NCP complexes 160

6.5.2 SAXS Experiments 161

6.5.3 pDC Stimulation Experiments 162

6.5.4 Potential of Mean Force Calculations 162

6.5.5 Coarse-Grained Simulations of DNA Superhelix Spacings and Energy Calculation 164

7 Structural studies of host defense peptides that lead to TLR9-based immune activation **165**

7.1 Abstract 165

7.2 Introduction 166

7.3 Results and Discussion . . . 169
7.3.1 Selection of candidate α-helical AMPs . . . 169
7.3.2 Melittin condenses dsDNA into a tetragonal lattice . . . 171
7.3.3 LL37 condenses dsDNA into a 2D columnar square lattice . . . 173
7.3.4 Buforin condenses dsDNA into a disordered columnar lattice . . . 176
7.3.5 Bent cationic antimicrobial peptide-dsDNA complexes amplify inflammation by binding to TLR9 . . . 177
7.3.6 Tuning immunomodulatory activity of AMPs by modifying their membrane activity . . . 178
7.4 Conclusions and Prospects . . . 185
7.5 Methods . . . 186
7.5.1 Preparation of self-assembled peptide-dsDNA complexes . . . 186
7.5.2 SAXS experiments with peptide-dsDNA complexes . . . 186
7.5.3 Domain Size Calculations . . . 187
7.5.4 Macrophage Stimulation Experiments . . . 187
7.5.5 IFN-β Bioassay . . . 188
7.5.6 pDC Stimulation Experiments . . . 188
7.5.7 SAXS Experiments with Model Membranes . . . 189
7.5.8 Computational Methods and Molecular Simulations . . . 190
7.5.9 Statistical Analyses . . . 193

8 Conclusions and Outlook . . . 194
8.1 Summary . . . 194
8.2 Conclusions and Outlook . . . 195

References . . . 201

LIST OF FIGURES

1.1 **Antimicrobial peptides and their mechanisms of action.** 3

1.2 **Common types of membrane curvature.** 5

1.3 **Structures of lipid phases.** 8

2.1 **Support vector machine learning and Pareto-optimization select for antimicrobial and membrane curvature-generating peptides.** 17

2.2 **Accuracy of the SVM.** 19

2.3 **Predictive performance of SVM.** 20

2.4 **Sequence atlas and Pareto frontier constructed by directed sampling of sequence space.** 23

2.5 **Distance-to-hyperplane of known AMPs does not correlate with antimicrobial efficacy.** 24

2.6 **Synthesized test peptides from directed Monte-Carlo search and SVM screening generate negative Gaussian curvature in model membranes.** 25

2.7 **Negative control peptides do not generate cubic phases in model membranes.** 28

2.8 **Visualization of NGC generation by test peptides relative to distance to margin and homology to known AMPs.** 30

2.9 **Distance-to-hyperplane of test peptides does correlate with strength of negative Gaussian curvature.** 31

2.10 **Test peptides spanning a large range of helicities and homologies to known AMPs can generate magnitudes of NGC similar to those of known AMPs.** 31

2.11 **Amphiphilicity is implicitly encoded within the SVM model and is a feature of positively-classified test peptides near Pareto frontier and α-helical AMPs in general.** 32

2.12 **Directed search of the sequence space discovers diverse families of membrane curvature-generating peptides.** 35

2.13 **2D projections of 3D sequence atlas and Pareto frontier.** 36

2.14 **Pareto-optimal sequences and newly discovered curvature-generating sequences follow the saddle-splay selection rule.** 38

3.1 **Machine learning screen of Dnm1 yields a putative membrane-destabilizing sequence in the conserved N-terminal domain.** 59

3.2 **Full sequence alignment of Drp1 and Dnm1.** 62

3.3 **Dnm1 generates NGC necessary for mitochondrial membrane fission.** 63

3.4 **Helix-rich Fis1 is not predicted to destabilize membranes and does not restructure model mitochondrial membranes.** 66

3.5 **Membrane activity fitness landscape of the Dnm1 N-terminal helix and evolution of membrane activity in the dynamin superfamily.** 69

3.6 **Mutability analysis of the Dnm1 N-terminal helix.** 70

3.7 **Multiple sequence alignment of dynamin superfamily proteins.** 73

3.8 **Histone H4 N-terminus generates negative Gaussian curvature in cell membranes.** 77

4.1 **Nucleic Acid Toll-Like Receptors and their Ligands.** 84

4.2 **Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation** 90

4.3 **TLR9 binding to DNA bundles exhibits "super-selective" behavior in agreement with experimental observations.** 91

4.4 **The structure of CsgA$_{R4-5}$-DNA complexes explains their ability to induce type I interferon responses via TLR9.** 102

4.5 **CXCL4 condenses human and bacterial DNA into columnar nanocrystalline complexes at an inter-DNA spacing optimal for TLR9 activation.** 106

4.6 **The C-terminus of CXCL4 can organize fragmented human DNA into more ordered lattices with optimal TLR9-activating inter-DNA spacings than unfragmented human DNA.** 108

4.7 **LL37 binds both human and bacterial DNA into square lattices that amplify TLR9 activation.** 110

5.1 **IL-6 production in keratinocytes through TLR3 activation depends on the inter-dsRNA spacing within self-assembled polycation-dsRNA electrostatic complexes.** 117

5.2 **Differences in structure of polycation-dsRNA complexes predict their abilities to amplify cytokine production.** 119

5.3 **Null-hypothesis significance test for statistical dependence between TLR3 activation and inter-dsRNA spacing.** 123

5.4 **Cytokine-inducing polycation-dsRNA complexes activate NHEK via TLR3.** 126

5.5 **Coarse-grained model of TLR3 binding to RNA bundles exhibits dependence on inter-dsRNA spacing and repeat number in agreement with experimental observations.** 128

5.6 **Effects of repeat number m and binding strength ϵ on TLR3 binding to peptide-dsRNA nanocrystalline clusters.** 130

5.7 **Dependence of TLR3 activation on first peak position q and repeat number m.** 131

5.8 **Monte-Carlo simulation snapshot of TLR3 dimers binding to a dsRNA-AMP nanocrystalline cluster.** 132

5.9 **Superselective binding of TLR3 to polycation-dsRNA complexes occurs only at inter-dsRNA spacings commensurate with the steric size of TLR3.** 134

6.1 **NCPs assemble into columnar complexes in the presence of calcium and activate pDCs via TLR9.** 149

6.2 **Antimicrobial peptides hBD3 and lysozyme also condense NCPs into columnar structures that induce IFN-α via TLR9 binding.** 151

6.3 **NCP columns assembled in the presence of polycations exhibit altered inter-DNA spacings.** 152

6.4 **LL37 organizes NCPs into hexagonal columnar nanostructures and amplifies TLR9-mediated inflammation.** 156

6.5 **TLR9 activation by remodeled chromatin depends on inter-DNA spacing and size of the NCP columns.** 158

6.6 **NCPs released from processed chromatin self-assemble with antimicrobial peptides into crystalline nanostructures that hyperactivate TLR9 in pDCs.** 160

7.1 **SAXS structure of melittin-dsDNA complexes.** 170

7.2 **Structure of melittin-dsDNA complexes from molecular simulations** . 172

7.3 **Structure of LL37-dsDNA Complexes from SAXS and molecular simulations.** 174

7.4 **Structure of buforin-dsDNA complexes.** 176

7.5 **LL37-dsDNA and buforin-dsDNA complexes induce Type 1 Interferon production in a TLR9-dependent manner in mouse macrophages and human pDCs.** 179

7.6 **Systematic reduction of the membrane activity of melittin in a series of mutants with conserved dsDNA-binding activity corresponds with a recovery of TLR9-activation potential.** 182

7.7 **SAXS Lattice Parameter Fits for Membrane Activity of Melittin-related Peptides.** 183

7.8 **SAXS Lattice Parameter Fits for DNA Complexes with Melittin-related Peptides.** 184

8.1 **Antimicrobial peptides can modulate immune responses through multiple mechanisms.** 196

8.2 **Nanocrystalline ordering of dsDNA and dsRNA by AMPs modulates TLR-mediated inflammation.** 198

LIST OF TABLES

2.1 **Subset of 12 bagged descriptors identified by L1-SVC variable selection procedure.** 18

2.2 **Measured physicochemical parameters and SVM classification of 16 synthesized positively-classified test peptides near the Pareto frontier.** 26

2.3 **Physicochemical descriptors.** 42

2.4 **Performance of trained SVMs.** 49

3.1 **Machine learning screen and phylogenetic distances of dynamins.** . . . 74

5.1 **Data from SAXS studies and cell experiments for dsRNA complexed with various polycationic molecules.** 135

5.2 **List of primers used for RT-qPCR.** 136

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the innumerable contributors who helped make this dissertation possible. First and foremost, I want to acknowledge my mentor Gerard C.L. Wong who, in many ways, served simultaneously as a father figure, colleague, friend, and thesis advisor. We shared shared similar upbringings and were both undergraduate physics majors at Caltech, 25 years apart. We also shared an especially strong passion for good food and drink. I was initially drawn to his lab by the multidisciplinary nature of his research, which involved collaborations among physicists, engineers, biologists, and clinicians. As an aspiring physician-scientist who was passionate about both basic science and medicine, I quickly realized that this was the best learning environment anyone could ask for. He taught me how to leverage my unique background to answer unsolved questions at the interface of physics and biomedicine, with real-world clinical implications. At the same time, he also allowed me to maintain a flexible schedule and was respectful of personal commitments. He gave me the freedom to pursue multiple projects, manage my own grant, take an active role in the mentorship of other students, and travel to countless conferences. My fondest memories are of the frequent, high-intensity one-on-one meetings in which we gradually outlined the overarching conceptual frameworks for the physical and chemical rules governing the immunomodulatory and membrane activity of host-defense peptides. In summary, I cannot express in words how grateful I am for his continued mentorship, support, and advocacy along my journey.

I would also like to thank each and every current and former member of the Wong group who I overlapped with during graduate school, including Michelle Lee, Calvin Lee, Jaime de Anda, Wujing Xian, Nathan Schmidt, and Ghee Hwee Lai. I want to especially acknowledge Michelle, who I have collaborated extensively with throughout graduate school, and who trained me on X-ray experiments when I first joined the group. I am extremely thankful for her companionship on synchrotron runs and her amazing food recommendations.

I want to thank all the undergraduate researchers in our lab who assisted me with killing assays, immune stimulation experiments, and bioinformatics/computational work, including Juelline Lieng, Xiang Ji, Will Connell, Jennifer Wang, Sandy Zarmer, Mandy Hung, Cole Malkoff, Yashes Srinivasan, and Veronica Veksler.

This thesis also would not be possible without the key contributions from over 20 collaborators from all around the world. I would like to specifically acknowledge Dr. Andrew Ferguson, Toshiya Takahashi, Tine Curk, Jeremy Di Domizio, Changsheng Zhang, Sarah Tursi, Amber Kaplan, Leila Abbaspour, Dr. Richard Gallo, Dr. Michel Gilliet, Dr. Jure Dobnikar, Dr. Robert Modliin, Dr. Nannette Yount, Dr. Michael Yeaman, Dr. Pengyu Ren, Dr. Cagla Tukel, Dr. Roberto Lande, Dr. Loredana Frasca, Dr. Andrea Armani, Dr. Hossein Fazli, and Dr. Lars Nordenskiold.

I thank the UCLA-Caltech MSTP for accepting me into this rigorous training program, the former directors Dr. Kelsey Martin and Dr. Steve Smale, and the current directors Dr. Carlos Portera-Cailliau and Dr. Leanne Jones. I want to particularly acknowledge Dr. Portera-Cailliau for his amazing mentorship through medical and graduate school. I thank all the members of my thesis committee, including Dr. Jeffery Miller, Dr. Tom Chou, Dr. Andrea Kasko, Dr. Jacob Schmidt, and Dr. Portera-Cailliau.

I also want to thank my previous research mentors from high school and college who accepted me into their labs and stoked my interest in pursuing research at an early age, including Dr. Bruce Hamilton, Dr. Charles Stevens, Dr. Joseph Noel, and Dr. Stephen Mayo.

Last but not least, I want to thank my parents Joan and Kuo-Fen and my sister Eudoria for their unwavering love and support throughout this journey, and for always providing me with opportunities to succeed.

Several chapters of this dissertation contain material from published work:

Parts of Chapter 1 are adapted with permission from Lee, E.Y., Lee, M.W., Fulan, B., Ferguson, A.L., Wong, G.C.L. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface Focus*, **7**(6), 20160153 (2017). Copyright © 2017 The Royal Society.

Parts of Chapter 2 are adapted with permission from the following publications: Lee, E.Y., Fulan, B., Wong, G.C.L., Ferguson, A.L. Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proc Natl Acad Sci USA*, **113**(48), 13588-13593 (2016). Copyright © 2016 National Academy of Sciences. Lee, E.Y., Wong, G.C.L., Ferguson, A.L. Machine learning-enabled discovery and design of membrane-active peptides. In press, *Bioorganic & Medicinal Chemistry* (2017). DOI: 10.1016/j.bmc.2017.07.012. Copyright © 2017 Elsevier.

Parts of Chapter 3 are adapted with permission from Lee, M.W., Lee, E.Y., Lai, G.H. Kennedy, N.W., Posey, A.E., Xian, W., Ferguson A.L., Hill, R.B., Wong, G.C.L. Molecular Motor Dnm1 Synergistically Induces Membrane Curvature To Facilitate Mitochondrial Fission. *ACS Central Science*, **3**(11), 1156-1167 (2017). Copyright © 2017 American Chemical Society.

Parts of Chapter 4 are adapted with permission from the following publications: Lee, E.Y., Lee, C.K., Schmidt, N.W., Jin, F., Lande, R., Curk, T., Frenkel, D., Dobnikar, J., Gilliet, M., Wong, G.C.L., A review of immune amplification via ligand clustering by self-assembled liquid-crystalline DNA complexes. *Advances in Colloid and Interface Science*, **232**, 17-24 (2016). Copyright © 2016 Elsevier. Tursi, S.A., Lee, E.Y., Medeiros, N.J., Lee, M.H., Butter B., Galluccci S., Wilson, R.P., Wong, G.C.L., Tukel C. Bacterial amyloid curli acts as a carrier for DNA to elicit an autoimmune response via TLR2 and

TLR9. *PLoS Pathogens*, **13**(4), e1006315 (2017). Copyright © 2017 Public Library of Science.

Parts of Chapter 5 are adapted with permission from Lee, E.Y., Takahashi, T., Curk, T., Dobnikar, J., Gallo, R.L., Wong, G.C.L. Crystallinity of Double-Stranded RNA-Antimicrobial Peptide Complexes Modulates Toll-Like Receptor 3-Mediated Inflammation. *ACS Nano*, **11**(12), 12145-12155 (2017). Copyright © 2017 American Chemical Society.

Parts of Chapter 8 are adapted with permission from Lee, E.Y., Lee, M.W., Wong, G.C.L. Modulation of Toll-like receptor signaling by antimicrobial peptides. In press, *Seminars in Cell & Developmental Biology*, (2018). DOI: 10.1016/j.semcdb.2018.02.002. Copyright © 2018 Elsevier.

Chapters 3, 4, 6, and 7 include data and text from unpublished manuscripts in preparation, submitted, or under review.

Work in this thesis was funded by the Systems and Integrative Biology Training Program (T32GM008185) at UCLA, the Medical Scientist Training Program (T32GM008042) at UCLA, the Dermatology Scientist Training Program (T32AR071307) at UCLA, and an Early Career Research Grant from the National Psoriasis Foundation. X-ray research was conducted at Stanford Synchrotron Radiation Lightsource, SLAC National Laboratory, supported by the US DOE Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

VITA

2012	B.S., Physics (with Honors) California Institute of Technology, Pasadena, California
2018	Ph.D., Bioengineering (expected) University of California, Los Angeles, Los Angeles, California
2020	M.D. (expected) University of California, Los Angeles, Los Angeles, California

PUBLICATIONS

Lee, E.Y., Lee, C.K., Schmidt, N.W., Jin, F., Lande, R., Curk, T., Frenkel, D., Dobnikar, J., Gilliet, M., Wong, G.C.L., A review of immune amplification via ligand clustering by self-assembled liquid-crystalline DNA complexes. *Advances in Colloid and Interface Science*, **232**, 17-24 (2016).

Sankhagowit, S., **Lee, E.Y.**, Wong, G.C.L., Malmstadt, N. Oxidation of Membrane Curvature-Regulating Phosphatidylethanolamine Lipid Results in Formation of Bilayer and Cubic Structures. *Langmuir*, **32**(10), 2450-2457 (2016).

Realegeno, S., Kelly-Scumpia, K.M., Dang, A.T., Lu, J., Teles, R., Liu, P.T., Schenk, M., **Lee, E.Y.**, Schmidt, N.W., Wong, G.C.L., Sarno, E.N., Rea, T.H., Ochoa, M.T., Pellegrini, M., Modlin, R.L. S100A12 Is Part of the Antimicrobial Network against *Mycobacterium leprae* in Human Macrophages. *PLoS Pathogens*, **12**(6), e1005705 (2016).

Lee, E.Y., Fulan, B., Wong, G.C.L., Ferguson, A.L. Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proc Natl Acad Sci USA*, **113**(48), 13588-13593 (2016).

Kaplan, A.*, Lee, M.W.*, Wolf, A.J., Limon-Tello, J., Becker, C.A., Ding, M., Murali, R., **Lee, E.Y.**, Liu, G.Y., Wong, G.C.L., Underhill, D.M. Direct Antimicrobial Activity of Interferon-β. *Journal of Immunology*, **198**(10), 4036-4045 (2017).
*Authors contributed equally

Tursi, S.A., **Lee, E.Y.**, Medeiros, N.J., Lee, M.H., Butter B., Galluccci S., Wilson, R.P., Wong, G.C.L., Tukel C. Bacterial amyloid curli acts as a carrier for DNA to elicit an autoimmune response via TLR2 and TLR9. *PLoS Pathogens*, **13**(4), e1006315 (2017).

Lee, E.Y., Wong, G.C.L., Ferguson, A.L. Machine learning-enabled discovery and design of membrane-active peptides. In press, *Bioorganic & Medicinal Chemistry* (2017). DOI: 10.1016/j.bmc.2017.07.012

Stolzenberg, E., Berry, D., Yang, D., **Lee, E.Y.**, Kroemer, A., Kaufman, S., Wong, G.C.L., Oppenheim, J., Sen, S., Fishbein, T., Bax, A., Harris, B., Barbut, D., Zasloff, M.A. A Role for Neuronal Alpha-Synuclein in Gastrointestinal Immunity. *Journal of Innate Immunity*, **9**(5), 456-463 (2017).

De Anda, J.*, **Lee, E.Y.***, Lee, C.K.*, Bennett, R.R., Ji, X., Soltani, S., Harrison, M.C., Baker, A.E., Luo, Y., Chou, T., O'Toole, G.A., Armani, A.M., Golestanian, R., Wong, G.C.L. High-Speed "4D" Computational Microscopy of Bacterial Surface Motility. *ACS Nano*, **11**(9), 9340-9351 (2017). *Co-first authorship

Lee, E.Y., Lee, M.W., Fulan, B., Ferguson, A.L., Wong, G.C.L. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface Focus*, **7**(6), 20160153 (2017).

Lee, M.W., **Lee, E.Y.**, Lai, G.H. Kennedy, N.W., Posey, A.E., Xian, W., Ferguson A.L., Hill, R.B., Wong, G.C.L. Molecular Motor Dnm1 Synergistically Induces Membrane Curvature To Facilitate Mitochondrial Fission. *ACS Central Science*, **3**(11), 1156-1167 (2017).

Lee, E.Y., Takahashi, T., Curk, T., Dobnikar, J., Gallo, R.L., Wong, G.C.L. Crystallinity of Double-Stranded RNA-Antimicrobial Peptide Complexes Modulates Toll-Like Receptor 3-Mediated Inflammation. *ACS Nano*, **11**(12), 12145-12155 (2017).

Takahashi, T., Kulkarni, N.N., **Lee, E.Y.**, Zhang, L-J., Wong, G.C.L., Gallo, R.L. Cathelicidin promotes inflammation by enabling binding of self-RNA to cell surface scavenger receptors. *Scientific Reports*, 8(1):4032 (2018).

Lee, C.K.*, De Anda, J.*, Baker, A.E., Bennett, R.R., Luo, Y., **Lee, E.Y.**, Keefe, J.A., Helali, J.S., Ma, J., Zhao, K., Golestanian, R., O'Toole, G.A., Wong, G.C.L. Multigenerational Memory and Adaptive Adhesion in Early Bacterial Biofilm Communities. In press, *Proc Natl Acad Sci USA* (2018). DOI: 10.1073/pnas.1720071115
*Authors contributed equally

Lee, E.Y., Lee, M.W., Wong, G.C.L. Modulation of Toll-like receptor signaling by antimicrobial peptides. In press, *Seminars in Cell & Developmental Biology* (2018). DOI: 10.1016/j.semcdb.2018.02.002

CHAPTER 1

Introduction to antimicrobial peptides, membrane biophysics, membrane curvature, and machine learning

1.1 Antimicrobial Peptides

Antimicrobial peptides (AMPs) are essential components of innate host defense [1, 2, 3, 4, 5, 6]. As of today, over 2000 natural and synthetic AMPs have been discovered and characterized [7, 8, 9, 10]. While these heterogeneous peptide-based broad spectrum antibiotics span an enormous diversity of sequences and secondary structures [1, 6], early studies have identified common characteristics among AMPs. They tend to have short amino acid sequences ($<$ 50 amino acids), net cationic charge (+2 to +9), and amphiphilicity [1, 2, 3, 4, 5, 6]. AMPs are often divided into three classes: α-helical AMPs [11, 12], β-sheet AMPs [13], and extended linear peptides enriched with specific amino acids [14, 15]. While AMPs in general are amphiphilic with segregated groups of polar and hydrophobic residues, a large class of AMPs can form α-helical structures that have polar (charged) and hydrophobic residues arranged onto opposite faces along the helical axis, thus creating facial amphiphilicity (Figure 1.1A). This unique presentation of residues is often described as being "amphipathic".

In vitro experiments suggest that AMPs generally function by selectively disrupting microbial membranes. In theory, this leads to cell death due to the loss of electrochemical gradients, reduction in resistance to osmotic stress, leakage of cellular contents, and disruption of metabolic processes [6, 2]. This bactericidal activity typically is dependent on interactions between AMPs and bacterial membranes [16], which has been demonstrated using a variety of experimental techniques, including X-ray scattering, nuclear magnetic resonance (NMR), dye leakage assays, electron microscopy, and circular dichroism [6]. Different

models describing membrane permeation have been proposed, including the "barrel-stave" model, the "carpet" model, and the "toroidal-pore" model among others (Figure 1.1B-D). In the "barrel-stave" model, amphipathic α-helical AMPs self-assemble into cylindrical bundles that are embedded perpendicularly into the cell membrane to form pores [17, 18, 19, 20, 21]. Within the membrane, the hydrophobic faces of the individual AMPs are oriented outward toward the hydrophobic interior of the bilayer, and the hydrophilic faces are oriented inward toward one another to form the lumen of the aqueous pore (Figure 1.1B) [3, 12]. In the "carpet" model, AMPs adsorb to the cell membrane in a parallel orientation. Once a critical local concentration is reached, the peptides disintegrate the membrane via micellization (Figure 1.1C) [22, 23, 3, 24, 25, 26]. Here, pore formation does not occur, unlike the "barrel-stave" model. In the "toroidal-pore" model, AMPs insert perpendicularly into the membrane to form a pore [27, 28] but differs from the "barrel-stave" model in that the membrane is integrated into the pore lining, forming a continuous interface with the peptides (Figure 1.1D) [3, 29, 30].

The selectivity of AMPs for bacterial membranes over eukyarotic membranes is generally believed to result from the compositional differences between their membranes [1, 6, 2]. More specifically, bacteria membranes contain large amounts of anionic lipids (e.g. phosphatidylglycerol (PG), phosphatidylserine (PS), and cardiolipin (CL)), while eukyarotic membranes contain mostly zwitterionic lipids (e.g. phosphatidylcholine (PC) and sphingomyelin (SM)) [34, 35, 36]. In human cells, cholesterol is particularly important [37]. Indeed, *in vitro* experiments have shown that the presence of anionic lipids results in increased membrane disruption and permeabilization by cationic membrane-active antimicrobials [1]. However, while the existence of anionic lipids is a necessary condition for permeation by AMPs, it is not a sufficient one. Bacterial membranes also contain high amounts of negative intrinsic curvature lipids, such as phosphatidylethanolamine (PE) and cardiolipin (CL), which predispose their membranes to poration [38, 37, 39, 40, 41, 42, 43]. In the simplest models, AMPs first interact with bacterial cells by binding electrostatically to their membrane surfaces, during which the cationic residues of the peptide bind to the anionic lipid head groups and other anionic surface components. After adsorbing onto the membrane with its helical

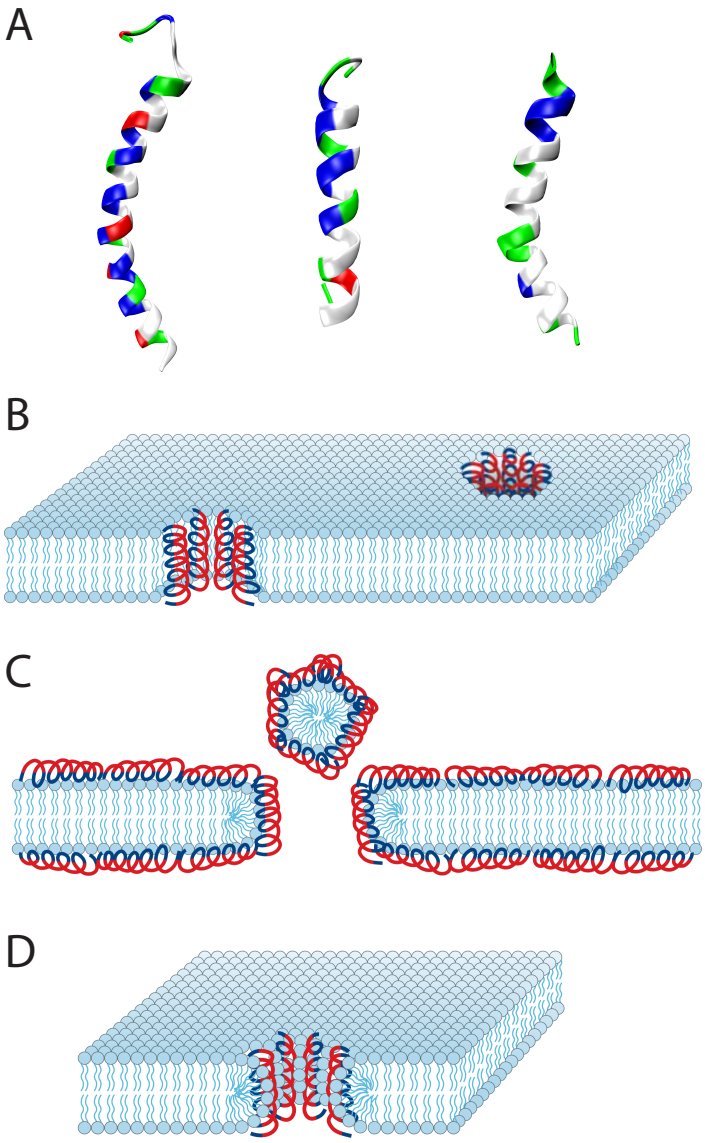


Figure 1.1: **Antimicrobial peptides and their mechanisms of action.** (A) Examples of cationic antimicrobial peptides: LL-37 [31] (left, PDB ID: 2K6O), magainin [32] (middle, PDB ID: 2MAG), and melittin [33] (right, PDB ID: 1MLT, 2MLT). Cationic residues are colored blue and hydrophobic residues are colored white. Structures were taken from the Protein Data Bank and visualized in VMD. Proposed mechanisms of AMP antimicrobial activity include the "barrel-stave" model (B), the "carpet" model (C), and the "toroidal pore" model (D). Reproduced with permission from [3], [12], and [28].

axis parallel to the surface, the AMP partitions into the lipid bilayer, driven primarily by hydrophobic interactions between its hydrophobic residues and the membrane core. The amphipathic nature of AMPs allows for this direct interaction with the cell membrane, which can then lead to membrane permeation and cell death [22]. In fact, the majority of early experimental studies conclude that membrane permeation underlies the primary mechanism of action of AMPs.

1.2 Phospholipid Membranes and Membrane Curvature

To understand how peptides can permeate membranes, we provide a brief overview of the physics of membrane curvature and deformation processes. Cell membranes are fluid bilayers containing two leaflets that can stretch and deform. Although cell membranes exist in three dimensions, they are better visualized as a curved two-dimensional surface. At any given point on the surface, the curvature can be defined by a plane tangent to that point. Any plane orthogonal to this tangent plane can intersect the surface with some curvature value $c = \frac{1}{R}$ defined by the radius of curvature R. Given all possible orthogonal planes, the two that give the maximal and minimal curvatures are defined as $c_1 = \frac{1}{R_{min}}$ and $c_2 = \frac{1}{R_{max}}$, respectively. These are called the two principal curvatures, and in combination, can be used to describe the shape of the surface at that point. Mathematically, the arithmetic mean of these two principal curvatures is the mean curvature $H = \frac{1}{2}(c_1 + c_2)$, and the product of the two principal curvatures is the Gaussian curvature $K = c_1 c_2$ [44]. Curvature can be positive or negative, however, when describing a membrane monolayer, the sign of curvature is conventionally defined by the direction of bending. Positive curvature is defined by bending of the monolayer to form a convex hydrophilic surface (such as the outside of a spherical cell), while negative curvature is defined as the bending of the monolayer to form a concave hydrophilic surface (such as on the neck of a budding vesicle) [45, 46, 47] (Figure 1.2). Furthermore, positive Gaussian ("dome-like") curvature, $K > 0$, results from principal curvatures of the same sign, while negative Gaussian ("saddle-like") curvature (NGC), $K < 0$, results from principal curvatures of the opposite sign (Figure 1.2). Membrane curvature can also be influenced by

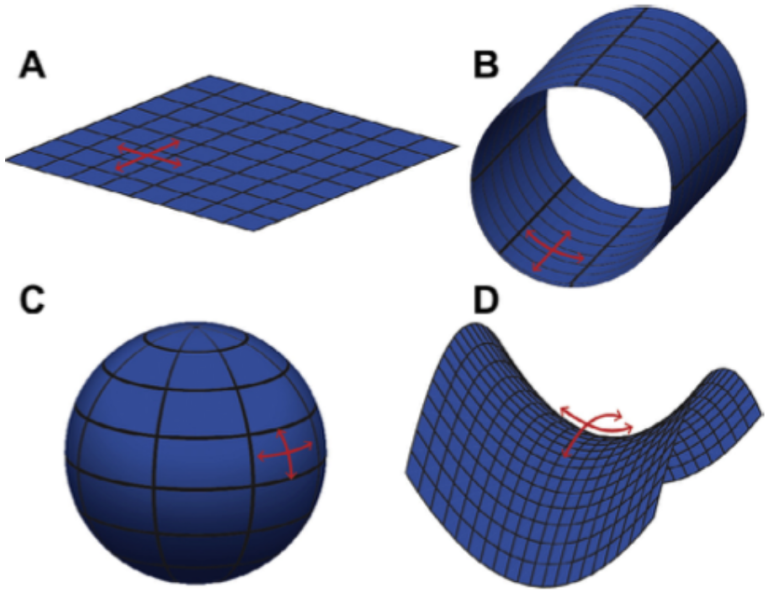


Figure 1.2: **Common types of membrane curvature.**
(A) A planar bilayer has both zero mean curvature and zero Gaussian curvature ($H = 0, K = 0$). (B) A cylindrical membrane is commonly found in inverted hexagonal phase. This has negative mean curvature and zero Gaussian curvature ($H < 0, K = 0$). (C) Spherical shapes like micelles or vesicles have positive Gaussian curvature ($K > 0$). (D) Saddle shapes are classic examples of negative Gaussian curvature ($K < 0$). These are commonly seen in cubic phases such as Pn3m, Im3m, and Ia3d. Adapted with permission from [47].

the geometry and packing behavior of the various phospholipids present in the bilayer, which together define the equilibrium shape of the membrane. Thus, deformation of the membrane from equilibrium exacts an energetic penalty that can be calculated from curvature changes and mechanical properties of the membrane.

1.3 Induction of Membrane Curvature and Negative Gaussian Curvature by Diverse Peptides and Proteins

Peptides and proteins can generate membrane curvature through several mechanisms, including membrane partitioning and insertion [48], membrane scaffolding [49, 37], curvature sensing [50], molecular crowding [51, 52], and membrane wrapping [53, 54]. Often, the mechanism of membrane deforming peptides and proteins involve one or more of the above. The specific structure of amphipathic α-helical AMPs, with separate polar and hydrophobic faces,

facilitates the interactions associated with these mechanisms. The hydrophobic domains contacting the hydrocarbon chains in the lipid core, driving positive curvature generation due to steric impingement in the membrane. Simultaneously, electrostatic interactions between the cationic groups of the peptide and anionic lipid head groups drive negative curvature by inducing wrapping of the membrane.

The variation in membrane disruption and curvature effects among amphipathic cationic α-helical AMPs has been attributed to differences in the charge distribution and sizes of the polar and hydrophobic faces of their helices [55, 56]. A large number of studies have explored correlations between the relative sizes of the polar and hydrophobic faces with biological effects, such as cell lysis [57, 55, 58, 56, 59, 60]. Specifically, lytic antibacterial helical peptides that are known to destabilize membranes can often be described as having an inverted wedge-shaped cross-section, with a narrow polar face to form the apex, and a wide hydrophobic face to form the base [2, 55, 57, 58]. A large hydrophobic face allows the peptide to deeply penetrate into the membrane, perturbing the packing of the bilayer core. This can lead to membrane thinning and cause the membrane to become more susceptible to curvature deformations as a result of reduced bending moduli [61, 62, 63, 28]. Moreover, previous work has found that increasing the bulkiness or angle subtended by the hydrophobic face further increases the ability of the peptide to destabilize membranes and cause lysis [55, 64, 57]. For peptides with narrow polar faces and wide hydrophobic faces, we expect the hydrophobic insertion effects to dominate, resulting in strong positive curvature. Conversely, for peptides with wide polar faces and narrow hydrophobic faces, electrostatic interactions between the cationic residues and anionic lipid head groups cause membrane wrapping (negative curvature) to dominate over the positive curvature from hydrophobic insertion.

This ability of a cationic amphipathic AMP to induce both positive and negative curvature in mutually orthogonal directions at a single location results in NGC, which is the type of curvature that is topologically necessary for membrane destabilization processes. In fact, AMPs have been shown to destabilize membranes through a variety of modes, including pore-formation [65, 17, 66], blebbing [67, 68], budding [69], and vesicularization [70, 71], all

of which require NGC. Since AMPs vary in relative amounts and distributions of cationic charge and hydrophobicity, and thus, can yield different levels of negative and positive curvature, an optimal balance among these properties would enable efficient generation of NGC. For instance, for many AMPs, the angle subtended by the polar face tends to be $\leq 100^\circ$ and the ratio of cationic/anionic residues is approximately 4-5 [57].

While the actual form of membrane destabilization is dependent on the physical chemistry of the specific AMP and the target membrane, AMP-induced disruption of bacterial membranes has been found to correlate with the ability of the peptides to generate NGC [72]. Synchrotron SAXS is an efficient way for measuring NGC in lipid membranes. Recent work has used SAXS to map out the phase behavior of lipids complexed with membrane deforming peptides and proteins. For example, three families of defensins, a class of AMPs, were assayed for their induction of membrane curvature deformations by incubating the peptides with SUVs composed of binary and ternary mixtures of phospholipids. By tuning the phospholipid compositions of the SUVs (e.g. modifying molar ratios between PS, PE, PC, and PG), one can model the lipid compositions of prokaryotic and eukaryotic membranes. For compositions mimicking bacterial membranes, defensins were found to restructure SUVs to liquid-crystalline cubic phases (e.g. Im3m, Pn3m, Ia3d, see Figure 1.3), which are characterized by a minimal surface with NGC at every point. [73, 72]. The ability to generate NGC has also been observed for cell-penetrating peptides [74], natural and synthetic AMPs [75, 76], and viral fusion proteins [77, 78]. Taken together, these results suggest that the generation of NGC is not only a feature of AMPs, but a common root mechanism for membrane destabilizing processes in general.

1.4 Introduction to Machine Learning

Machine learning leverages expertise from mathematics, statistics, and computer science to learn from data. Machine learning models can be divided into supervised and unsupervised methods. Supervised learning builds a prediction model based on existing "ground truth" data, which consists of actual measured outcomes for each object of interest. In turn,

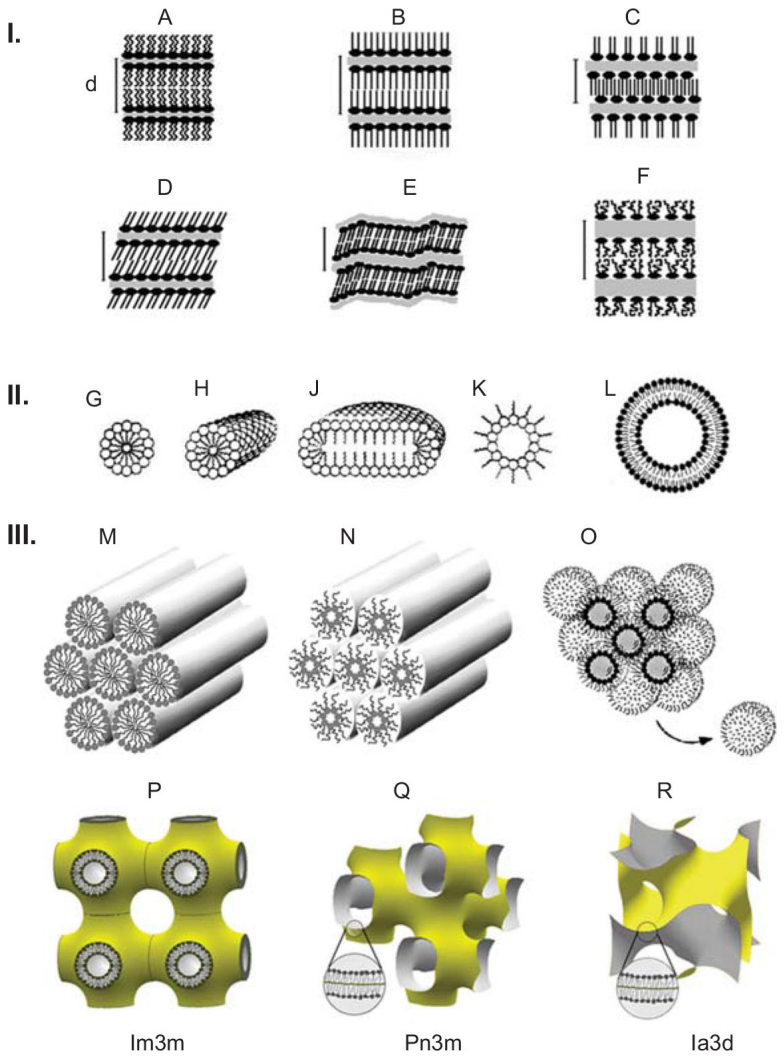


Figure 1.3: **Structures of lipid phases.**
I. Lamellar phases (A-C). II. Micellar aggregates (G-L). Liposomes (L) are used as mimics of eukaryotic and prokaryotic membranes in SAXS experiments. III. Liquid-crystalline phases with hexagonal (M, N)) or cubic symmetry (O, P, Q, R). The commonly observed lipid cubic phases rich in negative Gaussian curvature are the Im3m (P), Pn3m (Q), and Ia3d (R) phases. Adapted with permission from [79].

each object can be characterized by any number of input variables or "features". The presence of the outcome variable guides the learning process. Supervised models will take the "features" as input and output a prediction, which can be a binary, categorical, or continuous variable. In contrast, unsupervised learning approaches do not rely on measurements of outcome and rely purely on the input "features" of the objects. Unsupervised approaches must infer a function to discover hidden trends in the data. The majority of established learning methods revolve around supervised learning but literature on unsupervised methods is growing quickly. Most of the learning methods discussed in this review center upon supervised learning on validated datasets of AMPs.

The mathematical and statistical framework for machine learning has existed for centuries, far before computers were invented. Ideas and elements most recognizable today in machine learning, including Bayes' Theorem, principal component analysis (PCA), multiple linear regression (MLR), least squares fitting, and Markov chains, were established by mathematicians before 1950 [80]. Alan Turing's pioneering work on the Turing machine in 1950 [81] led to the development of the first artificial neural network (ANN) [33]. After the invention of the modern computer, research in machine learning exploded with the development of many of the modern methods used today, including partial least squares regression (PLSR), recurrent neural networks (RNN), hidden Markov models (HMM), support vector machines (SVM), and random forests (RF) [80]. Most recently, advances in scalable training algorithms and the availability of large data sets has spurred a resurgence of interest in ANNs with deep network architectures capable of tasks including handwriting recognition [82], cancer diagnosis [83], HIV classification [84], facial recognition [85], spam filtering [86], and quantum chemistry [87]. In 2016 the "AlphaGo" ANN rose to prominence by beating a 9-dan human grand master at the ancient board game Go [88].

1.5 Machine Learning on AMPs

As machine learning methods matured in the latter half of the century, the same was occurring to the field of antimicrobial peptides. As described in the first section, decades of

biophysical studies centered upon understanding AMP mechanisms of action and sequence rules gradually filled the literature. The advent of high-throughput screening coupled with decades of experimental data allowed for curation of large annotated datasets [9]. In the last 10-15 years, the focus of machine learning has shifted to an intensely data-driven approach. Significant advancements in computational power and easy-to-use statistical learning tools has made supervised machine learning a viable strategy for leveraging large datasets for the high-throughput and high-accuracy classification of AMPs. Typical readouts from biophysical assays on AMPs include calculations of minimum inhibitory concentrations, minimum bactericidal concentrations, and binding affinities. These quantities, coupled with sequence information about AMPs, allow for the training of various supervised learning models using peptide sequence information as an input. Before this era, methods for de novo AMP discovery relied on long-standing bioinformatics methods, including sequence alignment and homology modeling for prediction of biological activity. Now, the convergence of innovations in machine learning models, the presence of modern computational tools, and the availability of high quality datasets has enabled the machine learning-aided design of AMP candidates.

The diversity of AMP sequences and structures coupled with the time and expense associated with experimental design, production, and testing of antimicrobial peptide candidates precludes comprehensive experimental screening of peptide sequence space. Thus, the earliest machine learning models were quantitative structure-active relationship (QSAR) models that proved useful in efficient screening and optimization of a small number of promising sequences for experimental evaluation. QSAR models seek to use physicochemical descriptors to predict biological activity of a molecule that is typically expensive and/or time consuming to measure or calculate. In contrast, many physicochemical properties of a peptide can be inexpensively computed directly from its amino acid sequence [89, 90]. Models are trained and validated over experimentally characterized databases, then employed in high-throughput *in silico* screening to identify novel candidates with the desired biological activity. This approach relies on statistical learning to infer empirical relationships between physicochemical properties and biological activity, and is therefore contingent on an underlying relationship between (some subset) of the descriptors, the capacity of the machine learning model to

discover and encode this relationship in a mathematical expression, and sufficiently large and diverse training databases to produce robust predictive models [91, 89, 80].

QSAR models for computational AMP design have been developed using a variety of statistical learning approaches. The majority of prior AMP machine learning studies have tended to focus on either optimizing classification accuracy of AMPs or identifying potent AMP candidates with low minimum inhibitory concentrations. In one of the first applications of machine learning to AMPs, Lata et al. developed a QSAR AMP classification tool based on ANN, SVM, and quantitative matrix (QM) models based on unique motifs found in the C- and N-terminal residues of known AMPs [92]. In 2008, Chersakov et al. utilized high-throughput screening methods to train an ANN model on the measured antimicrobial efficacies of thousands of 9-residue peptides to discover potent antimicrobial that were potent against multi-drug resistant bacteria [93]. Fjell et al. published a 2008 study using hidden Markov models to screen for AMPs in the bovine genome, which led to the discovery of a previously unknown AMP and confirmed the absence of α-defensins [94]. In a similar vein, this group later developed an ANN model in 2009 to screen a larger number of synthetic AMP candidates, characterizing 18 sequences with high antimicrobial efficacy against multi-drug resistant bacteria [89]. In 2011, Wang et al. utilized a combination of sequence alignment and feature selection methods to design a computational model to more accurately classify AMPs [95]. Similarly, Torrent et al. trained an 8-descriptor SVM to classify AMPs with $75 - 90\%$ accuracy while taking into account new factors like peptide aggregation [96]. In 2013, Maccari et al. utilized RF models to design and validate the antimicrobial activity of two natural peptides and one peptide with non-natural amino acids [97], and Xiao et al. designed a two-level classifier to first classify peptide sequences as an AMP, and then sub-classify them into 10 functional AMP categories [98]. In 2015, Giguere et al. used a kernel method based on graph theory to train a 100 peptide dataset based on multiple measures of bioactivity to predict novel candidates [99]. Most recently in 2017, Schneider et al. reported the first application of unsupervised-supervised two-step models to classify AMPs. They used self-organizing maps to apply nonlinear dimensionality reduction to the training data, which was then used as an input for a supervised neural network model [100]. Together,

these studies highlight a diversity of methods and approaches that have been used to classify and design AMPs with great success.

1.6 Summary

From the discussion, it is evident that membrane curvature induction and membrane destabilization by antimicrobial peptides (and many other classes of peptides and proteins) all share a common molecular basis of activity. Specifically, the topological changes induced in membranes is a byproduct of their physical chemistry. In Chapters 2-3, we examine this interplay between chemistry and geometry using a machine-learning based QSAR approach. We develop a machine-learning algorithm to detect physicochemical features that underpin membrane activity, rather than antimicrobial potency. We use this search tool to discover hidden membrane activity in existing classes of multifunctional peptides and proteins.

CHAPTER 2

Mapping membrane activity in undiscovered peptide sequence space using machine learning

2.1 Abstract

There are some ∼1,100 known antimicrobial peptides (AMP), which permeabilize microbial membranes but have diverse sequences. Here, we develop a support vector machine (SVM)-based classifier to investigate α-helical AMPs and the interrelated nature of their functional commonality and sequence homology. The SVM is used to search the undiscovered peptide sequence space and identify Pareto-optimal candidates that simultaneously maximize the distance σ from the SVM hyperplane (thus maximize its "antimicrobial-ness") and its α-helicity, but minimize mutational distance to known AMPs. By calibrating SVM machine learning results with killing assays and small-angle X-ray scattering (SAXS), we find that the SVM metric σ correlates not with a peptide's minimum inhibitory concentration (MIC), but rather its ability to generate negative Gaussian membrane curvature. This surprising result provides a topological basis for membrane activity common to AMPs. Moreover, we highlight an important distinction between the maximal recognizability of a sequence to a trained AMP classifier (its ability to generate membrane curvature) and its maximal antimicrobial efficacy. As mutational distances are increased from known AMPs, we find AMP-like sequences that are increasingly difficult for nature to discover via simple mutation. Using the sequence map as a discovery tool, we find a unexpectedly diverse taxonomy of sequences that are just as membrane-active as known AMPs, but with a broad range of primary functions distinct from AMP functions, including endogenous neuropeptides, viral fusion proteins, topogenic peptides, and amyloids. The SVM classifier is useful

as a general detector of membrane activity in peptide sequences.

2.2 Introduction

The ∼1,100 known antimicrobial peptides (AMPs) [1, 2, 3, 4, 5, 6] are known collectively to have broad spectrum antimicrobial activity [1, 3, 5] via nonspecific interactions to target generic features in the many pathogen membranes [41, 1]. Machine learning can in principle be used to help discover the "blueprint" for natural AMP sequences; however such an enterprise presents significant structural difficulties. AMPs do not share a common core structure, but tend to be short (< 50 amino acids), cationic (+2 to +9), and amphiphilic [1, 2, 3, 4, 5, 6]. One of the principal components of AMP activity involves the selective permeabilization of microbial membranes [1, 2, 3, 5, 101, 102, 103, 47]. However, there is increasing evidence that membrane activity is but one of several modes of antimicrobial activity: Translocated AMPs can interact with intracellular targets to inhibit cell wall synthesis, nucleic acid synthesis, protein synthesis, and enzymatic activity [104, 105, 106, 107, 108]. Recent studies have shown that AMPs can be immunomodulatory [109, 110]: In fact, LL-37 plays a role in autoimmune disorders such as lupus and psoriasis [110]. These confounding factors make it difficult to implement adaptive learning for AMPs.

Prior AMP machine learning studies have focused primarily on empirical quantitative structure activity relationship (QSAR) models to evaluate large pools of candidate sequences and identify AMP candidates with improved minimum inhibitory concentrations (MIC) [111, 89, 91]. QSAR models for AMP discovery use a variety of statistical learning approaches, including multiple linear regression (MLR), linear discriminant analysis (LDA), principal component analysis (PCA), partial least squares regression (PLSR), artificial neural networks (ANN), support vector classifiers (SVC)/support vector machines (SVM), quantitative matrices (QM), hidden Markov models (HMM), and random forests (RF) [91, 112, 90, 113]. Lata et al. developed ANN, SVC, and QM models based on an analysis of the C and N-terminal residues in 486 antibacterial peptides [92]. Porto et al. reported the development of the CS-AMPPred predictor of cysteine-stabilized AMPs based on a SVC

trained over 5 physicochemical descriptors [111]. Fjell et al. developed a 44-descriptor ANN model to screen ∼100,000 candidates and produce 18 peptides that showed high activity against drug resistant bacteria [89]. Torrent et al. trained an 8-descriptor AMP and SVC to predict AMP sequences [96]. Maccari et al. employed RFs to design and validate the antimicrobial activity of two natural peptides, and one peptide incorporating non-natural amino acids [97]. Fjell et al. used an HMM to screen the bovine genome for AMPs, one of which was discovered in bovine intestinal tissue [94]. Chersakov et al. iteratively synthesized thousands of 9-residue peptides and trained ANN models to discover peptides with activities against drug-resistant superbugs [93]. The present work is similar in scope to these works—most notably those of Fjell et al. [89] and Cherkasov et al. [93]—in that we train QSAR models using limited experimental data to perform high-throughput virtual screening and identification of promising peptides for experimental synthesis and testing. Specifically, we train the SVM to recognize α-helical AMPs (Figure 2.1A) [114, 115, 116, 117, 118], so our work is cognate with computer-assisted antimicrobial peptide discovery and design [112, 119, 120], which have resulted in clinical trials of synthetic AMPs [112]. The goals of this work are quite different from the above. While a QSAR model with good predictive performance is expected to be able to identify physical determinants of AMP activity [112, 119, 120], our primary aim is not to use QSAR classifiers to find AMPs with improved activity, but rather use computational modeling in conjunction with calibrating experiments to examine the interrelated nature of AMP functional commonality, AMP sequence homologies, and general physicochemical mediators of AMP function at multiple length scales. We believe this approach will help uncover new unifying relationships between the discouragingly diverse peptide taxonomies that currently exist.

Here, we focus our study on α-helical AMPs, which have a structure common to many peptides and protein motifs. Specifically, we use our SVM to guide traversal and mapping of the undiscovered peptide sequence space. To help navigate this enormous space, we use known AMP sequences as "landmarks" and construct a "Pareto frontier" using the concept of Pareto optimality from microeconomics [121, 122] as the subset of sequences that simultaneously maximize the probability that the sequence is antimicrobial (the distance from

the SVM hyperplane σ) and the degree of α-helical secondary structure, and minimize the mutational distance to known AMPs (Figure 2.1B). Using a combination of killing assays and SAXS experiments on synthesized test peptide sequences that are not previously-known AMPs, we find a strong correlation between the SVM distance-to-hyperplane σ and the ability for peptides to generate negative Gaussian curvature (NGC) in model membranes. Since NGC is the type of membrane curvature topologically required for common membrane permeation mechanisms such as pore formation, blebbing, and budding (Figure 2.1C), it provides a structural basis for this common component of AMP activity. Using the SVM sequence map as a discovery tool for membrane activity, we increase mutational distances from known AMP sequences, probing the sequence space that is increasingly difficult for nature to discover via simple mutation from existing AMP sequences. What emerges is a diverse taxonomy of sequences that are expected to be not only just as membrane-active as known AMPs, but also have a broad range of putative primary functions beyond antimicrobial activity. We highlight several families, including endogenous neuropeptides, viral fusion proteins, topogenic peptides, and amyloids. Had their primary functions been undiscovered, these peptides could have been classified as AMPs. Not only is membrane activity not co-extensive with antimicrobial activity, it is surprisingly common for many classes of natural peptides as one component of multiplexed intracellular functions. Moreover, the calibrated SVM we construct is an efficient discovery tool to identify and discover membrane-active sequences.

2.3 Results and Discussion

2.3.1 SVM Model Development, Monte-Carlo Sampling, and Pareto Analysis

We constructed a linear support vector machine (SVM) classifier to predict whether a candidate peptide is likely to be antimicrobial. The SVM was trained by cross-validation over a balanced training set comprising 243 α-helical AMPs and 243 α-helical decoy peptides derived from natural and synthetic sources and a variety of microbial and multicellular species, and a balanced blind test set of 43 AMP and 43 decoy α-helical peptides (Figures

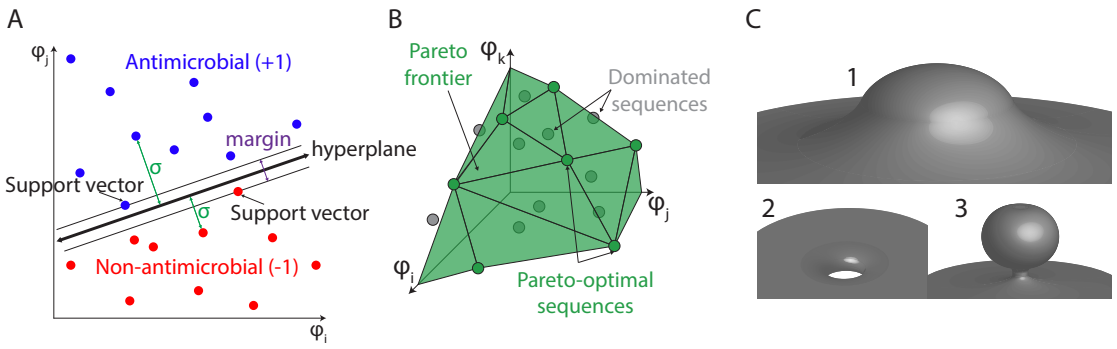


Figure 2.1: **Support vector machine learning and Pareto-optimization select for antimicrobial and membrane curvature-generating peptides.**
(A) Schematic depicting the use of a support vector machine binary classifier to partition hypothetical antimicrobial peptide sequences (blue circles) described by the $n = 2$ descriptors (ϕ_i, ϕ_j) from non-antimicrobial sequences (red circles) using a $(n-1)$-dimensional maximum-margin linear hyperplane. The support vectors are the sequences lying on the margins. The separating hyperplane lies midway between the margins. The metric σ (green arrows) indicates the distance-to-hyperplane for each peptide. Positive distances denote antimicrobial sequences while negative distances denote non-antimicrobial sequences. (B) Schematic demonstrating separation of Pareto-optimal sequences (green circles) from dominated sequences (gray circles) in an arbitrary three-dimensional subspace of descriptors. The Pareto frontier is the hypersurface containing the Pareto-optimal sequences. (C) Common biologically relevant manifestations of negative Gaussian curvature generation in cell membranes, including 1) blebbing, 2) pore formation, and 3) scission and budding.

2.2, 2.3). Variable selection was used to identify from a library of 1588 physicochemical descriptors (Table 2.3) a subset of 12 descriptors used to make classifications (Table 2.1). The classifier demonstrated excellent performance against the blind test set, with a prediction accuracy of 91.9%, specificity of 93.0%, and sensitivity of 90.7%.

Table 2.1: **Subset of 12 bagged descriptors identified by L1-SVC variable selection procedure.**
For each descriptor we provide a description of its physical interpretation (cf. Table 2.3 and Ref. [123]), and its weight in the linear SVC trained over the training data employing only these 12 descriptors. Positive (negative) weights correspond to a positive (negative) association of the descriptor with antimicrobial activity, and the magnitude of the weight indicates the relative importance of the Z-scored descriptor in the classification.

Rank	Descriptor	SVC Weight
1	netCharge	0.80
2	τ_2^G	0.48
3	p_{29}^G	0.36
4	SolventAccessD1025	-0.24
5	pc(M,K)	-0.21
6	p_{50}^G	0.20
7	AE	0.18
8	τ_4^G	-0.17
9	LW	0.17
10	NK	0.13
11	DP	-0.12
12	FC	-0.04

The trained SVM enables rapid computational screening of peptides for antimicrobial activity. Generation of the 12 descriptors and classification of all 572 AMP and decoy peptides in our training and test sets required only 80 s on a 2.13 GHz Intel Core 2 Duo processor, equal to 0.14 s per peptide. Nevertheless, comprehensive screening of all 10^{78} peptides of length 8-60 residues—the size range of training peptides—is intractable, so we instead traversed sequence space via a directed search according to the following four criteria: (i) Defined peptide length. The general mechanism of α-helical antimicrobial peptides involves a combination of electrostatic and hydrophobic interactions with negatively-charged bacterial cell membranes. Typical AMPs have a length of 20-25 amino acids, and generally generate membrane deformation by spanning the lipid bilayer and inducing opposite mem-

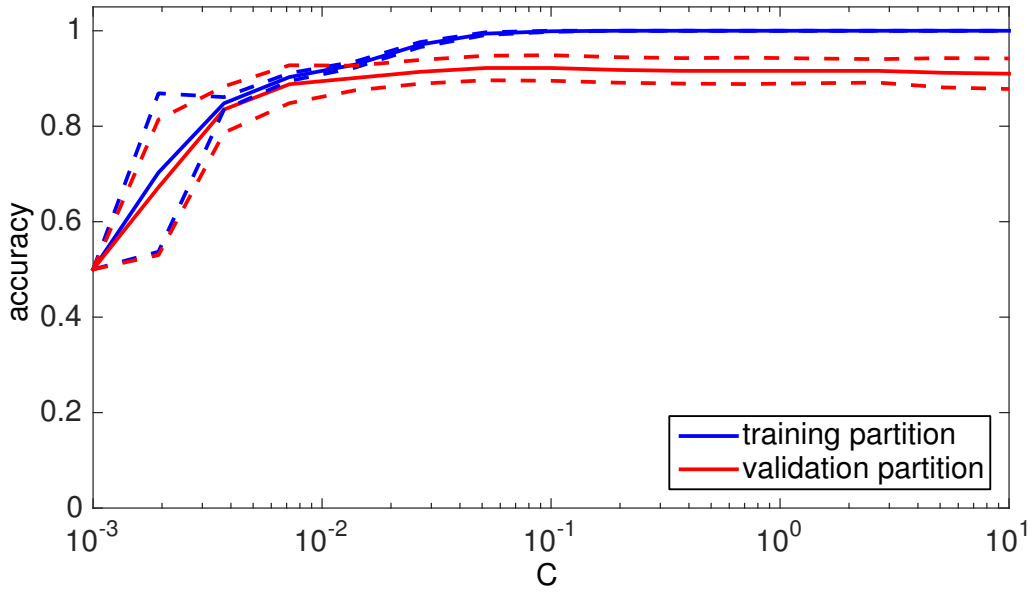


Figure 2.2: **Accuracy of the SVM.**
The accuracy of the L1-SVM over the 80% training partition (blue) and 20% validation partition (red) of the 486-peptide training set is plotted as a function of the hyperparameter C. The accuracy is defined as accuracy $= \frac{TP+TN}{TP+FP+TN+FN}$, where TP is the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives. The accuracy at each value of C was measured over $k = 15$ rounds of stratified shuffled cross-validation, and the mean (solid lines) and standard deviation (dashed lines) are reported. Our grid search followed by numerical maximization procedures identified an optimal value of $C_{opt} = 9.95 \times 10^{-2}$ with an associated cross-validation accuracy of 92.2%.

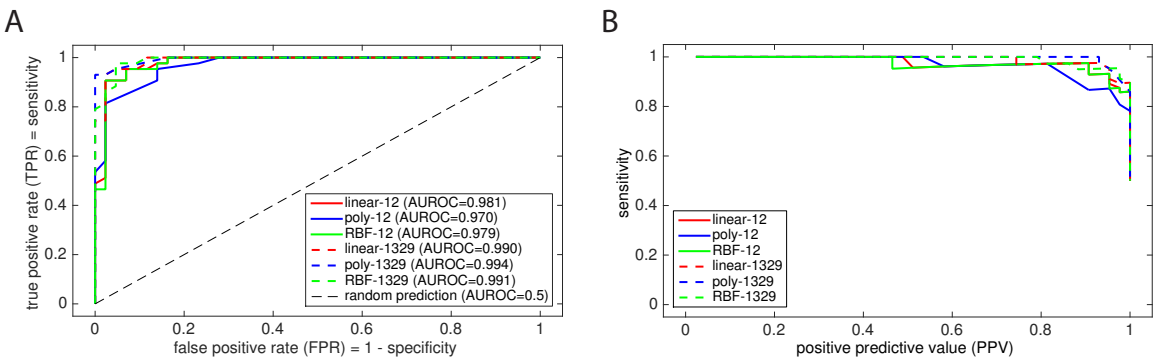


Figure 2.3: **Predictive performance of SVM.**
The performance of the SVM on the blind 86-peptide test dataset of six L2-SVCs employing three different kernels and trained on either the 1329 filtered descriptors or the 12-descriptor bagged subset listed in Table 2.1 as a function of the distance from the SVC hyperplane at which classification is performed. (A) The receiver operating characteristic (ROC) curve shows the trade-off between the true positive rate and false positive rate. (B) The precision-recall curve shows the trade-off between sensitivity (recall) and positive predictive value (precision).

brane curvature in orthogonal directions [2, 6, 124]. Accordingly, we restrict our screen to 20-25 residue peptides. (ii) Homology with a known AMP. Classifier accuracy is expected to diminish in regions of sequence space far from the AMP sequences upon which it was trained. Accordingly, we control within our screen peptide homology to known AMPs. (iii) Large positive distance from the SVM hyperplane. This criterion favors sequences for which the classifier possesses high positive predictive value (PPV, the expected proportion of positive results that are true positives) and specificity at the expense of sensitivity (left side of the receiver operating characteristic curve, right side of the precision-recall curve in Figure 2.3). In driving the false positive rate towards zero—at the expense of a high false negative rate—we focus our screen towards the most promising candidates. (iv) High helical content. Trained on α-helical peptides, classifier accuracy is expected to diminish for peptides with non-helical structure. Accordingly, we seek candidates that are predominantly helical. (We note that certain α-helical AMPs only adopt a helical structure upon interacting with the cell membrane [2, 6], and this criterion will necessarily disfavor such candidates.)

Using these criteria, we performed a guided search of sequence space to generate a subsampling of 242,110 candidates using two complementary approaches. To traverse the sequence

space with close homology to existing AMPs, we generated 33,079 sequences corresponding to all one-point mutants of the 76 AMPs in the database within the size range 20-25 residues. To probe the sequence space of AMPs that are less likely for nature to discover via simple mutation, we generated 208,955 additional sequences from Monte Carlo sampling by initializing 10 independent Monte Carlo chains with a randomly selected AMP in the size range of interest, and ran 25,000 rounds of random point mutation, insertion, and deletion. The 12 bagged descriptors were generated for each new trial sequence and the distance from the classifier hyperplane σ computed using the SVM classifier. Treating $(-\sigma)$ as an effective energy to bias the search towards promising candidates, trial sequences are accepted or rejected according to the Metropolis criterion $p_{acc} = min(1, e^{\frac{\Delta\sigma}{T}})$, where $\Delta\sigma = \sigma_{trial} - \sigma_{current}$ and T is an effective temperature [125, 126, 127]. In this work, we found that $T = 0.8$ provides a good compromise between focusing the search towards large σ candidates while also providing good sampling. All unique candidates were saved for the computational screen. Finally, the helical content of all 242,110 candidates were evaluated using the ab initio secondary structure prediction algorithm PSIPRED (`http://bioinf.cs.ucl.ac.uk/psipred/`) [128, 129] implemented in PROTEUS2 (`http://www.proteus2.ca/proteus2/`) [130]. We do not actively direct sampling towards high helical candidates, but identify them post hoc. We note that more comprehensive sampling of design space could be achieved using umbrella sampling in our Monte Carlo procedure [131].

After conducting this directed search, we wished to identify optimal sequences within our candidates. To do this, we borrowed the concept of Pareto optimality from microeconomics [121, 122] and employed multi-objective optimization to determine the Pareto optimal sequences that dominate all other candidates in simultaneously maximizing the distance from the SVM hyperplane and the degree of α-helical secondary structure, and minimizing the mutational distance to a known AMP. These sequences are optimal in the sense that no other candidates exist for which any one criterion can be improved without degrading at least one other. We term this Pareto frontier "physicochemically unrestricted" since we place no restriction on the value of the 12 bagged descriptors. It is possible that this Pareto set may contain candidates with descriptor values far outside the range of the training data. Accord-

ingly, we constructed a "physicochemically restricted" Pareto frontier constrained such that none of the 12 descriptors could lie more than 10% outside the range observed in the training set [89]. Together, these two frontiers serve as guides for our exploration of the sequence space, and our subsequent interpretation of discovered sequences.

We present a 3D scatterplot of all 242,110 candidate sequences (gray crosses) plotting the distance to the SVM hyperplane σ, minimum Jukes-Cantor distance to a known AMP $minHomologyAMP$, and predicted fractional helicity (Figure 2.4). Two-dimensional projections of the point cloud are presented in Figures 2.12 and 2.13. The candidate sequences partition into two point clouds. The cloud at low $minHomologyAMP$ corresponds to candidates most homologous to known AMPs comprising the 76 AMPs in the database within the size range of interest plus all of their point mutants. The cloud at high $minHomologyAMP$ corresponds to sequences generated by the directed Monte-Carlo search biased towards candidates with a large distance from the SVM hyperplane σ. Importantly, the latter may contain sequences with divergent physicochemical properties from known AMPs that are nonetheless predicted by our classifier to possess antimicrobial activity. We highlight the 13 sequences residing on the Pareto frontier of physicochemically restricted sequences (green diamonds), and the 85 sequences residing on the Pareto frontier of physicochemically unrestricted sequences (orange diamonds). We also indicate 16 peptides (red stars, Table 2.2) close to both Pareto frontiers that we selected to synthesize and test with SAXS and killing assays.

2.3.2 Distance to hyperplane of known AMPs does not correlate with antimicrobial efficacy

To engage the question of whether our SVM model can predict the efficacy of known AMPs, we analyzed a standardized database of MIC values of 478 known AMPs against Staphylococcus aureus collated from the literature (`https://www.antistaphybase.com/index.php`), and calculated their distance to the SVM hyperplane σ (Figure 2.5). We plotted their reported MIC values against σ and found poor and statistically insignificant correla-

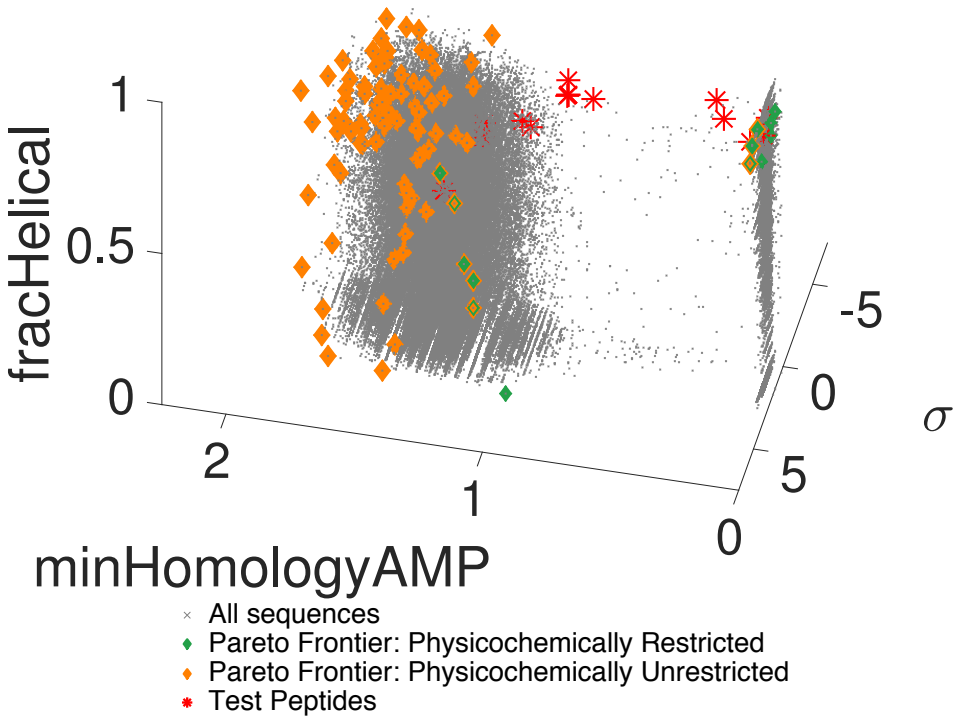


Figure 2.4: **Sequence atlas and Pareto frontier constructed by directed sampling of sequence space.**
Embedding of the 242,110 peptides generated by our directed sequence space search into the 3D space spanned by 1) predicted helical content, 2) Jukes-Cantor distance to known AMPs, and 3) distance-to-hyperplane (σ). Sequences with $\sigma > 0$ are predicted by the classifier to be antimicrobial or membrane active, whereas those with $\sigma < 0$ are not. The more positive σ becomes, the higher probability of being antimicrobial $P(+1)$. The orange diamonds highlight the 85 peptides lying on the physicochemically unrestricted Pareto frontier in which we place no restriction on the value of the descriptors generated for these candidates. Green diamonds highlight the 13 peptides on the physicochemically restricted Pareto frontier in which the descriptors are restricted to lie no more than 10% outside the range observed in the training data. Red stars are the 16 peptides proximate to the frontiers that were selected for experimental testing.

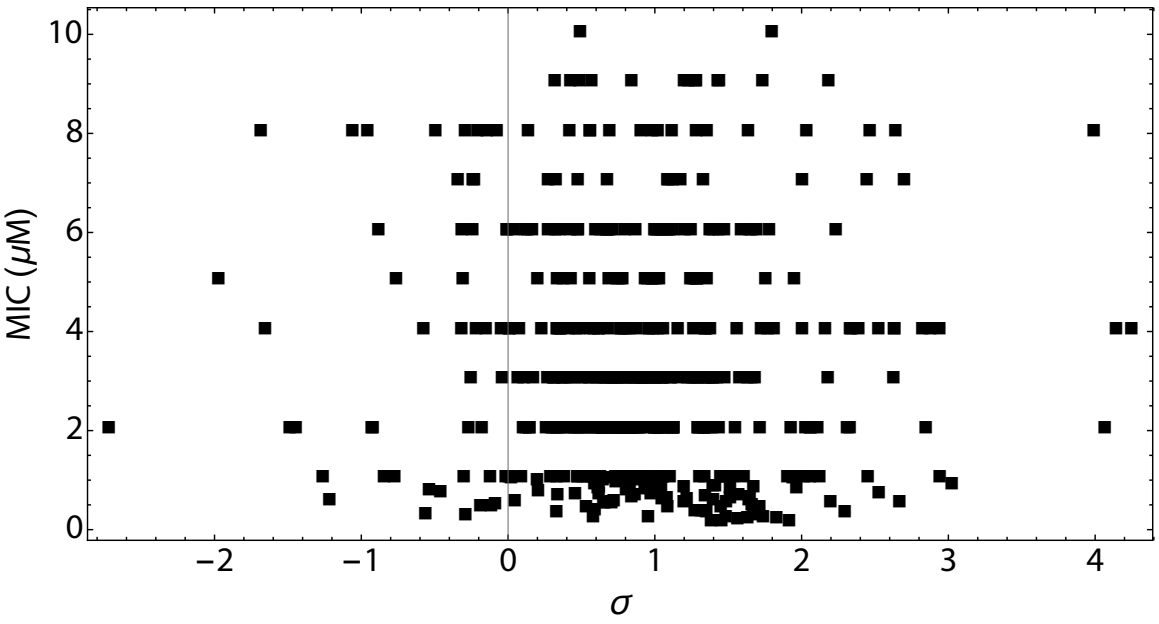


Figure 2.5: **Distance-to-hyperplane of known AMPs does not correlate with antimicrobial efficacy.**
We screened an online database of known AMPs with measured antimicrobial efficacy against *S. aureus* ($N = 478$). We plot the minimum inhibitory concentration (MIC) of each peptide against the SVM predicted distance to hyperplane σ. We find that there is no significant correlation between antimicrobial efficacy and predicted distance to hyperplane, indicating that antimicrobial efficacy of an AMP is uncorrelated with the recognizability of an AMP by our classifier ($R_{Spearman} = -0.060$ [-0.154, 0.034], $p = 0.187$).

tion ($R_{Spearman} = -0.060$ [-0.154, 0.034], $p = 0.187$). Analysis of the literature suggests that the majority of the known AMPs that we tested are compounds with multiplexed activities in addition to membrane penetration. This observation highlights a generic problem with machine learning approaches: High classification accuracy does not necessarily imply understanding or even amenability to traditional forms of understanding. Although it is known that AMPs can have immunologically relevant activity outside of membrane activity, there is currently no general way to identify AMPs with additional functions. We propose a way to identify candidate sequences with multiplexed functions in the next section.

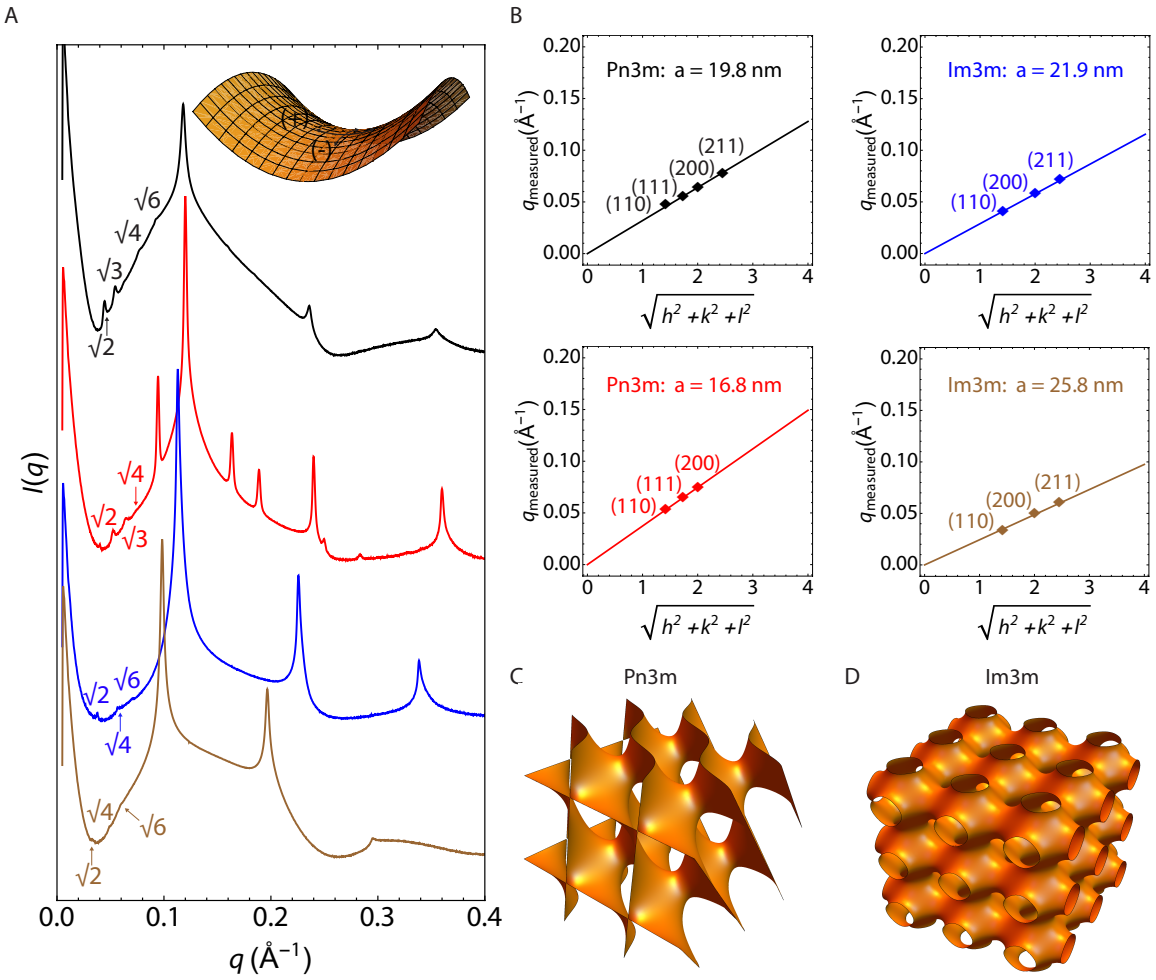


Figure 2.6: **Synthesized test peptides from directed Monte-Carlo search and SVM screening generate negative Gaussian curvature in model membranes.** (A) Representative SAXS data of four test peptides indicate ability to generate negative Gaussian curvature in model bacterial membranes. Peaks with cubic symmetry are labeled according to their x-coordinates $\sqrt{h^2+k^2+l^2}$ in (B). Unlabeled peaks correspond to coexisting lamellar and/or hexagonal phases induced by peptides. Inset depicts local topology of saddle-splay curvature. (B) Linear fits indicating the q positions of the Bragg peaks with cubic symmetry, their respective Miller indices (hkl), their respective space groups, and resulting lattice parameter a. (C) Contour surface representation of the Pn3m space group. (D) Contour surface representation of the Im3m space group.

Table 2.2: **Measured physicochemical parameters and SVM classification of 16 synthesized positively-classified test peptides near the Pareto frontier.** $|\langle k \rangle|$ is the experimentally measured average magnitude of negative Gaussian curvature in a unit cell induced by the peptide in model membranes. σ is the distance to hyperplane calculated from the SVM. $P(+1)$ is the probability of being antimicrobial, and $P(-1) = 1 - P(+1)$ is the probability of being non-antimicrobial. minHomologyAMP denotes the minimum mutational distance of the given sequence to any known AMP. fracHel is the helical content of the peptide.

Peptide	$\lvert\langle k \rangle\rvert$ (10^{-4} Å^{-2})	σ	$P(-1)$	$P(+1)$	minHomologyAMP	fracHel
1	2.5	1.25	0.016	0.984	0.2	0.864
2	1.113	1.27	0.015	0.986	0.71	0.864
3	2.224	1.22	0.017	0.983	0.81	0.864
4	2.131	1.2	0.018	0.982	0.81	0.864
5	1.994	1.24	0.016	0.984	0.81	0.864
6	1.458	1.29	0.014	0.986	0.81	0.864
7	1.543	1.15	0.022	0.978	0.81	0.909
8	0	0.93	0.043	0.957	0.04	0.87
9	0	0.36	0.227	0.773	0.25	0.87
10	4.03	1.63	0.005	0.996	0.98	0.778
11	1.655	2.17	0	1	0.08	0.853
12	3.289	1.95	0	1	0.03	0.867
13	2.242	2.35	0	1	1.27	0.55
14	1.615	1.01	0.034	0.966	1.15	0.667
15	1.22	0.72	0.084	0.916	1.15	0.7
16	2.382	1.44	0.009	0.992	0.95	0.75

2.3.3 Predicted distance-to-hyperplane of synthesized test peptides correlates with strength of negative Gaussian curvature generation

The distance-from-hyperplane σ possesses a clear mathematical interpretation as the distance of a candidate peptide from a hyperplane in the 12-dimensional space of the bagged descriptors (Table 2.1). Nevertheless, the high dimensionality of the space, opaque nature of some descriptors, and absence of a mechanistic model linking peptide sequence to function makes it challenging to assign physical interpretability to this discriminatory metric. To glean physical understanding, we synthesized peptide sequences with defined values of σ and assayed the peptide-membrane interactions. Specifically, we selected for synthesis 16 candidate sequences identified by the directed search procedure according to the following criteria: they did not feature in the AMP database, were classified by the SVM as antimicrobial, were predicted to have $> 50\%$ α-helical content by the secondary structure prediction algorithm PSIPRED [128, 129, 130] and were confirmed to possess antimicrobial activity *in vitro* (Table 2.2). Since a common contribution to antimicrobial activity is membrane

permeation, we investigated this aspect using SAXS. In our prior work, we have shown that the ability of a peptide to generate NGC in model membranes is an excellent proxy for antimicrobial activity through membrane-permeating properties [124, 47, 41]. We incubated these test sequences with small unilamellar vesicles (SUVs) mimicking bacterial membranes (compositions PS/PE = 20/80 and PG/PE = 20/80) under physiological salt and pH (100 mM NaCl, 10 mM HEPES, pH 7.4) at specific peptide-to-lipid charge ratios, and measured the resulting peptide-induced membrane curvature quantitatively using SAXS. We find that 14 of the synthesized peptides reorganize membranes into cubic phases rich in NGC (Figure 2.6A (inset), C, D). We show representative SAXS patterns for four of the test peptides, with labeled Bragg peaks corresponding to their cubic phases (Figure 2.6A). The best fit lines to determine the lattice parameter a of the cubic phases are shown, along with the Miller indices (hkl) of each cubic reflection (Figure 2.6B). We find these peptides typically reorganize model membranes into either Pn3m or Im3m cubic phases (Figure 2.6C, D). The amount of induced membrane curvature in these phases can be directly compared via calculation of the NGC from the lattice parameter and the Euler characteristic (see SAXS methods). The same procedure was carried out for the other ten sequences that generated cubic phases in these membranes. They also generated either Pn3m or Im3m phases (Figure 2.6C, D). The NGC values calculated from SAXS data (Table 2.2) indicate that these peptides generate similar quantities of NGC as natural AMPs, which already strongly suggest that these predicted peptides permeabilize membranes. To confirm that our classifier can also recognize decoy peptides that do not generate NGC, we selected 3 negatively classified peptides to synthesize and test with SAXS. We find that none of them generate NGC-rich cubic phases in the same membranes tested with the 16 test peptides (Figure 2.7). To test the ability of our classifier to detect peptide sequences lacking antimicrobial activity, we also screened 5 non-active granulysin fragment peptides from literature. We find that these peptides have low or negative values of σ and are far from the Pareto frontier, which is consistent with their inability to generate NGC. Our algorithm also correctly classifies the two active granulysin fragments [132], which are known to be membrane permeating.

To test the hypothesis that the SVM classifier has learned to discriminate peptides based

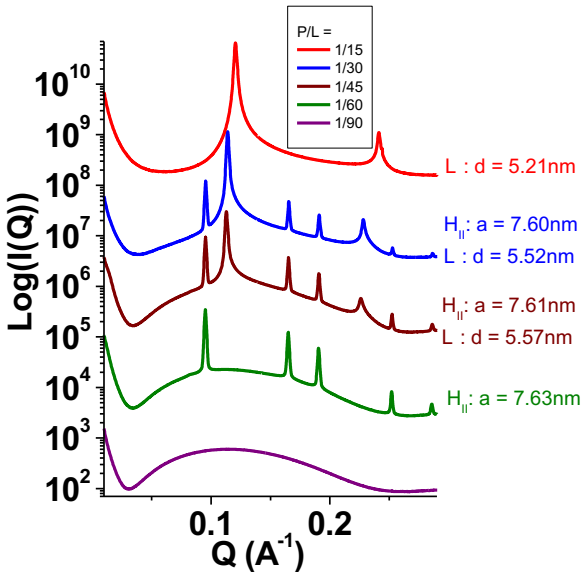


Figure 2.7: **Negative control peptides do not generate cubic phases in model membranes.**
We show a SAXS scattering curve of a negative control peptide #17 incubated with PS/PE 20/80 small unilamellar vesicles at 100 mM NaCl, 10 mM HEPES pH 7.4 at various P/L molar ratios. We find that the peptide generates non-NGC phases such as lamellar (L) and hexagonal structures (H), both of which have Gaussian curvatures of zero. This demonstrates that these control peptides do not generate NGC. The SAXS curves for all of the negative control peptides exhibit similar phase behavior. Values of $|\langle k \rangle| = 0$ indicate inability to generate NGC.

on their capacity to generate NGC, we computed the Spearman rank correlations of the magnitude of NGC calculated from the SAXS data for the 16 selected peptides with metrics utilized in the SVM classification algorithm (Table 2.2). We asked whether the magnitude of NGC correlates with the homology to known AMPs and/or the distance to hyperplane σ (Figure 2.8). We find that there is no significant correlation ($R_{Spearman} = 0.155$ [-0.425, 0.736], $p = 0.565$) between the magnitude of NGC generated by a peptide and its homology to a known AMP (Figure 2.9A). This finding implies that sequence homology to a known AMP is not a necessary requirement to generate NGC. In other words, sequences that are far from known AMPs via simple mutation have potential to generate membrane curvature. Given this result, we ask whether there is a correlation between σ and NGC ($|\langle k \rangle|$). In Figure 2.5, we showed that σ correlates poorly with antimicrobial efficacy. Here we observe a strong, statistically significant positive correlation between the distance to hyperplane of a peptide, and its ability to generate NGC ($R_{Spearman} = 0.653$ [0.234, 0.891], $p = 0.006$) (Figure 2.9B). This provides strong support for the hypothesis that our SVM classifier has learned to discriminate peptides based on their capacity to generate NGC. Since the distance to hyperplane σ also provides a measure of the confidence of the algorithm $P(+1)$ in whether a sequence is antimicrobial or not, we conjecture that a higher confidence in classification may correlate with an enhanced ability to generate NGC. Looking at the 16 sequences that were identified by the SVM as high-probability hits ($P(+1) > 0.95$), we find a strong positive and statistically significant correlation between the confidence in the prediction of the SVM $P(+1)$, and the magnitude of NGC ($R_{Spearman} = 0.653$ [0.231, 0.896], $p = 0.006$) (Figure 2.9C). This result makes sense since $P(+1)$ and σ are monotonically but nonlinearly related. This result has strong implications for the utility of our SVM classifier—we can now stratify the predicted quality of sequences by using the σ metric as a surrogate for expected NGC generating ability to efficiently explore sequence space and predict peptides with high membrane activity. Importantly, sequences less homologous to known AMPs with large values of σ that fall near, but not necessarily on, the Pareto frontier can generate the same level of membrane curvature expected of prototypical α-helical AMPs (such as Magainin, $|\langle k \rangle| = 2.536 \times 10^{-4} \text{Å}^{-2}$), and candidate AMPs spanning a large range of helicities and

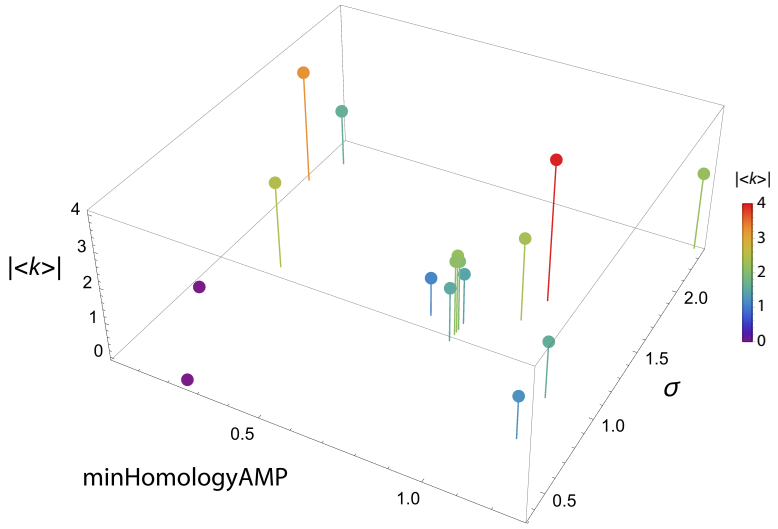


Figure 2.8: **Visualization of NGC generation by test peptides relative to distance to margin and homology to known AMPs.**
Tested sequences demonstrate positive correlation with σ, but no correlation with $minHomologyAMP$. The colors of the dots are chosen based on the magnitude of NGC ($|\langle k \rangle|$) generation, and stems to the x—y plane demonstrate where they fall along the selection rule. The units of $|\langle k \rangle|$ are 10^{-4}Å^{-2}. The sequences with larger NGC tend to fall at higher σ. The 2D projection into the σ—$minHomologyAMP$ plane is shown in Figure 2.9A while the 2D projection into the σ—$|\langle k \rangle|$ plane is shown in Figure 2.9B.

homologies to known AMPs can also generate magnitudes of NGC similar to those of known AMPs (Figure 2.10, Table 2.2). Accordingly, we expect to be able to use σ to predict the membrane activity of peptide families that may be very dissimilar to AMPs. We also investigated the relationship between amphiphilicity and ability to generate curvature in membranes. Interestingly, we find that 4 out of the 12 descriptors in our final SVM model enforce amphiphilicity in positively-classified peptides (#2, 3, 6, 8 in Table 2.1, Figure 2.11A), suggesting that the SVM encodes amphiphilicity implicitly in its selection criteria. To quantitatively compare amphiphilicity with $|\langle k \rangle|$, we calculate the mean hydrophobic moment of the test peptides and find that they compare favorably to known helical AMPs (Figure 2.11B) despite having large mutational distances from known AMPs. We find that amphiphilicity is highly correlated with the ability to generate NGC ($R_{Spearman} = 0.680$, $P = 0.0038$).

These findings strongly support the hypothesis that our SVM classifier has learned to

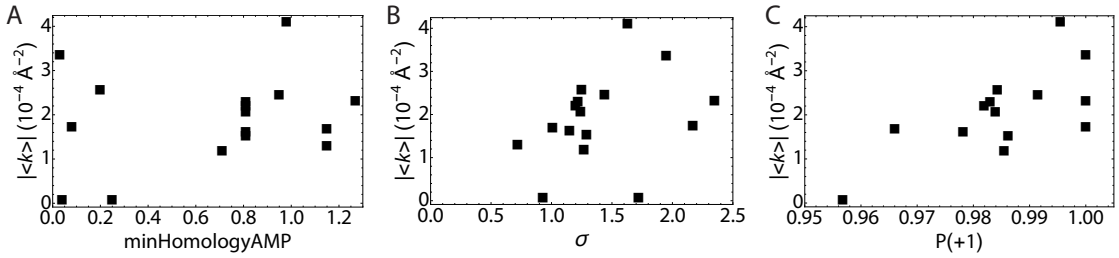


Figure 2.9: **Distance-to-hyperplane of test peptides does correlate with strength of negative Gaussian curvature.**
There is no significant correlation between the magnitude of NGC generation and homology of test peptides ($N = 16$) to known membrane-active peptides (A, $R_{Spearman} = 0.155$ [-0.425, 0.736], $p = 0.155$), but there is a statistically significant (B, $R_{Spearman} = 0.653$ [0.234, 0.891], $p = 0.006$) positive correlation between the magnitude of NGC generation and distance-to-hyperplane σ, as well as the probability of being antimicrobial (C, $R_{Spearman} = 0.653$ [0.231, 0.896], $p = 0.006$). This validates the use of σ as a proxy for optimization of curvature-generation as opposed to antimicrobial efficacy (Figure 2.5).

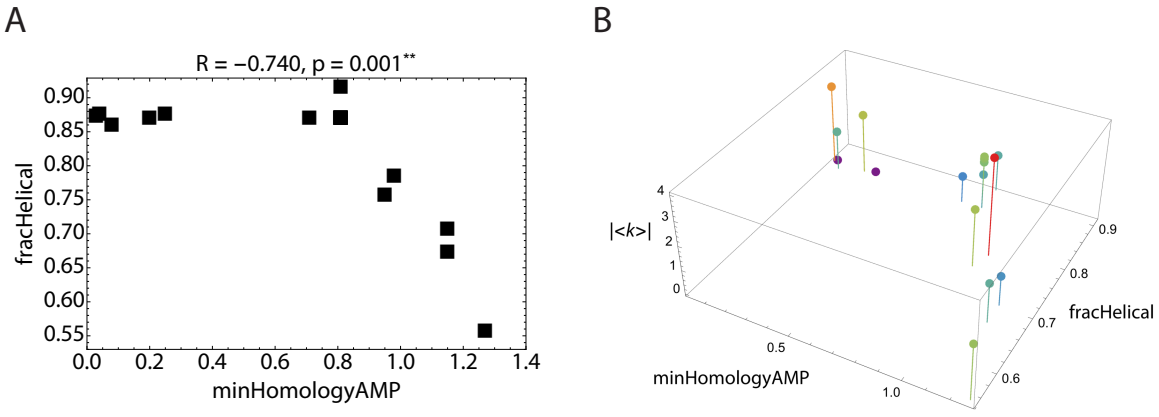


Figure 2.10: **Test peptides spanning a large range of helicities and homologies to known AMPs can generate magnitudes of NGC similar to those of known AMPs.**
We plot the helicities and homology to known AMPs of the peptides tested with SAXS ($n = 16$). There is a tradeoff between helical nature of peptide and its homology to known AMPs (A, $R_{Spearman} = -0.750$, $p = 0.001$) but there is consistent generation of NGC no matter where the peptides fall along the curve (B). This demonstrates that peptides far in mutational distance and a large range of helicities can generate NGC just as well as known AMPs. We observe that helicity is not a strong determinant of NGC generation $|\langle k \rangle|$ ($R_{Spearman} = -0.331$, $p = 0.210$) compared to σ ($R_{Spearman} = 0.653$, $p = 0.006$, Figure 2.9B).

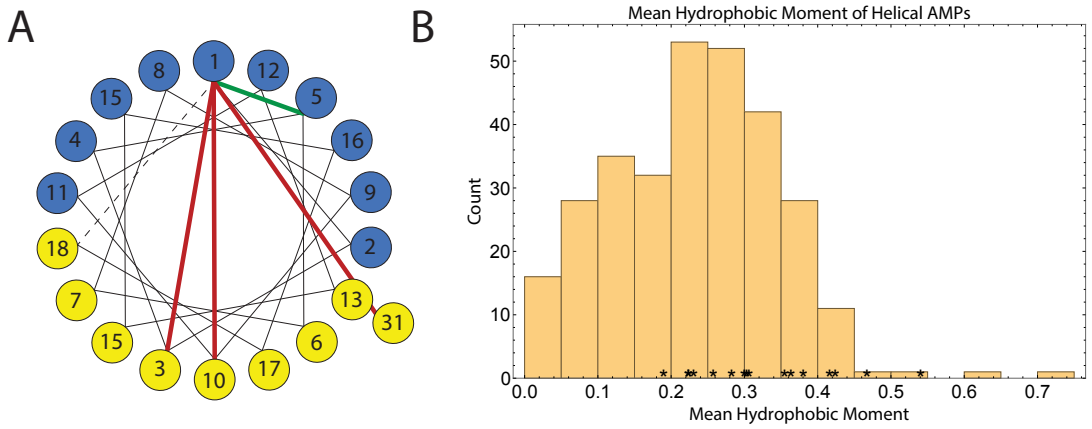


Figure 2.11: **Amphiphilicity is implicitly encoded within the SVM model and is a feature of positively-classified test peptides near Pareto frontier and α-helical AMPs in general.**
(A) Helical wheel plot demonstrating the amphiphilicity encoded in physicochemical descriptors #2, 3, 6, and 8 in Table 2.1. Any two residues spaced apart by 2, 9 and 30 positions along the helical wheel likely have opposite amino acid character (e.g. polar and hydrophobic), while residues spaced apart by 4 positions along the helical wheel likely have similar characteristics (e.g. polar and polar, or hydrophobic and hydrophobic). Blue residues are polar and yellow residues are hydrophobic. Red lines indicate pairs with opposite character while the green line indicates pairs with similar character. (B) Histogram of the mean hydrophobic moments of α-helical AMPs used in the training dataset of this paper (using Eisenberg consensus hydrophobicity scale [133]). Typical α-helical AMPs have a mean hydrophobic moment between 0.2-0.5, and the black stars indicate the mean hydrophobic moments of the 16 peptides we tested with SAXS (Table 2.2). Thus, our test peptides have amphiphilicity similar to that of α-helical AMPs. Hydrophobic moments were calculated using vector addition of residues around a helical wheel, and the magnitude of the hydrophobic moment vector is the mean hydrophobic moment.

distinguish membrane-permeating from non-membrane-permeating α-helical peptides. This result illustrates a simultaneous potentiality and deficiency of our SVM, and of QSAR approaches in general. Our classifier has discovered membrane permeation activity as a highly recognizable feature of AMPs within the training set, and used it to identify such AMPs with high accuracy and efficiency. A limitation of the classifier is that it is therefore capable of identifying and indexing membrane activity, but not necessarily antimicrobial activity. Although this result may at first blush appear to be a shortcoming of our QSAR approach, it emphasizes the transformative potential of the SVM classifier: It yields combinations of physicochemical properties that describe any peptide that can mediate membrane activity through induction of NGC. By using α-helical AMPs as a bootstrap dataset to learn about membrane curvature generating sequences in general, regardless of their primary known function, we now have a general tool for screening peptides for membrane crossing or permeating activity.

2.3.4 Directed search of the sequence space of physicochemically restricted and unrestricted peptides discovers diverse families of membrane curvature-generating peptides

To test the capacity of our SVM classifier to identify peptide sequences with membrane activity, we compiled from the Protein Data Bank (PDB) 31 peptides belonging to diverse families of known and unknown function including viral peptides that attack membranes, intrinsically disordered peptides, and exocytosis/endocytosis related sequences, and used our SVM classifier to project them into our sequence map (Figures 2.12, 2.13). Interestingly, a number of sequences were found to reside near the Pareto frontiers, suggesting the existence of membrane activity within these candidates. Specifically, we found several neuropeptides (purple stars), calcitonin peptide hormones (black stars), viral fusion proteins (yellow stars), membrane anchor proteins (green stars), membrane-permeating protein fragments (blue stars), and topogenic peptides (pink stars) proximate to the Pareto frontiers (Figure 2.12). This is exciting for several reasons. Several neuropeptides have been shown to be antimicrobial *in vitro* [134, 135, 136], but this is an indication that many members

of the family can generate the kind of membrane curvature required for permeabilization. This alludes to possible intracellular targets of these neuropeptides and receptor-independent mechanisms of signal transduction in addition to their regular mode of activity. This observation reinforces a previously known structural tendency of AMPs, since several of them are known to have endocrine and homeostatic functions. For example, hepcidin permeates membranes but is also involved in regulation of iron [137] while α-melanocyte stimulating hormone has antimicrobial and anti-inflammatory effects in addition to its signaling properties [138]. Calcitonin is a peptide hormone involved in calcium homeostasis, but is also a known amyloid that deposits in medullary thyroid carcinoma. Other studies of amyloid proteins have demonstrated the ability to increase membrane permeability [139], and this may provide a physicochemical basis for this observation [140]. We have also previously described the role of the M2 proton channel in budding and scission of the influenza virus [77]. Using this algorithm, we find that a variety of other viral fusion proteins likely share similar characteristics, including peptide domains from medically relevant viruses like ebolavirus, HIV, coronavirus, and Hepatitis C. Other diverse discovered proteins include membrane-permeating protein fragments from enzymes, DNA-binding proteins, and prion precursors. Our analysis also identified helical membrane-active sequences from topogenic peptides, which are known membrane curvature generating proteins. These special signal sequences present at the N-terminal portions of newly translated proteins help target and translocate large proteins across intracellular membranes [141]. In general, it can be shown that the sequence content of the Pareto-optimal and peptide sequences from the newly identified taxonomies follow the same sequence trends as existing AMPs (Figure 2.14). This demonstrates that our algorithm can efficiently and effectively identify candidates that can reorganize bacterial membranes from a large sequence space.

2.3.5 Pareto-optimal sequences and newly discovered curvature-generating sequences follow the saddle-splay selection rule.

We have established that σ correlates with the induction of NGC (Figures 2.6, 2.9), and showed that the SVM classifier can be used to predict NGC ability for arbitrary peptides and

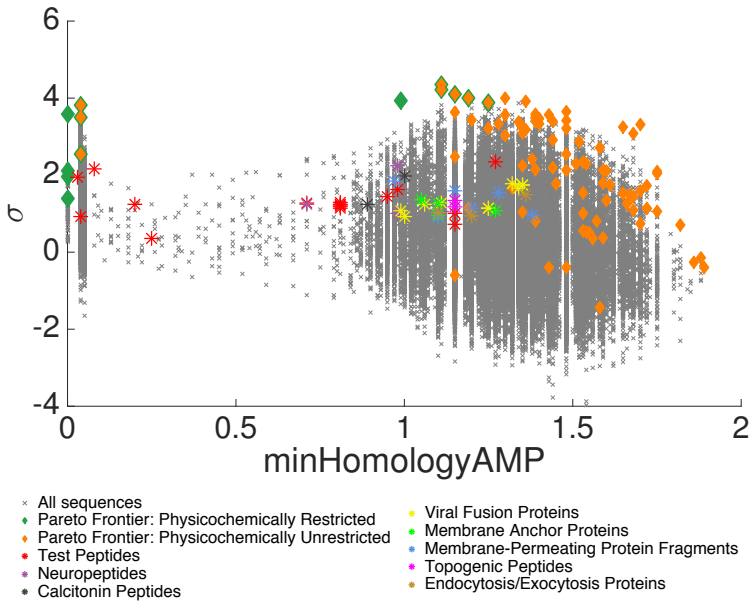


Figure 2.12: **Directed search of the sequence space discovers diverse families of membrane curvature-generating peptides.**
We visualize the 2D projection of the 242,110 candidate peptides generated by directed sampling of sequence space (Figure 2.4) into distance-to-hyperplane σ and Jukes-Cantor distance to known AMPs and supplemented by the 31 sequences belonging to diverse peptide families listed. To guide the interpretation of the discovered membrane-active sequences, we highlight the physicochemically restricted (13 peptides, green diamonds) and unrestricted Pareto frontiers (85 peptides, orange diamonds) For reference, the peptides experimentally tested are also shown (16 peptides, red stars). Screening of a variety of protein families yields sequences with predicted $\sigma > 0$ near the physicochemically unrestricted Pareto frontier. These sequences span a variety of protein families, including neuropeptides (purple stars), calcitonin peptides (black stars), viral fusion proteins (yellow stars), membrane anchor proteins (light green stars), membrane-permeating protein fragments (blue stars), and topogenic peptides (pink stars). Some of the proteins have unexpected predicted membrane activity, while others have confirmed experimental evidence for membrane permeation. In fact, these other classes of peptides are expected to be just as membrane active as AMPs. This demonstrates the power of the SVM directed search framework as a tool for discovery of new membrane reorganizing protein sequences.

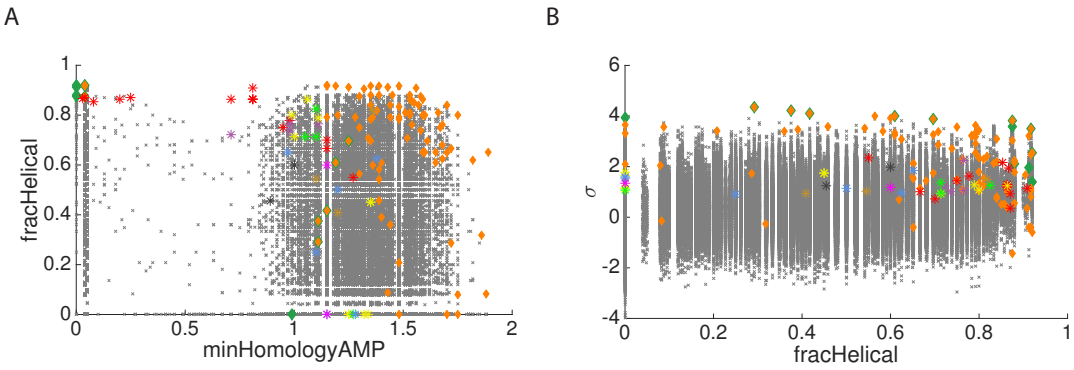


Figure 2.13: **2D projections of 3D sequence atlas and Pareto frontier.** Projection of the 3D Pareto plot of the 242,110 sequences generated by our directed sampling of sequence space (Figure 2.4, 2.12) under the 2D projections into (A) the fracHelical—minHomologyAMP plane and (B) the σ—fracHelical plane. The σ—minHomologyAMP projection is provided in Figure 2.12. We observe that the majority of the newly discovered membrane-active sequences have predicted fractional helicities larger than 0.2 positive values of σ, and large mutational distances from known AMPs such that they are unlikely to be explored by natural mutations within existing AMPs. The legend can be found in Figure 2.12.

also perform a directed search of sequence space to discover NGC-inducing peptides (Figure 2.12). We now investigate how the sequence content of the Pareto-optimal and newly discovered classes of membrane-active peptides compare to existing AMPs. Remarkably, the physicochemically restricted Pareto sequences (i.e., those with physicochemical descriptors no more than 10% outside the range observed in the training data) very closely follow the rule governing amino acid content previously discovered for AMPs [72] (Figure 2.14A, green diamonds). This "saddle-splay selection rule" delineates the tradeoffs between arginine, lysine, and hydrophobic content in order to retain membrane activity. We previously found that α-helical AMPs tend to be rich in lysine and hydrophobicity. This result also demonstrates the surprising result that our QSAR approach and SVM classifier trained solely on physicochemical descriptors "learned" a rule based on geometry and topology without human intervention.

For the physicochemically unrestricted Pareto optimal sequences (Figure 2.14B), we find that these sequences tend to segregate into two clusters, ones that follow the saddle-splay selection rule (orange diamonds, no border), and ones that fall below the curve (orange dia-

monds, black borders). These off-trend sequences suggest a greater plasticity in the peptide sequence and imply that these candidates may contain additional multiplexed functions in addition to membrane activity. In fact, comparing these Pareto sequences to the sequences of the newly discovered families shows a very similar distribution (Figure 2.14C). The majority fall along the trend (stars without borders), but have a subset that deviate with less lysine and more arginine content relative to their hydrophobic content (stars with black borders). This behavior makes sense since the purported functions of these proteins are more complicated than simply pore formation. For example, the topogenic peptides must penetrate through bilayers and deliver cargo rather than linger within membranes. Thus, the increased arginine content makes sense, since cell-penetrating peptides tend to have less hydrophobicity, and thereby deviate below the saddle-splay curve for AMPs. Conversely, the members of the membrane anchor proteins family should likely persist in the lipid bilayer, which is reflected by closer correlation to the selection rule.

Finally, we observe that the magnitude of NGC production does not appear to depend on where the sequences fall on the saddle-splay selection rule (Figure 2.14D). In other words, there is no correlation between position on this saddle-splay curve and degree of induced NGC in membranes, consistent with the idea that this curve represents a kind of tradeoff between using different combinations of amino acids to generate the same curvature. Using the directed search of the sequence space enabled by the SVM classifier, we can now efficiently identify and discover curvature-generating sequences.

2.4 Conclusions

In summary, we have trained a SVM classifier to recognize membrane activity and experimentally calibrated the recognition metric by peptide synthesis and characterization. The results, which highlight the difference between the efficacy of an antimicrobial and its recognizability as such, are surprising. A SVM classifier trained only on physicochemical information can effectively recapitulate geometric and topological principles required for membrane permeation. We use machine learning to not only predict unknown membrane

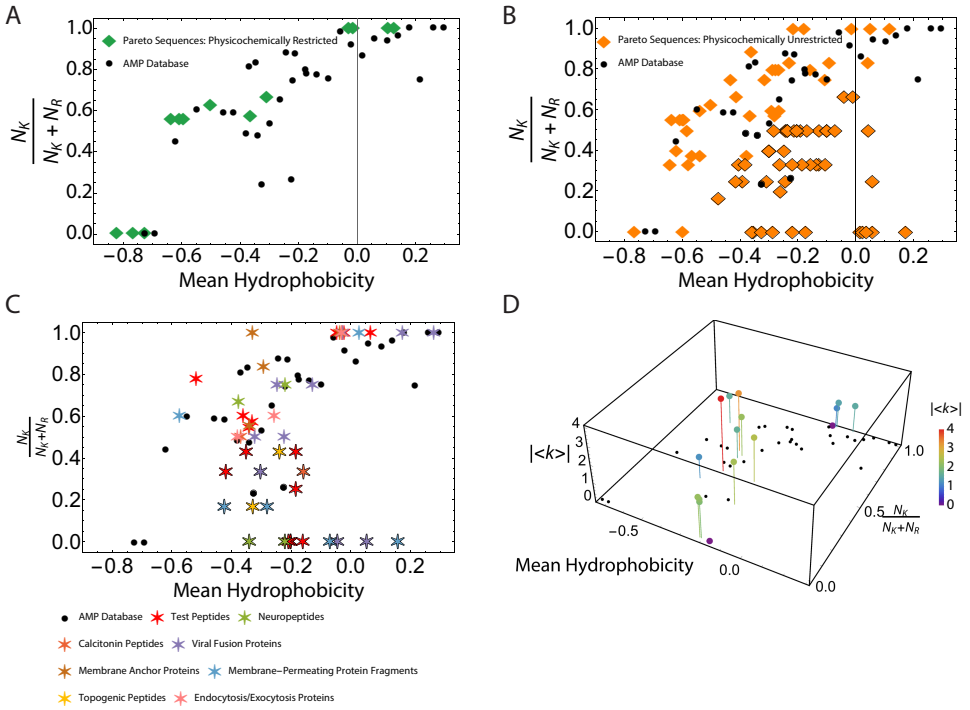


Figure 2.14: **Pareto-optimal sequences and newly discovered curvature-generating sequences follow the saddle-splay selection rule.**
(A) Pareto-optimal sequences that are physicochemically restricted to be close to known antimicrobials follow the saddle-splay selection rule denoted by the AMP database (black), as expected. (B) Pareto-optimal sequences that are physicochemically unrestricted (orange diamonds, no border) are partitioned into two clusters, ones that follow the saddle-splay selection rule, and ones that lie below and to the right of the curve (orange diamonds, black border), indicating a subset of membrane-active sequences that have multiplexed functions in addition to membrane activity. (C) Test peptides and newly discovered membrane-curvature generating sequences exhibit a similar distribution to physicochemically unrestricted Pareto-optimal sequences, which reflects the diverse identities of the newly discovered membrane-active peptide families. (D) Tested peptides (stems) demonstrate NGC generating ability regardless of their placement along the saddle-splay selection rule (black circles). The colors of the dots are chosen based on the magnitude of NGC ($|\langle k \rangle|$) generation, and stems to the plane demonstrate where they fall along the selection rule. The units of $|\langle k \rangle|$ are 10^{-4}Å^{-2}.

active peptides from known ones, but to also reflexively to identify peptides with multiple functions and discover previously unknown interrelations between existing peptide classifications. Using the SVM classifier as a discovery tool to map the sequence space of AMPs, we find a diverse taxonomy of sequences that are expected to be just as membrane-active as known AMPs, but with a broad range of primary functions outside of immunity. Finally, we show how our SVM classifier can be generalized to other fields and used as a search engine for membrane activity in peptide sequences and a detector of AMPs with multiplexed functions beyond membrane activity.

2.5 Methods

2.5.1 Computational Methods

2.5.1.1 QSAR Modeling

The development of a QSAR model comprises four principal steps: 1) the collation and curation of training and test data on which to construct and evaluate the model, 2) the choice and calculation of descriptors that will be used by the model to predict the property of interest, 3) filtering and selection of the descriptors upon which the model will be trained, and 4) the choice, construction, training, and testing of a predictive QSAR model [91, 142]. In this work, we train a support vector machine (SVM) to predict whether or not a candidate peptide possesses antimicrobial activity based on computationally generated physicochemical descriptors efficiently computed from knowledge of its sequence alone. We describe each step of this process in the following sections.

2.5.1.2 Datasets

A total of 286 α-helical AMPs were collated from The Antimicrobial Peptide Database (`http://aps.unmc.edu/AP/main.php` [8]). These data were curated to contain only peptides of length 8-60 residues with complete sequences, and for which antimicrobial activity had been experimentally verified by plate killing assays or broth microdilution assays.

These data constitute the positive dataset. A total of 2971 α-helical membrane active proteins of length 8-60 residues, which have been reported to have no antimicrobial activity, were downloaded from the Protein Data Bank of Transmembrane Proteins (PDBTM) (`http://pdbtm.enzim.hu` [143, 144, 145]). Using a Python script, we analyzed PDB files to extract α-helical portions of transmembrane proteins that were 8-60 residues long based on crystal structure data and the DSSP secondary-structure prediction algorithm. We included both types of peptides/proteins in the decoy dataset to sample a range of physicochemical diversity. To provide for balanced training of the QSAR model, this ensemble was randomly subsampled down to 286 peptides, which constituted the negative dataset [111]. This means of constructing a negative dataset of decoy sequences for the purposes of training a QSAR model has previously been employed in several prior works [111, 96, 146, 147].

From the aggregated ensemble of 572 peptides derived from natural ($\sim$ 98%) and synthetic ($\sim$ 2%) sources and a variety of microbial and multicellular species constituting the positive and negative datasets, we randomly extracted 85% to form a balanced training dataset comprising 243 antimicrobial and 243 decoy sequences. The remaining 15% of sequences were reserved for an 86-peptide balanced test dataset comprising 43 antimicrobial and 43 decoy sequences. The training data was used to perform feature selection and train our QSAR model employing cross-validation to guard against overfitting. The test data was used only to evaluate the final model performance against a blind dataset, and was not used at any stage of feature selection or model training [148].

2.5.1.3 Physicochemical Descriptors

The antimicrobial activity of a particular peptide can, in principal, be predicted by its sequence similarity to known AMPs. This approach typically fails, however, when the query protein has little sequence homology with any peptide in the AMP database [92, 149]. The variable length of AMPs together with the exponential explosion of sequence diversity with peptide size means that only a tiny fraction of sequence space will share sufficient homology with the 286 known AMPs in the training dataset, making this approach unsuitable for

high-throughput *in silico* screening of diverse candidate sequences.

Rather than attempting to predict antimicrobial activity directly from peptide sequence, one may avoid appealing to homology altogether, and instead formulate a model to predict activity based on physicochemical properties. This approach assumes that the antimicrobial activity of a peptide—a property that is expensive and time consuming to measure—may be predicted from physicochemical properties of the peptide—quantities that can be easily computed from a knowledge of only the peptide sequence [150, 111]. This hypothesis is well founded for AMPs, since despite the enormous diversity of cationic antimicrobial peptides sequences and provenance, they tend to be small, cationic, and hydrophobic [91, 151, 142]. Prior experimental work has also shown an ensemble of 1080 cationic AMP defensins to possess a conserved relationship between peptide hydrophobicity and the lysine to arginine ratio [72].

Under this paradigm, we generate for each peptide in the training set a vector of physicochemical descriptors from which to build a predictive model of antimicrobial activity [111, 152, 92]. For high-throughput computational screening, it is desirable that these properties be fast to compute, require no experimental input, and certainly be no more difficult or time consuming to calculate than the property we wish to predict [89]. The precise choice of descriptors is typically problem specific, but it has been shown that amino acid composition alone is typically insufficient for accurate prediction, and sequence order information must typically also be incorporated [149]. In this work we adopt a shotgun approach by generating a very large number of descriptors for each peptide, and then—as we describe in the next section—applying feature selection to systematically extract an optimal subset of descriptors [153, 114]. This procedure possesses the attractive quality that requires minimal knowledge or intuition about the underlying physicochemical mechanisms of antimicrobial, but instead allows the descriptors to "speak for themselves". Furthermore, we will show that reducing the number of descriptors to just a handful enhances both the predictive performance of the model and its intuitive interpretability [97, 153, 114, 154].

For each peptide in the training set, we generate an ensemble of 1588 physicochemical descriptors falling into the five categories listed in Table 2.3 [111, 91, 152, 123, 155]. The

particulars of each descriptor are described in detail below. All descriptors are efficiently generated using the freely available propy Python package [123, 155], requiring less than 2 s of CPU time per peptide on a single core of a 2.13 GHz Intel Core 2 Duo processor.

Table 2.3: **Physicochemical descriptors.**

Feature Group	Features	No. descriptors	No. descriptor values
Basic character	Length	1	1
	Charge	1	1
	Mean hydrophobicity	1	1
	Mean vdW volume	1	1
Residue composition	Amino acid composition	1	20
	Dipeptide composition	1	400
	Pairwise relative amino acid composition	1	380
Autocorrelation	Normalized Moreau-Broto autocorrelation	15	$15 \times \delta_{max}$
Physicochemical composition	Composition	7	21
	Transition	7	21
	Distribution	7	105
Sequence order features	Sequence order coupling number	2	$2 \times \delta_{max}$
	Pseudo-amino acid composition I	1	$20 + \lambda_{max}$
	Pseudo-amino acid composition II	2	$2 \times (20 + \delta_{max})$

Basic Character. The length is number of residues in the peptide, N. The charge is the net charge of the peptide assigning charges to each residue by specifying (H,K,R) to carry +1 charge, (D,E) to carry -1 charge, and the remaining 15 to be uncharged [156]. The mean hydrophobicity is defined as the per residue Eisenberg consensus hydrophobicity averaged over all residues in the peptide [133]. The mean vdW volume is defined as the average per residue normalized van der Waals volume [156].

Residue composition. The amino acid composition measures the fraction of each of the 20 natural amino acid types in the sequence, $f(X) = \frac{N_X}{N}$, where N_X is the number of residues of type X in the sequence of length N and $X \in (A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y)$ for a total of 20 descriptor values.

The dipeptide composition measures the relative fractions of the 400 possible contiguous dipeptides observed in the sequence, $f(X,Y) = \frac{N_{XY}}{N-1}$, where the indices X and Y run over the 20 natural amino acids. The pairwise relative amino acid composition is defined as $pc(X,Y) = \frac{N_X}{N_X+N_Y}$, which measures the relative fraction of each amino acid in the sequence relative to every other.

Autocorrelation. The autocorrelation measures the similarity in the properties of any pair of amino acid residues along the peptide chain, revealing any correlated distribution of amino acid properties as a function of the separation between residues along the peptide backbone. We elect to use the normalized Moreau-Broto autocorrelation, defined as a function of the separation between any pair of amino acid positions, δ, as

$$AC_{nMB}(\delta) = \sum_{i=1}^{N-\delta} \frac{P_i P_{i+\delta}}{N-\delta}$$

where $\delta = 1, ..., \delta_{max}$, P_i is the physicochemical property of amino acid at position i, and we make the standard choice of $\delta_{max} = 30$ [123, 155]. We consider 15 different amino acid physicochemical properties: (1) the Kyte-Doolittle hydropathy index [157], (2) the average flexibility index [158], (3) the polarizability parameter [159], (4) the free energy of solution in water [159], (5) the solvent accessible surface area [160], (6) the residue volume[161], (7) the steric parameter related to the van der Waals volume of the side chain [162], (8) the relative mutability [163], (9, 10, 11) the three principal property values (z_1, z_2, z_3) related to hydrophilicity, bulk, and electronic properties [164], (12) the isoelectric point [165], (13, 14) the positive and negative charge [156], and (15) the transfer free energy to the lipophilic phase [166].

Physicochemical composition. These descriptors seek to provide a coarse-grained representation of the distribution of particular physicochemical properties along the peptide backbone. Seven physicochemical features are considered—hydrophobicity, van der Waals volume, polarity, polarizability, charge, secondary structure, and solvent accessibility [123, 167, 168, 169]—each of which is coarse grained into the three categories listed in Table 2 in Ref. [123]. The composition descriptor, C, measures the fraction of amino acids in the

peptide that fall into each of the three categories for each of the seven descriptors, making a total of 21 descriptor values. The transition descriptor, T, measures the fraction of pairs of contiguous residues that belong to each of the three possible combinations of different categories (i.e., category 1 and 2, 1 and 3, or 2 and 3, where the order is immaterial). Applied to each of the seven descriptors, this also yields 21 descriptor values. The distribution descriptor, D, furnishes for each of the three categories five values: the fractional distance along the peptide sequence that must be traveled to encounter (i) the first residue belonging to the category, (ii) 25%, (iii) 50%, (iv) 75%, and (v) 100% of the residues belonging to the category. Applied to each of the three categories for each of the seven descriptors, this yields a total of 105 descriptor values.

Sequence order features. Similar to the autocorrelation descriptors, this class of features seeks to characterize patterns in physicochemical properties along the peptide backbone. The sequence order coupling number as a function of the separation between any pair of amino acid positions, δ, normalized by sequence length is defined as

$$\tau_\delta = \frac{1}{N-\delta} \sum_{i=1}^{N-\delta} (d_{i,i+\delta})^2$$

where $\delta = 1, ..., \delta_{max}$, $d_{i,i+\delta}$ is the physicochemical "distance" between the amino acid at position i and that at position $i + \delta$, and we make the standard choice of $\delta_{max} = 30$ [149, 123, 155]. We consider 3 different definitions of the physicochemical distance between a pair of amino acids defined by (1) the Schneider-Wrede context matrix [170] and (2) Grantham chemical distance matrix [171].

The pseudo amino acid composition (PseAAC) introduced by Chou is a descriptor that characterizes the amino acid composition of a peptide while simultaneously maintaining information on sequence and length [149, 172]. The original PseAAC defined by Chou associates to a peptide PseAAC = $(p_1, p_2, ..., p_{20}, p_{20+1}, ..., p_{20+\lambda_{max}})$, which is an ensemble of

$20 + \lambda_{max}$ numbers defined as:

$$p_i = \begin{cases} \frac{f_i}{\sum_{i=1}^{20} f_k + w \sum_{k=1}^{\lambda_{max}} (N-k)\tau_k} & 1 \leq i \leq 20 \\ \frac{w\tau_{i-20}}{\sum_{i=1}^{20} f_k + w \sum_{k=1}^{\lambda_{max}} (N-k)\tau_k} & 20+1 \leq i \leq 20 + \lambda_{max} \end{cases}$$

where $f_i = N_i/N$ is the fraction of residues of type i in the sequence of length N, and i indexes over the $i = 1, ..., 20$ natural amino acids, τ_δ is defined as before, and w is a weight factor. Following Chou, we choose $w = 0.05$ [149]. The first 20 features are associated with the amino acid composition of the sequence, whereas the next λ_{max} contain sequence order information describing the prevalence of pairwise physicochemical correlations between amino acids separated by $k = 1, ..., \lambda_{max}$ positions that Chou refers to as "tiers" [149].

In Chou's original formulation of the PseAAC, $\lambda_{max} \leq (N-1)$ reflecting the fact that tiers corresponding to residue separations exceeding the length of the peptide are undefined, limiting λ_{max} according to the shortest peptide in the ensemble. In this work $\lambda_{max} = 7$. Chou defines the physicochemical distances appearing in the factors of τ_δ as the weighted average of the Z-scored residue hydrophobicities H, hydrophilicities L, and side chain masses M assigned by Refs. [173, 174].

$$(d_{i,i+\delta})^2 = \frac{1}{3}\left((H_i - H_{i+\delta})^2 + (L_i - L_{i+\delta})^2 + (M_i - M_{i+\delta})^2\right)$$

We term the pseudo amino acid compositions computed in this manner PseAAC I, providing $20 + \lambda_{max}$) descriptor values.

We also consider a generalization of the PseAAC in which we (i) replace λ_{max} with $\delta_{max} = 30$, and (ii) compute physicochemical distances according to (1) the Schneider-Wrede context matrix [170] and (2) Grantham chemical distance matrix [171]. Values of p_{k+20} in short peptides for which $k > (N-1)$ are assigned a value of zero. We term the pseudo amino acid compositions computed according to this generalization as PseAAC II, which furnish an additional $2 \times (20 + \delta_{max})$ descriptor values.

2.5.1.4 Physicochemical Descriptor Selection

Variable selection techniques seek to identify from an ensemble of input variables a subset that is most predictive of the property of interest, and eliminating irrelevant or redundant descriptors that can diminish model generalization and predictive performance [153, 114, 154]. In addition to exhibiting improved performance, predictive models based on small number of variables are also more intuitively interpretable, providing insight and understanding of the underlying physicochemical mechanisms of activity.

Although there is the risk of rejecting descriptors that themselves are uninformative predictors but can be highly informative in combination with others, we stabilize the variable selection procedure by performing limited univariate and bivariate filtering out of irrelevant or redundant descriptors [153]. We remove two zero-variance descriptors that were invariant over all training examples, and 257 redundant descriptors possessing a Spearman correlation coefficient of $\rho \geq 0.95$ with another descriptor in the ensemble [89, 153]. For each pair of highly correlated descriptors, a non-parametric Mann-Whitney U test was conducted for each member of the pair to assess the probability that the value of the descriptor differed significantly between the AMP and decoy peptides [111]. The descriptor with the larger p-value—indicating lower confidence that the null hypothesis that the descriptor values do not differ between the AMP and decoy peptides can be rejected—was selected for elimination. The remaining 1329 filtered descriptors were Z-scored to center the descriptors, standardize their ranges, and render them dimensionless [153].

The large number of remaining descriptors makes exhaustive enumeration of all possible subsets computationally intractable. Instead, we appeal to the elegant approach of Bi et al. who propose training a linear support vector machine employing the L1-norm, rather than the conventional L2-norm, of the hyperplane normal to enforce sparsity in this weight vector and mediate selection of subsets of the input descriptors [114]. This approach may be considered an embedded variable selection method in which variable selection is performed on-the-fly during the process of training the predictor [153]. We then take the optimal subset of descriptors identified by this approach, and use them to train standard L2-norm linear

and nonlinear SVCs described in the next section. Full details of feature selection can be found in [175, 176, 177].

We perform feature selection at $C_{opt} = 9.95 \times 10^{-2}$ by performing $k = 15$ rounds of stratified shuffled cross-validation and identifying the descriptors selected in each round as those corresponding to non-zero elements of the w vector. Each round selects on average (115 ± 7) descriptors as non-zero elements within the sparse w vectors. Characteristic of the inherent sensitivity of the SVC to the training and validation partitions, the descriptors selected in each round are highly variable [178, 114]. We stabilize this inherent instability by bootstrap aggregation (bagging) [179], in which we selected those $m = 12$ descriptors that were selected in all $k = 15$ rounds of cross validation to define the bagged descriptor subset. The 12 bagged descriptors are listed in Table 2.1.

2.5.1.5 Model Training and Testing

The bagged subset of 12 descriptors in Table 2.1 were used to train a standard soft margin, hinge loss, L2-norm support vector classifier (L2-SVC) by minimizing the primal objective function [114, 115, 116, 117, 118, 80],

$$argmin(\mathbf{w}, b) \left[\frac{1}{2}||\mathbf{w}||_2 + C\frac{1}{n}\sum_{i=1}^{n} L_1(y_i, \phi(\mathbf{x}_i)) \right]$$

where

$$L_1(y_i, \phi(\mathbf{x}_i)) = \left[1 - y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b)\right]_+ = max\left[0, 1 - y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b)\right]$$

where $i = 1, ..., n$ indexes over the $n = 486$ training peptides, y_i is a binary indicator function indicating the class membership of peptide i as antimicrobial ($y_i = +1$) or decoy ($y_i = -1$), $\mathbf{x}_i$ is a d-dimensional vector containing the ordered set of $m = 12$ bagged descriptors associated with peptide i, ϕ is a (possibly nonlinear) function mapping the descriptor vectors to a transformed feature space. $\mathbf{w}$ is the weight vector defining a hyperplane separating the two classes of points in the transformed feature space descriptor space, b governs the perpendicular offset of the hyperplane from the origin, and $||\mathbf{w}||_2$ denotes the L2-norm

of the $\mathbf{w}$ vector. Using the kernel trick, the function ϕ is defined through the kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ expressing a dot product in the transformed feature space, leaving $\phi(\mathbf{x}_i)$ and $\mathbf{w}$ unknown but their L2-norms accessible and rendering the optimization problem soluble in the dual space [116, 117, 118]. $L_1(y_i, \phi(\mathbf{x}_i))$ is the hinge loss function defining the penalty for incorrect predictions, and C is a hyperparameter controlling the trade-off between maximizing the margin separating the two classes and prediction errors over the data.

We solved the L2-SVC optimization problem using scikit-learn [116] employing three standard kernels: (i) linear, $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$, (ii) polynomial, $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$, $d \in N$, and (iii) radial basis function (RBF), $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||_2^2)$. For each kernel, the hyperparameters C, d, and γ were optimized by grid search followed by—for the non-integer variables C and γ—numerical optimization using scipy [180, 181, 182] to maximize the accuracy of the classifier on a 20% validation partition averaged over $k = 15$ rounds of stratified shuffled cross-validation.

We assessed the predictive performance of each of the three L2-SVC classifiers on the blind test set of 43 AMP and 43 decoy peptides that were not used at any stage of model development. We report in Table 2.4 the performance of each classifier measured by the area under the receiver operating characteristic curve (AUROC) and the accuracy, specificity, sensitivity, positive predictive value, negative predictive value, and the Matthews correlation coefficient [111],

The positive predictive value (PPV) measures the expected proportion of positive results that are true positives, and the negative predictive value (NPV) measures the expected proportion of negative results that are true negatives. The Matthews correlation coefficient (MCC) can be interpreted as the correlation coefficient between the observed and predicted binary classifications, and is bounded over the range [-1,+1] [183]. Additionally, we present in Figure 2.3a the receiver operating characteristic (ROC) curve for each classifier showing the sensitivity-specificity trade-off, and in Figure 2.3b the precision-recall curves illustrating the PPV-sensitivity relationship. For comparative purposes, we present in Table 2.4 and Figure 2.3 the performance of the three classifiers trained on the filtered set of 1329 descriptors (i.e.,

the subset of all 1588 descriptors after removing zero variance and redundant descriptors) from which the 12 bagged descriptors were selected.

Table 2.4: **Performance of trained SVMs.**
Predictive performance on the blind 86-peptide test dataset of six L2-SVCs employing three different kernels and trained on either the 1329 filtered descriptors or the 12-descriptor bagged subset listed in Table 2.1. For each kernel we list the optimal hyperparameter values determined by grid search and numerical optimization. The largest values in each column are highlighted in bold.

# descriptors	Kernel	Acc.	Spec.	Sens.	PPV	NPV	MCC	AUROC
12	Linear $C = 1.27 \times 10^{-2}$	0.919	**0.930**	0.907	0.929	0.909	0.837	0.981
12	Polynomial $d = 3$ $C = 1.29 \times 10^{-1}$	0.884	0.860	0.907	0.867	0.902	0.768	0.970
12	RBF $\gamma = 4.92 \times 10^{-5}$ $C = 1.07 \times 10^{2}$	0.930	**0.930**	0.930	**0.930**	0.930	0.860	0.979
1329	Linear $C = 9.63 \times 10^{-4}$	0.930	0.906	0.953	0.911	0.951	0.861	0.990
1329	Polynomial $d = 3$ $C = 3.87 \times 10^{-2}$	0.907	0.814	**1.000**	0.843	**1.000**	0.828	**0.994**
1329	RBF $\gamma = 5.34 \times 10^{-4}$ $C = 1.12 \times 10^{1}$	**0.942**	0.907	0.977	0.913	0.975	**0.886**	0.991

The 12-descriptor linear SVC shows excellent performance along all metrics, with accuracy, specificity, sensitivity, PPV, and NPV all in excess of 90%. Most importantly for high throughput computational screening for novel AMP candidates, the expected fraction of positive predictions that are true positives is 92.9%, and the false positive rate (FPR = 1 - specificity) is only 7%. Relative to the linear kernel, the RBF kernel has comparable specificity and positive predictive value—the two most important metrics for our purposes—with slightly elevated accuracy, sensitivity, negative predictive value, and Matthews correlation, and slightly diminished AUROC. The polynomial kernel is poorer on all counts relative to the linear model.

Compared to the 12-descriptor linear SVC, the performance of the 1329-descriptor SVCs

implementing linear, polynomial, and RBF kernels possess poorer specificity and PPV. In the case of the linear and RBF kernels, this degradation is mitigated by slightly elevated performance along the other metrics. The 1329-descriptor polynomial classifier shows poorer performance along all metrics, with the exception of a comparable AUROC and perfect sensitivity and NPV values due to the absence of any false negatives (FN = 0). It is the objective of this work to train a classifier for high throughput computational screening of peptide sequence space to identify a small number of candidate peptides for experimental testing, so high specificity is a more important characteristic than high sensitivity for our application.

In sum, by paring down the large ensemble of descriptors to a 12-variable subset highly predictive of antimicrobial activity, the variable selection procedure permitted us to generate much simpler classifiers with good generalization and high predictive performance. We observed very little performance gains from the use of nonlinear kernels, and elect to use the linear classifier for its high performance and intuitive interpretability. The elements of the w vector corresponding to the weights accorded to each of the 12 descriptors by the linear L2-SVC are listed in Table 2.1, quantitatively relating each of these physicochemical descriptors to antimicrobial activity.

2.5.2 SAXS Experiments

Peptides were purchased (Lifetein, C3-Jian) and synthesized in high purity (> 95% HPLC) using solid-phase synthesis. Liposomes were prepared as previously described elsewhere Schmidt:2013gs. Lyophilized DOPS (1,2-dioleoyl-sn-glycero-3-phospho-L-serine (sodium salt)), DOPG (1,2-dioleoyl-sn- glycero-3-[phospho-rac-(1-glycerol)]), and DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine) from Avanti Polar Lipids were used. Stock solutions were dissolved in chloroform at 20 mg/mL. Mixtures of these lipids were used as a first-order model for bacterial membranes. DOPS and DOPE were mixed at a 1:4 molar ratio (DOPS/DOPE = 20/80), and DOPG and DOPE were mixed at a 1:4 molar ratio (DOPG/DOPE = 20/80). Chloroform was evaporated under N_2, and the lipids were fur-

ther dried overnight under vacuum. Dried lipids were resuspended in 100 mM NaCl + 10 mM HEPES at pH 7.4 to a final concentration of 20 mg/mL. Aqueous lipid solutions were incubated overnight at 37°C. SUVs were prepared by sonication of resuspended lipids until clear. Monodispersity in size was obtained via extrusion through a 0.2 μm filter.

Before use, peptides are solubilized in 100 mM NaCl + 10 mM HEPES at pH 7.4. Liposomes (DOPS/DOPE = 20/80, DOPG/PE 20/80) and peptides were mixed at a peptide-to-lipid charge ratio (P/L) of 1/2 and equilibrated at room temperature overnight. Precipitated peptide-lipid complexes were loaded into 1.5 mm glass quartz capillaries (Mark-tubes, Hilgenberg GmbH) and hermetically sealed. SAXS experiments were conducted at the Stanford Synchrotron Radiation Laboratory (BL 4-2) with monochromatic X-rays of energy 9 keV. Scattered radiation was collected using a Rayonix-MX225-HE detector (pixel size 73.2 μm). The 2D diffraction patterns were integrated using the Nika 1.68 package [184] for Igor Pro 6.37 (Wavemetrics) and FIT2D [185]. For all samples, multiple measurements were taken for consistency. Samples were incubated at 37°C before measurement. No changes were observed over multiple experiments and exposures.

SAXS data were analyzed by plotting the integrated scattering intensity $I(q)$ vs. q using Mathematica. To determine the phases present in each sample, the measured peak positions were obtained and their ratios were compared to the permitted reflections for different liquid-crystalline phases (e.g. lamellar, hexagonal, cubic). Curvature-generating ability of the peptides were determined by calculating the lattice parameter a for the measured cubic phases. Cubic phases observed in our experiments belonged to the Im3m and Pn3m space groups. Measured q positions for the Bragg peak reflections were fitted to the equation $q_{measured} = \frac{2\pi}{a}\sqrt{h^2+k^2+l^2}$ where (hkl) are the Miller indices. The quantity of induced NGC was calculated as the average NGC per unit cell volume using the equation $\langle k \rangle = \frac{2\pi\chi}{A_0 a^2}$, where χ is the Euler characteristic and A_0 is the surface area per cubic unit cell for each phase. Parameter values are $\chi = -4$ and $A_0 = 2.345$ for Im3m, and $\chi = -2$ and $A_0 = 1.919$ for Pn3m. This allows us to quantitatively compare curvature generating ability between peptides.

2.5.3 Antimicrobial Assays

Antimicrobial activity of the synthesized peptides was confirmed using growth inhibition assays with broth microdilution. Bacterial strains used in testing include Streptococcus mutans (UA140), Escherichia coli (D31), or Pseudomonas aeruginosa (PA01). Strains were cultured overnight at 37°C in LB, and then subcultured in fresh LB for 2-3 hours at 37°C until bacteria reach mid-log phase at OD600∼0.5. Bacteria is diluted to 105 CFU/mL in LB for the antimicrobial assay. In a 96-well assay plate, 90 μL of diluted peptides at varying concentrations are added to 10 μL of diluted bacteria, and incubated for 24 hours at 37°C in physiological buffer (100 mM NaCl, 10 mM HEPES, pH 7.4). The sequences are considered antimicrobial if the peptide inhibits bacterial growth in any strain by at least 90% compared to broth alone at a concentration of < 50 μM, as measured by OD600. We find that all of the synthesized test peptides exhibit antimicrobial activity which is in line with their binary SVM classification.

2.5.4 Statistical Analysis

Spearman correlations and p-values were calculated in R and plotted in Mathematica. 95% confidence intervals and p-values for the correlations were calculated using Monte Carlo bootstrapping methods in R with random sampling of the 16 peptide NGC dataset or the 478 peptide MIC dataset (n = 10,000 iterations). A significance threshold of $\alpha = 0.05$ was imposed for all p-value calculations.

2.6 Acknowledgements

We thank Nathan Schmidt and Xintong Lin for their helpful contributions to this project. E.Y.L. acknowledges support from the T32 Systems and Integrative Biology Training Grant at UCLA (T32GM008185) and the T32 Medical Scientist Training Program at UCLA (T32GM008042). B.M.F. acknowledges support from National Science Foundation grant DMS 1345032 "MCTP: PI4: Program for Interdisciplinary and Industrial Internships at

Illinois". G.C.L.W. acknowledges support from NSF DMR1411329. X-ray research was conducted at Stanford Synchrotron Radiation Lightsource, SLAC National Laboratory, supported by the US DOE Office of Basic Energy Sciences under Contract no. DE-AC02-76SF00515.

CHAPTER 3

Machine learning-enabled discovery of membrane-active domains in naturally occurring peptides and proteins

3.1 Introduction

In this section, we highlight several key examples of how the machine-learning based membrane-activity prediction tool described in the previous chapter is broadly enabling in a variety of fields. Here, we leverage the machine-learning classifier to discover hidden membrane activity in existing peptides and proteins, including the mitochondrial fission protein Dnm1 and nucleosomal histone H4, and validate their ability to generate negative Gaussian curvature using SAXS experiments.

3.2 Molecular Motor Dnm1 Synergistically Induces Membrane Curvature To Facilitate Mitochondrial Fission

3.2.1 Abstract

Dnm1 and Fis1 are prototypical proteins that regulate yeast mitochondrial morphology by controlling fission, the dysregulation of which can result in developmental disorders and neurodegenerative diseases in humans. Loss of Dnm1 blocks formation of fission complexes and leads to elongated mitochondria in the form of interconnected networks, while overproduction of Dnm1 results in excessive mitochondrial fragmentation. In the current model, Dnm1 is essentially a GTP hydrolysis-driven molecular motor that self-assembles into ring-

like oligomeric structures that encircle and pinch the outer mitochondrial membrane at sites of fission. In this work, we use machine learning and synchrotron small-angle X-ray scattering (SAXS) to investigate whether the motor Dnm1 can synergistically facilitate mitochondrial fission by membrane remodeling. A support vector machine (SVM)-based classifier trained to detect sequences with membrane restructuring activity identifies a helical Dnm1 domain capable of generating negative Gaussian curvature (NGC), the type of saddle-shaped local surface curvature found on scission necks during fission events. Furthermore, this domain is highly conserved in Dnm1 homologs with fission activity. Synchrotron SAXS measurements reveal that Dnm1 restructures membranes into phases rich in NGC, and is capable of inducing a fission neck with a diameter of 12.6 nm. Through *in silico* mutational analysis, we find that the helical Dnm1 domain is locally optimized for membrane curvature generation, and phylogenetic analysis suggests that dynamin superfamily proteins that are close relatives of human dynamin Dyn1 have evolved the capacity to restructure membranes via the induction of curvature. Taken together, these results suggest that Dnm1-induced membrane curvature and mechanochemical forces function cooperatively to catalyze efficient mitochondrial fission. In addition, we observe that Fis1, an adaptor protein, is able to inhibit the pro-fission membrane activity of Dnm1, which points to the antagonistic roles of the two proteins in the regulation of mitochondrial fission. In addition, we observe that Fis1, an adaptor protein, is able to inhibit the pro-fission membrane activity of Dnm1, which points to the antagonistic roles of the two proteins in the regulation of mitochondrial fission.

3.2.2 Introduction

The morphology and distribution of mitochondria, which vary among different cell types [186] and respond to different cellular conditions [187, 188, 189, 190], are crucial for maintaining normal cell function [191]. Mitochondrial morphology and intracellular distribution are primarily governed by a balance between the antagonistic processes of fusion and fission [192, 193, 194]. Excessive fusion results in elongated mitochondria that form highly interconnected net-like structures, whereas, uninhibited fission leads to mitochondrial fragmentation [192, 195, 196, 197]. Recent studies have associated perturbations in these dynamic pro-

cesses with developmental defects and neurodegenerative diseases [197, 198, 199]. Proteins that regulate and maintain mitochondrial morphology, therefore, play important roles in health and disease.

The machinery involved in mitochondrial fission was first identified in yeast to include Dnm1 and Fis1 [194, 200, 191, 196], with Drp1 and hFis1 as their respective conserved human homologs [201, 202, 203, 195, 204]. A highly conserved cytosolic dynamin-related GTPase, Dnm1 ($\sim$ 85 kDa) in yeast Saccharomyces cerevisiae and Drp1 ($\sim$ 82 kDa) in mammals, is the major essential protein involved in eukaryotic mitochondrial fission [194]. Loss of Dnm1/Drp1 blocks the formation of fission complexes [205, 200] and results in the formation of net-like structures [191, 194, 193]. Likewise, mutations in Dnm1 homologs were also shown to block fission [202, 206], while overproduction of Dnm1 leads to increased mitochondrial fragmentation [207]. Dnm1/Drp1 is characterized by an N-terminal GTPase, a middle domain, and a C-terminal GTPase effector domain involved in self-assembly [192, 208]. In the current model, Dnm1/Drp1 is essentially a molecular motor recruited from the cytosol and self-assembles into spiral-like or ring-like oligomeric structures that encircle the outer mitochondrial membrane at sites of fission. GTP hydrolysis leads to Dnm1/Drp1 conformational changes that constrict the membranes to drive membrane scission [209, 210, 211, 212, 213].

The process of budding and scission, however, requires the generation of significant local negative Gaussian curvature (NGC) in membranes [77]. NGC is the saddle-shaped surface membrane curvature found in the pinched regions of mitochondria undergoing fission. Interestingly, recent work has indicated that specific membrane lipids can facilitate mitochondrial fission [214, 215, 216, 217, 218], and that lipids can play a role in binding, recruiting, and activating proteins that mediate fission, including Drp1 [219, 220, 221, 214]. For example, a cryo-EM study has shown that the active helical conformation of Drp1 is stabilized through direct interactions with cardiolipin [222]. Furthermore, studies have shown that induced membrane curvature via protein crowding can also facilitate membrane fission [52, 51, 223]. In this work, we use machine learning and synchrotron small-angle X-ray scattering (SAXS) to investigate whether Dnm1-lipid interactions can play a role in facilitating mitochondrial

membrane remodeling and fission. To assess whether Dnm1 contains subdomains that can potentially induce membrane curvature necessary for scission, we screen the protein using a recently developed machine learning-based classifier [177] that predicts whether a given α-helical peptide sequence can generate NGC (the machine learning was performed on α-helical sequences.) Using this classifier, we find a high-scoring α-helical domain (LEDLIPTVN-KLQDVMYD) in the protein sequence of Dnm1, which suggests that this domain may be able to remodel membranes by generating NGC. Interestingly, this sequence is conserved in the subset of Dnm1 homologs with fission activity. To test experimentally whether Dnm1 has the capacity to facilitate fission by restructuring membranes in a manner synergistic with its molecular motor-based mitochondrial "pinching" activity, SAXS is used to investigate the full spectrum of Dnm1-induced membrane deformations in vesicles with mitochondrial-like lipid compositions. SAXS results show that Dnm1 indeed restructures membranes into phases rich in NGC, in agreement with the machine learning results. The capacity to induce NGC suggests that Dnm1 function is more complex than just simply mechanochemical constriction activity mediated by its oligomerization. In fact, the membrane deformations generated by Dnm1 promote fission by rendering membranes more amenable to the restructuring necessary for scission. The observed degree of induced NGC is consistent with a fission neck having a diameter of 12.6 nm. An *in silico* mutational analysis of the N-terminal helix of Dnm1 suggests that it is locally optimized for the induction of membrane curvature. In addition, phylogenetic analysis of dynamin superfamily members reveals that proteins close in mutational distance to human dynamin Dyn1 have evolved the ability to remodel membranes through curvature generation. These results suggest that Dnm1-induced membrane curvature and mechanochemical forces function cooperatively to efficiently catalyze mitochondrial fission, supporting the notion that the fission reaction depends on a combination of both protein and membrane mechanics. Remarkably, we find that the adaptor protein Fis1 inhibits the pro-fission membrane activity of Dnm1 by quantitatively suppressing Dnm1-induced NGC. Based on these antagonistic qualities of Dnm1 and Fis1, we suggest the possibility that the two proteins work in concert to modulate mitochondrial fission.

3.2.3 Results and Discussion

3.2.3.1 Machine learning predicts Dnm1 to contain membrane-destabilizing sequences

NGC can be described as saddle-splay curvature due to its shape—the surface curves upwards in one direction and downwards in the orthogonal direction. This specific type of curvature is topologically required for membrane-destabilizing processes, such as budding, pore formation. Moreover, it is the type of local surface curvature found on scission necks during fission events [72, 74, 224] (The pinched fission neck is a classic example of NGC). A large body of work has identified short peptides that can sense [225] or induce membrane curvature and remodeling [226, 227, 225, 228, 229]. Membrane active subsequences derived from viral fusion and fission proteins have been shown to disrupt and bend lipid bilayers [226, 230, 231]. In fact, for peptides that function through membrane destabilization, a strong correlation has been found between their ability to generate NGC and their activity [72]. For example, NGC generation has been experimentally observed using SAXS for antimicrobial peptides (AMPs) [47, 72, 73, 232, 75, 41], cell-penetrating peptides (CPPs) [124, 233, 74, 234], and viral fusion and fission proteins [78, 77]. These findings collectively suggest that generation of NGC is important in membrane-destabilizing processes.

To assess whether Dnm1 contains subdomains that may induce membrane curvature important for fission, we employed a recently developed machine learning classifier trained to predict the likelihood of membrane restructuring activity for a given α-helical peptide sequence [177]. This classifier, which was originally trained with AMP sequences, has been shown to be a good detector of NGC-generating α-helical domains in diverse families of proteins, including membrane-active segments of viral fusion proteins and endocytosis/exocytosis proteins. More notably, its effectiveness in detecting membrane-destabilizing activity in peptide sequences allows for the recognition of previously undetected membrane activity in existing proteins or peptides, such as the neuropeptide hormones [177].

The classifier is based on a linear support vector machine (SVM) that takes as input $n = 12$ physicochemical descriptors generated from the peptide sequence and outputs a

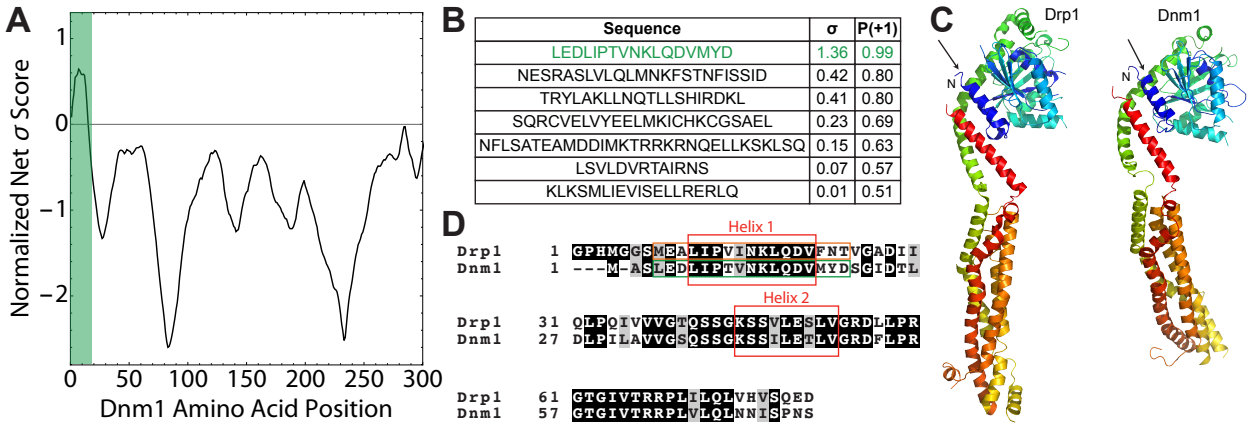


Figure 3.1: **Machine learning screen of Dnm1 yields a putative membrane-destabilizing sequence in the conserved N-terminal domain.**
(A) Normalized σ scores of a moving-window scan of Dnm1 for membrane activity as a function of amino acid position. Scores are shown for the first 300 amino acids of Dnm1. The green bar highlights the location of the top-scoring helical subsequence of Dnm1. (B) Table showing the top-scoring helical subsequences from Dnm1 with $\sigma > 0$ and their corresponding σ and $P(+1)$ assignments from the machine learning classifier. The top hit LEDLIPTVNKLQDVMYD corresponds to a segment of Dnm1 that contains the first N-terminal helix (highlighted in (A)). This sequence has a $P(+1)$ score $> 95\%$, which indicates a high likelihood of NGC generation and membrane restructuring ability. (C) 3D homology structure model of yeast Dnm1 (UniProtKB: P54861) (right) based on the known crystal structure of human Drp1 (PDB: 4BEJ) (left). The first N-terminal helices in both protein sequences are depicted in blue and indicated by the arrows. (D) Sequence alignment of the N-terminal regions of Drp1 and Dnm1 shows significant sequence homology and conservation. The first two conserved N-terminal helices in both sequences are boxed in red. The sequence boxed in green (LEDLIPTVNKLQDVMYD) corresponds to the top hit from the machine learning screen for membrane-destabilizing segments of Dnm1, while the corresponding aligned sequence from Drp1 is boxed in orange (MEALIPVINKLQDVFNT).

score σ specifying the distance of the peptide from a $(n-1) = 11$ dimensional hyperplane trained to optimally separate 243 known α-helical curvature generating peptides from 243 decoy peptides. A large, positive σ score correlates with increased ability to induce NGC in membranes, whereas a negative σ score indicates a lack of membrane-disruptive activity. This score can be converted through a monotonic function into a probability $0 < P(+1) < 1$ that the peptide induces membrane curvature. Large, positive values of σ ($P(+1) > 0.95$) indicate a high likelihood of membrane activity (ability to generate NGC), and negative values of σ ($P(+1) < 0.50$) indicate a low probability of membrane activity. Testing of the classifier on a blind balanced test set of 86 peptides demonstrated 91.9% prediction accuracy, 93.0% specificity, and 90.7% sensitivity. Experimental validation of computational predictions was carried out using SAXS. A subset of α-helical test peptides were incubated with model membranes, and induced NGC was measured. A strong correlation between the ability to generate NGC in membranes and the distance-to-hyperplane SVM metric σ was observed. Classification of a single peptide requires ~ 0.1 s of CPU time permitting high-throughput computational screening for membrane-active peptide discovery and design. Full details of dataset curation, physicochemical descriptor selection, and model training and validation is provided in Refs [177, 175].

We use the classifier to evaluate whether Dnm1 contains amino acid sequences that might remodel membranes through membrane curvature induction. To locate potential curvature-generating domains in Dnm1 (UniProtKB: P54861), we enumerated all possible 10-25 amino acid subsequences of the entire protein sequence and performed a moving window scan with the membrane activity prediction tool. To visualize the membrane activity landscape along the length of Dnm1, a normalized net σ score was calculated by aligning all windowed sequences and averaging over σ values (Figure 3.1A). Maxima at positive σ values in this landscape correspond to subsequences that have high likelihood of membrane activity. In conjunction with this scan, we predicted the secondary structure of Dnm1 using two methods: the DSSP secondary structure prediction algorithm [235] and sequence alignment with the known crystal structure of human homolog Drp1 (PDB: 4BEJ) [236]. Surprisingly, when we compared the results of the moving window scan with the secondary structure prediction

(Figure 3.2), we found a membrane-destabilizing α-helical sequence (LEDLIPTVNKLQDVMYD, $\sigma = 1.36$, $P(+1) = 0.99$) with an exceedingly high score (Figure 3.1B). This sequence lies in a highly conserved N-terminal region shared with Dnm1 homologs that also have membrane activity (Figure 3.1C, D, Table 3.1). Likewise, the corresponding aligned sequence for Drp1 was also predicted to be membrane-destabilizing (MEALIPVINKLQDVFNT, $\sigma = 1.51$, $P(+1) = 0.99$). In addition to this sequence, the algorithm predicted six other helices of Dnm1 to potentially possess membrane activity, but with lower confidence ($P(+1) < 0.95$). These machine learning results predict that Dnm1 contains multiple α-helical sequences with potential membrane-destabilizing activity, including one particularly high scoring domain (Figure 3.1B). These findings from machine learning suggest the testable hypothesis that in addition to promoting mitochondrial fission through mechanochemical constriction activity, the GTPase-driven molecular motor Dnm1 may have the capacity to enhance mitochondrial fission through induced NGC membrane deformations.

3.2.3.2 Dnm1 remodels membranes and induces negative Gaussian curvature to promote mitochondrial membrane fission

To test the predictions from the machine learning classifier, we assessed the ability of Dnm1 to generate membrane-destabilizing curvature. Using high-resolution synchrotron SAXS, we quantitatively characterized Dnm1-induced membrane deformations in model mitochondrial membranes. Small unilamellar vesicles (SUVs) were prepared from ternary phospholipid mixtures of phosphatidylethanolamine (PE), phosphatidylcholine (PC), and cardiolipin (CL) at molar ratios of 75/15/10 and 75/5/20 to mimic the lipid compositions of mitochondrial membranes [237, 238, 215, 239, 220]. Dnm1 was incubated with SUVs at a protein-to-lipid (P/L) molar ratio of 1/1000 and the resulting membrane structures were characterized using SAXS. The Dnm1 concentrations used are comparable to those used in a large number of prior Dnm1-lipid interaction studies (0.2-10 μM) [210, 212, 236, 216, 240, 241, 242].

We found that Dnm1 restructured the lipid vesicles into phases rich in NGC, while the control samples of SUVs only showed a broad characteristic feature consistent with the form

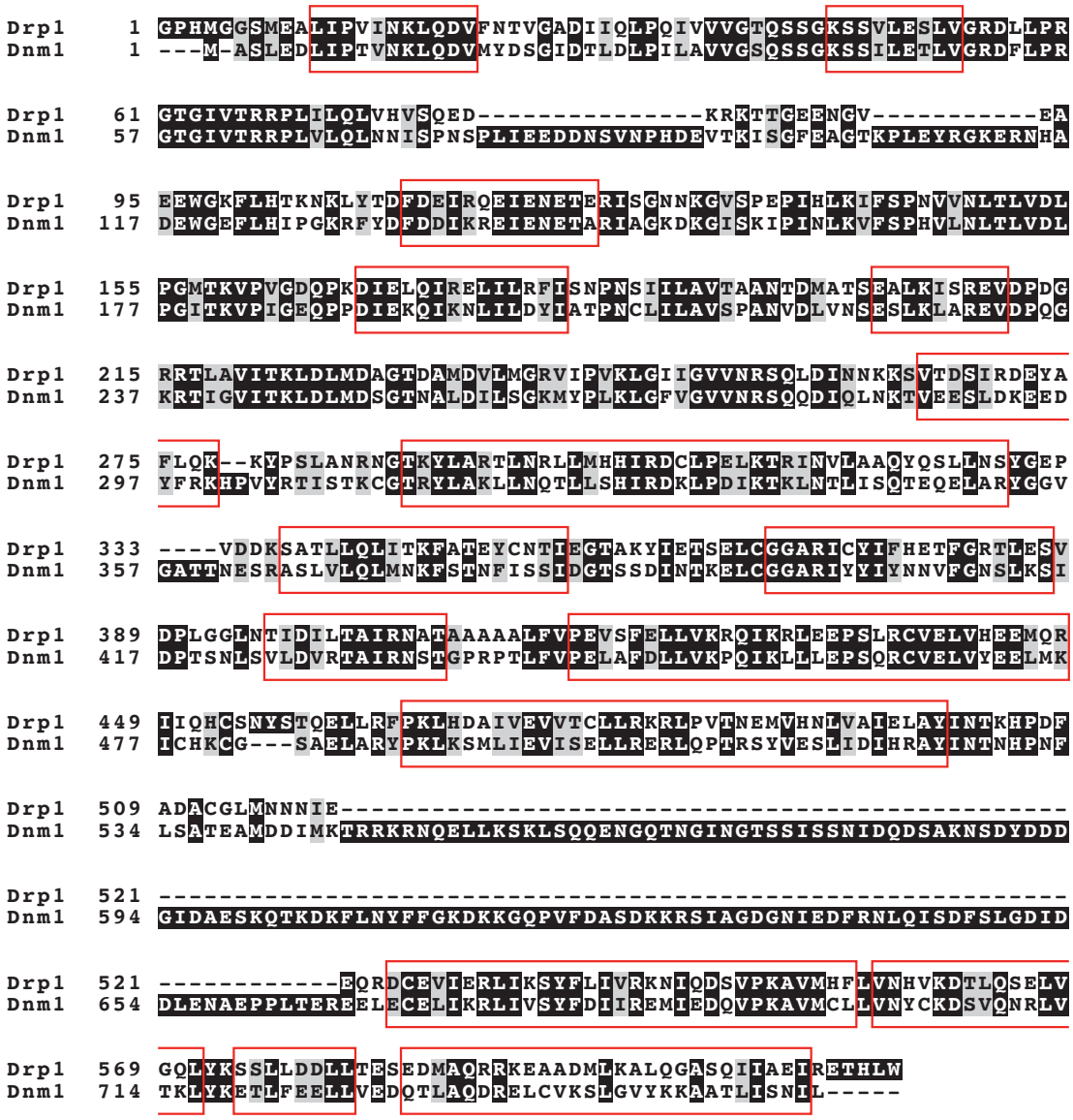


Figure 3.2: **Full sequence alignment of Drp1 and Dnm1.**
Putative helical domains are boxed in red. Identical amino acids are highlighted in black. Similar amino acids are highlighted in gray. Dissimilar amino acids are unhighlighted. Gaps in alignment are denoted with (-).

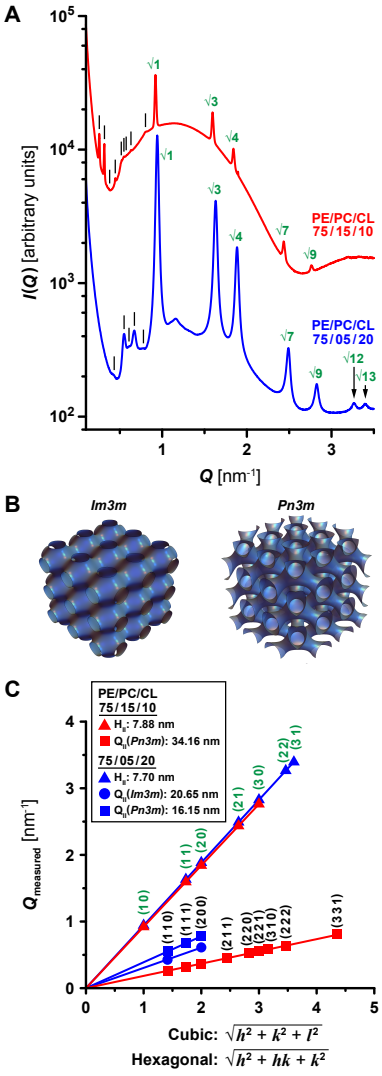


Figure 3.3: **Dnm1 generates NGC necessary for mitochondrial membrane fission.** (A) SAXS spectra from 75/15/10 and 75/5/20 PE/PC/CL model mitochondrial membranes incubated with Dnm1 at $P/L = 1/1000$. Correlation peaks corresponding to assigned reflections are indicated for hexagonal (green) and cubic phases (black lines). (B) 3D reconstructions depicting Im3m (left) and Pn3m (right) cubic phases, which have continuous surfaces with NGC at every point. (C) Indexing of Dnm1-induced $H_{||}$, and Pn3m and Im3m $Q_{||}$ phases for 75/15/10 (red) and 75/5/20 (blue) PE/PC/CL membranes. Plots of the measured Q positions, $Q_{measured}$, versus the assigned reflections in terms of Miller indices, $\sqrt{h^2 + hk + k^2}$ for hexagonal and $\sqrt{h^2 + k^2 + l^2}$ for cubic phases. The lattice parameters were calculated from the slopes of the linear regressions through the points and are provided in the legend.

factor of unilamellar vesicles. For the model membrane 75/15/10 PE/PC/CL, we observed a coexistence of two phases: (1) one set of correlation peaks with Q-ratios $\sqrt{1} : \sqrt{3} : \sqrt{4} : \sqrt{7} : \sqrt{9}$, consistent with an inverted hexagonal ($H_{||}$) phase with a lattice parameter of 7.88 nm; (2) a second set of peaks with Q-ratios $\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{6} : \sqrt{8} : \sqrt{9} : \sqrt{10}$, which indexed to a Pn3m "double-diamond" cubic ($Q_{||}$) phase with a lattice parameter of 34.16 nm (Figure 3.3). Similarly, the model membrane 75/5/20 PE/PC/CL exhibited $H_{||}$ and Pn3m $Q_{||}$ phases with lattice parameters of 7.70 nm and 16.15 nm, respectively. However, for this membrane composition, an additional third set of peaks were identified, which indexed to an Im3m "plumber's nightmare" $Q_{||}$ phase with a lattice parameter of 20.65 nm. These coexisting Pn3m and Im3m $Q_{||}$ phases for this membrane had lattice parameters with a ratio close to the Bonnet ratio of 1.279 [243], indicating that they are near equilibrium with the quantitative amount of membrane curvature balanced between the two cubic phases. Additional measurements on phosphatidylserine (PS)-containing SUVs incubated with Dnm1 having a maltose-binding protein (MBP) tag also showed the induction of $H_{||}$ and Im3m $Q_{||}$ phases, indicating that the NGC generating activity is robust, consistent with earlier findings that MBP, or other N-terminal fusion tags like GFP, do not interfere with Dnm1 activity [244, 245, 240]. A bicontinuous cubic phase, such as Pn3m and Im3m, consists of two nonintersecting aqueous regions that are separated by a lipid bilayer. The center of this bilayer traces out a periodic minimal surface that has NGC at every point. From our SAXS measurements, we found that Dnm1 promotes saddle-shaped membrane deformations to stabilize bulk cubic phases in model membranes with lipid compositions that closely resemble those of mitochondrial membranes, but not others (data not shown). These observations of a strong Dnm1-induced propensity to restructure a lamellar bilayer into a NGC-rich cubic phase mirror the required membrane remodeling events involved in mitochondrial fission. More specifically, NGC promotes membrane deformations that enable the adoption of the unique structures and morphologies that characterize mitochondrial fission. For instance, NGC manifests at the constriction necks of mitochondrial membranes that are formed during the fission process. Remarkably, this intrinsic membrane activity occurs in the absence of GTP binding or hydrolysis, and indicates that Dnm1 may help mediate membrane fission

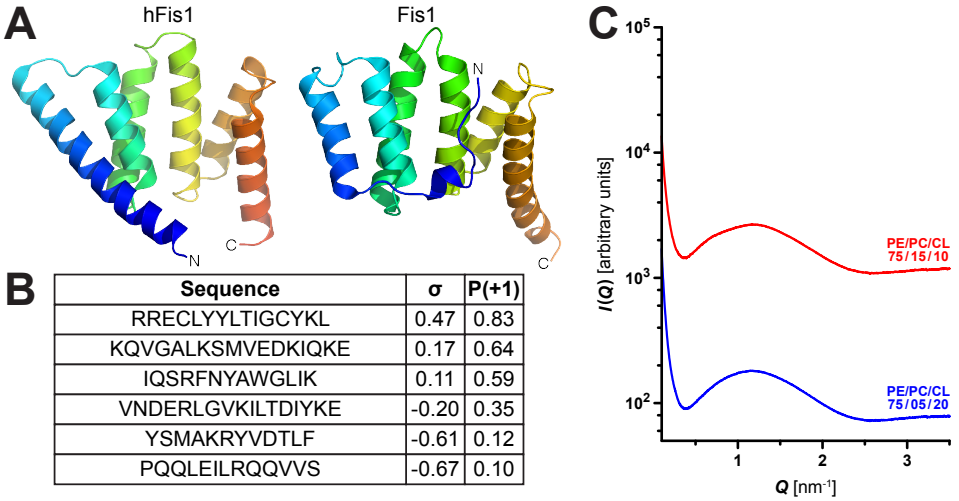
through the induction of NGC in a manner similar to that previously shown by viral scission proteins [77].

3.2.3.3 Fis1 is not predicted to contain membrane-destabilizing sequences and does not restructure mitochondrial-like membranes

The results presented here make contact with a key aspect of mitochondrial fission related to Fis1, a protein localized to the outer mitochondrial membrane. Previous studies have suggested that yeast Fis1 recruits Dnm1 from the cytosol to the mitochondrial membrane [191, 246, 245]. However, other studies have recently provided evidence that Fis1 is instead dispensable for mitochondrial membrane fission, supporting the notion that Dnm1/Drp1-mediated mitochondrial fission is not necessarily a conserved function of Fis1 [247]. Furthermore, the human homolog hFis1 does not appear to recruit Drp1 to mitochondria [248, 249]. Therefore, although Fis1 is a conserved factor for fission and believed to function with Dnm1, its specific role in mitochondrial fission is unclear. Using the same approach used for Dnm1, we examine Fis1 for possible membrane activity.

Fis1 ($\sim$ 16 kDa) is a highly conserved tail-anchored protein in the mitochondrial outer membrane that is believed to be involved in mitochondrial fission with Dnm1. Fis1 has a cytosolic N-terminal domain that contains a six-helix array with two tandem tetratricopeptide repeat motifs (TPR) and a C-terminal transmembrane (TM) domain serving as an anchor into the mitochondrial outer membrane [250, 203] (Figure 3.4A).

With the machine learning classifier, we evaluated Fis1 for potential membrane-destabilizing helical sequences. We ran a moving window scan on the protein sequence (UniProtKB: P40515) for subsequences of 10-25 residues and compared the outputs with the determined protein crystal structure (PDB: 3O48) [251]. While the algorithm predicted a few helical sequences to have a positive σ score, none were identified with a high probability ($P(+1) > 0.95$) for membrane activity (Figure 3.4B). Overall, these results indicate that Fis1 is unlikely to be membrane-destabilizing or have a direct role in the membrane restructuring that occurs during mitochondrial fission.



Sequence	σ	P(+1)
RRECLYYLTIGCYKL	0.47	0.83
KQVGALKSMVEDKIQKE	0.17	0.64
IQSRFNYAWGLIK	0.11	0.59
VNDERLGVKILTDIYKE	-0.20	0.35
YSMAKRYVDTLF	-0.61	0.12
PQQLEILRQQVVS	-0.67	0.10

Figure 3.4: **Helix-rich Fis1 is not predicted to destabilize membranes and does not restructure model mitochondrial membranes.**
(A) 3D structures of yeast Fis1 [251] (PDB: 3O48) (right) and human hFis1 [252] (PDB: 1NZN) (left) show strongly conserved helical content. Both proteins contain a six-helix bundle motif. (B) The six helical subsequences of Fis1 were screened for membrane activity using the machine learning classifier. A table of the Fis1 helical subsequences with their corresponding outputs σ and $P(+1)$ from the machine learning screen is reported. Predictions suggest that helical subsequences of Fis1 have a low probability ($P(+1) < 0.95$) of membrane-destabilizing activity and are unlikely to generate NGC. (C) Fis1 at $P/L = 1/500$ showed minimal membrane activity and did not significantly restructure either 75/15/10 or 75/5/20 PE/PC/CL membranes (no generation of NGC). Their SAXS profiles closely resembled those of control samples containing only lipid vesicles.

Using SAXS, we examined SUVs incubated with Fis1 at a P/L molar ratio of 1/500 and indeed found that the protein did not restructure the lipid vesicles. Instead, the SAXS profile for each model membrane, 75/15/10 and 75/5/20 PE/PC/CL, (Figure 3.4C) exhibited a broad feature that is consistent with the form factor of unilamellar vesicles and resembled the spectra observed for the control SUVs. More specifically, Fis1 does not generate NGC, and therefore, would not be expected to have direct membrane activity that mediates membrane fission in the manner of Dnm1, which is consistent with our machine learning predictions.

3.2.3.4 Membrane activity fitness landscape of the Dnm1 N-terminal helix and evolution of membrane activity in the dynamin superfamily

Now that we have experimentally validated the ability of Dnm1 to directly generate NGC in mitochondrial-like membranes, we apply the machine learning classifier to explore the membrane activity fitness landscape of the Dnm1 N-terminal helix, and trace the evolution of the membrane remodeling ability of this helical domain in members of the dynamin protein superfamily. In our previous work, we combined machine learning with experimental SAXS measurements to demonstrate that a positive, monotonic relationship exists between σ and the NGC-generating ability of membrane-active helical peptides [177]. Conveniently, the observed NGC induced by Dnm1 (-2.5×10^{-2} nm^{-2}) and its σ score (1.36) fall exactly within the predictive range established by previous SAXS experiments. This relationship thus allows us to evaluate the local membrane activity fitness landscape of Dnm1 and its related family members without the cost of experimentally screening hundreds of mutants (this strategy has been previously used in the context of protein stability and viral fitness [253, 254]).

First, we conducted an *in silico* mutational analysis of the N-terminal helical sequence corresponding to the top hit from the initial machine learning screen (LEDLIPTVNKLQD-VMYD, Figure 3.1A, B). 323 unique single mutants (17×19) were generated and screened using the SVM-based membrane activity classifier [177]. The σ values of all single mutants were quantitatively compared to the wild type (WT) sequence by the metric $\Delta\sigma = \sigma_{mutant} - \sigma_{WT}$.

An increase in σ upon mutation ($\Delta\sigma > 0$) corresponded to conservation of NGC-generating activity, and a decrease in σ upon mutation ($\Delta\sigma < 0$) predicted a loss of NGC-generating activity. The sign and magnitude of $\Delta\sigma$ for each single mutant is plotted on a heat map (Figure 3.5A) and the distribution of $\Delta\sigma$ is quantified (Figure 3.5B). Interestingly, residues N9 and K10 were found to be especially important for membrane activity, with all 19 possible mutations leading to large reductions in σ. Other important residues include L1, L4, L11, D13, and D17 (Figure 3.6). This demonstrates that the majority of single mutations led to reductions in σ and the predicted ability to generate NGC. The average value of $\Delta\sigma$ for all single mutants was -0.29, and the distribution of $\Delta\sigma$ values is heavily skewed to negative values (skewness $= -1.43$) (Figure 3.5B). In fact, a third of the single mutants (108) led to drastic reductions in σ ($\Delta\sigma \leq -0.3$). Overall, this mutational analysis of the Dnm1 N-terminal helix demonstrates that the WT sequence is well-optimized for membrane activity.

Next, to examine the membrane activity of Dnm1 in the global scope of dynamin superfamily proteins, we conducted machine learning screens on the homologous helical domains of each member. A sequence alignment (Figure 3.7) and phylogenetic reconstruction of 33 members of the dynamin superfamily from various organisms (Figure 3.5C) revealed that dynamins fall into two previously unrecognized groups: membrane-bound (containing a TM domain) and amphitropic, which are freely soluble proteins that reversibly interact with membranes (lacking a TM domain). We calculated σ for the homologous segments of the 33 dynamin superfamily members (Table 3.1) and found 15 of the sequences were predicted to have membrane activity ($\sigma > 0$) while 18 did not ($\sigma < 0$) (Figure 3.5C, D). Based on sequence alignment and secondary structure prediction, these conserved regions are likely helical. While 14 of the 15 positive hits belonged to amphitropic dynamins, 14 of the 18 proteins that were predicted to lack NGC-generating domains belonged to the membrane-tethered dynamins (MGM1_Sc is a member of both families since it has a long membrane-bound form and a short amphitropic form [255]). All family members with positive σ scores except for MFN2_Dr are amphitropic (lacking a TM domain), while most family members with negative σ contain TM domains (except for MXA_Hs, MXB_Hs, MXA_Dr, and MXB_Dr). Based

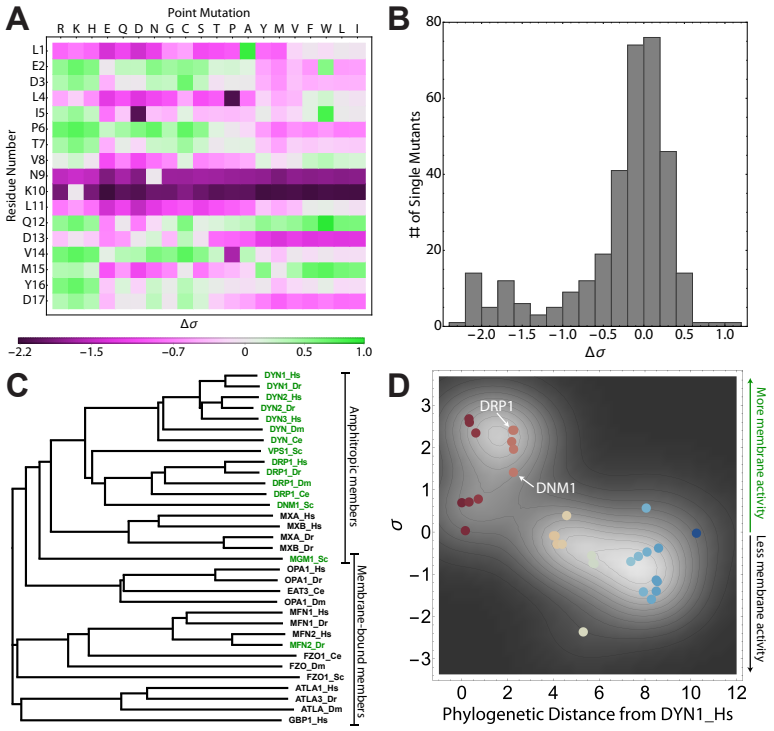


Figure 3.5: **Membrane activity fitness landscape of the Dnm1 N-terminal helix and evolution of membrane activity in the dynamin superfamily.**
(A) 2D heat map of the membrane activity fitness landscape of 323 single mutants of the Dnm1 N-terminal helix ($\Delta\sigma = \sigma_{mutant} - \sigma_{WT}$). Mutations predicted to reduce membrane activity are magenta, while those that conserve membrane activity are shown in green. (B). Distribution of $\Delta\sigma$ shows that the mutational landscape is skewed towards reduced σ (negative $\Delta\sigma$), indicating that the Dnm1 N-terminal helix is well-optimized for membrane activity. (C) A phylogram is constructed via multiple sequence alignment of 33 dynamin superfamily members. Machine learning screens for membrane activity were carried out on an aligned 16-amino acid span corresponding to the conserved N-terminal helix in Dnm1 and Drp1 (Figure 3.7). Those highlighted in green are predicted to be able to generate NGC ($\sigma > 0$), while those in black are not ($\sigma < 0$). (D) Phylogenetic distance from the reference human dynamin DYN1_Hs is inversely related to the predicted ability to generate NGC (σ). As phylogenetic distance increases, ability to generate NGC is decreased ($R_{Pearson} = -0.728$ [-0.859, -0.587], $P < 10^{-5}$, $R_{Spearman} = -0.750$ [-0.863, -0.560], $P < 10^{-5}$, $N = 33$).

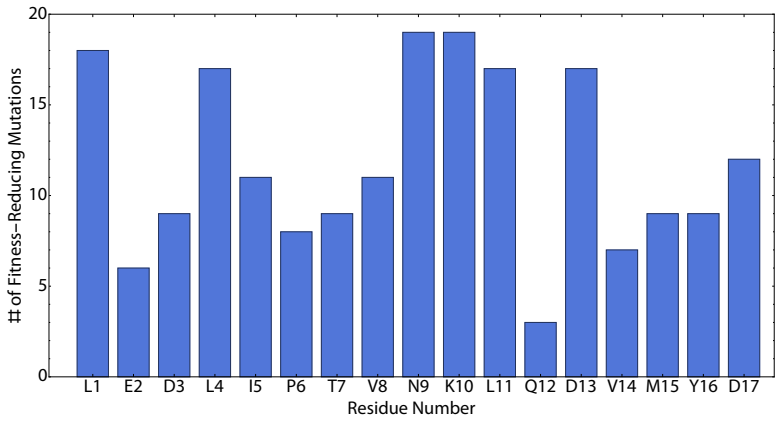


Figure 3.6: **Mutability analysis of the Dnm1 N-terminal helix.**
Plot of the total number of fitness-reducing single mutations ($\Delta\sigma < 0$) at each amino acid along the length of the N-terminal helix of Dnm1. 19 total mutations are possible at each of the 17 positions. Residues L1, L4, N9, K10, L11, D13, and D17 exhibit strong aversion to mutation. A 2D heat map and distribution of all single mutants of this helix are found in Figure 3.5A and 3.5B.

on this finding, we hypothesized that amphitropic dynamin superfamily members evolved in the ability to generate NGC to augment their membrane-remodeling functions. To probe the quantitative relationship between the ability to generate NGC and the evolutionary history of dynamin-related proteins, we calculated the strength of correlation between σ and the phylogenetic distance of each family member from the human classical dynamin DYN1_Hs (Figure 3.5D). We calculated Pearson and Spearman correlations, as well as the newly developed nonlinear correlation metrics distance correlation (dCor) [256] and maximal information coefficient (MIC) [257]. Surprisingly, we find a strong, statistically significant negative correlation between phylogenetic distance and predicted ability to generate NGC (σ) ($R_{Pearson} = -0.728$ [-0.859, -0.587], $P < 10^{-5}$; $R_{Spearman} = -0.750$ [-0.863, -0.560], $P < 10^{-5}$). Nonlinear correlations suggest a strong statistical dependence between σ and phylogenetic distance ($R_{dCor} = 0.781$ [0.678, 0.883], $P < 10^{-5}$; $R_{MIC} = 0.832$ [0.786, 0.999], $P < 10^{-5}$). Indeed, these results suggest that the amphitropic members of the family evolved the ability to generate NGC, and the predicted strength of NGC generation scales inversely with evolutionary distance. A density analysis of σ vs. phylogenetic distance reveals two clusters of dynamin superfamily members, which closely recapitulates the classification of the proteins as amphitropic (Figure 3.5D, upper left) and membrane-tethered (Figure 3.5D,

lower right).

3.2.4 Conclusion and Prospects

In summary, results from machine learning and synchrotron SAXS indicate that the fission protein Dnm1 may catalyze mitochondrial fission by both its GTPase hydrolysis-driven mitochondrial mechanical pinching activity and its synergistic membrane remodeling activity. The observations here provide a framework to reconcile diverse extant results. The existence of strong membrane remodeling activity provides a point of contact with recently observed roles for specific membrane lipids in mitochondrial fission: The ability of Dnm1 to form oligomeric spirals and induce NGC suggests that the protein drives mitochondrial membrane constriction through a combination of both mechanical pinching effects and membrane remodeling by creating the curvature needed for scission neck formation. We note that it is possible that the two effects are synergistic, since mechanical constriction can locally enrich negative spontaneous curvature PE lipids in the neck region [258], which will in turn lower the absolute value of the Gaussian modulus and thereby make it even easier to form narrower necks [41]. Finally, that the sequence domain for membrane curvature generation falls within a highly conserved region between Dnm1 and Drp1 suggests that such membrane activity may be a general feature of dynamin superfamily GTPases [259].

3.2.5 Methods

3.2.5.1 Machine Learning Methods

Full details of the membrane-activity prediction tool are described in prior work [177]. Here, we employ the classifier to identify putative membrane-active subsequences within proteins Dnm1/Drp1 and Fis1. Using Python, all subsequences of 10-25 amino acids in length were generated for these two proteins and cross-referenced with either the known secondary structure from crystallographic data or from a secondary structure prediction algorithm. Helical segments from Dnm1/Drp1 and Fis1 were screened using the machine learning classifier, and σ and $P(+1)$ values were sorted and tabulated. The top helical

subsequences for each protein were sorted by $P(+1)$. In other analyses, sequences from 33 dynamin superfamily members that aligned with the top helical segments of Dnm1/Drp1 were screened in a similar manner.

To generate a probability map for segments of Dnm1 that contain a high probability for membrane activity, all subsequences were aligned with the entire Dnm1 sequence. A normalized net σ score was assigned to each residue, equal to the sum of all σ scores corresponding to subsequences in which that residue appears, divided by the number of times that amino acid appears in a subsequence (Figure 3.1A). This ensures that residues are weighted by the magnitude of σ, which is proportional to the likelihood of membrane activity.

3.2.5.2 Sequence Alignment and 3D Modeling

Initial sequence alignment of Dnm1 and Drp1 was carried out using T-Coffee (`http://tcoffee.crg.cat/apps/tcoffee/index.html`), and the outputs were formatted using BoxShade (`http://www.ch.embnet.org/software/BOX_form.html`). Homology model of Dnm1 was generated from Drp1 using MODELLER [260]. All 3D images for Dnm1, Drp1, Fis1, and hFis1 were generated with MODELLER or the Tachyon plugin for VMD.

3.2.5.3 Phylogenetic Analysis

A multiple sequence alignment of 33 dynamin superfamily members was carried out with default parameters using MUSCLE [261, 262]. Newick tree topologies were extracted from the multiple sequence alignment using Simple Phylogeny (`http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/`). A phylogram was reconstructed using TreeVector [263].

3.2.5.4 Statistical Analysis

Bootstrap analysis for linear and nonlinear correlations were done in R. Correlation coefficients, 95% confidence intervals, and P-values were calculated with $N = 100{,}000$ sampling iterations. The 'minerva' [264] and 'energy' [256] packages were used to calculate MIC and dCor respectively. Null-hypothesis significance tests were carried out with a significance

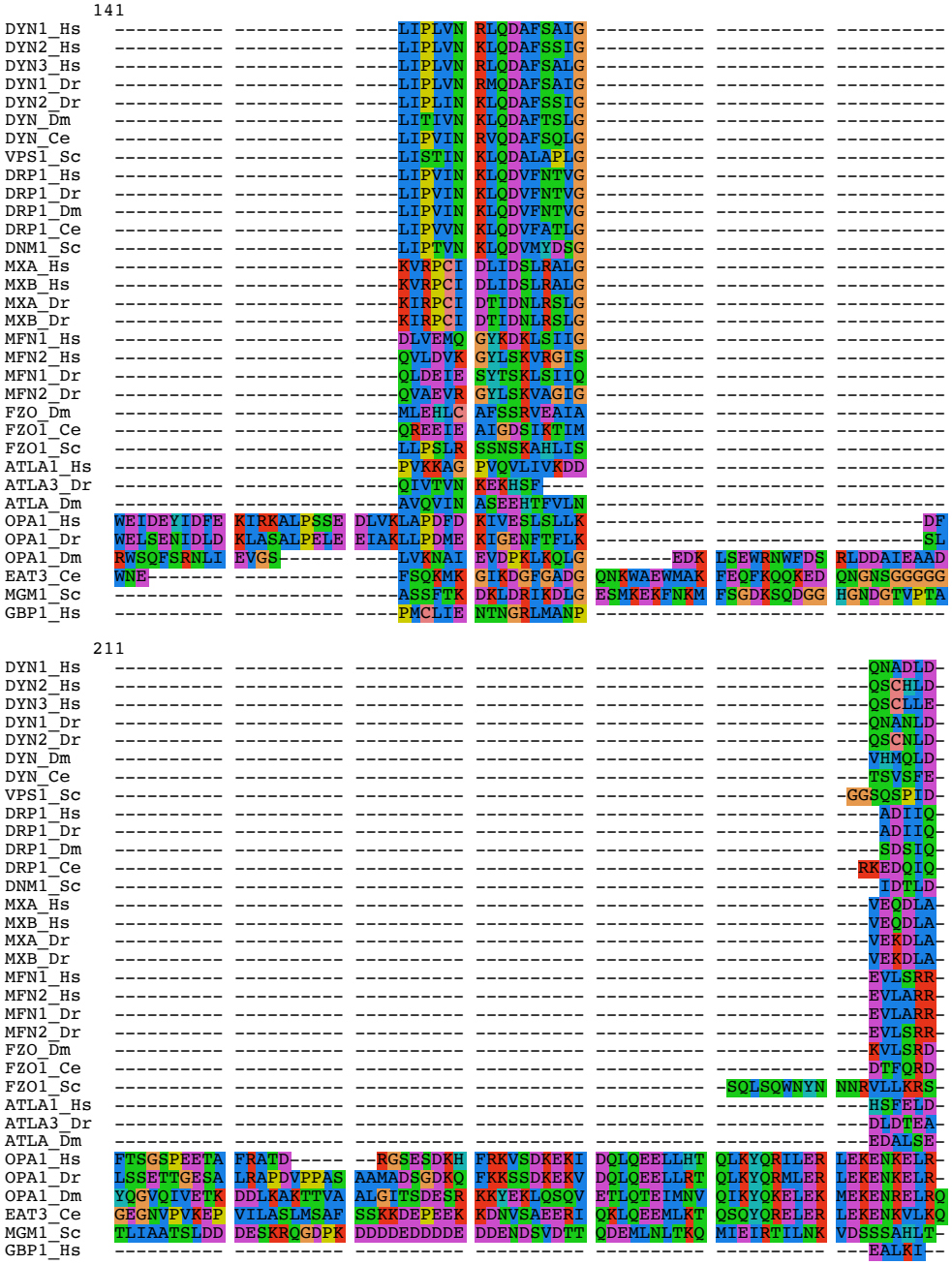


Figure 3.7: **Multiple sequence alignment of dynamin superfamily proteins.** Multiple sequence alignment of the conserved N-terminal helix in Dnm1/Drp1 with other dynamin superfamily members (top half of the page).

threshold $\alpha = 0.05$.

Table 3.1: **Machine learning screen and phylogenetic distances of dynamins.** 33 dynamin family members were screened for membrane activity using the previously described machine learning algorithm. Membrane activity prediction (+1 or -1), distance-to-hyperplane (σ), probability of classification ($P(+1)$), aligned sequence, phylogenetic distance, and accession codes are reported for each sequence. This data is plotted in Figure 3.5D.

ID	Prediction	σ	$P(+1)$	Sequence	Phylogenetic Distance
DYN1_Hs	1	0.69	0.91	LIPLVNRLQDAFSAIG	0
DYN2_Hs	1	2.6	1	LIPLVNKLQDAFSSIG	0.33
DYN3_Hs	1	0.71	0.91	LIPLVNRLQDAFSALG	0.32
DYN1_Dr	1	0.03	0.53	LIPLVNRMQDAFSAIG	0.15
DYN2_Dr	1	2.68	1	LIPLINKLQDAFSSIG	0.31
DYN_Dm	1	2.35	1	LITIVNKLQDAFTSLG	0.61
DYN_Ce	1	0.78	0.93	LIPVINRVQDAFSQLG	0.72
VPS1_Sc	1	2.14	1	LISTINKLQDALAPLG	2.18
DRP1_Hs	1	2.41	1	LIPVINKLQDVFNTVG	2.23
DRP1_Dr	1	2.41	1	LIPVINKLQDVFNTVG	2.23
DRP1_Dm	1	2.41	1	LIPVINKLQDVFNTVG	2.27
DRP1_Ce	1	1.96	1	LIPVVNKLQDVFATLG	2.26
DNM1_Sc	1	1.41	0.99	LIPTVNKLQDVMYDSG	2.25
MXA_Hs	-1	-0.29	0.29	KVRPCIDLIDSLRALG	4.16
MXB_Hs	-1	-0.29	0.29	KVRPCIDLIDSLRALG	4.38
MXA_Dr	-1	-0.09	0.43	KIRPCIDTIDNLRSLG	4
MXB_Dr	-1	-0.09	0.43	KIRPCIDTIDNLRSLG	4.07
MFN1_Hs	-1	-1.42	0.01	DLVEMQGYKDKLSIIG	7.93
MFN2_Hs	-1	-0.47	0.18	QVLDVKGYLSKVRGIS	8.09
MFN1_Dr	-1	-0.58	0.13	QLDEIESYTSKLSIIQ	7.71
MFN2_Dr	1	0.57	0.87	QVAEVRGYLSKVAGIG	8.05
FZO_Dm	-1	-0.7	0.09	MLEHLCAFSSRVEAIA	7.38
FZO1_Ce	-1	-1.59	0.01	QREEIEAIGDSIKTIM	8.28
FZO1_Sc	-1	-0.03	0.49	LLPSLRSSNSKAHLIS	10.25
ATLA1_Hs	-1	-1.14	0.02	PVKKAGPVQVLIVKDD	8.49
ATLA3_Dr	-1	-0.38	0.23	QIVTVNKEKHSFDLDT	8.59
ATLA_Dm	-1	-1.4	0.01	AVQVINASEEHTFVLN	8.49
OPA1_Hs	-1	-0.75	0.08	LAPDFDKIVESLSLLK	5.78
OPA1_Dr	-1	-0.74	0.08	LLPDMEKIGENFTFLK	5.73
OPA1_Dm	-1	-2.36	0	LVKNAIEVDPKLKQLG	5.3
EAT3_Ce	-1	-0.55	0.14	FSQKMKGIKDGFGADG	5.67
MGM1_Sc	1	0.39	0.79	ASSFTKDKLDRIKDLG	4.58
GBP1_Hs	-1	-1.17	0.02	PMCLIENTNGRLMANP	8.53

3.2.6 Acknowledgements

We thank Dr. Jeffery F. Miller for helpful discussions regarding the phylogenetic analysis. This work is supported by grants from the National Science Foundation (DMR-1411329 to G.C.L.W.) and the National Institutes of Health (R01GM067180 to R.B.H.) Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is sup-

ported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393).

3.3 Histone H4 permeates membranes by generating negative Gaussian curvature

3.3.1 Introduction

Chronic inflammation drives a large number of diseases, including lupus, psoriasis, and atherosclerosis. Neutrophils, while typically regarded as mediators of acute inflammation, have recently been implicated in several forms of chronic inflammation, including cell death. Recent work has shown that neutrophils can orchestrate vascular smooth muscle cell death in atherosclerosis. Neutrophils exert direct cytotoxic effects on target cells by releasing nuclear histone H4 from neutrophil extracellular traps, leading to permeabilization of plasma membranes and induction of cell death. However, the mechanism of H4-mediated cell death is unknown. Here, we combine machine learning and small-angle X-ray scattering to illuminate the biophysical mechanism of histone H4-mediated membrane disruption, and demonstrate that membrane activity is localized to the N-terminal helix of H4.

3.3.2 Machine learning predicts that membrane activity of histone H4 is localized to the N-terminus

Negative Gaussian curvature (NGC) is a specific type of membrane curvature that is topologically required for membrane remodeling processes like budding, scission, and pore formation. The ability to generate NGC in membranes has been broadly observed for cell-penetrating peptides [234, 233], antimicrobial peptides [47, 73, 72], and viral fusion/fission proteins [77, 78]. Interestingly, histone H4 has previous reported antimicrobial activity

[265, 266]. Notably, the antimicrobial peptide buforin is a histone fragment and has well-characterized membrane activity [267, 105].

To examine the mechanistic basis for histone H4-mediated membrane damage in smooth muscle cells, we employed a recently-developed machine learning classifier trained to identify domains of proteins that can destabilize membranes and generate NGC [177]. The machine learning classifier was originally trained on α-helical antimicrobial peptides and has been shown to be a powerful predictor of membrane activity in diverse families of peptides and proteins, including viral fusion proteins, topogenic peptides, endocytosis/exocytosis proteins, and mitochondrial remodeling proteins [268]. The classifier's ability to detect membrane-active sequences will enable us to pinpoint previously-undetected membrane-destabilizing portions of histone H4.

Briefly, the classifier is a linear support vector machine (SVM) that takes amino acid sequences as input and outputs a score σ that scales proportionally with membrane activity. Large, positive values of $\sigma >\sim 0.8$ indicate a high probability of membrane activity, while negative values of σ indicate a low probability of membrane activity. The classifier was experimentally validated with calibrating NGC measurements from small-angle X-ray scattering (SAXS) experiments on model membranes, and a strong correlation was found between σ and the ability to induce NGC. Full computational details describing the design, training, and validation of the machine learning classifier can be found here [177, 175, 176].

To isolate the membrane-active domains of histone H4, we carried out a moving window scan of 10-25 amino acids along the entire 103 amino acid length of the protein (Uniprot ID: P62805) using the membrane activity prediction tool. Out of a total of 1368 enumerated sequences, 260 were found to have very high σ scores ($\sigma > 1$), and the majority were localized toward the N-terminus of H4. Since histone H4 contains several helical domains, we correlated the machine learning results with the known secondary structure of H4. We created a visual heat map to locate the most likely membrane-active domains of H4 based on sequence alignments of subsequences and aggregate normalized σ scores (Figure 3.8A). Interestingly, the top consensus sequence contains the first 24 amino acids of H4 and forms an amphipathic helix (Figure 3.8B) with a charge of +8.03 and a hydrophobic moment of

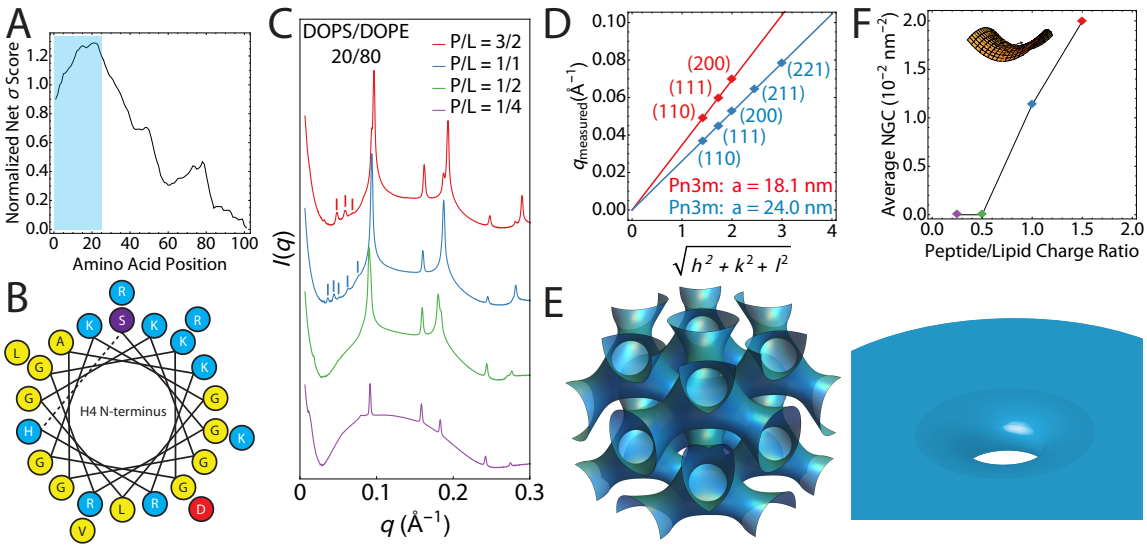


Figure 3.8: **Histone H4 N-terminus generates negative Gaussian curvature in cell membranes.**
(A) Machine learning screen of the entire histone H4 sequence predicts that potent membrane activity resides at the N-terminus (Residues 1-24 highlighted in blue). (B) Helical wheel projection of residues 1-24 of H4. Blue residues are cationic, red residues are anionic, yellow residues are hydrophobic, and purple residues are hydrophilic and uncharged. (C) SAXS data demonstrates that H4 N-terminus induces NGC in cell membranes at P/L ratios 1/1 and 3/2. Red and blue vertical lines correspond to the labeled correlation peaks of the induced Pn3m cubic phases indexed in (D). Miller indices of reflections and lattice parameters of the cubic phases are labeled in (D). (E) Schematic of the Pn3m cubic phase (left) and membrane pore (right) induced by H4 N-terminus. (F) Plot of the absolute value of NGC induced by H4 N-terminus as a function of P/L charge ratio.

0.247 (SGRGKGGKGLGKGGAKRHRKVLRD, $\sigma = 1.266$).

3.3.3 Histone H4 N-terminus generates negative Gaussian curvature in cell membranes

To validate the machine learning predictions, we quantified the ability of H4 N-terminus to generate NGC in model membranes. The 24-amino acid N-terminus peptide obtained from machine learning (SGRGKGGKGLGKGGAKRHRKVLRD) was synthesized and incubated with small unilamellar vesicles (SUVs) at increasing peptide-to-lipid (P/L) charge ratios (1/4, 1/2, 1/1, and 3/2), and the induced membrane deformations were characterized using SAXS.

At $P/L = 1/4$, the H4 N-terminus restructured DOPS/DOPE (PS/PE) = 20/80 SUVs into an inverted hexagonal phase with a lattice parameter of $d = 7.9$ nm, consistent with correlation peaks at q-ratios of $\sqrt{1} : \sqrt{3} : \sqrt{4} : \sqrt{7} : \sqrt{9}$ (Figure 3.8C, purple). At $P/L = 1/2$, the H4 N-terminus restructured SUVs into a coexistence of an inverted hexagonal phase with a lattice parameter of $d = 7.9$ nm and a lamellar phase with a lattice parameter of $d = 6.9$ nm, consistent with q-ratios of 1:2:3 (Figure 3.8C, green). However, at $P/L = 1/1$, the H4 N-terminus induced a coexistence of three phases: (1) a Pn3m "double-diamond" cubic phase with a lattice parameter of $a = 24.0$ nm consistent with correlation peaks at q-ratios of $\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{6} : \sqrt{9}$, (2) an inverted hexagonal phase with a lattice parameter of $d = 7.8$ nm, and (3) a lamellar phase with a lattice parameter of 6.7 nm (Figure 3.8C, D, blue). At $P/L = 3/2$ with PS/PE 20/80, we observed a similar coexistence of phases, with the H4 N-terminus inducing a Pn3m cubic phase with correlation peaks at q-ratios of $\sqrt{2} : \sqrt{3} : \sqrt{4}$ and a lattice parameter of $a = 18.1$ nm, an inverted hexagonal phase with a lattice parameter of $d = 7.7$ nm, and a lamellar phase with a lattice parameter of 6.5 nm (Figure 3.8C, D, red). Thus, at $P/L = 1/1$ and 3/2, the H4 N-terminus restructured PS/PE 20/80 SUVs into Pn3m cubic phases rich in NGC (Figure 3.8E), consistent with machine learning predictions and in accordance with previous work [177]. To quantify the effect of increasing peptide concentrations on membrane-destabilization, we calculated the average amount of Gaussian curvature $\langle K \rangle$ generated by H4 N-terminus for all four P/L charge ratios according to the equation $\langle K \rangle = \frac{2\pi\chi}{A_0 a^2}$, where the surface area per unit cell, A_0, and Euler characteristic, χ, are constants specific to the cubic phase, and a is the lattice parameter. For Pn3m cubic phases, $\chi = -2$ and $A_0 = 1.919$. For P/L ratios 1/4 and 1/2, the H4 N-terminus only induced inverted hexagonal phases, so $\langle K \rangle = 0$. However, the H4 N-terminus generated Pn3m cubic phases at $P/L = 1/1$ and 3/2 with lattice parameters 24.0 nm and 18.1 nm, corresponding to average NGC values of $\langle K \rangle = -1.13 \times 10^{-2}$ nm^{-2} and $\langle K \rangle = -1.99 \times 10^{-2}$ nm^{-2}, respectively. As increasing amounts of H4 N-terminus are added to the lipids (from $P/L = 1/4$ to $P/L = 3/2$), we observe a dose-dependent increase in the absolute magnitude of NGC (Figure 3.8F). Altogether, this data demonstrates that H4-induced pore formation and cytotoxicity can be localized to the N-terminus of H4, and that

the magnitude of induced NGC compares favorably to other membrane-remodeling peptides and proteins.

3.3.4 Methods

3.3.4.1 Machine Learning Methods

Full details of the SVM-based machine learning classifier for membrane activity are described in prior work [177, 175, 176]. The classifier is used here to identify membrane-active segments of histone H4. All subsequences of length 10-25 amino acids are generated with Python and screened with the classifier. σ scores are tabulated for each peptide alongside other physicochemical characteristics like net charge and mean hydrophobic moment. Top hits ($\sigma >\sim 0.8$) are sorted by decreasing σ values and cross-referenced with both secondary structure predictions [128] and known secondary structure from crystallographic data (Uniprot ID: P62805). To generate a "heat map" for segments of H4 that contain a high probability for membrane activity, all "hit" subsequences with $\sigma > 0$ are aligned with the entire H4 sequence. A normalized net σ score is assigned to each residue, equal to the sum of all σ scores corresponding to peptide "hits" in which that residue appears, divided by the number of times that amino acid appears in a "hit" (Figure 3.8A). This ensures that residues are weighted by the magnitude of σ, which is proportional to the likelihood of membrane activity. The subsequence of amino acids that envelop the global maximum (threshold $\sigma > 0.8$) is the most likely membrane-active segment.

3.3.4.2 SAXS Methods

Peptides were purchased and synthesized in high purity (> 95% HPLC) using solid-phase synthesis (Pepscan). SUVs were prepared as previously described elsewhere [77]. Briefly, lyophilized DOPS (1,2-dioleoyl-sn-glycero-3-phospho-L-serine (sodium salt)) and DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine) from Avanti Polar Lipids were purchased. Lipids were dissolved in chloroform to 20 mg/mL stock solutions. DOPS and DOPE were mixed at a 1:4 molar ratio (DOPS/DOPE = 20/80). Chloroform was evaporated un-

der nitrogen gas and the lipids were further dried overnight under vacuum. Dried lipids were resuspended in 140 mM NaCl + 10 mM HEPES at pH 7.4 to a final concentration of 20 mg/mL. Aqueous lipid solutions were incubated overnight at 37°C. SUVs were prepared by sonication of resuspended lipids until clear (3 × 10 min). Monodispersity in size was obtained via extrusion through a 0.2 μm filter. Before use, peptides are solubilized in nuclease-free water. SUVs and peptides were mixed together at fixed peptide-to-lipid charge ratios and equilibrated at room temperature overnight. Precipitated peptide-lipid complexes were loaded into 1.5 mm glass quartz capillaries (Mark-tubes, Hilgenberg GmbH) and hermetically sealed with an oxygen torch. SAXS experiments were conducted at the Stanford Synchrotron Radiation Lightsource (BL 4-2) with monochromatic X-rays of energy 9 keV. Samples were incubated at 37°C and centfifuged before measurement. Scattered radiation was collected using a Rayonix-MX225-HE detector (pixel size 73.2 μm). The 2D powder diffraction patterns were azimuthally integrated into 1D patterns using the Nika 1.76 package [184] for Igor Pro 7.04 (Wavemetrics) and FIT2D [185]. For all samples, multiple measurements were taken at different times to ensure consistency. To determine the phases present in each sample, the integrated scattering intensity $I(q)$ vs. q was plotted using Mathematica. The q-measurements corresponding to peak positions were obtained and their ratios were compared to the permitted reflections for different liquid-crystalline lipid phases (e.g. lamellar, hexagonal, cubic). Lamellar phases index to integer ratios of 1:2:3 and hexagonal phases index to ratios of $\sqrt{1} : \sqrt{3} : \sqrt{4} : \sqrt{7} : \sqrt{9}$. Cubic phases observed in our experiments belonged to the Pn3m space group, which permits reflections at ratios of $\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{6} : \sqrt{8} : \sqrt{9}$. Measured q positions for the Bragg peak reflections were fitted to the equation $q_{measured} = \frac{2\pi}{a}\sqrt{h^2 + k^2 + l^2}$ where (hkl) are the Miller indices. The curvature-generating ability of the peptides were determined by calculating the lattice parameter a for the measured cubic phases. The quantity of induced NGC was calculated as the average NGC per unit cell volume using the equation $\langle K \rangle = \frac{2\pi\chi}{A_0 a^2}$, where χ is the Euler characteristic and A_0 is the surface area per cubic unit cell for each phase. Parameter values are $\chi = -2$ and $A_0 = 1.919$ for Pn3m cubic phases.

3.4 Summary

In Chapters 2-3, we discussed the design and development of a machine learning-based search tool to discover hidden membrane-remodeling activity in new and existing families of peptides and proteins. The applications are numerous, ranging from design of next-generation antimicrobials, renovation of obsolete antibiotics, and improvement of our understanding of the molecular basis of disease. Here, we summarized two key examples of how this paradigm of computational discovery and experimental validation can be combined to impact diverse fields. Dysfunctional mitochondrial remodeling is central to certain causes of acquired and congenital neurodegeneration, and histone antimicrobial and cytotoxic activities are relevant to neutrophil extracellular traps, atherosclerosis, and autoimmune diseases. It is not possible to summarize here the multitude of additional molecules discovered by this search tool to possess hidden membrane activity, but these include bacterial secretion systems, membrane repair proteins, neuropeptide hormones, mitochondrial apoptosis proteins, and many more. However, another strength of this machine learning classifier is its unexpected ability to identify molecules with multiple functions. As it turns out, many antimicrobial peptides are multifunctional, and possess receptor binding abilities in addition to direct membrane activity. In the next portion of the thesis, we transition our focus towards a less understood aspect of antimicrobial peptides, namely their immunomodulatory activities.

CHAPTER 4

Introduction to innate immunity, Toll-like receptors, electrostatics of polyelectrolytes, and statistical mechanics of receptor-ligand interactions

4.1 Innate Immunity, Cytokines, and Type I Interferons

The innate immune system consists of a variety of broadly nonspecific effector cells and molecules, which include the complement system, antimicrobial peptides, neutrophils, dendritic cells, and macrophages [269]. This is distinct from the adaptive immune system which relies on the memory and hyperspecificity of antibody-producing B cells and antigen-recognizing T cells. The innate immune system can sense common pathogen-associated molecular patterns (PAMPs) characteristic of infectious organisms, directly attack invading microbes, and also direct the adaptive immune system to mount a more precise response. The innate immune system can do this by inducing the secretion of the class of signaling molecules known as cytokines. Cytokines are small proteins that mediate cell-cell signaling in the immune system. Type I interferons (IFNs) are a subtype of cytokines that were initially described as important for anti-viral defense [270]. However, we now recognize that the function of interferons is much more general, and that other infectious agents can also induce type I IFN production. For example, type I IFNs play an important part in host defense against many bacterial infections. There are two components to type I IFN signaling—the first is gene induction and release of the cytokine, and the second is binding of the cytokine to a type I IFN receptor and its activation, leading to type I IFN gene activation and downstream orchestration of immunity. (There are over 400 known interferon stimulated genes.)

In addition to host defense, type I IFNs also contribute to maintenance of the haematopoietic stem cell niche, are relevant in certain cancers, play pivotal roles in many autoimmune disorders.

4.2 Toll-like Receptors

TLRs were first discovered in Drosophila and were found to comprise one arm of the innate immune system [271, 272]. Innate immune cells can detect microbial pathogens through pattern-recognition receptors (PRRs), which sense signature PAMPs from microorganisms, signal to the host the presence of infection, and trigger inflammatory responses. The TLRs, a family of PRRs, are transmembrane proteins that have specialized recognition of multiple ligand classes. TLR2 senses a wide range of lipoproteins/lipopeptides, peptidoglycans, and lipoteichoic acids from bacteria. Moreover, it heterodimerizes with either TLR1 or TLR6 to specifically recognize triacylated or diacylated lipopeptides, respectively [273, 274]. LPS, a major constituent of the outer wall Gram-negative bacteria, is detected by TLR4, and bacterial flagellin, a structural protein component of flagella, is recognized by TLR5 [275, 276]. Several TLRs are also known to detect nucleic acids and are implicated in antiviral responses. Double-stranded RNA (dsRNA), typically generated from viral replication, is recognized by TLR3 [277, 278], while TLR7 and TLR8, which are similar in structure, both respond to single-stranded RNA (ssRNA) [279, 280, 281]. TLR9 has been identified as the receptor for unmethylated CpG-rich DNA motifs of bacteria [282] and viral double-stranded DNA (dsDNA) [283, 284].

The interaction of TLRs with PAMPs activates signal transduction pathways that lead to the production of pro-inflammatory cytokines. In general, TLRs 1, 2, 4, 5, and 6 primarily detect molecules unique to bacteria, and thus, allows for efficient discrimination of self and non-self. On the other hand, TLRs 3, 7, 8, and 9 recognize nucleic acids, which are not exclusive to microbial pathogens (Figure 4.1). In these cases, discrimination of self and non-self is typically achieved by the accessibility of the ligands to TLRs. More specifically, this subgroup of TLRs are localized to intracellular compartments [287, 288, 289, 290], and

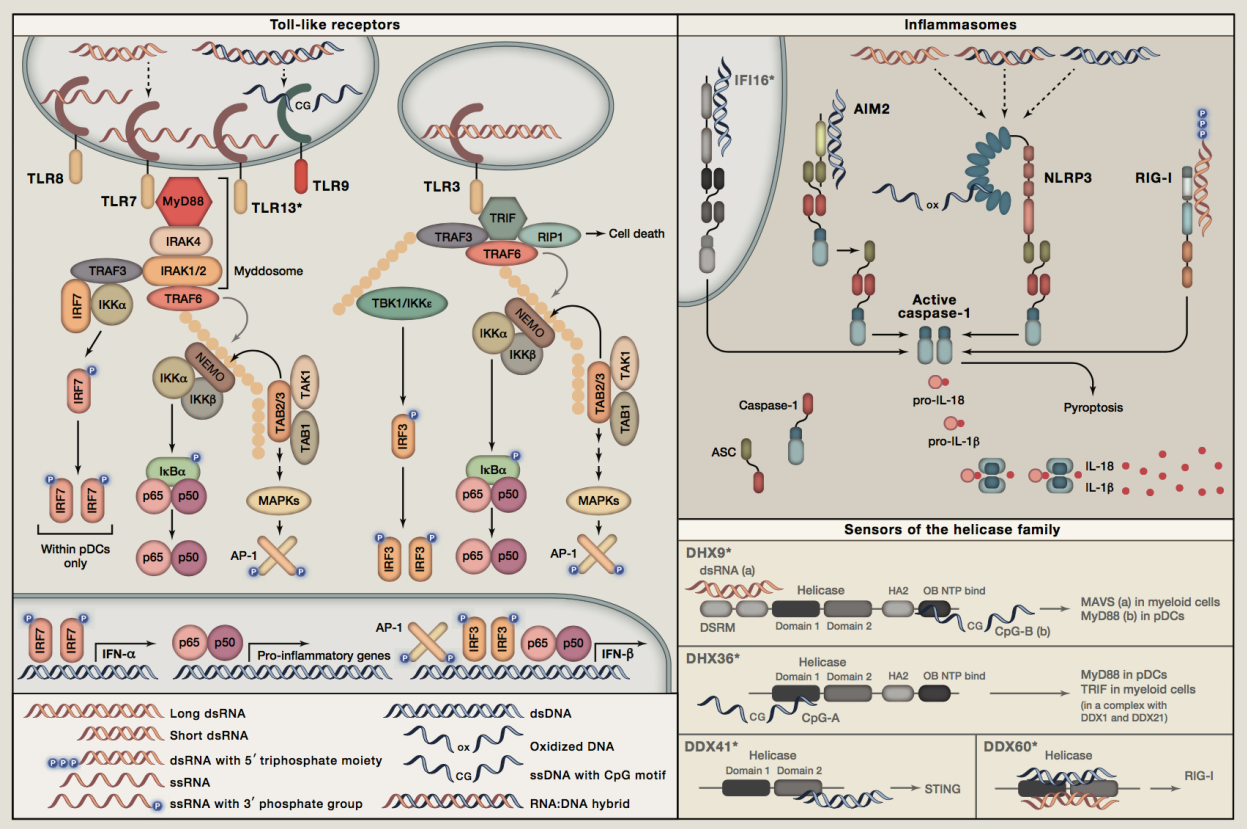


Figure 4.1: **Nucleic Acid Toll-Like Receptors and their Ligands.**
Toll-like receptors 3, 7, 8, and 9 are the primary endosomal nucleic acid receptors. TLR3 binds to dsRNA, TLR7/8 bind to ssRNA, and TLR9 binds to CpG DNA and dsDNA. Activation of these receptors drives intermediate signal transduction pathways that lead to nuclear translocation of transcription factors NF-κB (p65-p50 heterodimer), AP-1, IRF3, and IRF7, and upregulation of pro-inflammatory genes and type I interferon production. Reproduced with permission from [285, 286].

therefore, are exposed to foreign nucleic acids released into the compartments by endocytosed pathogens but not to the self-nucleic acids. However, under certain circumstances, such as poor clearance of apoptotic cells or cellular damage, self-nucleic acids may become aberrantly accessible to TLRs and can lead to autoimmunity [291, 292, 275, 293, 294]. Structural studies of the receptor—ligand complexes for TLR1—TLR2 heterodimer [295, 296], TLR2—TLR6 heterodimer [297], TLR3 homodimer [298, 299], TLR4 homodimer [295, 296], and TLR5 homodimer [300] revealed that they all share an "m"-shaped architecture, composed of two individual horseshoe-shaped TLR ectodomains [301]. These findings together suggest that the activation of TLRs may be governed by a common mechanism. The structural basis for activation of these TLRs by nucleic acids have been studied using X-ray crystallography [298, 302, 301, 303].

4.3 Immunomodulatory Properties of AMPs

Based on experiments both *in vitro* and *in vivo*, the immunomodulatory profiles of AMPs have been extensively studied, with human cathelicidin LL37 and defensins being the best-characterized [304]. Typical experiments that are conducted to assess the immunomodulatory behavior of AMPs include measurements of *in vitro* cytokine production from immune cells, real-time quantitative PCR monitoring of the expression of genes involved in inflammation, and challenge with TLR-ligands in mouse models of infection or inflammatory disease [305]. AMPs are usually co-administered with canonical immune ligands such as lipopolysaccharide (LPS), CpG oligonucleotides, and Poly(I:C) (viral double-stranded RNA mimic), and resulting changes in cytokine production are measured. From the body of literature, it is clear that AMPs are involved in triggering both pro-inflammatory and anti-inflammatory immune responses, and do so through distinct mechanisms that depend on biological context.

Examples of AMP-induced immunomodulation are legion: An early study in mice demonstrated that LL37 is a potent antisepsis agent that inhibits macrophage stimulation by bacterial endotoxins like LPS and lipoteichoic acid (LTA), and protects mice from endotoxemia [306, 307]. LL37 upregulated production of MCP-1, IL-8, but not TNF-α, and led to recruit-

ment of immune cells to the site of infection [308]. More specifically, LL37 chemoattracts leukocytes via the formyl peptide receptor-like 1 (FPRL1) [309]. In a similar manner, human β-defensins are chemotactic for immune cells by binding to the CCR6 receptor, a chemokine receptor found on dendritic cells and memory T cells [310]. In the human airway, defensins induce significant IL-8 production from airway epithelial cells, which lead to neutrophil chemotaxis [311]. In addition, human neutrophil peptide (HNP) defensins demonstrate significant direct chemotactic activity for monocytes [312]. HNP defensins are also involved in modulation of host antibody responses. HNPs directly induced proliferation of and cytokine production from T helper cells, enhanced systemic IgG responses, and fostered B cell-T cell interactions to bridge innate and adaptive immunity [313]. In mice, the murine β-defensin 2 (mDF2β), a small AMP produced in response to bacterial infections of mucosal tissue and skin, has also been shown to directly activate immature dendritic cells as an endogenous ligand for TLR4, resulting in the production of pro-inflammatory chemokines and cytokines [314]. The rhesus macaque θ-defensin RTD-1 inhibits production of pro-inflammatory cytokines induced by a variety of TLR ligands *in vitro*, including CpG DNA and LPS [315]. In a mouse model of bacterial sepsis, RTD-1 also suppressed pro-inflammatory cytokines and enhanced survival [316]. Granulysin is a human AMP produced by cytotoxic T-cells that forms membrane pores and induces apoptosis [317]. During Propionibacterium acnes infection, a fragment of granulysin (31-50v44w) was reported to inhibit production of IL-12 and chemokines MCP-1, IP-10, and MDC from human monocytes [132]. The porcine cathelicidin PR-39 is a broad-spectrum AMP that was observed to induce IL-8 production from macrophages. Truncation mutants of PR-39 were also found to differentially regulate the production of TNF-α [318]. Dual-functional synthetic host defense peptide mimics have been recently developed that capture both antimicrobial and immunomodulatory properties, while circumventing the issues common to peptide-based antimicrobials, such as hemolysis and short half-lives due to rapid degradation *in vivo*. These synthetic mimics of AMPs have demonstrated the ability to effectively kill Staphylococcus aureus and inhibit TLR2-mediated inflammation normally associated with the bacterial infection [319, 320]. One synthetic design in particular, Clavanin-MO, derived from a marine tunicate AMP, was shown to recruit

leukocytes, induce production of GM-CSF, IFN-γ, and MCP-1, stimulate secretion of the anti-inflammatory cytokine IL-10, and inhibit production of pro-inflammatory cytokines IL-12 and TNF-α [321].

We have summarized only $\sim$10 examples of endogenous and synthetic AMPs that possess immunomodulatory activity above; there are more that exhibit this behavior. To make sense of the vast diversity of evidence documenting the immune signaling activity of AMPs, we suggest a conceptual framework for AMP activity based on recent structural studies, and ask whether we can deduce deterministic rules governing AMP-mediated immunomodulation. To do this, we first analyze the molecular interactions between AMPs, nucleic acid immune ligands, and their respective TLRs.

4.4 AMP Interactions with TLRs and Relevance to Autoimmune Disease

While activation of TLRs has been shown to modulate and induce the production of AMPs from immune cells [322, 323, 324, 325], the direct immunomodulatory effects of AMPs on TLRs are less clear. AMPs have been found within or co-localize and interact with a variety of immune cells, including dendritic cells, monocytes, B cells, and T cells [326, 327, 308, 328]. AMPs are also produced by a variety of non-canonical immune cells such as keratinocytes [329]. A multitude of studies have demonstrated links between AMP expression, modulation of innate immune receptors like TLRs, and the subsequent upregulation or downregulation of pro-inflammatory and anti-inflammatory cytokines. We describe recent studies of AMP interactions with TLRs, and how an understanding of the structural rules underlying TLR activation can enable prediction of immunomodulatory abilities of a broad range of molecules.

4.4.1 dsDNA binding and TLR9 activation

TLR9 is typically thought of as a sensor for CpG DNA, which is relevant to viral and bacterial infections [303] (Figure 4.1). However, in a landmark study, Lande et al. demonstrated that human cathelicidin LL37 could potently activate plasmacytoid dendritic cells (pDCs) by forming immune complexes with human genomic dsDNA, accessing endosomal compartments, and binding to TLR9 [330, 110]. In the normal course of infection, LL37 enhances appropriate immune responses by binding to microbial DNA and inducing TLR9 activation. However, binding of LL37 to self-DNA and inappropriate activation of pDCs can break immune tolerance to self-nucleic acids and lead to autoimmunity. Indeed, overexpression of LL37 in autoimmune diseases like lupus [331] and psoriasis [332] has been linked to exacerbation of inflammation via TLR9 hyperactivation in pDCs and keratinocytes [333]. Interestingly, neutrophil extracellular traps (NETs), which are prevalent in autoimmune diseases [334, 335], have been shown to release LL37-DNA complexes, which contribute to dysregulation of inflammation in lupus via pDC activation [331]. Intracellular LL37 can also enhance DNA-mediated TLR9 activation in B cells [336], in addition to dendritic cells and macrophages [337].

However, the ability to enable immune recognition of DNA and induce TLR9 activation in immune cells is not unique to LL37. Subsequent studies have shown that the cationic AMPs human β-defensin 2 (hbD2), human β-defensin 3 (hbD3), and lysozyme can also co-assemble with dsDNA to activate pDCs, in a manner analogous to LL37 [338]. In the context of the immune response to malaria, histone-DNA complexes from Plasmodium falciparum were shown to activate dendritic cells via TLR9 [339]. Histones have also been well-characterized as antimicrobial peptides [266, 265], in addition to their DNA-binding properties, suggesting that histones play dual roles in microbial defense and immunomodulation. Other DNA-binding proteins have also demonstrated the ability to enable TLR9 recognition of DNA. High-mobility group box-1 protein (HMGB1) is a nuclear DNA-binding protein released from necrotic cells that forms immune complexes with DNA and stimulates cytokine production through a TLR9-dependent pathway [340]. Other members of the

HMGB family also facilitate recognition of nucleic acids by PRRs [341, 342, 343]. Interestingly, even pathogenic amyloids, some of which have been shown to have hidden antimicrobial activity [344, 345], are able to induce chemoattraction [346] and bind to dsDNA and induce type I IFN production via TLR9 [347].

One potentially interesting question to ask is whether TLR9-activation by AMP-DNA immune complexes and the subsequent induction of an adaptive immune response resulting in autoantibody production against dsDNA leads to a self-amplifying pathway that further potentiates TLR9 activation. In a study of immune complexes obtained from lupus patients, anti-dsDNA autoantibodies were found to form complexes with DNA, enter endosomal and lysosomal compartments, and activate dendritic cells through cooperation of CD32 and TLR9 [348]. In a separate study, immune complexes formed between anti-nucleosome autoantibodies and mammalian DNA were shown to activate TLR9 in B cells [291].

From the examples above, it is evident that immune amplification of TLR9-mediated inflammation can be triggered by the formation of various protein-DNA complexes. Furthermore, this phenomenon is general and occurs across a broad range of molecules, including but clearly not limited to AMPs. What is equally clear from these studies is that a large number of natural AMPs and related molecules can induce TLR9 signaling through dsDNA binding, but many other molecules that also bind to dsDNA cannot. To delineate the necessary and sufficient criteria for immune activation by AMP-dsDNA complexes, we utilized high-resolution small-angle synchrotron X-ray scattering (SAXS) to solve the structures of complexes formed between dsDNA and AMPs, and correlated these findings with measurements of pDC IFN-α production [349]. We found that cationic molecules (such as AMPs) can form columnar nanocrystalline complexes with dsDNA, and that the inter-DNA spacing between parallel DNA columns changes with the identity of the cationic molecule. For 15 AMP-DNA and polycation-DNA complexes, measured inter-DNA spacings were correlated with induced IFN-α production from pDCs (Figure 4.2). Two functional classes of complexes were found: Complexes with inter-DNA spacings in the range of d = 3-4 nm led to strong amplification of IFN production (e.g., LL37-DNA and hbD3-DNA), while those with inter-DNA spacings outside of this range, $d < 3$ nm or $d > 4$ nm, resulted in significantly lower

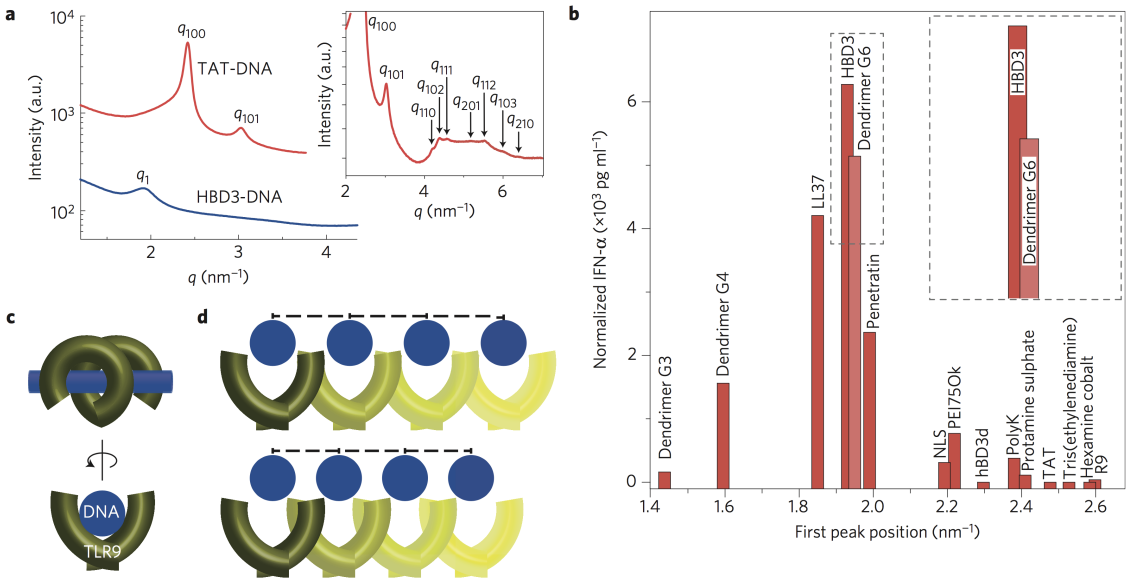


Figure 4.2: **Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation**
(A) SAXS data showing differences in structure between highly immunogenic complexes (HBD3-DNA) and non-activating complexes (TAT-DNA). (B) Structural selection rule for TLR9 activation by nanocrystalline columnar polycation-DNA complexes. Complexes with inter-DNA spacings in the range 3-4 nm amplify IFN-α production several orders of magnitude, but complexes much smaller or larger than this range do not. (C, D) Schematic of polycation-DNA complexes binding to TLR9. Complexes with inter-DNA spacings well-matched with the steric size of TLR9 lead to amplification, while those that are mismatched do not. Reproduced with permission from [349].

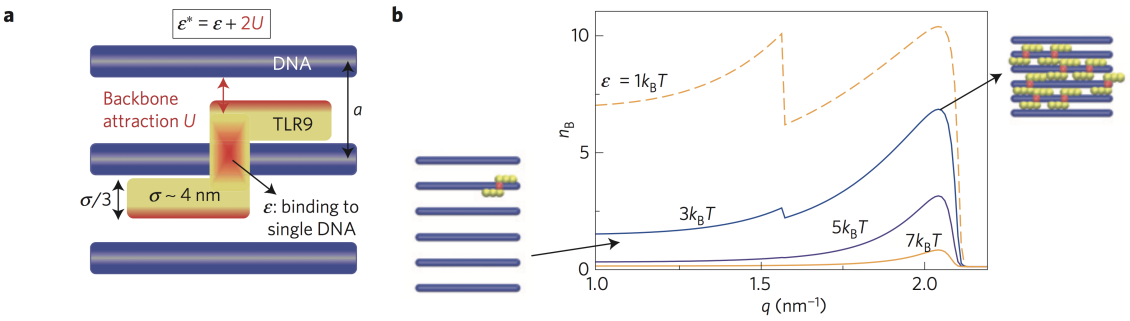


Figure 4.3: **TLR9 binding to DNA bundles exhibits "super-selective" behavior in agreement with experimental observations.**
(A) Schematic representation of coarse-grained model of DNA binding to a TLR9 dimer. (B) At very large inter-DNA spacings, few TLR9 dimers are bound to the DNA rods (snapshot, left). However, as the inter-DNA spacing is reduced into the range well-matched with the steric size of TLR9, the total number of bound TLR9 drastically increases (snapshot, right). Reproduced with permission from [349].

levels of IFN production, despite the fact that both classes of complexes have roughly equivalent levels of endosomal access. Interestingly, the optimum spacing of 3-4 nm is strongly suggestive, as it is approximately commensurate with the steric size of TLR9. Electrostatic interactions between the AMP-dsDNA complex and TLR9 are optimized when the inter-DNA spacing is in this range. Accordingly, the presentation of a spatially-periodic "grill" of parallel dsDNA by the nanocrystalline complex allows multivalent intercalative binding of the AMP-dsDNA complex to a cluster of TLR9 embedded in the endosomal membrane [350].

To understand these effects further, we formulated a mathematical model and conducted Monte-Carlo simulations of TLR9 binding to AMP-dsDNA complexes. As the inter-DNA spacing between parallel DNA ligands was varied, we observed different regimes of binding. At inter-DNA spacings smaller than the steric size of TLR9 (< 3 nm), low numbers of TLR9 were bound to the DNA due to sterically reduced accessibility to binding sites. Similar effects were observed when the inter-DNA spacings were too large (> 4 nm). However, as the inter-DNA spacing approached the optimal range (3-4 nm), we observed a drastic increase in the number of bound TLR9, experimentally correlating with the two orders of magnitude higher levels of immune activation and cytokine production [349, 350] (Figure 4.3).

4.4.2 dsRNA binding and TLR3 activation

Similar to TLR9 and CpG DNA, TLR3 was initially shown to sense viral dsRNAs in the context of infection [351] (Figure 4.1). However, recent work has shown that TLR3 also plays an important role in sensing skin injury by binding to non-coding self-dsRNA [352, 353]. TLR3 is expressed at especially high levels in human keratinocytes, and plays a role in inducing regeneration after skin injury [354]. Similarly, keratinocytes also produce large amounts of AMPs, and are important cellular mediators of microbial defense in the skin [355, 356, 329]. Although many studies have shown that AMPs can influence TLR9 activation via DNA binding, AMPs have also been reported to modulate TLR3-mediated immune responses to dsRNA. LL37 and other cationic peptides have been shown to enhance or inhibit TLR3 signaling by viral dsRNAs [357, 358, 359]. Other RNA-binding proteins have also been implicated in enhancement of TLR3 activation in immune cells [360, 361]. In Chapter 5, we describe in more detail how the crystallinity of dsRNA-AMP complexes can influence TLR3 activation, cognate to our findings for TLR9 [362].

4.4.3 ssRNA binding and TLR7/8 activation

As with nucleic-acid-sensing TLR9 and TLR3, the recognition of ssRNA by TLR7 and TLR8 [279, 280, 281] is also crucial in the defense against invading pathogens. Studies have demonstrated that TLR7 and TLR8 recognize guanosine (G)- and uridine (U)-rich viral ssRNA [279], ssRNA viruses such as vesicular stomatitis virus (VSV) and influenza [281], and more recently, bacterial and mitochondrial RNA [363]. However, the aberrant transportation of self-nucleic acids to endosomes enables the activation of TLRs, which can promote autoimmunity. In fact, self-ssRNA has been shown to act as ligands for TLR7 and TLR8 [364] and their internalization into endosomal compartments exacerbates autoimmune diseases such as systemic lupus erythematosus (SLE) [365], psoriasis [366], and antiphospholipid syndrome (APS) [367]. Interestingly, expression of TLR8, but not TLR7, is correlated to disease activity in patients with systemic arthritis, which suggests that independent signaling through TLR8 contributes to the pathogenesis of the disease [368, 369]. Furthermore,

oligodeoxynucleotides have been shown to antagonize nucleic-acid-sensing TLRs to alleviate disease conditions for psoriasis [370] and SLE [371] mouse models. Like we have observed for TLR9 and TLR3, these findings also implicate a role of TLR7 and TLR8 in the development of autoimmunity and excessive inflammation.

Normally, self-nucleic acids are rapidly degraded in the extracellular environment and are unable to reach the endosomal compartments [287]. However, self-nucleic acids that are aberrantly transported into endosomes can access nucleic-acid-sensing TLRs and trigger pDC activation, promoting autoimmunity (Figure 4.1). In the case of psoriasis, self-DNA forms complexes with LL37 and gains access to endosomal TLR9, which triggers pDCs to produce IFN-α and initiate the autoimmune process [330]. LL37-self-RNA complexes have been observed in psoriatic skin in vivo [366] and high levels of LL37 have also been reported in a variety of chronic inflammatory diseases, such as rosacea [372], rheumatoid arthritis [373], ulcerative colitis [374], chronic nasal inflammatory disease [375], sarcoidosis [376], and cystic fibrosis [377]. To examine the possible involvement of TLR7 and TLR8 in the pathogenesis of psoriasis, Ganguly et al. investigated whether LL37 is also capable of interacting with self-ssRNA released by dying cells to activate immune cells. Similarly, LL37 was found to complex with self-ssRNA via electrostatic interactions and reach the endosomal compartments of dendritic cells, inducing TLR7 activation in pDCs and secretion of IFN-α. Moreover, LL37-self-ssRNA complexes also triggered the activation of TLR8 in mDCs to produce TNF-α and IL-6. Interestingly, TLR7 in pDCs and TLR8 in mDCs were activated by self-ssRNA mixed with LL37, but not when self-ssRNA was administered alone or mixed with scrambled peptide GL37. Furthermore, complexation of self-ssRNA with LL37 was shown to protect self-ssRNA against degradation from extracellular RNases, which normally prevent RNA from accessing intracellular compartments [366]. From these findings, it will be interesting to see if the complexation of AMPs with RNA ligands activate TLR7 and TLR8 in a cognate, structural manner similar to that identified for AMP-dsDNA complexes with TLR9 [349] and AMP-dsRNA complexes with TLR3 [378]. Present efforts are under way to do such studies.

4.5 Electrostatics of Biological Polyelectrolytes

A large amount of evidence suggests that cationic proteins like AMPs can condense nucleic acids into insoluble complexes with marked effects on the immune system. To fully understand and appreciate the underlying physics of AMP-nucleic acid interactions, it helps to summarize prior work done on condensed phases of biological polyelectrolytes. DNA and other biological polyelectrolytes (such as F-actin and microtubules) carry uncompensated negative charge. In biologically relevant aqueous conditions, these charges are strongly screened by the dielectric response of water, and by salt solutions. However, this simple picture is incomplete. Counterion entropy can result in surprisingly strong interactions between charged objects in water despite strong screening, via coupling between osmotic and electrostatic interactions. Many experiments have demonstrated that like charged objects repel because of the osmotic pressure of squeezed counterions, and oppositely charged objects repel because of the entropy gain of counterion release. For example, the free energy gain upon binding between two macro-ions scales as kT multiplied by the number of counterions released (if one neglects ion hydration effects), which can constitute a large thermodynamic driving force.

Poisson-Boltzmann (PB) theory provides quantitative predictions of ion distribution about charged rods [379, 380, 381]. In this mean field approximation, properties associated with discrete ions are not considered, including correlations between ions or finite ion size. PB predicts that like-charged objects, such as polyelectrolytes, repel in all salt conditions, including multivalent salts. Increasing the valence of the ions increases the screening contribution and weaken repulsion, but does not invert the sign of the repulsive interaction to attraction.

The above prediction is clearly not in agreement with experiments. Condensation of polyelectrolytes by multivalent ions or/and macroions is widely observed in biological and biomedical contexts. Polyelectrolytes chains such as DNA can "condense" or collapse into a compact phase from solution as the concentration of oppositely charged multivalent ions or macroions increases. Examples include DNA packing in viruses [382, 383], in bacte-

ria [384, 385], and in many of the DNA complexes that do not lead to TLR9 activation [349]. From a considerable body of theoretical and computational work, we know why highly charged polyelectrolytes can form aggregates. In physical situations with high surface charge densities or with multivalent ions, the organization and dynamics of condensed ions surrounding the polyelectrolyte are important. A series of pioneering Monte Carlo simulations [386] demonstrated attraction between like-charged DNA. In the last two decades, a large number of theoretical investigations have focused on the physical origins of like-charge attraction [387, 388], and on the changes induced in the organization of the polyelectrolyte [389, 390]. Correlations between condensed counterions on the polyelectrolyte surface can generate attractions [380, 391, 386, 392] and organize condensed structures of polyelectrolytes. Oosawa [393] showed that correlations between thermal fluctuations of the condensed counterion layers can result in attractions. These ideas have been refined by various groups [394, 395]. At close distances, local fields due to the spatial arrangement of charges on a macroion can lead to patterns of counterion binding [396, 397, 398]. If these ions are ordered, attractions can result as counterions arrange themselves along the surfaces of adjacent macroions in complementary patterns. An elegant picture of interacting Wigner crystals was developed by Rouzina and Bloomfield [399], Shklovskii [400], and Lau and Pincus [401]. Recent work in the "strong coupling" limit [402, 403], where the counterion charge or surface charge density exceed the range of applicability of Poisson-Boltzmann theories [404] and where correlations between counterions are strong, has predicted the spatial dependence of attractive forces [405, 404, 406]. Using a new analytical Wigner-crystal based formulation of strong coupling, amaj and Trizac [407] were able to eliminate divergences from the virial expansion approach, and obtain results on charged plates that agree well with Monte Carlo simulations.

The condensing agent for the DNA polyelectrolyte can be a multivalent macroion like a peptide or protein rather than a simple multivalent cation. Examples include chromosomes [408, 409, 410, 411, 412], and synthetic gene delivery systems based on cationic polymers [413, 414] and dendrimers [415, 416]. The behavior of these condensed DNA phases becomes even richer with increasing complexity in the condensing cations. Lipids are amphiphiles that are structurally quite different from facially amphiphilic antimicrobial peptides con-

sidered in this review [349]. DNA-cationic lipid complexes have recently received extensive experimental and theoretical scrutiny, since they are empirically known to be efficient gene delivery systems [417, 418, 419, 420, 421, 422, 423, 424, 425, 53]. A polymorphism of different self-assembled structures in these DNA-lipid complexes (such as the lamellar, hexagonal, and inverted hexagonal phases) with different transfection efficiencies has been found using synchrotron x-ray scattering [53, 426, 427]. It is interesting to note that other biological polyelectrolytes can participate in similar forms of electrostatic self-assembly. Cytoskeletal polyelectrolytes include F-actin, intermediate filaments, and microtubules. These polyelectrolytes have different effective diameters, surface charge densities, and flexibilities, and can interact with oppositely charged cations, or cationic proteins, or cationic amphiphilies to generate a diverse range of structures [428, 429, 54, 430, 431, 432].

4.6 Statistical Mechanics of Multivalent, Superselective Receptor-Ligand Interactions

Now that we have reviewed the electrostatic interactions underlying the formation of nucleic acid nanocrystals, we now explore the statistical mechanics underlying their interactions with target Toll-like receptors, using TLR9 as an example. The immune system is often thought of as an example of molecular recognition par excellence. The adaptive immune system employs antibodies that bind to very specific motifs on infectious agents, marking them for destruction. However, molecular recognition in the innate immune system is more primitive, with non-specific recognition of a variety of ligands that have similar structural and physiochemical character. In molecular and cell biology, there is a common notion that there is a kind of dichotomy between specific and non-specific interactions. We think of the failure of molecular recognition in terms of specific interactions becoming non-specific. Recent work, including our work here, show examples of the opposite, where non-specific interactions of PAMPs (such as dsDNA) with the innate immune system can actually be organized into amplified, specific interactions, with consequences for autoimmune disorders and targeted drug delivery.

Ligand-receptor signaling is a cornerstone of molecular biology. Cells and tissues communicate by sending and receiving messages in the form of proteins, small molecules, and nucleic acids. When engaging stochastic cellular phenomena like signaling, the first sensible step is to describe such interactions as a single ligand binding to a single receptor, triggering downstream activation of other signaling molecules. There are many examples where multivalent interactions are required in order to obtain a cellular response. These are interactions in which multiple ligands of one object bind to multiple receptors of another. There are many examples of this behavior in the immune system, where it is important to discern signal from noise. Multivalent antigens can cross-link multiple IgM receptors at once on the surface of B cells, leading to a T-cell independent activation of the immune system [269]. Multivalent antigen-receptor interactions by multivalent CD40 ligands can also trigger B cell activation and immunoglobulin synthesis [433]. Another example of multivalent interactions includes the binding of hyaluronic acid (HA) to CD44 receptors on the surfaces of cells, mediating cell adhesion to the extracellular matrix [434]. Lesley et. al. found that an increase in the oligomer size of HA led to a concomitant superlinear increase in avidity of CD44 binding. Interestingly, proliferation of the HA-rich pericellular matrix has been associated with amplification of CD44 signaling in inflammation [435] and tumor formation [436]. Thus, the HA-CD44 interaction has become an important target for the development of anti-cancer drugs. Cancer-targeting nanoparticles coated with multiple ligands also take advantage of superselectivity by inducing multivalent interactions with rogue cells that express a higher density of a particular receptor relative to normal cells [437]. In the context here, we have found that a self-assembled LL37-DNA complex can cross-link multiple TLR9 receptors within the endosomes of pDCs, leading to receptor recruitment and amplification of the immune response [349]. These examples highlight the importance of multivalent ligand-receptor signaling in normal and abnormal physiology, and in the treatment of disease.

The statistical mechanics of multivalent ligand-receptor interactions has been recently examined experimentally, theoretically, and computationally, and it leads naturally to the concept of superselectivity. Superselectivity is defined by a superlinear (steeper than linear) relationship between the surface density of receptor sites and the surface density of bound

ligands. This under-appreciated aspect of multivalent interactions allows a small change in receptor density to drive a large change in the number of bound ligands, in a manner that depends on the degree of multivalency. Using a model system of multivalent "guest/host" interactions, in which "host" β-cyclodextrin (β-CD) is grafted onto a hyaluronic acid polymer backbone, and "guest" ferrocene functionalized on a self-assembled monolayer (SAM) surface, Dubacheva et al. discovered that the amount of bound ferrocene increased faster than linearly with increases in the density of surface binding sites [438]. Interestingly, receptor recruitment and clustering was also observed, which is analogous to the TLR9 work above. In a different experiment, spatiotemporal control over the monomer ordering in a supramolecular polymer was achieved, via the binding of ssDNA to cationic receptors in a multivalent manner. The experimental system is based on 1,3,5-benzenetricarboxamide derivatives (BTA). Supramolecular copolymers are made by assembling two BTA monomer subtypes, neutral and charged. The charged BTA monomers can interact electrostatically and act as receptors. The introduction of ssDNA resulted in the recruitment and binding of receptors within the assembled copolymer. FRET results showed that exposure of ssDNA triggered clustering of the cationic subtypes of BTA receptors via reorganization of the polymer, forming "islands" of similar receptors in a superselective fashion [439]. In both cases, these model systems are validated using statistical mechanical models. For example, these models predict that a relatively low receptor-ligand binding affinity and a threshold receptor concentration are required for superselectivity. Above the threshold concentration, ligands saturate the receptors very quickly despite each individual interaction being weak. Below the threshold concentration, however, ligands do not bind appreciably to the targets [440]. With this perspective, one can see that what is often called binding specificity (greatly increased binding of one specific component) and binding amplification are in fact related concepts. Superselectivity, a downstream consequence of multivalent interactions, can increase binding specificity, receptor recruitment, and subsequent amplification of downstream responses.

Assuming the size and the shape of the bundles as observed in the experiments, it is possible to predict the amplification of the receptor activation based on a simple coarse-grained theory and statistical mechanics of multivalent binding. The model contains steric

and electrostatic interactions between TLR9 receptors and DNA bound in a bundle of a fixed size (a crystallite with 6×6 DNA helices of length 20 nm) but with variable DNA-DNA spacing a. As measured in the experiments, we assume that the electrostatic attraction between a negatively charged single DNA molecule and a TLR9 receptor is not sufficiently strong to result in appreciable binding affinity ϵ. The core assumption of the model is that the electrostatic interactions are amplified by the backbone attraction between the TLR9 receptor and DNA molecules arranged in a grill-like pattern. The effective binding affinity for a bundle interacting with a receptor is then $\epsilon^* = \epsilon + 2U(a)$, where the backbone attraction contribution $U(a)$ is

$$U(a) = -B\frac{e^{-\kappa(a-\frac{\sigma}{3})}}{a-\frac{\sigma}{3}} + \left(\frac{\sigma}{12\left(a-\frac{2\sigma}{3}\right)}\right)^{12}$$

Here σ is the typical width of the TLR9 receptor, $B \sim 20k_BT$ (see SI of [349]) is the magnitude of the electrostatic interaction and $\kappa^{-1} = 0.8$ nm is the Debye screening length at physiological conditions. The backbone attraction depends on the lattice spacing a: at large a, it is negligible, but it increases sharply when the spacing becomes of the order of σ. At very small DNA separations, due to the steric effects, it decreases again. Through our model, the binding affinity–and with it the expected number of bound receptors n_B to one DNA bundle—clearly depend on the lattice spacing. A bundle is considered bound to the cell, if there is at least one receptor bound to it. The expected number of bound bundles is related to the free energy F of bundle-receptor binding, $\langle N_{DNA}\rangle \propto e^{-\frac{F}{k_BT}}$, and the total number of active receptors, N_A, is

$$N_A(a) = n_B(a)\langle N_{DNA}\rangle \propto n_B(a)e^{-\frac{F(a)}{k_BT}}$$

The free energy F can be obtained from computer simulations and also estimated by a simple Langmuir adsorption theory. It has been shown in [349] that the number of bound receptors to a bundle exponentially depends on the binding affinity ϵ^* and through it on the spacing a, while the total activation N_A exhibits a much sharper double exponential dependence, which is a basis of the observed superselectivity of the TLR9 activation towards the DNA-DNA spacing in the DNA-peptide bundles (Figure 4.3). Thus, statistical mechanical modeling and

molecular simulation of TLR9 binding to AMP-DNA nanocrystals strongly recapitulates the experimental observations.

4.7 Nanocrystalline dsDNA complexes influence the pathophysiology of autoimmune disease

From the work outlined above, it is quite evident that nanocrystalline organization of DNA by various peptides and proteins have strong immunological consequences. Here, we describe two experimental examples of how the nanocrystalline ordering of dsDNA complexes can lead to amplification of inflammation in two unique disease contexts. In the first, we study the structure of curli amyloid-DNA complexes found in *Salmonella* enteric biofilms, which stimulate systemic autoimmunity via production of type I interferons. In the second, we study the structure of the chemokine CXCL4 bound to DNA, which is a key step in the pathogenesis of systemic sclerosis, a chronic autoimmune disease characterized by plasmacytoid dendritic cell hyperactivation.

4.7.1 Bacterial amyloid curli acts as a carrier for DNA to elicit an autoimmune response via TLR2 and TLR9

4.7.1.1 Abstract

Bacterial biofilms are associated with numerous human infections. The predominant protein expressed in enteric biofilms is the amyloid curli, which forms highly immunogenic complexes with DNA. Infection with curli-expressing bacteria or systemic exposure to purified curli-DNA complexes triggers autoimmunity via the generation of type I interferons (IFNs) and anti-double-stranded DNA antibodies. Here, we show that DNA complexed with amyloid curli powerfully stimulates Toll-like receptor 9 (TLR9) through a two-step mechanism. First, the cross beta-sheet structure of curli is bound by cell-surface Toll-like receptor 2 (TLR2), enabling internalization of the complex into endosomes. After internalization, the curli-DNA immune complex binds strongly to endosomal TLR9, inducing production of type

I IFNs. Structural analysis using small-angle X-ray scattering revealed that incorporation of DNA into curli fibrils resulted in the formation of ordered curli-DNA immune complexes. Curli organizes parallel, double-stranded DNA rods at an inter-DNA spacing that matches up well with the steric size of TLR9, which enables multivalent intercalative binding to ordered arrays of TLR9 in the endosomal membrane.

4.7.1.2 Curli forms nanocrystalline complexes with DNA to activate TLR9

Previous data demonstrates that activation of TLR9 and induction of type I IFNs by the curli-DNA complex requires activation of TLR2 [441]. Molecules such as LL37 and defensins can amplify TLR9-mediated interferon production by binding DNA and forming immune complexes, although other molecules that bind DNA equally well do not [349]. Recent work has shown that a key determinant for potent TLR9-dependent interferon secretion is the inter-DNA spacing within the protein-DNA complexes. Inter-DNA spacings that match the steric size of TLR9 receptors (3.0-4.0 nm) allow for multivalent presentation of DNA ligands to TLR9, driving receptor recruitment, binding amplification, and immune activation [349]. We therefore measured the inter-DNA spacing within curli-DNA complexes.

For these experiments, we used a synthetic peptide corresponding to the fourth and fifth repeats of CsgA ($CsgA_{R4-5}$) [442]. The peptide was fibrillized in the presence or absence of CpG DNA. We measured the inter-DNA spacing within $CsgA_{R4-5}$-DNA complexes using synchrotron small-angle X-ray scattering (SAXS). In the scattering profiles, we observed a sharp diffraction peak at 0.151 $Å^{-1}$ for the $CsgA_{R4-5}$-DNA complex but no significant scattering for the $CsgA_{R4-5}$ alone (Figure 4.4A). We found that $CsgA_{R4-5}$ organizes CpG DNA into an ordered columnar array that interacts with TLR9 at an inter-DNA spacing of 4.16 nm (Figure 4.4B). Although this value of inter-DNA spacing is slightly larger than the optimal inter-DNA spacings for interaction with TLR9 of 3.0-4.0 nm, it results in potent amplification of interferon production. By organizing DNA into a columnar lattice with an inter-DNA spacing commensurate with the steric size of TLR9, curli can maximize TLR9 binding to DNA, leading to the amplified type I IFN response observed (Figure 4.4C, bottom). Other

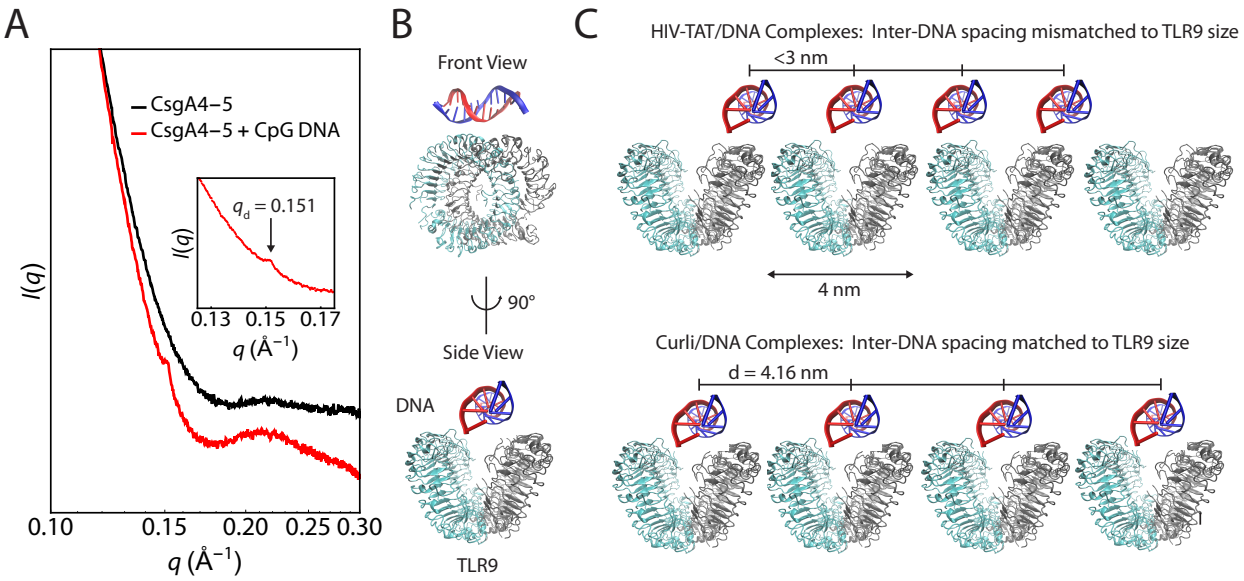


Figure 4.4: **The structure of CsgA$_{R4-5}$-DNA complexes explains their ability to induce type I interferon responses via TLR9.**
(A) SAXS diffraction profiles for CsgA$_{R4-5}$ amyloid fibrillized in the absence (black curve) and presence (red curve) of CpG dsDNA. No diffraction peaks indicating columnar structure are observed for the CsgA$_{R4-5}$ fibrils in absence of DNA, but fibrillization of CsgA$_{R4-5}$ in the presence of DNA results in a sharp diffraction peak at $q_d = 0.151$ Å^{-1} (inset). This indicates that CsgA$_{R4-5}$ organizes the CpG DNA into an ordered columnar lattice with inter-DNA spacing $d = \frac{2\pi}{q_d} = 4.16$ nm. (B) Schematic depicting CpG DNA binding to TLR9 from two orthogonal views. The anionic DNA binds electrostatically to the cationic paddles of TLR9. (C) Closest approach of adjacent TLR9 receptors on the endosomal membrane does not allow size-matching and multivalent binding to HIV-TAT/DNA complexes (top). This data was adopted from Schmidt et al [349]. In contrast, the inter-DNA spacing for CsgA$_{R4-5}$-DNA complexes (bottom) allows for amplification of binding from multivalent and electrostatic effects.

peptides, such as HIV-TAT, are capable of endosomal access but do not efficiently activate TLR9 [349] because they organize DNA into lattices with inter-DNA spacings significantly smaller than 3 nm, which does not allow effective contact with the cationic domains of TLR9 (Figure 4.4C, top). Finally, our data suggests that $CsgA_{R4-5}$ amyloid fibrillized in the absence of DNA does not form the ordered columnar lattices that are observed in the $CsgA_{R4-5}$-DNA complex (Figure 4.4A). This emphasizes that fibrillization of $CsgA_{R4-5}$ in the presence of DNA is required for its immunogenic activity.

4.7.1.3 Methods

$CsgA_{R4-5}$ was purchased from Biosynthesis Incorporated (Lot #T876-2) and dissolved in 500 μL hexafluoroisopropanol (HIFP) and concentrated by speedvac (Savant) for approximately 1 hour. The pellet was then resuspended in 500 μL DMSO. $CsgA_{R4-5}$ monomers were then purified using a Hi-Trap Column according to manufacturer's protocol (GE Healthcare, 17140701). $CsgA_{R4-5}$ was fibrillized alone or in the presence of 10 ng/mL CpG DNA (Invivogen, ODN 1826). Fibrils with and without CpG DNA were pelleted in Eppendorf tubes with centrifugation at 10,000 rpm for 20 minutes. All but 50 μL of supernatant was discarded. The fibrils were resuspended in remaining supernatant and loaded and sealed in quartz capillaries (Mark-tubes, Hilgenberg, GmbH). The samples were stored at 4°C until measurement. SAXS experiments were performed at the Stanford Synchrotron Radiation Lightsource (SSRL, Beamline 4-2) using monochromatic X-rays with an energy of 9 keV. The scattered radiation was measured using a Rayonix MX-225-HE detector (pixel size 73.2 μm). 2D powder diffraction patterns were integrated using the Nika 1.74 [184] package for Igor Pro 6.37 and FIT2D [185]. SAXS data were analyzed by plotting integrated scattering intensity against the momentum transfer $q$using Mathematica. Peak positions were measured by fitting diffraction peaks to a Lorentzian. The inter-DNA spacing d was obtained from the first peak position q_d by the formula $d = \frac{2\pi}{q_d}$.

4.7.1.4 Acknowledgements

Work in this manuscript was supported by National Institutes of Health grant R21AI126133 to C.T., R21AI119947 and Novel Research Grant from Lupus Research Institute to S.G. and C.T. E.Y.L. and G.C.L.W. acknowledge support from NIH R21AI122212 the T32 Systems and Integrative Biology Training Grant at UCLA (T32GM008185) and the T32 Medical Scientist Training Program at UCLA (T32GM008042). X-ray research was conducted at Stanford Synchrotron Radiation Lightsource, SLAC National Laboratory, supported by the US DOE Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

4.7.2 CXCL4-DNA nanocrystal formation enables multivalent recognition of self-DNA in the pathogenesis of systemic sclerosis

4.7.2.1 Abstract

The chemokine (C-X-C motif) ligand 4 (CXCL4) plays key roles in immunity and autoimmunity, as well as in hemostasis and antimicrobial defense. CXCL4 is overexpressed in the serum of patients with systemic sclerosis (SSc), a chronic autoimmune disease characterized by fibrosis and vasculopathy. CXCL4 has been proposed as an early serum biomarker of severe disease, but its contribution to the type-I interferon (IFN-I) gene signature in SSc and its role in disease pathogenesis is unclear. Here, we combine synchrotron small angle X-ray scattering (SAXS) and plasmacytoid dendritic cell (pDC) stimulation experiments to elucidate a previously unanticipated mechanism for CXCL4-mediated immune amplification in SSc. Synchrotron SAXS reveals that CXCL4 organizes "self" and microbial DNA into proinflammatory nanocrystalline complexes that drive TLR9-dependent pDC hyperactivation and IFN-α production. Furthermore, CXCL4 scaffolds clustered DNA ligands at a spacing well-matched with the steric size of TLR9, enabling multivalent intercalative binding to clustered arrays of TLR9 receptors in the endosomal membrane.The evidence adduced here illuminates a hidden function of CXCL4 and demonstrates a direct link between CXCL4

overexpression and the IFN-I gene signature in SSc.

4.7.2.2 CXCL4 condenses both self-DNA and bacterial DNA into columnar nanocrystalline complexes at an inter-DNA spacing optimal for TLR9 activation

CXCL4 complexes amplify inflammation by hyperactivating pDCs in a TLR9-dependent manner. We investigate the structural basis for CXCL4-enabled immune amplification via TLR9. CXCL4 is a hybrid α-helical/β-sheet molecule with a net cationic charge (+3.07 at pH 7.4). Secondary structure analysis reveals several cationic, amphipathic helices arranged around its circumference that can potentially bind to anionic DNA (Figure 4.5A, top). Calculation of the electrostatic surface potential of CXCL4 uncovers several solvent-exposed cationic surfaces available for multivalent interactions with the DNA phosphate backbone (Figure 4.5A, bottom). These findings indicate that CXCL4 has the requisite physicochemical properties to electrostatically bind to DNA in a multivalent fashion, similar to other AMPs like LL37 and human β-defensins.

To examine the mechanism for CXCL4-mediated autoimmune hyperactivation in systemic sclerosis, we solved the structures of CXCL4-DNA complexes using synchrotron small-angle X-ray scattering (SAXS). Previous studies have demonstrated that TLR9 activation in pDCs is strongly influenced by the inter-DNA spacing d between parallel DNA ligands in antimicrobial peptide-DNA complexes [349]. Specifically, inter-DNA spacings close to the steric side of TLR9 (d = 3-4 nm) allow for optimized multivalent binding of columnar DNA lattices to clustered arrays of TLR9 receptors. In addition to the inter-columnar spacing, the crystalline size (domain size L, which reflects the average number of dsDNA repeats in the nanocrystal) of each complex also influences Toll-receptor activation [378]. Nanocrystals with larger domain sizes have more DNA repeat units, allowing for a greater degree of multivalent presentation and binding to TLR9 dimers, which drives a stronger IFN response. To rationalize the observed immunological outcomes in the context of this mechanistic model, we measured the crystallinity parameters (inter-DNA spacing d and domain size L) of CXCL4-DNA com-

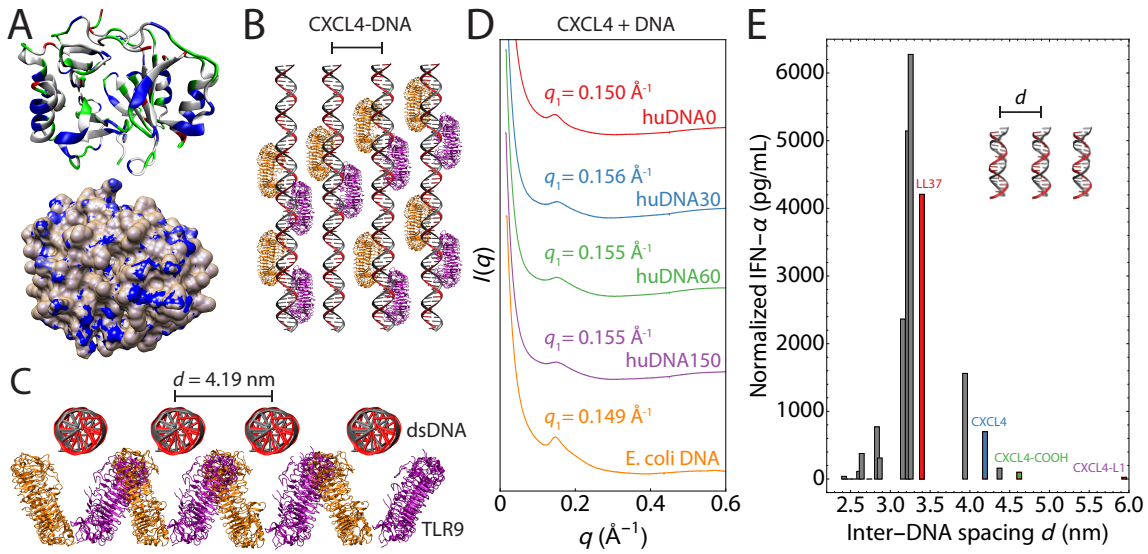


Figure 4.5: **CXCL4 condenses human and bacterial DNA into columnar nanocrystalline complexes at an inter-DNA spacing optimal for TLR9 activation.** (A) Structure of CXCL4 (PDB ID: 1F9Q) with cationic residues highlighted in blue and hydrophobic residues highlighted in white (top). Electrostatic surface potential of CXCL4 reveals multiple solvent-exposed cationic surfaces (blue) available for multivalent interactions with the DNA phosphate backbone (bottom). (B) Top-down view and (C) end-on view of CXCL4-DNA complexes binding to TLR9 dimers in the endosome. CXCL4 organizes human DNA into a columnar lattice with short-ranged order. The lattice is characterized by an average inter-DNA spacing of $d = 4.19$ nm, which is near the sterically optimal value for an arrangement that permits intercalative multivalent binding to a clustered array of TLR9 receptors, consistent with high levels of IFN production. (D) SAXS data of CXCL4 bound to human and bacterial DNA demonstrate that CXCL4 forms similar columnar lattices with full-length (huDNA0) and fragmented human DNA (huDNA30, huDNA60, and huDNA150 are human DNA sonicated for 30, 60, and 150 minutes respectively), as well as with bacterial DNA. First peak positions q_1 are indicated. (E) Relative levels of IFN-α production induced by CXCL4, CXCL4-COOH, and CXCL4-L1. Molecular structures were visualized in VMD. TLR9 structure (PDB ID: 3WPG) was obtained from previous work [303]. Electrostatic potential map of the CXCL4 surface was calculated in Chimera using the APBS plugin.

plexes with SAXS.

We find that CXCL4-DNA forms nanocrystalline columnar lattices with short-ranged order (Figure 4.5B, C). A DNA correlation peak is measured at $q_1 = 0.150$ Å^{-1} for unfragmented DNA and $q_1 = 0.149$ Å^{-1} for E. coli DNA, equivalent to inter-DNA spacings $d = 4.19$ nm and 4.21 nm, respectively. These inter-DNA spacings are near the sterically optimal value for interdigitation of clustered TLR9 receptors with CXCL4-DNA complexes, and previous studies suggest that complexes with these inter-DNA spacings should exhibit significant activation [349]. The observation that CXCL4-E. coli DNA complexes induce higher levels of IFN-α production than CXCL4-human DNA complexes is consistent with a larger measured domain size ($L = 17.2$ nm vs. 13.3 nm, Figure 4.5D) at similar inter-DNA spacings. In the presence of fragmented human DNA, CXCL4 formed columnar lattices with slightly smaller inter-DNA spacings ($q_1 = 0.155$-0.156 Å^{-1}, $d = 4.03$-4.05 nm) (Figure 4.5D). Interestingly, this smaller spacing (which sterically fits the size of TLR9 better and is closer to the optimal d of 3-4 nm) agrees with the finding that CXCL4 bound to fragmented DNA induces IFN-α more efficiently than CXCL4 bound to unfragmented DNA.

Since CXCL4 contains a highly cationic helical C-terminal heparin-binding domain (CXCL4-COOH: amino acids 47-70, Figure 4.6A), we investigated whether this portion of the molecule could also bind to DNA and induce IFN-α in pDCs. SAXS studies show that CXCL4-COOH binds unfragmented DNA into a columnar lattice with an inter-DNA spacing of $d = 4.62$ nm (Figure 4.6A), which is much larger than the CXCL4-DNA spacing ($d = 4.19$ nm). pDC activation experiments confirm that CXCL4-COOH induces less IFN-α than native CXCL4, although both induce significant IFN-α levels. Strikingly, fragmentation of human DNA shifts the inter-DNA spacings of CXCL4-COOH-DNA complexes to $d = 3.7$-3.8 nm, which is in the highly stimulatory range for TLR9 (Figure 4.6A). To determine whether mutation of cationic residues in CXCL4-COOH impacts DNA binding and TLR9 activation, a 3-amino acid mutant of CXCL4-COOH was synthesized and tested (CXCL4-L1, Figure 4.6B). CXCL4-L1 bound unfragmented human DNA into a hexagonal columnar lattice with a large inter-DNA spacing of $d = 5.94$ nm, far outside the stimulatory range. CXCL4-L1 was also unable to bind to fragmented DNA (Figure 4.6B). Both findings agree with the

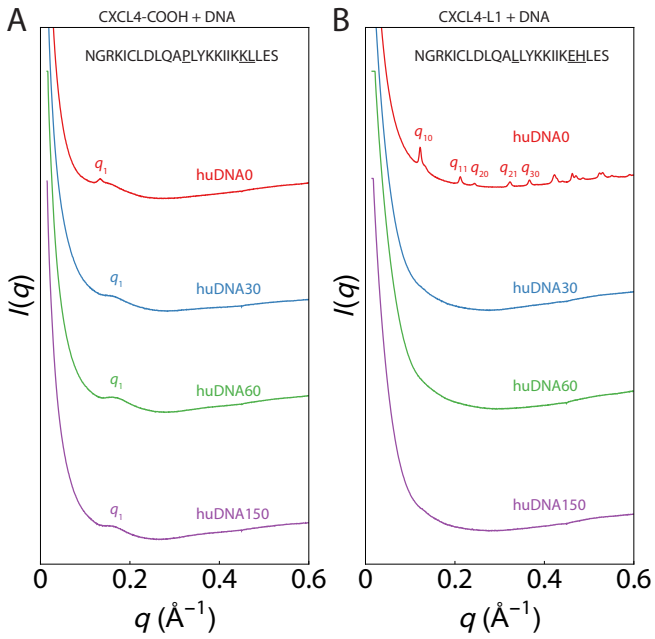


Figure 4.6: **The C-terminus of CXCL4 can organize fragmented human DNA into more ordered lattices with optimal TLR9-activating inter-DNA spacings than unfragmented human DNA.**
(A) CXCL4 C-terminus (CXCL4-COOH) contains 24 amino acids from the C-terminus of CXCL4. CXCL4-COOH condenses unfragmented human DNA (huDNA0) into a columnar lattice with an inter-DNA spacing of $d = 4.62$ nm, which is predicted to induce low activation of TLR9. However, sonication of human DNA (huDNA30, 60, 150) shifts the inter-DNA spacing to 3.7-3.8 nm which is in the hyperstimulatory range for TLR9. Fragmented human DNA can in principle organize into ordered lattices with less conformational distortion than unfragmented human DNA. First peak positions q_1 are indicated for all complexes. (B) Mutating 3 amino acids in the CXCL4 C-terminus reduces the charge from +4 to +2 and leads to the formation of an ordered hexagonal lattice with a significantly increased inter-DNA spacing of $d = 5.94$ nm (huDNA0). Although the lattice is well-ordered, this value of the inter-DNA spacing is far outside the range of spacings that strongly activate TLR9. Interestingly, this mutation also ablates the ability of the C-terminus to bind to fragmented human DNA. The first peak position and four higher order reflections of the ordered hexagonal lattice of unfragmented DNA are labeled as q_{hk}.

observation that CXCL4-L1 is poorly stimulatory for pDCs. Thus, the cationic C-terminus of CXCL4 is able to bind DNA into highly stimulatory complexes provided that the DNA is highly fragmented, and mutation of cationic residues in the C-terminus abolishes this activity.

LL37 has previously been demonstrated to be a potent autoantigen that induces high levels of TLR9-mediated IFN-α production from pDCs [330]. To compare the ability of LL37-DNA and CXCL4-DNA complexes to induce TLR9 activation in pDCs, we correlated IFN-α production in pDCs with their SAXS structures. LL37 forms square columnar lattices with inter-DNA spacings of d = 3.56-3.57 nm with unfragmented human and E. coli DNA (Figure 4.7A). However, LL37-E. coli DNA also exhibits a larger nanocrystal size than LL37-huDNA0 (L = 22.4 nm vs. 18.2 nm), an observation that is consistent with higher levels of IFN production. With fragmented DNA, LL37 forms square columnar lattices with slightly smaller inter-DNA spacings of d = 3.34-3.39 nm. LL37 also binds to fragmented DNA more efficiently than non-fragmented DNA, as indicated by an increase in the domain size of the nanocrystals (L = 22-24 nm vs. 18.2 nm). Overall, the observation that LL37-DNA complexes exhibit a broader capacity to induce IFN-α production from pDCs compared to that of CXCL4 together with the measured structural parameters of these complexes suggest a conceptual framework in which the crystallinity of these complexes directly influence their ability to activate TLR9: These data suggest that the structure of CXCL4-DNA complexes is critical for hyperactivation of pDCs via TLR9, and that such effects are cognate to other previously characterized immunogenic DNA nanoplexes like LL37-DNA [330], hBD3-DNA [349], and *Salmonella* Curli-DNA [441].

4.7.2.3 Methods

Lyophilized full-length human CXCL4, CXCL4-COOH (NGRKICLDLQAPLYKKI-IKKLLES), CXCL4-L1 (NGRKICLDLQALLYKKIIKEHLES), and LL37 (LLGDF-FRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES) were custom synthesized ($\leq$ 95% purity by HPLC). All peptides were dissolved in nuclease-free water (Ambion) to 10 mg/mL. For

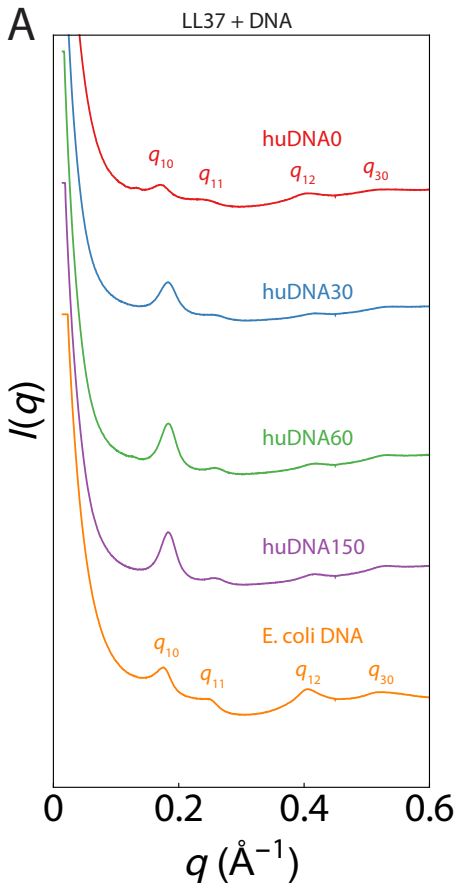


Figure 4.7: **LL37 binds both human and bacterial DNA into square lattices that amplify TLR9 activation.**
(A) First peak positions q_{10} and higher order reflections q_{hk} for LL37 bound to unfragmented human DNA (huDNA0) and E. coli DNA are indicated. Inter-DNA spacings for LL37-huDNA0 and LL37-E. coli DNA complexes are $d = 3.56$ nm and 3.57 nm, respectively. LL37 binds to fragmented DNA more efficiently than non-fragmented DNA, indicated by an increase in the domain size of the nanocrystals (22-24 nm vs. 18.2 nm). LL37-E. coli DNA also exhibits a larger nanocrystal size than LL37-huDNA0, in concordance with higher levels of IFN production (22.4 nm vs. 18.2 nm).

SAXS experiments, E. coli DNA (Affymetrix) was ethanol precipitated and resuspended in physiological buffer (140 mM NaCl + 10 mM HEPES, pH 7.4) to 5 mg/mL. Human DNA was purchased (Biochain) and used directly (huDNA0) or fragmented via sonication for 30, 60, and 150 minutes (huDNA30, huDNA60, huDNA150). Self-assembled protein-DNA complexes were formed by incubating the cationic proteins with human or E. coli DNA at isoelectric peptide-to-DNA charge ratios (P/DNA = 1/1) in microcentrifuge tubes. Complexes were vortexed at low speeds (900 RPM) for 1 hr or until strong precipitates formed. After thorough mixing and centrifugation, precipitated complexes are transferred to 1.5 mm quartz capillaries (Hilgenberg GmbH, Mark-tubes) and hermetically sealed using an oxygen torch. The structures of CXCL4-DNA and related complexes were solved using SAXS. Experiments were performed at the Stanford Synchrotron Radiation Lightsource (SSRL, Beamline 4-2) using monochromatic X-rays with an energy of 9 keV. A Rayonix MX225-HE detector (pixel size 73.2 μm) was used to measure the scattered radiation. Independent identical samples were prepared and measured over multiple separate experiments ($n \geq 3$). 2D powder diffraction patterns were azimuthally integrated using the Nika 1.76 [184] package for Igor Pro 7.04 and FIT2D [185]. SAXS data were analyzed by plotting scattering intensity $I(q)$ against the momentum transfer q in Mathematica.

Structures of the nanocrystalline protein-DNA complexes were solved by measuring the q-positions of all peaks and comparing them with the permitted reflections for phases with varying symmetries. Peak positions were accurately measured by fitting diffraction peaks to a squared-Lorentzian. SAXS patterns with a single major peak at position q_1 corresponds to a columnar DNA lattice with short-range order. The lattice parameter d indicates the inter-DNA spacing between parallel DNA columns and is calculated from the first peak position by the formula $q_1 = \frac{2\pi}{d}$. For square columnar DNA lattices, we observe peaks at $q_{hk} = \frac{2\pi}{d}\sqrt{h^2+k^2}$ where (h,k) are the Miller indices and d is the lattice parameter corresponding to the inter-DNA spacing. Typical square lattices will have reflections at q_{10} and q_{11} with a ratio of $1:\sqrt{2}$. For hexagonal columnar DNA lattices, we observe peaks at $q_{hk} = \frac{2\pi}{d}\sqrt{\frac{4}{3}(h^2+k^2+hk)}$. Typical hexagonal lattices will have reflections at q_{10} , q_{11}, and q_{20} with ratios of $1:\sqrt{3}:2$. The lattice parameters were calculated by linear regression

through points corresponding to measured and theoretical peaks. Procedures to assign these liquid-crystalline phases are similar to those found here [349, 378].

To calculate the domain size of the nanocrystals, the first peak was fitted to a squared-Lorentzian using least-squares regression in Mathematica (q_1 is the location of the first peak and h is the peak width).

$$S(q) = \frac{h^3}{4\pi \left(|q - q_1|^2 + \left(\frac{h}{2}\right)^2 \right)^2}$$

We extract the peak width h and calculate the average domain size L using Warren's approximation [443]. The domain size is related to h via the following equation [430].

$$L(h) = \frac{(8\pi)^{1/2}}{\frac{h}{2}}$$

4.8 Summary

In this Chapter, we described several examples of how the nanocrystalline ordering of DNA by amyloids, chemokines, and antimicrobial peptides can drastically lead to immune amplification in diseases like lupus, psoriasis, and systemic sclerosis. In Chapter 5, we extend this paradigm to TLR3, another important innate immune receptor relevant to autoimmune diseases and the normal immune response to dsRNA. In Chapter 6, we revisit the structural rules for TLR9 activation by higher order assemblies of chromatin components, which comprise the predominant form of cellular DNA. Finally, in Chapter 7, we combine ideas presented in Chapters 2-3 and Chapters 4-6 to begin to understand the competing molecular requirements for membrane activity and immunomodulatory activity of helical AMPs.

CHAPTER 5

Crystallinity of dsRNA-AMP immune complexes modulates TLR3-mediated inflammation

5.1 Abstract

Double-stranded RNA (dsRNA) induces production of pro-inflammatory cytokines in normal human epidermal keratinocytes (NHEK) by specific binding to endosomal Toll-like receptor-3 (TLR3). Recently, it has been shown that hyperactivation of TLR3 in psoriatic keratinocytes by dsRNA can occur in the presence of human antimicrobial peptide (AMP) LL37. Here, we combine synchrotron X-ray scattering, microscopy, computer simulations, and measurements of NHEK cytokine production to elucidate a previously unanticipated form of specific molecular pattern recognition. LL37 and similar α-helical AMPs can form proinflammatory nanocrystalline complexes with dsRNA that are recognized by TLR3 differently than dsRNA alone. dsRNA complexes that activate IL-6 production in NHEK and those that do not are both able to enter cells and colocalize with TLR3. However, the crystallinity of these AMP-dsRNA complexes, specifically the geometric spacing between parallel dsRNA and the repeat number of ordered dsRNA, strongly influence the level of TLR3 activation. Crystalline complexes that present dsRNA at a spacing that matches with the steric size of TLR3 can recruit and engage multiple TLR3 receptors, driving receptor clustering and immune amplification, whereas crystalline complexes that exhibit poor steric matching do not. Reverse-transcription quantitative PCR (RTqPCR) of IL-6 during siRNA knockdown of TLR3 confirms that cytokine production is due to TLR3: High levels of IL-6 transcription are observed for sterically-matched complexes without TLR3 knockdown, whereas such activity is abrogated with TLR3 knockdown.

5.2 Introduction

TLR3 recognition of dsRNA is a fundamental example of molecular recognition [271]. Binding of viral dsRNA to TLR3 in endosomes triggers downstream NF-κB activation and production of pro-inflammatory cytokines [351, 269]. Recently, non-coding host-derived dsRNA has also been shown to break immune tolerance and activate TLR3, but these dsRNAs require presence of the antimicrobial peptide (AMP) LL37 [444]. In psoriasis, self-dsRNA can complex with LL37, access endosomes, potently activate TLR3 in keratinocytes [354, 355, 445, 358] and amplify production of the pro-inflammatory cytokines IL-6 and IFN-β [446], thereby exacerbating disease [447]. A variety of RNA-binding proteins [360, 361], AMPs [366], cell-penetrating peptides [448], cationic lipids [449], and synthetic polymers [450] have also been reported to bind to dsRNA. However, there is no general criterion to determine which of these self-assembled dsRNA complexes induce immune responses and which do not. A common hypothesis explaining the ability of complexed nucleic acids to activate endosomal immune receptors is that formation of aggregates confers increased efficiency of endosomal access and protection from nuclease degradation [110]. Although this model can explain why self-antigens can activate TLRs, it does not explain why some complexes are pro-inflammatory while others are non-inflammatory. Recent results on activation of TLR9 by dsDNA complexes suggest that cooperativity of ligand binding may be important [349], but it is not clear whether these results are generalizable. At present, the mechanism for TLR3 activation by polycation-dsRNA complexes is not well understood.

It is not surprising that anionic dsRNA and cationic molecules can form electrostatic complexes, which have been studied extensively [451, 452, 453, 431, 413, 416, 454]. Perhaps a little more surprising is the observation that polycation-dsRNA complexes that activate IL-6 production and those that do not are both able to enter cells efficiently and colocalize with endosomal TLR3 in normal human epidermal keratinocytes (NHEK), which shows that the ability to access endosomal compartments, a necessary condition of activation, is not the only determinant for strong immune activation. What is unanticipated is the discovery against expectation that TLR3 immune activation depends strongly on the recognition of

crystallinity in dsRNA-AMP complexes, and metrics of crystalline order can drive immune outcomes. The two major criteria defining TLR3 activation and amplification are 1) the inter-dsRNA spacing between parallel dsRNA ligands within the dsRNA complex, and 2) the average size (the average number of unit cells) of ordered crystalline domains within dsRNA complexes. These criteria, the lattice constant and the domain size, correspond to two fundamental measures of crystallinity. We rationalize this behavior using a coarse-grained computational model and show that TLR3 binding to dsRNA is optimized when the inter-dsRNA spacings are in an optimum range: Crystalline complexes that present dsRNA at a spacing well-matched with the TLR3 steric size can recruit and engage multiple TLR3 receptors, effectively driving receptor clustering and immune amplification, whereas complexes that exhibit poor steric matching do not. Our observations are consistent with a conceptual framework in which activation levels are modulated by different inter-RNA spacings and by different crystalline RNA domain sizes in AMP-dsRNA complexes. Importantly, reverse-transcription quantitative PCR (RTqPCR) of IL-6 during siRNA knockdown of TLR3 confirms that observed cytokine production is due to TLR3: High levels of IL-6 transcription are observed for sterically-matched complexes without knockdown, whereas abrogation of such activity is observed for these complexes with TLR3 knockdown. Recent work on TLR9 has shown that complexes between dsDNA and polycations at optimal inter-DNA spacings (such as LL37) can lead to a drastic increase of interferon-α secretion by plasmacytoid dendritic cells, whereas those at mismatched spacings do not [349]. Taken together with the evidence adduced here, these cognate observations suggest that AMP-induced ligand clustering and crystalline ordering can lead to an unexpected channel of pattern recognition and amplification for TLR responses, and a potentially general mode of AMP immunomodulatory function.

5.3 Results and Discussion

5.3.1 IL-6 production in keratinocytes through TLR3 activation depends on the inter-dsRNA spacing within self-assembled polycation-dsRNA electrostatic complexes

The structural basis of dsRNA binding by TLR3 has been demonstrated via X-ray crystallography [298] (Figure 5.1A). dsRNA can bind electrostatically via the anionic phosphate backbone to cationic portions of TLR3 dimers. Since complexes of dsRNA and various cationic molecules can lead to high levels of immune activation [357], we investigate their structures. The structures of dsRNA complexes with immunologically relevant polycations, many of which are amphiphilic and impart an additional principle of self-assembly, are largely unknown. To assess the relationship between keratinocyte activation and these peptide-dsRNA complexes, we characterized 15 polycation-dsRNA structures with SAXS and correlated these structures with their ability to activate NHEK cells. Consistent with the behavior of other electrostatic complexes composed of nucleic acids [417], the observed structures do not depend strongly on the global concentration of these complexes. The structures of these complexes, however, can in principle change with the composition of the complex. To mitigate this uncertainty, we solved each individual structure at three or more different polycation-to-dsRNA molar ratios on both sides of the isoelectric point, the stoichiometry at which the negative charge of the dsRNA is compensated by the positive charge of the polycation (typically obtained from the polycation to dsRNA charge ratio, P/RNA = 1/4, 1/2, 1/1, 2/1, 4/1). This library of over 50 solved structures, which covers a large range of physiologically relevant scenarios, is organized into a "phase diagram" for each type of complex. We find that the self-assembled structures of these dsRNA complexes are remarkably conserved for a broad range of stoichiometry with minor exceptions. Typical structures are dsRNA columnar phases with different in-plane symmetries (square lattices, hexagonal lattices, disordered lattices), and with different spacings between dsRNA (center-to-center distances between dsRNA). From each structure, we measure the inter-dsRNA spacing a, the linear domain size L (and the repeat number m) (Table 5.1).

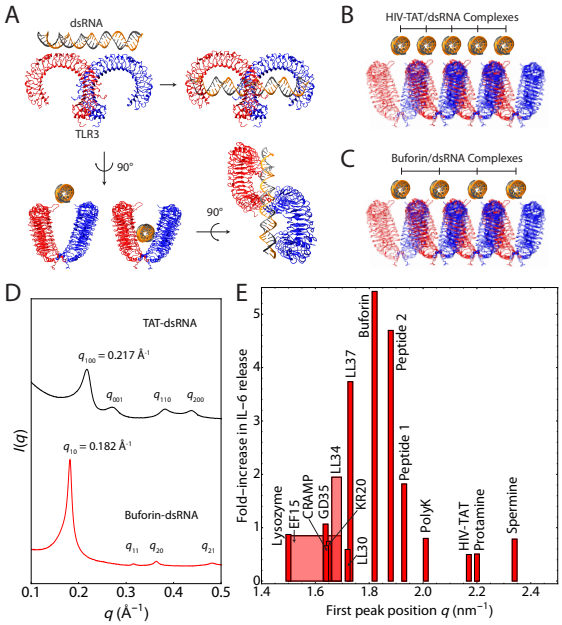


Figure 5.1: **IL-6 production in keratinocytes through TLR3 activation depends on the inter-dsRNA spacing within self-assembled polycation-dsRNA electrostatic complexes.**
A) Illustration of dsRNA binding to TLR3 from two orthogonal views based on the known crystal structure of mouse TLR3 (PDB ID: 3CIY [298]). (B) Smaller geometric spacings between ordered dsRNA columns (such as in HIV-TAT/dsRNA complexes) reduce TLR3 activation because they sterically hinder accessibility of dsRNA to the receptors. (C) In contrast, larger geometric spacings between ordered dsRNA columns (such as in buforin-dsRNA complex) optimize electrostatic interactions with TLR3, leading to binding of multiple TLR3s simultaneously and amplification of the immune response. (D) Representative SAXS data from peptide-dsRNA complexes that induce strong IL-6 production (buforin-dsRNA, red), and complexes that do not (TAT-dsRNA, black). The TAT-dsRNA complex forms a 3D hexagonal lattice with an approximate inter-dsRNA spacing of $a \sim 2.90$ nm. The buforin-dsRNA complex forms a 2D hexagonal lattice with an inter-dsRNA spacing of $a \sim 3.45$ nm. (E) Normalized fold-increase in IL-6 production by NHEK induced by polycation-dsRNA complexes. The immune response correlates strongly with the SAXS measurements of their first diffraction peak positions, which are inversely proportional to the inter-dsRNA spacing.

To assess the level of immune activation induced by each complex, IL-6 protein production is measured as a quantitative readout for NHEK activation using ELISA (Table 5.1). IL-6 is one of several primary cytokines implicated in the pathogenesis of psoriasis, and frequently serves as a reporter cytokine for TLR-mediated activation of inflammation through NF-κB. While dendritic cells and T cells are important cell types in the pathogenesis of psoriasis, keratinocytes have been clearly implicated as key participants in the immune response to nucleic acids, including acting as a main source of type 1 interferon production [353]. The data shows that some dsRNA complexes strongly amplify TLR3 activation while others do not. To investigate the origin of this difference, we compare the measured level of IL-6 production with the inter-dsRNA spacing and unit cell repeat number of each complex for each set of stoichiometric conditions in our phase diagram. We find that there is a strong relationship between the structure of polycation-dsRNA immune complexes and their ability to amplify IL-6 production. Specifically, immune complexes with inter-dsRNA spacings much smaller than TLR3 do not activate strongly (Figure 5.1B), while complexes with inter-dsRNA spacings roughly commensurate with the size of TLR3 maximize activation (Figure 5.1C). We have chosen to illustrate these differences by highlighting four prototypical examples of polycation-dsRNA complexes with varying structures. HIV-TAT is a well-characterized cell-penetrating peptide that has been explored as a nucleic acid transfection reagent, and is known to have strong endosomal access [234, 455]. Buforin is an antimicrobial peptide obtained from a histone fragment that is known to penetrate cell membranes and bind to nucleic acids [105, 267]. Spermine is an endogenous polyamine known to be biologically important in the folding and binding of nucleic acids to their targets [456, 457, 458], and is a structural relative of many cationic polymers used for gene delivery applications. LL37 is a multifunctional endogenous innate immunity peptide that is implicated in the pathogenesis of autoimmune disease such as lupus and psoriasis [332, 110, 459]. It can also form pores in membranes, enable endosomal access, and has been reported to signal to the immune system via TLR9 [338].

From SAXS measurements, the TAT-dsRNA complex forms a close-packed 3D hexagonal columnar lattice with a first peak position at $q_{100} = 0.217$ Å^{-1} and higher order reflections at

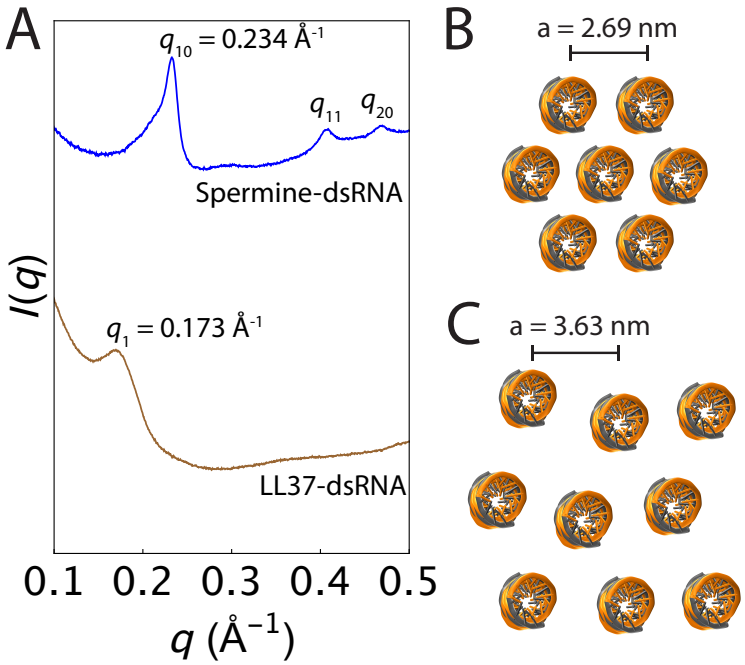


Figure 5.2: **Differences in structure of polycation-dsRNA complexes predict their abilities to amplify cytokine production.**
Ordered dsRNA columnar complexes have diverse structures with varying inter-dsRNA spacings. The presentation of "open" dsRNA complexes with spacings commensurate with the steric size of TLR3 (a = 3.3-3.7 nm) highly amplify IL-6 production in NHEK, while "closed" dsRNA complexes with spacings smaller than the steric size of TLR3 ($d < 3.2$ nm) result in reduced TLR3 activation and IL-6 production. (A) Representative SAXS data of spermine and LL37-dsRNA complexes. LL37-dsRNA complexes induce $\sim 5\times$ the IL-6 production from NHEK compared to spermine-dsRNA complexes. Spermine-dsRNA complexes (blue) form a 2D hexagonal columnar phase with a first peak position at $q_{10} = 0.234$ Å$^{-}1$, corresponding to an inter-dsRNA spacing of $a = 2.69$ nm (B). All higher order reflections of the spermine-dsRNA complex are labeled with their Miller indices. In comparison, the LL37-dsRNA complexes (brown) form a disordered columnar phase with a first peak position at $q_1 = 0.173$ Å^{-1} and an average inter-dsRNA spacing of $a = 3.63$ nm (C). This difference in inter-dsRNA spacing drastically changes the ability of the complex to activate TLR3, as observed in Figure 5.1E.

$q_{001} = 0.271$, $q_{110} = 0.384$, and $q_{200} = 0.441$ Å^{-1}, corresponding to an inter-dsRNA spacing of $a \sim 2.90$ nm, which is much smaller than the steric size of TLR3 (Figure 5.1B, D). In contrast, the structure of the buforin-dsRNA complex is an "open" 2D hexagonal columnar lattice with a first peak position at $q_{10} = 0.182$ Å^{-1} and higher order reflections at $q_{11} = 0.314$, $q_{20} = 0.362$, and $q_{21} = 0.482$ Å^{-1}, corresponding to a larger inter-dsRNA spacing of $a \sim 3.45$ nm that can accommodate the size of TLR3 (Figure 5.1C, D). dsRNA complexes based on TAT induce low levels of IL-6 even though TAT is known to exhibit pervasive endosomal access in a broad range of cells. In comparison, dsRNA complexes based on buforin exhibits high levels of IL-6 production, with the latter inducing an order of magnitude ($\sim 10\times$) more IL-6 production than TAT-dsRNA complexes (Figure 5.1E). SAXS measurements show that spermine-dsRNA complexes form tightly packed hexagonal columnar arrays, corresponding to an inter-dsRNA spacing of $a \sim 2.69$ nm (Figure 5.2A, B). In contrast, LL37-dsRNA complexes exhibited a single diffraction peak at $q_1 = 0.173$ Å^{-1} corresponding to an average inter-dsRNA spacing of 3.63 nm, consistent with an "open" disordered columnar complex (Figure 5.2A, C). Like buforin, LL37-dsRNA complexes induce high levels of IL-6 production from NHEK, consistent with prior experimental observations [357], but spermine-dsRNA complexes, like TAT-dsRNA complexes, do not. Quantitatively, complexes based on LL37 induce $\sim 5\times$ more IL-6 from NHEK than those based on spermine. The disparity between IL-6 expression levels induced by LL37- and buforin-dsRNA with those induced by TAT- and spermine-dsRNA complexes illustrate two distinct families of structures that differentially predict immune activation (Figure 5.1, 5.2). In fact, our immune activation data for the 15 different dsRNA complexes suggest that the measured inter-dsRNA spacings in these complexes correlate strongly with TLR3 activation (Figure 5.1E). Generally, we observed that complexes with first peak positions $\sim$ 1.7-1.9 nm^{-1} strongly activated TLR3, while those outside this range had reduced abilities to activate TLR3. We note that although most complexes follow the trend line, a small subset of complexes appears to deviate slightly. This is not surprising given the complex nature of TLR activation: For example, it is possible that small quantitative differences in the degree of endosomal access between different types of dsRNA-peptide complexes can influence the level of the IL-6 response. In particular, from

our previous SAXS experiments with model membranes, helical AMPs like LL37, CRAMP, and their mutants exhibit different levels of membrane curvature generating activity due to differences in charge, hydrophobicity, and amphipathicity [177], which can alter their intake pathways and thereby TLR access. Differential degradation rates of these peptides in biological environments may also play a role in the observed variation. In fact, our characterization of the structures of a cohort of LL37 fragments bound to dsRNA mimics the distribution of LL37 degradation products likely present *in vivo* [460, 461].

The majority of the characterized polycation-dsRNA complexes did not exhibit significant changes in the first-peak positions across the wide range of tested peptide-to-dsRNA stoichiometric ratios (and by implication charge ratios), indicating that most of these structures are robust to changes in relative stoichiometry. Moreover, the LL37-dsRNA complex was found to be stable in human serum, which indicates that the self-assembly process is robust even in the presence of other proteins. Indeed, this is not unexpected on physics grounds: Electrostatic self-assembly at physiological conditions favors binding between components with equal and opposite surface charge densities due to the resultant maximal entropy gain of counterion release [462], which is the case for dsRNA and LL37. However, a small subset of the tested peptide-dsRNA complexes (EF15 and LL34) exhibited a small shift in their first peak position from a lower to higher q as the peptide-to-dsRNA charge ratio increased from below the isoelectric point (P/RNA = 1/4, 1/2) to isoelectric (P/RNA = 1/1) and beyond (P/RNA = 2/1, 4/1). This behavior indicates that the structure of the complex is changing with stoichiometry. However, it is interesting to note that these data points also follow the basic trend line (Figure 5.1E)

5.3.2 Potential relationship between TLR3 activation and the inter-RNA spacing within dsRNA complexes

The potential structural criterion outlined above is an interesting finding, given that the prevailing paradigm of immune activation is that the spatial organization of dsRNA ligands is not expected to influence the magnitude of activation. To test the statistical significance

of these trends, we construct a null hypothesis based on the assumption that there is no relationship between TLR3 activation (as measured by IL-6 secretion) and the inter-dsRNA spacing. The inter-dsRNA spacings (inversely proportional to first-peak position q) and the magnitude of TLR3 activation are assumed to be two random variables. That the present problem involves nonlinear correlations dictates that it is necessary to go beyond the usual linear statistical methods. To this end, we used the recently developed metric distance correlation (R_{dCor}) [256, 463] to quantify the association between our paired variables without assuming any functional form of the correlation (e.g. fitting the data parametrically to a function). By definition, R_{dCor} takes on values between 0 and 1. $R_{dCor} = 0$ if and only if the variables are statistically independent, while $R_{dCor} = 1$ if strongly dependent. Using a Monte-Carlo simulation ($N = 100,000$ sampling iterations), we randomly sampled the experimental dataset of measured IL-6 production for polycation-dsRNA complexes with the null hypothesis that there is no correlation between the structure of the complexes and IL-6 activation. We find that the distance correlation between the inter-dsRNA spacing and TLR3 activation is $R_{dCor} = 0.371$ [0.327, 0.492] with $P = 0.022$ at a 95% significance level of $\alpha = 0.05$ (Figure 5.3A), indicating a likely relationship between TLR3 activation and inter-RNA spacing. To test this finding further, we also calculated the recently developed maximal information coefficient (MIC), which non-parametrically measures the strength of the non-linear association between paired variables [257]. Like dCor, MIC takes on values between 0 and 1. MIC values close to 0 indicate statistical independence while values near 1 indicate a strong functional relationship. Using a similar Monte-Carlo simulation ($N = 100,000$), we conducted a null-hypothesis significance test using MIC [264]. The MIC correlation between inter-dsRNA spacing and TLR3 activation was MIC = 0.652 [0.414, 0.894] with $P < 10^P - 5$ at a 95% significance level of $\alpha = 0.05$ (Figure 5.3B). Taken together, these two results suggest that the null hypothesis is highly unlikely, and that there is clearly a statistically significant nonlinear relationship between TLR3 activation and the inter-RNA spacings of self-assembled RNA complexes.

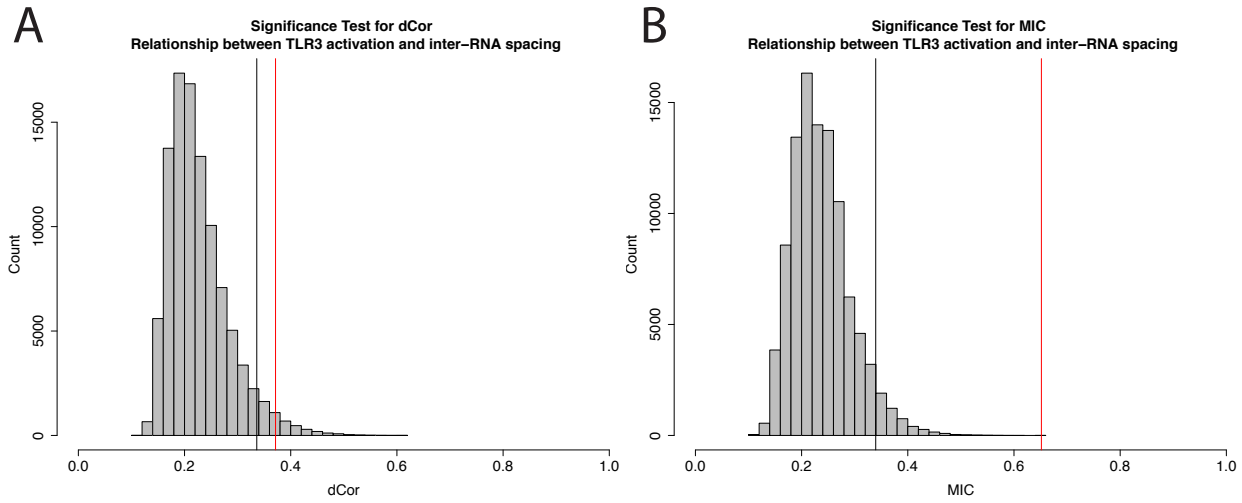


Figure 5.3: **Null-hypothesis significance test for statistical dependence between TLR3 activation and inter-dsRNA spacing.**
(A) We calculated the distance correlation (d_{Cor}) between fold-increase in IL-6 production and inter-dsRNA spacing ($q = 2\pi/d$) for the 15 complexes (IL-6 measured in triplicate). Using 100,000 Monte-Carlo iterations with $N = 45$, we plot the null distribution and show our experimental findings relative to the significance threshold. We find that the observed $d_{Cor} = 0.371$ [0.327, 0.492] (vertical red line) exceeds the 95% significance level of $\alpha = 0.05$ (vertical black line) with $P = 0.022$. Analysis was done in R using the 'energy' package. (B) Using the same dataset, we did a null-hypothesis simulation test (100,000 Monte-Carlo iterations) for the maximal information coefficient (MIC), which also measures the strength of nonlinear correlations. We find that the MIC = 0.652 [0.414, 0.894] (vertical red line) also exceeds the 95% significance level of $\alpha = 0.05$ (vertical black line) with with $P < 10^{-5}$. This indicates that it is unlikely that TLR3 activation and inter-RNA spacings are statistically independent, and that there is a clear relationship between the structure of crystalline RNA complexes and immune activation. Analysis was done in R using the 'minerva' package.

5.3.3 Cytokine-inducing polycation-dsRNA complexes activate NHEK via TLR3

From the data in Figure 5.1E, we observed that IL-6 amplifying and non-IL-6-amplifying molecules activate TLR3 to varying degrees. To assay whether polycation-dsRNA complexes indeed co-localize with and activate specifically through TLR3, we carried out fluorescence microscopy experiments and RTqPCR experiments with NHEK. We tested one IL-6 amplifying complex (LL37) and two non-amplifying complexes (TAT and spermine). From microscopy studies of prototypical examples of dsRNA-polycation complexes, we find that dsRNA complexes that amplify NHEK production of IL-6 and those that do not both enter cells and colocalize with TLR3 (Figure 5.4A). Interestingly, LL37-dsRNA, TAT-dsRNA, and spermine-dsRNA all enter NHEK cells and apparently bind to TLR3. However, only LL37-dsRNA complexes seem to amplify cytokine production, while the other complexes do not. These results suggest that internalization and access to immune cells, while clearly important to be sure, do not by themselves explain differences in levels of immune activation. Entry and binding to TLR3 is necessary but not sufficient for immune amplification. Rather, the inter-dsRNA spacing appears to play an important role also. To confirm that the cytokine production is indeed specifically due to TLR3, we conducted RTqPCR of IL-6 during siRNA knockdown of TLR3 in the presence and absence of the complexes. We observe high levels of IL-6 transcription for IL-6-amplifying LL37-dsRNA, Peptide 2-dsRNA, and buforin-dsRNA complexes without knockdown, and the abrogation of such activity with TLR3 knockdown, suggesting that activation of TLR3 by these complexes directly leads to cytokine production (Figure 5.4B) ($p^* < 0.05$ for dsRNA only, $p^{***} < 0.001$ for buforin-dsRNA, $p^{**} < 0.01$ for Peptide 2-dsRNA, $p^* < 0.05$ for LL37-dsRNA). To confirm that the siTLR3 indeed knocked down TLR3 expression, we also measured TLR3 expression via RTqPCR (Figure 5.4C) ($p < 0.001$). Although it is generally recognized that siRNA knockdown is not always completely efficient, we observe significant knockdown of IL-6 transcription. We note that it is possible that some residual IL-6 production may be present due to endosomal escape and activation of the cytoplasmic MAVS signaling pathway by dsRNA complexes. In summary, IL-6 induction by polycation-dsRNA complexes are mediated at least in part by TLR3, and

maximally-inducing complexes exhibit similar degrees of endosomal access as non-amplifying complexes.

5.3.4 Coarse-grained model of TLR3 binding to RNA bundles exhibits dependence on inter-dsRNA spacing and repeat number in agreement with experimental observations

To investigate TLR3 interactions with ordered dsRNA immune complexes, we developed a coarse-grained theoretical model, which aims to predict the fraction of ligand-bound, active TLR3 receptors $f_A(q, m)$ as a function of the crystallinity parameters $q = 2\pi/a$ (where a is the inter-dsRNA spacing) and m (repeat number of RNA in the complex) (Figure 5.5A). Experimentally, the repeat number m is calculated from the linear domain size L (see Methods). The dsRNA complex/TLR3 system is cognate to the dsDNA complex/TLR9 system, which we have previously engaged [349, 350]. Based on crystallography data, TLR3 [298] has a larger opening angle compared to TLR9 [303]. Therefore, we predict that the optimal lattice spacing resulting in maximal TLR3 activation will be larger for RNA-peptide complexes compared to DNA-peptide complexes. That dsRNA complexes exhibit different domain sizes affords us the opportunity to explicitly take into consideration the effect of varying the repeat number of polycation-dsRNA complexes on TLR3 activation, in addition to the inter-dsRNA spacing. We also take into account the variable strength of the TLR3-dsRNA backbone interactions and explicitly include the density of TLR3 in the endosome and average dsRNA length. To do this, we use a Langmuir-like approach and assume that there are k possible binding sites for TLR3 receptors on each dsRNA cluster (i.e. k is the maximum number of TLR3 receptors that can bind to each peptide-dsRNA cluster), with all binding sites identical and independent. The number k is determined by the repeat number m and the individual dsRNA length. Here, the TLR3 receptor molecules interact only through the hard-core excluded volume interaction forbidding their overlap. The total number of bound TLR3 is given by a Langmuir isotherm

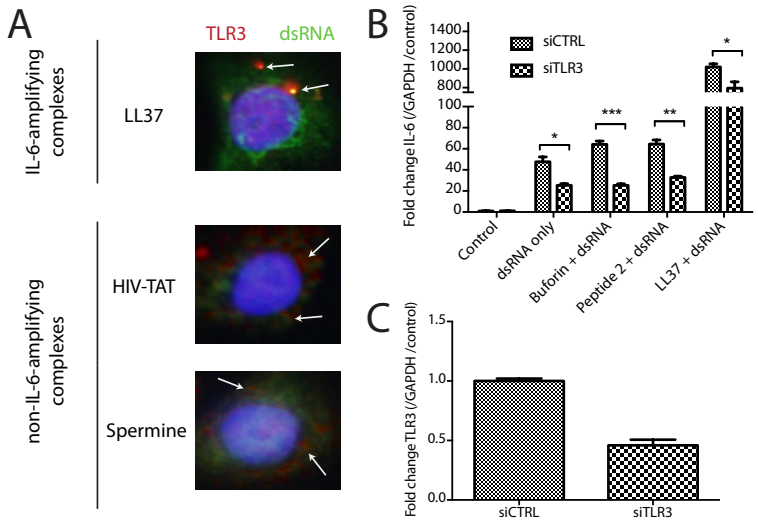


Figure 5.4: **Cytokine-inducing polycation-dsRNA complexes activate NHEK via TLR3.**
NHEK were stimulated for 4 hrs with biotinylated poly(I:C) complexed with either IL-6-amplifying LL37 or non-IL-6-amplifying HIV-TAT and spermine. Cells were imaged using an Olympus BX41 microscope. (A) We observe strong co-localization between TLR3 (red) and polycation-dsRNA complexes (green) within NHEK (white arrows indicate overlap). Both IL-6-amplifying and non-IL-6-amplifying polycation-dsRNA complexes co-localize with TLR3, indicating that endosomal access is a necessary but not sufficient criterion for high levels of cytokine production. (B) RTqPCR of IL-6 transcripts during siRNA knockdown of TLR3 confirms that TLR3 activation by strongly-activating polycation-dsRNA complexes are predominantly responsible for IL-6 production. The buforin-poly(I:C), Peptide 2-Poly(I:C), and LL37-poly(I:C) complexes all demonstrate statistically significant reduction in IL-6 production in the presence of siTLR3 compared to control siRNA ($p^* < 0.05$ for dsRNA only, $p^{***} < 0.001$ for buforin-dsRNA, $p^{**} < 0.01$ for Peptide 2-dsRNA, $p^* < 0.05$ for LL37-dsRNA). Data is shown relative to expression of the GAPDH housekeeping gene. (C) Data showing successful knockdown of TLR3 in NHEK using siTLR3 ($p < 0.001$). Pairwise two-tailed t-tests were used to analyze siRNA knockdowns in both (B) and (C) (Error bars show mean $\pm$ SEM, $n = 3$).

$$n_B = k\frac{(1-f_A)e^{-\frac{\epsilon(a)}{k_BT}}}{1+(1-f_A)e^{-\frac{\epsilon(a)}{k_BT}}}$$

The single TLR3 binding free energy $\epsilon(a) = \Delta G + \mu^{TLR} + 2U(a)$ includes 3 terms. The backbone electrostatic interaction $U(a)$ varies with the lattice spacing a due to the presence of adjacent anionic RNA rods in the nanocrystalline cluster [349]. $U(a)$ is characterized by two parameters: B, the magnitude of the screened electrostatic interactions, and σ, the steric length scale associated to the size of TLR3 receptors. The functional form of U and the realistic values of the parameters are given in the Methods. The first two terms in the expression for the binding free energy, $\epsilon^* = \Delta G + \mu^{TLR}$ do not depend on crystallinity and are treated as a constant parameter. Here ΔG is the "pure" TLR3-dsRNA Gibbs interaction free energy and μ^{TLR} the activation-independent part of the chemical potential of TLR3 in the endosomal membrane. The part of the chemical potential that depends on the fraction f_A of activated (bound) receptors is included explicitly in the theory and results in the prefactors $(1 - f_A)$ in the above expression n_B for the number of bound TLR3. Note that in our model we explicitly consider receptor depletion: more bound TLR3 leads to fewer available TLR3, increasing the cost of recruitment. The fraction of all active receptors is $f_A = \frac{n_B N_c}{N^{tot}}$, with N_c the average number of clusters on the membrane and N^{tot} the total number of TLR3 receptors. Assuming equilibrium between clusters in solution and those adsorbed to a membrane, we can determine f_A using a self-consistent relationship from multivalent adsorption theory [349, 438, 464].

In Figure 5.5B, we present a contour plot of the fraction of active TLR3 receptors f_A (and predicted immune activation profile) as a function of the first peak position q and the repeat number m (a contour plot showing the fraction of active TLR3 receptors as function of the interaction strength ϵ and repeat number m is found in Figure 5.6). We assume that f_A is a good proxy for TLR3 activation and IL-6 production, with a larger fraction of active receptors correlating with increased immune activation and cytokine secretion. Using the approximate model here, a strong dependence on the inter-dsRNA spacing is observed.

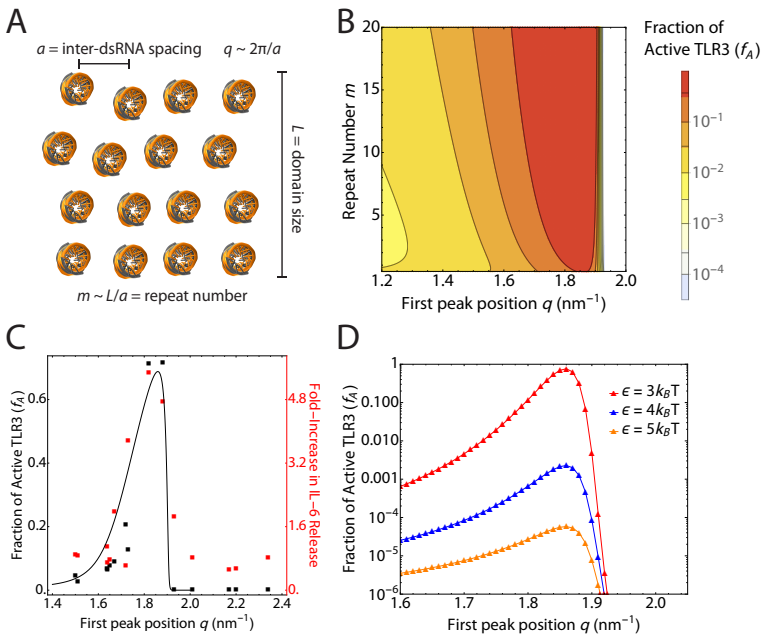


Figure 5.5: **Coarse-grained model of TLR3 binding to RNA bundles exhibits dependence on inter-dsRNA spacing and repeat number in agreement with experimental observations.**
(A) Schematic depicting the crystallinity parameters of an example peptide-dsRNA complex. (B) Contour plot of the fraction of active TLR3 receptors f_A as a function of the repeat number m (e.g. number of dsRNA rods per cluster) and the first peak position $q = 2\pi/a$ ($C^* = l\frac{c_{RNA}}{\rho_{TLR}} = 10^{-2}$, $\epsilon^* = 2k_BT$, $B = 30k_BT$ nm, and $\sigma = 4.4$ nm). High levels of TLR3 activation are denoted in orange, and low levels of binding and TLR3 activation are denoted in yellow to blue. Maximal activation occurs at inter-dsRNA spacings $q \sim$ 1.7-1.9 nm^{-1} ($a \sim$ 3.3-3.7 nm). Inter-dsRNA spacings that are much smaller than the steric size of TLR3 do not lead to increased activation ($q >$ 2.0 nm^{-1}) due to steric repulsion between adjacent receptors and dsRNA rods. (C) We find good quantitative agreement between the theoretical prediction of active TLR3 (left vertical axis, black) and measured IL-6 production from NHEK (right vertical axis, red) for experimentally measured q and m (Figure 5.1E). The black line denotes the theoretical curve calculated for $m = 6$. (D) Grand canonical Monte Carlo simulations of TLR3 binding to crystalline lattices of dsRNA recapitulate features observed in the theoretical model and experimental data. f_A is plotted against q at $m = 4$ for various interaction energies.

Moreover, a dependence on the repeat number m is also clearly observed. At larger m, dsRNA complexes with a broader range of inter-dsRNA spacings can maximally activate TLR3, while at smaller m, a narrower range of spacing is allowed (Figure 5.7A). At an inter-dsRNA spacing that is highly commensurate (well-matched) with the steric size of TLR3, even relatively small ligand clusters with a low number of repeats m can significantly activate due to strong multivalency and resultant "superselective" effects [464, 438]. Interestingly, at spacings slightly larger than the steric size, a large m can compensate for reduced effective binding energy between sterically-mismatched dsRNA ligands and TLR3 (Figure 5.7B), while at spacings smaller than the steric size this is not possible. These observations suggest that the elastic response of dsRNA-AMP nanocrystals to deformations can influence binding to TLR3 and thereby immune activation. For spacings much larger or smaller than the TLR3 size, we observe a sharp drop in number of bound TLR3. This trend is reflected in the experimental data: polycation-dsRNA clusters with small inter-dsRNA spacings ($q > 2.0$ nm^{-1}) like HIV-TAT and spermine cannot potentiate TLR3 activation despite having a large domain size ($L \sim 17$ nm for both TAT and spermine). In contrast, buforin-dsRNA and LL37-dsRNA complexes have quite different domain sizes and number of repeats, but since both have ideal inter-dsRNA spacing, they both activate NHEK. The buforin complex has a larger repeat number m than the LL37 complex (Table 5.1), explaining the higher relative levels of IL-6 induction by buforin-dsRNA. Indeed, in Figure 5.5C, we show that the fraction of active TLR3 receptors f_A can be predicted from both the experimentally measured inter-dsRNA spacing and the repeat number m, and that there is good correspondence between the model (black) and experimentally measured IL-6 secretion (red, data reproduced from Figure 5.1E).

To understand how the fraction of activated TLR3 depends on the crystallinity of TLR3-AMP complexes, we conducted coarse-grained Grand canonical Monte-Carlo simulations on TLR3-dsRNA binding (details in Methods). The fraction of bound and activated TLR3 receptors f_A for multiple dsRNA-AMP complexes with varying inter-dsRNA spacings a is calculated. In Figure 5.5D, simulation results for f_A vs. $q = 2\pi/a$ are shown for cluster size $m = 4$. For an interaction energy of $\epsilon^* = 3k_BT$, the fraction of active TLR3 is maximized

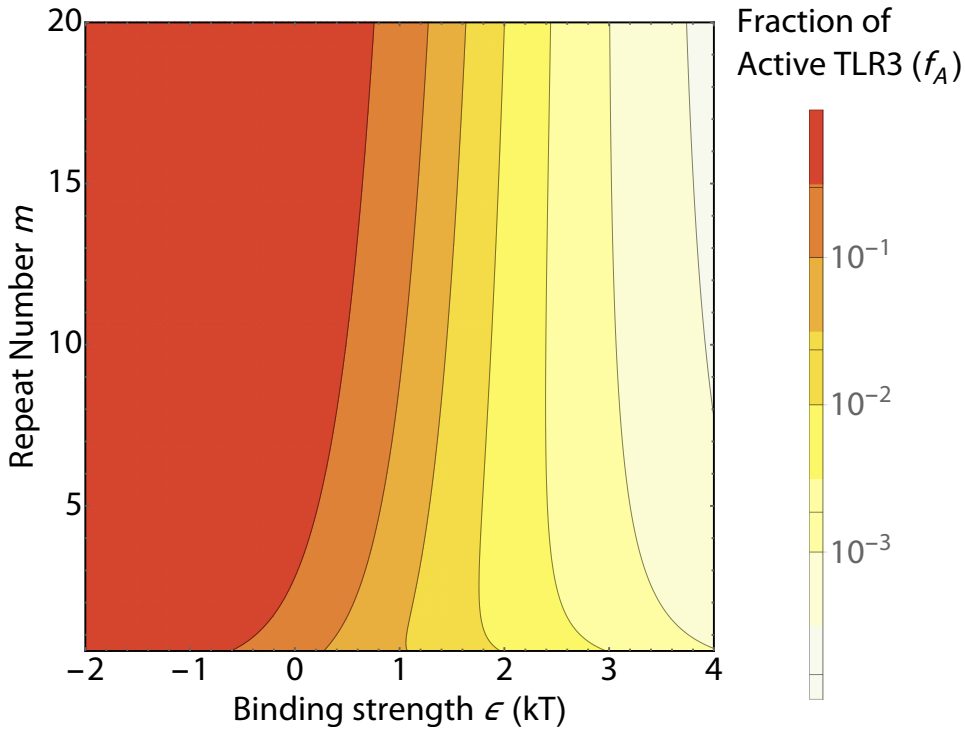


Figure 5.6: **Effects of repeat number m and binding strength ϵ on TLR3 binding to peptide-dsRNA nanocrystalline clusters.**
Contour plot showing the fraction of active TLR3 receptors f_A as a function of the repeat number m and the binding strength ϵ (units of k_BT) for a single TLR3 binding to the nanocrystalline cluster. The prefactor was chosen as $C^* = 10^{-3}$. Changing the prefactor results in only minor quantitative differences. Figure 5.5B shows the binding efficiency as a function of the repeat number m and the lattice spacing indicated by the first peak position q. The binding strength depends on the inter-dsRNA spacing as specified by $\epsilon = \epsilon^* + 2U(a)$. With the value $\epsilon^* = 2k_BT$ chosen here, the optimal spacings correspond to negative values of ϵ (net attraction) while positive values of ϵ and net repulsion are indicating suboptimal spacings.

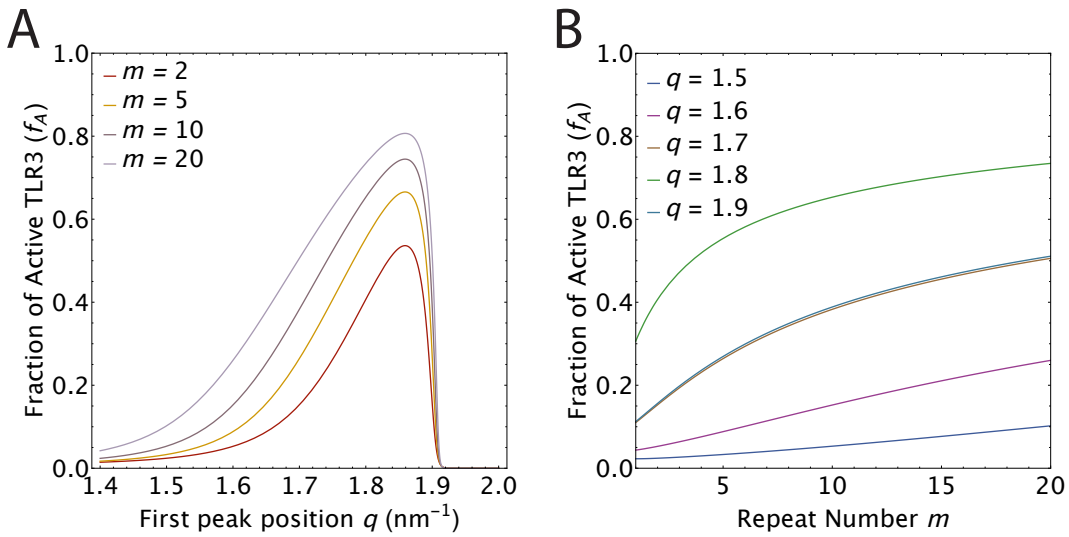


Figure 5.7: **Dependence of TLR3 activation on first peak position q and repeat number m.**
(A) Slices taken from Figure 5.5B at constant values of repeat number $m = 2$, 5, 10, and 20. The broadening of the peaks as m increases indicates that the range of inter-dsRNA spacings that can accommodate strong TLR3 activation increases as the size of the nanocrystalline cluster increases. For smaller clusters, the optimal range of spacings is more confined. For clusters near the optimal value of q, increase in m leads to a larger fraction of bound TLR3. (B) Slices taken from Figure 5.5B at constant values of inter-dsRNA spacings corresponding to $q = 1.5$, 1.6, 1.7, 1.8, and 1.9 nm^{-1}. Note that $q = 1.8$ nm^{-1} has a higher ceiling of bound TLR3 for a given repeat number compared to $q = 1.7$ or 1.9 nm^{-1}. The dependence is much stronger for clusters close to the optimal spacing ($q = 1.8$ nm^{-1}), but weaker for those further away from the optimal spacing ($q = 1.5$ nm^{-1}).

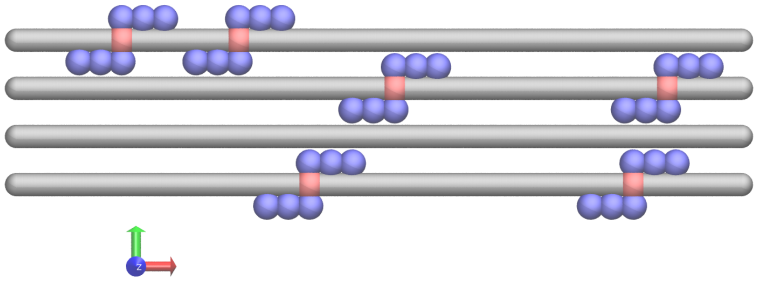


Figure 5.8: **Monte-Carlo simulation snapshot of TLR3 dimers binding to a dsRNA-AMP nanocrystalline cluster.**
Parallel dsRNA within the crystal are shown as four grey rods. dsRNA binding TLR3 dimers are modeled as two blue paddles (each represented with 3 balls) connected by a red linker. The length of dsRNA rods is 50 nm and the size of each TLR3 paddle is 4.4 nm. For this snapshot, the spacing between dsRNA rods is commensurate with the steric size of TLR3 (such as with LL37-dsRNA), leading to a drastic amplification of bound receptors via a variation of the superselectivity mechanism.

for complexes with 1st order diffraction peaks near $q \sim 1.85$ nm^{-1}, which agrees well with experiment. A simulation snapshot of the dsRNA-TLR3 system near this corresponding optimal a spacing (such as for LL37-dsRNA) is shown in Figure 5.8. For interaction energies of $\epsilon^* = 4k_BT$ and $5k_BT$, the dependence of f_A on q is weaker, but activation is still maximized at a comparable inter-dsRNA spacing. Note that the optimal spacing observed here is slightly larger than that of the cognate TLR9 system, due to geometric differences in size between TLR3 and TLR9. Comparison to experiment (Figure 5.1E, 5.5C) indicates that an interaction energy of $\epsilon^* = 3k_BT$ is appropriate. The results also show that TLR3 binding and activation are more permissive for complexes with spacings that are sterically larger than the receptor than complexes with spacings that are smaller, in qualitative agreement with both the theoretical model and experiments.

The theoretical and simulation models address two limiting regimes of behavior: The former calculates TLR3 activation in a system where the total activation is limited by the number of available TLR3, whereas the latter assumes that the number of available TLR3

is large and the limiting constraint is the concentration of RNA-peptide complexes. The fact that we observe consistent behavior in these two limiting cases of TLR3 saturation and excess gives us confidence that we can address the basic aspects of the observed phenomenon with these simple models. Taken together, the dependence on the inter-dsRNA spacing, on the domain size and repeat number of dsRNA ordering, the residual ability of complexes with sterically larger dsRNA spacings to activate (especially with larger numbers of unit cells), and the inability of complexes with sterically smaller dsRNA spacings to activate, strongly suggest that the crystallinity of the dsRNA complex can structurally influence TLR3 activation significantly (Figure 5.9).

5.4 Conclusions and Prospects

In summary, we demonstrate that the presentation of ordered lattices of dsRNA leads to amplification of TLR3 mediated IL-6 cytokine production in NHEKs. That it was recently shown that the presentation of ordered crystalline lattices of dsDNA leads to amplification of TLR9 mediated interferon-α secretion in plasmacytoid dendritic cells suggests that these may not be unrelated phenomena [349, 441, 362] (It will be interesting to examine whether any other proinflammatory cytokines or cell types also exhibit similar amplification mechanisms.) These results suggest a general picture in which immune signaling can be strongly influenced by AMP-immune ligand interactions. The resultant "ligand clustering" can in principle drive receptor clustering. The data here suggest that cationic proteins, peptides, or other cationic agents less than $\sim$ 0.5 nm in size or much greater than $\sim$ 3 nm in size do not significantly activate TLR3, since steric constraints will prevent the crystalline dsRNA rods from effectively recruiting TLR3. It is evident that immune modulation is complex and multifactorial. Quantitative differences in the degree of endosomal access between different types of dsRNA-peptide complexes can influence the level of TLR3 response in a manner independent of the nanocrystalline multivalent binding model. What is also equally clear is that not only can AMPs organize dsRNA into nanocrystalline structures recognizable by TLR3, the quantitative degree of immune activation is strongly influenced by crystalline

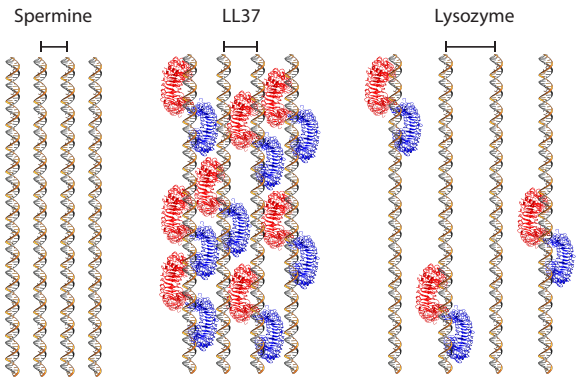


Figure 5.9: **Superselective binding of TLR3 to polycation-dsRNA complexes occurs only at inter-dsRNA spacings commensurate with the steric size of TLR3.** This schematic demonstrates the drastic differences in TLR3 binding and activation as the inter-dsRNA spacing is varied from "close-packed" to "open". This is a top-down view of mobile TLR3 receptors (red and blue) in the endosomal membrane binding to clustered dsRNA ligands (gray). Spermine-dsRNA complexes with tightly packed dsRNA rods does not allow for steric access of TLR3 paddles between rods. LL37-dsRNA complexes have optimal inter-dsRNA spacings well-matched to the width of the TLR3 paddles, which drives TLR3 receptor recruitment and clustering and drastically amplifies immune activation. "Wide-packed" lysozyme-dsRNA complexes can accommodate TLR3 binding and activation, but loss of multivalency between adjacent dsRNA rods due to reduced backbone interactions leads to a lower density of TLR3 binding.

parameters such as the inter-ligand spacing and the domain size of the ordering.

5.5 Methods

5.5.1 Preparation of self-assembled polycation-dsRNA complexes

Lyophilized cationic peptides LL37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES), buforin (TRSSRAGLQFPVGRVHRLLRK), CRAMP (GLLRKGGEKIGEKLKKIGQKIKNFFQKLVPQPEQ), and HIV-TAT (47-57, YGRKKRRQRRR) were purchased from Anaspec or Genemed Synthesis Inc. ($\geq$ 95% purity by HPLC). The following cationic peptides were also custom synthesized (LifeTein, $\geq$ 95% purity by HPLC): GD35 (GDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES), LL34 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPR), LL30 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRN), KR20 (KRIVQRIKDFLRNLVPRTES), EF15 (EFKRIVQRIKDFLRN), Peptide 1, and Peptide 2.

Table 5.1: **Data from SAXS studies and cell experiments for dsRNA complexed with various polycationic molecules.**
We report crystallinity parameters of the structurally-characterized polycation-dsRNA complexes, including the first peak position q, the inter-RNA spacing ($a = 2\pi/q$), the domain size L, the number of repeats m, the measured structure, and the mean normalized fold-increase in IL-6 production from keratinocytes relative to baseline production.

Peptide	IL-6	q_1 (nm^{-1})	a (nm)	L (nm)	m	Phase
LL37	3.74	1.63	3.63	2.11	0.58	Disordered columnar
Buforin	5.43	1.82	3.45	39.36	11.40	Hexagonal
Peptide 1	1.82	1.93	3.26	13.67	4.20	Disordered columnar
Peptide 2	4.70	1.88	3.34	36.40	10.89	Hexagonal
GD35	1.07	1.64	3.84	4.92	1.28	Disordered columnar
LL34	1.95	1.67-1.70	3.69-3.76	4.59	1.22	Square
KR20	0.75	1.66	3.79	5.50	1.45	Square
EF15	0.85	1.51-1.70	3.69-4.16	5.93	1.42	Disordered columnar
LL30	0.59	1.72	3.65	8.30	2.27	Square
Lysozyme	0.87	1.50	4.19	31.17	7.44	Hexagonal
Spermine	0.79	2.34	2.69	17.80	6.63	Hexagonal
Protamine	0.51	2.20	2.86	13.46	4.71	Hexagonal
PolyK	0.80	2.01	3.13	15.98	5.11	Hexagonal
HIV-TAT	0.50	2.17	2.90	17.18	5;93	Hexagonal
CRAMP	0.66	1.64	3.83	4.37	1.14	Disordered columnar

Other polycationic agents used include human lysozyme, polylysine (MW $\sim$ 70,000), spermine, and protamine sulfate (Sigma Aldrich). All cationic agents were dissolved in nuclease-free water (Ambion) to 10 mg/mL. For SAXS and cell experiments, poly(I:C) (high molecular weight) and poly(A:U) (Invivogen) were directly resuspended in physiological buffer (100 mM NaCl + 10 mM HEPES, pH 7.4) to 5 mg/mL. Polycation-dsRNA complexes were formed by incubating the peptide with dsRNA (1-5 mg/mL for SAXS experiments, 10-20 μg/mL for cell experiments) at specific peptide-to-RNA charge ratios (P/RNA = 1/4, 1/2, 1/1, 2/1, 4/1). For a specific cationic reagent, the same charge ratios were used in both SAXS and cell experiments.

Table 5.2: **List of primers used for RT-qPCR.**
Gene symbols and forward and reverse primer sequences for each gene are shown in the table below.

Gene	Strand	Primer Sequence
GAPDH	Forward	TGGGCTACACTGAGCACCAG
GAPDH	Reverse	GGGTGTCGCTGTTGAAGTCA
IL-6	Forward	AATTCGGTACATCCTCGACGG
IL-6	Reverse	TTGGAAGGTTCAGGTTGTTTTCT

5.5.2 SAXS Experiments and Data Analysis

We mapped out the structural phase diagrams of peptide-dsRNA complexes by incubating peptides (10 mg/mL) with dsRNA (5 mg/mL) at specific charge ratios in microcentrifuge tubes as described above. After thorough mixing and centrifugation, precipitated complexes are hermetically sealed in 1.5 mm quartz capillaries (Hilgenberg GmbH, Mark-tubes). SAXS experiments were performed at Stanford Synchrotron Radiation Lightsource (SSRL, Beamline 4-2) using monochromatic X-rays with an energy of 9 keV. A Rayonix MX225-HE detector (pixel size 73.2 μm) was used to measure the scattered radiation. Independent identical samples were prepared and measured over multiple timepoints to ensure consistency. 2D powder diffraction patterns were integrated using the Nika 1.74 [184] package for Igor Pro 6.37 and FIT2D [185]. SAXS data were analyzed by plotting integrated scattering intensity against the momentum transfer q using Mathematica. Peak positions were measured by fitting diffraction peaks to a Lorentzian. Structures of complexes were solved by calculating ratios between the q-positions of all measured peaks and comparing them with the permitted reflections for known liquid-crystalline phases. The lattice parameter(s) of each phase were calculated by linear regression through points corresponding to measured and theoretical peaks. The lattice parameter a indicates the inter-dsRNA spacing between RNA columns. For each complex, the inter-dsRNA spacing a is estimated from the first peak position by the formula $a = 2\pi/q_1$.

To determine the phase and lattice parameters for each nanocrystalline complex, we measure the q peak positions and relate them to the Miller indices for powdered averaged

phases. $q_{hk} = \frac{2\pi}{a}\sqrt{h^2+k^2}$ for square columnar lattices, $q_{hk} = \frac{2\pi}{a}\sqrt{\frac{4}{3}\left(h^2+k^2+hk\right)}$ for hexagonal columnar lattices, and $q_1 = 2\pi/a$ for disordered columnar lattices. Square and hexagonal lattices were assigned based on fitting the measured q positions to the above equations. Typical square lattices will have reflections at q_{10} and q_{11} with a ratio of $1:\sqrt{2}$. In contrast, typical hexagonal lattices will have reflections at q_{10} , q_{11}, and q_{20} with a ratio of $1:\sqrt{3}:2$. Procedures to assign these liquid-crystalline phases are similar to those found here [413].

In addition to quantifying the lattice parameters and inter-dsRNA spacings in each complex, we also measured average domain size L of each complex. From that value, we extracted the number of repeat units m in each complex. To obtain the domain size, we approximated the structure factor peaks as squared-Lorentzians. Comparing Gaussian and squared-Lorentzian lineshapes, the Lorentzian lineshape yielded the best fit for all of the complexes. The form of the equation used for the fit was

$$S(q) = \frac{w^3}{4\pi\left(|q-q_0|^2 + \left(\frac{w}{2}\right)^2\right)^2}$$

where q_0 is the location of the peak and w is the peak width [349]. The experimental SAXS data was background subtracted, and the first peak for each complex was fitted using non-linear least square regression in Mathematica. The extracted value for peak width w can be related to the average linear domain size L using Warren's approximation [443]. For the squared-Lorentzian lineshape, the domain size is related to w via the equation [430].

$$L = \frac{(8\pi)^{1/2}}{\frac{w}{2}}$$

We assume that a typical bundle has an area proportional to L^2. To calculate the number of repeat units in each complex m, we estimated this by dividing the linear domain size L by the measured inter-dsRNA spacings a ($m = L/a$). The resulting values for L and m for each liquid-crystalline complex are reported in Table 5.1.

5.5.3 Reverse transcription-quantitative PCR (RTqPCR) analyses

Control siRNA and TLR3 siRNA were purchased from Dharmacon. During each keratinocyte stimulation, 5 nM of each siRNA was electroporated into NHEKs as described previously [465]. Total cellular RNA was extracted using the PureLink RNA isolation kit and RNA was reverse transcribed to cDNA using iScript cDNA synthesis kit (Bio-rad). Quantitative, real-time PCR was performed on the CFX96 real time system (Biorad) using TaqMan® Gene Expression Master Mix. The expression of GAPDH gene was used as a house keeping gene to normalize data in human cells. Specific primer sequences are shown in Table 5.2.

5.5.4 Keratinocyte Stimulation Experiments

Primary neonatal human epidermal keratinocytes (NHEKs) were purchased from Invitrogen. NHEKs were grown in serum free EpiLife medium supplemented with 0.06 mM $CaCl_2$, EpiLife Defined growth supplements (EDGS) (Invitrogen) and antibiotics, and passage 4-6 cells were used for experiment. Cells at 60%-80% confluence were starved overnight without EDGS prior to treatment. NHEKs were treated with polycationic agent/peptide (10 μg/mL) or vehicle control mixed with poly(I:C) (0.3 μg/mL) for 6 hrs (RTqPCR) and 18 hrs (ELISA). IL-6 protein volume using supernatant for NHEKs after indicated stimulation were measured by human IL-6 ELISA kit (R&D systems, Minneapolis, MN) per manufacturer's protocol. All stimulation experiments were done in triplicate. IL-6 release for each polycation-dsRNA complex was normalized relative to control stimulation and reported in Table 5.1.

5.5.5 NHEK Immunofluorostaining Experiments

NHEKs were grown on 8-well chamber slides (Thermo Fisher Scientific, Waltham, MA, USA). NHEKs were stimulated with biotinylated poly(I:C) (Invivogen, San Diego, CA, USA) with LL37 (Genemed Synthesis), HIV-TAT, or Spermine for 4 hr. Then NHEK were fixed in 4% PFA for 10 mins prior to immunofluorescence staining. Fixed and permeabilized cells were

incubated with anti-TLR3 antibody (ab137722, abcam, Cambridge, MA, USA) followed by streptavidin-conjugated AlexaFluor 488 (Thermo Fisher Scientific) and 576-coupled anti-rabbit secondary antibody (Thermo Fisher Scientific). Nuclei were counterstained with DAPI. All images were taken with an Olympus BX41 microscope.

5.5.6 Statistical Analyses

Bootstrapping calculations to compute confidence intervals and P-values were done using the distance correlation function in the 'energy' package in R. MIC bootstrapping calculations were done using the 'minerva' package in R. We tested the relationship between the inter-dsRNA spacing and cytokine production by conducting a null hypothesis significance test with a significance level of $\alpha = 0.05$. Analysis of RTqPCR siRNA knockdown data was done in Mathematica using pairwise two-tailed t-tests with a significance level of $\alpha = 0.05$.

5.5.7 Analytical Theory Derivation

The model builds on the TLR9-DNA-peptide interaction model [349] generalized to take into account the limited number of TLR3 receptors in the endosomal membrane. Furthermore, the effect of varying nanocrystalline cluster sizes between different peptide-RNA complexes is explicitly considered.

We use a Langmuir-like approach and assume that there are k possible binding sites for TLR receptors on each RNA cluster (i.e. k is the maximum number of TLR receptors that can bind to each peptide-RNA cluster), all binding sites are identical and independent. The number k will be determined by the cluster size (i.e. the repeat number m) and the individual RNA length. We assume that clusters have a cuboidal shape, all RNA rods in the cluster are aligned parallel and only one face of a presenting a RNA-grill to the membrane interacts with membrane bound TLR3. Therefore, the number of possible bound TLR is $k = k_0 m$, where m is the repeat number and k_0 denotes an effective number of TLR that can bind along a single RNA rod. The TLR3 receptor molecules do not interact with each other, except for the hard-core excluded volume interaction forbidding their overlaps. The partition

function counting all possible bound/unbound states of a RNA-peptide cluster lying on a TLR3 decorated membrane is

$$Q = (1 + (1 - f_A)e^{-\frac{\epsilon}{k_B T}})^k$$

where ϵ is the binding free energy including the chemical potential μ of the TLR3 on the endosomal membrane: $\epsilon = \Delta G + \mu$, with ΔG the Gibbs free energy of binding of a single TLR3 to a RNA-peptide binding site. The number of available (unbound) receptors on the membrane is $N^{tot}(1-f_A)$ where $f_A = \frac{N_A}{N^{tot}}$ is the fraction of bound (activated) receptors with N_A and N^{tot} the number of activated (bound) and total number of receptors, respectively. The binding free energy for the first bond is $\epsilon = \Delta G - k_B T \log\left(\frac{N^{tot}}{A\rho_0}\right)$ with ρ_0 the reference density and $k_B T$ the thermal energy. Subsequent bonds are weaker: $\epsilon^* = \epsilon - k_B T \log(1 - f_A)$ due to the depletion of available receptors. The average number of bound TLR per RNA-peptide cluster is given by the Langmuir adsorption isotherm

$$n_B = k\frac{(1 - f_A)e^{-\frac{\epsilon(a)}{k_B T}}}{1 + (1 - f_A)e^{-\frac{\epsilon(a)}{k_B T}}}$$

Furthermore, the mean number of bound clusters is determined in terms of the membrane coverage:

$$\Theta = \frac{zQ}{1 + zQ}$$

, or the surface density of clusters. z is the dimensionless activity, i.e. the probability that an unbound cluster is found on a membrane within area a_k that a single cluster occupies. The total number of bound clusters: $N = \Theta\frac{A}{a_k}$ is proportional to the membrane coverage, but depends also on the (unknown) size of the membrane A and the cluster excluded area a_k. Larger clusters occupy more space on the membrane, therefore $a_k = ka_1$, with a_1 a constant specifying the space occupied per single TLR binding site, where a_1 is approximately of the size of a single TLR3 receptor.

It is not possible at present to measure the full size distribution of all the endosomal compartments in keratinocytes used in our experiments; therefore, we cannot predict the

absolute number of bound TLR3. However, we can calculate the number of bound TLR3 up to proportionality constant. Finally, we can obtain the average number of all active receptors by multiplying the number of clusters with the average number of bound TLR per cluster:

$$N_A = n_B N = n_B \Theta \frac{A}{a_1 k}$$

Dividing by the total number of receptors and expanding we get

$$f_A = \frac{zA}{N^{tot} a_1} \frac{(1-f_A)e^{-\frac{\epsilon(a)}{k_B T}} \left(1 + (1-f_A)e^{-\frac{\epsilon(a)}{k_B T}}\right)^{k-1}}{1 + z\left(1 + (1-f_A)e^{-\frac{\epsilon(a)}{k_B T}}\right)^{k}}$$

which needs to be solved self consistently for the fraction of bound (activated) receptors . This theory is similar to the one we previously presented for the binding of the DNA-peptide complexes to TLR9 [349]. However, we now consider the number of TLR3 receptors to be limited, while previously, we have considered TLR9 to be in excess compared to DNA-peptide clusters ($f_A << 1$). We expect that the present theory more accurately models the realistic system where the number of TLR3 in endosomal compartments is limited. However, we stress that both approaches give qualitatively similar results, and that both approaches can explain the experimental observations.

Since we expect the number of TLR3 receptors to be the limiting factor in RNA-peptide cluster adsorption, rather than the total coverage of the membrane with clusters ($zQ << 1$), the previous equation can be well approximated with

$$f_A = \frac{zA}{N^{tot} a_1} (1-f_A)e^{-\frac{\epsilon(a)}{k_B T}} \left(1 + (1-f_A)e^{-\frac{\epsilon(a)}{k_B T}}\right)^{k-1}$$

The dimensionless activity z is the probability that an unbound cluster is found on a membrane within lattice size a_k, hence we estimate it is $z = l a_k c_{clust}$, where l is the length scale in our system—the interaction range (height of a single TLR ~ 1 nm) and c_{clust} is the concentration of clusters in solution. In experiments the RNA strand concentration c_{RNA} and

molar ratios to peptides was kept fixed. Therefore, the density of clusters will be inversely proportional to the number of RNA strands per cluster m^2: $c_{clust} = \frac{c_{RNA}}{m^2}$. We remember that the number of possible TLR3 attachment sites is also determined by the repeat number $k = k_0 m$. Using these relations, we can compute the fraction of activated TLR3 as a function of the repeat number m and interaction energy ϵ:

$$f_A = C\frac{k_0}{m}(1-f_A)e^{-\frac{\epsilon(a)}{k_BT}}\left(1+(1-f_A)e^{-\frac{\epsilon(a)}{k_BT}}\right)^{k_0 m-1}$$

The prefactor $C = l\frac{c_{RNA}}{\rho_{TLR}}$ depends on the ratio between the densities of RNA in solution and TLR3 on the membrane. Experiments were performed at $c_{RNA} = 5$ g/L with the average RNA molecular weight: $M_{RNA} = 408$ kDa. We do not know the absolute density of TLR3 receptors in endosomal compartments, however, a reasonable value would be about one TLR3 per 100 square nanometers: $\rho_{TLR} \approx \frac{1}{100\ \mathrm{nm}^2}$. The length scale (TLR3-RNA interaction range) is about a nanometer, $l \approx 1$ nm. Therefore, the prefactor can be estimated as: $C \approx 10^{-3}$. The effective number of TLR3 that can bind along a single RNA rod depends on the RNA length that is presented to the membrane, assumed to be one persistence length ($l_p^{RNA} \approx 50$ nm). The TLR3 horseshoe size is $\sigma = 4.4$ nm; The TLR3 dimer is composed of two horseshoes (or "paddles", see Figure 5.8) that assemble in the presence of RNA. Therefore, $k_0 \approx \frac{l_p^{RNA}}{2\sigma} \approx 5$.

Following Ref. [349], the TLR3-dsRNA interaction has two contributions: $\epsilon = \epsilon^* + 2U(a)$. The main part, ϵ^*, is the cognate binding of RNA to the TLR dimer. However, there is also a weak nonspecific contribution due to electrostatics between an RNA and an adjacent TLR3 paddle: $U(a)$, which depends on a, the RNA spacing within the cluster. The factor 2 reflects that TLR3 is a dimer of two "paddles" (see Figures 5.1 and 5.9). This contribution is only present when the RNA is organized into clusters:

$$U(a) = -B\frac{e^{-\kappa(a-\frac{\sigma}{3})}}{a-\frac{\sigma}{3}} + \left(\frac{\sigma}{12\left(a-\frac{2\sigma}{3}\right)}\right)^{12} k_BT$$

where the first term captures the electrostatic attraction and the second term the steric

repulsion between a TLR3 paddle and an adjacent RNA strands in a cluster. $\sigma_{TLR3} = 4.4$ nm is the TLR3 size which is slightly larger than the TLR9 size $\sigma_{TLR9} = 4.0$ nm [349]. The Debye screening length is $\kappa^{-1} = 0.8$ nm at the physiological salt concentration. The prefactor B defines the strength of electrostatic interaction, and a value $B = 30k_BT$ nm is adopted from TLR9 [349].

Combining expressions for ϵ, $U(a)$, and f_A, we obtain the fraction of activated receptors as a function of the cluster spacing expressed with the lattice vector $q = 2\pi/a$, and the unit cell repeat number m, $f_A = f_A(q, m)$.

5.5.8 Monte Carlo Simulation Model

Grand canonical Monte Carlo simulations were performed by simulating an array of RNA chains lying flat on a membrane as shown in Figure 5.8. We simulate a single persistence length of dsRNA $\sim$ 50 nm. On length scales larger than persistence length, the cluster is likely more orientationally disordered. The membrane acts as a reservoir for TLR3. TLR3 can bind to DNA with binding energy $\epsilon^*(a)$ which includes the chemical potential of TLR3 (μ^{TLR}). Receptor depletion is not considered in MC simulations and μ^{TLR} is treated as constant. We vary ϵ and a, and evaluate the average number of TLR3-RNA bonds per RNA bundle, n_B. We also calculate the probability that no TLR3 attach to the RNA cluster, p_0, with the subscript denoting zero bonds. This probability is proportional to the (cluster-membrane) dissociation rate, and hence, to the dissociation constant. The total number of bound clusters N is inversely proportional to the dissociation constant, therefore, neglecting TLR3 depletion and cluster steric repulsion: $N \propto \frac{1}{p_0}$. The total number of bound (activated) receptors is proportional to the ratio $N_A = n_B N \propto \frac{n_B}{p_0}$.

5.6 Acknowledgements

E.Y.L. acknowledges support from the T32 Systems and Integrative Biology Training Grant at UCLA (T32GM008185) and the T32 Medical Scientist Training Program at UCLA (T32GM008042). R.L.G. acknowledges support from NIH R01AR064781, R01AI116576, and

R01AI052453. G.C.L.W. acknowledges support from NIH 1R21AI122212. X-ray research was conducted at Stanford Synchrotron Radiation Lightsource, SLAC National Laboratory, supported by the US DOE Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

CHAPTER 6

Neutrophil extracellular traps and necrotic cell death: structural basis of chromatin-mediated inflammation

6.1 Abstract

The molecular pathogenesis of psoriasis is characterized in part by immune dysregulation driven by recognition of self-DNA by Toll-like receptor-9 (TLR9). In one mechanism of aberrant immune activation, the overexpressed endogenous cationic antimicrobial peptide (AMP) LL37 organizes self-DNA into immunogenic complexes that potently hyperactivate TLR9 in plasmacytoid dendritic cells (pDCs). Recent work has demonstrated that cationic molecules that organize DNA into lattices with inter-DNA spacings commensurate with the steric size of TLR9 can amplify inflammation, while complexes poorly matched to TLR9 cannot. However, at present, it is unclear how and why nucleosomal DNA activates TLR9 in psoriasis. Here, we combine synchrotron X-ray scattering, computer simulations, and mathematical modeling with measurements of pDC interferon production to illuminate how nucleosome core particles (NCPs) in neutrophil extracellular traps (NETs) and NCPs released during necrotic cell death amplify TLR9-mediated inflammation. Interestingly, in the presence of common NETs components, NCPs self-assemble into liquid-crystalline columnar nanostructures that exhibit altered DNA wrapping states. Computer simulations and X-ray measurements suggest that the superhelical pitch of DNA within the NCP column relaxes into a value that is well-matched with the steric size of TLR9, allowing for optimal interdigitation with multiple TLR9 simultaneously like threads on a cylindrical screw, leading to multivalent electrostatic interactions that amplify binding and TLR9 activation. Our results elucidate the structural mechanism for TLR9 activation by NCPs and may more generally

explain why nucleosomes released from different modes of cell death can induce immune activation through a shared pathway.

6.2 Introduction

Self-nucleic acids can be inadvertently exposed to the immune system during necrotic cell death (e.g. myocardial infarction), controlled neutrophil cell death (NETosis) [466], and cell death related to chronic inflammation and autoimmune disease [467]. In necrosis, pathological damage to cellular membranes lead to uncontrolled release of damage-associated molecular patterns (DAMPs) into the extracellular milieu. These DAMPs include calcium, histones, DNA, nucleosome core particles (NCPs) (histone-DNA complexes), and heat-shock proteins [467]. These are recognized by pattern recognition receptors (PRRs) in antigen presenting cells such as macrophages and dendritic cells, which lead to aberrant immune amplification. NETosis is a controlled form of neutrophil cell death that is usually triggered by pathogens [466, 468, 469]. Induced by chemotactic agents like IL-8 and lipopolysaccharide, neutrophils undergo apoptosis, forming neutrophil extracellular traps (NETs) by extruding their chromatin and pre-formed granule proteins to form a self-assembled extracellular complex [470, 471]. This web serves to trap pathogens, expose them to a high local concentration of antimicrobial proteins, and form a physical barrier to prevent the spread of infection [468] (Figure 6.1A). Interestingly, excessive NET formation is also strongly associated with autoimmune diseases, including psoriasis [335, 472], lupus [473], systemic sclerosis, and chronic vasculitis [474]. Common autoantibodies observed in these diseases are directed against self-antigens found within NETs, including histones, DNA, NCPs, myeloperoxidase, and other cationic neutrophil proteins. Common to inflammation associated with cell death and autoimmune disease is the activation of the same immune pathways by DNA. In fact, elevated circulating plasma cells of calcium [475], histones [476, 477, 478], and DNA [478, 479] are found not only in autoimmune diseases but acute inflammatory conditions like systemic inflammatory response syndrome (SIRS) and sepsis [480], and chronic inflammatory conditions like cancer [481] and atherosclerosis [482]. While it is known that TLR9 activation

in plasmacytoid dendritic cells (pDCs) by DNA is implicated in these situations, it is not known how the structure of DNA bound to histones impact presentation to immune cells and the resultant inflammation. In prior studies, we explained how and why naked DNA bound to cationic antimicrobial peptides (AMPs) such as LL37 and hbD3 can amplify the immune system [110, 338, 330, 349]. Here, we focus on studying the immunogenicity of DNA bound to cationic histones in the NCP, which is far more abundant than naked DNA. Chromatin itself consists of chains of NCPs linked together by histone H1 like "beads on a string". An individual NCP monomer contains ~ 2 superhelical turns of anionic dsDNA wrapped around an octomeric cationic histone core [410, 483, 484]. The NCP is shaped like a cylindrical hockey-puck roughly 6 nm in height and 11 nm in diameter, and possesses a net negative charge (-144) [485] (Figure 6.1B). Since nucleosomes are found in native chromatin, they are released from cells in apoptosis, necrosis, and NETosis during subsequent chromatin decondensation. The inflammatory response to the release of nucleosomes from cell death leads to calcium elevation [475, 486], neutrophil recruitment, and release of cationic neutrophil proteins like LL37 [471]. The electrostatic self-assembly of NCPs by small polycations such as calcium and magnesium has been well studied [487, 488, 489], but condensation of chromatin by other polycationic proteins, such as AMPs, has never been characterized. Using small angle X-ray scattering (SAXS), we first solve the structure of self-assembled complexes of NCPs under conditions that mimic an inflammatory environment. After solving the structures of these complexes, we test their immunogenicity by incubating them with pDCs isolated from humans and measuring the subsequent interferon-α (IFN-α) production. Finally, we confirm the mechanism by conducting coarse-grained molecular simulations on isolated and columnar complexes of NCPs. We find that NCPs, in the presence of multivalent cations, form liquid-crystalline columnar complexes with DNA arranged in a superhelical structure with defined pitch. Individual isotropic NCPs do not activate pDCs, but when incubated with multivalent cationic molecules like antimicrobial peptides, they become highly immunogenic. Assembly of NCPs into nanostructures enables endosomal access and allows them to be recognized by antigen presenting cells. Furthermore, we show that electrostatic interactions between the TLR9 and the DNA in the NCP columns are optimized when the columnar complexes have

an inter-DNA pitch spacing corresponding to the steric size of TLR9 receptors. We propose that columns of NCPs released from dying cells can self-assemble under inflammatory conditions, enabling them to engage multiple TLR9 like threads on a screw sliding perfectly into a slot with a matching helical pitch, leading to superselective binding to TLR9, IFN-α production, and immune amplification.

6.3 Results and Discussion

6.3.1 NCPs form columnar nanocrystalline aggregates in the presence of calcium and antimicrobial peptides and amplify TLR9-mediated inflammation

During NET formation, chromatin, multivalent ions and cationic antimicrobial proteins are released, forming an electrostatically self-assembled web. Chromatin decondensation during NET formation is necessarily heterogenous, with a subpopulation of free NCPs released via cleavage and linker histones, intermediate multinucleosome aggregates, and digested NCPs with free DNA and cleaved core histones. Here, we choose to study the case of chromatin decondensation into individual intact NCPs in the presence of various NET components (Figure 6.1A-B). Using small-angle X-ray scattering (SAXS), we studied the self-assembly of NCPs with cationic molecules *in vitro*. First, we solved the structure of isolated NCPs in complex with physiologically relevant divalent cations like calcium, which are locally elevated in inflammatory responses. In fact, calcium surges precede formation of NETs in neutrophils [490]. Divalent cations are also abundant intracellularly, and are highly relevant to chromatin wrapping within the nucleus [491]. During cellular damage, divalent cations can be released into the extracellular millieu. When NCPs are exposed to physiological concentrations of divalent cations (1-5 mM), they self-assemble into isotropic columns with a domain size of $\sim$ 15-20 nm, corresponding to several NCPs stacked together (Figure 6.1C). The first correlation peak appears at $q_h = 0.108$ Å^{-1} which corresponds to the height of the individual NCP monomer, $h = 6.0$ nm. The second peak is a broad peak corresponding to $q_{p'} = 0.180$ Å^{-1}, which is derived from the inter-DNA pitch between consecutive turns

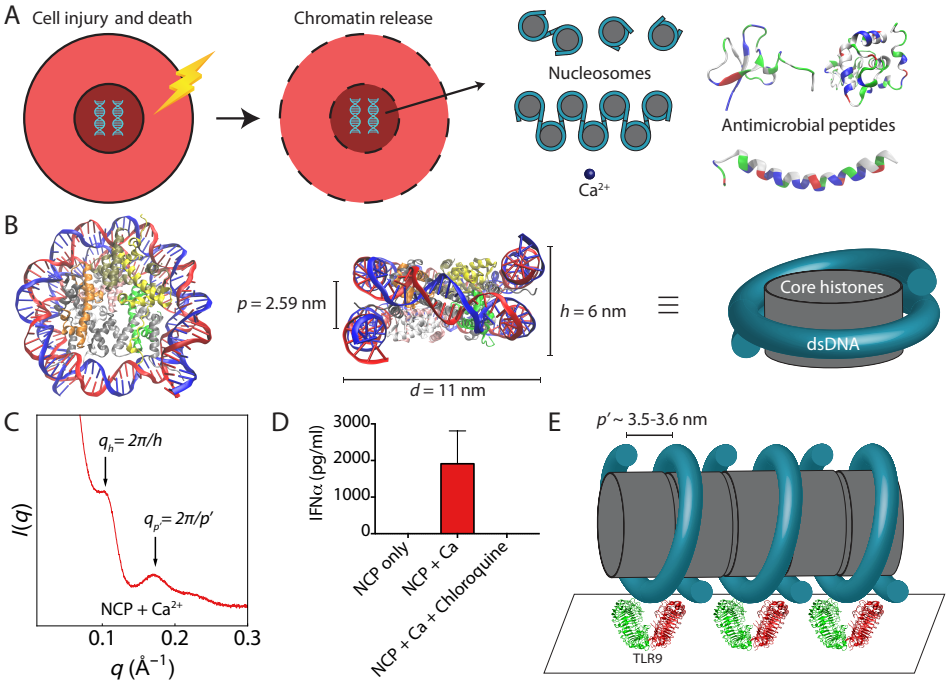


Figure 6.1: **NCPs assemble into columnar complexes in the presence of calcium and activate pDCs via TLR9.**
(A) Chromatin in the form of linked nucleosomes is released from dying cells in necrosis or neutrophils in NETosis. During this process, the chromatin is differentially processed into a heterogeneous mixture that contains monomeric NCPs. NCPs co-localize with divalent cations and antimicrobial peptides at the site of inflammation. (B) Top and side views of the crystal structure of the human nucleosome core particle (PDB ID: 2CV5). An NCP monomer has a hockey puck-like cylindrical shape with a diameter of 11 nm and height of 6 nm. From the crystal structure, the superhelical pitch of wrapped DNA around the histone core has a pitch of 2.59 nm. (C) SAXS data demonstrate that Ca^{2+} induces like-charge attraction between adjacent NCPs, leading to the formation of isotropic columnar complexes containing 2-3 NCPs in height. Correlation peaks indicating NCP height (q_h) and inter-DNA spacing between consecutive superhelical turns of DNA ($q_{p'}$) are labeled. (D) pDCs are stimulated with NCP alone or with NCP complexed with Ca^{2+}. NCP-Ca^{2+} columnar complexes amplify IFN-α secretion from pDCs in a TLR9-dependent manner. (E) Schematic of NCP-Ca^{2+} columns binding multivalently to TLR9.

of the DNA superhelix wrapped around the NCP column $p' = 3.49$ nm (Figure 6.1C). Next, we incubated NCPs alone and NCP-Ca complexes with pDCs in the presence and absence of chloroquine, an endosomal inhibitor of TLR signaling. In the presence of calcium, the NCP-Ca complexes induced strong TLR9-dependent IFN-α production, while NCPs alone did not. Knockdown with chloroquine abrogated signaling (Figure 6.1D). This led us to hypothesize that formation of nanocrystalline NCP complexes leads to multivalent presentation of wrapped DNA to clustered of TLR9 like threads in a cylindrical screw. The inter-DNA spacing between consecutive superhelical turns of DNA relaxes into a larger spacing of 3.5-3.6 nm, which is well-matched with the steric size of TLR9 [349] (Figure 6.1E). This drives receptor recruitment, multivalent binding of DNA to TLR9, and immune amplification.

To determine whether this phenomenon occurs in the presence of other cationic molecules found in NETs, we solved the structures of NCPs in complex with human β defensin-3 (hbD3) and lysozyme, and tested their ability to activate pDCs (Figure 6.2). SAXS data of isoelectric NCP-hbD3 and NCP-lysozyme complexes show that these cationic antimicrobial peptides form columnar nanostructures similar to NCP-Ca (Figure 6.2A, 6.2B). However, additional lateral inter-columnar ordered was observed, indicated by the appearance of a peak at around $q_d \sim 0.06$ Å^{-1}, corresponding to the rough diameter of an NCP monomer (and NCP column). Spacings between NCPs within each columnar structures are indicated by $q_h \sim 0.11$ Å^{-1}, simiilar to NCP-Ca. An altered inter-DNA superhelix pitch was also observed, ranging between 3.4-3.6 nm (corresponding to peak at $q_{p'}$. This relaxed inter-DNA spacing is also well-matched with the steric size TLR9, which predicts a strong ability to cross-link multiple TLR9 receptors. To test the ability of NCP-hbD3 and NCP-lysozyme to induce TLR9 activation, complexes were incubated with pDCs and IFN-α production was measured. Interestingly, NCP-hbD3 and NCP-lysozyme were both able to amplify TLR9-mediated inflammation similar to NCP-Ca, and activation was knocked down with chloroquine. The complexes also induced significantly more cytokine production than NCP alone, confirming that NCPs only induce inflammation when complexed with hBD3 and lysozyme (Figure 6.2C, 6.2D). The NCP columns formed in the presence hBD3 and lysozyme are larger than NCP-Ca complexes, with 4-5 NCPs in each stack, corresponding to an increase

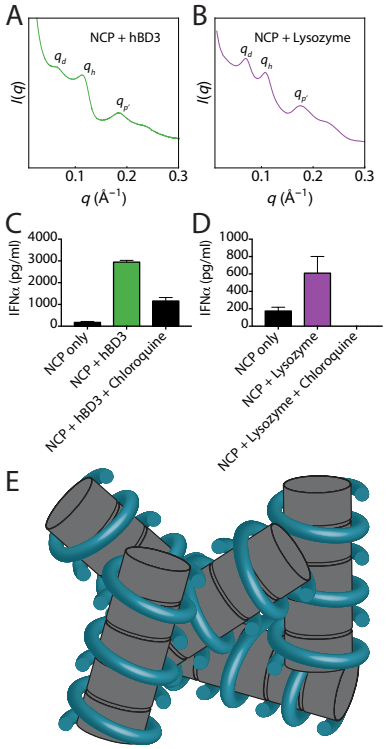


Figure 6.2: **Antimicrobial peptides hBD3 and lysozyme also condense NCPs into columnar structures that induce IFN-α via TLR9 binding.**
SAXS data of isoelectric (A) NCP-hBD3 complexes and (B) NCP-lysozyme complexes show that cationic antimicrobial peptides enriched in NETs can organize and remodel chromatin into columnar nanostructures. Labeled peaks correspond to the NCP height (h), inter-DNA spacing (p'), and inter-columnar lateral spacing (d). Some short-range lateral ordering is observed between parallel NCP columns, indicated by the emergence of a peak at q_d (this peak is absent in NCP-Ca^{2+} complexes, indicating that hBD3 and lysozyme induce stronger order). NCP-hBD3 and NCP-lysozyme complexes both also exhibit altered DNA wrapping states, indicated by a shift of the position of $q_{p'}$ to a q-value within the range that is predicted to be highly activating for TLR9 (3.4-3.6 nm). (C, D) Stimulation of pDCs with NCP-hBD3 and NCP-lysozyme complexes results in amplification of IFN-α production, consistent with prediction from structural studies. Treatment with chloroquine abolishes TLR9-dependent activation of pDCs by the NCP nanostructures. (E) Schematic of the columnar phase induced by self-assembly of NCPs with hBD3 and lysozyme (4-5 NCPs per stack).

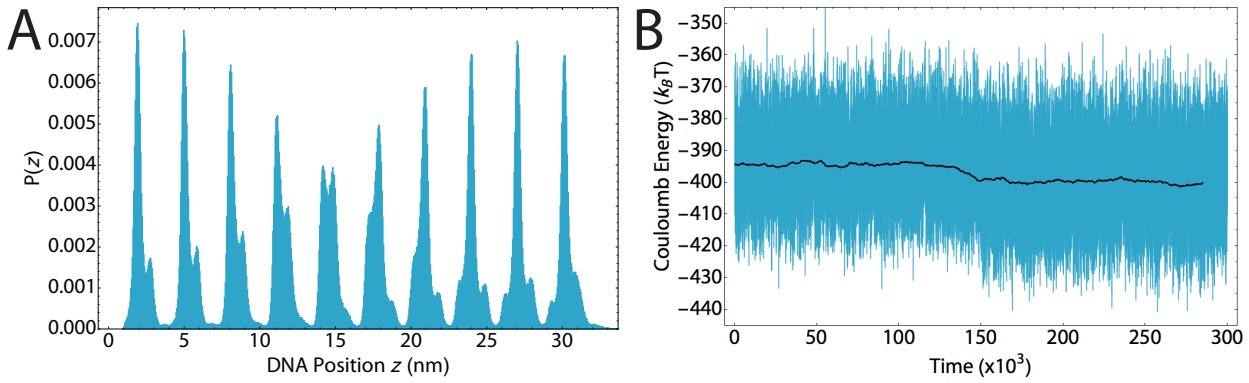


Figure 6.3: **NCP columns assembled in the presence of polycations exhibit altered inter-DNA spacings.**
(A) Computer simulations of NCP columns demonstrate that self-assembly driven by antimicrobial peptides changes the electrostatic landscape of histone-DNA interactions, leading to relaxation of the superhelical pitch of wrapped DNA into a spacing optimal for TLR9 activation. 10 turns of the DNA superhelix is simulated using a harmonic potential between adjacent turns. The probability of consecutive DNA turns along the Z-axis (long axis of the NCP column) is plotted versus Z-position. Although the positions of the DNA turns fluctuate, each peak corresponds to the most probable locations. The inter-DNA spacing predicted from simulation is ∼3.1-3.2 nm, which is close to previously observed values from cryo-EM [492], and similar to values measured from SAXS. (B) Plot of the free energy of the entire NCP complex shows that the average free energy of NCP self-assembly is $10k_BT$ lower when the superhelical pitch of wrapped DNA relaxes into a value optimal for TLR9 activation, than when the DNA spacing is fixed at the native value.

in domain size to $\sim$ 25 nm (Figure 6.2E).

6.3.2 Coarse-grained simulations demonstrate that columnar aggregates of NCPs exhibit altered DNA wrapping states well-matched to the steric size of TLR9

To investigate the physical origins of the altered inter-DNA spacing within the NCP column, we calculate an approximation of the X-ray scattering of the NCP column and conduct coarse-grained computer simulations to simulate the assembly of NCP columns in the presence of multivalent cationic molecules. Using a theoretical form-factor comparison of a helical screw wrapped with a groove DNA, we investigate the change in superhelical pitch of the DNA wrapped around the histone core of the NCP column. The form factor calculation was computed using the following derivation by Vaĭnshteĭn [493]. For a continuous helix of

length L_z, radius r_0, and pitch P, with helical width δ, the scattering amplitude is given by

$$A(q_r, q_z) = \sum_n \left[J_n(q_r\delta) \frac{J_1(q_r r_0)}{q_r r_0} \right]^2 \exp\left[-\frac{L_z^2 \left(q_z - \frac{2\pi n}{P}\right)^2}{4\pi} \right]$$

Taking the dominant Bessel functions $n = -1, 0, 1$ and taking the orientational average, the scattering intensity $I(q)$ as a function of q is

$$\langle I(q) \rangle \propto \sum_{n=-1}^{1} \int_{-1}^{1} \left[J_n\left(q\delta\sqrt{1-x^2}\right) \frac{J_1\left(qr_0\sqrt{1-x^2}\right)}{qr_0\sqrt{1-x^2}} \right]^2 \exp\left[-\frac{L_z^2 \left(qx - \frac{2\pi n}{P}\right)^2}{4\pi} \right] dx$$

We chose to fix our parameters based on the known crystal structure of an NCP (L_z = 100 nm, $r_0 = 4.2$ nm, $\delta = 1.0$ nm) and varied the pitch parameter P. The native monomeric NCP has a pitch of $p = 2.59$ nm, and the resulting form factor is shown in black. As the pitch increases to $p' = 3.49$ nm, the location of the first Bessel function peak shifts from $q_p = 0.24$ Å^{-1} to $q_{p'} = 0.18$ Å^{-1}. Using this approximation of a helical screw, it is suggestive that the superhelical pitch of DNA on the NCP shifts from 2.59 nm in the monomer to 3.4-3.6 nm in the column during the process of assembling into columns (Figure 6.1, 6.2). This likely results from an optimization of charge-matching between the cationic histone core and the anionic DNA that results in release of additional counterions into the bulk. As the NCPs stack up in a column, the wound DNA superhelix has freedom to equilibrate to the new charge distribution landscape. We hypothesize that this entropically driven process occurs due to elimination of edge effects in isolated NCPs that can affect the electrostatic interactions between the histone core and the DNA. This agrees with simple theoretical order-of-magnitude estimations for the charge matching conditions for a column of NCPs, as well as coarse grained molecular simulations.

30 positively charged monomers corresponding to histones and 10 negatively charged monomers corresponding to DNA were simulated in the presence of divalent ions (approximating Ca^{2+}). The initial configuration of the monomers corresponds to a non-altered inter DNA spacing of the native NCP. The positively charged monomers are fixed and negatively charged monomers are free to fluctuate in the Z-direction (their X and Y coordinates are also

fixed). The restoring force of stretching and compressing the DNA superhelix is modeled with a harmonic spring with a classical harmonic potential. The DNA monomers were first fixed, and then released from their fixed Z positions and allowed to vary. The Langevin equation for the particles was solved iteratively, and the Z-position of all DNA monomers during the simulation were recorded as a histogram. The simulations suggest that the inter-helical pitch tends to equilibrate towards a value in the 3.0-3.5 nm range (Figure 6.3A), well-matched with the steric size of TLR9, and consistent with SAXS findings and with recent measurements of superhelix pitch in NCP columns from cryo-electron microscopy [492]. Next, the Coulomb energy of the entire system was calculated over the simulation time frame, and the average energy was calculated before and after the DNA monomers were allowed to fluctuate. The average Coulomb energy of the system after the DNA monomers were released was $10k_BT$ lower than the average Coulomb energy of the system in the non-altered, fixed DNA-spacing (Figure 6.3B). Together, these results suggest that the DNA superhelix prefers to be in an energetically favorable state in which the pitch relaxes into a slightly larger inter-DNA spacing, which "turns on" the ability of these complexes to engage with TLR9 (Figure 6.1E).

6.3.3 LL37 organizes NCPs into hexagonal columnar nanostructures and amplifies TLR9-mediated inflammation

In addition to calcium, hBD3, and lysozyme, LL37 is a critical component of NETs and an important antimicrobial produced by neutrophils and other immune cells. LL37 has previously described TLR9 immunomodulatory activity via DNA binding, but its ability to amplify inflammation the presence of NCPs is not known. LL37 is a well-known autoantigen in lupus and psoriasis, and may play a role in amplifying inflammation from NETs. Here, we solve the structures of NCPs in complex with LL-37. We find that LL-37 condenses NCPs into a hexagonal columnar arrangement (Figure 6.4A-C). Unlike the more liquid-crystalline nematic ordering observed with calcium, hBD3, or lysozyme, LL37 can glue together adjoining NCP columns into a nanocrystalline aggregate, which increases ligand clustering and increases the potential for multivalent electrostatic interactions with TLR9. The diffraction peaks for the LL37-NCP complexes are located at $q_1 = 0.106, q_2 = 0.208$ Å^{-1} for the inter-

NCP spacing in the column, and $q_{10} = 0.064, q_{11} = 0.114, q_{20} = 0.130, q_{21} = 0.175, q_{22} = 0.229, q_{31} = 0.240, q_{41} = 0.305$ Å^{-1} for the hexagonal columnar phase (Figure 6.4B). The hexagonal lattice has an inter-columnar spacing of $d = 11.0$ nm and an inter-NCP spacing within the column of $h = 6.1$ nm. These lattice parameters correspond to the diameter and height of the NCP monomer (Figure 6.4C). To test the ability of LL37-NCP to activate immune cells, complexes were incubated with pDCs in the presence and absence of chloroquine. As expected, and consistent with observations for Ca, hbD3, and lysozyme, LL37-NCP immunocomplexes drastically amplified TLR9-dependent IFN-α production (Figure 6.4D).

To understand how LL37 can enable the formation of a hexagonal lattice while calcium cannot, we calculate the potential of mean force between adjacent NCP columns using a similar coarse-grained model described above for simulations of the inter-DNA spacings. Mobile multivalent ions of valence +2 and +6 are incubated with sectors of positive and negative monomers corresponding to adjacent NCP columns, and the potential energy is measured (U/k_BT) as a function of inter-columnar spacing R/σ. Simulations are conducted for NCP columns that have relaxed inter-DNA spacings (in the optimal range for TLR9 activation), and non-altered inter-DNA spacings (native NCP monomer). The potential of mean force shows that LL37 induces strong inter-columnar attraction, evidence by the energy minima observed at around $2R/\sigma$, and that the average energy is lower for the relaxed inter-DNA spacing (blue, Figure 6.4E) compared to the native non-altered inter-DNA spacing (red, Figure 6.4E).

6.3.4 TLR9 activation by remodeled chromatin depends on inter-DNA spacing and size of the NCP columns

Given the experimental results described, nanocrystalline ordering of NCPs by calcium, hBD3, lysozyme, and LL37 drastically amplifies TLR9 inflammation in pDCs, suggesting that multiivalent electrostatic effects are at play. To make sense of how and why these complexes amplify TLR9 biinding, we use a simple statistical mechanical model. We model the system of mobile membrane-bound TLR9 receptors. We assume that there are N_{tot}

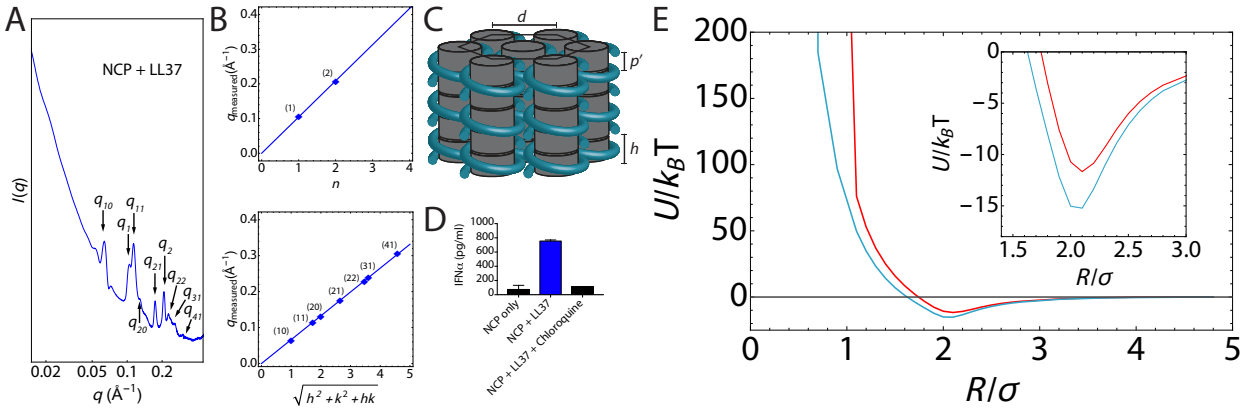


Figure 6.4: **LL37 organizes NCPs into hexagonal columnar nanostructures and amplifies TLR9-mediated inflammation.**
(A) SAXS data of isoelectric NCP-LL37 complexes demonstrate that LL37 can remodel NCPs into a hexagonal columnar lattice with long-range order. q_{hk} indices correspond to the hexagonal ordering, while the q_i indices correspond to the inter-NCP spacing within each NCP column. (B) Linear regression fits to reflections corresponding to the vertical inter-NCP spacing $h \sim 6$ nm (top) and the lateral inter-columnar hexagonal ordering $d \sim 11$ nm (bottom). These lattice parameters correspond to the known height and width of NCP monomers. Miller indices of observed reflections in (A) are indicated for each of the fits. (C) Schematic of the columnar hexagonal phase formed by NCP-LL37 with the lattice parameters measured from SAXS. (D) Stimulation of pDCs with NCP-LL37 amplify IFN-α production via TLR9, and the response is knocked down with an endosomal inhibitor. (E) Coarse-grained computer simulations of self-assembly of parallel NCP columns demonstrate that LL37-NCP nanostructures energetically prefer a "relaxed" pitch between superhelical turns of wrapped DNA, in concordance with previous simulations on isolated NCP columns. Potential of mean force calculations show that LL37-NCP structures with altered DNA spacings (at the range for optimal TLR9 activation) (green) have a lower free energy than the same structures with non-altered spacings (outside the range for optimal TLR9 activation) (blue). Furthermore, the difference in free energy between the two becomes greater at small inter-columnar distances, consistent with the strong attraction between parallel columns.

receptors on the membrane of area A. The receptors are activated upon forming a bond to dsDNA, which is wrapped around the NCPs. The strength of a single bond $E = G - \mu$ depends on the hybridization free energy G between a binding site and a receptor at the membrane and on the chemical potential μ, i.e. on the number of available (unbound) receptors $N_{tot}(1 - f_A)$ where f_A is the fraction of bound (activated) receptors. The binding free energy for the first bond is $\epsilon = \Delta G - k_B T \log(N_{tot}/A\rho_0)$ with ρ_0 the reference density and $k_B T$ the thermal energy. The subsequent bond formation is increasingly weaker: $E = \epsilon - k_B T \log(1 - f_A)$. A single NCP is wrapped with a 145 bp dsDNA and based on its size and shape, the number k_0 of possible bonds it can make with the receptors should be $1 < k_0 < 4$ (if TLRs can only bind to the dangling dsDNA ends than $k_0 \approx 2$). In the presence of multivalent cations the NCPs assemble into clusters, i.e. stacks (Ca^{2+}) or columnar lattices (LL37). In our model we assume that there are k discrete identical and independent binding sites on each cluster (this number is roughly proportional to the cluster area facing the membrane and scales with the cluster size) with two possible states: bound and unbound. The partition function counting all possible bonding arrangements is then

$$Q = \left(1 + e^{-E/k_B T}\right)^k = \left(1 + (1 - f_A)e^{-\epsilon/k_B T}\right)^k$$

The fraction $f_A = n_B N_{ads}/N_{tot}$ is proportional to the number of all bonds, which is the product of the expected number of adsorbed clusters N_{ads} and the mean number of bonds per cluster n_B, both of which can be derived from Q (see SI). In chemical equilibrium, this fraction can thus be determined by self-consistently solving the equation

$$f_A = C(1 - f_A)e^{-\epsilon/k_B T}\left(1 + (1 - f_A)e^{-\epsilon/k_B T}\right)^{k-1}$$

where the prefactor C is determined by the ratio of the NCP density in the solution and TLR9 surface density on the membrane. An estimate for the present system is $C \approx 10^{-5}$.

A contour plot showing the fraction of activated TLR9 receptors f_A depending on the NCP cluster size k and individual bond strength ϵ is shown in Figure 6.5. In order to observe appreciable activation and consequent response (IFN production), the activation level needs

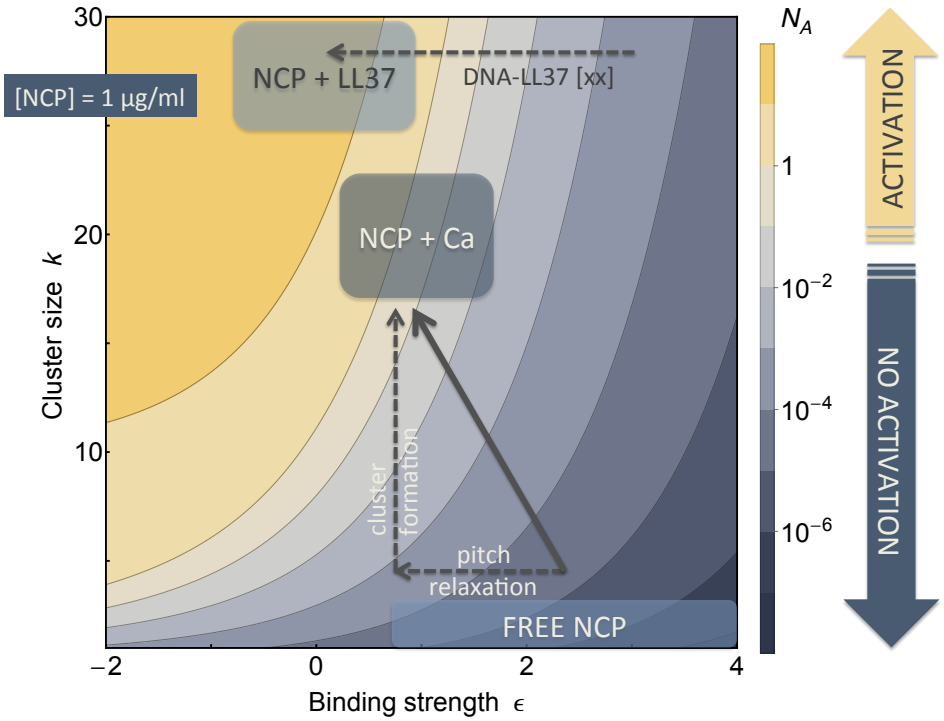


Figure 6.5: **TLR9 activation by remodeled chromatin depends on inter-DNA spacing and size of the NCP columns.**
A statistical mechanical model of NCP-TLR9 binding explains why NCP aggregates are immunogenic. TLR9 activation is assumed to be scale with the total number of bound TLR9 (N_A). The contour plot of the immune activation landscape of NCP nanocrystals predicts NA as a function of the TLR9-NCP complex binding strength ϵ (optimal inter-DNA spacing) and the size of the aggregates k (cluster size). Isotropic, monomeric NCPs exhibit low activation levels (blue) due to low binding strengths (suboptimal pitch of wrapped DNA) and the lack of multivalency (no columnar stacking). Incubation of NCPs with calcium or antimicrobial peptides (e.g. LL37) in NETs leads to dramatic remodeling and reorganization of NCPs into highly immunogenic columnar nanostructures (yellow). These NCP complexes exhibit increased binding strength (relaxation of pitch into an optimal inter-DNA spacing) and an increase in their domain size (increase in multivalent binding sites due to ligand clustering), which both drive a dramatic increase in the total number of bound TLR9.

to be above a certain threshold. This roughly corresponds to the yellow part of our phase diagram and the crossover from no activation to activation coincides with the saturation towards NCP concentration. Free NCPs are located in the lower right part of the diagrams with a very low activation level, while the clusters are close to the activation threshold. Taking into account the observations from the experiments (Figure 6.5), we can specify the region of the phase diagram where different clusters are located: since for LL37 the signal is generally stronger and almost independent of the NCP concentration, NCP-LL37 clusters have to be larger and lie in the yellow region as marked on the plots. The NCP- Ca^{2+} clusters only trigger a response for higher [NCP], therefore their position in the plot is not in the saturated regime. We also marked the comparison between LL37-DNA clusters and other peptide-DNA clusters studied previously [349], where only the selectivity towards the DNA lattice spacing – and consequently towards the effective binding energy – has been considered. The theory presented here is based on the assumption that the number of TLR9 receptors on the membrane is limited and thus the binding is anti-cooperative since the receptors are depleted within the binding process. The behavior is slightly different (but qualitatively the same) in the other limit, where the number of membrane-bound receptors is very large and the limiting factor is the saturation of adsorbed NCPs. This theory can also account for why the hbD3-NCP and lysozyme-NCP complexes induce strong IFN-α production, since

6.4 Conclusions and Prospects

In summary, the data presented here illuminates a structural mechanism for chromatin-mediated immune amplification in autoimmune diseases like lupus and psoriasis. Cationic molecules like antimicrobial peptides produced by neutrophils can self-assemble with nucleosomes released from decondensed chromatin into nanocrystalline immunocomplexes that drastically upregulate IFN-α production in pDCs (Figure 6.6). X-ray measurements, computer simulations, and mathematical modeling confirm that the nanocrystalline NCP columns exhibit altered DNA superhelix wrapping states that are well-matched with the steric size of TLR9, enabling multivalent intercalative binding to clustered arrays of TLR9.

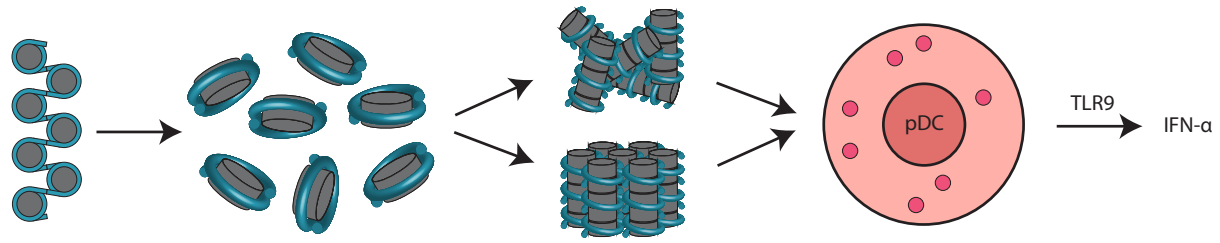


Figure 6.6: **NCPs released from processed chromatin self-assemble with antimicrobial peptides into crystalline nanostructures that hyperactivate TLR9 in pDCs.** Schematic demonstrating the mechanism of chromatin-mediated immune hyperactivation leading to IFN-α production in autoimmune disease. Chromatin in the form of linked NCPs is decondensed and processed into a heterogenous population of NCP monomers and higher order multimers. NCP monomers co-assemble with antimicrobial peptides like hbD3, lysozyme, and LL37 into columnar nanostructures that are recognized by pDCs. Internalization of complexes into endosomes enables multivalent binding to clustered arrays of TLR9 receptors, leading to production of IFN-α.

This has broad implications for understanding immune responses to chromatin in many additional contexts, including necrotic cell death, pyroptosis, and apoptosis, and provides a therapeutic interventional opportunity for the disruption and control of inflammation.

6.5 Methods

6.5.1 Preparation of polycation-NCP complexes

NCPs were formed *in vitro* via reconstitution from recombinantly expressed core histones H2A, H2B, H3, H4 and 145 bp DNA and dialyzing a 1:0.9 ratio of DNA and protein against a buffer with decreasing concentration of KCl. The NCP stock solution was diluted down from 35.686 mg/mL in 20 mM Tris pH 7.5, 1 mM EDTA and 10 mM KCl. To confirm purity, a 5% PAGE gel was run in 0.25× TBE buffer at 100 V. 200 ng of NCP was loaded per well with final 5% sucrose as a loading dye. Bands corresponding to NCP was observed at 350 bp with no free 145 bp DNA observed. Protein component of the NCP is 108.5 kDa (+146 charge), and the DNA component is 89.5 kDa (-144 charge), with net charge -144. Calcium chloride and human lysozyme were purchased from Sigma Aldrich. LL37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES) and hbD3 (GI-

INTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK, Disulfide bridge: 11-40, 18-33, 23-41) were purchased from Anaspec. Peptides are solubilized in nuclease-free water and physiological salt and buffer (100 mM NaCl + 10 mM HEPES at pH 7.4) to a final concentration of 10 mg/mL. Polycation-NCP complexes were formed by incubating fixed amounts of NCP with varying amounts of polycation (1-5 mg/mL for SAXS experiments, 10-20 μg/mL for cell experiments) at specific polycation-to-NCP charge ratios (P/NCP = 1:4, 1:2, 1:1, 2:1, 4:1). For a particular peptide, the same charge ratios were used in both SAXS and cell experiments. Calcium was mixed with NCPs at fixed concentrations ranging from 1 mM to 10 mM.

6.5.2 SAXS Experiments

We measured the phase diagram of the structures of polycation-NCP complexes by incubating polycations with NCPs at specific charge ratios in microcentrifuge tubes as described above. These complexes were equilibrated at room temperature overnight. Precipitated NCP-polycation complexes were loaded into 1.5 mm glass quartz capillaries (Mark-tubes, Hilgenberg GmbH) and hermetically sealed. SAXS experiments were conducted at the Stanford Synchrotron Radiation Lightsource (SSRL, BL 4-2) with monochromatic X-rays of energy 9 keV. Scattered radiation from complexes was collected using a Rayonix-MX225-HE detector (pixel size 73.2 μm). The 2D powder diffraction patterns were integrated using the Nika 1.76 package [184] for Igor Pro 7.04 (Wavemetrics) and FIT2D [185]. For all samples, multiple measurements were taken for consistency ($n \geq 3$). Samples were incubated at 37°C and centrifuged at 6000 RPM for 20 min before measurement. No significant changes were observed in scattering patterns over multiple experiments and exposures. SAXS data were analyzed by plotting the integrated scattering intensity $I(q)$ vs. q using Mathematica. To determine the phases present in each sample, the measured Bragg peak q-positions were obtained and their ratios were compared to the permitted reflections for isotropic columnar or hexagonal columnar liquid crystalline phases. Lattice parameters were obtained by fitting the peak positions with linear regression. For isotropic columnar complexes, a single peak is observed at q_1, corresponding to inter-NCP spacings along the column axis ($h = 2\pi/q_1$). For hexag-

onal columnar lattices, we observe peaks at $q_{hk} = \frac{2\pi}{d}\sqrt{\frac{4}{3}(h^2 + k^2 + hk)}$ where (h, k) are the Miller indices, and d is the lattice parameter corresponding to inter-columnar packing. Typical hexagonal lattices will have characteristic reflections at ratios of $\sqrt{1} : \sqrt{3} : \sqrt{4} : \sqrt{7} : \sqrt{9}$. Inter-DNA spacings corresponding to the pitch of the wrapped superhelical DNA are measured in a similar manner, with peak position q_p corresponding to inter-DNA spacings along the column axis ($p = 2\pi/q_p$). To calculate the domain size of the observed periodicities, the peaks corresponding to intra-columnar and inter-columnar ordering were fitted to squared-Lorentzians using least-squares regression in Mathematica. For more details of the fits, see section 5.5.2.

6.5.3 pDC Stimulation Experiments

Blood buffy coats of healthy donors were obtained from the Interregionale Blutspende SRK, Bern, Switzerland. Peripheral blood mononuclear cells (PBMCs) were obtained by Ficoll centrifugation of blood buffy coats. Plasmacytoid dendritic cells (pDCs) were then isolated by magnetic separation using the Diamond pDC Isolation kit II (Miltenyi). Purified pDCs were cultured in 96-well round-bottom plates at 5×10^4 per well in 200 μL RPMI 1640 (GIBCO) supplemented with 10% FCS (Atlanta Biologicals) and Pen/Strep in the presence of polycation-NCP complexes. Polycation-NCP complexes were generated by mixing 10 μg of NCP with 2 μg of LL37 or buforin in 40 μL of nuclease-free water (Ambion) and then diluted into 200 μL of complete medium for cell stimulation (final concentrations: 50 μg/mL hDNA and 10 μg/mL LL37 or buforin). In some instances, pDCs were pre-incubated with 1 μM chloroquine diphosphate (Sigma) for 15 min before stimulation. The different supernatants were collected after overnight culture and the levels of type I IFN production were measured by ELISA (human IFN-α pan, Mabtech).

6.5.4 Potential of Mean Force Calculations

Electrostatic interaction is short range for 100 mM NaCl, Debye length is 0.96 nm. $\lambda_D = 0.96$ nm is noticeably smaller than diameter of NCPs ($D_{NCP} = 11$ nm): $\frac{\lambda_D}{D_{NCP}} = 0.087$.

Accordingly, only the electrostatic interaction between the closest sectors of the two parallel NCPs columns is of considerable importance.

We use molecular dynamic simulation to describe a 2 nm sector of NCPs (containing 10 NCPs). We model the two sectors by two rods. In our simulations, each rod is composed of 30 spherical monomers of diameter σ ($\sigma = 2$ nm) separated at fixed intervals. The two rods are perfectly rigid and fixed at a specified center-to-center separation, R. In all cases, rods lie parallel to the face diagonal of the enclosing periodic box. In addition, we introduce mobile multivalent ions of charge +2 (as Ca^{+2}) and +6 (as LL37), and mobile monovalent ions of charges +1 and -1. The system is always electrostatically neutral. We reference salt concentrations to the total charge on the two rods for example, a 2:1 salt concentration of $c_{2:1} = 1$ means that the total charge due to +2 ions is equal to the total charge on the two rods. Similarly, a 6:1 salt concentration of $c_{6:1} = 1$ means that the total charge due to +6 ions from the monovalent salt is equal to the total charge of the two rods. We include two types of pair interactions between particles. First, we use the truncated Lennard-Jones potential to allow for short-range repulsion. This introduces the energy scale ϵ and the particle size σ. Second, we include the Coulomb interaction, $\frac{Z_1 Z_2}{\epsilon r_{12}}$, where Z_i is the charge on particle i and ϵ is the dielectric constant. To handle the long-range Coulomb interaction in our system with periodic boundaries, we use the particle-particle particle-mesh (P3M) method. Our simulations are carried out in the canonical (NVT) ensemble with the temperature fixed at $k_B T = 1$ (All energies are scaled with $k_B T$) using the Langevin thermostat. The separation between monomers is $l = \sigma$ and the Bjerrum length, $l_B = \frac{1}{k_B T}$ (the distance at which the electrostatic interaction between two electronic charges is equal to the thermal energy), is $l_B = 0.7$ nm $= 0.35\sigma$. For NCP altered/non-altered-DNA spacing, we assume that charge distribution is as represented in **Figures 4, 5**. We keep the rods fixed and the distance between rods is R. All of the other charged particles are allowed to diffuse inside the simulation box via thermal fluctuation. The characteristic time scale for the simulation is $\tau = \sqrt{\frac{m\sigma^2}{\epsilon}}$ where m is the particle mass. We calculate electrostatic force on the rods and obtain its average over simulation. Then we integrate the force with rod-rod separation R and obtain the potential of mean force.

6.5.5 Coarse-Grained Simulations of DNA Superhelix Spacings and Energy Calculation

We assume that 30 positive and 10 negative monomers, referred to Histone and DNA respectively in the presence of divalent ions (as Ca^{+2}). The initial configuration of the monomers corresponds to no-altered DNA spacing. The positive monomers are fixed and negative monomers are free to fluctuate in the Z-direction and their X and Y coordinates are fixed. Also, we suppose bonded form in which all negative monomers are connected to their neighbors explicitly by the harmonic spring with a classical harmonic potential. In which $r = R = 3.2\sigma$ is the equilibrium separation between neighboring monomers, and the prefactor is $K = \frac{50k_BT}{\sigma^2}$. It is given by $V(r) = \frac{1}{2}K(r-R)^2$. We solved the Langevin equation for the particles and recorded the position of all negative monomers during the simulation. Our results demonstrate that the separation between monomers is approximately, $H = 3\sigma$. We calculate the system Coulomb energy in two ways; 1) We keep all of the monomers fix in no-altered DNA configuration and leave the counterions to fluctuate and calculate Coulomb energy after equilibrium of the system. 2) After equilibrium of the counter-ions, we remove the constraints of the negative monomers and leave them to fluctuate and calculate the Coulomb energy.

CHAPTER 7

Structural studies of host defense peptides that lead to TLR9-based immune activation

7.1 Abstract

Alpha helices are ubiquitous. A subset of them can form self-assembled complexes with double-stranded DNA (dsDNA) and potently activate the innate immune system via Toll-like Receptor 9 (TLR9). For example, electrostatic complexes between the endogenous antimicrobial peptide (AMP) LL37 and dsDNA can activate TLR9 in plasmacytoid dendritic cells (pDCs), thus triggering production of type-I interferons (IFNs) that can exacerbate inflammation in diseases such as lupus and psoriasis. Clearly, most α-helical motifs (and many α-helical AMPs) do not exhibit this immunological activity. Recent work showed that dsDNA immunocomplexes that are pro-inflammatory via the TLR9 pathway have inter-dsDNA lattice spacings that match the steric size of TLR9, so that spatially-organized parallel dsDNA ligands can amplify recruitment and multivalent binding to clustered arrays of TLR9. However, at present, it is unclear what physicochemical parameters of α-helices allow for geometric organization of dsDNA into these optimal structures, and why peptides with similar charges lead to drastically different immune activation profiles. In this work, we examine the self-assembled structures of immunocomplexes formed between dsDNA and three different α-helical AMPs (mellitin, LL37, buforin) using high resolution synchrotron small angle X-ray scattering (SAXS). By correlating these measured structures to computer simulations and to their induced immunological activity in pDCs and macrophages, we deduce some fundamental structural rules for α-helices that are proinflammatory via the TLR9 pathway. We find that the two necessary structural criteria are the peptide's ability to condense ds-

DNA and to form large-diameter coiled-coil super helices, both of which can be met by bent, facially-amphiphilic α-helices that are cationic with charge densities near the Manning limit. It is likely not a coincidence that these criteria are satisfied by a number of AMPs, at least some of which are known to be immunomodulatory. Using a subset of melittin-related peptides, we conducted a detailed dissection of the tradeoffs between membrane activity and dsDNA-binding/TLR9 activating ability of α-helical AMPs. Interestingly, we discover that the molecular rules governing membrane activity and immunomodulation via TLRs are overlapping but distinct, which has implications for the design of multifunctional AMPs with therapeutic potential.

7.2 Introduction

Antimicrobial peptides (AMPs) are multifunctional molecules that play a key role in microbial host defense [1, 355]. AMPs are polycationic, amphipathic, and kill pathogens primarily by disrupting their cell membranes and inhibiting intracellular processes [1, 6, 2]. However, recent evidence suggests that AMPs also play a dual role in modulating inflammation [110, 494, 355, 304, 305]. AMPs can trigger cytokine production, induce chemoattraction, and elicit immune cell differentiation and proliferation [494]. AMPs have also been shown to modulate Toll-like receptor (TLR) activation by endogenous and microbial immune ligands such as lipopolysaccharide and nucleic acids [495, 315, 316, 318, 496, 321, 353]. In the context of normal immune responses to microbial invasion, AMP immunomodulation synergizes with its direct antimicrobial activity. However, aberrant expression of AMPs can lead to undesirable amplification of inflammation in autoimmune diseases [338, 330, 332]. For example, in lupus and psoriasis, the endogenous α-helical AMP LL37 breaks immune tolerance to self-DNA released from dying cells by forming insoluble complexes, driving type I interferon production via TLR9 in plasmacytoid dendritic cells (pDCs) [330, 338, 331].

Until recently, the molecular basis for AMP-mediated TLR hyperactivation was unknown. In two complementary studies, we demonstrated that AMPs like LL37 can modulate TLR9 [349] and TLR3 [378, 362] signaling by scaffolding and patterning double-stranded DNA

(dsDNA) and double-stranded RNA (dsRNA), respectively, into ordered nanocrystalline immunocomplexes. AMPs electrostatically self-assemble with dsDNA to form ordered AMP-nucleic acid nanocrystals. TLRs detect and respond to the crystallinity of these AMP-nucleic acids complexes (e.g. inter-ligand spacing and crystal size), which quantitatively determines the degree of inflammation [378]. AMP-dsDNA immunocomplexes with an inter-dsDNA spacing well matched to the steric size of TLR9 (3-4 nm) maximally activated TLR9 in pDCs (e.g. LL37-dsDNA), while those with spacings mismatched with the size of TLR9 did not (e.g. TAT-dsDNA) [349, 350, 441]. While a diverse range of polycationic molecules and peptides were able to condense dsDNA, only a small number were able to scaffold dsDNA at the optimal spacing. Interestingly, immune complexes formed between LL37 and dsDNA had larger inter-dsDNA spacings measured from SAXS than the inter-dsDNA spacings predicted purely from the size of LL37 and electrostatic charge matching. Furthermore, we observed that some α-helical AMPs were capable of condensing dsDNA into this narrow range of spacings that allowed for TLR9 activation, while others were unable to [349]. Although the α-helix is a common secondary structure element in proteins, it is clear that not all α-helices can bind to dsDNA and induce inflammation via TLR9.

Motivated by this observation, we set out to determine which physicochemical parameters of α-helices allow for geometric organization of dsDNA into nanocrystalline immunocomplexes at optimal inter-dsDNA spacings for TLR9 activation. To do this, we selected three naturally-occurring α-helical AMPs (LL37, melittin, and buforin) and solved their structures in complex with dsDNA using small-angle X-ray scattering (SAXS). Then, using the measured diffraction pattern as a guide, all-atom molecular dynamics (MD) simulations were conducted to model the self-assembly of the AMPs with dsDNA. Using an iterative process, several energy-minimized AMP-dsDNA lattice structures recapitulating our SAXS findings were obtained for each complex. To validate the ability of these complexes to activate TLR9, mouse macrophages and human pDCs were stimulated and type I IFN production was measured. We found that common core features in AMPs which allow them to punch holes in bacterial membranes also optimize their ability to condense dsDNA into immunologically-active complexes. Bent, amphipathic α-helical peptides with a hydrophobic

face $>\sim$ 90-100° and a sufficient charge density near the Manning limit can oligomerize into a protofibril via hydrophobic interactions in the interior. This AMP protofibril minimizes solvent exposure of the hydrophobic core by efficiently burying hydrophobic volume while simultaneously clustering cationic residues enabling multivalent binding to dsDNA phosphate backbones. LL37, melittin, and buforin have sufficient charge density to condense dsDNA into nanocrystalline complexes. However, only LL37 and melittin are able to condense dsDNA into columnar square/tetragonal lattices with four-fold coordination, driven by self-assembly of the peptides into protofibrils with four-fold symmetry. In contrast, buforin induces formation of a disordered columnar dsDNA lattice with reduced crystallinity. The average inter-dsDNA spacings for all three complexes fall into the immunogenic range previously established [349, 350]. These spacings are commensurate with the size of TLR9, which optimizes multivalent electrostatic interactions between the dsDNA immunocomplexes and clustered TLR9.

In human and murine immune cell stimulation experiments, we find that the LL37-dsDNA and buforin-dsDNA complexes induce inflammation via TLR9, and that differences in their crystallinity quantitatively affect cytokine production. Interestingly, melittin-dsDNA complexes were unable to induce TLR9 activation despite the requisite structural parameters due to cellular toxicity. Systematic reduction of the membrane activity of melittin in a series of mutants with conserved dsDNA-binding activity corresponded with a recovery of TLR9-activating ability, indicating that the requirements for membrane activity and dsDNA-binding/TLR9 activation are overlapping but distinct. Taken together, these results provide insight into how the immunomodulatory and membrane active properties of α-helical AMPs can be engineered. This has broad implications for the design of new immunotherapies that can modulate the immune system via geometric patterning of immune ligands.

7.3 Results and Discussion

7.3.1 Selection of candidate α-helical AMPs

LL37 (+6 charge at pH 7.4) has a curved amphipathic helix-bend-helix motif with a glycine at the 14th position [497, 31]. This facilitates its antimicrobial properties, allowing it to adsorb onto bacterial membranes and induce pore formation. However, these same properties also allow it to condense dsDNA into nanocrystalline immune complexes. We hypothesized that LL37-like bent α-helical AMPs with sufficient charge density and amphipathicity would be able to condense dsDNA into a lattice with the optimal spacing for TLR9 activation. Typically, bent α-helices are found in soluble peptides and transmembrane helices [498]. The main kink-inducing amino acids are glycine and proline [499] and are typically present toward the center of the α-helix. The presence of proline and glycine can bend helices by 10-20 Å and allow for increased backbone flexibility [500]. We hypothesized that this shape also allows it to self-assemble into a coiled-coiled in the presence of dsDNA. This self-assembly allows for organization of parallel dsDNA rods at a spacing larger than expected for single peptides alone. To look for other candidate peptides that share a similar structural morphology, we collated 301 α-helical AMPs from the Antimicrobial Peptide Database [9] (`http://aps.unmc.edu/AP/about.php`) and analyzed their sequence content. A total of 76/301 AMPs contained a single proline kink, 50 of which were within 5 amino acids of the peptide center. 61/301 AMPs contained a single glycine kink, 41 of which were within 5 amino acids of the peptide center. As a comparison to the glycine-kinked LL37, we selected two well-known α-helical AMPs from this subpopulation with single proline kinks. Melittin (+5 charge at pH 7.4) is 26 amino acids with a proline at the 14th position, and buforin (+6 charge at pH 7.4) is 21 amino acids with a proline at the 11th position. Using this set of three model peptides, we mapped out a phase diagram of self-assembled AMP-dsDNA nanostructures with potential to activate TLR9. Using SAXS, we proceeded to solve the structures of LL37, melittin, and buforin-dsDNA complexes at varying peptide-to-dsDNA charge ratios.

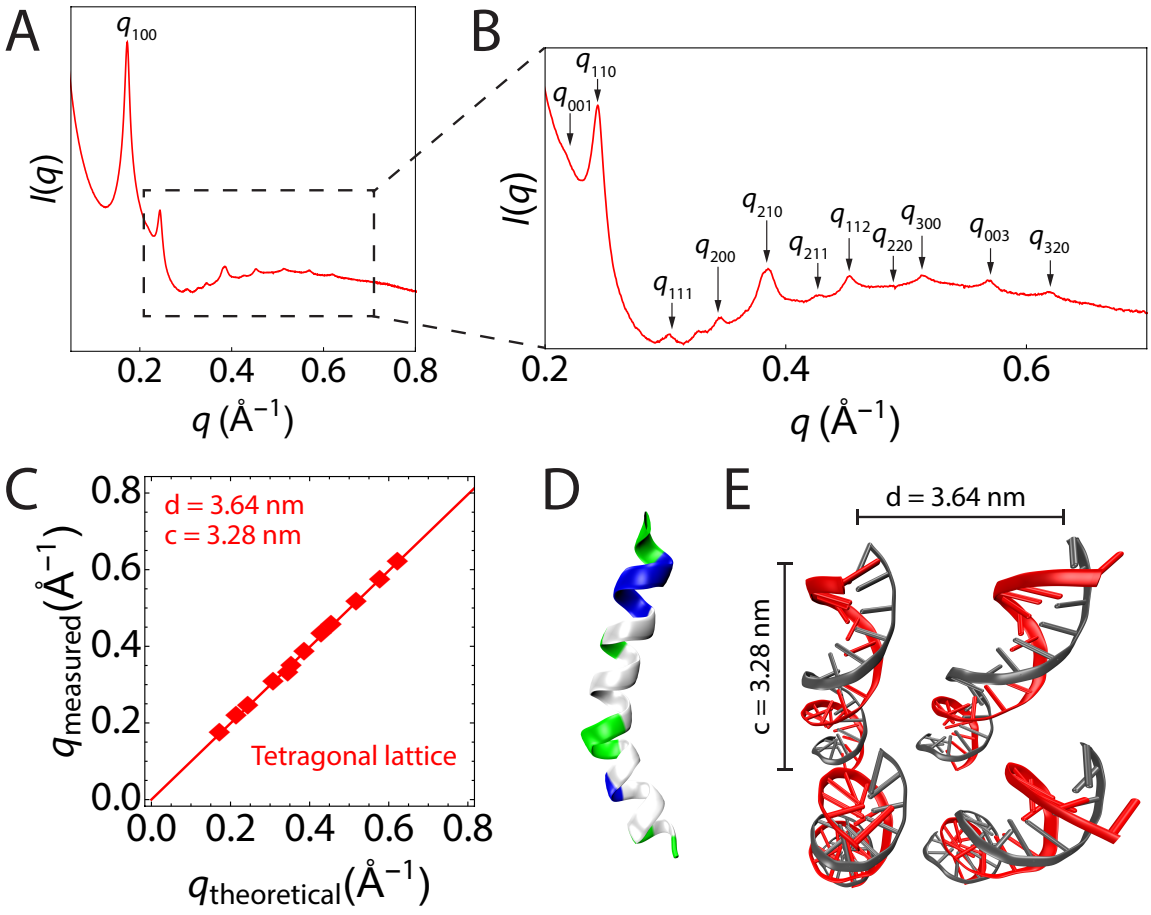
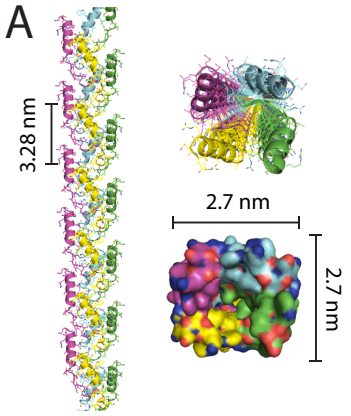


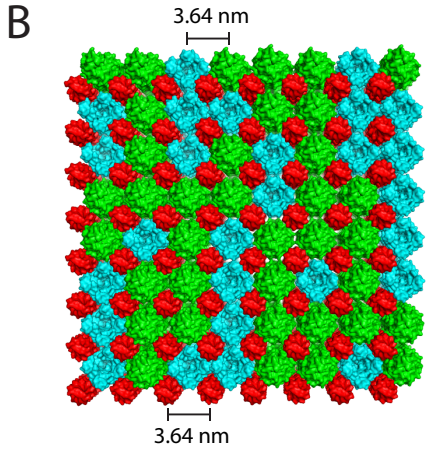
Figure 7.1: **SAXS structure of melittin-dsDNA complexes.**
(A) Representative SAXS spectra of an isoelectric melittin-dsDNA complex with the first Bragg peak q_{100} indicated. (B) The magnified portion of the spectra shows the higher order reflections (hkl) indicative of a 3D tetragonal lattice. (C) We present the higher order reflections of the 3D tetragonal melittin-dsDNA lattice (q_{hkl}). To obtain the lattice parameters, we fit the measured Bragg peak positions using multiple linear regression. The lattice parameters of the melittin-dsDNA complex are $d = 3.64$ nm and $c = 3.28$ nm ($R^2 = 0.999$). (D) Structure of the melittin monomer (PDB: 2MLT_A [33]). Cationic residues are colored blue, and hydrophobic residues are colored white. (E) Schematic of parallel dsDNA rods within the melittin-dsDNA tetragonal lattice labeled with lattice parameters obtained from SAXS. The average domain size of the melittin-dsDNA lattice is $L \sim 31.2$ nm.

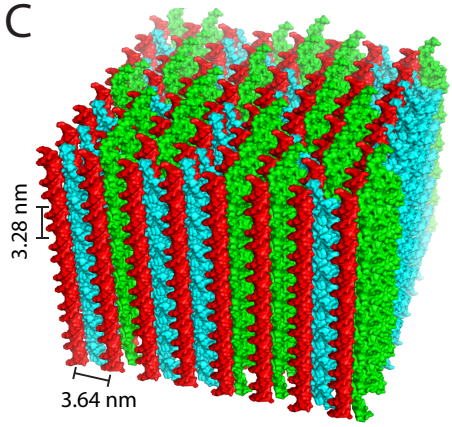
7.3.2 Melittin condenses dsDNA into a tetragonal lattice

The formation of condensed phases of dsDNA are important in biology, including the wrapping of dsDNA in chromosomes [409, 411] and the delivery of genes into cells [414, 415, 417]. Polyelectrolyte chains like dsDNA can collapse into a compact phase from solution as increasing concentrations of polycations are added, driven by the entropy gain of counterion release into bulk solution [388, 389, 390, 462, 451]. Condensation of anionic dsDNA by small multivalent cations like hexammine cobalt and spermine and large gene delivery dendrimers have been studied in-depth [452, 501, 502], and their structures have been well characterized using X-ray scattering and molecular simulation [416, 413, 503, 504, 505]. However, the molecular basis for dsDNA condensation mediated by biological macroions such as AMPs and other large cationic proteins is poorly understood.

To understand the structural basis of dsDNA condensation and immune stimulation by cationic, amphipathic α-helical AMPs, we begin by studying the structure of melittin-dsDNA complexes using a combination of SAXS and molecular simulations. Melittin is a well-studied prototypical α-helical AMP from honey bee venom that also has diverse functions, including cell lysis, antifungal and antibacterial activity, and disruption of ion conductance [17, 507, 508]. Melittin has been reported to bind dsDNA in the context of gene transfer [509], but the structure of the complex has never been studied. Melittin itself is structurally similar but sequentially non-homologous to LL37. Melittin is cationic and amphipathic, with cationic residues more concentrated towards the C terminus, and hydrophobic residues more concentrated toward the N terminus [510]. From SAXS, we find that the melittin-dsDNA complex forms a 3D tetragonal lattice with Bragg reflections at $q_{100} = 0.173$, $q_{001} = 0.215$, $q_{110} = 0.244$, $q_{111} = 0.308$, $q_{200} = 0.345$, $q_{210} = 0.386$, $q_{211} = 0.429$, $q_{112} = 0.454$, $q_{220} = 0.492$, $q_{300} = 0.518$, $q_{003} = 0.577$, and $q_{320} = 0.622$ Å^{-1} (Figure 7.1A, 7.1B). The lattice parameters are $d = 3.64$ nm and $c = 3.28$ nm (Figure 7.1C). This indicates that the inter-dsDNA spacings between parallel dsDNA rods is 3.64 nm. The out of plane parameter indicates that the complex has additional order along the dsDNA axis, with an average spacing of $c = 3.28$ nm. The melittin monomer (Figure 7.1D) has a diameter of ~ 1.2 nm. At first glance, the inter-







D

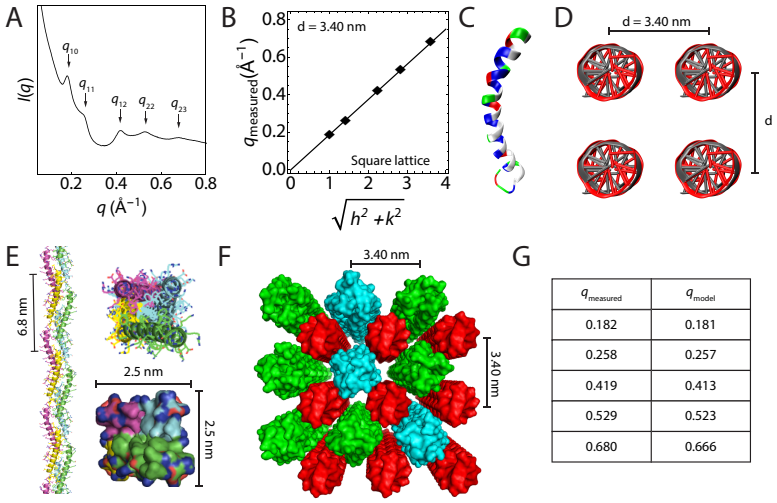
$q_{measured}$	q_{model}
0.173	0.168
0.215	0.204
0.244	0.246
0.308	0.302
0.345	0.344
0.386	0.384
0.429	0.431
0.454	0.457
0.492	0.497
0.518	0.525
0.577	0.578
0.622	0.621

Figure 7.2: **Structure of melittin-dsDNA complexes from molecular simulations** (A) Side and top views of the melittin superhelix self-assembled in the presence of dsDNA. The cross-sectional dimensions of the superhelix is 2.7 nm × 2.7 nm. The pitch of the superhelix is 3.28 nm per turn. Top (B) and side (C) views of an 8×8 melittin-dsDNA square lattice from MD simulations. Columnar dsDNA is colored red, and the melittin superhelix is colored green (N to C terminus polarity) and teal (C to N terminus polarity). Random polarity of the melittin superhelix is assumed. The inter-dsDNA spacing from the molecular model obtained from simulation is $d = 3.64$ nm, which agrees with the predicted inter-dsDNA spacing from SAXS. The out-of-plane lattice parameter of $c = 3.28$ nm also agrees with the simulated pitch of the melittin superhelix. (D) There is good agreement between the observed structure factor peaks from SAXS with peaks recovered from theoretical X-ray scattering of the lattice model in (B) and (C) from CRYSOL [506].

dsDNA spacing of 3.64 nm is larger than the $\sim$ 3.2 nm one might expect from melittin alone (Figure 7.1E). To better understand why this occurs, we conducted MD simulations on the self-assembly of melittin. We hypothesized that melittin can self-assemble into a coiled coil in the presence of dsDNA, forming a "superhelix" that can organize dsDNA into a tetragonal lattice. Starting from the NMR monomer structure of melittin (PDB: 2MLT_A), we modeled a 26mer of melittin self-assembled into a superhelix in the presence of parallel dsDNA rods using TINKER and the AMBER force field (see Computational Methods) (Figure 7.2A). We obtained an optimized monomer structure within the melittin superhelix. The superhelix has a cross sectional dimensions of 2.7 nm $\times$ 2.7 nm, with a pitch between monomers of 0.82-0.86 nm ($\sim$ 3.28 nm/turn). Melittin packs into a superhelix with hydrophobic residues in the interior, and cationic residues exposed to the exterior. We constructed an 8×8 square lattice of dsDNA rods with the energy-minimized melittin superhelices (Figure 7.2B, 7.2C). To compare the MD-derived melittin-dsDNA lattice structure to the SAXS data, we use CRYSOL to diffract the atomic model. Theoretical diffraction of the energy-minimized structure (Figure 7.2B-7.2C) reproduces most of the scattering peaks seen in the experimental SAXS data ($q_{100} = 0.168$, $q_{001} = 0.204$, $q_{110} = 0.246$, $q_{111} = 0.302$, $q_{200} = 0.344$, $q_{210} = 0.384$, $q_{211} = 0.431$, $q_{112} = 0.457$, $q_{220} = 0.497$, $q_{300} = 0.525$, $q_{003} = 0.578$, and $q_{320} = 0.621$ Å^{-1}) (Figure 7.2D). This provides a plausible mechanism for the observed inter-dsDNA spacing formed by melittin-dsDNA complexes.

7.3.3 LL37 condenses dsDNA into a 2D columnar square lattice

LL37 is a multifunctional human α-helical AMP that is essential to innate host defense. LL37 kills bacteria by forming pores in their cell membranes [34]. However, LL37 is also central to the pathogenesis of autoimmune diseases [332]. LL37 has been observed to enable immune recognition of self-antigens in lupus and psoriasis by forming immune complexes with dsDNA [332, 330, 331]. Whereas melittin has a proline kink, LL37 has a glycine kink. Here, we present the structure of the isoelectric LL37-dsDNA immune complex. The LL37-dsDNA complex forms a 2D square columnar lattice with Bragg reflections at $q_{10} = 0.182$, $q_{11} = 0.258$, $q_{12} = 0.419$, $q_{22} = 0.529$, and $q_{23} = 0.680$ Å^{-1}, corresponding to a lattice



$q_{measured}$	q_{model}
0.182	0.181
0.258	0.257
0.419	0.413
0.529	0.523
0.680	0.666

Figure 7.3: **Structure of LL37-dsDNA Complexes from SAXS and molecular simulations.**
(A) Representative SAXS spectra from an isoelectric LL37-dsDNA complex. All observed reflections are labeled with their respective Miller indices (hk). (B) 2D square lattice reflections (q_{hk}) from (A) are fitted, yielding a lattice parameter of $d = 3.40$ nm ($R^2 = 0.999$). (C) Structure of the LL37 monomer (PDB: 2K6O [31]). Cationic residues are colored in blue and hydrophobic residues are colored in white. (D) Schematic of the LL37-dsDNA square columnar lattice with an inter-dsDNA spacing of $d = 3.40$ nm. The average domain size of the LL37-dsDNA lattice is $L \sim 8.1$ nm. (E) Side and top views of the self-assembled LL37 superhelix structure formed in the presence of dsDNA. The superhelix has rough cross-sectional dimensions of 2.5 nm $\times$ 2.5 nm. The pitch of the superhelix is 6.8 nm per turn. (F) Top view of a 3×3 LL37-dsDNA square lattice from MD simulations. Columnar dsDNA is colored red, and the LL37 superhelix is colored green (N to C terminus polarity) and teal (C to N terminus polarity). Random polarity of the LL37 superhelix is assumed. The inter-dsDNA spacing from the molecular model is 3.4 nm, which agrees with the inter-dsDNA spacing predicted from SAXS. (G) There is good agreement between the observed structure factor peaks from SAXS with peaks recovered from theoretical X-ray scattering of the lattice model in (F) from CRYSOL.

parameter of d = 3.40 nm (Figure 7.3A, 7.3B). This indicates that the spacing between parallel columns of dsDNA rods is 3.40 nm in this complex, which is slightly smaller than the inter-dsDNA spacing of melittin-dsDNA (3.64 nm). However, the known diameter of LL37 is around $\sim$ 1.2 nm (Figure 7.3C), which is too small to explain the observed inter-dsDNA spacing from SAXS (3.4 nm). We hypothesize that LL37 can also oligomerize into a coiled coil superhelix in the presence of dsDNA, organizing dsDNA into a square columnar lattice (Figure 7.3D). We conducted MD simulations in a similar manner to the melittin-dsDNA complex to develop a physical model of LL37 self-assembly. Starting from the crystal structure of LL37 (PDB: 2K6O), we optimized the structure of an LL37 26mer. The final superhelix structure yields 4 LL37 per helix turn, with a superhelix pitch of 6.8 nm and cross section of 2.5 nm $\times$ 2.5 nm (Figure 7.3E). Similar to melittin, the bent, amphipathic nature of LL37 allows for favorable interactions between the hydrophobic faces of adjacent LL37 in the interior of the superhelix, allowing for outward exposure of the cationic residues on the opposite face. The cationic backbone of the superhelix can then coordinate 4 dsDNA rods simultaneously, forming a square lattice. Next, we create a full LL37-dsDNA square lattice model by superimposing 8 $\times$ 8 dsDNA rods coordinated with multiple LL37 superhelices (Figure 7.3F). We observe that the LL37-dsDNA lattice from simulation has an inter-dsDNA spacing of 3.40 nm, which is almost identical to the measured value from SAXS. Using CRYSOL, we observe good correspondence between the major peaks observed in the SAXS diffraction spectrum and the predicted peaks from the atomic structure of simulated lattice (q_{10} = 0.181, q_{11} = 0.257, q_{12} = 0.413, q_{22} = 0.523, and q_{23} = 0.666 Å^{-1}) (Figure 7.3G). We studied effects of various structural parameters on the diffraction pattern. Changing the chirality of the LL37 superhelix (left vs. right handed), the direction distribution (C to N vs. N to C), and phase distribution did not appreciably change the predicted diffraction pattern, suggesting that the observed peaks are directly due to the formation of a square lattice of dsDNA rods intercalated with LL37. Here, we find that the LL37-dsDNA square lattice has a similar inter-dsDNA spacing to the melittin-dsDNA tetragonal lattice. However, the melittin-dsDNA complex has stronger order along the longitudinal axis of the dsDNA rod compared to LL37-dsDNA.

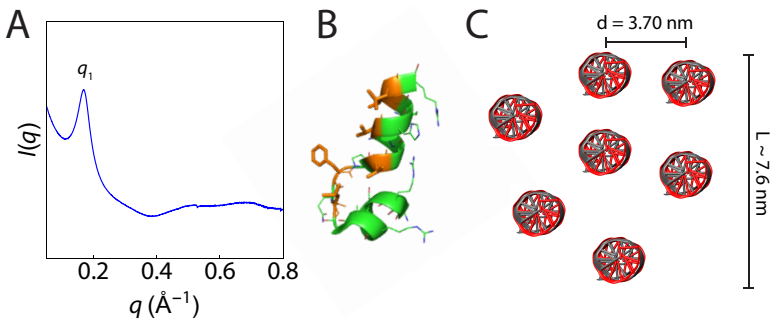


Figure 7.4: **Structure of buforin-dsDNA complexes.**
(A) Representative SAXS spectra of an isoelectric buforin-dsDNA complex. One reflection is observed at $q_1 = 0.170$ Å^{-1}, consistent with a columnar lattice with short-ranged order. The average domain size is extracted from the first peak by fitting the peak width. (B) Structure of the buforin monomer (taken from the H2A histone, PDB: 4KGC [511]). (C) Illustration of the buforin-dsDNA columnar lattice with average inter-dsDNA spacing of $d = 3.70$ nm and average domain size of $L \sim 7.6$ nm.

7.3.4 Buforin condenses dsDNA into a disordered columnar lattice

To determine whether the size of the hydrophobic face is essential to the formation of square columnar DNA lattices with the optimal inter-DNA spacing for TLR9 activation, we chose to study buforin, which is an antimicrobial peptide derived from histone H2A of the Asian toad Bufo bufo garagrizans. Buforin is structurally homologous to LL37 and melittin with a core α-helical region linked by a proline hinge to an extended helical region with a random coil [267, 512]. Buforin is known to penetrate cell membranes [513, 514] and bind to dsDNA and RNA within bacterial cells, disrupting replication [105]. In addition, buforin is a known degradation product of histone H2A, and presence of buforin mRNA in humans indicate a role in innate host defense [515]. Compared to melittin and LL37, buforin has similar net charge and charge density, but has a much smaller hydrophobic face ($\sim 80°$). Unlike LL37 and melittin, buforin is not a full α-helix. The helical portion distal to the proline kink is a random coil. SAXS studies of the buforin-dsDNA complex revealed a single diffraction peak at $q_1 = 0.170$ Å^{-1}, indicating the formation of a columnar phase with short-ranged order at an average inter-dsDNA spacing of $d = 3.70$ nm (Figure 7.4A). The buforin-dsDNA lattice is unique from the LL37- and melittin-dsDNA lattices because it only has one dimensional order. Buforin organizes dsDNA into parallel ligands, but the structure does not

conform to a periodic two-dimensional lattice. In contrast, the LL37 and melittin-dsDNA complexes have multiple higher order reflections, indicating stronger crystallinity than the buforin-dsDNA complexes (Figure 7.1, 7.2, 7.3). The increased disorder of the buforin-DNA complex implies that buforin cannot assemble into protofibril with 4-fold symmetry, which is required for formation of the crystalline square/tetragonal lattice, like melittin and LL37. Instead of stacking periodically like melittin or LL37, buforin (Figure 7.4B) likely forms multimers intermediate coordination numbers between 3-5, which then condenses dsDNA into a columnar lattice with short-ranged order (Figure 7.4C). The inability of buforin to pack hydrophobic volume efficiently and form a square lattice despite sufficient charge density to condense DNA indicates that there is a minimum constraint on hydrophobic face size ($>\sim$90-100°) for helical AMPs to coordinate DNA with 4-fold symmetry.

7.3.5 Bent cationic antimicrobial peptide-dsDNA complexes amplify inflammation by binding to TLR9

To investigate the ability of melittin-, LL37-, and buforin-dsDNA complexes to activate TLR9, we incubated pre-mixed AMP-dsDNA complexes with murine wild type (WT) and TLR9-knockout (TLR9-KO) macrophages and measured subsequent IFN-β production (Figure 7.5A). In a parallel assay, we also stimulated TLR9-expressing human pDCs in the absence and presence of chloroquine, and measured subsequent IFN-α production (Figure 7.5B). LL37 and buforin were mixed with mouse genomic dsDNA (mdsDNA) in the macrophage experiments and human genomic dsDNA (hdsDNA) in the pDC experiments. Melittin was unable to induce IFN-α production because it was highly cytotoxic to macrophages and dendritic cells (see SI). In both assays, we observed significant levels of TLR9-dependent type I IFN production in response to both LL37-dsDNA ($p < 0.01$ macrophage, $p < 0.05$ pDC) and buforin-dsDNA complexes ($p < 0.05$ macrophage, $p < 0.05$ pDC) (Figure 7.5A-7.5B). In the murine macrophage system, both complexes produced IFN levels comparable to or greater than stimulation with LPS (TLR4 agonist) or ODN 1826 (TLR9 agonist). No significant activation was observed by peptides alone or mouse genomic dsDNA alone. In the human pDC system, both LL37-dsDNA and buforin-dsDNA complexes

produced significant TLR9-dependent IFN production compared to controls. The presence of 1 μM chloroquine, an endosomal acidification inhibitor, abrogates all TLR9-dependent signaling. This data confirms that both LL37 and buforin, two prototypical cationic amphipathic AMPs can condense dsDNA into liquid-crystalline immune complexes and drive immune cell activation in a TLR9-dependent manner. Genomic dsDNA by itself is non-immunogenic, but complexation with AMPs enables breakdown of immune tolerance [330]. The consistency in results between the murine and human systems suggest this phenomenon is general, and reinforces the idea that the crystallinity of these AMP-dsDNA complexes are the primary determinants governing TLR9 activation. We observed that the LL37-dsDNA complex activates 3-5$\times$ more strongly than the buforin-dsDNA complex. To understand this difference in activation, we measured the domain sizes of each of the two complexes and calculated the total number of unit cells in each complex. The average domain sizes for the LL37-dsDNA, melittin-dsDNA, and buforin-dsDNA complexes are 8.07 nm, 31.21 nm, and 7.58 nm respectively, corresponding to roughly 2.37, 8.57, and 2.04 unit cells. The LL37-dsDNA complex has a larger number of repeat units than the buforin-dsDNA complex, which could partially explain the difference in activation level. However, another major difference between the LL37 and buforin-dsDNA lattices is that the LL37-dsDNA lattice exhibit more crystalline structure. Whereas the buforin-dsDNA lattice is disordered columnar, the LL37-dsDNA lattice is a square lattice with well-defined higher order correlation peaks. It is likely that AMPs which condense dsDNA into more crystalline complexes lead to stronger activation. The buforin-dsDNA complex has the right inter-dsDNA spacing for TLR9 amplification, but it does not possess the crystallinity of the LL37-dsDNA complex that also has an ideal inter-dsDNA spacing for TLR9 activation.

7.3.6 Tuning immunomodulatory activity of AMPs by modifying their membrane activity

The observation that melittin has poor IFN-inducing abilities via TLR9 is an intriguing one, especially since melittin-dsDNA complexes form a tetragonal complex with an optimal inter-dsDNA spacing similar to that of the LL37-dsDNA complex. The physicochemical

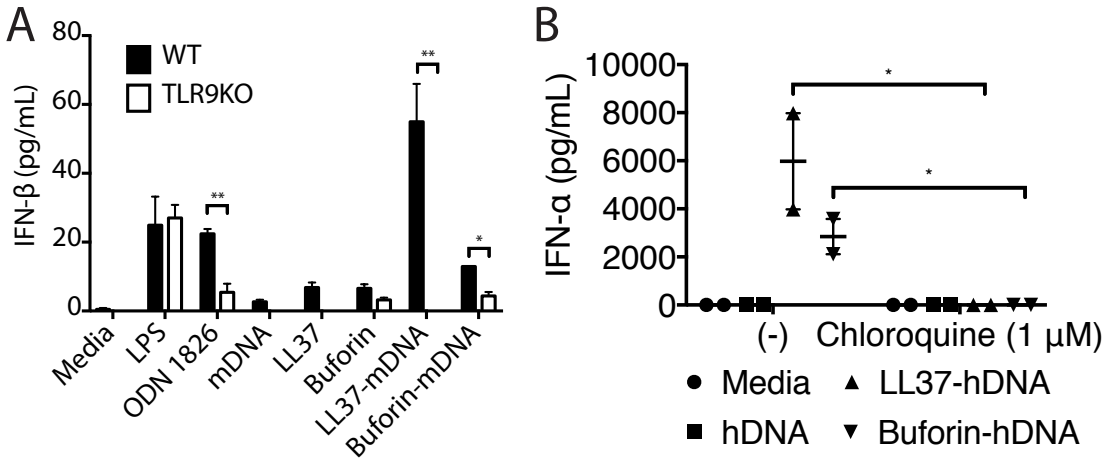


Figure 7.5: **LL37-dsDNA and buforin-dsDNA complexes induce Type 1 Interferon production in a TLR9-dependent manner in mouse macrophages and human pDCs.**
(A) To measure the ability of AMP-dsDNA complexes to activate TLR9, complexes were incubated with wild type (WT) and TLR9-knockout (TLR9KO) murine macrophages. IFN-β production was measured and quantified from supernatants using a Bio-ELISA. Mouse genomic dsDNA was used to generate the complexes. Data shows TLR9-dependent activation of WT macrophages by LL37-dsDNA and buforin-dsDNA complexes, with minimal to no activation with peptides or dsDNA alone. As a positive control, mouse-specific TLR9 agonist ODN 1826 was used. Both WT and TLR9KO cells responded to the TLR4 agonist LPS ($n = 3$, $p^* < 0.05$, $p^{**} < 0.01$). (B) Similarly, AMP-dsDNA complexes were incubated with human pDCs in the absence or presence of chloroquine (1 μM), and IFN-α production was quantified with ELISA. Data shows that pDCs potently release IFN-α in a TLR9-dependent manner in response to LL37-dsDNA and buforin-dsDNA complexes, and this response is attenuated in the presence of chloroquine, which inhibits endosomal acidification and TLR signaling ($n = 2$, $p^* < 0.05$).

properties of melittin are similar to LL37 in that they both are strongly cationic, amphipathic helices with a large hydrophobic face to facilitate self-assembly into a protofibril in the presence of dsDNA, as shown in Figures 1-3. However, closer inspection reveals that the hydrophobic face of melittin ($\sim 180°$) is larger than that of LL37 ($\sim 120°$) with a similar overall charge. From previous work on AMPs, a large hydrophobic face corresponds with increased membrane residence and nonspecific membrane activity, which is consistent with the increased toxicity of melittin to immune cells relative to that of LL37 [59]. We hypothesized that perhaps melittin represents an extreme example of an intrinsic evolutionary tradeoff between the ability of α-helical AMPs to porate membranes and their ability to induce strong TLR9 activation via dsDNA binding.

To probe the relationship between membrane activity, dsDNA binding, and TLR9 activation by α-helical AMPs, we selected three melittin-related peptides to study, AR23, RV23, and MM1, in addition to melittin itself. AR23 and RV23 are 23-amino acid α-helical structural homologues of melittin, with both containing a proline hinge similar to melittin. Both AR23 and RV23 possess reduced hemolytic and cytotoxic activity compared to melittin [516, 517], and have either reduced charge or reduced overall hydrophobicity. MM1 is a melittin point mutant with several modified amino acids, including removal of the proline hinge and reduced overall hydrophobicity. Melittin, AR23, RV23, and MM1 were incubated with PS/PE = 20/80 model membranes and the induced negative Gaussian curvature (NGC) was measured using SAXS (Figure 7.6A, B), which is a well-validated metric of AMP membrane activity [518, 47, 72, 177, 176, 175, 519] and membrane permeation processes in general [268, 78]. In parallel, the peptides were assayed for their ability to bind to dsDNA, and the resulting nanocrystalline structures were solved using SAXS (Figure 7.6C). The TLR9-activating abilities of the AMP-dsDNA complexes were determined via immune stimulation experiments with human pDCs (Figure 7.6D).

Melittin, AR23, and RV23 induced cubic phases rich in NGC in model membranes, evidenced by peaks satisfying the Pn3m q-ratios of $\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{6} : \sqrt{8} : \sqrt{9}$ (Figure 7.6A). In contrast, MM1 did not induce appreciable topological changes in the model membranes compared to controls. Pn3m cubic phases have continuous surfaces with NGC at every

point, which reflects the membrane topology enforced in membrane poration processes. The quantity of NGC induced by the peptides is calculated from the lattice parameters (a) of the Pn3m cubic phases measured, according to the formula $\langle k \rangle = \frac{2\pi\chi}{A_0 a^2}$ ($\chi = -2$ and $A_0 = 1.919$ for Pn3m). Melittin generated the largest amount of NGC $\langle k_{melittin} \rangle = -2.75 \times 10^{-2}$ nm^{-2}, corresponding to a lattice parameter of $a_{melittin} = 15.4$ nm. In contrast, AR23 and RV23 induced reduced quantitative amounts of NGC, with $\langle k_{AR23} \rangle = -1.85 \times 10^{-2}$ nm^{-2} and $\langle k_{RV23} \rangle = -1.18 \times 10^{-2}$ nm^{-2} corresponding to lattice parameters of $a_{AR23} = 18.8$ nm and $a_{RV23} = 23.6$ nm (Figure 7.6B, fits in Figure 7.7). MM1 did not generate NGC in model membranes. Thus, melittin, AR23, RV23, and MM1 are representative examples of α-helical AMPs that span a range of NGC-generating abilities, with melittin being the strongest and MM1 being the weakest.

To determine whether the differences in charge and hydrophobicity of melittin-related AMPs affected their ability to bind to DNA, we measured the structures of AR23-, RV23-, and MM1-DNA complexes with SAXS and compared them to the structure of the melittin-DNA complex (Figure 7.6C, fits in Figure 7.8) (melittin-DNA SAXS data is reproduced from Figure 7.1 for clarity). AR23 and RV23 form square columnar lattices with dsDNA, in a similar manner to melittin and LL37 (Figure 7.1,7.2,7.3). AR23-DNA and RV23-DNA complexes exhibit inter-DNA spacings ($d_{AR23} = 3.78$ nm, $d_{RV23} = 3.87$ nm) similar to melittin-DNA ($d_{melittin} = 3.64$ nm) and slightly larger than LL37-DNA ($d_{LL37} = 3.40$ nm). This indicates that, despite amino acid changes leading to reduced membrane activity, AR23 and RV23 retain the ability to organize DNA into a 4-fold square lattice. In contrast, MM1-DNA forms a columnar lattice with short-ranged order (Figure 7.6C). Like buforin-DNA complexes (Figure 7.4), MM1-DNA complexes exhibit more disorder than melittin, AR23, RV23, and LL37, and have a much larger inter-DNA spacing ($d_{MM1} = 4.19$ nm).

Based on these measured inter-DNA spacings, we would predict that AR23-DNA and RV23-DNA complexes possess the ability to induce TLR9-mediated inflammation in pDCs, while MM1 would not. The above complexes were incubated with human pDCs and resulting IFN-α production was measured (Figure 7.6D). As expected, AR23 and RV23 both induced increased IFN-α production from pDCs compared to both melittin and MM1. Interestingly,

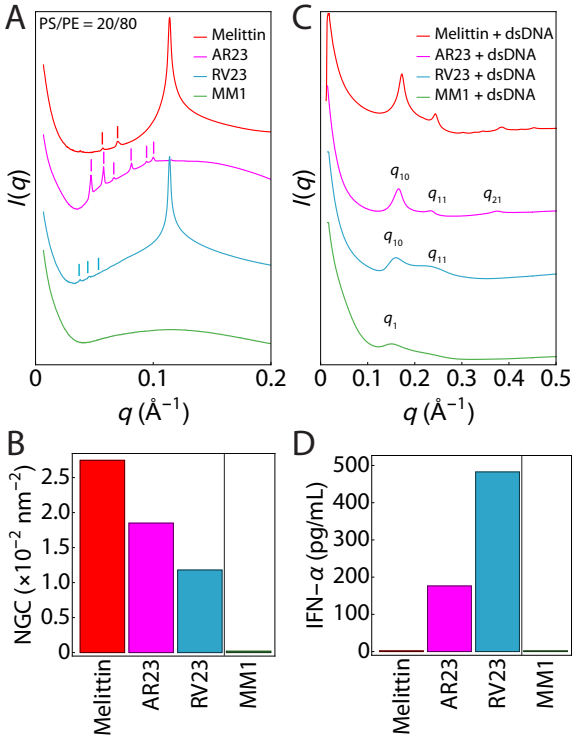


Figure 7.6: **Systematic reduction of the membrane activity of melittin in a series of mutants with conserved dsDNA-binding activity corresponds with a recovery of TLR9-activation potential.**
(A) SAXS spectra from PS/PE = 20/80 model membranes incubated with melittin, AR23, RV23, and MM1 at P/L = 1/2. Melittin, AR23, and RV23 induce Pn3m cubic phases in model membranes, while MM1 does not induce appreciable deformations in the membranes. (B) Compared to melittin, AR23 and RV23 induce decreasing amounts of NGC in membranes, while MM1 does not induce NGC at all. (C) SAXS spectra corresponding to the structures of melittin-, AR23-, RV23-, and MM1-DNA complexes. AR23 and RV23 form square columnar lattices with dsDNA (similar to melittin and LL37), while MM1 forms a columnar lattice with short-ranged order (similar to buforin). MM1-DNA complexes have a larger inter-DNA spacing (4.19 nm) and less crystalline order than melittin-, AR23-, and RV23-DNA complexes (3.64 nm, 3.78 nm, 3.87 nm). (D) Measured TLR9-specific IFN-α production (pg/mL) from human pDCs. Systematic reduction of NGC-inducing ability in melittin-related peptides AR23 and RV23 (B) correlate with a recovery in ability to induce IFN-α from pDCs. MM1 is unable to induce NGC or IFN-α production from pDCs.

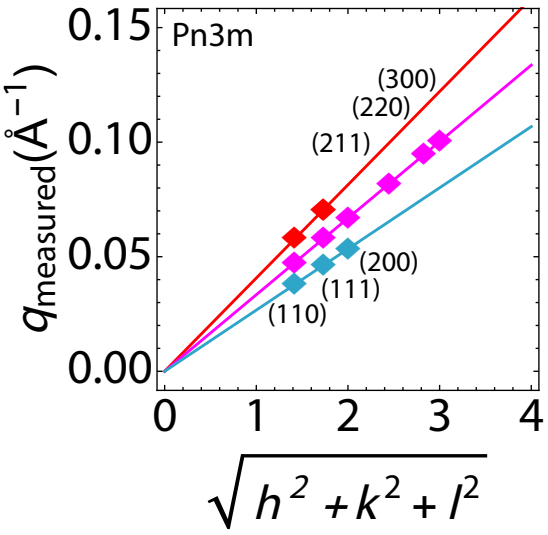


Figure 7.7: **SAXS Lattice Parameter Fits for Membrane Activity of Melittin-related Peptides.**
Pn3m cubic phases are induced by melittin (red), AR23 (magenta), and RV23 (teal) peptides in PS/PE = 20/80 membranes. Observed reflections are labeled with their Millier indices (hkl). Lattice parameters a are extracted from linear fits, and induced NGC is calculated via the formula $\langle k \rangle = \frac{2\pi\chi}{A_0 a^2}$, where χ is the Euler characteristic, A_0 is the surface area per cubic unit cell for each phase, and a is the lattice parameter. The decreasing slope of the fits from melittin to AR23 corresponds to a decrease in NGC.

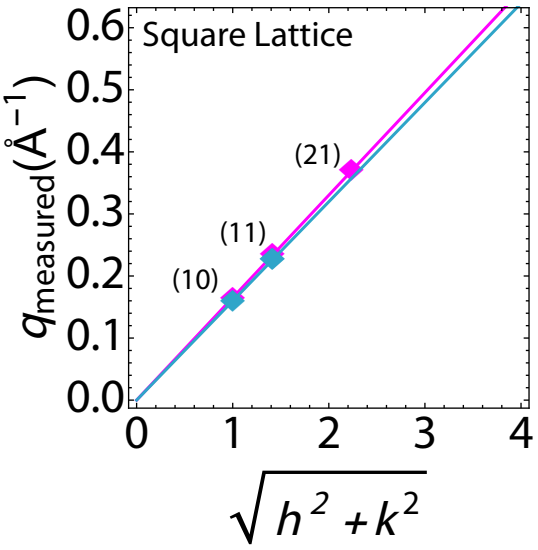


Figure 7.8: **SAXS Lattice Parameter Fits for DNA Complexes with Melittin-related Peptides.**
AR23 (magenta) and RV23 (teal) form square columnar complexes with dsDNA. The observed higher order reflections characteristic of the $1:\sqrt{2}:2$ ratios are labeled. AR23 and RV23 have lattice parameters of 3.78 nm and 3.87 nm respectively, and the fit demonstrates that they bind DNA similarly.

the increase in TLR9-activating ability in AR23 and RV23 relatively to melittin negatively correlates with induced NGC in cell membranes (Figure 7.6B, 7.6D). For peptides that have conserved DNA-binding activity (AR23, RV23), systematic reduction of NGC-inducing ability corresponds to a recovery in ability to induce IFN-α from pDCs. MM1, however, is unable to organize DNA into a square lattice with an inter-DNA spacing in the optimal range, and is also unable to induce NGC. This is consistent with mutational changes in its sequence that lead to dual abolishment of membrane activity and DNA-binding/TLR9-activating ability. Taken together, these findings suggest that the physiochemical requirements for membrane activity and dsDNA-binding/TLR9 activation of α-helical AMPs are overlapping but distinct.

7.4 Conclusions and Prospects

LL37 is a multifunctional α-helical human antimicrobial peptide that can not only kill bacteria via membrane permeation, but also modulates Toll-like receptor signaling in the innate immune system. LL37 forms electrostatic complexes with dsDNA, leading to amplification of TLR9-mediated signaling in pDCs and macrophages, triggering production of type-I IFNs. Other α-helical AMPs have also been reported to modulate TLRs in a similar manner, but the criterion for such immunological activity is unknown. Here, we use X-ray scattering, computer simulations, and immune cell stimulation experiments to deduce some fundamental structural rules for α-helices that are proinflammatory via the TLR9 pathway, based on case studies of 6 prototypical helical AMPs that span a spectrum of charge, charge density, hydrophobic content, and membrane activity.

To activate TLR9, α-helical AMPs must first condense dsDNA into nanocrystalline columnar immunocomplexes. This requires a sufficiently high cationic charge density near the Manning limit. Secondly, they must form complexes with an inter-DNA spacing well-matched with the steric size of TLR9. Computer simulations suggest that such peptides require a sufficiently large hydrophobic face, enabling oligomerization via hydrophobic interactions into a protofibril in the presence of DNA. X-ray measurements show that activating complexes have sufficient crystallinity, and form via 4-fold coordination of DNA by an AMP superhelix into square-columnar complexes. Finally, the α-helical AMP must possess sufficient but not excessive membrane activity, enabling endosomal access but not inducing cytotoxicity in immune cells.

The cationic and amphipathic properties of α-helical AMPs are primarily associated with their ability to permeate cell membranes. Here, we show that these same properties also facilitate their ability to self-assemble in the presence of dsDNA to produce nanocrystalline immune complexes that can activate TLR9 in immune cells. In fact, it is precisely their ability to organize and scaffold dsDNA into clustered immune ligands at the right inter-dsDNA spacing that enables multivalent binding to TLR9 and subsequent immune activation. These findings are consistent with a conceptual framework in which the crystallinity of AMP-

DNA complexes impact immune activation, and have broad implications for the design of multifunctional antimicrobial and immunomodulatory peptides.

7.5 Methods

7.5.1 Preparation of self-assembled peptide-dsDNA complexes

Lyophilized peptides LL37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES), melittin (GIGAVLKVLTTGLPALISWIKRKRQQ), buforin (TRSSRAGLQFPV-GRVHRLLRK), MM1 (GIGAVLKALTTGLGALASAIKRKRQQ), AR23 (AIGSIL-GALAKGLPTLISWIKNR), and RV23 (RIGVLLARLPKLFSLFKLMGKKV) were purchased from Anaspec or Lifetein ($\geq$ 95% purity by HPLC) and dissolved in nuclease-free water (Ambion) to 10 mg/mL. For SAXS experiments, monodispersed Lambda dsDNA (New England BioLabs) or E. coli genomic dsDNA (Affymetrix) was precipitated and resuspended in 100 mM NaCl + 10 mM HEPES (pH 7.4) to 5 mg/mL. For cell experiments, endotoxin-free human (hdsDNA) and mouse (mdsDNA) genomic self-dsDNA were purchased (Biochain) and used without further purification. Peptide-dsDNA complexes were formed by incubating the peptide with dsDNA (1-5 mg/mL for SAXS experiments, 10-20 μg/mL for cell experiments) at specific peptide-to-dsDNA charge ratios (P/dsDNA = 1:4, 1:2, 1:1, 2:1, 4:1). For a particular peptide, the same charge ratios were used in both SAXS and cell experiments.

7.5.2 SAXS experiments with peptide-dsDNA complexes

We measured the phase diagram of the structures of peptide-dsDNA complexes by incubating peptides (10 mg/mL) with dsDNA (5 mg/mL) at specific charge ratios in microcentrifuge tubes. After thorough mixing and centrifugation, precipitated complexes are hermetically sealed in 1.5 mm quartz capillaries (Hilgenberg GmbH, Mark-tubes). SAXS experiments were performed at Stanford Synchrotron Radiation Lightsource (SSRL, Beamline 4-2) using monochromatic X-rays with an energy of 9 keV. A Rayonix MX225-HE detector

(pixel size 73.2 μm) was used to measure the scattered radiation. Independent identical samples were prepared and measured over multiple timepoints to ensure consistency. 2D powder diffraction patterns were integrated using the Nika 1.74 [184] package for Igor Pro 6.37 and FIT2D [185]. SAXS data were analyzed by plotting integrated scattering intensity against the momentum transfer q using Mathematica. Peak positions were measured by fitting diffraction peaks to a Lorentzian. Structures of complexes were solved by calculating ratios between the q-positions of all measured peaks and comparing them with the permitted reflections for known liquid-crystalline phases. The lattice parameter(s) of each phase were calculated by linear regression through points corresponding to measured and theoretical peaks. The lattice parameter d indicates the inter-dsDNA spacing between dsDNA columns. For a square columnar lattice, $q_{hk} = \frac{2\pi}{d}\sqrt{h^2 + k^2}$. For a tetragonal columnar lattice, $q_{hkl} = 2\pi\sqrt{\frac{h^2+k^2}{a^2} + \frac{l^2}{c^2}}$, where c is the out-of-plane lattice parameter along the dsDNA axis. For a columnar lattice with short-ranged order, the inter-dsDNA spacing a is obtained from the first peak position by the formula $d = 2\pi/q_1$.

7.5.3 Domain Size Calculations

For full details of domain size calculations, see section 5.5.2.

7.5.4 Macrophage Stimulation Experiments

To validate the ability of peptide-dsDNA complexes to activate TLR9, we stimulated wild type (WT) and TLR9-knockout (TLR9KO) BL6 mouse macrophages *in vitro* and measured subsequent IFN-β production using a Luciferase Bio-ELISA. Peptide-dsDNA complexes were generated as described above, using mouse genomic self-dsDNA. Complexes were incubated at room temperature prior to stimulation. WT and TLR9KO macrophages were grown to confluency in RPMI + Bovine Growth Serum (BGS) + Penicillin/Streptomycin (Pen/Strep) + L-glutamine (Gibco), washed, and seeded into 96-well microplates with a final concentration of 10^5 cells/well. Complexes were added to WT and TLR9KO cells, diluted to a final volume of 150 μL, and incubated at 37°C for 10 hours. Supernatants were removed and

frozen at −80°C overnight. All stimulations were done in triplicate.

7.5.5 IFN-β Bioassay

IFN-β bioassay was carried out similarly to prior work [520]. Supernatants were assayed for type I IFN production (IFN-β) using ISRE-L929 IFN reporter cells (Bio-ELISA system with IFN-β receptor coupled to a luciferase promoter). ISRE-L929 cells were grown to confluency in RPMI + BGS + Pen/Strep + L-glutamine and plated to a final concentration of 105 cells/well in a 96-well microplate. 50 μL of supernatant was incubated with ISRE-L929 cells for 7 hours at 37°C. An 8-point standard curve with two-fold serial dilutions of known IFN-β concentrations (R&D Systems) was incubated with ISRE-L929 cells at the same time. ISRE-L929 cells are lysed overnight at −80°C in Passive Lysis Buffer (Promega). Luciferase production in ISRE-L929 is quantified using a Luciferase Assay System (Promega) with a luminometer plate reader (Biotek). Luciferase production was converted to IFN-β production using the calibrated IFN-β standard curve. Positive controls include ODN 1826 (TLR9 agonist, Invivogen), LPS (TLR4 agonist, Sigma-Aldrich), and negative controls include peptides and genomic dsDNA alone. Peptide-dsDNA complexes that induce a TLR9-specific response produce a significantly larger amount of IFN-β from WT stimulation compared to TLR9KO stimulation (unpaired Student's t-test).

7.5.6 pDC Stimulation Experiments

Blood buffy coats of healthy donors were obtained from the Interregionale Blutspende SRK, Bern, Switzerland. Peripheral blood mononuclear cells (PBMCs) were obtained by Ficoll centrifugation of blood buffy coats. Plasmacytoid dendritic cells (pDCs) were then isolated by magnetic separation using the Diamond pDC Isolation kit II (Miltenyi). Purified pDCs were cultured in 96-well round-bottom plates at 5x104 / well in 200 μL RPMI 1640 (GIBCO) supplemented with 10% FCS (Atlanta Biologicals) and Pen/Strep in the presence of LL37-hdsDNA or buforin-hdsDNA complexes. LL37-hdsDNA and buforin-hdsDNA complexes were generated by mixing 10 μg of human dsDNA (BioChain) with 2 μg of LL37 or

buforin in 40 μL of nuclease-free water (Ambion) and then diluted into 200 μL of complete medium for cell stimulation (final concentrations: 50 μg/mL hdsDNA and 10 μg/mL LL37 or buforin). In some instances, pDCs were pre-incubated with 1 μM chloroquine diphosphate (Sigma) for 15 min before stimulation. The different supernatants were collected after overnight culture and the levels of type I IFN production were measured by ELISA (human IFN-α pan, Mabtech).

7.5.7 SAXS Experiments with Model Membranes

Peptides were purchased (Anaspec, Lifetein) and synthesized in high purity (> 95% HPLC) using solid-phase synthesis. Liposomes were prepared as previously described elsewhere [77]. Lyophilized DOPS (1,2-dioleoyl-sn-glycero-3-phospho-L-serine (sodium salt)) and DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine) from Avanti Polar Lipids were used. Stock solutions were dissolved in chloroform at 20 mg/mL. Mixtures of these lipids were used as a first-order model for bacterial membranes. DOPS and DOPE were mixed at a 1:4 molar ratio (DOPS/DOPE = 20/80). Chloroform was evaporated under N_2, and the lipids were further dried overnight under vacuum. Dried lipids were resuspended in 100 mM NaCl + 10 mM HEPES at pH 7.4 to a final concentration of 20 mg/mL. Aqueous lipid solutions were incubated overnight at 37°C. Liposomes were prepared by sonication of resuspended lipids until clear. Monodispersity in size was obtained via extrusion through a 0.2 μm filter. Before use, peptides are solubilized in nuclease-free pure water. Liposomes (DOPS/DOPE = 20/80) and peptides were mixed at a peptide-to-lipid charge ratio (P/L) of 1/2 and equilibrated at room temperature overnight. Precipitated peptide-lipid complexes were loaded into 1.5 mm glass quartz capillaries (Mark-tubes, Hilgenberg GmbH) and hermetically sealed. SAXS experiments were conducted at the Stanford Synchrotron Radiation Laboratory (BL 4-2) with monochromatic X-rays of energy 9 keV. Scattered radiation was collected using a Rayonix-MX225-HE detector (pixel size 73.2 μm). The 2D diffraction patterns were integrated using the Nika 1.74 [184] package for Igor Pro 6.37 (Wavemetrics) and FIT2D [185]. For all samples, multiple measurements were taken for consistency. Samples were incubated at 37°C and centrifuged at 6000 RPM for 20 min before measurement. No

changes were observed over multiple experiments and exposures. SAXS data were analyzed by plotting the integrated scattering intensity $I(q)$ vs. q using Mathematica. To determine the phases present in each sample, the measured peak positions were obtained and their ratios were compared to the permitted reflections for different liquid-crystalline lipid phases (e.g. lamellar, inverse hexagonal, cubic). Membrane curvature-generating ability of the peptides were determined by calculating the lattice parameter a for the measured cubic phases. Cubic phases observed in our experiments belonged to the Pn3m space group. Measured q positions for the Bragg peak reflections were fitted to the equation $q_{measured} = \frac{2\pi}{a}\sqrt{h^2+k^2+l^2}$ where (h,k,l) are the Miller indices. The quantity of induced NGC was calculated as the average NGC per unit cell volume using the equation $\langle k \rangle = \frac{2\pi\chi}{A_0 a^2}$, where χ is the Euler characteristic and A_0 is the surface area per cubic unit cell for each phase. Parameter values are $\chi = -2$ and $A_0 = 1.919$ for Pn3m.

7.5.8 Computational Methods and Molecular Simulations

7.5.8.1 Structure factor of a 2D coarse-grained lattice

If a periodic cell contains n coarse-grained particles and the fractional coordinate of the ith particle is (x_i, y_i), then the structure factor of (h,k) is:

$$F(h,k) = \sum_{i=1}^{n} f_i \exp(hx_i + ky_i)$$

where f_i is the diffraction coefficient, which is proportional to electron density. If the electron density of all the particles are equal, then,

$$F(h,k) = f \sum_{i=1}^{n} \exp(hx_i + ky_i)$$

7.5.8.2 Construction of the DNA lattice model

A GC repeat sequence was used for modeling the BDNA double helix. The helix has about 10.5 base pairs per turn and the pitch was 34.5 Å. Each of the 64 double helices were

rotated around its axis on a random angle between 0° and 360°, and placed on an $N \times N$ square column lattice with a 0-2 Å deviation (for melittin) or 0-4 Å (for LL37) deviation.

7.5.8.3 Estimation of the pitch of the LL37 superhelix

Per helical turn, the volume of the square column is $d^2 \times c$ and the weight of the column is $4 \times w$, where d is the periodic length in the square plane, c is the periodic length prediction in the axis direction, and w is the molecular weight of monomer. Thus, the density was $\rho = \frac{d^2 c}{4w}$. We assume that density of the LL37 lattice and the melittin lattice should be similar. So, the LL37 periodic length prediction in the axis direction was

$$c(LL37) = c(Melittin) \times \frac{w(LL37)}{w(Melittin)} \times \left(\frac{d(Melittin)}{d(LL37)} \right)^2$$

7.5.8.4 Modeling the Melittin and LL37 superhelices

First, we built the initial superhelix structure from the monomeric structures. Both melittin and LL37 are amphipathic α-helices. In every 4 residues (about 1 helical turn), there are one or two hydrophobic residues (VILFW) and one or two charged residues (DEKR). PDB structures 2MLT_A [33] and 2K6O [31] were used as the building blocks of the melittin and LL37 superhelices, respectively. Since they both have a hydrophobic and hydrophilic face, a "hydrophobic" line and a "hydrophilic" line were first determined. The "hydrophobic" line denotes the minimum distance to all the hydrophobic side chains, and the "hydrophilic" line denotes the minimum distance to all the charged-residue side chains. Furthermore, these two lines must be parallel. The initial axis of the superhelix was placed in the plane of the two lines and parallel to them, and the distance to the "hydrophilic" line was observed to be 1 nm. The helical twist angle for both melittin and LL37 is 90°, and the helical pitch per monomer was 0.82 nm for melittin and 1.7 nm for LL37 (see above section on "Estimation of the pitch of the LL37 superhelix"). This was a quarter of the periodic length in the axis dirction (c value). 26mer superhelix models were built for both melittin or LL37. Left handed superhelix models exhibited a greater percentage of buried hydrophobic residues

than that of the right-handed models. Thus, we hypothesized that the predominant species was the left-handed superhelix.

After the initial superhelix structure was constructed, we optimized the structure by several iterations of structure relaxation and helical symmetry reconstruction. Energy minimization was first used for structure relaxation. TINKER MINIMIZE program [521] and the AMBER-03 [522] force field with GBSA implicit solvent were used for superhelix energy minimization with a threshold gradient 0.5 kcal/mol/Å. The helical symmetry was reconstructed by averaging the structure of each of the middle 18 monomers, and then repeating the energy minimization again. This iterative process continued until the root mean square deviation (RMSD) of the superhelix model became smaller than 0.5 Å. Then, a 400 ps explicit-solvent NVT molecular dynamics simulation was carried out for further structure relaxation using the AMBER and AMBER-03 force fields. About 1.0 mol/L NaCl was added after charge neutralization. The system was gradually heated up from 50°K to 300°K at 50°K intervals. After the MD simulation, the helical symmetry was reconstructed. These steps were repeated iteratively until the RMSD thresholds were met. After 4 total MD-reconstruction iteration cycles, the final superhelix column (26mer) structure was obtained. This final model was used for the lattice construction.

7.5.8.5 Building the Melittin-DNA and LL37-DNA lattices

The melittin-DNA and LL37-DNA complexes are square columnar lattices (melittin is tetragonal with an implicit square columnar lattice). The nearest column-column distance is the periodic length ($a = b$). 8×8 lattice models were built for melittin, and 3×3 lattice models were built for LL37 in accordance with their domain sizes. The melittin superhelix has a rounded square cross section, while the LL37 superhelix column is near square. Base on the extinction in diffraction, the diagonal of column-square was parallel to the lattice vector and the adjacent two squares showed a corner-to-corner configuration. The superhelix column is directional. In the lattice, the two column directions (N terminal up or C terminal up) were randomly distributed. To simulate the defect in the lattice, three more random parameters

were added. First, the column was rotated around its axis 0-±20° and its diagonal was placed offset to the lattice vector. Second, the columns were randomly deviated (0-2 Å) from the lattice position on the square plane. Third, on the axis direction, the lattice is not perfectly ordered. We modeled this by inducing a 0-$\pm\frac{c}{4}$ deviation in the center of melittin superhelix columns, and a 0-$\pm\frac{c}{2}$ deviation in the center of the LL37 superhelix columns.

7.5.8.6 Diffraction simulation using CRYSOL

The X-ray diffraction was simulated using the software CRYSOL [506] on the all-atom lattice models derived above. Because the side chains of the long-chain hydrophilic residues (DNEQKR) are very flexible, only the backbone and Cβ atoms of these residues contribute to the diffraction simulation study. The CRYSOL diffraction of 50 different 8×8 lattice structures were calculated and averaged for the melittin-DNA and LL37-DNA lattices.

7.5.9 Statistical Analyses

Analysis of immune stimulation data was done in Mathematica or R using pairwise two-tailed t-tests with a significance level of $\alpha = 0.05$.

CHAPTER 8

Conclusions and Outlook

8.1 Summary

Antimicrobial peptides have been studied for over 40 years in the context of innate host defense. Initial studies focused on understanding their microbicidal mechanisms and identifying common structural and physicochemical features essential to this activity [1, 6]. In the past, machine learning methods strived to improve identification and classification of AMPs with the goal of designing potent next-generation antibiotics to tackle drug-resistant infections. In this thesis, we endeavored to do something different. In Chapter 2, we leveraged machine-learning methods to illuminate the physicochemical basis of membrane activity, using α-helical AMPs as a bootstrap dataset. We conducted a Monte-Carlo search of the undiscovered sequence space, coupled with calibrating X-ray measurements of induced membrane curvature, and discovered that our machine learning algorithm can identify α-helical membrane-active peptides in general. In Chapter 3, We employed the machine learning classifier as a search tool to discover hidden membrane activity in new and existing peptide and protein taxonomies, including curvature-generating domains in mitochondrial fission proteins, membrane repair proteins, histones, and neuropeptides, among many others.

In a way, machine learning allowed us to identify molecules that have multiplexed functions in addition to membrane activity. Throughout the latter half of the thesis, we deeply explored the multifunctional nature of AMPs, specifically their nucleic-acid-binding and immune receptor-modulating abilities. Within the last 10 years, dramatic progress has been made in elucidating how AMPs interact with the host innate and adaptive immune systems, and how this synergizes with their direct activity on the pathogen. With the rise of drug-

resistant infections, there is a resurgence of interest in the design and development of AMPs with potent, selective activity and favorable immunomodulatory profiles [523, 524, 525, 526]. In a complementary direction of inquiry, AMPs have also been repeatedly implicated in the dysregulation of inflammation, such as in chronic autoinflammatory diseases like lupus and psoriasis. Therefore, improving AMP-mediated immune signaling can bolster responses to infections [527], while disruption of this behavior could provide a therapeutic opportunity for autoimmune diseases [459]. Because of the diversity in sequence and structure of AMPs, it is difficult to predict immunomodulatory behavior for all AMPs and identify universal themes. In Chapters 4 and 5, we described work that suggests a deterministic mechanism for how a surprisingly broad range of AMPs signals to the immune system. Using an unconventional combination of techniques from physics, chemistry, computational biology, and immune activation experiments, we describe a novel physical basis for AMP-induced immunomodulation via TLRs. We find that AMPs can structurally organize and scaffold immune ligands into spatially periodic nanocrystalline and sometimes liquid crystalline complexes, and that the crystallinity of AMP—nucleic acid complexes can determine the degree of immune amplification in well-defined *in vitro* systems [349, 362, 378, 528]. In Chapter 6, we explore how the structural rules underpinning TLR9 activation can extend to chromatin released during cell death and NETosis. This represents a significant generalization of the central paradigm in immunology. Innate immune receptors can recognize not just pathogen-associated molecular patterns of single ligand molecules, but also recognize nanocrystalline arrangements of AMPs and ligands, ultimately leading to potent immunomodulation that depends sensitively on crystallinity parameters such as the inter-ligand spacing and the number of repeating ligands.

8.2 Conclusions and Outlook

AMPs are known to kill microbes through direct activity, including membrane permeation, disruption of electrochemical gradients, and inhibition of metabolic machinery. However, recent evidence shows that AMPs can also orchestrate host immune responses by

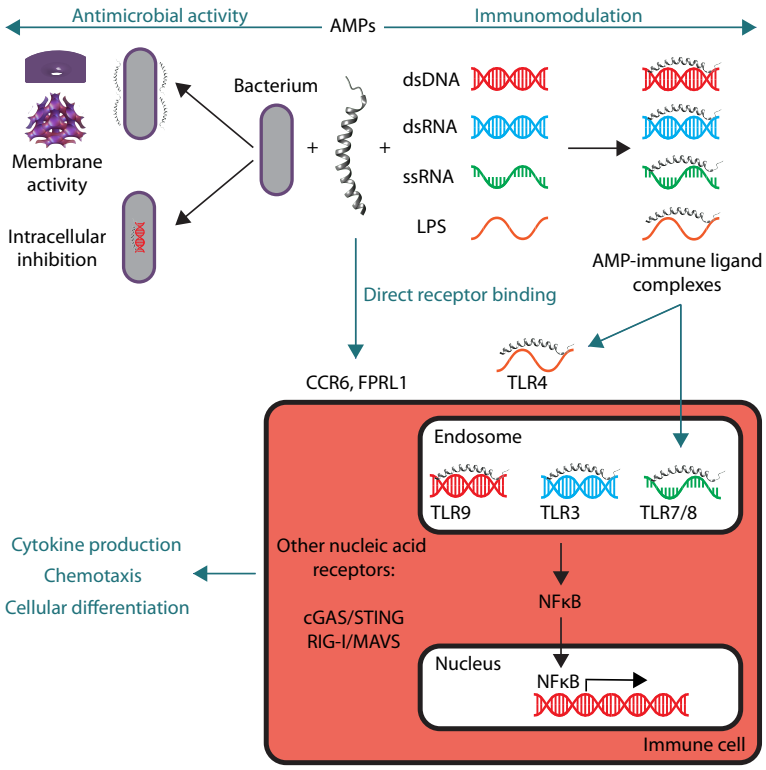


Figure 8.1: **Antimicrobial peptides can modulate immune responses through multiple mechanisms.**
AMPs are multifunctional molecules that can kill microbes by punching holes in their membranes and inhibiting intracellular machinery (upper left). AMPs can also signal to the host immune system (upper right), and exhibit both pro-inflammatory and anti-inflammatory properties. AMPs can interact directly with receptors on immune cells, or form complexes with immune ligands to enable modulation of multiple PRRs like TLRs. (Note that the illustrations of AMP-immune ligand complexes here are purely schematic in nature and do not reflect the actual structures of the nanocomplexes.). AMP-dsDNA, -dsRNA, and -ssRNA complexes can enter endosomes of immune cells and bind to their respective TLRs, triggering activation of signal transduction pathways that lead to modulation of inflammation (lower right). Downstream consequences of AMP immune signaling include activation of transcription factors, cytokine production, immune cell chemotaxis, and cellular differentiation and proliferation, eventually leading to induction of the adaptive immune response (lower left).

communicating with the innate and adaptive immune systems through receptor modulation. AMPs can directly bind to various cell-surface or intracellular receptors on immune cells, triggering chemotaxis, differentiation/maturation, and cytokine production (Figure 8.1).

We also highlighted an underappreciated facet of AMP immunomodulation: AMPs can also organize immune ligands such as dsDNA and dsRNA into nanocrystalline complexes that amplify TLR9 and TLR3 activation (Figure 8.2). Furthermore, the degree of immune amplification can be quantitatively correlated from the crystallinity parameters of AMP-nucleic acid complexes. Indeed, the same physicochemical features of AMPs that enable their antimicrobial and membrane activity (e.g., cationic charge, amphipathicity) also enable strong electrostatic binding to nucleic acids like dsDNA and dsRNA. It will be interesting to see whether these general mechanisms extend beyond TLRs to various cytosolic nucleic acid sensors. AMP involvement in these corollary pathways have been implicated in earlier studies. For example, LL37 can induce IFN-α production from keratinocytes by binding to cytosolic DNA [529]. LL37 drives IFN-β production in epidermal keratinocytes during skin injury via the mitochondrial antiviral-signaling protein (MAVS). LL37 enables recognition of self-dsRNA through MAVS in addition to TLR3 [353]. In pDCs, LL37 transports dsDNA into endosomal compartments, triggering TLR9 activation [330]. However, in non-pDC dendritic cells, LL37 enables cytosolic DNA sensing via STING and TBK1 kinase [530].

In parallel direction of inquiry, it will be informative to critically dissect the optimal structural and physicochemical requirements for AMP antimicrobial and immunomodulatory activity via TLRs, and compare and contrast these competing requirements. Perhaps it is worthwhile to revisit old questions regarding the structural requirements of AMPs in the context of immunomodulatory effects reviewed here. We have known for some time that cationic charge and hydrophobicity are required for AMP activity, and have interpreted that in the perspective of membrane activity. We can ask broader questions now. How does immunomodulatory behavior change with cationic charge and hydrophobicity? Can we predict which AMPs will be more strongly antimicrobial or immunomodulatory and which innate immune receptors will be involved? Can we rationally design AMPs and AMP-like molecules, and tune both antimicrobial and immunomodulatory properties independently?

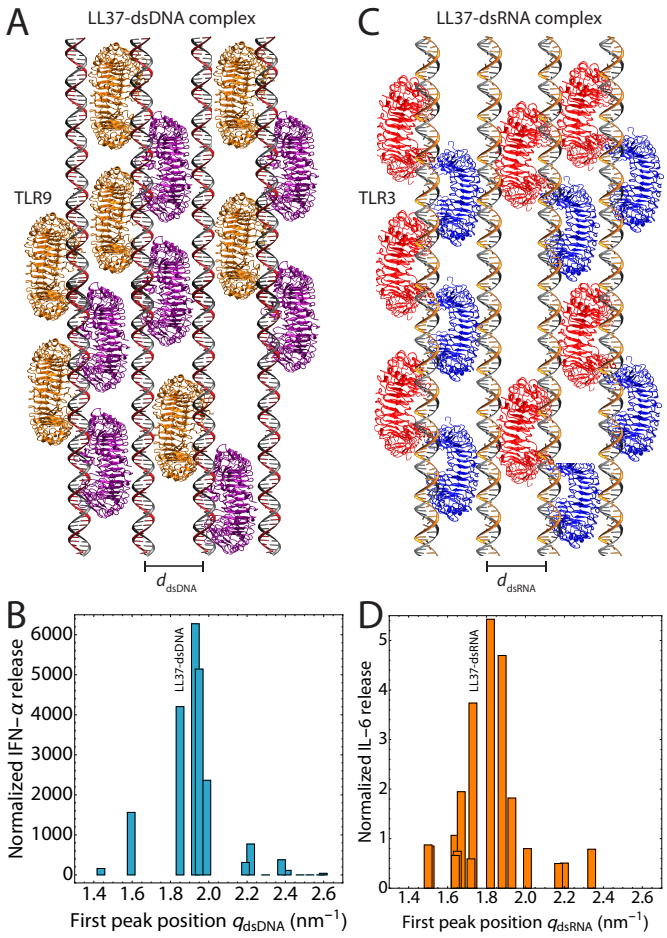


Figure 8.2: **Nanocrystalline ordering of dsDNA and dsRNA by AMPs modulates TLR-mediated inflammation.**
(A) Schematic of nanocrystalline LL37-dsDNA complexes binding to clustered arrays of TLR9 dimers in the endosomal membrane (d_{dsDNA} = 3.40 nm). (B) Complexes with inter-DNA spacings (d_{dsDNA}) well-matched with the steric size of TLR9 leads to drastic amplification of inflammation, due to optimized multivalent intercalative binding of LL37-dsDNA to TLR9. Maximal IFN-α production occurs for complexes with spacings in the range of 3.1-3.5 nm. (C) Schematic of nanocrystalline LL37-dsRNA complexes binding to clustered arrays of TLR3 dimers in the endosomal membrane (d_{dsRNA} = 3.63 nm). (D) Complexes with inter-RNA spacings (d_{dsRNA}) well-matched with the steric size of TLR3 leads to strong amplification of inflammation, while those with poorly matched spacings do not. Maximal IL-6 production occurs for complexes with spacings in the range of 3.3-3.7 nm.

In Chapter 7, we begin to answer some of these broad questions, comparing and contrasting the requirements of membrane activity and DNA binding/TLR-activating ability.

Looking forward, there are a multitude of opportunities to expand on this work. The majority of immunomodulatory AMPs were first characterized as antimicrobial peptides, and then were later found to possess immunomodulatory activity. However, recent work shows that many traditional immune-signaling molecules that exert their functions through specific membrane-bound receptors (such as cytokines) also possess hidden direct antimicrobial activity. In fact, an increasing body of evidence shows that this categorical delineation between antimicrobial molecules and immunomodulatory molecules is becoming progressively blurred [531]. Perhaps the emerging "kinocidin" family of proteins can provide insight into how nature multiplexes antimicrobial and immunomodulatory activity. Kinocidins are cytokines with intrinsic antimicrobial activity [532], and many share core physicochemical and structural similarities with AMPs. For example, various members of the chemokine family of cytokines contain a β-sheet rich "γ-core" motif that is also characteristic of classical defensins [532, 533], and exhibit facial amphiphilicity and net cationic charge [531, 534]. Many of these molecules contain modular components that share structural homology with α-helical and β-sheet AMPs [535, 536]. For example, in a pioneering screen of over 30 human chemokines from these families, 21 were found to possess potent broad-spectrum antimicrobial activity comparable to human defensins, suggesting that chemokines may also directly kill microbes in addition to their chemotactic functions [537]. In another series of studies, the platelet chemokine (C-X-C motif) ligand 4 (CXCL4)/platelet factor 4 (PF-4), its rabbit analogue thrombin-induced platelet microbicidal protein 1 (tPMP-1), and a synthetic congener peptide derived from CXCL4 C-terminal domain was found to have potent antimicrobial properties [538, 539, 536, 540, 541]. Other studies have shown that IL-8 and MCP-1, two important mediators of chemotaxis, also have antimicrobial activity [542]. Strikingly, IL-8, CXCL4/PF4, and MCP-1 were all found to share strong structural homology with human β-defensins, namely the γ-core motif, indicating that a common scaffold has been optimized by nature for dual antimicrobial and immunomodulatory activity. Recently, the cytokines IL-26 and IFN-β were also discovered to have hidden antimicrobial activity

[543, 544, 545, 519]. Other molecules in innate immunity such as complement C3a and the S100 class of endogenous calcium-binding proteins have also exhibited dual antimicrobial and immunomodulatory activities [546, 547, 548, 549]. The development of machine learning tools to detect hidden antimicrobial activity in proteins has enabled the possibility of uniting disparate peptide taxonomies under a common umbrella of multifunctional proteins that exhibit different degrees of antimicrobial activity, receptor-binding activity, and immunomodulatory ability [177, 175, 176, 550].

In our view, it will be worthwhile in future work to work towards a more nuanced understanding of the self-assembly of AMPs and related molecules with traditional immune ligands in the context of inflammation. These studies also question the conventional demarcation between "pro-inflammatory" and "anti-inflammatory" AMPs and cytokines. Context matters especially to pleiotropic molecules, whether it is the tissue-specific expression of immunomodulatory mediators or activation of multiple receptors in specific cells in different tissues. For example, an immune signaling molecule may induce an anti-inflammatory response via binding to its cell-surface receptor, but could simultaneously induce a pro-inflammatory response by binding to nucleic acids and signaling through TLRs. Instead of measuring a binary response, we propose models to quantitatively predict and measure the degree of inflammation on a continuous scale as a function of the biophysical and structural properties of AMPs, immune ligands, and their targets.

Finally, the ability of AMPs to modulate the innate system via mechanisms described here is a double-edged sword: In the normal context of infection, AMPs bolster the immune response by enabling immune cell signaling, recruitment, and proliferation. However, overexpression of AMPs has clearly been associated with exacerbation of many additional chronic inflammatory diseases like systemic sclerosis, vasculitis, atherosclerosis, and others. Further structure-function and mechanistic studies will be required to fully map out the immunomodulation landscapes of AMPs, their pathophysiological role in disease, and strategies to regulate their immunomodulatory abilities.

REFERENCES

[1] Zasloff M (2002) Antimicrobial peptides of multicellular organisms. *Nature* 415(6870):389–395.

[2] Shai Y (1999) Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1462(1-2):55–70.

[3] Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology* 3(3):238–250.

[4] Hancock REW, Lehrer R (1998) Cationic peptides: a new source of antibiotics. *Trends in biotechnology* 16(2):82–88.

[5] Hancock REW, Sahl HG (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology* 24(12):1551–1557.

[6] Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. *Pharmacological reviews* 55(1):27–55.

[7] Wang Z, Wang G (2004) APD: the Antimicrobial Peptide Database. *Nucleic Acids Research* 32(Database issue):D590–2.

[8] Wang G, Li X, Wang Z (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Research* 37(Database):D933–D937.

[9] Wang G, Li X, Wang Z (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research* 44(D1):D1087–93.

[10] Skarnes RC, Watson DW (1957) Antimicrobial factors of normal tissues and fluids. *Bacteriological reviews* 21(4):273–294.

[11] Zasloff M (1987) Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proceedings of the National Academy of Sciences* 84(15):5449–5453.

[12] Oren Z, Shai Y (1998) Mode of action of linear amphipathic α-helical antimicrobial peptides. *Biopolymers* 47(6):451–463.

[13] Ganz T (2003) Defensins: antimicrobial peptides of innate immunity. *Nature reviews. Immunology* 3(9):710–720.

[14] Selsted ME et al. (1992) Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *Journal of Biological Chemistry* 267(7):4292–4295.

[15] Agerberth B et al. (1991) Amino acid sequence of PR-39. Isolation from pig intestine of a new member of the family of proline-arginine-rich antibacterial peptides. *Eur J Biochem* 202(3):849–854.

[16] Hancock REW, Rozek A (2002) Role of membranes in the activities of antimicrobial cationic peptides. *FEMS microbiology letters* 206(2):143–149.

[17] Yang L, Harroun TA, Weiss TM, Ding L, Huang HW (2001) Barrel-Stave Model or Toroidal Model? A Case Study on Melittin Pores. *Biophysical Journal* 81(3):1475–1485.

[18] Ehrenstein G, Lecar H (1977) Electrically gated ionic channels in lipid bilayers. *Quarterly reviews of biophysics* 10(1):1–34.

[19] Lee MT, Chen FY, Huang HW (2004) Energetics of pore formation induced by membrane active peptides. *Biochemistry* 43(12):3590–3599.

[20] Spaar A, Münster C, Salditt T (2004) Conformation of peptides in lipid membranes studied by x-ray grazing incidence scattering. *Biophysical Journal* 87(1):396–407.

[21] He K, Ludtke SJ, Worcester DL, Huang HW (1996) Neutron scattering in the plane of membranes: structure of alamethicin pores. *Biophysical Journal* 70(6):2659–2666.

[22] Bechinger B et al. (1991) Orientations of amphipathic helical peptides in membrane bilayers determined by solid-state NMR spectroscopy. *Journal of Biomolecular NMR* 1(2):167–173.

[23] Pouny Y, Rapaport D, Mor A, Nicolas P, Shai Y (1992) Interaction of antimicrobial dermaseptin and its fluorescently labeled analogs with phospholipid membranes. *Biochemistry* 31(49):12416–12423.

[24] Yamaguchi S et al. (2001) Orientation and dynamics of an antimicrobial peptide in the lipid bilayer by solid-state NMR spectroscopy. *Biophysical Journal* 81(4):2203–2214.

[25] Bechinger B (1999) The structure, dynamics and orientation of antimicrobial peptides in membranes by multidimensional solid-state NMR spectroscopy. *Biochimica et biophysica acta* 1462(1-2):157–183.

[26] Ladokhin AS, White SH (2001) 'Detergent-like' permeabilization of anionic lipid vesicles by melittin. *Biochimica et biophysica acta* 1514(2):253–260.

[27] Matsuzaki K, Murase O, Fujii N, Miyajima K (1996) An antimicrobial peptide, magainin 2, induced rapid flip-flop of phospholipids coupled with pore formation and peptide translocation. *Biochemistry* 35(35):11361–11368.

[28] Huang HW, Chen FY, Lee MT (2004) Molecular Mechanism of Peptide-Induced Pores in Membranes. *Physical Review Letters* 92(19):198304.

[29] Hallock KJ, Lee DK, Ramamoorthy A (2003) MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain. *Biophysical Journal* 84(5):3052–3060.

[30] Yamaguchi S, Hong T, Waring A, Lehrer RI, Hong M (2002) Solid-state NMR investigations of peptide-lipid interaction and orientation of a beta-sheet antimicrobial peptide, protegrin. *Biochemistry* 41(31):9852–9862.

[31] Wang G (2008) Structures of human host defense cathelicidin LL-37 and its smallest antimicrobial peptide KR-12 in lipid micelles. *Journal of Biological Chemistry* 283(47):32637–32643.

[32] Gesell J, Zasloff M, Opella SJ (1997) Two-dimensional 1H NMR experiments show that the 23-residue magainin antibiotic peptide is an alpha-helix in dodecylphosphocholine micelles, sodium dodecylsulfate micelles, and trifluoroethanol/water solution. *Journal of Biomolecular NMR* 9(2):127–135.

[33] Terwilliger TC, Eisenberg D (1982) The structure of melittin. I. Structure determination and partial refinement. *Journal of Biological Chemistry* 257(11):6010–6015.

[34] Epand RM, Epand RF (2011) Bacterial membrane lipids in the action of antimicrobial agents. *Journal of peptide science : an official publication of the European Peptide Society* 17(5):298–305.

[35] Zachowski A (1993) Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. *The Biochemical journal* 294 (Pt 1):1–14.

[36] van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology* 9(2):112–124.

[37] Zimmerberg J, Kozlov MM (2006) How proteins produce cellular membrane curvature. *Nature reviews. Molecular cell biology* 7(1):9–19.

[38] Epand RF, Savage PB, Epand RM (2007) Bacterial lipid composition and the antimicrobial efficacy of cationic steroid compounds (Ceragenins). *Biochimica et biophysica acta* 1768(10):2500–2509.

[39] Siegel DP, Kozlov MM (2004) The Gaussian Curvature Elastic Modulus of N-Monomethylated Dioleoylphosphatidylethanolamine: Relevance to Membrane Fusion and Lipid Phase Behavior. *Biophysical Journal* 87(1):366–374.

[40] Som A, Yang L, Wong GCL, Tew GN (2009) Divalent metal ion triggered activity of a synthetic antimicrobial in cardiolipin membranes. *Journal of the American Chemical Society* 131(42):15102–15103.

[41] Yang L et al. (2008) Mechanism of a prototypical synthetic membrane-active antimicrobial: Efficient hole-punching via interaction with negative intrinsic curvature lipids. *Proceedings of the National Academy of Sciences of the United States of America* 105(52):20595–20600.

[42] Yang L et al. (2007) Synthetic antimicrobial oligomers induce a composition-dependent topological transition in membranes. *Journal of the American Chemical Society* 129(40):12141–12147.

[43] Sankhagowit S, Lee EY, Wong GCL, Malmstadt N (2016) Oxidation of Membrane Curvature-Regulating Phosphatidylethanolamine Lipid Results in Formation of Bilayer and Cubic Structures. *Langmuir* 32(10):2450–2457.

[44] Kreyszig E (1991) *Differential geometry.* (New York : Dover Publications).

[45] Koller D, Lohner K (2014) The role of spontaneous lipid curvature in the interaction of interfacially active peptides with membranes. *Biochimica et biophysica acta* 1838(9):2250–2259.

[46] Deserno M (2015) Fluid lipid membranes: from differential geometry to curvature stresses. *Chemistry and physics of lipids* 185:11–45.

[47] Schmidt NW, Wong GCL (2013) Antimicrobial peptides and induced membrane curvature: geometry, coordination chemistry, and molecular engineering. *Current Opinion in Solid State & Materials Science* 17(4):151–163.

[48] Bechinger B (2009) Rationalizing the membrane interactions of cationic amphipathic antimicrobial peptides by their molecular shape. *Current Opinion in Colloid & Interface Science* 14(5):349–355.

[49] Baumgart T, Capraro BR, Zhu C, Das SL (2011) Thermodynamics and Mechanics of Membrane Curvature Generation and Sensing by Proteins and Lipids. *Annual review of physical chemistry* 62(1):483–506.

[50] Antonny B (2011) Mechanisms of membrane curvature sensing. *Annual review of biochemistry* 80(1):101–123.

[51] Stachowiak JC, Hayden CC, Sasaki DY (2010) Steric confinement of proteins on lipid membranes can drive curvature and tubulation. *Proceedings of the National Academy of Sciences of the United States of America* 107(17):7781–7786.

[52] Stachowiak JC et al. (2012) Membrane bending by protein–protein crowding. *Nature Cell Biology* 14(9):944–949.

[53] Koltover I, Salditt T, Radler JO, Safinya CR (1998) An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery. *Science* 281(5373):78–81.

[54] Raviv U et al. (2005) Cationic liposome-microtubule complexes: pathways to the formation of two-state lipid-protein nanotubes with open or closed ends. *Proceedings of the National Academy of Sciences of the United States of America* 102(32):11167–11172.

[55] Tytler EM et al. (1993) Reciprocal effects of apolipoprotein and lytic peptide analogs on membranes. Cross-sectional molecular shapes of amphipathic alpha helixes control membrane stability. *Journal of Biological Chemistry* 268(29):22112–22118.

[56] Epand RM, Shai Y, Segrest JP, Anantharamiah GM (1995) Mechanisms for the modulation of membrane bilayer properties by amphipathic helical peptides. *Peptide Science* 37(5):319–338.

[57] Segrest JP, De Loof H, Dohlman JG, Brouillette CG, Anantharamaiah GM (1990) Amphipathic helix motif: Classes and properties. *Proteins: Structure, Function, and Bioinformatics* 8(2):103–117.

[58] Zemel A, Ben-Shaul A, May S (2008) Modulation of the Spontaneous Curvature and Bending Rigidity of Lipid Membranes by Interfacially Adsorbed Amphipathic Peptides. *The Journal of Physical Chemistry B* 112(23):6988–6996.

[59] Wieprecht T et al. (1997) Influence of the angle subtended by the positively charged helix face on the membrane activity of amphipathic, antibacterial peptides. *Biochemistry* 36(42):12869–12880.

[60] Drin G, Antonny B (2009) Amphipathic helices and membrane curvature. *FEBS Letters* 584(9):1840–1847.

[61] Chen FY, Lee MT, Huang HW (2003) Evidence for membrane thinning effect as the mechanism for peptide-induced pore formation. *Biophysical Journal* 84(6):3751–3758.

[62] Tristram-Nagle S et al. (2010) HIV Fusion Peptide Penetrates, Disorders, and Softens T-Cell Membrane Mimics. *Journal of molecular biology* 402(1):139–153.

[63] Szleifer I, Kramer D, Ben-Shaul A, Roux D, Gelbart WM (1988) Curvature Elasticity of Pure and Mixed Surfactant Films. *Physical Review Letters* 60(19):1966–1969.

[64] Dathe M, Wieprecht T (1999) Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1462(1-2):71–87.

[65] Yang L, Weiss TM, Lehrer RI, Huang HW (2000) Crystallization of Antimicrobial Pores in Membranes: Magainin and Protegrin. *Biophysical Journal* 79(4):2002–2009.

[66] Ludtke SJ et al. (1996) Membrane pores induced by magainin. *Biochemistry* 35(43):13723–13728.

[67] Saiman L et al. (2001) Cathelicidin peptides inhibit multiply antibiotic-resistant pathogens from patients with cystic fibrosis. *Antimicrob Agents Chemother* 45(10):2838–2844.

[68] Kalfa VC et al. (2001) Congeners of SMAP29 Kill Ovine Pathogens and Induce Ultrastructural Damage in Bacterial Cells. *Antimicrob Agents Chemother* 45(11):3256–3261.

[69] Yu Y, Vroman JA, Bae SC, Granick S (2010) Vesicle budding induced by a pore-forming peptide. *Journal of the American Chemical Society* 132(1):195–201.

[70] Falagas ME, Kasiakou SK (2005) Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. *Clinical Infectious Diseases* 40(9):1333–1341.

[71] Gidalevitz D et al. (2003) Interaction of antimicrobial peptide protegrin with biomembranes. *Proceedings of the National Academy of Sciences* 100(11):6302–6307.

[72] Schmidt NW et al. (2011) Criterion for amino acid composition of defensins and antimicrobial peptides based on geometry of membrane destabilization. *Journal of the American Chemical Society* 133(17):6720–6727.

[73] Schmidt NW et al. (2012) Arginine in α-defensins: differential effects on bactericidal activity correspond to geometry of membrane curvature generation and peptide-lipid phase behavior. *J Biol Chem* 287(26):21866–21872.

[74] Schmidt NW, Mishra A, Lai GH, Wong GCL (2010) Arginine-rich cell-penetrating peptides. *FEBS Letters* 584(9):1806–1813.

[75] Lee MW et al. (2014) Two interdependent mechanisms of antimicrobial activity allow for efficient killing in nylon-3-based polymeric mimics of innate immunity peptides. *Biochimica et biophysica acta* 1838(9):2269–2279.

[76] Xiong M et al. (2015) Helical antimicrobial polypeptides with radial amphiphilicity. *Proceedings of the National Academy of Sciences of the United States of America* 112(43):13155–13160.

[77] Schmidt NW, Mishra A, Wang J, DeGrado WF, Wong GCL (2013) Influenza Virus A M2 Protein Generates Negative Gaussian Membrane Curvature Necessary for Budding and Scission. *Journal of the American Chemical Society* 135(37):13710–13719.

[78] Yao H, Lee MW, Waring AJ, Wong GCL, Hong M (2015) Viral fusion protein transmembrane domain adopts β-strand structure to facilitate membrane topological changes for virus-cell fusion. *Proceedings of the National Academy of Sciences of the United States of America* 112(35):10926–10931.

[79] Koynova R, Tenchov B (2013) Transitions between lamellar and nonlamellar phases in membrane lipids and their physiological roles. *OA Biochemistry* 1(1).

[80] Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning*, Data Mining, Inference, and Prediction, Second Edition. (Springer Science & Business Media, New York, NY).

[81] Turing AM (1950) Computing machinery and intelligence. *Mind* 59(236):433–460.

[82] Graves A et al. (2009) A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(5):855–868.

[83] Ganesan N, Venkatesh K, Rama MA (2010) Application of neural networks in diagnosing cancer disease using demographic data. *International Journal of Computer Theory and Engineering* 1:81–97.

[84] Betechuoh BL, Marwala T, Tettey T (2006) Autoencoder networks for HIV classification. *Current Science* 91(11):1467–1473.

[85] Agarwal M, Jain N, Kumar MM (2010) Face recognition using eigen faces and artificial neural network. *International Journal of Computer Theory and Engineering* 2:624–629.

[86] Rothwell AC, Jagger LD, Dennis WR, Clarke DR, Networks Associates Technology, Inc. (2004) Intelligent SPAM detection system using an updateable neural analysis engine (US Patent Office).

[87] Balabin RM, Lomakina EI (2009) Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies. *The Journal of Chemical Physics* 131(7):074104.

[88] Silver D et al. (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–489.

[89] Fjell CD et al. (2009) Identification of novel antibacterial peptides by chemoinformatics and machine learning. *Journal of medicinal chemistry* 52(7):2006–2015.

[90] Mitchell JBO (2014) Machine learning methods in chemoinformatics. *Wiley interdisciplinary reviews. Computational molecular science* 4(5):468–481.

[91] Hilpert K, Fjell CD, Cherkasov A (2008) Short linear cationic antimicrobial peptides: screening, optimizing, and prediction. *Methods in molecular biology (Clifton, N.J.)* 494(Chapter 8):127–159.

[92] Lata S, Sharma BK, Raghava G (2007) Analysis and prediction of antibacterial peptides. *BMC Bioinformatics* 8(1):1.

[93] Cherkasov A et al. (2008) Use of Artificial Intelligence in the Design of Small Peptide Antibiotics Effective against a Broad Spectrum of Highly Antibiotic-Resistant Superbugs. *ACS Chemical Biology* 4(1):65–74.

[94] Fjell CD et al. (2008) Identification of novel host defense peptides and the absence of α-defensins in the bovine genome. *Proteins: Structure, Function, and Bioinformatics* 73(2):420–430.

[95] Wang P et al. (2011) Prediction of Antimicrobial Peptides Based on Sequence Alignment and Feature Selection Methods. *PLoS ONE* 6(4):e18476.

[96] Torrent M, Andreu D, Nogués VM, Boix E (2011) Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS ONE* 6(2):e16968.

[97] Maccari G et al. (2013) Antimicrobial Peptides Design by Evolutionary Multiobjective Optimization. *PLOS Comput Biol* 9(9):e1003212.

[98] Xiao X, Wang P, Lin WZ, Jia JH, Chou KC (2013) iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Analytical biochemistry* 436(2):168–177.

[99] Giguère S et al. (2015) Machine learning assisted design of highly active peptides for drug discovery. *PLOS Comput Biol* 11(4):e1004074.

[100] Schneider P et al. (2017) Hybrid Network Model for "Deep Learning" of Chemical Data: Application to Antimicrobial Peptides. *Molecular Informatics* 36(1-2):1600011.

[101] Matsuzaki K (1999) Why and how are peptide–lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1462(1-2):1–10.

[102] Matsuzaki K et al. (1998) Relationship of membrane curvature to the formation of pores by magainin 2. *Biochemistry* 37(34):11856–11863.

[103] Huang HW (2000) Action of Antimicrobial Peptides: Two-State Model†. *Biochemistry* 39(29):8347–8352.

[104] Brötz H, Bierbaum G, Leopold K, Reynolds PE, Sahl HG (1998) The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II. *Antimicrob Agents Chemother* 42(1):154–160.

[105] Park CB, Kim HS, Kim SC (1998) Mechanism of Action of the Antimicrobial Peptide Buforin II: Buforin II Kills Microorganisms by Penetrating the Cell Membrane and Inhibiting Cellular Functions. *Biochemical and Biophysical Research Communications* 244(1):253–257.

[106] Yonezawa A, Kuwahara J, Fujii N, Sugiura Y (2002) Binding of tachyplesin I to DNA revealed by footprinting analysis: significant contribution of secondary structure to DNA binding and implication for biological action. *Biochemistry* 31(11):2998–3004.

[107] Patrzykat A, Friedrich CL, Zhang L, Mendoza V, Hancock REW (2002) Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in Escherichia coli. *Antimicrob Agents Chemother* 46(3):605–614.

[108] Otvos L et al. (2000) Interaction between Heat Shock Proteins and Antimicrobial Peptides †. *Biochemistry* 39(46):14150–14159.

[109] Bowdish DME, Davidson DJ, Hancock REW (2005) A Re-evaluation of the Role of Host Defence Peptides in Mammalian Immunity. *Current protein & peptide ...* 6(1):35–51.

[110] Gilliet M, Lande R (2008) Antimicrobial peptides and self-DNA in autoimmune skin inflammation. *Current opinion in immunology* 20(4):401–407.

[111] Porto WF, Pires ÁS, Franco OL (2012) CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides. *PLoS ONE* 7(12):e51444.

[112] Fjell CD, Hiss JA, Hancock REW, Schneider G (2012) Designing antimicrobial peptides: form follows function. *Nature Reviews Drug Discovery* 11(1):37–51.

[113] Yee LC, Wei YC (2012) Current Modeling Methods Used in QSAR/QSPR in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany), pp. 1–31.

[114] Bi J, Bennett K, Embrechts M, Breneman C, Song M (2003) Dimensionality reduction via sparse support vector machines. *The Journal of Machine Learning Research* 3:1229–1243.

[115] Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning* 20(3):273–297.

[116] Pedregosa F et al. (2011) Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research* 12:2825–2830.

[117] Aizerman A, Braverman EM, Rozoner LI (1964) Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control* 25:821–837.

[118] Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers in *Proceedings of the fifth annual workshop on Computational learning theory*. (ACM Press, New York, New York, USA), pp. 144–152.

[119] Engel T (2006) Basic Overview of Chemoinformatics†. *Journal of Chemical Information and Modeling* 46(6):2267–2277.

[120] Schneider G, Baringhaus KH (2008) *Molecular Design*, Concepts and Applications. (John Wiley & Sons).

[121] Arora JS (2011) *Introduction to Optimum Design*. (Academic Press).

[122] Shoval O et al. (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 336(6085):1157–1160.

[123] Li ZR et al. (2006) PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Research* 34(Web Server issue):W32–7.

[124] Schmidt NW et al. (2012) Molecular basis for nanoscopic membrane curvature generation from quantum mechanical models and synthetic transporter sequences. *Journal of the American Chemical Society* 134(46):19207–19216.

[125] Gilks WR (2005) *Markov Chain Monte Carlo*. (John Wiley & Sons, Ltd, Chichester, UK).

[126] Gilks WR, Richardson S, Spiegelhalter D (1995) *Markov Chain Monte Carlo in Practice*. (CRC Press).

[127] Geyer CJ (1992) Practical markov chain monte carlo. *Statistical Science*.

[128] McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404–405.

[129] Zhang H et al. (2011) Critical assessment of high-throughput standalone methods for secondary structure prediction. *Briefings in Bioinformatics* 12(6):672–688.

[130] Montgomerie S, Sundararaj S, Gallin WJ, Wishart DS (2006) Improving the accuracy of protein secondary structure prediction using structural alignment. *BMC Bioinformatics* 7(1):1.

[131] Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23(2):187–199.

[132] McInturff JE et al. (2005) Granulysin-derived peptides demonstrate antimicrobial and anti-inflammatory effects against Propionibacterium acnes. *Journal of Investigative Dermatology* 125(2):256–263.

[133] Eisenberg D, Weiss RM, Terwilliger TC, Wilcox W (1982) Hydrophobic moments and protein structure. *Faraday Symposia of the Chemical Society* 17:109.

[134] El Karim IA, Linden GJ, Orr DF, Lundy FT (2008) Antimicrobial activity of neuropeptides against a range of micro-organisms from skin, oral, respiratory and gastrointestinal tract sites. *Journal of neuroimmunology* 200(1-2):11–16.

[135] Hansen CJ, Burnell KK, Brogden KA (2006) Antimicrobial activity of Substance P and Neuropeptide Y against laboratory strains of bacteria and oral microorganisms. *Journal of neuroimmunology* 177(1-2):215–218.

[136] Brogden KA, Guthmiller JM, Salzet M, Zasloff M (2005) The nervous system and innate immunity: the neuropeptide connection. *Nature Immunology* 6(6):558–564.

[137] Maisetta G et al. (2013) pH-dependent disruption of Escherichia coli ATCC 25922 and model membranes by the human antimicrobial peptides hepcidin 20 and 25. *FEBS Journal* 280(12):2842–2854.

[138] Singh M, Mukhopadhyay K (2014) Alpha-Melanocyte Stimulating Hormone: An Emerging Anti-Inflammatory Antimicrobial Peptide. *BioMed research international* 2014(6):1–10.

[139] Friedman R, Pellarin R, Caflisch A (2009) Amyloid aggregation on lipid bilayers and its impact on membrane permeability. *Journal of molecular biology* 387(2):407–415.

[140] Caillon L, Killian JA, Lequin O, Khemtémourian L (2013) Biophysical investigation of the membrane-disrupting mechanism of the antimicrobial and amyloid-like peptide dermaseptin S9. *PLoS ONE* 8(10):e75528.

[141] von Heijne G (1986) Towards a comparative anatomy of N-terminal topogenic protein sequences. *Journal of molecular biology* 189(1):239–242.

[142] Taboureau O (2010) Methods for building quantitative structure-activity relationship (QSAR) descriptors and predictive models for computer-aided design of antimicrobial peptides. *Methods in molecular biology (Clifton, N.J.)* 618:77–86.

[143] Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* 40(Database issue):D370–6.

[144] Kozma D, Simon I, Tusnády GE (2012) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Research* 41(D1):gks1169–D529.

[145] Tusnády GE, Dosztányi Z, Simon I (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* 20(17):2964–2972.

[146] Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S (2009) CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Research* 38(Database):gkp1021–D780.

[147] Fernandes FC, Rigden DJ, Franco OL (2012) Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. *Peptide Science* 98(4):280–287.

[148] Golbraikh A, Tropsha A (2002) Beware of q2! *Journal of molecular graphics & modelling* 20(4):269–276.

[149] Chou KC (2009) Pseudo Amino Acid Composition and its Applications in Bioinformatics, Proteomics and System Biology. *Current Proteomics* 6(4):262–274.

[150] Porto WF, Fernandes FC, Franco OL (2010) An SVM Model Based on Physicochemical Properties to Predict Antimicrobial Activity from Protein Sequences with Cysteine Knot Motifs in *Advances in Bioinformatics and Computational Biology*. (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 59–62.

[151] Frecer V, Ho B, Ding JL (2004) De novo design of potent antimicrobial peptides. *Antimicrob Agents Chemother* 48(9):3349–3357.

[152] Mauri A, Ballabio D, Consonni V, Manganaro A, Todeschini R (2008) Peptides multivariate characterisation using a molecular descriptor based approach. *Match-Communications in Mathematical and in Computer Chemistry* 60(2):671–690.

[153] Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3:1157–1182.

[154] Kittler J (1986) Feature selection and extraction. *Handbook of Pattern Recognition and Image Processing*.

[155] Cao DS, Xu QS, Liang YZ (2013) propy: a tool to generate various modes of Chou's PseAAC. *Bioinformatics* 29(7):960–962.

[156] Fauchere JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. *International journal of peptide and protein research* 32(4):269–278.

[157] Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* 157(1):105–132.

[158] Bhaskaran R, Ponnuswamy PK (1988) Positional flexibilities of amino acid residues in globular proteins. *International journal of peptide and protein research* 32(4):241–255.

[159] Charton M, Charton BI (1982) The structural dependence of amino acid hydrophobicity parameters. *Journal of theoretical biology* 99(4):629–644.

[160] Radzicka A, Wolfenden R (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* 27(5):1664–1670.

[161] Bigelow CC (1967) On the average hydrophobicity of proteins and the relation between it and protein structure. *Journal of theoretical biology* 16(2):187–211.

[162] Charton M (1981) Protein folding and the genetic code: An alternative quantitative model. *Journal of theoretical biology* 91(1):115–123.

[163] Dayhoff MO, Schwartz RM (1978) Chapter 22: A model of evolutionary change in proteins, in 'in Atlas of Protein Sequence and Structure'.

[164] Wold S et al. (2011) Principal property values for six non-natural amino acids and their application to a structure–activity relationship for oxytocin peptide analogues. *Canadian Journal of Chemistry* 65(8):1814–1820.

[165] Zimmerman JM, Eliezer N, Simha R (1968) The characterization of amino acid sequences in proteins by statistical methods. *Journal of theoretical biology* 21(2):170–201.

[166] von Heijne G, Blomberg C (1979) Trans-membrane translocation of proteins. *Eur J Biochem.*

[167] Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Research* 31(13):3692–3697.

[168] Dubchak I, Muchnik I, Mayor C, Dralyuk I, Kim SH (1999) Recognition of a protein fold in the context of the SCOP classification. *Proteins: Structure, Function, and Bioinformatics* 35(4):401–407.

[169] Dubchak I, Muchnik I, Holbrook SR, Kim SH (1995) Prediction of protein folding class using global description of amino acid sequence. *Proceedings of the National Academy of Sciences* 92(19):8700–8704.

[170] Schneider G, Wrede P (1994) The rational design of amino acid sequences by artificial neural networks and simulated molecular evolution: de novo design of an idealized leader peptidase cleavage site. *Biophysical Journal* 66(2 Pt 1):335–344.

[171] Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185(4154):862–864.

[172] Chou KC (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins: Structure, Function, and Bioinformatics* 43(3):246–255.

[173] Tanford C (1962) Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. *Journal of the American Chemical Society* 84(22):4240–4247.

[174] Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences* 78(6):3824–3828.

[175] Lee EY, Wong GCL, Ferguson AL (2017) Machine learning-enabled discovery and design of membrane-active peptides. *Bioorganic & medicinal chemistry* p. DOI: 10.1016/j.bmc.2017.07.012.

[176] Lee EY, Lee MW, Fulan BM, Ferguson AL, Wong GCL (2017) What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface focus* 7(6):20160153.

[177] Lee EY, Fulan BM, Wong GCL, Ferguson AL (2016) Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proceedings of the National Academy of Sciences of the United States of America* 113(48):13588–13593.

[178] Breneman C, Bennett K, Embrechts M, Cramer S (2002) *Descriptor generation, selection and model building in quantitative structure-property analysis.* (Experimental Design for Combinatorial and High Throughput Materials Development).

[179] Breiman L (1996) Bagging predictors. *Machine learning* 24(2):123–140.

[180] Oliphant TE (2007) Python for Scientific Computing. *Computing in Science & Engineering* 9(3):10–20.

[181] Powell MJD (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal* 7(2):155–162.

[182] Press WH (1992) *Numerical Recipes in C*, The Art of Scientific Computing. (Cambridge University Press).

[183] Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et biophysica acta* 405(2):442–451.

[184] Ilavsky J (2012) Nika: software for two-dimensional data reduction. *Journal of Applied Crystallography* 45(2):324–328.

[185] Hammersley AP (1997) FIT2D: an introduction and overview. *European Synchrotron Radiation Facility Internal Report: ESRF97HA02T*.

[186] Youle RJ, van der Bliek AM (2012) Mitochondrial fission, fusion, and stress. *Science* 337(6098):1062–1065.

[187] Otera H, Ishihara N, Mihara K (2013) New insights into the function and regulation of mitochondrial fission. *Biochimica et biophysica acta* 1833(5):1256–1268.

[188] Chang DTW, Honick AS, Reynolds IJ (2006) Mitochondrial trafficking to synapses in cultured primary cortical neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26(26):7035–7045.

[189] Saxton WM, Hollenbeck PJ (2012) The axonal transport of mitochondria. *Journal of cell science* 125(Pt 9):2095–2104.

[190] Li Z, Okamoto KI, Hayashi Y, Sheng M (2004) The Importance of Dendritic Mitochondria in the Morphogenesis and Plasticity of Spines and Synapses. *Cell* 119(6):873–887.

[191] Shaw JM, Nunnari J (2002) Mitochondrial dynamics and division in budding yeast. *Trends in cell biology* 12(4):178–184.

[192] Westermann B (2010) Mitochondrial fusion and fission in cell life and death. *Nature reviews. Molecular cell biology* 11(12):872–884.

[193] Sesaki H, Jensen RE (1999) Division versus fusion: Dnm1p and Fzo1p antagonistically regulate mitochondrial shape. *Journal of Cell Biology* 147(4):699–706.

[194] Bleazard W et al. (1999) The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. *Nature Cell Biology* 1(5):298–304.

[195] Smirnova E, Shurland DL, Ryazantsev SN, van der Bliek AM (1998) A human dynamin-related protein controls the distribution of mitochondria. *Journal of Cell Biology* 143(2):351–358.

[196] Otsuga D et al. (1998) The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. *The Journal of Cell Biology* 143(2):333–349.

[197] Chan DC (2006) Mitochondrial fusion and fission in mammals. *Annual review of cell and developmental biology* 22(1):79–99.

[198] Itoh K, Nakamura K, Iijima M, Sesaki H (2013) Mitochondrial dynamics in neurodegeneration. *Trends in cell biology* 23(2):64–71.

[199] Knott AB, Perkins G, Schwarzenbacher R, Bossy-Wetzel E (2008) Mitochondrial fragmentation in neurodegeneration. *Nature reviews. Neuroscience* 9(7):505–518.

[200] Mozdy AD, McCaffery JM, Shaw JM (2000) Dnm1p GTPase-mediated mitochondrial fission is a multi-step process requiring the novel integral membrane component Fis1p. *The Journal of Cell Biology* 151(2):367–379.

[201] James DI, Parone PA, Mattenberger Y, Martinou JC (2003) hFis1, a novel component of the mammalian mitochondrial fission machinery. *Journal of Biological Chemistry* 278(38):36373–36379.

[202] Smirnova E, Griparic L, Shurland DL, van der Bliek AM (2001) Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Molecular Biology of the Cell* 12(8):2245–2256.

[203] Yoon Y, Krueger EW, Oswald BJ, McNiven MA (2003) The Mitochondrial Protein hFis1 Regulates Mitochondrial Fission in Mammalian Cells through an Interaction with the Dynamin-Like Protein DLP1. *Molecular and cellular biology* 23(15):5409–5420.

[204] Stojanovski D, Koutsopoulos OS, Okamoto K, Ryan MT (2004) Levels of human Fis1 at the mitochondrial outer membrane regulate mitochondrial morphology. *Journal of cell science* 117(Pt 7):1201–1210.

[205] Tieu Q, Nunnari J (2000) Mdv1p is a WD repeat protein that interacts with the dynamin-related GTPase, Dnm1p, to trigger mitochondrial division. *Journal of Cell Biology* 151(2):353–365.

[206] Labrousse AM, Zappaterra MD, Rube DA, van der Bliek AM (1999) C-elegans dynamin-related protein DRP-1 controls severing of the mitochondrial outer membrane. *Molecular cell* 4(5):815–826.

[207] Fukushima NH, Brisch E, Keegan BR, Bleazard W, Shaw JM (2001) The GTPase effector domain sequence of the Dnm1p GTPase regulates self-assembly and controls a rate-limiting step in mitochondrial fission. *Molecular Biology of the Cell* 12(9):2756–2766.

[208] Sesaki H, Adachi Y, Kageyama Y, Itoh K, Iijima M (2014) In vivo functions of Drp1: lessons learned from yeast genetics and mouse knockouts. *Biochimica et biophysica acta* 1842(8):1179–1185.

[209] Ingerman E et al. (2005) Dnm1 forms spirals that are structurally tailored to fit mitochondria. *The Journal of Cell Biology* 170(7):1021–1027.

[210] Lackner LL, Horner JS, Nunnari J (2009) Mechanistic Analysis of a Dynamin Effector. *Science* 325(5942):874–877.

[211] Bhar D, Karren MA, Babst M, Shaw JM (2006) Dimeric Dnm1-G385D interacts with Mdv1 on mitochondria and can be stimulated to assemble into fission complexes containing Mdv1 and Fis1. *Journal of Biological Chemistry* 281(25):17312–17320.

[212] Mears JA et al. (2011) Conformational changes in Dnm1 support a contractile mechanism for mitochondrial fission. *Nature Structural & Molecular Biology* 18(1):20–26.

[213] Daumke O, Praefcke GJK (2016) Invited review: Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Peptide Science* 105(8):580–593.

[214] Stepanyants N et al. (2015) Cardiolipin's propensity for phase transition and its reorganization by dynamin-related protein 1 form a basis for mitochondrial membrane fission. *Molecular Biology of the Cell* 26(17):3104–3116.

[215] Frohman MA (2015) Role of mitochondrial lipids in guiding fission and fusion. *Journal of molecular medicine (Berlin, Germany)* 93(3):263–269.

[216] Ugarte-Uribe B, Müller HM, Otsuki M, Nickel W, García-Sáez AJ (2014) Dynamin-related Protein 1 (Drp1) Promotes Structural Intermediates of Membrane Division. *Journal of Biological Chemistry* 289(44):30645–30656.

[217] Morlot S, Roux A (2013) Mechanics of dynamin-mediated membrane fission. *Annual Review of Biophysics* 42:629–649.

[218] Shnyrova AV et al. (2013) Geometric catalysis of membrane fission driven by flexible dynamin rings. *Science* 339(6126):1433–1436.

[219] Montessuit S et al. (2010) Membrane Remodeling Induced by the Dynamin-Related Protein Drp1 Stimulates Bax Oligomerization. *Cell* 142(6):889–901.

[220] Macdonald PJ et al. (2014) A dimeric equilibrium intermediate nucleates Drp1 reassembly on mitochondrial membranes for fission. *Molecular Biology of the Cell* 25(12):1905–1915.

[221] Heymann JAW, Hinshaw JE (2009) Dynamins at a glance. *Journal of cell science* 122(Pt 19):3427–3431.

[222] Francy CA, Clinton RW, Fröhlich C, Murphy C, Mears JA (2017) Cryo-EM Studies of Drp1 Reveal Cardiolipin Interactions that Activate the Helical Oligomer. *Scientific Reports* 7(1):R283.

[223] Snead WT et al. (2017) Membrane fission by protein crowding. *Proceedings of the National Academy of Sciences of the United States of America* 114(16):E3258–E3267.

[224] Gelbart WM, Ben-Shaul A, Roux DD (1994) *Micelles, membranes, microemulsions, and monolayers*. (New York : Springer-Verlag).

[225] Gómez-Llobregat J, Elías-Wolff F, Lindén M (2016) Anisotropic Membrane Curvature Sensing by Amphipathic Peptides. *Biophysical Journal* 110(1):197–204.

[226] Lear JD, DeGrado WF (1987) Membrane binding and conformational properties of peptides representing the NH2 terminus of influenza HA-2. *Journal of Biological Chemistry* 262(14):6500–6505.

[227] Wenzel M et al. (2014) Small cationic antimicrobial peptides delocalize peripheral membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* 111(14):E1409–18.

[228] Sato H, Feix JB (2006) Peptide-membrane interactions and mechanisms of membrane destruction by amphipathic alpha-helical antimicrobial peptides. *Biochimica et biophysica acta* 1758(9):1245–1256.

[229] Makovitzki A, Avrahami D, Shai Y (2006) Ultrashort antibacterial and antifungal lipopeptides. *Proceedings of the National Academy of Sciences* 103(43):15997–16002.

[230] Lorizate M, Huarte N, Sáez-Cirión A, Nieva JL (2008) Interfacial pre-transmembrane domains in viral proteins promoting membrane fusion and fission. *Biochimica et biophysica acta* 1778(7-8):1624–1639.

[231] Epand RM (2003) Fusion peptides and the mechanism of viral fusion. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1614(1):116–121.

[232] Hu K et al. (2013) A critical evaluation of random copolymer mimesis of homogeneous antimicrobial peptides. *Macromolecules* 46(5):1908–1915.

[233] Mishra A et al. (2011) Translocation of HIV TAT peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. *Proceedings of the National Academy of Sciences* 108(41):16883–16888.

[234] Mishra A, Gordon VD, Yang L, Coridan R, Wong GCL (2008) HIV TAT forms pores in membranes by inducing saddle-splay curvature: potential role of bidentate hydrogen bonding. *Angewandte Chemie (International ed. in English)* 47(16):2986–2989.

[235] Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637.

[236] Fröhlich C et al. (2013) Structural insights into oligomerization and mitochondrial remodelling of dynamin 1-like protein. *The EMBO journal* 32(9):1280–1292.

[237] Daum G (1985) Lipids of Mitochondria. *Biochimica et biophysica acta* 822(1):1–42.

[238] Horvath SE, Daum G (2013) Lipids of mitochondria. *Progress in lipid research* 52(4):590–614.

[239] Ardail D et al. (1990) Mitochondrial contact sites. Lipid composition and dynamics. *Journal of Biological Chemistry* 265(31):18797–18802.

[240] Koppenol-Raab M et al. (2016) A Targeted Mutation Identified through pKa Measurements Indicates a Postrecruitment Role for Fis1 in Yeast Mitochondrial Fission. *J Biol Chem* 291(39):20329–20344.

[241] Chang CR et al. (2010) A lethal de novo mutation in the middle domain of the dynamin-related GTPase Drp1 impairs higher order assembly and mitochondrial division. *J Biol Chem* 285(42):32494–32503.

[242] Francy CA, Alvarez FJD, Zhou L, Ramachandran R, Mears JA (2015) The mechanoenzymatic core of dynamin-related protein 1 comprises the minimal machinery required for membrane constriction. *J Biol Chem* 290(18):11692–11703.

[243] Shearman GC, Ces O, Templer RH, Seddon JM (2006) Inverse lyotropic phases of lipids and membrane curvature. *Journal of Physics: Condensed Matter* 18(28):S1105–24.

[244] Naylor K et al. (2006) Mdv1 interacts with assembled dnm1 to promote mitochondrial division. *Journal of Biological Chemistry* 281(4):2177–2183.

[245] Wells RC, Picton LK, Williams SCP, Tan FJ, Hill RB (2007) Direct binding of the dynamin-like GTPase, Dnm1, to mitochondrial dynamics protein Fis1 is negatively regulated by the Fis1 N-terminal arm. *Journal of Biological Chemistry* 282(46):33769–33775.

[246] Cerveny KL, Jensen RE (2003) The WD-repeats of Net2p interact with Dnm1p and Fis1p to regulate division of mitochondria. *Molecular Biology of the Cell* 14(10):4126–4139.

[247] Koirala S et al. (2013) Interchangeable adaptors regulate mitochondrial dynamin assembly for membrane scission. *Proceedings of the National Academy of Sciences* 110(15):E1342–E1351.

[248] Gandre-Babbe S, van der Bliek AM (2008) The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells. *Molecular Biology of the Cell* 19(6):2402–2412.

[249] Otera H et al. (2010) Mff is an essential factor for mitochondrial recruitment of Drp1 during mitochondrial fission in mammalian cells. *The Journal of Cell Biology* 191(6):1141–1158.

[250] Suzuki M, Jeong SY, Karbowski M, Youle RJ, Tjandra N (2003) The Solution Structure of Human Mitochondria Fission Protein Fis1 Reveals a Novel TPR-like Helix Bundle. *Journal of molecular biology* 334(3):445–458.

[251] Tooley JE et al. (2011) The 1.75 Å resolution structure of fission protein Fis1 from Saccharomyces cerevisiae reveals elusive interactions of the autoinhibitory domain. *Acta crystallographica. Section F, Structural biology and crystallization communications* 67(Pt 11):1310–1315.

[252] Dohm JA, Lee SJ, Hardwick JM, Hill RB, Gittis AG (2003) Cytosolic domain of the human mitochondrial fission protein fis1 adopts a TPR fold. *Proteins: Structure, Function, and Bioinformatics* 54(1):153–156.

[253] Ferguson AL et al. (2013) Translating HIV Sequences into Quantitative Fitness Landscapes Predicts Viral Vulnerabilities for Rational Immunogen Design. *Immunity* 38(3):606–617.

[254] Romero PA, Krause A, Arnold FH (2013) Navigating the protein fitness landscape with Gaussian processes. *Proceedings of the National Academy of Sciences of the United States of America* 110(3):E193–201.

[255] DeVay RM et al. (2009) Coassembly of Mgm1 isoforms requires cardiolipin and mediates mitochondrial inner membrane fusion. *Journal of Cell Biology* 186(6):793–803.

[256] Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35(6):2769–2794.

[257] Reshef DN et al. (2011) Detecting novel associations in large data sets. *Science* 334(6062):1518–1524.

[258] Ostrowski SG, Van Bell CT, Winograd N, Ewing AG (2004) Mass spectrometric imaging of highly curved membranes during Tetrahymena mating. *Science* 305(5680):71–73.

[259] Praefcke GJK, McMahon HT (2004) The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature reviews. Molecular cell biology* 5(2):133–147.

[260] Webb B, Sali A (2014) Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics* 47:5.6.1–32.

[261] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792–1797.

[262] Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5(1):113.

[263] Pethica R, Barker G, Kovacs T, Gough J (2010) TreeVector: scalable, interactive, phylogenetic trees for the web. *PLoS ONE* 5(1):e8934.

[264] Albanese D et al. (2013) Minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics* 29(3):407–408.

[265] Lee DY et al. (2009) Histone H4 is a major component of the antimicrobial action of human sebocytes. *The Journal of investigative dermatology* 129(10):2489–2496.

[266] Kawasaki H, Iwamuro S (2008) Potential roles of histones in host defense as antimicrobial agents. *Infectious disorders drug targets* 8(3):195–205.

[267] Park CB, Yi KS, Matsuzaki K, Kim MS, Kim SC (2000) Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: the proline hinge is responsible for the cell-penetrating ability of buforin II. *Proceedings of the National Academy of Sciences* 97(15):8245–8250.

[268] Lee MW et al. (2017) Molecular Motor Dnm1 Synergistically Induces Membrane Curvature To Facilitate Mitochondrial Fission. *ACS central science* 3(11):1156–1167.

[269] Murphy K (2014) *Janeway's Immunobiology.* (Garland Pub).

[270] Platanias LC (2005) Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nature Reviews Immunology* 5(5):375–386.

[271] Uematsu S, Akira S (2008) Toll-Like receptors (TLRs) and their ligands. *Handbook of experimental pharmacology* 183(183):1–20.

[272] Blasius AL, Beutler B (2010) Intracellular Toll-like Receptors. *Immunity* 32(3):305–315.

[273] Takeuchi O et al. (2002) Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology* 169(1):10–14.

[274] Takeuchi O et al. (2001) Discrimination of bacterial lipoproteins by Toll-like receptor 6. *International immunology* 13(7):933–940.

[275] Iwasaki A, Medzhitov R (2004) Toll-like receptor control of the adaptive immune responses. *Nature Immunology* 5(10):987–995.

[276] O'Neill LAJ, Golenbock DT, Bowie AG (2013) The history of Toll-like receptors - redefining innate immunity. *Nature reviews. Immunology* 13(6):453–460.

[277] Akira S, Hemmi H (2003) Recognition of pathogen-associated molecular patterns by TLR family. *Immunology letters* 85(2):85–95.

[278] Carpenter S, O'Neill LAJ (2007) How important are Toll-like receptors for antimicrobial responses? *Cellular microbiology* 9(8):1891–1901.

[279] Heil F et al. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* 303(5663):1526–1529.

[280] Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C (2004) Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* 303(5663):1529–1531.

[281] Lund JM et al. (2004) Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the National Academy of Sciences* 101(15):5598–5603.

[282] Hemmi H et al. (2000) A Toll-like receptor recognizes bacterial DNA. *Nature* 408(6813):740–745.

[283] Lund J, Sato A, Akira S, Medzhitov R, Iwasaki A (2003) Toll-like receptor 9-mediated recognition of Herpes simplex virus-2 by plasmacytoid dendritic cells. *Journal of Experimental Medicine* 198(3):513–520.

[284] Krug A et al. (2004) TLR9-dependent recognition of MCMV by IPC and DC generates coordinated cytokine responses that activate antiviral NK cell function. *Immunity* 21(1):107–119.

[285] Hornung V (2014) SnapShot: nucleic acid immune sensors, part 1. *Immunity* 41(5):868–868.e1.

[286] Hornung V (2014) SnapShot: Nucleic acid immune sensors, part 2. *Immunity* 41(6):1066–1066.e1.

[287] Barton GM, Kagan JC, Medzhitov R (2006) Intracellular localization of Toll-like receptor 9 prevents recognition of self DNA but facilitates access to viral DNA. *Nature Immunology* 7(1):49–56.

[288] Heil F et al. (2003) The Toll-like receptor 7 (TLR7)-specific stimulus loxoribine uncovers a strong relationship within the TLR7, 8 and 9 subfamily. *European journal of immunology* 33(11):2987–2997.

[289] Lee J et al. (2003) Molecular basis for the immunostimulatory activity of guanine nucleoside analogs: activation of Toll-like receptor 7. *Proceedings of the National Academy of Sciences* 100(11):6646–6651.

[290] Matsumoto M et al. (2003) Subcellular Localization of Toll-Like Receptor 3 in Human Dendritic Cells. *Journal of Immunology* 171(6):3154–3162.

[291] Marshak-Rothstein A et al. (2004) Comparison of CpG s-ODNs, chromatin immune complexes, and dsDNA fragment immune complexes in the TLR9-dependent activation of rheumatoid factor B cells. *Journal of endotoxin research* 10(4):247–251.

[292] Leadbetter EA et al. (2002) Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature* 416(6881):603–607.

[293] Crowl JT, Gray EE, Pestal K, Volkman HE, Stetson DB (2017) Intracellular Nucleic Acid Detection in Autoimmunity. *Annual review of immunology* 35(1):313–336.

[294] Theofilopoulos AN, Kono DH, Beutler B, Baccala R (2011) Intracellular nucleic acid sensors and autoimmunity. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* 31(12):867–886.

[295] Jin MS, Lee JO (2008) Structures of the toll-like receptor family and its ligand complexes. *Immunity* 29(2):182–191.

[296] Jin MS, Lee JO (2008) Structures of TLR-ligand complexes. *Current opinion in immunology* 20(4):414–419.

[297] Kang JY et al. (2009) Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* 31(6):873–884.

[298] Liu L et al. (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320(5874):379–381.

[299] Choe J, Kelker MS, Wilson IA (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309(5734):581–585.

[300] Yoon Si et al. (2012) Structural basis of TLR5-flagellin recognition and signaling. *Science* 335(6070):859–864.

[301] Botos I, Segal DM, Davies DR (2011) The Structural Biology of Toll-like Receptors. *Structure* 19(4):447–459.

[302] Kang JY, Lee JO (2011) Structural biology of the Toll-like receptor family. *Annual review of biochemistry* 80(1):917–941.

[303] Ohto U et al. (2015) Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9. *Nature* 520(7549):702–705.

[304] Bowdish DME, Davidson DJ, Hancock REW (2006) Immunomodulatory properties of defensins and cathelicidins. *Antimicrobial Peptides and Human Disease* 306:27–66.

[305] Bowdish DME, Davidson DJ, Scott MG, Hancock REW (2005) Immunomodulatory activities of small host defense peptides. *Antimicrob Agents Chemother* 49(5):1727–1732.

[306] Mookherjee N et al. (2006) Modulation of the TLR-mediated inflammatory response by the endogenous human host defense peptide LL-37. *Journal of Immunology* 176(4):2455–2464.

[307] Ruan Y et al. (2013) Antimicrobial peptide LL-37 attenuates LTA induced inflammatory effect in macrophages. *International immunopharmacology* 15(3):575–580.

[308] Scott MG, Davidson DJ, Gold MR, Bowdish D, Hancock REW (2002) The human antimicrobial peptide LL-37 is a multifunctional modulator of innate immune responses. *Journal of Immunology* 169(7):3883–3891.

[309] Kurosaka K, Chen Q, Yarovinsky F, Oppenheim JJ, Yang D (2005) Mouse cathelin-related antimicrobial peptide chemoattracts leukocytes using formyl peptide receptor-like 1/mouse formyl peptide receptor-like 2 as the receptor and acts as an immune adjuvant. *Journal of Immunology* 174(10):6257–6265.

[310] Yang D et al. (1999) Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* 286(5439):525–528.

[311] Van Wetering S et al. (1997) Effect of defensins on interleukin-8 synthesis in airway epithelial cells. *The American journal of physiology* 272(5 Pt 1):L888–96.

[312] Territo MC, Ganz T, Selsted ME, Lehrer R (1989) Monocyte-chemotactic activity of defensins from human neutrophils. *Journal of Clinical Investigation* 84(6):2017–2020.

[313] Lillard JW, Boyaka PN, Chertov O, Oppenheim JJ, McGhee JR (1999) Mechanisms for induction of acquired host immunity by neutrophil peptide defensins. *Proceedings of the National Academy of Sciences* 96(2):651–656.

[314] Biragyn A et al. (2002) Toll-like receptor 4-dependent activation of dendritic cells by beta-defensin 2. *Science* 298(5595):1025–1029.

[315] Tongaonkar P et al. (2015) Rhesus macaque θ-defensin RTD-1 inhibits proinflammatory cytokine secretion and gene expression by inhibiting the activation of NF-κB and MAPK pathways. *Journal of Leukocyte Biology* 98(6):1061–1070.

[316] Schaal JB et al. (2012) Rhesus macaque theta defensins suppress inflammatory cytokines and enhance survival in mouse models of bacteremic sepsis. *PLoS ONE* 7(12):e51337.

[317] Stenger S et al. (1998) An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* 282(5386):121–125.

[318] Veldhuizen EJA et al. (2014) Antimicrobial and immunomodulatory activities of PR-39 derived peptides. *PLoS ONE* 9(4):e95939.

[319] Som A et al. (2012) Identification of synthetic host defense peptide mimics that exert dual antimicrobial and anti-inflammatory activities. *Clinical and vaccine immunology : CVI* 19(11):1784–1791.

[320] Thaker HD et al. (2012) Synthetic mimics of antimicrobial peptides with immunomodulatory responses. *Journal of the American Chemical Society* 134(27):11088–11091.

[321] Silva ON et al. (2016) An anti-infective synthetic peptide with dual antimicrobial and immunomodulatory activities. *Scientific Reports* 6(1):35465.

[322] Aderem A, Ulevitch RJ (2000) Toll-like receptors in the induction of the innate immune response. *Nature* 406(6797):782–787.

[323] Birchler T et al. (2001) Human Toll-like receptor 2 mediates induction of the antimicrobial peptide human beta-defensin 2 in response to bacterial lipoprotein. *European journal of immunology* 31(11):3131–3137.

[324] Liu PT et al. (2006) Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* 311(5768):1770–1773.

[325] Tauszig S, Jouanguy E, Hoffmann JA, Imler JL (2000) Toll-related receptors and the control of antimicrobial peptide expression in Drosophila. *Proceedings of the National Academy of Sciences* 97(19):10520–10525.

[326] Davidson DJ et al. (2004) The cationic antimicrobial peptide LL-37 modulates dendritic cell differentiation and dendritic cell-induced T cell polarization. *Journal of Immunology* 172(2):1146–1156.

[327] Zhang XJ et al. (2017) B Cell Functions Can Be Modulated by Antimicrobial Peptides in Rainbow Trout Oncorhynchus mykiss: Novel Insights into the Innate Nature of B Cells in Fish. *Frontiers in Immunology* 8(5):388.

[328] Agerberth B et al. (2000) The human antimicrobial and chemotactic peptides LL-37 and alpha-defensins are expressed by specific lymphocyte and monocyte populations. *Blood* 96(9):3086–3093.

[329] Braff MH, Di Nardo A, Gallo RL (2005) Keratinocytes store the antimicrobial peptide cathelicidin in lamellar bodies. *The Journal of investigative dermatology* 124(2):394–400.

[330] Lande R et al. (2007) Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide. *Nature* 449(7162):564–569.

[331] Lande R et al. (2011) Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide complexes in systemic lupus erythematosus. *Science Translational Medicine* 3(73):73ra19–73ra19.

[332] Lande R et al. (2014) The antimicrobial peptide LL37 is a T-cell autoantigen in psoriasis. *Nature communications* 5:5621.

[333] Morizane S, Yamasaki K, Mühleisen B (2012) Cathelicidin antimicrobial peptide LL-37 in psoriasis enables keratinocyte reactivity against TLR9 ligands. *Journal of Investigative Dermatology* 132(1):135–143.

[334] Radic M, Marion TN (2013) Neutrophil extracellular chromatin traps connect innate immune response to autoimmunity. *Seminars in immunopathology* 35(4):465–480.

[335] Kaplan MJ, Radic M (2012) Neutrophil extracellular traps: double-edged swords of innate immunity. *Journal of Immunology* 189(6):2689–2695.

[336] Hurtado P, Peh CA (2010) LL-37 promotes rapid sensing of CpG oligodeoxynucleotides by B lymphocytes and plasmacytoid dendritic cells. *Journal of immunology (Baltimore, Md. : 1950)* 184(3):1425–1435.

[337] Nakagawa Y, Gallo RL (2015) Endogenous intracellular cathelicidin enhances TLR9 activation in dendritic cells and macrophages. *Journal of Immunology* 194(3):1274–1284.

[338] Lande R et al. (2015) Cationic antimicrobial peptides in psoriatic skin cooperate to break innate tolerance to self-DNA. *European journal of immunology* 45(1):203–213.

[339] Gowda NM, Wu X, Gowda DC (2011) The Nucleosome (Histone-DNA Complex) Is the TLR9-Specific Immunostimulatory Component of Plasmodium falciparum That Activates DCs. *PLoS ONE* 6(6):e20398–14.

[340] Tian J et al. (2007) Toll-like receptor 9-dependent activation by DNA-containing immune complexes is mediated by HMGB1 and RAGE. *Nature Immunology* 8(5):487–496.

[341] Yanai H, Ban T, Taniguchi T (2011) Essential role of high-mobility group box proteins in nucleic acid-mediated innate immune responses. *Journal of internal medicine* 270(4):301–308.

[342] Yanai H et al. (2009) HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* 462(7269):99–103.

[343] Urbonaviciute V et al. (2008) Induction of inflammatory and immune responses by HMGB1-nucleosome complexes: implications for the pathogenesis of SLE. *The Journal of experimental medicine* 205(13):3007–3018.

[344] Kagan BL et al. (2011) Antimicrobial Properties of Amyloid Peptides. *Molecular Pharmaceutics* 9(4):708–717.

[345] Soscia SJ et al. (2010) The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide. *PLoS ONE* 5(3):e9505.

[346] Stolzenberg E et al. (2017) A Role for Neuronal Alpha-Synuclein in Gastrointestinal Immunity. *Journal of Innate Immunity* 9(5):456–463.

[347] Di Domizio J et al. (2012) Nucleic acid-containing amyloid fibrils potently induce type I interferon and stimulate systemic autoimmunity. *Proceedings of the National Academy of Sciences of the United States of America* 109(36):14550–14555.

[348] Means TK et al. (2005) Human lupus autoantibody–DNA complexes activate DCs through cooperation of CD32 and TLR9. *Journal of Clinical Investigation* 115(2):407–417.

[349] Schmidt NW et al. (2015) Liquid-crystalline ordering of antimicrobial peptide-DNA complexes controls TLR9 activation. *Nature materials* 14(7):696–700.

[350] Lee EY et al. (2016) A review of immune amplification via ligand clustering by self-assembled liquid-crystalline DNA complexes. *Advances in colloid and interface science* 232:17–24.

[351] Alexopoulou L, Holt AC, Medzhitov R, Flavell RA (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* 413(6857):732–738.

[352] Adase CA et al. (2016) Non-coding double-stranded RNA and LL-37 induce growth factor expression from keratinocytes and endothelial cells. *J Biol Chem* 291(22):jbc.M116.725317–11646.

[353] Zhang Lj et al. (2016) Antimicrobial Peptide LL37 and MAVS Signaling Drive Interferon-β Production by Epidermal Keratinocytes during Skin Injury. *Immunity* 45(1):119–130.

[354] Nelson AM et al. (2015) dsRNA Released by Tissue Damage Activates TLR3 to Drive Skin Regeneration. *Cell Stem Cell* 17(2):139–151.

[355] Zhang Lj, Gallo RL (2016) Antimicrobial peptides. *Current biology : CB* 26(1):R14–9.

[356] Büchau AS, Gallo RL (2007) Innate immunity and antimicrobial defense systems in psoriasis. *Clinics in dermatology* 25(6):616–624.

[357] Lai Y et al. (2011) LL37 and Cationic Peptides Enhance TLR3 Signaling by Viral Double-stranded RNAs. *PLoS ONE* 6(10):e26632.

[358] Chen X et al. (2013) Human antimicrobial peptide LL-37 modulates proinflammatory responses induced by cytokine milieus and double-stranded RNA in human keratinocytes. *Biochemical and Biophysical Research Communications* 433(4):532–537.

[359] Chen X et al. (2013) Modulation of double-stranded RNA- and cytokine- induced responses of human keratinocytes by LL-37. *Journal of Dermatological Science* 69(2):e14.

[360] Yang Y et al. (2016) The RNA-binding protein Mex3B is a coreceptor of Toll-like receptor 3 in innate antiviral response. *Cell research* 26(3):288–303.

[361] Lai Y et al. (2011) Viral double-strand RNA-binding proteins can enhance innate immune signaling by toll-like Receptor 3. *PLoS ONE* 6(10):e25837.

[362] Lee EY et al. (2017) 070 Liquid crystalline ordering of antimicrobial peptide-RNA complexes controls TLR3 activation. *Journal of Investigative Dermatology* 137(5):S12.

[363] Krüger A et al. (2015) Human TLR8 senses UR/URR motifs in bacterial and mitochondrial RNA. *EMBO reports* 16(12):1656–1663.

[364] Vollmer J et al. (2005) Immune stimulation mediated by autoantigen binding sites within small nuclear RNAs involves Toll-like receptors 7 and 8. *Journal of Experimental Medicine* 202(11):1575–1585.

[365] Marshak-Rothstein A, Rifkin IR (2007) Immunologically active autoantigens: the role of toll-like receptors in the development of chronic inflammatory disease. *Annual review of immunology* 25:419–441.

[366] Ganguly D et al. (2009) Self-RNA-antimicrobial peptide complexes activate human dendritic cells through TLR7 and TLR8. *The Journal of experimental medicine* 206(9):1983–1994.

[367] Prinz N et al. (2011) Antiphospholipid antibodies induce translocation of TLR7 and TLR8 to the endosome in human monocytes and plasmacytoid dendritic cells. *Blood* 118(8):2322–2332.

[368] Sacre SM et al. (2008) Inhibitors of TLR8 Reduce TNF Production from Human Rheumatoid Synovial Membrane Cultures. *Journal of Immunology* 181(11):8002–8009.

[369] Guiducci C et al. (2013) RNA recognition by human TLR8 can lead to autoimmune inflammation. *The Journal of experimental medicine* 210(13):2903–2919.

[370] Jiang W et al. (2013) A Toll-like receptor 7, 8, and 9 antagonist inhibits Th1 and Th17 responses and inflammasome activation in a model of IL-23-induced psoriasis. *The Journal of investigative dermatology* 133(7):1777–1784.

[371] Zhu FG et al. (2013) A novel antagonist of Toll-like receptors 7, 8 and 9 suppresses lupus disease-associated parameters in NZBW/F1 mice. *Autoimmunity* 46(7):419–428.

[372] Yamasaki K et al. (2007) Increased serine protease activity and cathelicidin promotes skin inflammation in rosacea. *Nature medicine* 13(8):975–980.

[373] Paulsen F et al. (2002) Antimicrobial peptides are expressed and produced in healthy and inflamed human synovial membranes. *The Journal of pathology* 198(3):369–377.

[374] Schauber Jr et al. (2006) Heterogeneous expression of human cathelicidin hCAP18/LL-37 in inflammatory bowel diseases. *European Journal of Gastroenterology & Hepatology* 18(6):615–621.

[375] Kim ST et al. (2003) Antimicrobial peptide LL-37 is upregulated in chronic nasal inflammatory disease. *Acta oto-laryngologica* 123(1):81–85.

[376] Agerberth B et al. (1999) Antibacterial components in bronchoalveolar lavage fluid from healthy individuals and sarcoidosis patients. *American journal of respiratory and critical care medicine* 160(1):283–290.

[377] Bucki R, Byfield FJ, Janmey PA (2007) Release of the antimicrobial peptide LL-37 from DNA/F-actin bundles in cystic fibrosis sputum. *The European respiratory journal* 29(4):624–632.

[378] Lee EY et al. (2017) Crystallinity of Double-Stranded RNA-Antimicrobial Peptide Complexes Modulates Toll-Like Receptor 3-Mediated Inflammation. *ACS Nano* 11(12):12145–12155.

[379] Draper DE, Grilley D, Soto AM (2005) Ions and RNA folding. *Annual review of biophysics and biomolecular structure* 34(1):221–243.

[380] Anderson CF, RECORD MT (1990) Ion Distributions Around Dna and Other Cylindrical Polyions - Theoretical Descriptions and Physical Implications. *Annual Review of Biophysics and Biophysical Chemistry* 19(1):423–465.

[381] Sharp KA, Honig B (1990) Electrostatic Interactions in Macromolecules - Theory and Applications. *Annual Review of Biophysics and Biophysical Chemistry* 19(1):301–332.

[382] Odijk T (2004) Statics and dynamics of condensed DNA within phages and globules. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 362(1820):1497–1517.

[383] Kindt J, Tzlil S, Ben-Shaul A, Gelbart WM (2001) DNA packaging and ejection forces in bacteriophage. *Proceedings of the National Academy of Sciences* 98(24):13671–13674.

[384] Reich Z, Wachtel E, Wachtel EJ, Minsky A (1994) Liquid-crystalline mesophases of plasmid DNA in bacteria. *Science* 264(5164):1460–1463.

[385] Englander J et al. (2004) DNA toroids: framework for DNA repair in Deinococcus radiodurans and in germinating bacterial spores. *Journal of Bacteriology* 186(18):5973–5977.

[386] Guldbrand L, Nilsson LG, Nordenskiöld L (1986) A Monte Carlo simulation study of electrostatic forces between hexagonally packed DNA double helices. *The Journal of Chemical Physics* 85(11):6686–6698.

[387] Barbosa MC, Deserno M, Holm C (2000) A stable local density functional approach to ion-ion correlations. *Europhysics Letters (EPL)* 52(1):80–86.

[388] Ray J, Manning GS (1994) An attractive force between two rodlike polyions mediated by the sharing of condensed counterions. *Langmuir* 10(7):2450–2461.

[389] Schiessel H, Pincus P (1998) Counterion-Condensation-Induced Collapse of Highly Charged Polyelectrolytes. *Macromolecules* 31(22):7953–7959.

[390] Solis FJ, de la Cruz MO (2000) Collapse of flexible polyelectrolytes in multivalent salt solutions. *The Journal of Chemical Physics* 112(4):2030–2035.

[391] Lamm G, Wong L, Pack GR (1994) Monte Carlo and Poisson–Boltzmann calculations of the fraction of counterions bound to DNA. *Biopolymers* 34(2):227–237.

[392] MacKerell Jr. AD, Nilsson L (2008) Molecular dynamics simulations of nucleic acid–protein complexes. *Current opinion in structural biology* 18(2):194–199.

[393] Oosawa F (1968) Interaction between parallel rodlike macroions. *Biopolymers* 6(11):1633–1647.

[394] Golestanian R, Kardar M, Liverpool TB (1999) Collapse of Stiff Polyelectrolytes due to Counterion Fluctuations. *Physical Review Letters* 82(22):4456–4459.

[395] Ha BY, Liu AJ (1997) Counterion-Mediated Attraction between Two Like-Charged Rods. *Physical Review Letters* 79(7):1289–1292.

[396] de la Cruz MO et al. (1995) Precipitation of highly charged polyelectrolyte solutions in the presence of multivalent salts. *The Journal of Chemical Physics* 103(13):5781–5791.

[397] Tan ZJ, Chen SJ (2006) Ion-Mediated Nucleic Acid Helix-Helix Interactions. *Biophysical Journal* 91(2):518–536.

[398] Travesset A, Vaknin D (2006) Bjerrum pairing correlations at charged interfaces. *Europhysics Letters (EPL)* 74(1):181–187.

[399] Rouzina I, Bloomfield VA (1996) Macroion Attraction Due to Electrostatic Correlation between Screening Counterions. 1. Mobile Surface-Adsorbed Ions and Diffuse Ion Cloud. *The Journal of Physical Chemistry* 100(23):9977–9989.

[400] Shklovskii BI (1999) Wigner Crystal Model of Counterion Induced Bundle Formation of Rodlike Polyelectrolytes. *Physical Review Letters* 82(16):3268–3271.

[401] Lau AWC, Levine D, Pincus P (2000) Novel Electrostatic Attraction from Plasmon Fluctuations. *Physical Review Letters* 84(18):4116–4119.

[402] Moreira AG, Netz RR (2007) Strong-coupling theory for counter-ion distributions. *Europhysics Letters (EPL)* 52(6):705–711.

[403] Moreira AG, Netz RR (2001) Binding of Similarly Charged Plates with Counterions Only. *Physical Review Letters* 87(7):078301.

[404] Naji A, Netz RR (2005) Counterions at Charged Cylinders: Criticality and Universality beyond Mean-Field Theory. *Physical Review Letters* 95(18):185703.

[405] Deserno M, Arnold A, Holm C (2003) Attraction and ionic correlations between charged stiff polyelectrolytes. *Macromolecules*.

[406] Naji A, Jungblut S, Moreira AG, Netz RR (2005) Electrostatic interactions in strongly coupled soft matter. *Physica A: Statistical Mechanics and its Applications* 352(1):131–170.

[407] Šamaj L, Trizac E (2011) Wigner-crystal formulation of strong-coupling theory for counterions near planar charged interfaces. *Physical Review E* 84(4):041401.

[408] Mohammad-Rafiee F, Golestanian R (2005) Elastic Correlations in Nucleosomal DNA Structure. *Physical Review Letters* 94(23):238102.

[409] Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311(5986):532–537.

[410] Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389(6648):251–260.

[411] Schiessel H (2006) The nucleosome: A transparent, slippery, sticky and yet stable DNA-protein complex. *The European Physical Journal E* 19(3):251–262.

[412] Chen Y et al. (2014) Revealing transient structures of nucleosomes as DNA unwinds. *Nucleic Acids Research* 42(13):8767–8776.

[413] DeRouchey J, Netz RR, Radler JO (2005) Structural investigations of DNA-polycation complexes. *The European Physical Journal E* 16(1):17–28.

[414] Boussif O et al. (1995) A Versatile Vector for Gene and Oligonucleotide Transfer Into Cells in Culture and in-Vivo - Polyethylenimine. *Proceedings of the National Academy of Sciences* 92(16):7297–7301.

[415] Kukowska-Latallo JF et al. (1996) Efficient transfer of genetic material into mammalian cells using Starburst polyamidoamine dendrimers. *Proceedings of the National Academy of Sciences* 93(10):4897–4902.

[416] Evans HM et al. (2003) Structural polymorphism of DNA-dendrimer complexes. *Physical Review Letters* 91(7):075501.

[417] Radler JO, Koltover I, Salditt T, Safinya CR (1997) Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes. *Science* 275(5301):810–814.

[418] Felgner PL et al. (1987) Lipofection - a Highly Efficient, Lipid-Mediated Dna-Transfection Procedure. *Proceedings of the National Academy of Sciences* 84(21):7413–7417.

[419] Felgner PL, Rhodes G (1991) Gene therapeutics. *Nature* 349(6307):351–352.

[420] Gustafsson J, Arvidson G, Karlsson G, Almgren M (1995) Complexes between cationic liposomes and DNA visualized by cryo-TEM. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1235(2):305–312.

[421] Sternberg B, Sorgi FL, Huang L (1994) New structures in complex formation between DNA and cationic liposomes visualized by freeze—fracture electron microscopy. *FEBS Letters* 356(2-3):361–366.

[422] Danilo D Lasic et al. (1997) The Structure of DNA-Liposome Complexes. *Journal of the American Chemical Society* 119(4):832–833.

[423] May S, BenShaul A (1997) DNA-lipid complexes: Stability of honeycomb-like and spaghetti-like structures. *Biophysical Journal* 73(5):2427–2440.

[424] Bruinsma R (1998) Electrostatics of DNA-cationic lipid complexes: isoelectric instability. *The European Physical Journal B - Condensed Matter and Complex Systems* 4(1):75–88.

[425] Harries D, May S, Gelbart WM, Ben-Shaul A (1998) Structure, Stability, and Thermodynamics of Lamellar DNA-Lipid Complexes. *Biophysical Journal* 75(1):159–173.

[426] Safinya C (2001) Structures of lipid–DNA complexes: supramolecular assembly and gene delivery. *Current opinion in structural biology* 11(4):440–448.

[427] Kai K Ewert et al. (2006) *A Columnar Phase of Dendritic Lipid-Based Cationic Liposome-DNA Complexes for Gene Delivery: Hexagonally Ordered Cylindrical Micelles Embedded in a DNA Honeycomb Lattice.* (American Chemical Society) Vol. 128.

[428] Wong GCL et al. (2003) Lamellar phase of stacked two-dimensional rafts of actin filaments. *Physical Review Letters* 91(1):018103.

[429] Wong GCL et al. (2000) Hierarchical self-assembly of F-actin and cationic lipid complexes: Stacked three-layer tubule networks. *Science* 288(5473):2035–+.

[430] Needleman DJ et al. (2004) Higher-order assembly of microtubules by counterions: from hexagonal bundles to living necklaces. *Proceedings of the National Academy of Sciences of the United States of America* 101(46):16099–16103.

[431] Sanders LK et al. (2007) Control of electrostatic interactions between F-actin and genetically modified lysozyme in aqueous media. *Proceedings of the National Academy of Sciences* 104(41):15994–15999.

[432] Sanders LK et al. (2005) Structure and stability of self-assembled actin-lysozyme complexes in salty water. *Physical Review Letters* 95(10):108302.

[433] Snapper CM, Kehry MR, Castle BE, Mond JJ (1995) Multivalent, but not divalent, antigen receptor cross-linkers synergize with CD40 ligand for induction of Ig synthesis and class switching in normal murine B cells. A redefinition of the TI-2 vs T cell-dependent antigen dichotomy. *Journal of Immunology* 154(3):1177–1187.

[434] Lesley J, Hascall VC, Tammi M, Hyman R (2000) Hyaluronan binding by cell surface CD44. *J Biol Chem* 275(35):26967–26975.

[435] Puré E, Cuff CA (2001) A crucial role for CD44 in inflammation. *Trends in Molecular Medicine* 7(5):213–221.

[436] Toole BP (2009) Hyaluronan-CD44 Interactions in Cancer: Paradoxes and Possibilities. *Clinical cancer research : an official journal of the American Association for Cancer Research* 15(24):7462–7468.

[437] Weissleder R, Kelly K, Sun EY, Shtatland T, Josephson L (2005) Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology* 23(11):1418–1423.

[438] Dubacheva GV, Curk T, Auzély-Velty R, Frenkel D, Richter RP (2015) Designing multivalent probes for tunable superselective targeting. *Proceedings of the National Academy of Sciences of the United States of America* 112(18):5579–5584.

[439] Albertazzi L et al. (2013) Spatiotemporal control and superselectivity in supramolecular polymers using multivalency. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12203–12208.

[440] Martinez-Veracoechea FJ, Martinez-Veracoechea FJ, Frenkel D (2011) Designing super selectivity in multivalent nano-particle binding. *Proceedings of the National Academy of Sciences of the United States of America* 108(27):10963–10968.

[441] Tursi SA et al. (2017) Bacterial amyloid curli acts as a carrier for DNA to elicit an autoimmune response via TLR2 and TLR9. *PLoS Pathogens* 13(4):e1006315.

[442] Gallo PM et al. (2015) Amyloid-DNA Composites of Bacterial Biofilms Stimulate Autoimmunity. *Immunity* 42(6):1171–1184.

[443] Warren BE (1941) X-Ray Diffraction in Random Layer Lattices. *Physical Review* 59(9):693–698.

[444] von Herrath MG, Bot A (2003) Immune responsiveness, tolerance and dsRNA: implications for traditional paradigms. *Trends in immunology* 24(6):289–292.

[445] Miller LS (2008) Toll-like receptors in skin. *Advances in dermatology* 24:71–87.

[446] Grossman RM et al. (1989) Interleukin-6 Is Expressed in High-Levels in Psoriatic Skin and Stimulates Proliferation of Cultured Human Keratinocytes. *Proceedings of the National Academy of Sciences* 86(16):6367–6371.

[447] Goodman WA et al. (2009) IL-6 signaling in psoriasis prevents immune suppression by regulatory T cells. *Journal of immunology (Baltimore, Md. : 1950)* 183(5):3170–3176.

[448] Endoh T, Ohtsuki T (2009) Cellular siRNA delivery using cell-penetrating peptides modified for endosomal escape. *Advanced drug delivery reviews* 61(9):704–709.

[449] Nathan F Bouxsein et al. (2007) Structure and Gene Silencing Activities of Monovalent and Pentavalent Cationic Lipid Vectors Complexed with siRNA†. *Biochemistry* 46(16):4785–4792.

[450] Singha K, Namgung R, Kim WJ (2011) Polymers in small-interfering RNA delivery. *Nucleic acid therapeutics* 21(3):133–147.

[451] Wong GCL, Pollack L (2010) Electrostatics of strongly charged biological polymers: ion-mediated interactions and self-organization in nucleic acids and proteins. *Annual review of physical chemistry* 61(1):171–189.

[452] Pollack L (2011) SAXS studies of ion-nucleic acid interactions. *Annual Review of Biophysics* 40(1):225–242.

[453] Livolant F, Leforestier A (1996) Condensed phases of DNA: Structures and phase transitions. *Progress in polymer science* 21(6):1115–1164.

[454] Rau DC, Parsegian VA (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. *Biophysical Journal* 61(1):246–259.

[455] Brooks H, Lebleu B, Vivès E (2005) Tat peptide-mediated cellular delivery: back to basics. *Advanced drug delivery reviews* 57(4):559–577.

[456] Feuerstein BG, Williams LD, Basu HS, Marton LJ (1991) Implications and concepts of polyamine-nucleic acid interactions. *Journal of Cellular Biochemistry* 46(1):37–47.

[457] Tabor CW, Tabor H (1984) Polyamines. *Annual review of biochemistry.*

[458] Moinard C, Cynober L, de Bandt J (2005) Polyamines: metabolism and implications in human diseases. *Clinical Nutrition* 24(2):184–197.

[459] Morizane S, Gallo RL (2012) Antimicrobial peptides in the pathogenesis of psoriasis. *The Journal of Dermatology* 39(3):225–230.

[460] Yamasaki K et al. (2006) Kallikrein-mediated proteolysis regulates the antimicrobial effects of cathelicidins in skin. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 20(12):2068–2080.

[461] Murakami M, Lopez-Garcia B, Braff M, Dorschner RA, Gallo RL (2004) Postsecretory processing generates multiple cathelicidins for enhanced topical antimicrobial defense. *Journal of Immunology* 172(5):3070–3077.

[462] Wong GCL (2006) Electrostatics of rigid polyelectrolytes. *Current Opinion in Colloid & Interface Science* 11(6):310–315.

[463] Székely GJ, Rizzo ML (2013) Energy statistics: A class of statistics based on distances. *Journal of statistical planning and inference* 143(8):1249–1272.

[464] Dubacheva GV et al. (2014) Superselective Targeting Using Multivalent Polymers. *Journal of the American Chemical Society* 136(5):1722–1725.

[465] Borkowski AW, Park K, Uchida Y, Gallo RL (2013) Activation of TLR3 in keratinocytes increases expression of genes involved in formation of the epidermis, lipid accumulation, and epidermal organelles. *The Journal of investigative dermatology* 133(8):2031–2040.

[466] Brinkmann V et al. (2004) Neutrophil extracellular traps kill bacteria. *Science* 303(5663):1532–1535.

[467] Rock KL, Kono H (2008) The Inflammatory Response to Cell Death. *Annual Review of Pathology: Mechanisms of Disease* 3(1):99–126.

[468] Brinkmann V, Zychlinsky A (2007) Beneficial suicide: why neutrophils die to make NETs. *Nature reviews. Microbiology* 5(8):577–582.

[469] Brinkmann V, Zychlinsky A (2012) Neutrophil extracellular traps: is immunity the second function of chromatin? *The Journal of Cell Biology* 198(5):773–783.

[470] Halverson TWR, Wilton M, Poon KKH, Petri B, Lewenza S (2015) DNA is an antimicrobial component of neutrophil extracellular traps. *PLoS Pathogens* 11(1):e1004593.

[471] Knight JS, Carmona-Rivera C, Kaplan MJ (2012) Proteins derived from neutrophil extracellular traps may serve as self-antigens and mediate organ damage in autoimmune diseases. *Frontiers in Immunology* 3:380.

[472] Hu SCS et al. (2016) Neutrophil extracellular trap formation is increased in psoriasis and induces human β-defensin-2 production in epidermal keratinocytes. *Scientific Reports* 6(1):31119.

[473] Hakkim A et al. (2010) Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proceedings of the National Academy of Sciences of the United States of America* 107(21):9813–9818.

[474] Darrah E, Andrade F (2012) NETs: the missing link between cell death and systemic autoimmune diseases? *Frontiers in Immunology* 3:428.

[475] Zhivotovsky B, Orrenius S (2011) Calcium and cell death mechanisms: A perspective from the cell death community. *Cell calcium* 50(3):211–221.

[476] Chen R, Kang R, Fan XG, Tang D (2014) Release and activity of histone in diseases. *Cell Death and Disease* 5(8):e1370.

[477] Huang H et al. (2011) Endogenous histones function as alarmins in sterile inflammatory liver injury through Toll-like receptor 9 in mice. *Hepatology* 54(3):999–1008.

[478] Kim JE, Lee N, Gu JY, Yoo HJ, Kim HK (2015) Circulating levels of DNA-histone complex and dsDNA are independent prognostic factors of disseminated intravascular coagulation. *Thrombosis research* 135(6):1064–1069.

[479] Rumore PM, Steinman CR (1990) Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone. *Journal of Clinical Investigation* 86(1):69–74.

[480] Camicia G, Pozner R, de Larrañaga G (2014) Neutrophil Extracellular Traps in Sepsis. *Shock* 42(4):286–294.

[481] Holdenrieder S et al. (2008) Clinical Relevance of Circulating Nucleosomes in Cancer. *Annals of the New York Academy of Sciences* 1137(1):180–189.

[482] Borissoff JI et al. (2013) Elevated levels of circulating DNA and chromatin are independently associated with severe coronary atherosclerosis and a prothrombotic state. *Arteriosclerosis, thrombosis, and vascular biology* 33(8):2032–2040.

[483] Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* 423(6936):145–150.

[484] Widom J, Klug A (1985) Structure of the 3000Å chromatin filament: X-ray diffraction from oriented samples. *Cell* 43(1):207–213.

[485] Berezhnoy NV et al. (2012) Supramolecular organization in self-assembly of chromatin and cationic lipid bilayers is controlled by membrane charge density. *Biomacromolecules* 13(12):4146–4157.

[486] Kristián T, Siesjö BK (1998) Calcium in ischemic cell death. *Stroke* 29(3):705–718.

[487] Leforestier A, Livolant F (1997) Liquid crystalline ordering of nucleosome core particles under macromolecular crowding conditions: evidence for a discotic columnar hexagonal phase. *Biophysical Journal* 73(4):1771–1776.

[488] Livolant F et al. (2006) Are liquid crystalline properties of nucleosomes involved in chromosome structure and dynamics? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 364(1847):2615–2633.

[489] Berezhnoy NV et al. (2016) The Influence of Ionic Environment and Histone Tails on Columnar Order of Nucleosome Core Particles. *Biophysical Journal* 110(8):1720–1731.

[490] Gupta AK, Giaglis S, Hasler P, Hahn S (2014) Efficient Neutrophil Extracellular Trap Induction Requires Mobilization of Both Intracellular and Extracellular Calcium Pools and Is Modulated by Cyclosporine A. *PLoS ONE* 9(5):e97088.

[491] Davey CA, Richmond TJ (2002) DNA-dependent divalent cation binding in the nucleosome core particle. *Proceedings of the National Academy of Sciences* 99(17):11169–11174.

[492] Leforestier A, Lemercier N, Livolant F (2016) *Local order and conformation of nucleosomes in solution and in interphase nuclei analysed by cryo-electron microscopy of vitreous sections*. (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany) Vol. 5.

[493] Vaĭnshteĭn BK (1966) *Diffraction of X-rays by chain molecules*. (Elsevier).

[494] Salzet M (2002) Antimicrobial peptides are signaling molecules. *Trends in immunology* 23(6):283–284.

[495] Di Nardo A et al. (2007) Cathelicidin Antimicrobial Peptides Block Dendritic Cell TLR4 Activation and Allergic Contact Sensitization. *Journal of Immunology* 178(3):1829–1834.

[496] Veldhuizen EJA et al. (2017) Antimicrobial and Immunomodulatory Activity of PMAP-23 Derived Peptides. *Protein and Peptide Letters* 24(7):609–616.

[497] Porcelli F et al. (2008) NMR Structure of the Cathelicidin-Derived Human Antimicrobial Peptide LL-37 in Dodecylphosphocholine Micelles. *Biochemistry* 47(20):5565–5572.

[498] Wilman HR, Shi J, Deane CM (2014) Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins: Structure, Function, and Bioinformatics* 82(9):1960–1970.

[499] Deville J, Rey J, Chabbert M (2008) Comprehensive analysis of the helix-X-helix motif in soluble proteins. *Proteins: Structure, Function, and Bioinformatics* 72(1):115–135.

[500] Jacob J, Duclohier H, Cafiso DS (1999) The role of proline and glycine in determining the backbone flexibility of a channel-forming peptide. *Biophysical Journal* 76(3):1367–1376.

[501] Allahyarov E, Gompper G, Löwen H (2004) Attraction between DNA molecules mediated by multivalent ions. *Physical Review E* 69(4):041904–13.

[502] Qiu X, Andresen K, Lamb JS, Kwok LW, Pollack L (2008) Abrupt Transition from a Free, Repulsive to a Condensed, Attractive DNA Phase, Induced by Multivalent Polyamine Cations. *Physical Review Letters* 101(22):228101–4.

[503] Ziebarth J, Wang Y (2009) Molecular dynamics simulations of DNA-polycation complex formation. *Biophysical Journal* 97(7):1971–1983.

[504] Prévost S et al. (2011) Colloidal Structure and Stability of DNA/Polycations Polyplexes Investigated by Small Angle Scattering. *Biomacromolecules* 12(12):4272–4282.

[505] Su CJ et al. (2009) Columnar mesophases of the complexes of DNA with low-generation poly(amido amine) dendrimers. *Biomacromolecules* 10(4):773–783.

[506] Svergun D, Barberato C, Koch MHJ, IUCr (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* 28(6):768–773.

[507] Lee MT, Sun TL, Hung WC, Huang HW (2013) Process of inducing pores in membranes by melittin. *Proceedings of the National Academy of Sciences of the United States of America* 110(35):14243–14248.

[508] Rozengurt E, Gelehrter TD, Legg A, Pettican P (1981) Melittin stimulates Na entry, Na-K pump activity and DNA synthesis in quiescent cultures of mouse cells. *Cell* 23(3):781–788.

[509] Chang-Po Chen et al. (2006) Gene Transfer with Poly-Melittin Peptides. *Bioconjugate chemistry* 17(4):1057–1062.

[510] Terwilliger TC, Eisenberg D (1982) The structure of melittin. II. Interpretation of the structure. *Journal of Biological Chemistry* 257(11):6016–6022.

[511] Adhireksan Z et al. (2014) Ligand substitutions between ruthenium-cymene compounds can control protein versus DNA targeting and anticancer activity. *Nature communications* 5:3462.

[512] Yi GS, Park CB, Kim SC, Cheong C (1996) Solution structure of an antimicrobial peptide buforin II. *FEBS Letters* 398(1):87–90.

[513] Xie Y, Fleming E, Chen JL, Elmore DE (2011) Effect of proline position on the antimicrobial mechanism of buforin II. *Peptides* 32(4):677–682.

[514] Kobayashi S et al. (2004) Membrane translocation mechanism of the antimicrobial peptide buforin 2. *Biochemistry* 43(49):15610–15616.

[515] Minn I, Kim HS, Kim SC (1998) Antimicrobial peptides derived from pepsinogens in the stomach of the bullfrog, Rana catesbeiana. *Biochimica et biophysica acta* 1407(1):31–39.

[516] Urbán E, Nagy E, Pál T, Sonnevend A, Conlon JM (2007) Activities of four frog skin-derived antimicrobial peptides (temporin-1DRa, temporin-1Va and the melittin-related peptides AR-23 and RV-23) against anaerobic bacteria. *International journal of antimicrobial agents* 29(3):317–321.

[517] Zhang SK et al. (2016) RV-23, a Melittin-Related Peptide with Cell-Selective Antibacterial Activity and High Hemocompatibility. *Journal of microbiology and biotechnology* 26(6):1046–1056.

[518] Lee MW et al. (2017) Interactions between Membranes and "Metaphilic" Polypeptide Architectures with Diverse Side-Chain Populations. *ACS Nano* 11(3):2858–2871.

[519] Kaplan A et al. (2017) Direct Antimicrobial Activity of IFN-β. *Journal of immunology (Baltimore, Md. : 1950)* 198(10):4036–4045.

[520] McWhirter SM et al. (2009) A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *Journal of Experimental Medicine* 206(9):1899–1911.

[521] Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*.

[522] Duan Y et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry* 24(16):1999–2012.

[523] Jenssen H, Hancock REW (2010) Therapeutic potential of HDPs as immunomodulatory agents. *Methods in molecular biology (Clifton, N.J.)* 618(Chapter 20):329–347.

[524] Easton DM, Nijnik A, Mayer ML, Hancock REW (2009) Potential of immunomodulatory host defense peptides as novel anti-infectives. *Trends in biotechnology* 27(10):582–590.

[525] Nijnik A, Hancock R (2009) Host defence peptides: antimicrobial and immunomodulatory activity and potential applications for tackling antibiotic-resistant infections. *Emerging health threats journal* 2:e1.

[526] Haney EF, Hancock REW (2013) Peptide design for antimicrobial and immunomodulatory applications. *Peptide Science* 100(6):572–583.

[527] Nicholls EF, Madera L, Hancock REW (2010) Immunomodulators as adjuvants for vaccines and antimicrobial therapy. *Annals of the New York Academy of Sciences* 1213(1):46–61.

[528] Takahashi T et al. (2018) Cathelicidin promotes inflammation by enabling binding of self-RNA to cell surface scavenger receptors. *Scientific Reports* 8(1):4032.

[529] Chi Z, Wang Z, Wang K, Zhu Y, Qin S (2014) Cathelicidin Antimicrobial Peptide LL-37 in Cholesteatoma Enables Keratinocyte Reactivity with Cytosolic DNA. *Scandinavian Journal of Immunology* 79(3):214–221.

[530] Chamilos G et al. (2012) Cytosolic sensing of extracellular self-DNA transported into monocytes by the antimicrobial peptide LL37. *Blood* 120(18):3699–3707.

[531] Yount NY, Yeaman MR (2006) Structural congruence among membrane-active host defense polypeptides of diverse phylogeny. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1758(9):1373–1386.

[532] Yount NY, Yeaman MR (2004) Multidimensional signatures in antimicrobial peptides. *Proceedings of the National Academy of Sciences* 101(19):7363–7368.

[533] Yount NY et al. (2009) Selective reciprocity in antimicrobial activity versus cytotoxicity of hBD-2 and crotamine. *Proceedings of the National Academy of Sciences of the United States of America* 106(35):14972–14977.

[534] Yeaman MR, Yount NY (2007) Unifying themes in host defence effector polypeptides. *Nature Reviews Microbiology* 5(9):727–740.

[535] Yeaman MR, Yount NY (2005) Code among chaos: Immunorelativity and the AEGIS model of antimicrobial peptides. *Asm News* 71(1):21–27.

[536] Yeaman MR et al. (2007) Modular determinants of antimicrobial activity in platelet factor-4 family kinocidins. *Biochimica et biophysica acta* 1768(3):609–619.

[537] Yang D et al. (2003) Many chemokines including CCL20/MIP-3alpha display antimicrobial activity. *Journal of Leukocyte Biology* 74(3):448–455.

[538] Yount NY et al. (2011) Context Mediates Antimicrobial Efficacy of Kinocidin Congener Peptide RP-1. *PLoS ONE* 6(11):e26727.

[539] Xiong YQ, Bayer AS, Elazegui L, Yeaman MR (2006) A synthetic congener modeled on a microbicidal domain of thrombin- induced platelet microbicidal protein 1 recapitulates staphylocidal mechanisms of the native molecule. *Antimicrob Agents Chemother* 50(11):3786–3792.

[540] Yeaman MR (2014) Platelets: at the nexus of antimicrobial defence. *Nature Reviews Microbiology* 12(6):426–437.

[541] Yount NY et al. (2004) Platelet microbicidal protein 1: structural themes of a multifunctional antimicrobial peptide. *Antimicrob Agents Chemother* 48(11):4395–4404.

[542] Yount NY et al. (2007) Structural correlates of antimicrobial efficacy in IL-8 and related human kinocidins. *Biochimica et biophysica acta* 1768(3):598–608.

[543] Meller S et al. (2015) T(H)17 cells promote microbial killing and innate immune sensing of DNA via interleukin 26. *Nature Immunology* 16(9):970–979.

[544] Lee JS, Cua DJ (2015) IL-26 AMPs up the TH17 arsenal. *Nature Immunology* 16(9):897–898.

[545] Poli C et al. (2017) IL-26 Confers Proinflammatory Properties to Extracellular DNA. *Journal of immunology (Baltimore, Md. : 1950)* 198(9):1600594–3661.

[546] Nordahl EA et al. (2004) Activation of the complement system generates antibacterial peptides. *Proceedings of the National Academy of Sciences* 101(48):16879–16884.

[547] Ruan BH et al. (2010) Complement C3a, CpG oligos, and DNA/C3a complex stimulate IFN-α production in a receptor for advanced glycation end product-dependent manner. *Journal of immunology (Baltimore, Md. : 1950)* 185(7):4213–4222.

[548] Realegeno S et al. (2016) S100A12 Is Part of the Antimicrobial Network against Mycobacterium leprae in Human Macrophages. *PLoS Pathogens* 12(6):e1005705.

[549] Yang Z et al. (2001) Proinflammatory properties of the human S100 protein S100A12. *Journal of Leukocyte Biology* 69(6):986–994.

[550] Lee EY, Lee MW, Wong GCL (2018) Modulation of Toll-like receptor signaling by antimicrobial peptides. *Seminars in cell & developmental biology* p. DOI: 10.1016/j.semcdb.2018.02.002.